GENERATION OF 3D-CITY MODELS AND THEIR UTILISATION IN IMAGE SEQUENCES

Uwe STILLA, Uwe SOERGEL, Klaus JAEGER
FGAN-FOM Research Institute for Optronics and Pattern Recognition
D 76275 Ettlingen, Germany
{usti, soe,jae}@fom.fgan.de

Working Group III/3

KEY WORDS: Image sequence, GIS, Data fusion, Buildings, Navigation.

ABSTRACT

In this paper we describe the construction of a city model and its support for the analysis of image sequences. Our city model consists of building models which were generated from large scale vector maps and laser altimeter data. First the vector map is analysed to group the outlines of buildings and to obtain a hierarchical description of buildings or building complexes. The base area of single buildings are used to mask the corresponding elevation data. Depending on the task prismatic or polyhedral object models are reconstructed from the masked elevation data. The interpretation of image sequences taken by an airborne sensor in oblique view can be supported by a 3D city model. Possible GIS applications could be automated overlaying of selected buildings or query detailed building information from a database by interactively pointing on a frame in a sequence. The projection parameters of the model data are derived from GPS and INS. In practice the projected building contours do not exactly coincide with their image location. To overcome this problem an automated matching approach of image and model description is required. After image and vector map analysis correspondences between both scene descriptions can be found. These correspondences are used to correct the navigation data. This approach can easily be extended to image sequences. The corrected navigation data of a frame can be used as prediction for subsequent frames.

1 INTRODUCTION

Three-dimensional city models find more and more interest in city and regional planning (Danahy, 1997). They are used for visualisation (Gruen, 1998)(Gruber et al., 1997), e.g. to demonstrate the influence of a planned building to the surrounding townscape. Furthermore there is a great demand for such models in civil and military mission planning, disaster management (Kakumoto et al., 1997) and as basis for simulation e.g. in the fields of environmental engineering for microclimate investigations (Adrian & Fiedler, 1991) or telecommunications for transmitter placement (Kürner et al., 1993).

However, detailed 3D-descriptions may be used as well for interpretation of scenes by image sequences, which were taken by different sensors or with different views. Knowing the sensor geometry, the position and orientation of the platform from Global Positioning System (GPS) and Inertial Navigation System (INS) measurements, visible structures of the city model can be calculated and projected into the image sequence. This can be used for GIS aided interpretation, offline or online. Such a GIS aided interpretation of ground-based video sequences is described by Shibasaki [1998]. We focus on airborne taken imagery.

First, we present a simple approach for the automatic generation of 3D city models (Chapter 2). In Chapter 3 two possible applications for utilisation of a city models in image sequences taken by airborne sensors are proposed. Due to incorrect navigation data in practice the projected model contours do not exactly coincide with their image location.. To overcome this problem an automated matching approach of image and model data is described in Chapter 4. Finally, the results are discussed and an extended utilisation is proposed in Chapter 5

2 GENERATION OF 3D-CITY MODELS

The manual construction and update of 3D building models is time consuming and expensive. That is why some authors propose approaches to automate the process of exploiting scene data. In contrast to semi-automatic approaches we pursue approaches which allow a fully automatic reconstruction of buildings. For this reconstruction detailed elevation data are required which typically are derived from stereo image matching or laser scanning. An approach to derive building models with simple roof structures is e.g. proposed in Weidner & Förstner [1995]. The combined exploitation of laser elevation data and maps is shown in Haala & Brenner [1997]. In our approach we also combine digital vector maps and laser altimeter data to extract building models.

2.1 Digital Vector Maps

We use a large scale vector map which is organised in several layers, each of which contains a different class of objects (e.g. streets, buildings, etc.). Fig. 1a show a section of the scene, the large scale raster map (1:5000) (Fig. 1b) and the building layer of the vector map (Fig. 1c). The topological properties connectivity, closedness, and containment of the non-ordered set of map-lines are tested by a production net of a generic model [Stilla & Michaelsen, 1997]. The aim of the analysis is to separate parts of buildings, determine encapsulated areas and group parts of buildings. Output of the analysis is a hierarchical description of the buildings or complexes of buildings.

2.2 Laser altimeter data

Nowadays elevation data are commercially available from airborne laser scanners. Knowing the precise position and orientation of the airborne platform from differential Ground Positioning System (dGPS) and Inertial Navigation System (INS) measurements, the geographic position of the surface points in three spatial dimensions can be calculated to decimeter accuracy (Huising & Pereira, 1998). With current systems, points can be measured at approximately one point each 0.5 x 0.5 m^2 (Lohr, 1998). The sampled surface points distributed over a strip of 250-500m width allows the generation of a geocoded 2D array with elevation data in each cell (elevation image). Single flight strips are merged to a consistent digital surface model (DSM) of the whole survey area. A simple way to visualise the elevation data is to assign a brightness value to the z-coordinate of each raster element. Fig. 1d shows a section of an elevation image

2.3 Building models

The result of the map analysis step is used to mask the elevation data (Fig. 1f). Combining the brightness (Fig. 1f) with the z-coordinate in a 3D-view (Fig. 1g) leads to a plastic appearance of raster data and reveals roof details. For each building object of the map, a coarse 3D-description is constructed by a prismatic model (Fig. 1h). Depending on the task, the height of the covering plane can be calculated from the histogram using the (i) mean, (ii) median (for suppression of distortions), (iii) minimum (to get the height of the eaves) or (iv) maximum (to obtain the bounding box). Method (iii) was chosen for our purpose. The reconstruction of more complex roof structures is described in [Stilla & Jurkiewicz K, 1999]. For a visualisation the resulting wire-frame model is transformed into a surface model using an automatic triangulation. A perspective view of the city model (test area Karlsruhe) is shown in Fig. 2.

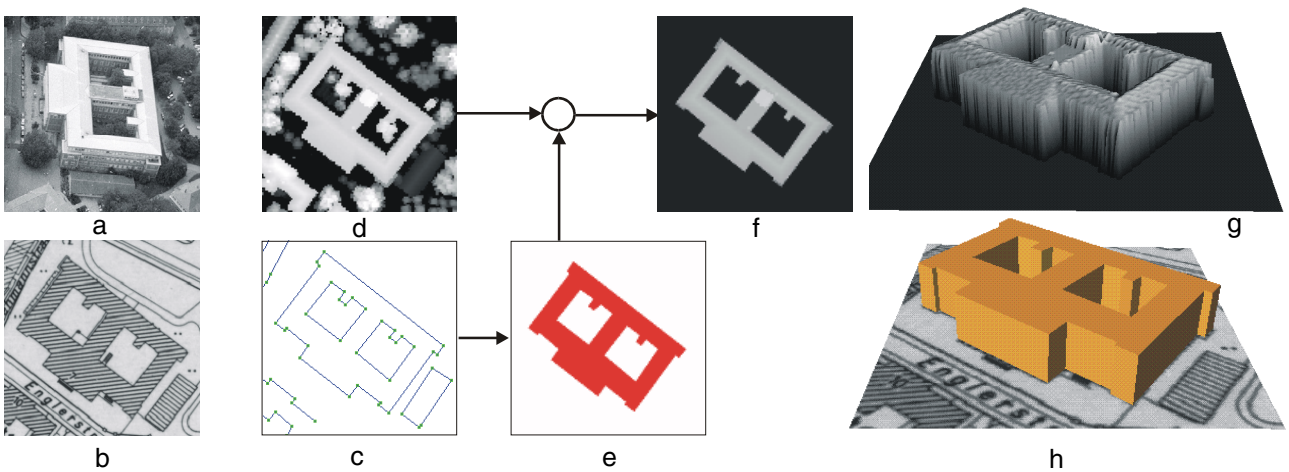


Fig. 1: Generation of prismatic objects from maps and elevation data

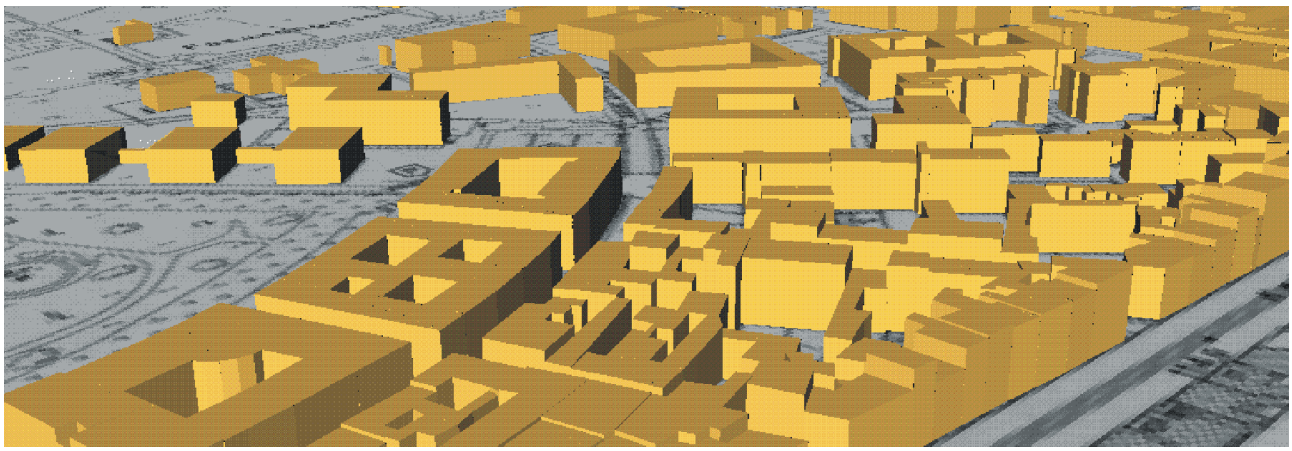


Fig. 2: City model of test area Karlsuhe

3 UTILIZATION OF 3D-CITY MODELS IN IMAGE SEQUENCES

In this paper we focus on video sequences of urban areas taken by an airborne sensor in oblique view (side looking). As test data video sequences by multiple flights over the campus of the university of Karlsruhe were taken by an infrared camera. The sequences differ in carrier elevation, flight direction and focal length of the camera. In Fig. 3 one test sequence containing 250 frames is shown. For the visualisation of the sequence a mosaic image is assembled from n columns taken from the centre of each frame. Parameter n is estimated from the carrier velocity. The geometric distortions in the right image part are obviously caused by a variation of the roll angle

For visual interpretation of image sequences (e.g. video) additional scene information may be helpful. This knowledge can be given in form of vector maps or 3D city models which are projected into the imagery. If image sequences are taken in ground-based or oblique views, then the incorporation of three-dimensional data is required. The correct projection of object contours and occlusions by other objects in the scene can not be handled properly using two-dimensional vector maps. Two possible applications are proposed.

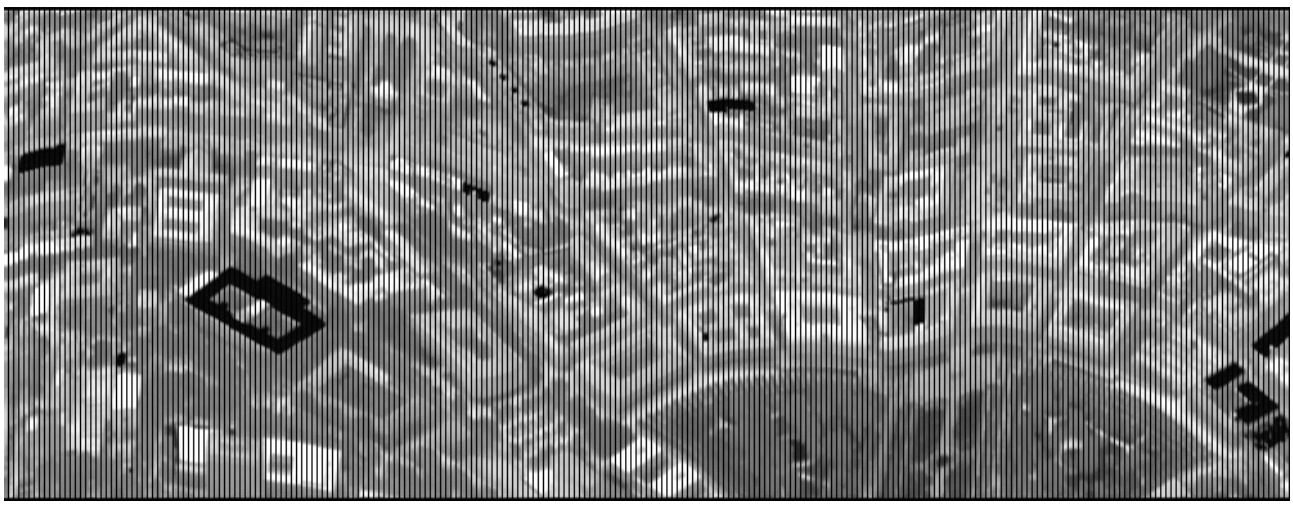


Fig. 3: Mosaicing of an infrared image sequence containing 250 frames with geometric distortions.

3.1 Interpretation of selected objects in image sequences

Human interpretation of video sequence taken from an airborne based sensor in oblique view is unfeasible in case of a small field of view, which makes the orientation in the scene difficult. The situation becomes worse in sequences with fast optical flow and unexpected carrier movements (e.g. rotation around roll axis). Additional problems occur if clouds partially occlude the scene. The interpreter can benefit by highlighting objects of interest for detailed analysis. In this application it is assumed, that the objects are selected a priori.

Fig. 4 shows some frames of an infrared image sequence tracking a selected object. To focus the attention of the interpreter to the expected location of the object, rays are drawn from the principal point of the image to four object vertices. The 3D vertices are selected using a bounding box oriented parallel to North-South direction. In frames not containing the selected object just the rays are drawn to inform about the direction in which the object is expected to appear (Fig. 4a). If the object is still far away the rays are almost parallel. Additionally, the object contours are highlighted (Fig. 4b-d).

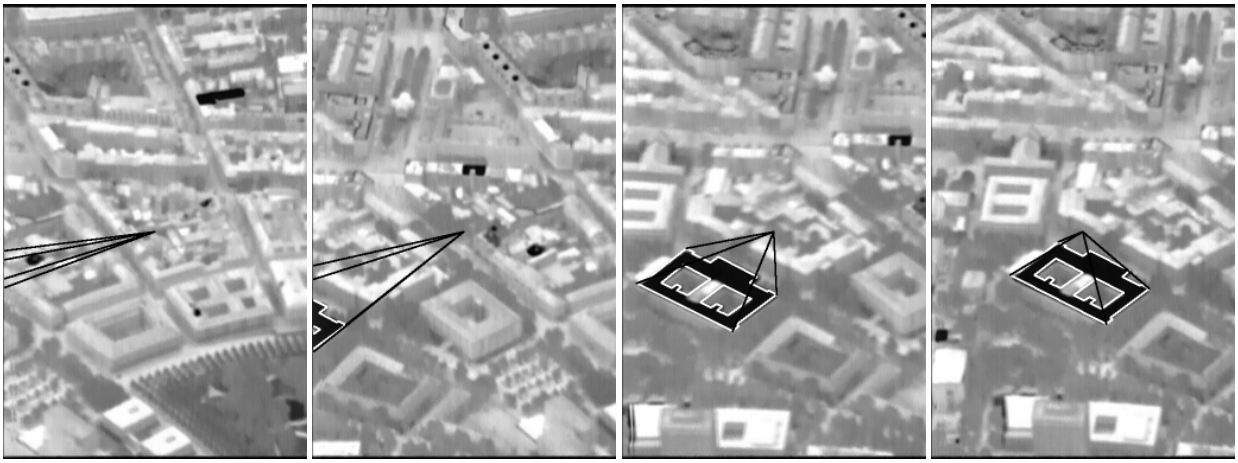


Fig. 4: Tracking of a selected object in an infrared image sequence

3.2 Interactive query of objects in image sequences

Another possible GIS application could be observing an image sequence without selection specific object a priori. Viewing an image sequence a prominent object or situation could attract the attention and produce an interest for more information. In case of buildings such additional information could for example refer to functional aspects (usage as residential building, industrial building, public building, etc.) or to geographical aspects (address, district etc.). A way to get this information is pointing interactively the cursor on the building of interest and starting a query by a mouse-click. For the query it has to be determined, which of the visible objects contains the selected position as surface point. An object surface can be bounded by several closed polygons. For example, the object in Fig. 2 highlighted contains three polygons. The stored attributes can be accessed by the object index.

IMAGE BASED POSE ESTIMATION OF THE SENSOR

The examined image data were taken from dense urban area in oblique view. Typical man made structures which are visible in this case are buildings with their roofs, in contrast to e.g. street crossings which are often occluded. Therefore, it is expected that vertices of roof structures are proper tie points for matching of image data and model data (city model). Using the given navigation data and camera parameters the 3D polygons of the model are projected into the image plane. Tie points are extracted independently in the image data and the projected model data by structural analysis methods. Afterwards, correspondences between the two tie point sets are searched. Corresponding pairs of tie points are used to recalculate the navigation data.

3.3 Tie points in image data

Line extraction. A variety of different approaches for edge detection are known from literature. We achieved promising results incorporating Burns edge detector [Burns et al., 1986]. Parameters of the algorithm are adjusted by camera parameters and analysis of the imagery. One parameter is the threshold th_m of the minimal required magnitude of the gradient to build a line. The threshold th_m is calculated from the mean μ and standard deviation σ_d of the magnitude of the gradient image: $th_m = \mu + 3\sigma_d$. Pixels of the magnitude image, which have a magnitude $m > th_m$ are edge pixel candidates. Another parameter is the minimal length l_{min} of a connected set of edge pixels in a direction. The parameter l_{min} is derived from camera parameters.

Selection. Two lines which fulfil two requirements built an angle structure CORNER. The distance d between the endpoints of the lines has to be $d < d_{min}$ and the value of the enclosing angle φ has to lay in-between the borders $\varphi_{min} <$

$\varphi < \varphi_{max}$. The vertices of the angle structures are calculated. If the vertices of different objects CORNER are close, the longest is kept and the other is rejected. The remaining objects CORNER are possible image tie points

To limit the computational load of the matching step, the overall number of image tie points has to be restricted to a maximum number N_T. The sum of sides of an angle is chosen as criteria for the selection of the N_T tie points. In case of short focal length the pinhole camera model is not necessarily valid anymore, because of geometric distortions especially found in the border areas of the image. Hence, only tie points inside a centred image window are considered. The size of the image window is adjusted with the focal length as parameter.

3.4 Tie points in model data

Suitable objects. City models may consist of different types of objects, e.g. buildings, streets, rivers, bridges, vegetation. Because of the oblique view we consider roof structures of buildings only. Depending on the expected sensor position, aspect angle, and field of view possible visible buildings are selected. In some cases a large amount of buildings would have to be considered, according to the field of view or the aspect. This may happen e.g. when the focal length is shortened by zoom out or the plane rotates around the roll axis and the horizon appears. Far distant objects are considered as unreliable. Furthermore, the computational effort of the subsequent process increases with the number of objects. Hence, far distant objects are clipped. In strong oblique views the problem of occlusion by other roofs and self-occlusion of roof parts is inherent. For the remaining objects a hidden line algorithm is applied.

Selection. In the hidden line step new vertices (endpoints of lines) are derived from the intersection of borders of different objects or object parts. These are treated in the same way as visible vertices of the object. From the set of projected lines objects corner are composed. The value of the enclosing angle φ has to lay in-between the borders $\varphi_{min} < \varphi < \varphi_{max}$ and the length l of each side has to be $l > l_{min}$. If the vertices of different objects corner are close, the longest is kept and the other is rejected. Analogous to the image the overall number of map tie points has to be restricted to n_t The remaining set of vertices selected as reference tie point set for the matching step.

3.5 Point Correspondences (Registration of model and image)

The two sets of tie points are matched with a Geometric Hashing algorithm [Hummel & Wolfson, 1988]. The search space of the matching approach can be approximated by rigid transformation, because we apply the matching in the image co-ordinate system. The Geometric Hashing algorithm works in two steps examining independently first the map and then the image. Assuming rigid transformation between map and image, two points are sufficient to form a rigid base. In the first step every possible combination of two map points define a base line. All other map points are expressed in terms of angle and ratio, referring to the base line (complexity $o(n^3)$). A pointer to the two base points is written to every coordinate of a linked list. This coordinate is a related position to corresponding values of angle and ratio in a hash table, with suitable quantisized axes ratio and angle.

In the second step every possible combination of two image tie points forms a base line. Analogous to the first step, angle and ratio of the remaining image tie points are represented in the base coordinates. The map bases found in the linked lists at the related positions of the hash table contain possible point partners of the image base. For each entry of the list a vote is cast and accumulated in a histogram matrix (complexity $o(n^3)$). Map and image base pairs which achieved peak values of votes above a minimum threshold are assumed to be correspondences.

3.6 Sensor Pose from Point Correspondences (Determination of outer orientation)

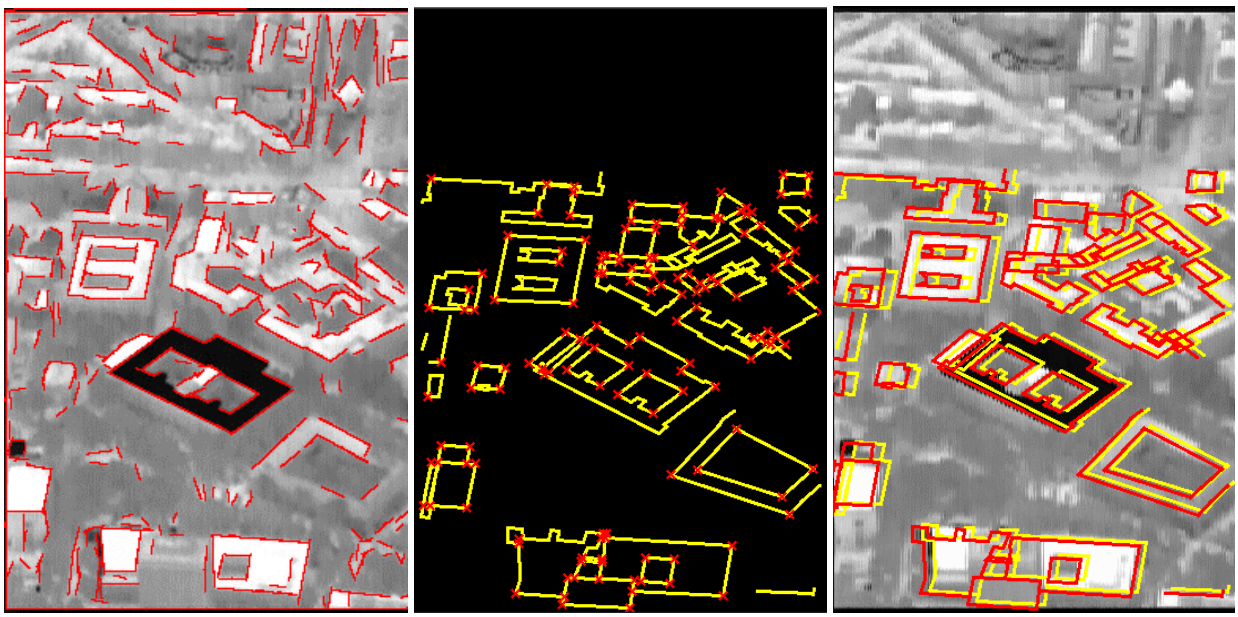
Knowing corresponding model and image points it is possible to calculate the camera orientation and navigation data of the platform. Several numerical approaches are known from literature, divided in closed-form and numerical solutions. Whenever the number of point correspondences is large (greater then 6) iterative numerical solutions are generally required. They are considered as being more robust because the measurement errors and image noise average out between the feature points and because the pose information content becomes highly redundant. Two alternative numerical methods were used to assess their suitability for the determination of platform orientation by finding sensor pose from single image interpretation.

The first method is resection in space [see e.g. Kraus, 1992], representing the classical approach making use of Newton´s method. Starting from an initial guess not faraway from the true values, this non-linear minimisation process converges in principal to a very accurate solution. The second method developed by Horaud et al. [1997] is based on a

linear algorithm, which does not require a proper initialisation. It is based on iteratively improving the pose computed with a paraperspective camera model to converge to a pose estimation computed with a perspective camera model.

Compared to the non-linear technique, the linear method usually computes the pose with fewer floating point operations, which may be useful for real-time applications. On the other hand non-linear methods are generally more stable and accurate in respect to noisy data and matching errors. Initial approximations of camera orientation required for the non-linear pose determination method can be obtained in two ways: Either from recorded navigation data given for each single frame or by prediction based on the camera parameters of preceding frames.

Applying a numerical method for camera pose determination results in an image based determination or update of navigation data. Using the calculated orientation parameters the projection of the building contours can be improved compared to the quality given by GPS accuracy only.



a b c

Fig. 5: Matching of image data and model data. a) Extracted lines of image, b) Projected lines of model, c) Line of the model before and after correction

4 DISCUSSION

At present the results of both calibration methods were merely visually assessed. Hereby, a numerical stable determination of sensor position and orientation was stated for all frames of a test sequence. Moreover, an improved alignment of projected model lines and image structures was achieved. We explain this by the fact that in our case not only a small number of tie points is available but a multitude of possible tie points is provided by numerous objects of the city model. Hence, the large number of tie points allows to eliminate false correspondences as outliers. Furthermore, in the investigated image sequences, the initial camera parameters derived from navigation data turned out to be already close to the real parameters. As a consequence, the search area for the correspondence matching can be chosen small to decrease the number of possible mismatches.

Subject of future work is to assess the calculated navigation data under consideration of assumptions about the flight trajectory. This assumptions could incorporated in context based analysis of the calculated camera parameters by considering results of preceding frames to eliminate outliers. Subject of ongoing investigations is to consider scaling effects in the field of view by projecting the objects in different levels of detail.

In the proposed investigations for every single image of the sequences the outer orientation was determined. This approach is time consuming. For real time application different possible reductions of the computational load could be considered. Assuming that to error of the navigation data changes only slowly over the sequences, an calculated error correction can be used for a prediction for some subsequent frames.

Short interruptions in the of navigation data stream (GPS, INS) could be bridged by analysing the last available navigation data for predicting missing navigation data. Errors occurred in this prediction are corrected by image based

pose estimation as described in chapter 4. This strategy fails if not enough structures are visible e.g. occlusion by clouds or if no man made structures are available in the scene or the data base.

REFERENCES

Adrian G, Fiedler F (1991) Simulation of unstationary wind and temperature fields over complex terrain and comparison with observations. Beitr. Phys. Atmosph., 64, 27-48

Burns JB, Hanson AR und Riseman EM (1986) Extracting straight lines. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 8(4):425-455.

Christy S, Horaud R (1999) Iterative pose computation from line correspondences. Computer vision and image understanding: 73(1): 137-144

Danahy J (1997) A set of visualization data needs in urban environmental planning & design for photogrammetric data. In: Gruen et al. (eds) Automatic extraction of man-made objects from aerial and space images (II), 357-366, Basel: Birkhäuser

DeMenthon DF, Davis LS (1995) Model-based object pose in 25 lines of code. International journal of computer vision, 15(1/2): 123-141

Gruber M, Kofler M, Leberl F (1997) Managing large 3D urban database contents supporting phototexture and levels of detail. In: Gruen et al. (eds) Automatic extraction of man-made objects from aerial and space images (II), 377-386, Basel: Birkhäuser

Gruen A (1998) TOBAGO - a semi-automated approach for the generation of 3-D building models. ISPRS Journal of photogrammetry & remote sensing, 53(2): 108-118

Horaud R, Dornaika F, Lamiroy B, Christy S (1997) Object pose: The link between weak perspective, paraperspective, and full perspective. International Journal of Computer Vision, 22(2): 173-189

Huising EJ, Gomes Pereira LM (1998) Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. ISPRS Journal of photogrammetry and remote sensing, 53: 245-261

Hummel R and Wolfson H (1988), Affine Invariant Matching. DARPA Image Understanding Workshop, 351-361

Kakumoto S, Hatayama M, Kameda H, Taniguchi T (1997) Development of disaster management spatial information system. Proc. GIS'97 Conf., 595-598

Kraus K (1992) Photogrammetry. Vol 2: Bonn: Dümmler

Kürner T, Cichon DJ, Wiesbeck W (1993) Concepts and results for 3D digital terrain-based wave propagation models: An overview. IEEE Journal on selected areas in communications, 11: 1002-1012

Lohr U (1998) Laserscan DEM for various applications. In: Fritsch D, Englich M, Sester M (eds) GIS - Between Visions and Applications. ISPRS, International archives of photogrammetry and remote sensing, Vol. 32, Part 4, 353-356

Phong TQ, Horaud R, Yassine A, Pham DT (1995) Object pose from 2-D to 3-D point and line correspondences. international journal of computer vision, 15(3): 225-243

Shibasaki R, Takuma A, Zhao H, Tianen C (1998) A mobile user interface for 3D spatial database based on the fusion of live landscape imagery. Proceedings of international workshop on urban multi-media/3D mapping, UM3'98, 23-30

Stilla U, Michaelsen E (1997)) Semantic modeling of man-made objects by production nets. In: Gruen A, Baltsavias EP, Henricsson O (eds) Automatic extraction of man-made objects from aerial and space images (II). Basel: Birkhäuser, 43-52

Stilla U, Jurkiewicz K (1999) Reconstruction of building models from maps and laser altimeter data. In: Agouris P, Stefanidis A (eds) Integrated spatial databases: Digital images and GIS. Berlin: Springer, 34-46