

# Real-time Indoor Positioning Using Range Imaging Sensors

Ph.D. Tobias K. Kohoutek\*, Dr. Rainer Mautz, Dr. Andreas Donaubauer  
ETH Zurich – Institute of Geodesy and Photogrammetry  
Wolfgang-Pauli-Strasse 15, CH-8093 Zurich, Switzerland

## ABSTRACT

This paper considers a novel indoor positioning method that is currently under development at the ETH Zurich. The method relies on a digital spatio-semantic interior building model CityGML and a Range Imaging sensor. In contrast to common indoor positioning approaches, the procedure presented here does not require local physical reference infrastructure, such as WLAN hot spots or reference markers.

**Keywords:** Range Imaging, CityGML, Indoor Navigation, Building Information Modeling

## 1. INTRODUCTION

The development of indoor positioning techniques is booming at the moment. For industrial applications such as automation, warehousing and logistics there is a significant demand for systems that have the capability to determine the 3D location of objects in indoor environments without the requirement of physically deployed infrastructure. In particular, tracking of persons in indoor environments has become vital during fire fighting operations, in hospitals and in homes for vulnerable people especially vision impaired or elderly people.

In contrast to the majority of indoor positioning methods, the novel method described in this paper does not require any physical reference infrastructure (e.g. Wi-Fi hot spots) inside buildings which can be a decisive advantage in respect to other methods at the same level of accuracy. Instead of a locally deployed reference infrastructure, the method relies on a digital spatio-semantic interior building model, based on the CityGML [4] scheme. The paper describes the status of current research conducted at the ETH Zurich. The research questions behind this work are:

- What should be the ideal concept for an indoor positioning method that is based on range imaging and semantically rich geospatial data (CityGML) instead of relying on physically deployed infrastructure (e.g. Wi-Fi access points)?
- Which level of accuracy can be achieved and what application scenario has the method?
- What are the strengths and drawbacks of the method compared to methods that rely on physical reference infrastructure and do not make use of a spatio-semantic model for indoor environments?

## 2. CITYGML AS A METHOD FOR MODELING INDOOR ENVIRONMENTS

### 2.1. Overview of approaches for interior building models

Different disciplines such as Computer Aided Architectural Design (CAAD) / Building Information Modeling (BIM), Computer Graphics and Geographic Information Systems (GIS) are dealing with three-dimensional interior building models with each of the disciplines having developed a different design that has been tailored to their application. The following criteria can be used for classifying these approaches:

\* kohoutek | mautz | donaubauer @geod.baug.ethz.ch; www.igp.ethz.ch

1. Creation process: CAAD/BIM models are normally generated during the planning process of a building before the construction phase and therefore represent the building as it was designed before it has been built. The state of the building after its completion is often represented by Computer Graphic Models (e.g. for virtual tours) and by 3D GIS (e.g. for Computer Aided Facility Management applications). These models are often derived from measurements taken inside the building. The models that were created during the planning phase of a building often contain information that is no longer visible after completion of the building (e.g. columns which are integrated in walls, cables and pipes or beams under suspended ceilings). In contrast, models that have been created after completion of the building normally contain only the visible parts of the building interior.
2. Geometric modeling: There are two paradigms for modeling 3D vector geometry [11]. In the Boundary Representation (b-rep) paradigm, where a room is represented by its bounding surfaces and the boundaries being lines or curves. The 3D coordinates for each vertex of a boundary are stored individually. B-rep is the common modeling paradigm for vector geometries in GIS. CAAD/BIM models often apply the Constructive Solid Geometry (CSG) paradigm. In CSG, complex 3D solids are derived by combining 3D primitives such as cuboids, spheres, cylinders with the help of Boolean operators *union*, *difference* or *intersection*. The parameters of the primitives and coordinates of insertion points are stored.
3. Semantic modeling: Models of indoor environments that are used for visualization purposes only, usually contain little explicitly modeled semantic information. Those models focus on the geometry and the data that is used for controlling the graphical representation of geometry such as textures. In contrast, models from the CAAD or BIM domain provide a large amount of semantic information as they rely on detailed semantic information for each construction component of a building. For example, the standard IFC defines more than 600 semantic object classes [5] in a building model.

Depending on the application different data formats are used for the codes and data transfer. In applications for the purpose of visualization, VRML or X3D formats are common. The standard IFC format fulfils the requirements for applications in architecture and construction. The standard CityGML [4] that is designed for GIS applications and used in our application is discussed in more detail below.

## 2.2. The CityGML indoor space model

The standard CityGML [4] defines a data model and an XML data format for 3D city and topography models. CityGML defines several Levels of Detail (LoD) with the highest LoD 4 having the capability for modeling the interior of buildings. In particular for the purpose of indoor modeling, the semantic model provides an object class room that can capture semantic data and contains the attributes class for the classification of rooms, function for the intended use and usage for the current use of the room such as living room or office. An object of the class room can be associated with its geometry in two different ways. One way of defining the outer shell of a room is to establish a link to a geometric object of type Solid or MultiSurface (both types are defined by the GML 3.1.1 specification [3]). Alternatively, the outer shell can be decomposed into semantic objects of the types InteriorWallSurface, CeilingSurface and FloorSurface. These semantic objects refer to geometric objects of type MultiSurface. Openings in the outer shell of a room can be modeled by the use of the object classes Window and Door that can belong to one or two InteriorWallSurfaces. This data structure can be used to express topological relationships between rooms. Permanent fixed objects belonging to a room (e.g. radiators, columns, beams) can be modeled using the semantic object class IntBuildingInstallation. In order to model the mobile components of a room such as desks and chairs, the object class BuildingFurniture can be used. IntBuidlingInstallation and BuildingFurniture provide the attributes class and function and usage for semantically describing the objects. The geometry of these fixed installed objects can be defined by the standard GML 3.1.1. In addition, the geometries of the variable components of a room can be modeled using the so called implicit geometries. Hereby the shape of an object is stored only once in the library even if multiple objects of the same shape are present (e.g. pieces of furniture). For each occurrence of such an object, only the local coordinates of an insertion point and a transformation matrix are stored. They are then linked to the geometry that is captured in the CityGML. Using this mechanism, the model could have a direct link to the 3D-CAD-drawings of pieces of furniture in the manufacturer's catalog. Figure 1 shows the indoor model of a room using the semantic classification of CityGML.

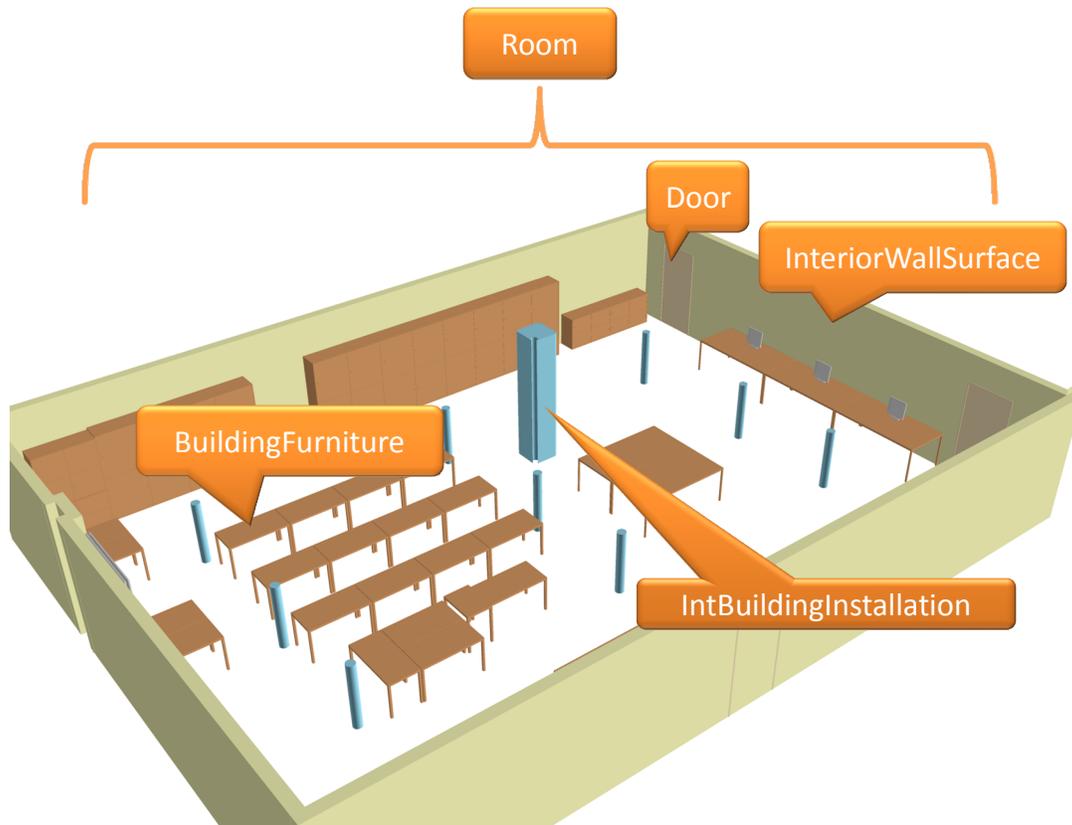


Fig. 1 ETH Zurich lecture room HIL C71.3 in CityGML

### 3. RANGE IMAGING AS A METHOD FOR THE PURPOSE OF INDOOR-POSITIONING

#### 3.1. Current indoor-positioning methods and Range Imaging as an optical method

Positioning awareness and navigation capabilities have become increasingly important for many applications in all environments. Global Navigation Satellite Systems (GNSS) and surveying total stations are able to cover the positioning requirements for outdoor applications. However, these systems have weaknesses indoors. Due to the importance to deliver position in indoor environments, various alternative approaches such as those exploiting signal strengths indicators, internodal ranges by time-of-flight for trilateration or angular measurements for triangulation have been developed with not yet satisfying performance. A drawback of these approaches is the often missing connection to the global geodetic coordinate reference that is used outdoors.

Figure 2 gives an overview of current positioning systems according to their specific coordinate accuracies and coverage. On the left side of the graphic, the high precision systems used for applications in industrial metrology are shown. The drawback of systems with positioning capabilities in sub-mm precision is the small coverage and the requirements for expensive local installations. In contrast, inexpensive systems that have been developed for low-accuracy applications are shown on the right side of Figure 2. These systems exploit the Received Signal Strength Indicators (RSSI) in order to obtain positioning capabilities within meter accuracy or to resolve the position of a device within room level.

For many applications the requirements of positioning accuracy is within millimeters to centimeters. This level of accuracy can be reached with geodetic methods such as total stations or rotational lasers. In recent years, network based methods that obtain range or time of flight measurements between the network nodes have become significant for applications at decimeter level accuracy. The measured distances can be used to determine the 3D position of a device by spatial resection or multilateration.

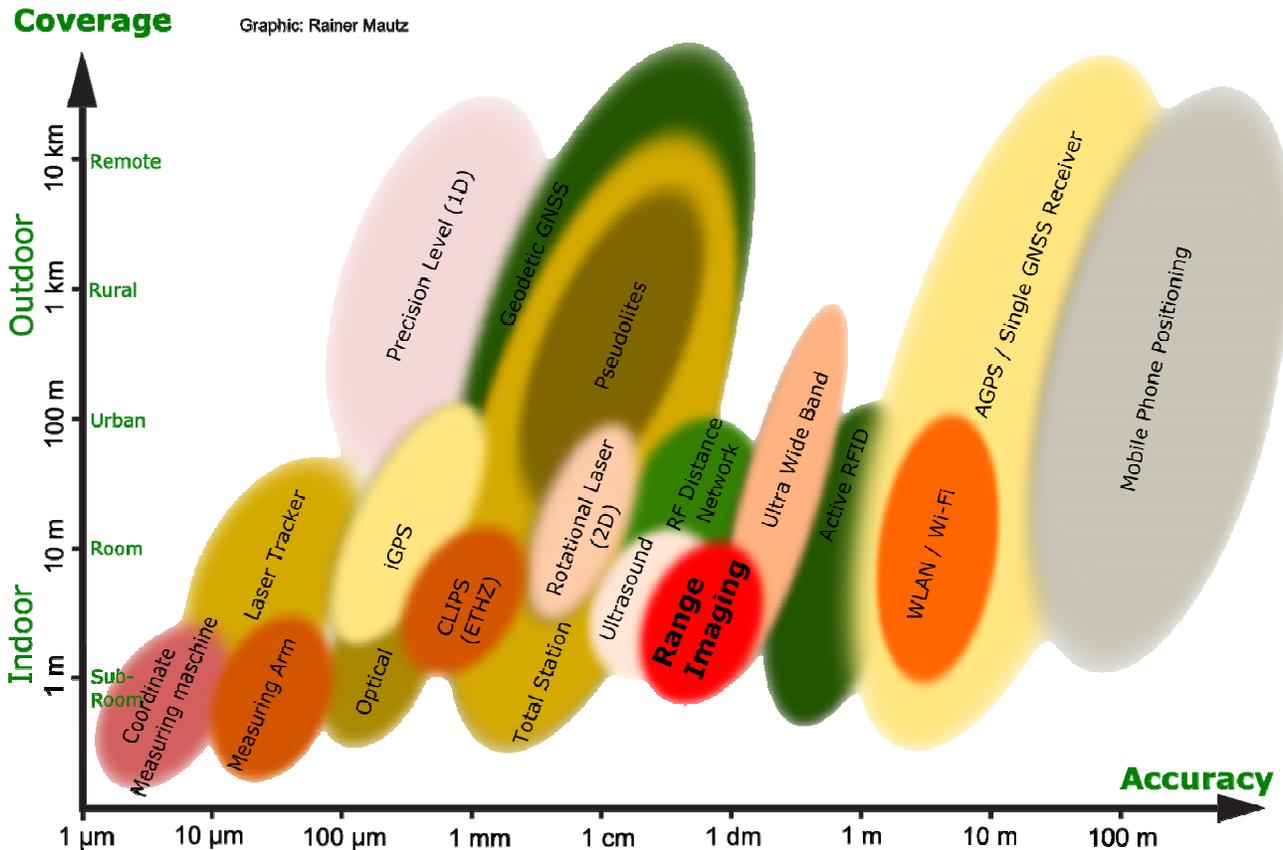


Fig. 2 Today's positioning systems in dependence to accuracy and coverage.

In particular vision based methods have become an interesting alternative for indoor positioning and navigation. The performance, the size and the speed of CCD and CMOS sensors have grown rapidly in the last few years. The computing speed and the algorithms for feature recognition in images obtained by digital cameras have reached an unprecedented performance. Optical methods use digital images to recognize points, codes, features and objects in order to determine their image coordinates. These 2D image coordinates can be transformed into the reference coordinate system with the goal to determine the camera position by spatial resection.

Manufacturers of optical indoor positioning systems such as AICON [1] with their camera system ProCam offer high precision positioning systems in the range of 1/10 mm for applications in optical metrology, in particular for surface inspection or reverse engineering. However, these systems require the installation of an active field of reference points and a prior calibration of the system.

One approach to avoid the dependency on a reference field is the project CLIPS (Camera and Laser Indoor Positioning System) which is currently under development at the ETH Zurich. The system uses the fundamentals of stereo photogrammetry, where the position and the rotation of a camera relative to another camera are derived. But instead of using a second camera, it is replaced with a device called laser-hedgehog that projects well-distributed laser spots as flexible reference points on the ceiling, walls and furnishings in any indoor environment. The projection creates a flexible field of reference points that can be observed by the real digital camera. Tilch [12] has shown that the CLIPS camera could be located with an accuracy of sub-mm.

Matsumoto et al. [9] and Desouza et al. [4] present an overview of attempts that have been made to exploit view- or map-based indoor positioning systems for mobile robot navigation in indoor environments.



Fig. 3 Range Imaging sensors, left: PMD Tech: CamCube, right: MESA Imaging: SR 4000

The method based on Range Imaging that is subject of this article belongs to the optical map-based indoor positioning systems. In contrast to traditional optical sensors, the range image does not reflect the brightness of the objects in the scene, but the distance of these objects to the range imaging camera. The expected 3D position accuracy for objects seen by a range imaging camera (in terms of a  $1-\sigma$  standard deviation) is 1 cm for distances of 2 m and 1 dm for distances of 10 m. The largest error budget contributes the low-accuracy distance measurement. According to the manufacturer the ranging accuracy of the current model SR4000 is 1.5 cm for objects in 8 m distance at a level of reflectivity of 100%. Range Imaging can be particularly used for the purpose of indoor positioning, because in contrast to the other methods mentioned above range imaging can exploit semantic 3D geoinformation models. The methods will be detailed in Chapter 4.

### 3.2. Range Imaging

Common methods to measure 3D point clouds are stereo triangulation, sheet of light triangulation, structured light projection or interferometry. In contrast to these techniques, Range imaging uses the Time-Of-Flight (TOF) principle. There are two different ways to measure distances using TOF [7]. One method is the time of flight measurement of a laser pulse, which is reflected at an object. This principle is used by most laser scanners. The other method that is applied by the sensor used in this research, the phase delay of a modulated infrared light source is measured.

Range imaging cameras provide real time distance observations at video frame rates up to 50 frames per second. They acquire an amplitude image and a range image. The local brightness as well as the distances for every pixel are coded in 16 bit gray values. For each pixel the distance is measured directly by calculating the phase shift between the emitted and reflected signal [8]. The phase map and finally a complete distance map can be acquired by detecting the phase delay between both signals as shown in Figure 4a. The three unknown parameters amplitude  $A$ , offset  $I$  and the phase  $\varphi$  of the modulated signal can be determined using the equations (1) to (3), where  $m_1 \dots m_4$  are four samples of the signal recorded at different times during one period.

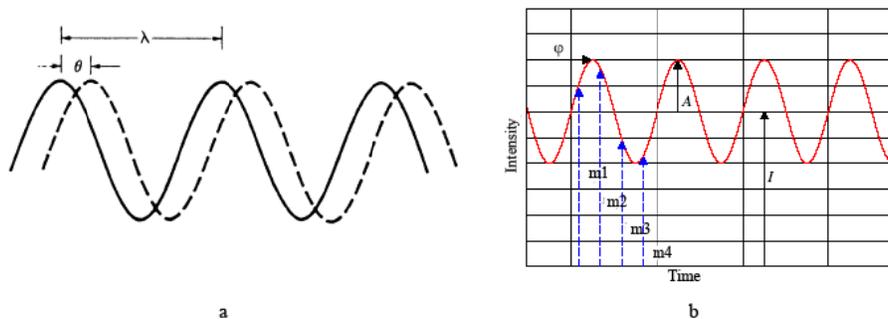


Fig. 4 a) phase delay, b) modulated signal [14]

Phase: 
$$\varphi = \arctan\left(\frac{m_4 - m_2}{m_1 - m_3}\right) \quad (1)$$

Amplitude: 
$$A = \frac{\sqrt{(m_3 - m_1)^2 + (m_4 - m_2)^2}}{2} \quad (2)$$

Offset: 
$$I = \frac{m_1 + m_2 + m_3 + m_4}{4} \quad (3)$$

The distance  $D$  and the achievable depth resolution  $\Delta D$  can be determined by

$$D = D_{\max} * \frac{\varphi}{2\pi} \quad \text{with} \quad \Delta D = \frac{D_{\max}}{\sqrt{8}} \frac{\sqrt{I}}{2A}, \quad (4)$$

where the maximal operating range of the camera is denoted by  $D_{\max}$ . Every object which is further away than  $D_{\max}$  will be assigned a false gray value due to the repetition of the modulated signal. State-of-the-art range imaging cameras operate at a wavelength of 850 nm and modulated frequencies between 25 and 30 MHz. The current array sizes are available between 60 x 60 pixels up and 205 x 205 pixels.

## 4. POSITIONING USING RANGE IMAGING AND CITYGML

### 4.1. Overview of the proposed method

The presented positioning method consists of two main components, the range imaging sensor (described in Chapter 3) and a semantic-geometric 3D-data base that is modeled in CityGML (details given in Chapter **Error! Reference source not found.**2). Details on the processing, data storage, analyzing and data transfer remain critical for the realization of the system, but cannot be given within the scope of this paper. A description of the method is given below.

### 4.2. Room identification through object detection and processing using the CityGML data base

This first step has the goal to identify the room in the CityGML data base, where the camera is located. The detection and identification of objects is the key part of this step, which can be achieved from the amplitude image of the range imager that is similar to a grayscale optical image of the scene. In order to identify the objects such as chairs, tables, etc., known or “learned” primitives, features and image samples from the libraries (that are described in Chapter 2.2) are matched with the image data from the camera. The detected object properties such as the size, geometry or quantity of a certain object are the main criteria for the comparison with the data base. This way, the unknown camera position can be reduced to a small number of possible rooms. By detecting distinct properties the room can be identified uniquely and additional semantic and geographic information can be extracted from the 3D geo-data base. Figure 5 shows the comparison between an observed 3D point cloud from the range imaging sensor and a form primitive of a data base model.

### 4.3. Accurate positioning from distance measurements

The second step of camera localization is the precise positioning part, described in this section. This step compares and transforms the local coordinates of the objects that have been recognized by the camera into the reference coordinate system of the database. The reference points for the transformation are the corners of the room, vertices of doors, windows and other fixed installation or furniture. The accuracy of the objects in CityGML should be at centimeter level<sup>1</sup> and should lead to position determination of the camera with cm-accuracy using a least squares adjustment with a redundant number of reference points to determine the 3D camera position. One requirement for the camera is that its inner orientation has been determined previously. The outer camera orientation and position are determined by a technique that combines trilateration (based on the distance measurements) and spatial resection (based on the image coordinates that are translated into horizontal and vertical angles). If there has been an ambiguous solution in the

---

<sup>1</sup> The Standard CityGML at level of detail 4 (LoD 4) defines a horizontal and vertical accuracy of 0.2 m.

identification at room level in step 1, the precise positioning step has the potential to disambiguate and deliver only one unique solution for the correct room. Further research needs to be investigated with the goal to exploit the semantic information that the CityGML data base holds.

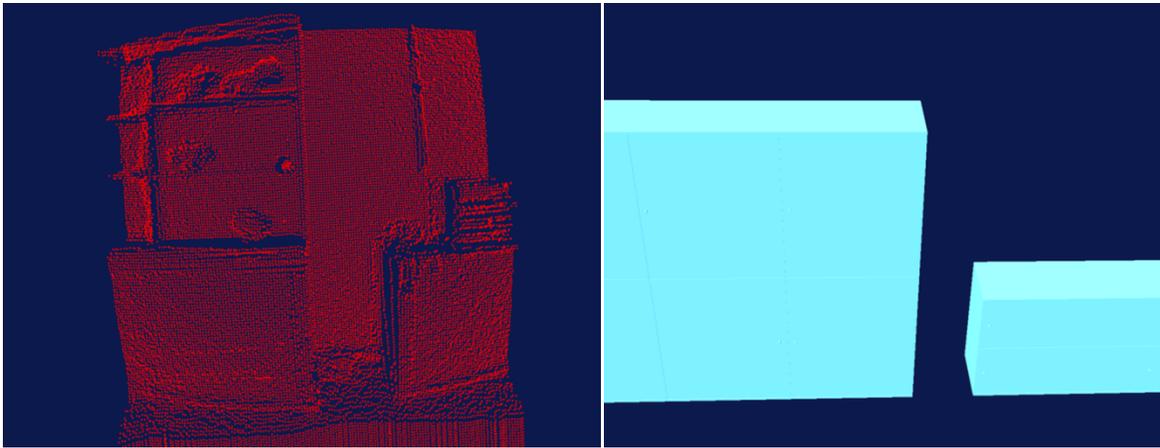


Fig. 5 Object comparison between a range image (left) and form primitives from data base (right)

#### 4.4. Opportunities and limits of the proposed method

Kinematic acquisition of 3D-coordinates in real-time allow for efficient recognition of rooms and the position of objects in those rooms in relation to a given model. The identification of objects can be trained with the help of neuronal networks. Currently, the proposed method is limited by the relatively small distance measurement range. Modern Range Imaging sensors are able to measure distances unambiguously between 5 – 10 m at an accuracy level of centimeters. The ambiguity problem arises from the frequency of the modulated signal of the Range Imaging sensor. For example, the SR4000 camera has a unique distance range of 5 m, i.e. an object in 6 m distance from the camera could also be in 1 m or 11 m distance. The ambiguity problem can only be solved with additional prior information.

Another problem pose the so-called mixed pixels, that are obtained when the signal from the Range Imaging camera hits an edge of an object. Then, the signal is partially reflected at the foreground, but also reflected at the background. Both signal parts arrive at a single CCD element. As a result, the values of the mixed pixels consist of an average between the foreground and background distance. In the point cloud, these pixels appear as single unconnected points that seem to float in the air and that do not belong to any object. This is also a common problem in terrestrial laser scanning. Note that systematic optical influences such as focussing, vignetting<sup>2</sup> and aberation<sup>3</sup> must also be determined by a prior calibration and need to be corrected accordingly.

CityGML seems to be an appropriate basis for the positioning method for the following reasons:

1. The coherent spatio-semantic model with its high level of detail and the positioning method are complementing one another. Range-imaging and other imaging methods used for positioning can capture visible objects. Also in CityGML, only visible objects are modeled. Partially visible objects are split in such a way that only their visible part is modeled. This approach is different from semantic interior building models from the CAAD/BIM domain which contain a high amount of objects which are invisible (like cables or pipes) or just partially visible (like beams spanning several rooms) and therefore do not fit the range imaging method very well.
2. For CityGML there is a standardized Web access interface, the OGC WFS (VRETANOS et al. 2005) which provides access to data and associated operations (e.g. geometric and thematic filtering of data).

<sup>2</sup> Vignetting is the effect reduction in an image's brightness or saturation at the periphery compared to the image center.

<sup>3</sup> Aberaction is the optical distortion that leads to a local variation in scale due to geometric errors of an optical system.

3. In contrast to models from the CAAD/BIM domain, CityGML supports any geodetic reference system. If an interior building model is already in the geodetic reference system, the coupling of indoor and outdoor positioning methods is straightforward.

Further practical tests and an assessment will show whether our optimistic view for the use of CityGML turns out to be justified or whether CityGML needs to be further extended as shown in BLEIFUß et al. 2009 in the facility management domain.

## 5. OUTLOOK

First steps towards a realization of the proposed indoor positioning method have been carried out with a Range Imaging camera. In parallel, parts of an office building at the ETH Zurich have been modeled in CityGML. The next steps are the implementation of the coarse and the fine positioning method. These methods need to be tested in order to have an answer to the questions that have been raised in the beginning of this paper.

## REFERENCES

- [1] AICON 3D Systems (2009): <http://www.aicon.de>, last accessed on January, 10, 2010.
- [2] BLEIFUß, R., DONAUBAUER, A., LIEBSCHER, J., SEITL, M. (2009): „Entwicklung einer CityGML-Erweiterung für das Facility Management am Beispiel Landeshauptstadt München“, In: Symposium und Fachmesse für Angewandte Geoinformatik (AGIT)
- [3] COX, S., DAISEY, P., LAKE, R., PORTELE, C., WHITESIDE, A. (2004): “OpenGIS® Geography Markup Language (GML) Implementation Specification” Version 3.1.1, International OGC Standard. Open Geospatial Consortium, Doc. No. 03-105r1.
- [4] DESOUZA, G. and A. KAK (2002), “Vision for Mobile Robot Navigation: A Survey”, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 24, No. 2, pp. 237-267, 2002
- [5] GRÖGER, G., KOLBE, T., CZERWINSKI, A., NAGEL, C. (2008): „OpenGIS City Geography Markup Language (CityGML) Encoding Standard“ Version 1.0.0, International OGC Standard. Open Geospatial Consortium, Doc. No. 08-007r1.
- [6] ISO/PAS 16739:2005: „Industry Foundation Classes, Release 2x, Platform Specification (IFC2x Platform)“
- [7] KAHLMANN, T., INGENSAND, H. (2005): „Untersuchungen des hochauflösenden Range-Imaging-Sensors SwissRanger™“; Photogrammetrie Laserscanning Optische 3D-Messtechnik, Beiträge der Oldenburger 3D-Tage 2005, S. 156; Luhmann, Th. (Hrsg), Herbert Wichmann Verlag, Heidelberg.
- [8] LANGE, R. et al. (1999): „Time-of-Flight Entfernungskamera in CMOS/CCD-Technik mit pixelintegrierten Lock-in-Verstärkern“; DGZfP congress Optische Formerfassung, Stuttgart.
- [9] MATSUMOTO, Y, SAKAI, K, INABA, M. and INOUE, H. (2000) “View-Based Approach to Robot Navigation”, in Intelligent Robots and Systems, Vol. 3, pp. 1702-1708, 2000
- [10] Oggier, T., Lehmann, M., et al., “An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger™)”, St. Etienne, SPIE Optical Design and Engineering, (2004)
- [11] PLÜMER, L. & KOLBE, T. (2004): „Bridging the Gap between GIS and CAAD“. In: GIM International, July 2004, Lemmer, The Netherlands.
- [12] TILCH, S. (2010): „Entwicklung eines optischen Innenraum-Positionierungssystems“, Master Thesis at ETH Zürich, January 2010.
- [13] VRETANOS, P. et al. (2005): „Web Feature Service Implementation Specification“ Version 1.1.0, International OGC Standard, Open Geospatial Consortium, Doc. No. OGC 04-094
- [14] WEINGARTEN, J.W. (2004): „A State-of-the-Art 3D Sensor for Robot Navigation“; IROS, Sendai.