

# ON THE REFINEMENT OF URBAN MODELS BY TERRESTRIAL DATA COLLECTION

Norbert Haala

Institute for Photogrammetry (ifp), University of Stuttgart, Germany  
Geschwister-Scholl-Strasse 24D, D-70174 Stuttgart  
Norbert.Haala@ifp.uni-stuttgart.de

Commission III WG 7

**KEY WORDS:** Visualisation, Virtual Reality, Photo-realism, Terrestrial

## ABSTRACT:

Urban models can be collected area covering and efficiently based on aerial data like stereo images or LIDAR. While a large number of applications like simulations are already feasible based on the available three-dimensional building representations, a further quality improvement is required for some other tasks. Especially, if very realistic visualisations from pedestrian viewpoints have to be generated, the quality and amount of detail, which is available for urban models generated from aerial data has to be improved. As an example, due to the viewpoint restrictions of airborne platforms, image texture for the facades of the buildings frequently is not available from airborne sensors. Since this information is crucial to improve the visual appearance of urban models for pedestrian perspectives, alternative data sources have to be applied. Thus, image texture for the faces of existing building models is frequently generated based on manual mapping of terrestrial images. For this purpose, correspondences between the existing model and the terrestrial image have to be provided. As it will be discussed within the paper, the efficiency of the manual mapping process can be increased considerably by the application of panoramic scenes. If these panoramic scenes are collected from a high-level system based on a rotating CCD line scanner, large areas are covered at high resolution and superb image quality. Thus, the processing of a single scene is sufficient to extract texture for a considerable number of buildings. Texture mapping can additionally be sped up if the required correspondences between image and model are provided from automatic tools. If multiple images are collected, these tools should also allow for an automatic selection of the image, which is most suitable for texture mapping of a specific part of the building. While the main emphasis of the paper is on the refinement of existing urban models based on terrestrial images, the final part of the paper will briefly discuss the benefits of integrating existing building models to the processing of data from terrestrial LIDAR.

## 1. INTRODUCTION

The development of tools for the efficient and area covering collection of 3D city models has been a topic of intense research for the past years. A good overview on the current state-of-the-art is for example given in (Baltasvias, Grün, van Gool 2001). Meanwhile, a number of algorithms are available, which are usually based on 3D measurements from airborne stereo imagery or LIDAR. For this reason, the production of virtual city models is more and more becoming a standard task of photogrammetric data collection, resulting in a growing availability of data sets, which include 3D representations of buildings.

Frequently, these 3D models are used to generate photorealistic visualisations of the urban environment. Typically, this type of application is required in the context of urban planning, tourism or entertainment, like games based on real locations. Airborne data collection is suitable to efficiently provide a complete set of 3D building models, mainly representing the footprints and the roof shapes of all buildings at sufficient detail and accuracy. On the other hand, viewpoint restrictions of airborne platforms frequently limit the amount of information, which can be made available for the facades of the buildings. For this reason, building models collected based on airborne data can only be applied for high quality visualisations if virtual images from elevated viewpoints have to be generated. While this type of images is preferable in order to generate overviews of larger areas, a number of applications in the context of urban planning or 3D navigation require visualisations at very high degree of realism

for pedestrian viewpoints. This is only feasible, if data from terrestrial platforms is additionally available during the generation of the urban models. As an example, photorealistic visualisations of building façades require the availability of texture from terrestrial images. Of course, approaches like close-range photogrammetry or the evaluation of terrestrial LIDAR measurement allow for an autonomous collection of virtual city models. Nevertheless, the efficiency of terrestrial data processing can be increased considerably, if existing models are integrated during this evaluation. For this reason, we are aiming on the refinement of existing urban models, which have been made available from airborne data sources by the application of data from terrestrial platforms.

As already discussed, a common task to be solved in this context is the mapping of terrestrial images to the facades of existing 3D building models. Frequently, this is realised by a GUI, which allows a human operator to select corresponding primitives between the available building model and the respective images. As it will be discussed in section 2, this time consuming process of manually mapping can be sped up considerably, if panoramic scenes are applied. In our experiments the imagery is collected from a high-level panoramic camera, which is based on a rotating CCD line scanner. By these means large areas can be covered at high resolution and superb image quality. Thus, a single scene is sufficient to provide texture for a considerable number of buildings, especially if suitable camera stations have been selected during image collection. Since the number of images to be processed for texture mapping is cut down significantly, the overall effort is reduced, even though manual inter-

action is still necessary. For geometric processing of the panoramic scenes, the exterior orientation is determined from control points, which are measured manually. Control point information is provided from the available building models. As it will be demonstrated in section 3 this georeferencing process can be sped up, if these correspondences between image and building model are provided automatically. This can be realised by a shape based matching of standard terrestrial images to the existing building models. Additionally, this section demonstrates the automatic selection of optimal texture for the respective parts of the building if multiple images have been captured for a single building. The potential of integrating existing building models in the processing of terrestrial LIDAR will be discussed briefly in the final part of the paper.

## 2. TEXTURE MAPPING FROM PANORAMIC IMAGES

One general problem for the provision of façade texture from terrestrial images is the large amount of terrestrial scenes to be processed. As an example, within the project “World Models for Mobile Context-Aware Systems” (Stuttgart University 2003) a detailed virtual landscape model of the city of Stuttgart had to be made available. Within the first phase of the project, façade texture was collected for a number of buildings in the historical central area of the city by manual mapping. This manual mapping was based on approximately 5,000 terrestrial images collected by a standard digital camera. In order to extract the façade structures from these images, which were available for approximately 500 buildings several man months were required.

One approach to increase the efficiency of this process is to reduce the number of scenes to be processed for texture mapping by the application of panoramic images. Frequently, these panoramic images are used for tourist purposes in internet applications. For this type of application, the scenes can be generated by the combination of several overlapping images from simple digital cameras. In contrast, for our investigations a high resolution CCD line scanner, which is mounted to a turntable was applied. By these means high quality panoramas can be captured, which allow for precise photogrammetric processing, i.e. in applications like 3D reconstruction or mobile mapping. Especially, if the panoramic scene is collected from an elevated point in order to minimize occlusion, a large area can be covered. Thus, texture for a considerable number of buildings is available from a single scene. In addition to the reduced amount of images to be processed, the application of panoramic scenes allows to minimize changes in illumination, since image collection at different epochs is avoided.

### 2.1 Data Collection

For our investigations, panoramic scenes were captured by the camera system EYSCAN. An exemplary scene collected from this system is depicted in Figure 1. In order to demonstrate the available resolution for this type of imagery, an enlarged section of the complete scene is additionally presented.

Figure 2 depicts a 3D view generated from the 3D city model of Stuttgart, which was used as basic dataset within our investigations. The model was collected on behalf of the City Surveying Office of Stuttgart semi-automatically by photogrammetric stereo measurement from images at 1:10000 scale (Wolf 1999). For data collection, the outline of the buildings from the public Automated Real Estate Map (ALK) was additionally used. Thus, a horizontal accuracy in the centimetre level as well as a relatively large amount of detail could be achieved. The result-

ing model contains the geometry of 36,000 buildings. In addition to the majority of relatively simple buildings in the suburbs, some prominent historic buildings in the city center are represented in detail by more than 1000 triangles, each. The panoramic image depicted in Figure 1 was collected from the top of the building, which is represented by the black wire-frame lines in Figure 2.

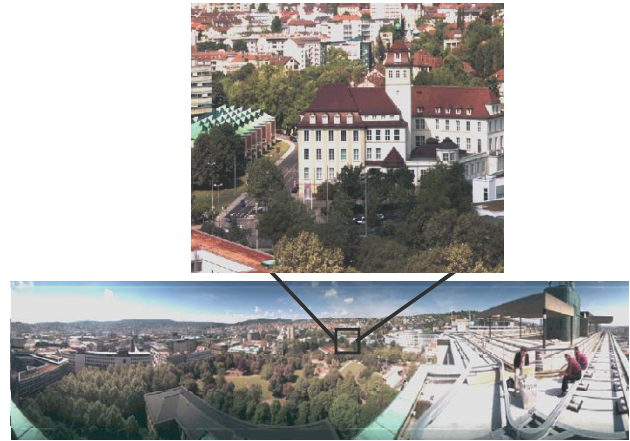


Figure 1: Image collected from panoramic camera EYSCAN.



Figure 2: Available 3D urban model.

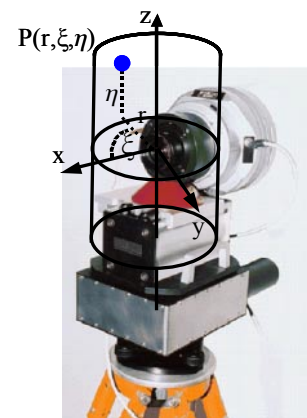


Figure 3: EYSCAN camera with cylindrical coordinate system.

The camera system EYSCAN used within our investigations was developed by the KST GmbH in a cooperation with the German Aerospace Centre. The system, which is depicted in Figure 3, is based on CCD line mounted parallel to the rotation axis of a turntable. Thus, the height of the panoramic image is determined by the number of detector elements of the CCD line. In contrast, the width of the image is related to the number of single image lines, which are captured during the rotation of the turntable while collecting the panorama. In our experiments, this resulted in an image height of 10.200 pixels, while during a 360° turn of the camera more than 40.000 columns were captured. Since the CCD is a RGB triplet, true color images are available after data collection. The spectral resolution of each channel is 14 bit, the focal length of the camera is 60mm, and the pixel size is 7 $\mu$ m.

## 2.2 Geometric Processing

In order to map the visible faces of the buildings to the respective image patches of the panoramic scene, corresponding image coordinates have to be provided for the 3D object points of the building models. In accordance to the processing of standard perspective images, the object coordinates  $\mathbf{X}_p$  are linked to the corresponding camera coordinates  $\mathbf{x}$  based on the well known collinearity equation

$$\mathbf{x} = \frac{1}{\lambda} \mathbf{R}^{-1} \cdot (\mathbf{X}_p - \mathbf{X}_0)$$

which defines a transformation between two Cartesian coordinates system. In accordance to the approach described by (Schneider & Maas 2003), a cylindrical coordinate system is additionally introduced to simplify the transformation of panoramic imagery. In this system, which is overlaid to the picture of the EYSCAN camera in Figure 3, the parameter  $\xi$  represents the scan angle of the camera with respect to the first column of the scene. The radius of the image cylinder is given by the parameter  $r$ . In the ideal case, this parameter is equal to the principal distance of the camera. The parameter  $\eta$  represents the height of an image point above the xy-plane. Thus, this parameter is related to the vertical distance of the object point to the camera station. The transformation between the cylindrical camera coordinates  $r, \xi, \eta$  and the Cartesian camera coordinates is then given by

$$\mathbf{x} = [x \ y \ z]^T = [r \cdot \cos \xi \quad -r \cdot \sin \xi \quad \eta]^T$$

In the final step, the transformation between the cylindrical camera coordinates and the pixel coordinates  $m, n$  is defined by

$$n = -\frac{\eta + \eta_0}{d_v}, \quad m = \frac{\xi}{d_h}$$

Similar to the processing of frame imagery, the pixel coordinate  $n$  in vertical direction is determined by the corresponding component  $\eta_0$  of the principal point and the vertical resolution  $d_v$ , which is defined by the pixel size. In contrast, the horizontal resolution  $d_h$  required to compute pixel coordinate  $m$  in horizontal direction is defined by the rotation angle of the CCD line per column during collection of the panoramic image.

The required exterior orientation parameters of the scene were computed from a spatial resection. In order to allow for an interactive measurement of the required control points the avail-

able 3D building models were approximately mapped to the panoramic image. Based on this approximate mapping, a sufficient number of visible faces, which were distributed over the complete scene were selected and manually measured in the image. In principle, the geometric quality of the EYSCAN camera allows for accuracies of point measurement in the sub-pixel level (Schneider & Maas 2003). Still, in our experiments only accuracies of several pixel could be achieved. In our opinion, this results from the fact, that the fit between model and image after spatial resection is not only influenced by the geometric quality of the image, but also from the accuracy, level of detail and visual quality of the available 3D building models used for the provision of the control points. While our urban model provides a reliable representation of the overall shape of the visible buildings, the amount of detail is limited especially for the facades. As it is discussed earlier, this situation is typical for 3D city model data sets, which are collected from airborne data.



Figure 4: Building models mapped to panoramic image.

Figure 4 exemplarily demonstrates the geometric quality of the mapping process based on the result of spatial resection for a part of the panoramic image.



Figure 5: Scenes generated using buildings with texture from panoramic image.



Figure 5 exemplarily demonstrates the visual quality of virtual images generated from the available building models, which are textured using the collected panoramic scene. In order to generate the virtual image a standard tool for 3D visualisation was applied. For this purpose correspondences between image and object points were provided after the visibility of each building polygon was controlled based on an occlusion detection and analysis of normal vector of each building polygon.

Usually, these standard tools for visualisation do not allow for projective transformation during texture mapping. Thus, perspective distortions are frequently eliminated in advance by a rectification of the respective image sections. For our scenes these distortions could be neglected during the visualisation process, since the facades of the buildings only cover small sections of the applied panoramic images.

### 3. MODEL BASED GEOREFERENCING AND IMAGE SELECTION

As it was demonstrated in the previous section, panoramic images can provide texture for a considerable number of buildings from a single camera station. If this camera station can be selected carefully, the quality of image texture is at least sufficient for visualisations of buildings in the background of the virtual scenes. Nevertheless, the realisation of closer virtual viewpoints will require the integration of additional images in order to minimize disturbing effects i.e. due to occluding objects or unfavourable perspectives. In order to allow for an efficient processing of these image, which were collected from standard terrestrial cameras, additional tools were developed to support the georeferencing and image selection process.

#### 3.1 Model based refinement of direct georeferencing

One important prerequisite for texture mapping is the availability of image position and orientation at a sufficient accuracy. While in our experiments the exterior orientation of the panoramic scenes is determined by spatial resection using manually measured control points, direct georeferencing can be applied alternatively if suitable sensors are available. In principle, the collection of georeferenced terrestrial images in urban areas is feasible at high accuracies by integrated DGPS/INS systems. This type of sensor system is for example applied by (Bosse et al 2000) for the collection of images, which were subsequently used in order to collect the geometry and texture of the depicted buildings. Alternatively to the application of such precise but expensive sensor systems, the accuracy requirements for direct georeferencing can be reduced considerably if – as in our case – a 3D model of the buildings at the site is already available. In that case, the approximately measured exterior orientation can be refined by an automatic alignment of the terrestrial image to the depicted building models.

In order to provide an approximate exterior orientation from direct georeferencing, a low-cost system based on a GPS receiver, an electronic compass and a tilt sensor is applied. Despite the good theoretical accuracy of differential GPS, there are a number of practical limitations of this techniques within built-up areas. Shadowing from high buildings can result in poor satellite configurations, in the worst case the signal is lost completely. Additionally, signal reflections from close buildings result in multipath effects, which are further reducing the accuracy of GPS measurement. In our experiments position accuracies of 7-10 m were achieved. The applied digital compass is specified to provide the azimuth with a standard deviation of  $0.6^\circ$  to  $1.5^\circ$ .

However, compasses are vulnerable to distortion, because especially in build-up areas the Earth's magnetic field can be influenced by cars or electrical installations. These disturbance can reduce the accuracy of digital compasses to approximately  $6^\circ$  (Hoff & Azuma 2000).

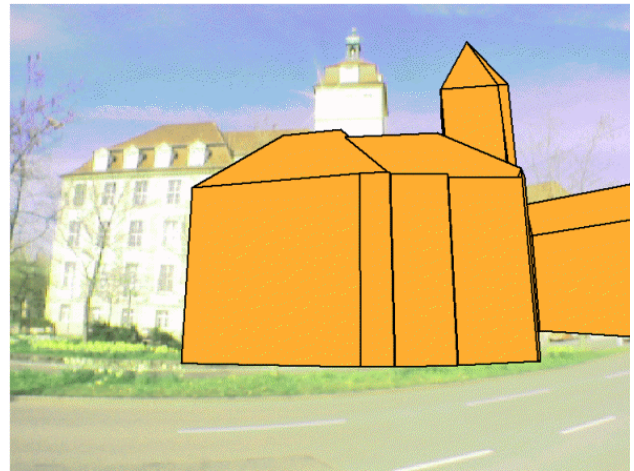


Figure 6: Coarse model to image mapping based on direct georeferencing



Figure 7: Building localization based on shape based matching.

Figure 6 shows an overlay of the available building model to the captured image based on the measured orientation parameters. The limited mapping quality, which results from the restricted accuracy of the directly measured exterior orientation is clearly visible. This coarse model to image mapping can be improved by a Generalized Hough Transform (GHT) was applied (Haala & Böhm 2003). For this purpose, the coarse model to image mapping is used to extract the visible silhouette of the depicted building. The outline of the projected building model then can then be localised based on the GHT, which automatically provides the relevant two-dimensional transformation parameters for shift, rotation and scale in relation to the respective image. The result of this step is depicted in Figure 7.

The GHT additionally allows for a certain tolerance in shape deviation. This is useful since the CAD model of the building provides only a coarse generalization of its actual shape as it is appearing in the image. After the localization of the outline of the building model, check points can be generated automatically based on the 3D coordinates of the visible building and used for

the improvement of the exterior orientation by a spatial resection. Based on this information, the mapping between image and model can be refined and image texture can be extracted automatically for the visible facades.

Within our model based refinement of direct georeferencing, the localization is based on the overall shape of the buildings. This shape is of course only depicted, if a sufficient distance between camera station and the depicted building is available. Despite these limitations, this technique could be used as one component of a hybrid system for the area covering provision of oriented images. Based on complete sequences of images the relative camera motion can be reconstructed (M. Pollefeys et al 2000). Within such a sequence, the availability of the exterior orientation is only required for some images in order to relate the complete trajectory to the object coordinate system, which is defined by the building models. One application based on such image sequences and existing building models is the elimination of occlusions during façade texturing (Böhm 2004).

### 3.2 Automatic texture selection

If a building model is depicted in multiple images, an automatic selection of the scene, which is most suitable to provide the required texture information is feasible. In principle, this selection can be performed independently for each face of the building.



Figure 8: Wire-frame model mapped to different images

Figure 8 exemplarily demonstrates this situation by mapping a wire-frame model of a building to two different images. Based on this transformation, for each image the visibility and the corresponding area can be computed for every façade of the 3D building model. If this information is put to a database, the most suitable surface texture, which provides the highest available

resolution can be selected during the visualisation process. The result of this step is depicted in Figure 9.

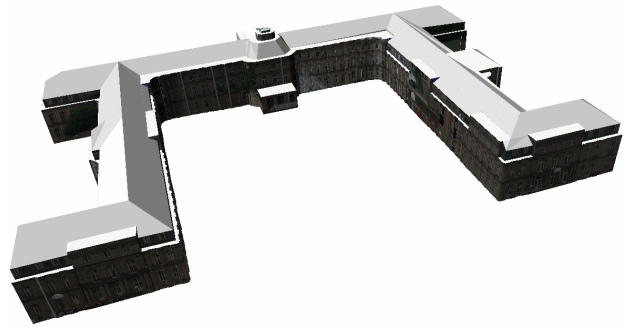


Figure 9: Automatic texture selection from multiple images.

Currently, the surface texture is only selected depending on the orientations and distances of the single building facades in relation to the camera stations. In our future work, we are aiming on the selection of texture depending on the actual virtual viewpoint. For this reason, during visualisation texture will be selected from camera stations, which are as close as possible to the respective virtual viewpoint. By these means distortions resulting from the displacement of image texture i.e. due to protrusions or balconies, which are not modelled geometrically in the available building model can be minimised.

## 4. MODEL BASED REGISTRATION OF TERRESTRIAL LIDAR

For applications aiming on the highest visual quality the provision of façade texture for existing building models might not be sufficient. In that case the geometry of the available building model has to be improved i.e. by terrestrial LIDAR. By these means a precise measurement of 3D point clouds at the facades of the building is feasible, which allows for an refinement of the model in the following steps.

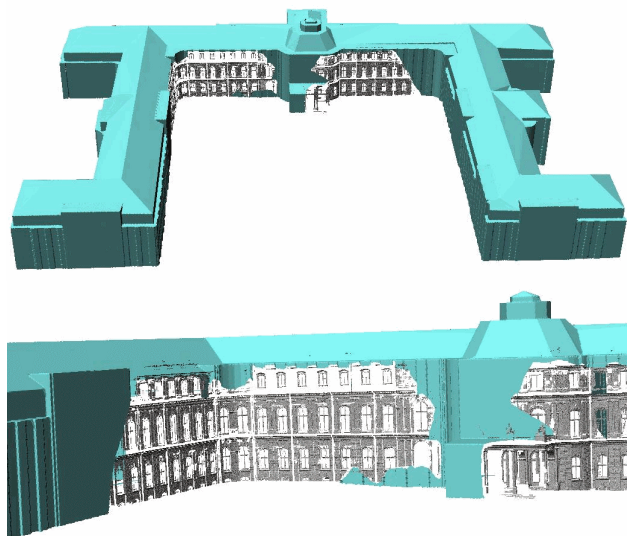


Figure 10: Terrestrial LIDAR measurement registered to existing CAD model.

During data collection, the complete coverage of complex objects like buildings can only be guaranteed, if data from differ-

ent viewpoints is available. In order to connect point clouds from different stations, frequently tie point information is provided using reflecting targets or spheres with defined diameter, which are distributed in the object area. After measurement, the operator manually specifies the approximate position of these targets. Based on this information, the target can be precisely localised automatically in the laser data and used as tie-point information in the subsequent registration process.

As it is demonstrated in Figure 10, this coregistration can alternatively be performed based on a ICP (iterative closest point) algorithm. By these means a coregistration of the laser data to the coarse three-dimensional model of the building is feasible. Thus, similar to the orientation of the collected images, control point information is provided from the existing building model in order to speed up the data processing. In principle, this technique can also be used for the georeferencing of LIDAR data collected from a moving platform (Früh & Zakhor 2003).

## 5. CONCLUSION

If existing urban models, which are frequently available from airborne data collection are used for applications like realistic visualisations of urban scenes from terrestrial viewpoints, a refinement using terrestrial data collection is required. The evaluation of the terrestrial data can be simplified considerably by integrating these existing building models to the respective processing steps. Within the paper, the combined processing was mainly demonstrated for georeferencing the terrestrial data sets by coregistration to the given models in order to provide image texture. However, this combined processing is even more important, if terrestrial data sets are used for geometric improvement of the given models.

## 6. REFERENCES

- Baltsavias, E., Grün, A. & van Gool, L. [2001]. *Automatic Extraction of Man-Made Objects From Aerial and Space Images (III)*. A.A. Balkema Publishers
- Bosse, M., De\_Couto, D. & Teller, S. [2000]. Eyes of Argus: Georeferenced Imagery in Urban Environments. *GPS World* **11**(4), pp.20-30.
- Böhm, J. [2004]. Multi-image fusion for occlusion-free facade texturing. to be published in IAPRS Vol. 35 Part B5.
- Früh, C. & Zakhor, A. [2003]. Constructing 3D City Models by Merging Ground-Based and Airborne Views. *IEEE Computer Graphics and Applications*, Special Issue Nov/Dec
- Haala, N. & Böhm, J. [2003]. A multi-sensor system for positioning in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing* **58**(1-2), pp.31-42.
- Hoff, B. & Azuma, R. [2000]. Autocalibration of an Electronic Compass in an Outdoor Augmented Reality System. Proceedings of International Symposium on Augmented Reality, pp.159-164.
- M. Pollefeys, R. Koch, M. Vergauwen & L. Van Gool [2000]. Automated reconstruction of 3D scenes from sequences of images. *ISPRS Journal Of Photogrammetry And Remote Sensing* **55**(4), pp.251-267.
- Schneider, D. & Maas, H.-G. [2003]. Geometric Modelling and Calibration of a High Resolution Panoramic Camera. *Optical 3-D Measurement Techniques VI*. Vol. II, pp.122-129.
- Stuttgart University [2003]. Nexus World Models for Mobile Context-Based Systems. <http://www.nexus.uni-stuttgart.de/>.
- Wolf, M. [1999]. Photogrammetric Data Capture and Calculation for 3D City Models. *Photogrammetric Week '99*, pp.305-312.