

AN INTEGRATED APPROACH FOR THE ORIENTATION OF TERRESTRIAL OUTDOOR SCENES

Norbert Haala, Jan Böhm, Darko Klinec

Institute for Photogrammetry (ifp), University of Stuttgart, Germany

Commission II, WG II/1

KEY WORDS: Navigation, Orientation, Augmented Reality, Generalisation, Visualisation

ABSTRACT:

Within the paper a low-cost system for the provision of georeferenced terrestrial images in urban environments is presented. Based on an image with approximate exterior orientation from a low-cost GPS and a digital compass and a 3D CAD model of a visible model as provided from an 3D virtual city model, the exact location of the building in the image is detected automatically and used for a refined orientation of the image. The work presented in this paper is part of a project aiming on the development of a mobile device, which enables access to location based services in a complex urban environment. The intuitive access to object related information is realised by so-called telepointing. For that purposes a spatial model of the user's environment is mapped to the oriented image on order to allow the access to object related information by pointing to the respective image sections. The camera used for image collection is integrated to a mobile device, which has to be capable for data transmission in order to provide the required location based services. Since the provision of location based services currently is one of the most promising markets for the application of spatial data as they are collected by photogrammetric techniques, these applications will also be discussed briefly within the paper.

1. INTRODUCTION

Applications aiming on the provision of location-dependent access to spatial data are currently emerging to the consumer market. In order to provide information in the framework of a location aware environment, the actual position and orientation of the user has to be available. Hence, to reach the goal of an area-covering provision of location aware applications like personal navigation to any user anywhere, these systems have to be based on small, yet powerful, mobile devices, which integrate features like the accurate localization of the user as well as a location-dependent processing and transmission of spatial data via wireless networks.

Up to now photogrammetry has mainly benefited from the dynamic development triggered by these application scenarios due to the resulting demand for the collection and update of geographic data bases. Additionally, the emerging spread of location based services also stimulated the development of new products currently not provided by standard 2D GIS environments. One of the most obvious examples are current developments in car navigation systems, aiming on the superseding of 2D map-like representations by a 3D visualization of the environment. Thereby a more realistic and intuitive presentation of the path to be followed by the user can be achieved. Since this type of visualization of course presumes the availability of a complete 3D model of the environment, this will result in an increasing demand for area covering 3D data collection mainly in urban environments. In addition to a realistic visualization of the 3D environment the intuitive provision of supplementary information related to the visible objects is another key feature of location based services. In order to enable a very realistic presentation of object related information as well as a intuitive realization of spatial queries, so-called augmented reality (AR) techniques can be applied. By this techniques object related information as it is represented by

computer graphics is overlaid to the user's current field of view based on optical see-through head-mounted displays.

One of the most valuable applications of AR, which can also be expected to become of considerable interest for the consumer market, is its capacity to provide situation awareness in build-up areas. This results from the fact that urban environments are usually complicated, dynamic, and inherently three-dimensional. Within an urban environment, AR can for example be applied for the presentation of name labels or additional alphanumeric data appearing to be attached to a side of a building. Thereby a user is supported if he navigates through an unknown build-up area. As an example (Höllner et al. 1999) present the names of buildings to the user depending on his actual field of view by see-through data glasses. By pointing to the respective buildings, additional information is made accessible via an integrated wireless access to the internet. This feature is also realized within the project NEXUS (Fritsch et al 2000), which is the basis of the work presented within this paper. For simplification of the overall system within this project the head-mounted display is replaced by an image of the user's environment. This image can for example be captured by a camera integrated into a small hand-held display.

If the exterior orientation of the captured image is available, the augmented world data, i.e. the spatial model of the user's environment enriched by additional objects, can be directly mapped to the corresponding sections of the image. After this step access to object related information can be realized by pointing to respective regions of interest directly on the image display (telepointing). The required mapping between virtual objects and the real world as it is depicted by the image must guarantee a correct overlay of the generated computer graphics to the corresponding objects in the perceived environment. For this reason, position and orientation of the camera or the head

mounted display, respectively, has to be determined at an sufficient accuracy.

Within the paper the automatic alignment between terrestrial images and an available spatial model in an urban environment will be described. The main component of this spatial model is provided by a 3D representation of the visible buildings. The alignment is based on the direct measurement of exterior orientation by low cost components. For refinement of this coarse measurement, the visible silhouettes of the depicted buildings are localized automatically in the image based on a Generalized Hough Transform (GHT). After this matching check points can be generated automatically based on the 3D coordinates of the visible building primitives and subsequently used for the improvement of the exterior orientation by a spatial resection. Even though our work is currently motivated by the provision of location dependent services, the approach can also be applied if image pose has to be automatically determined accurately in urban environments for a data collection like the automatic provision of image texture for the facades of the building models.

2. MODEL TO IMAGE MAPPING

In order to enable spatial queries by pointing to object related regions of interest directly on the image display, the augmented world data is directly mapped to the captured image based on the exterior orientation of the current viewpoint. This augmented world data consist of a spatial model of the user's environment enriched by additional objects and object related information.

2.1 Spatial Model

In an urban environment, the main component of the required spatial model, which has to represent the perceived environment, is provided by a 3D city model. This 3D city model essentially contains the 3D wire-frames of the visible buildings. In addition to other applications in the context of visualizations, 3D city models are also required for network planning or city climate and environmental research. Also triggered by the requirements of these applications, recent years have shown great effort in the development of tools for the accurate and efficient reconstruction of 3D city models. For this reason a rapidly growing number of data bases is becoming available. Even though the quality of these 3D city models is an important factor for the accuracy of the model to image mapping, the description of available approaches for the required 3D building reconstruction is beyond the scope of this paper.

For our test area in the city of Stuttgart two different 3D city model datasets are available. The first dataset was derived automatically at Stuttgart University based on a combination of laser DSM and groundplan information (Haala & Brenner 1999). The second data set has been collected manually by photogrammetric stereo measurement of images at scale 1:10.000 (Wolf 1999). This data has been provided by the City Surveying Office of Stuttgart. For both datasets the outline of each building is defined by the public Automated Real Estate Map (ALK), which provides accuracies in the centimetre level.

2.2 Required Mapping Accuracies

In general, the mapping accuracy, which is required for the



Figure 1. 3D city model of Stuttgart test area

overlay of supplementary information to the real environment, is tightly coupled to the type of data to be integrated. The accuracy requirements are fairly low, if annotations like the name of a building and its known function are overlaid to the corresponding object in the real world, or if the translation of road signs are projected to the respective signposts. The same order of accuracy is also sufficient for the presentation of routing information, for example if the path that has to be followed to reach a particular destination is overlaid to the visible street. Although all this type of information is of course coupled to the user's current position, a relatively coarse mapping is sufficient since the generated graphics have to be aligned only roughly to relatively large scale features.

In contrast, if the system shows highly localized information like the wire frame versions of buildings, a user cannot afford to see them far apart from the real edges of the observed objects. If the graphics are not exactly aligned in this type of application, the result will be annoying or possibly even misleading. Similar problems will occur if the presentation of otherwise hidden features such as the location of power lines, water supplies or other infrastructure and utility information is aspired. If the reconstructed virtual environment is directly overlaid to the observers view, even small errors in the model in the order of tens of centimeters can lead to significant errors, undermining the effectiveness of an AR system. For this reason all of this data can only be accurately registered to the corresponding observed object primitives, if both a detailed model of the environment and an accurate tracking system is available.

2.3 Image Georeferencing in Urban Environments

For airborne imagery, the exterior orientation usually is determined indirectly based on a spatial resection or a bundle block adjustment, respectively. For this purpose commercial software tools are available, which enable an efficient and almost autonomous processing. In contrast to that, the required procedures for highly automated tie and control point measurement do not exist for terrestrial images of natural outdoor scenes. Hence, direct georeferencing, i.e. the direct measurement of camera position and orientation at the time of image capture by a suitable sensor system usually is the preferable solution for the processing of the outdoor scenes to be processed by our system.

A commercial system, which integrates directly geocoded image sequences together with supplementary information into electronic city maps is described by (Sood & Fahrenhorst 1999). Their system mainly aims on a coarse inspection of an area of interest based on the collected images, which are linked to a digital 2D map. Thus, the quality of their collected exterior orientation provided by GPS measurement is not sufficient for a task like telepointing. A system for the collection of georeferenced terrestrial images in urban areas at high accuracies using integrated DGPS/INS measurements is presented by (Bosse et al. 2000). In their application the collected images are used for the subsequent measurement of building geometry. Therefore the accuracy demands to be met by their system hardware are considerable high and thus would also meet the specifications for precise object to image mapping. Alternatively to the application of very precise and expensive sensors for direct georeferencing, the accuracy requirements of the measured camera position and orientation can be reduced if – as in our case - a 3D model of the buildings at the site is already available. In that case the terrestrial imagery can be aligned to the reconstructed buildings by an automatic matching of corresponding primitives between object and image space.

2.4 Exemplary Application

The platform we used for data collection within our current system is depicted in Figure 2. It consists of a standard resolution color video camera with extreme wide-angle lens, a GPS receiver, an electronic compass and a tilt sensor. By combining image data and orientation data we obtain an image with an approximate exterior orientation. The camera is a consumer style Sony DFW-500 video camera connected to the system via IEEE 1394 also known as FireWire. The camera was calibrated on a test-field using a ten parameter camera model. The GPS receiver is a Garmin LP-25, which can be operated both in normal and differential mode. We used the ALF service (Accurate Positioning by Low Frequency) of Deutsche Telekom for differential mode, thus obtaining a correction signal every three seconds. All the devices are connected to a laptop. While the camera and compass/tilt sensor are hand held, the GPS is attached to a backpack.



Figure 2. Prototype of the mobile photogrammetry device.

An exemplary application based on our current prototype is depicted in Figure 3. Simultaneously to the capture of the image, the position and orientation of the camera is determined.

Afterwards the viewing frustum is calculated for the captured image and projected to an ortho image or a map. Based on this information, the visible building is selected from the available data base and corresponding object related information as it is for example provided by a website is presented by the graphical user interface. These websites then give access to services like ticket sales if for example a theatre is visible. For demonstration of the telepointing functionality, the system is currently realized within a standard GIS software package. In the final system the NEXUS platform will provide both the management of the positioning components and the provision of the spatial models. A small mobile device (PDA) will be utilized as personal NEXUS station and information between platform and station will be exchanged by wireless communication.

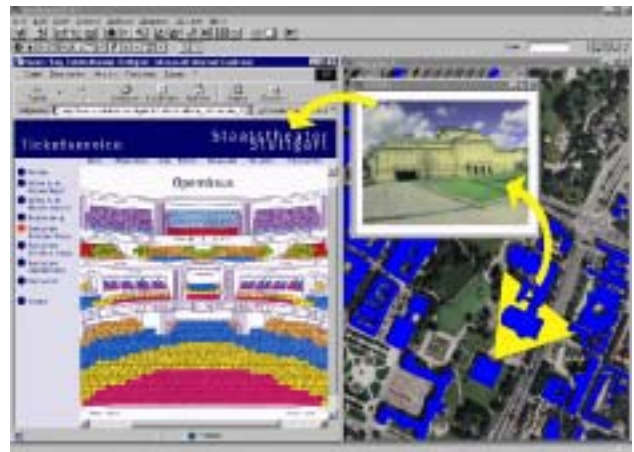


Figure 3. Exemplary application of the NEXUS prototype

2.5 Direct Georeferencing Based on Low-Cost Components

While the theoretical accuracy of differential GPS as it is used in our prototype is very high, there are many practical limitations. This is especially true when using GPS in built-up areas. Shadowing from high buildings causes poor satellite configurations, sometimes the signal is lost completely. Additionally, signal reflections from buildings nearby cause so called multipath effects, which are further reducing the accuracy of GPS measurement. Our experience shows that the system as it is depicted in Figure 2 allows for a determination of the exterior orientation of the camera to a precision of 7-10 m in planar coordinates. In our system, the vertical component of the GPS measurement was discarded and substituted by height values from a digital elevation map due to the higher accuracy of that data source. The orientation accuracy provided by the digital compass and the tilt sensor resulted in an error of approximately 1° – 2°.

An example of the accuracy which can be achieved for model to image mapping based on the directly measured parameters of exterior orientation is given in Figure 4. For the given orientation and calibration of the camera, a rendered view of the depicted building computed. The outline of this rendered view, i.e. the visible silhouette of the building then has been overlaid to the image as a red polygon. As it is clearly visible in Figure 4, the accuracy of direct georeferencing based on the applied low-cost components only allows for an approximate model to image mapping.



Figure 4. Silhouette of the building as projected to the image based on the exterior orientation from GPS and digital compass.

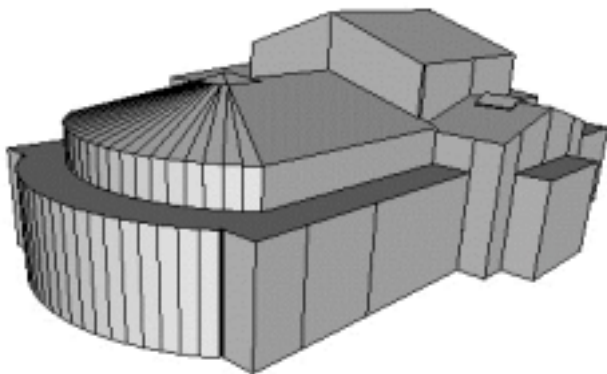


Figure 5. 3D model used for generation of building silhouette.

In order to additionally show the amount of detail, which is represented by the available building model, Figure 5 gives a 3D view of the building already depicted in Figure 4 from a slightly different viewpoint.

Even though the accuracy of direct georeferencing by our low-cost system is sufficient for some applications, this coarse mapping as it is depicted in Figure 4 is not sufficient if highly localized information has to be presented to the user. Thus, in order to enable a precise access object related information by pointing to corresponding image regions, the model to image mapping has to be refined. For the refinement of the directly measured exterior orientation a method for automated appearance-based detection of buildings in terrestrial images is applied. The problem is stated as follows: From an image with a given approximated exterior orientation and a three-dimensional CAD model of the building, detect the exact location of the building in the image and use this information for a refined pose estimation of the camera.

3. IMAGE BASED LOCALIZATION OF BUILDING SHAPES

When the task is to detect a three-dimensional shape in an image, two general strategies for object representation are available. One is the mapping of the inherent three dimensional representation of the object, which leads to a 3D to 2D matching problem, the other is the applications of a two-dimensional representation, which leads to a 2D to 2D matching

problem. While the former is the more general and theoretically more appealing approach, there are several practical problems, which often prevent its use. One of the problems is the reliability of feature extraction, the other the exponential complexity of the matching task. For the later approach in order to have a two-dimensional representation of a three-dimensional shape, one has to decompose the shape into several views and store a two-dimensional representation for each view. This approach is referred to as an aspect-graph. For our system we do not have to build the whole aspect graph. As an approximated exterior orientation of the imaging device is available, a single view of the shape can be created on-the-fly for each image in correspondence to the respective orientation data.

Additionally, when designing a object recognition system one has to choose the type of features used for recognition. The decision on the feature type is often guided by the available model data. In our case, the buildings are modelled as polyhedrons, no in-plane facade detail or texture information is available. This strong discrepancy in feature detail in-between model and image data, prevented us from using edge or corner detection. Since there is no texture information available, image correlation was also not an option. To achieve a robust detection we chose to detect the overall shape of the building in the image rather than extracting single features.

As already demonstrated in Figure 4, the overall shape of the building as it is represented by its silhouette can be calculated based on the available directly measured exterior orientation and the 3D model of the building. Now, this representation has to be detected and exactly localized within the corresponding image. For this purpose a Generalized Hough Transform (GHT) is applied.

3.1 Generalized Hough Transform

The well known Hough transform is a technique which can be used to isolate features of a particular shape within an image. Since the desired features have to be specified in some parametric form, the classical Hough transform is most commonly used for the detection of regular curves such as lines, circles, ellipses, etc. (Hough 1962). The GHT (Ballard & Brown 1982) is a generalization of this concept for the employment in applications, where a simple analytic description of a feature is not possible. In this case, instead of using a parametric equation of the curve, a look-up table is applied to define the relationship between the boundary positions and orientations and the Hough parameters.

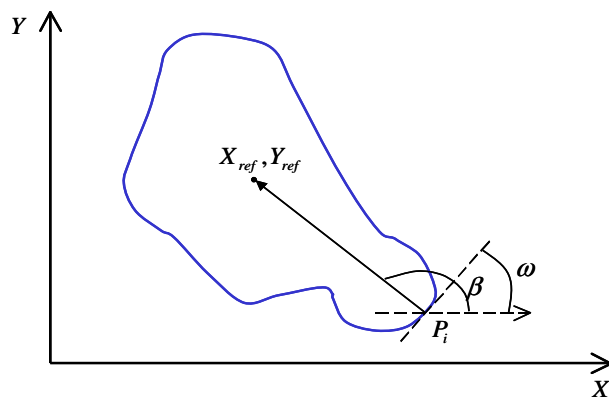


Figure 6. Representation of prototype shape as recorded in the framework of the Generalized Hough Transform.

During the offline phase, which has to be computed once for each 2D prototype shape, a look-up table, the so-called R-table is generated. In our application this object shape is defined by the silhouette of the building as it is calculated from the approximate orientation. First, an arbitrary reference point x_{ref}, y_{ref} is defined within the feature. Usually the centroid of the shape is selected for that purpose. As it is depicted in Figure 6 the shape of the feature can then be defined with respect to this point of reference by the distance r and angle β of normal lines drawn from the boundary. In order to generate the R-table, first for every point of the prototype shape the edge orientation ω is calculated. The R-table then consists of distance and direction pairs r, β , which are indexed by the orientation ω of all points along the boundary of the given shape. The Hough transform space is now defined in terms of the possible locations of the shape in the image, i.e. the possible range of the centroid's position x_{ref}, y_{ref} .

In the online phase, when the actual localization of the shape in the collected image is performed, gradients are computed by an arbitrary edge operator within the search image. After this step for each edge pixel i at position x_i, y_i the orientation ω_i is available. Now for each edge pixel $p_i(x_i, y_i)$ the corresponding values for r_i and β_i can be selected from the generated R-table based on the available orientation ω_i . These values can now be used to calculate the position x_{ref}, y_{ref} of the prototype shape in the image by

$$x_{ref} = x_i + r \cos \beta_i$$

$$y_{ref} = y_i + r_i \sin \beta_i$$

During processing now each edge pixel votes for a certain position x_{ref}, y_{ref} of the given shape. Each vote is represented by an update of the accumulator array at this calculated position. The position in the image receiving the most votes at the end is selected as the position of the shape in the image. If the orientation of the object is allowed to vary, as is the case in our application, a separate R-table has to be computed for each discrete rotation angle. The same is true for scaling, which has also to be enabled in our environment. Thus the formation of the R-tables in this case is quite complex and computationally expensive.

3.2 Implementation and Results



Figure 7. Detected silhouette of the building

For our implementation the HALCON image processing environment was used, which provides a shape detection mechanism based on the GHT as it is discussed above (Ulrich et al 2001). In order to compensate for the computational costs of large R-tables, this operator includes several modifications to the original GHT. For example it uses a hierarchical strategy generating image pyramids to reduce the size of the tables. By transferring approximation values to the next pyramid level the search space is drastically reduced. Additionally, the expected accuracy of the shape's location can be applied for an further reduction of the search space.



Figure 8. Building model projected to the image after refinement of orientation.

Figure 7 shows the silhouette of a building as automatically detected by the GHT within the captured image. Based on the estimated parameters for shift, rotation and scale within this process, the approximate image coordinates of the visible object model already shown in Figure 4 can now be improved. After the refinement of the shape's position in the image, corresponding points in object and image space are available. These control points can now be applied in order to improve the original exterior orientation by a spatial resection. In principle, the complete process, extraction of building silhouette, improvement of image coordinates by GHT and spatial resection then has to be iteratively repeated in order to avoid errors resulting from the simplification of the original 3D to 2D matching to a 2D to 2D problem. Nevertheless, for our application the differences between the projected wire-frame and the image were mainly caused by errors of the available 3D building model. Thus this iteration was not applied.

In addition to measurement errors during the collection of the 3D city model, these errors mainly result from generalisation effects. Figure 8 shows the original 3D building model projected to the image after orientation refinement. A second example of the algorithm for a different building is given in Figure 9. In this example the original model is overlaid to the collected image based on the directly measured exterior orientation (top), the result of the GHT is represented by the overlay of the building's silhouette (bottom).

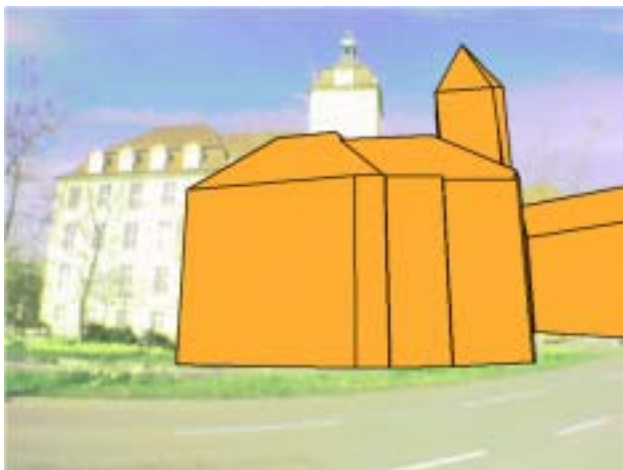


Figure 9. Building localization based on Generalized Hough Transform. Top image shows the 3D model projected by measured orientation, bottom depicts the detected silhouette .

4. CONCLUSION

Within the paper the use of oriented images as an interface to provide location based services has been discussed. Since the spatial model of the environment is mapped to the corresponding image, an intuitive access to additional object related information is feasible for a user by pointing to the respective image sections. Even though mapping accuracies of a few meters can be sufficient for some application, the directly measured position and orientation of the captured imagery by standard low-cost hardware has to be improved if highly localized information has to be presented.

For that purpose the successful implementation of a fully automated process for the detection of buildings in terrestrial images has been demonstrated. Based on the GHT the shape of the depicted building can be detected no matter whether it is shifted, rotated or optionally even scaled in relation to the image. Since the orientation is only known approximately, these degrees of freedom are required for our application. Additionally the GHT allows for a certain tolerance in shape deviation. This is also necessary, since the CAD model of the

building provides only a coarse generalization of its actual shape as it is appearing in the image.

In addition to the realisation of spatial queries by telepointing as it has been discussed within this paper, this accurately oriented terrestrial image can also be applied for image based data collection in an urban environment like a geometric refinement of the collected 3D building model or an automatic texture mapping for the facades of the depicted buildings.

REFERENCES

- Ballard, D.H. and Brown, C.M., 1982, *Computer Vision*, Prentice-Hall Inc., Englewood Cliffs, NJ.
- Bosse, M., De_Couto, D., and Teller, S. 2000 Eyes of Argus: Georeferenced Imagery in Urban Environments. *GPS World* April, 20-30.
- Fritsch, D., Klinec, D. & Volz, S. 2000, NEXUS - Positioning and Data Management Concepts for Location Aware Applications in Proc. of the 2nd International Symposium on Telegeoprocessing , pp. 171-184.
- Haala, N. and Brenner, C. 1999 Virtual City Models from Laser Altimeter and 2D Map Data. *Photogrammetric Engineering and Remote Sensing* (657), 787-795.
- Höllner, T., Feiner, S., Terauchi, T., Rashid, G. & Hallaway, D. 1999. *Computers & Graphics* 23, pp. 779-785.
- Hough, P. V. C. 1962 Method and means for recognizing complex patterns.
- Sood, R-A. & Fahrenhorst, C. 1999. Geocoded Image Sequences in *Photogrammetric Week 99* , pp. 313-316.
- Ulrich, M., Steger, C., Baumgartner, A. & Ebner, H. 2001, Real-Time Object Recognition in Digital Images for Industrial Applications in 5th Conference on Optical 3D Measurement Techniques, Vienna , pp. 308-318.
- Wolf, M. 1999. *Photogrammetric Data Capture and Calculation for 3D City Models in Photogrammetric Week '99* , pp. 305-312.