Crowd-sourced reconstruction of building interiors

A thesis accepted by the Faculty of Aerospace Engineering and Geodesy of the
University of Stuttgart
in partial fulfilment of the requirements for the degree of
Doctor of Engineering Sciences (Dr.-Ing.)

by
Michael Peter
born in Alzenau

Committee chair:	Prof. Dr.-Ing. Dieter Fritsch
Committee member:	Prof. Dr. rer. nat. Kurt Rothermel
Date of defence:	November 20th, 2015

Institute for Photogrammetry
University of Stuttgart
2015

DGK Deutsche Geodätische Kommission
der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 768

Crowd-sourced reconstruction of building interiors

Von der Fakultät Luft- und Raumfahrttechnik und Geodäsie
der Universität Stuttgart
zur Erlangung der Würde eines
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
genehmigte Abhandlung

Vorgelegt von

Dipl.-Ing. Michael Peter

aus Alzenau

München 2016

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission beim Verlag C. H. Beck

ISSN 0065-5325 ISBN 978-3-7696-5180-5

Adresse der Deutschen Geodätischen Kommission:

Deutsche Geodätische Kommission

Alfons-Goppel-Straße 11 ● D – 80 539 München

Telefon +49 – 89 – 23 031 1113 ● Telefax +49 – 89 – 23 031 - 1283 / - 1100

e-mail hornik@dgfi.badw.de ● http://www.dgk.badw.de

Hauptberichter: Prof. Dr.-Ing. habil. Dieter Fritsch
Mitberichter: Prof. Dr. rer. nat. Kurt Rothermel

Tag der mündlichen Prüfung: 20.11.2015

Diese Dissertation ist auf dem Server der Deutschen Geodätischen Kommission unter <http://dgk.badw.de/>
sowie auf dem Server der Universität Stuttgart unter <http://elib.uni-stuttgart.de/opus/doku/e-diss.php>
elektronisch publiziert

ISSN 0065-5325 ISBN 978-3-7696-5180-5

Abstract

Location-based services (LBS) have gained huge commercial and scientific interest in recent years, due to the ubiquitous and free availability of maps, global positioning systems, and smartphones. To date, maps and positioning solutions are mostly only available for outdoor use. However, humans spend most of their time indoors, rendering indoor LBS interesting for applications such as location-based advertisement, customer tracking and customer flow analysis. Neither of the two prerequisites for indoor LBS - a map of the user's environment and a positioning system - is currently generally available: Most positioning methods currently under scientific investigation are based either on fingerprint maps of electro-magnetic signals (e.g. WiFi) or inertial measurement units. To overcome the flaws of these methods, they are often supported by models for the human movement which in turn rely on indoor maps. Ready-made maps, on the other hand, are generally unavailable due to indoor mapping being mostly manual, expensive and tedious. The vast amount of unmapped indoor space therefore calls for the transfer of methods used by Volunteered Geographic Information (VGI) communities like OpenStreetMap to indoor mapping. These methods comprise the digitization of features of interest such as building outlines from aerial images released to the community and the use of position traces. In this thesis, approaches are illustrated which can serve to enable this transfer.

On the one hand, the thesis shows how photographs of evacuation plans - which are a compulsory part of the safety equipment of publicly used buildings in many countries - can substitute for the aerial images in the indoor domain. Due to the standardised nature of such plans, the manual digitization employed by VGI mappers in the outdoor domain can be replaced by an automatic reverse-engineering pipeline. To this end, the image is pre-processed and symbols, which depict evacuation routes or emergency equipment, are detected. Subsequently, foreground objects (i.e. walls) are distinguished from the background using an adequate binarisation operation. Based on the binary image, the sought-after vector information can be extracted by skeletonisation and skeleton tracing. The model is finalised by a bridging operation of the previously detected symbols which occlude parts of walls or stairs. As the model resulting from these operations is only available in a coordinate system defined by the original image, the transformation to a world-coordinate system or, at least, the unknown scale has to be determined. To this end, the indoor model is matched to an available model of the building's external shell. By detection of stairs, an approximate floor height can be computed and the 2D model is extruded to a 3D model.

On the other hand, geometric features and semantic annotations may be added to existing models using pedestrian traces recorded by an indoor positioning system. As suitable generally available and low-cost systems do not exist yet, their existence is simulated in this work by a dead-reckoning system basing on a foot-mounted inertial measurement system. Methods for the derivation of the initial position and orientation necessary for the application of such a system are shown, as well as methods enabling the correction of remaining errors. The latter comprise an alignment approach using the external building shell and a map-matching method which employs the existing coarse model derived from the evacuation plan. Building

on the collected pedestrian traces, semi-automatic and automatic approaches for the existing models' semantic and geometric refinement are presented which range from semantic annotation using the analysis of photographed doorplates to automatic door reconstruction. Furthermore, a geometric update of single rooms by conjoint analysis of the coarse model, pedestrian traces and a hand-held low-cost range camera is described.

Lastly, works of indoor mapping are presented which are based on pedestrian traces and higher-level knowledge about the interior structure of the building modelled in an indoor grammar. Due to the differing characteristics of the two central elements of building interiors, corridors and rooms, the grammar is composed of a Lindenmayer system modelling the floor's corridor system and a split grammar describing the room layout which is found in the non-corridor spaces. The grammar is put to the test by applying it to distributedly collected noisy trace data.

Kurzfassung

Das kommerzielle und wissenschaftliche Interesse an ortsbezogenen Diensten (location-based services, LBS) hat in den vergangenen Jahren stark zugenommen, verursacht durch die allgegenwärtige und kostenlose Verfügbarkeit von Karten, globalen Positionierungssystemen und die weite Verbreitung von Smartphones. Gegenwärtig sind Karten und Positionierungslösungen hauptsächlich für die Verwendung im Außenraum nutzbar. Menschen verbringen jedoch den Großteil ihrer Zeit in Gebäudeinnenräumen, was LBS für Anwendungen wie ortsbezogene Werbung oder Verfolgung und Analyse von Kundenbewegungen interessant macht. Jedoch ist keine der beiden Voraussetzungen für LBS - eine Karte der Umgebung des Benutzers und ein System für die Bestimmung seiner Position - derzeit in der Breite verfügbar: Die meisten Positionierungslösungen, die gegenwärtig wissenschaftlich untersucht werden, basieren entweder auf Karten, in welchen die Fingerabdrücke elektromagnetischer Signale (wie WLAN) gespeichert sind, oder auf inertialen Messeinheiten. Um die Schwachstellen solcher Systeme auszugleichen, werden diese häufig durch Modelle der menschlichen Bewegung unterstützt, welche ihrerseits Karten des Innenraums benötigen. Zum gegenwärtigen Zeitpunkt fehlen jedoch direkt verwendbare Karten im Allgemeinen, da die Kartierung von Innenräumen meist einen großen manuellen und finanziellen Aufwand erfordert. Die große Menge nicht kartierten Raumes motiviert daher eine Übertragung der Methoden, die von Internetcommunities wie OpenStreetMap eingesetzt werden, welche sich mit der Erfassung von Geodaten durch Freiwillige (Volunteered Geographic Information, VGI) beschäftigen. Diese Methoden bestehen zum Einen aus der Digitalisierung von Karteninhalten wie Gebäudegrundrissen aus Luftbildern, die der Community zur Verfügung gestellt wurden, zum Anderen aus der Verwendung von Benutzerfußspuren. In dieser Arbeit werden Ansätze beschrieben, die die Realisierung der Übertragung dieser Methoden unterstützen können.

Zunächst wird vorgestellt, wie Fotografien von Fluchtplänen, welche laut Vorschrift in vielen Ländern Teil der Sicherheitsausstattung von öffentlich genutzten Gebäuden sind, die im Außenraum genutzten Luftbilder ersetzen können. Aufgrund des standardisierten Aussehens solcher Pläne kann die manuelle Digitalisierung, welche von VGI-Kartographen im Außenraum eingesetzt wird, durch einen automatisierten Reverse-Engineering-Prozess ersetzt werden. Hierfür wird das Bild einer Vorprozessierung unterworfen und Symbole, welche Fluchtwege oder Sicherheitsausrüstung darstellen, werden detektiert. In der Folge können mittels einer geeigneten Binarisierungs-Operation Objekte im Vordergrund (Wände) vom Hintergrund unterschieden werden. Auf Basis des Binärbildes wird die gesuchte Vektorinformation mit Hilfe von Skelettierung und Skelettverfolgung extrahiert. Anschließend wird das Modell durch die Überbrückung der vorher detektierten Symbole, welche Teile von Wänden oder Treppen überdecken, finalisiert. Da das resultierende Modell in einem Koordinatensystem definiert ist, welches durch das Originalbild vorgegeben wird, muss die Transformation in ein Weltkoordinatensystem, bzw. als Mindestanforderung der Maßstabsunterschied, bestimmt werden. Die Transformation wird durch einen Abgleich des Innenraummodells mit einem verfügbaren Modell der Gebäudeaußenhülle erhalten. Indem eine genäherte Stockwerkshöhe aus der Anzahl detektierter Treppenstufen berechnet wird, kann das 2D-Modell zu einem 3D-Modell erweitert werden.

Darüber hinaus kann eine Erweiterung existierender Innernaummodelle durch geometrische und semantische Informationen anhand von Benutzerfußspuren, welche mittels eines Innernaumpositionierungssystems aufgezeichnet wurden, erfolgen. Da jedoch zum jetzigen Zeitpunkt keine allgemein verfügbaren, preiswerten Systeme verfügbar sind, wird ihre Existenz in dieser Arbeit durch ein System simuliert, welches die Koppelnavigation auf Basis einer am Fuß angebrachten Inertialen Messeinheit realisiert. Zunächst werden Ansätze vorgestellt, welche dazu dienen können, die benötigten Startwerte in Bezug auf Orientierung und Position zu bestimmen, sowie Methoden, welche die Korrektur verbleibender Fehler erlauben. Letztere umfassen einen Ansatz zur Korrektur der Ausrichtung des Benutzerpfads auf Basis der Gebäudeaußenhülle und eine Methode zum Kartenabgleich unter Verwendung des Grobmodells, welches beispielsweise aus dem Fluchtplan abgeleitet wurde. Basierend auf den aufgezeichneten Benutzerfußspuren werden halbautomatische und automatische Methoden beschrieben, die eine semantische und geometrische Verfeinerung der vorliegenden Modelle erlauben. Diese reichen von der semantischen Anreicherung anhand der Analyse fotografierter Türschilder bis zur automatischen Türrekonstruktion. Des weiteren wird die Aktualisierung einzelner Raumgeometrien auf Basis einer gemeinsamen Analyse des Grobmodells, der Benutzerfußspuren und einer handgeführten low-cost Tiefenkamera wird darüber hinaus beschrieben.

Abschließend werden Arbeiten vorgestellt, welche die Innenraumkartierung unter Verwendung von Benutzerfußspuren und in einer Innenraumgrammatik gespeichertem höherwertigem Wissen über die Struktur des Gebäudeinnenraums erlauben. Aufgrund der unterschiedlichen Eigenschaften der beiden zentralen Elemente von Gebäudeinnenräumen setzt sich diese Grammatik aus einem Lindenmayer-System, welches das Korridorsystem des Stockwerks modelliert, und einer Split-Grammatik, die die Anordnung der Räume in den anderen Bereichen beschreibt, zusammen. Die Tests dieser Grammatik, welche mittels fehlerbehafteter, verteilt erfasster Benutzerfußspuren durchgeführt wurden, werden darüber hinaus dargestellt.

Contents

1 Introduction

1.1 Motivation

Location based services (LBS) such as pedestrian navigation rely on two central components - a model of the user's local environment and a positioning device. During the past decade LBS have attracted huge interest and been tapped for both commercial and private applications. This interest has been fuelled by a) the ubiquitous availability of low-cost positioning systems based on Global Navigation Systems (GNSS) like GPS, GLONASS, and the upcoming BeiDou and GALILEO, and b) freely and globally available maps from major suppliers like Google, Microsoft, Nokia and Apple or the OpenStreetMap community.

All the commercial map providers are interested in keeping their maps up to date in order to attract commercial customers interested in advertising as well as private customers whom the advertisements are aimed at. OpenStreetMap, on the other hand, being a community-driven alternative, can rely on a loyal and active mapper base ensuring its up-to-dateness. To this end, the voluntary mappers utilize manually annotated GNSS traces as well as aerial images released to the community by their owners (e.g. Bing Maps and Yahoo Maps).

Thus, with positioning and up-to-date maps being available, outdoor LBS seem to be a solved problem. However, people spend more than 90% of their time in building interiors like their home, offices or shopping malls (EPA, 1989). Consequently, this induces increased interest in the possibility to extend LBS to these currently unmapped spaces. However, neither of the two requirements for indoor LBS are fulfilled to date.

The map providers mentioned above have started to map spacious and highly frequented indoor areas such as malls and airports, either by employing manual efforts or involving the building owners[1]. The problem of a lack of indoor maps, in principle, could be left to commercial map providers who are interested in making profit of indoor LBS. Public buildings like administrative offices, however, while being of public interest in terms of e.g. routing, do not provide profit to those map providers. Automatic mapping of building interiors is still a major problem: while as-planned data mostly is available (in the form of CAD plans or the like), indoor environments are often subject to rearrangements. The registration of different data sets covering parts of the indoor environment is problematic as is the geo-referencing respectively the registration to the external building shell. These problems could be overcome using an available indoor positioning system. A very promising solution for the indoor mapping problem seems to be the transfer of community-driven approaches in order to enable crowd mapping of the vast unmapped indoor spaces. Analogous to the outdoor OpenStreetMap mapping methods, such a crowd-sourced indoor mapping method needs base data like images and traces collected using a positioning system.

[1] like in Google's approach: `https://www.google.com/maps/about/partners/indoormaps/` (last visited 2nd February 2016)

Furthermore, the positioning problem is only partially solved to date, despite the highly increased interest visible in the scientific community in the recent years. Various methods have been presented, however, the most promising seem to be found in electro-magnetic signals such as Wi-Fi, the analysis of inertial measurements or combinations of both[2]. To bound the errors of these positioning approaches, mobility models based on the environment model have been proven very effective. Thus, the positioning and mapping problems are interconnected: mapping would profit from an available indoor positioning system in the registration of different data sets, while most upcoming or available positioning systems rely on an environment model in order to deliver the highest accuracy possible.

In order to overcome the lack of digital models of building interiors, this thesis investigates methods enabling or facilitating crowd-sourced indoor mapping. To this end, it is shown that the aerial images used by Volunteered Geographic Information (VGI) communities like OpenStreetMap can be replaced by photographed evacuation plans. Evacuation plans are a compulsory part of all public buildings' emergency equipment in many countries. They may be used in a manual modelling strategy analogous to the OpenStreetMap method in the outdoor domain. However, as will be shown, they can also be analysed automatically due to their standardised features. The GNSS position traces which form the base of outdoor VGI mapping can be replaced by indoor traces collected using a foot-mounted inertial sensor. These indoor traces can serve in the geometric and semantic refinement of existing models, or be analysed using a grammar which contains rules for the architectural style of the building's interior structure.

1.2 Objectives

The aim of this thesis is the feasibility study of a system for the crowd-sourced acquisition of building interiors. To this end, the system bases on the two aforementioned types of data, represented here by images of evacuation plans and indoor position traces.

Objective 1: Make evacuation plans usable for indoor mapping, navigation and positioning support.

On the one hand, this can be done by geo-referencing the image in order to display pedestrian traces and support human way-finding. On the other hand, the main objective is the derivation of explicit polygonal models which can be used for routing purposes and be applied in the context of positioning support methods like map-matching. In contrast to the aerial images used in the manual outdoor OpenStreetMap mapping strategy, the images depict pre-processed model information about the building interior's layout, rendering an automated analysis possible. The difficulties which have to be overcome in order to reach this goal are the following: 1) there is no measurement redundancy. All the information for the derivation of the model must be extracted from the photographed evacuation plan. 2) In contrast to existing approaches which are based on scanned paper plans, the photographs suffer from a low resolution and flaws resulting from the lighting and image capturing conditions. Lastly, 3) geo-referencing and scaling should be derived automatically.

Objective 2: Develop methods for the refinement of coarse models (as e.g. stemming from the automated analysis of photographed evacuation plans) by use of pedestrian traces.

[2] apart from commercial systems basing on Bluetooth low energy beacons introduced recently

Low-cost indoor positioning systems with the needed accuracy are not available yet. However, as will be shown, expensive sensor equipment like foot-mounted inertial measurement units together with suitable signal processing methods can be used as a preview, having the sought-after characteristics like infrastructure-independence and high accuracy. As such systems deliver coordinates relative to a pre-defined initial orientation and position, these values shall be obtained e.g. by analysis of the photographed evacuation plan. The position traces, additionally improved by use of an available model of the building's external shell and the coarse model, open the possibility to refine and update the coarse model in various ways. As will be shown, semantic information can be added manually at the user's current position, but also by analysing the content of photographed doorplates. Furthermore, the model can be refined geometrically by the automatic detection of door openings in a conjoint analysis of model and user trace, as well as by processing point clouds collected using a low-cost range camera.

Objective 3: Reconstruct the indoor space using pedestrian traces and an indoor grammar.

Using knowledge about the indoor structure's layout modelled in a rule system - i.e. grammar - enables the reconstruction of full floorplans based solely on pedestrian traces. To this end, the approach is based on a large number of traces collected distributedly by people walking around in the building.

1.3 Structure of the thesis

The thesis is structured as follows: Chapter 2 gives an overview over existing indoor positioning methods, subdivided into infrastructure-based and infrastructure-independent approaches. In chapter 3, the related work in indoor mapping is presented.

According to the first objective described above, chapter 4 is dedicated to an in-depth description of the implemented processing pipeline for the reconstruction of building interiors from photographed evacuation plans. Apart from the presentation of steps like the pre-processing of the acquired image, binarisation, symbol detection and bridging, this chapter includes a short review of existing methods for the reverse-engineering of scanned paper plans.

Ojectives two and three are comprised in chapter 5. Thus, this chapter initially describes methods for the geometric update and semantic refinement of the resulting coarse models by means of an indoor positioning system. The implemented positioning system and methods supporting it are presented. Apart from manual methods, the refinement and update of the coarse model includes the annotation by semantic information using the analysis of photographed door plates and the automatic reconstruction of doors from the combination of the pedestrian trace and the coarse model. Additionally, the geometric refinement using a low-cost range camera as well as the reconstruction solely based on user traces employing an indoor grammar are presented.

Chapter 6 concludes the main part with an evaluation and discussion of the results of all presented reconstruction methods. Chapter 7 gives the conclusion and a short outlook on future research in these areas.

2 Indoor positioning and tracking

GNSSs played a major role in the success of location based services (LBS) experienced in the recent years. While the ubiquitous availability of positioning systems fosters LBS in the outdoor domain, indoor LBS are hindered by the absence of an analogously pervasive technology. However, there is a wide variety of indoor applications which could either benefit or would be rendered possible by the knowledge of a person's current position[1].

Applications like "user mobility" or "nomadic computing", i.e. digital workplaces following their users to the respective nearest computer (Want et al., 1992), lost much of their relevance in times of ubiquitous computing using smartphones and tablets. Some of the concepts, however, still remain interesting, e.g. data security scenarios such as the ones described by Blankenbach et al. (2007): automatic proximity-based unlocking of computers or the printing of confidential documents which is delayed until the authorized recipient is near the printer.

In emergency cases, the positions of both a person seeking help and the helpers can prevent injuries or fatalities. The former example is governed by directives like "Enhanced 911 (E911)" in the United States and "Enhanced 112 (E112)" in the EU with the aim of ensuring that an emergency caller's position is determined and transmitted to the action forces. For example, the US Federal Communications Commission (FCC) claims the horizontal indoor position of an emergency caller to be determinable to within 50 meters in 80% of cases in the near future (FCC, 2014). Then again, the positions of members of the action forces have to be known in order to coordinate them and guide them in situations with low visibility like smoke-filled areas.

In health care, especially in hospitals, localization and tracking of patients, personnel, and equipment are areas where commercial systems are currently put to use. Similarly, the ageing societies in post-industrial countries motivate research in the area of Assisted Living (Schindhelm and MacWilliams, 2011) where tracking of a person's exact position is part of the situation recognition. Furthermore, real-time tracking of players during indoor sports events is of interest (e.g. in basketball[2]).

Asset management as well as tracking of products, production resources, and personnel inside factories can render production more efficient, as e.g. in the SmartFactory (Zuehlke, 2010). Turning off the lights and heating in rooms which were left empty could save energy and money. Construction and building management could profit e.g. from augmented reality enhanced inspection of wires and pipes behind walls. This demands a very high accuracy of indoor positioning methods (as well as accurate and complete models).

However, the biggest interest is in commercial location-based services (LBSs), as this market has been growing for years in outdoor areas. These LBS applications include navigation and location-based reminders, as

[1] robotic applications and methods are left aside as being out of the scope of this thesis

[2] `www.sportiq.fi` (last visited 2nd February 2016)

well as location check-ins and location-based games such as geo-caching and Google's Ingress[3]. Apart from these customer-centered applications, benefits for shop owners can be found in location-based advertisement, customer tracking and customer flow analysis. On the customer side, however, many enhanced applications pose threats to personal privacy.

At the highest level, Methods for the position determination inside a building or enclosed structure can be distinguished by their autonomy from or dependence on existing infrastructure. This distinction defines the structure of this chapter which recapitulates the related work concerning infrastructure-based as well as infrastructure-independent indoor positioning methods.

In contrast to the definition found in Chen et al. (2012), where infrastructure-based methods are defined as relying on hardware actively deployed for the purpose of positioning, infrastructure-based methods here shall comprise all methods in need of external signals, hardware, or information. An overview over related work in this field is given in section 2.1.

The infrastructure-independent methods described in section 2.2, on the other hand, are self-contained. To this end, they use inertial measurement units as well as other sensors in order to implement pedestrian dead-reckoning. As the errors in dead-reckoning are likely to grow out of bounds quickly, this section furthermore outlines some aiding methods which were presented in the related work.

2.1 Infrastructure-based

Despite the relatively short period of time in which pedestrian positioning systems for building interiors have been in the focus of the research community, the number and diversity of approaches motivated a quite large number of survey papers. In the area of infrastructure-based systems, these are Hightower and Borriello (2001), Gu et al. (2009), Mautz (2009), Koyuncu and Yang (2010), Deak et al. (2012) and Stojanović and Stojanović (2014). Figure 2.1 gives an overview over infrastructure-based positioning methods with their according accuracies as available in literature respectively in the market at the time of writing of the survey paper (Mautz, 2009). It still represents a relatively complete overview apart from some excluded research fields such as approaches based on light, images, or FM radio/TV signals. Furthermore, the recent research concerning geo-magnetism fingerprinting as well as current hardware developments (e.g. bluetooth low energy) is not reflected. Most of the depicted ranges and accuracies are true up to now. However, newer developments using cell-phone signals (GSM) in the context of received signal strength fingerprinting improved the achievable accuracies to match those of WiFi-based fingerprinting.

[3] `https://www.ingress.com/` (last visited 2nd February 2016)

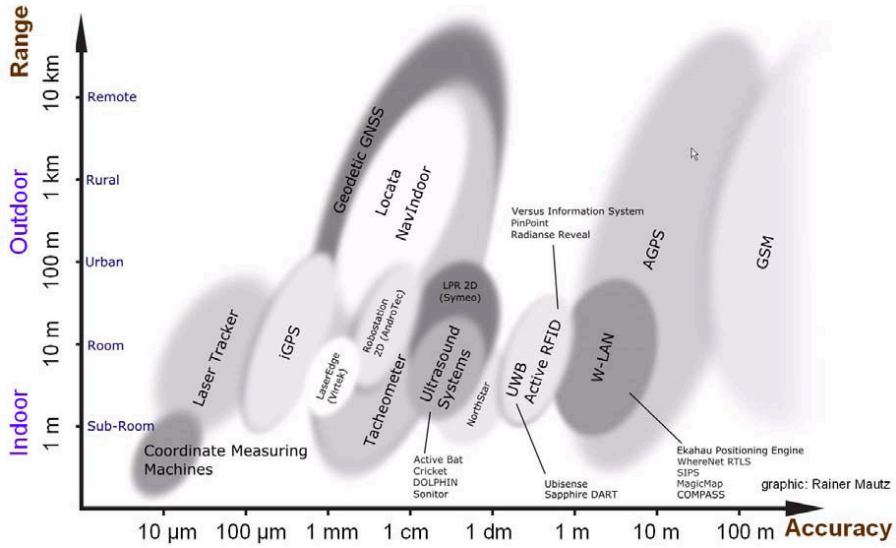


Figure 2.1: Overview of infrastructure-based indoor positioning methods and according accuracies as presented by Mautz (2009)

Apart from their range, accuracy, and the used signals, infrastructure-based indoor positioning methods can be further subdivided according to various characteristics (compare also Hightower and Borriello, 2001):

- tracking vs. ego-motion positioning: in tracking systems the coordinates of the subject to be positioned are computed by systems in the infrastructure. Ego-motion positioning systems, on the other hand, allow their determination by the subject itself, thus offering a higher level of privacy
- the system's ability to deliver coordinates continuously or only at discrete points
- the system's ability to deliver metric or symbolic coordinates (e.g. room identifiers)
- method used for the coordinate determination:
  - proximity: discrete location updates whenever the system detects proximity to a beacon
  - angulation (angle-of-arrival, AOA): coordinates are computed by intersecting lines reconstructed from the different angles in which the signals arrive. Mostly used in tracking systems due to the necessity of rotating antennas or antenna arrays
  - lateration: coordinates are computed by intersection of circles/spheres (time-of-arrival, TOA) or hyperboles (time-difference-of-arrival, TDOA) which are reconstructed using the runtimes or runtime differences of the signals, respectively. While TOA needs sender-receiver synchronization, TDOA relies on sender synchronization which can be established more easily as only affecting the infrastructure
  - fingerprinting: in an offline phase, a map of fingerprints (e.g. consisting of received signal strengths) at known positions is constructed. During the online phase, coordinates are computed using feature matching
- necessity of a direct line-of-sight between sender and receiver (line-of-sight vs. NLOS), where the definition of the line-of-sight is dependent on the used signals

- costs and scalability

There are various, obvious ways of obtaining a person's indoor position without intermediate measurements. Examples are the push of a button, use of a key card, drawing money at an ATM or paying with a credit card in combination with the knowledge of the according facilities' positions (Hightower and Borriello, 2001).

Another straightforward method for direct pedestrian localization was presented in Orr and Abowd (2000): the SmartFloor. As the name suggests, it consists of floor tiles measuring the ground reaction force profiles of the pedestrians' steps. While this enables a direct position determination, the major problem to be solved is the identification and tracking of individual users of the system. The identification is carried out using features of the ground reaction force profiles like mean, standard deviation, total area under the profile curve etc., resulting in an accuracy of 93% independent of differing shoes, while the authors left tracking for future work. Building on direct, tactile measurements, the system does not scale well and the installation costs are high.

Apart from these direct methods, all other infrastructure-based indoor positioning systems are based on contactless measurements. In the remaining subsections of this section, an overview of the wide range of related work is given, covering light-based methods as well as methods using images, sound, and electromagnetic signals. Electro-magnetic signals can be further subdivided into TV, FM radio, or cell phone signals as well as GPS, RFID, bluetooth, UltraWideBand, magnetism, and WiFi.

2.1.1 Light

There are various ways in which light can be leveraged for the determination of indoor positions. They can be distinguished by the use of visible vs. infrared light as well as their characteristics such as ego-motion vs. tracking systems using e.g. stereo-vision and their use of dedicated deployed lights vs. fingerprinting of the available ambient light. Obviously, all of these methods need a direct line-of-sight in order to work.

The "Active Badge" system (Want et al., 1992) was one of the earliest efforts to enable user tracking inside of buildings. Its badges transmit a unique code using pulse-width modulated infrared light every 15 seconds for 1/10 of a second. This signal is captured by sensors distributed throughout the building which are polled by a server. In order to reach the goal of room-granular localization infrared signals were chosen because they are blocked by walls. In contrast to triangulating methods signal reflections are welcome in this system. Aitenbichler and Muhlhauser (2003) similarly use infrared LEDs emitting an identifier code, however, tracking the LEDs by stereo cameras with known intrinsic and extrinsic parameters. This way, they reach an accuracy of 16.67 centimeters in a lecture hall of 15.1x9 square metres. Because of the high effort for the installation of the infrastructure and the limited range, on the one hand, and the achievable high accuracy, on the other hand, commercial systems employing infrared light can mostly be found in the fields of motion capturing, asset management, or health care. The systems offered by OptoTrak[4] and Versus[5] shall be named here exemplarily.

[4] `http://www.ndigital.com/lifesciences/products-motioncapturesystems.php` (last visited 2nd February 2016)

[5] `www.versustech.com` (last visited 2nd February 2016)

In the aforementioned systems, infrared light sources installed solely for this purpose are tracked and used to compute the position in the infrastructure. In contrast, ego-motion positioning systems use already available sources of visible light or modulate signals onto the light emitted by LEDs installed for the purpose of lighting.

A very early attempt of using visible light fingerprinting can be found in Golding and Lesh (1999). However, this approach uses fingerprints which are a combination of light intensity, accelerations, magnetometer readings and temperature. Using handcrafted high-level features derived from these data sources, the authors report a room detection error rate of 2%, without those features it is around 50%. Ravi and Iftode (2007) describe visible light fingerprinting for the purpose of room identification in an environment with static lighting, i.e. no windows. In a test bed of 20 rooms, their approach is able to achieve up to 91.2% respectively 71.6% detection accuracy using Bayesian or range-max fingerprinting and a sensor worn either on a hat or as a pendant. Despite not being a genuine fingerprinting approach and the need to be combined with an available PDR dead-reckoning method, Jiménez et al. (2013) deserves note here as well. They replace the tedious process of fingerprint collection by a known map of the positions and orientations of all lights installed in the area. The information that the user passed under a switched-on light is fed into a particle filter together with the PDR results (and optionally WiFi fingerprints, magnetometer data and a map). In the best case, this method delivers coordinates with an error smaller than one metre (70%).

Today, LEDs are considered the future in lighting and are expected to soon replace light bulbs. In addition to being much more efficient than light bulbs, LEDs can produce precisely controllable visible light signals which can be used for positioning purposes. Zhou et al. (2012) proved in simulation that position estimation can be achieved using signals modulated onto visible LED light used for illumination. By placing four LEDs in the upper corners of a cube, which use Time Division Multiplexing to alternately emit a position code or just illuminate, a simulated accuracy of 0.5 to 30 millimetres (depending on the noise) can be achieved. Nakajima and Haruyama (2013) leverage LED light communication in combination with magnetometer data for the benefit of indoor navigation for visually impaired people. The identification of the LED under which the user is standing by the emitted signal triggers a position lookup in a database. In order to provide the accurate orientation, available magnetometer readings are corrected using values collected in an offline phase.

Presumably, a similar method (except for the magnetometer corrections) is used by the commercial system offered by ByteLight[6]. By equipping LED light bulbs with a processing unit and providing a smartphone application which decodes the signals captured by the mobile devices' front cameras, they claim sub-metre accuracy. A similar system is in development at Philips[7].

2.1.2 Image-based methods

An extensive part of the related work in indoor positioning concentrates on image-based methods, despite the unfavourable characteristics such as bad lighting and missing textures. Mautz and Tilch (2011) give an overview of the works published until the year 2011, grouping them according to the kind of reference data

[6] `www.bytelight.com` (last visited 2nd February 2016)

[7] `http://www.gim-international.com/news/positioning/gps/id7900-philips_launches_led_lighting_indoor_navigation.html` (last visited 2nd February 2016)

used. The most important categories are building models or floorplans, images (also called content-based image retrieval, CBIR), coded markers, and projected markers/patterns. As mentioned in the previous section, light-based methods partly employ cameras for the tracking of light signals. Furthermore, the concept of ego-motion localization by matching online data to an available map overlaps with the map-based positioning methods described in section 2.1.5. Thus, additional related work for this concept can be found there.

In an early paper, Krumm et al. (2000) present the person localization and tracking methods behind Microsoft's EasyLiving research project. In their experiments, the authors use two stereo cameras for a room and determine the 2D transformation parameters between them as well as the background which has to be subtracted for a robust blob detection. The detected blobs are merged to person-shaped areas which are identified for tracking using their color histograms. The accuracy of the 2D position is quoted as 10 centimetres at a frequency of 3.5Hz. Ten years later, the emergence of the Microsoft Kinect marked the arrival of stereo-imaging based people tracking in the consumer market. The Kinect software can detect and track skeletons of up to six persons in a range between 0.8 and 4 meters, however, it is not possible to recognize individuals once they leave and re-enter this area or the visibility cone of another Kinect sensor. Using the RGB image it would be possible to overcome this flaw, but this comes at the cost of harming the users' privacy. Additionally, the use of multiple Kinect sensors will lead to interferences.

Quigley et al. (2010) present a complete system consisting of methods for map-building, collection of reference data and a localization method basing on low-cost smartphone sensors. In the map-building step, a robot equipped with laser scanners builds the required map using Simultaneous Localization and Mapping (SLAM), before a human operator visits the mapped area again with a sensor backpack and the consumer-grade device to be used during localization. The online localization uses the smartphone camera by extracting SURF keypoints (looked up in the reference data), and combines this information with WiFi signal strength based positioning. This is complemented by a particle filtering operation in order to reach submeter precision.

Elloumi et al. (2013) fully concentrate on images and their vanishing points. In the offline phase the vanishing points in each key frame are extracted using the Canny edge detector (Canny, 1986), the probabilistic Hough transform (Kiryati et al., 1991) and a RANSAC (RANdom SAmple Consensus) filtering operation. From the information about the vanishing points of the scene, the camera orientation is estimated. For the localization during the online phase, the closest key frame is identified using the Harris corner detector and zero-mean normalized cross-correlation. The position associated with the key frame and an additional angular deviation estimation result in the final position. The experiments carried out by use of a calibrated smartphone camera reveal errors of around 0.5 metres, even at higher walking speeds and with an offset from the offline phase's path.

An example for image-based positioning by matching to an available floorplan can be found in Hile and Borriello (2008). The authors describe the detection of the floor in the image captured by a camera phone. The position estimation with respect to an available floorplan is carried out by matching points of both representations of the floor using a RANSAC method.

Commercially, none of these methods have been put to use. Instead, companies like PinPoint[8] employ camera scanned QR code tags in order to provide the user's current (discrete) position with respect to a map.

2.1.3 Sound

Examples from animality - e.g. whales and bats - demonstrate the applicability of (ultra)sound for distance measurements. The knowledge of the sender's or receiver's location(s) together with the distance measurement enables the position computation for the navigating subject using lateration in sound-based indoor positioning systems. Furthermore, the specific acoustic background fingerprint of rooms can be used for room-level positioning.

While Madhavapeddy et al. (2003) illustrate the possibility to leverage audible sound signals for position determination with room granularity, a method for absolute positioning was presented by Mandal et al. (2005), aptly named "Beep". Its infrastructure consists of microphones distributed throughout the desired working volume of the system and an existing Wi-Fi network, at the client no additional hardware to the one existing e.g. in off-the-shelf mobile phones is needed. Upon the user request for a position update, time synchronization over the Wi-Fi network is established before the client emits an audio signal with pre-defined frequency and duration. After detection using the microphones, the system's central server estimates the time of flight and computes the distances, which can be used to compute the client's position using 3D multilateration up to an accuracy of 0.6 metres in 97% of cases in a testbed of 10x5 square meters (compare figure 2.2a). Hoflinger et al. (2012) use the same principle, but add a self-calibration step instead of manual referencing of the receivers. Their ASSIST system reaches 0.34 metres of error.

Rishabh et al. (2012) add audible signals to controlled ambient sounds like music using pseudo-random noise. The signal recorded by the client is sent to the server which performs cross-correlation in order to determine the time-of-arrival and estimate the client's location and time offset. Due to the unfavourable signal-to-noise ratio, long recordings or continuous tracking are necessary, however, a positioning error of 0.2 metres can be achieved.

More common than audible sound, however, is the use of ultrasound. The most obvious motivation for its use over audible sound is the removal of user disturbance. Furthermore, the signals are less prone to interferences by talking people, music or other human-generated sound. An overview over various methods can be found in Holm (2012).

Harter et al. (2002) present a positioning system based on ultrasound as part of their context-aware system enabling digital working environments to follow their users to the respectively nearest available workstation. The clients are mobile units, the so-called Bats (see figure 2.2b), which contain a radio transmitter and an ultrasonic transducer and have a unique identifier. The infrastructure consists of a base station and receivers placed on the ceilings. The base station periodically resets the receivers and transmits a radio message containing the identifier of the Bat which is requested to emit an ultrasonic impulse. The following distance computation from the signal's travelling times takes the ambient temperature into account. Additionally, an outlier rejection was incorporated into the multilateration position computation. The orientations of clients can be determined using multiple Bats or by the use of the object's ultrasonic shadow.

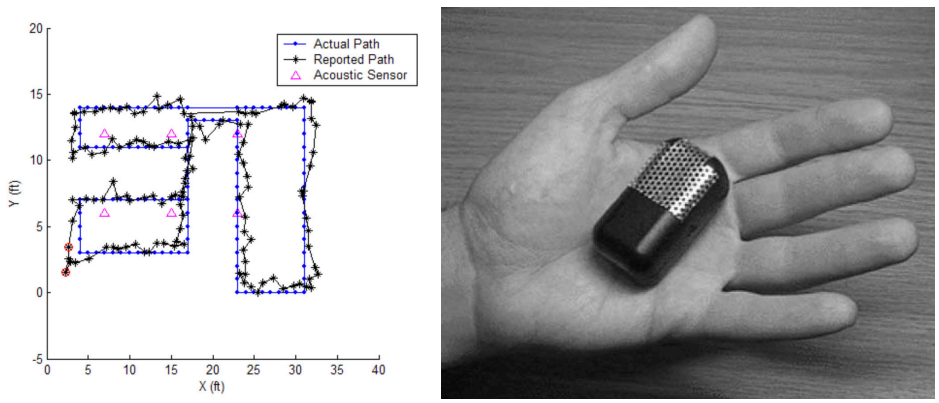
[8] `http://www.pinpoint-navigation.com/?l=en` (last visited 2nd February 2016)

The evaluation in a 10m³ testbed yielded an accuracy of 9 centimeters. Using mobile phones as transmitters (instead of dedicated mobile units) and Time Difference of Arrival in order to eliminate the need for synchronization, Filonenko et al. (2013) reach 8.2-18.1 centimetres of error. Errors of 4.2 centimetres are reported by Sato et al. (2011), who densely pack the receivers and thus have to increase the distance measurement accuracy using the extended phase accordance method.

Instead of active clients answering to a radio frequency (RF) request, Priyantha et al. (2000)'s "Cricket" system makes use of the difference between the speeds of sound and light and also adds the transmission of the ultrasonic pulse to the beacons. The beacons concurrently send a RF and ultrasonic pulse, the clients, upon receiving the RF pulse, turn on the ultrasonic receiver and listen for the ultrasonic pulse. The distance can be derived from the difference between the respective times-of-arrival of both pulses. To ensure privacy and ease of implementation, however, the locations of the beacons are unknown, and only the nearest beacon is identified. Thus, the system delivers symbolic coordinates and the positioning accuracy is dependent on the arrangement of the beacons, a 2.4x2.4 metres grid was proven to be feasible. Using smartphones as receivers, Lopes et al. (2013) can establish the system's synchronization over Wi-Fi. Thus, 7.3 centimetres of error in a 7x7 square metres test bed can be reached.

Taking the principle described in Priyantha et al. (2000) one step further, the DOLPHIN system presented in Fukuju et al. (2003) consists of dual-purpose nodes acting as senders and receivers, some of which act as reference nodes. The signals from the reference nodes as well as the previously localized nodes are used to subsequently compute the position of other nodes using multilateration. The system was tested in a small testbed of 2x2.5 square metres and yielded an error of less than 5 centimetres for nodes directly positioned by reference nodes. All other nodes suffer from error propagation, which results in 10-15 centimetres of error for nodes on the second layer.

Moving away from lateration, Tarzia et al. (2011) use the acoustic background spectrum audible in every room in a fingerprinting method. Therefore, they compute the power spectogram of a 30-second sound recording at different positions in the room, ensure transient noise rejection, and normalize it. In the online phase, the user's position is determined as the nearest neighbour in the fingerprint database. The system is able to identify the correct room in 69% of cases, independent of the use of professional audio recording equipment or their iOS BatPhone implementation. Probably more significant is the achieved 92% adjacent room distinction which makes the method complementary to Wi-Fi fingerprinting, promising good results for a system combining both methods.



(a) (b)

Figure 2.2: Sound-based systems: a) results from the Beep system presented by Mandal et al. (2005), b) an Active Bat unit as presented in Harter et al. (2002)

Medina et al. (2012), in contrast, evaluate the use of ultrasound in a RSSI-based system, reaching a standard deviation of 4.5 centimetres using RSSI values derived by signal propagation modeling. As the cell size of ultrasound systems does not exceed single rooms, genuine fingerprinting methods are rare. However, Ben-Moshe et al. (2013) developed a method basing on TDOA fingerprints and evaluted it using off-the-shelf smartphones, speakers and a quad channel sound card. They come to the conclusion, that, while TDOA measurements are insufficiently accurate for the purpose of lateration, they are highly stable and usable for fingerprinting.

Sonitor[9] offers a commercial system basing on ultrasound that is aimed at asset and patient tracking in hospitals. While mainly delivering room granularity, the company states that an accuracy of 30 centimetres is feasible.

2.1.4 Electro-magnetic signals

As shown above, indoor positioning systems using light or sound can deliver high accuracies. However, in the confined space in buidling floors divided into rooms and corridors, their range is restricted by the limited line-of-sight. Electro-magnetic signals, on the other hand, are able to penetrate walls and thus enlarge the area covered by any positioning system based on them. Furthermore, for many signals, e.g. bluetooth, FM radio or WiFi signals, senders are either cheap (off-the-shelf) or already installed for communication purposes. Thus, receivers for a range of electro-magnetic signals are already found in common modern hardware such as smart phones.

GPS and Pseudolites

Most publications reporting on scientific progress in the field of indoor positioning motivate the described efforts by the unavailability of GNSSs in indoor environments. Due to signal attenuation, interferences, and multipath effects, GNSS signals either completely vanish in noise, cannot be tracked in order to derive stable position fixes, or deliver results prone to very high errors. Lachapelle (2004) assesses these generally known facts regarding the signals' availability, accuracy, continuity and reliability. The low availability of satellite signals in indoor environments can be overcome by an available clock constraint and a low-cost MEMS barometer, resulting in two satellites being sufficient for a 2D position fix. The largest influences on accuracy are high-noise, echo-only or multipath signals as well as the degraded geometry of the available satellite constellations. Continuity depends on rapid temporal and spatial decorrelations of multipath signals and rapidly changing satellite constellations. Using four typical building interiors, Lachapelle (2004) describes the achievable accuracies as around 5 metres (wooden building, kinematic measurement mode), 10-30 metres (north american residence, static), 25-45 metres (sports arena made of concrete, kinematic) and 75 metres (office building, static). Errors of these dimensions are confirmed by Eissfeller et al. (2005) as well as Schon and Bielenberg (2008).

While reliably achievable indoor positioning accuracies in this scale would be acceptable e.g. in emergency, safety-of-life scenarios in order to locate the building the user is in, most other intended purposes for indoor positioning require much higher accuracies. The use of Assisted GNSS (AGNSS), basing on Satellite Based

[9] `www.sonitor.com` (last visited 2nd February 2016)

Augmentation Systems like EGNOS and MSAS, is reported to enable indoor accuracies down to 10 metres (Mautz, 2009). For further improvement, terrestrial time-synchronized transceivers - so-called pseudolites - have to be used. Kuusniemi et al. (2012) present an evaluation of the combination of high-sensivity GNSS and pseudolites. In their scenario, the pseudolites are used as beacons, i.e. contribute to the particle filter only by their proximity information. In a three floor building, the use of pseudolites and a simple map-matching approach results in an accuracy of 7.6/6.6 metres (1st/3rd floor) in contrast to 15.8/11.7 metres for the GNSS-only solution. Locata[10] is one example for commercially available pseudolites systems. According to the tests presented in Rizos et al. (2013) (open indoor space, using a directional antenna, 2D positioning), the achievable accuracy is 1.5 centimetres (95%).

RFID and Bluetooth

With SpotON, Hightower et al. (2000) presents an RFID-(radio frequency identification)-based system using off-the-shelf base stations and active badges. The signal strengths measured by the base stations are reported to a central server which computes the position using lateration and the inverse square law. According to the authors, the evaluated 3-metre-accuracy is due to the limited resolution of the signal strength measurements. Experiments with a prototype system with new hardware indicated sub-metre accuracy. Similarly, Ni et al. (2004) employ commercially available hardware, however, try to improve the accuracy by adding reference tags with known positions. This method keeps the needed number of more expensive readers low and allows for a dynamic assessment of the current environmental conditions (interferences induced by moving people etc.). Using four readers and 16 reference tags, one meter of error (50%) is reached in a 4x9 square metres testbed.

Commercial systems using RFID are sold by vendors including WhereNet[11], Radianse[12] and Versus[13]. The primary designated uses for these systems are found in asset management, retail, and the health system, i.e. asset and patient tracking in hospitals. WhereNet specifies 1.6 metres (95%) as location accuracy of the WhereLAN system.

Before the emergence of Bluetooth low-energy (BLE), commercial systems leveraging bluetooth signals were sold by Topaz[14] and Quuppa[15] (the former indoor positioning team of Nokia). Topaz states a room level accuracy of up to 2 metres and a latency of 15-30 seconds. In contrast, Quuppa's High Accuracy Indoor Positioning system is able to achieve 0.3-1 metres in real-time using AOA and dedicated tags. Soon, however, the same accuracy level is expected to be achieved with off-the-shelf BLE-enabled handsets.

The very low energy consumption of hardware following the BLE standard enables the production of beacons which run for several months or even years on a single coin cell battery. Furthermore, BLE receivers barely affect the run times of smartphones or similar devices and can be left running all the time. Combining these

[10] `http://www.locata.com` (last visited 2nd February 2016)

[11] `http://www.zebra.com/gb/en.html` (last visited 2nd February 2016)

[12] `www.radianse.com` (last visited 2nd February 2016)

[13] `www.versustech.com` (last visited 2nd February 2016)

[14] `http://www.tadlys.co.il/` (last visited 2nd February 2016)

[15] `http://quuppa.com/` (last visited 2nd February 2016)

two characteristics, smartphones in the proximity of a beacon can detect it and location-based events attached to the beacon like e.g. an advertisement or a price discount can be triggered. If the IDs of the beacons are not included in a database stored locally in the smartphone, however, their meaning has to be requested from a network database, which reveals the user's position. While this raises privacy issues on the user side, it is a welcome feature for shop owners eager to analyse customer behaviour. Proximity-based identification is the intended use of BLE beacons, however, first experiments employing them in more accurate multilateration strategies were presented, in research (Salas, 2014) and also commercially available[16]. BLE beacons are sold by Apple (iBeacons), estimote[17], stick'n'find[18] and qualcom[19]. Recently, Samsung joined with their Proximity[20] system.

UltraWideBand (UWB)

According to the FCC (2002), UltraWideBand is a frequency band with absolute bandwidth wider than 500MHz or a relative bandwidth wider than one fourth of its mean frequency. Blankenbach et al. (2007) furthermore mention that, while the frequency band is license free in the United States, power restrictions apply due to the overlap with other frequency bands (e.g. WiFi). Despite these power restrictions, UWB bears a low sensitivity to interferences, multipath effects, and affords a good penetration of materials. Thus, this signal band seems to be an ideal base for indoor positioning systems.

As an example of ongoing UWB indoor positioning research, Blankenbach et al. (2007) describe experiments covering distance measurements using UWB without obstacles as well as through different materials, resulting in differences to ground truth measurements of below 10 centimetres up to 30 centimetres, respectively. Concerning the feasibility of 3D positioning in indoor environments, the authors give a short description of an experiment delivering errors in the centimetre range.

Due to the characteristics of UWB, there are some commercial systems already available in the market, e.g. by Zebra Technologies[21], BlinkSight[22] or UbiSense[23]. For example, the UbiSense Real-Time Location System (RTLS) leverages a combination of TOA and AOA in order to track assets or personnel equipped with UWB emitters in 3D with an accuracy of down to 30 centimetres and a high precision and reliability.

FM radio, TV and cell phone signals

The main advantage of the use of FM radio, TV and cell phone (GSM, UMTS, LTE) signals for positioning purposes lies in the fact that their frequencies are licensed. Thus, they do not interfere with signals emitted

[16] `http://www.nextome.org/en/index.php` (last visited 2nd February 2016)

[17] `http://estimote.com/`

[18] `https://www.sticknfind.com/` (last visited 2nd February 2016)

[19] `https://www.gimbal.com/` (last visited 2nd February 2016)

[20] `https://placedge.samsung.com/` (last visited 2nd February 2016)

[21] `http://www.zebra.com/id/zebra/na/en/index/products/location/ultra_wideband.htm` (last visited 2nd February 2016)

[22] `http://www.blinksight.com/` (last visited 2nd February 2016)

[23] `http://www.ubisense.net/en/products-and-services/rtls-products.html` (last visited 2nd February 2016)

by other electronic devices such as microwave ovens. Because their emitters are located outside of the building the position shall be computed in, the signals are more likely to be available in events of emergency than those of locally installed systems. Lastly, the networks are reconfigured very infrequently and thus the signals should bear a better long-time stability than e.g. WiFi.

Despite being licensed, the transmission of low-power short-range FM radio signals is permitted in many countries. Matic et al. (2010) and Popleteev (2011) make use of this permission, deploying FM radio transmitters in the measuring volume and using an off-the-shelf smartphone as the client of their fingerprinting-based indoor positioning system. Matic et al. (2010) describe position estimation experiments using k nearest neighbours as well as Gaussian process regression, reaching accuracies of 3.88 metres respectively 2.65 metres (95%). In combination with Wi-Fi fingerprints, these results can be improved to 0.85 metres (95%).

In contrast to this approach which is mainly usable for the low-cost improvement of Wi-Fi fingerprinting, Moghtadaiee et al. (2011) present results from a similar approach using only available FM radio signals. Taking offline-phase sample measurements at 150 points for a 11x23m² test bed, they reach an accuracy of 6.4 metres (95%). Chen et al. (2012), on the other hand, augment the available FM fingerprints with additional information about the signal's physical layer: signal-to-noise ratio, multipath and frequency offset. Concentrating on room-level positioning, because this information can be crowdsourced using people checking in to stores, they also assess fine-grain positioning with fingerprints sampled at 30 centimetres distance. While positioning using Wi-Fi reaches an accuracy of around 3 meters in this scenario, FM fingerprints deliver an impressive accuracy of 30 centimetres in 90% of cases.

Rabinowitz and Spilker (2005) describe a system building on TV signals in combination with AGPS, which - reaching 30-50 metres of indoor accuracy - is supposed to be E911 compliant. The authors tried to commercialize the system in a company called Rosum, however, without success.

Cell phone signals - GSM as well as UMTS, LTE - are another class of signals which are ubiquitously available. Depending on the country, GSM operates on different frequency bands in the range between 380MHz and 2000MHz, e.g. in Germany on 900MHz and 1800MHz, which are subdivided into physical channels[24], which in turn are split into logical channels[25]. GSM cells are defined by the areas covered by the different antennas forming a base station. The accuracies of GSM positioning methods Mautz (2009) presents (see figure 2.1) refer to network-based methods as described in Schwieger (2007).

More recent research, however, builds on fingerprinting methods employing cell phone signals. While earlier work on outdoor positioning only used the strongest or the six strongest cells for fingerprinting, Otsason et al. (2005) present "wide signal strength fingerprints" consisting of the six strongest cells extended by readings from 29 additional channels. For the online position estimation, they use k nearest neighbours search, additionally geographically clustered by k-means. The evaluation comprises floor classification and within-floor positioning using GSM as well as Wi-Fi fingerprints for comparison. The within-floor GSM positioning shows an accuracy similar to Wi-Fi between 2.5 and 5.4 metres (50%). Denby et al. (2009), despite using the full GSM band for fingerprinting and more sophisticated position retrieval techniques like Support Vector Machines and Gaussian Processes, only evaluated the room identification potential of their

[24] using Frequency Division Multiple Access (FDMA)

[25] using Time Division Multiple Access (TDMA)

system, reaching 97.8% in a testbed with five rooms. Birkel and Weber (2012) deploy short-range UMTS senders at the same locations as Wi-Fi access points, reaching an accuracy of 7 metres (80%) with a received signal strength fingerprinting method.

Magnetism

Due to the various interferences produced by electronic installations, e.g. elevator motors or electro-magnetic fields, magnetometers measuring the earth's magnetic field tend to be of marginal value for the orientation determination in building interiors. However, it has been shown that artificial magnetic fields as well as fingerprinting approaches basing on the local perturbances of the geomagnetic field can be leveraged for indoor positioning.

One solution found in related work consists of utilizing synthetic magnetic fields. The foundations to position estimation using these fields were published by Raab et al. (1979), however, with the emergence of 3-axis magnetometers as standard sensor equipment of smartphones this method was revisited by Blankenbach and Norrdine (2010); Norrdine and Blankenbach (2013). As infrastructure, static inductors are set up which sequentially emit artificial magnetic fields following the Time Division Multiplexing (TDOA) method. In order to eliminate the influence of interfering fields, a differential measuring principle is applied by sequential reversal of the receiver's polarity. In the authors' 2D experiments, three static inductors were sufficient to reach a position error of up to 0.5 metres, even with low-cost smartphone sensors.

Commercial positioning systems employing artificial magnetic fields are available, e.g. by Ascension[26] and Polhemus[27]. However, these aim at low-range positioning in medical and motion capturing applications and therefore are not applicable for indoor positioning.

In contrast to generating artificial fields in order to provide a stable, known field to mobile magnetometers, there have been efforts to leverage the effects of local variations interfering with geomagnetism. For example, Chung et al. (2011) describe an experimental system building on a magnetic fingerprint map. As magnetometer readings alone provide at most a three-dimensional feature vector, the optimal number of sensors and their displacement are identified experimentally. Using three testbeds, the fingerprints are collected in various orientations on a 0.6 metres grid. During the online phase the position is selected as the least RMS neighbour, leading to a mean error of between 2.84 and 6.28 metres for the different testbeds, while 75% of all measurements show less than one metre of error. Adding a rudimentary mobility model causes an increase to one metre error in 88% of cases. The authors also examined the long-term stability as well as the sensitivity to nearby objects like elevators, concluding that in combination with Wi-Fi fingerprint data an accuracy improvement to 0.45 metres would be possible. Li et al. (2012) report similar results, in addition they evaluated the feasibility to detect the correct floor in a 5-floor building to be between 72 and 90%. Song et al. (2012) come to the conclusion, that geo-magnetism based systems are able to outperform Wi-Fi fingerprint-based systems.

A commercial system leveraging fingerprints of the geomagnetic field's disturbances is offered by indoorAtlas[28]. The web page presents tutorials explaining how to upload and georeference an available indoor map

[26] `http://www.ascension-tech.com/` (last visited 2nd February 2016)

[27] `http://polhemus.com/` (last visited 2nd February 2016)

[28] `http://www.indoorAtlas.com` (last visited 2nd February 2016)

in a similar way as the Google Indoor Maps method. Additionally, the collection of the needed data using a smartphone is covered. In the FAQ the achievable accuracy using off-the-shelf hardware is specified as less than 3 metres (90%).

Wi-Fi and combinations of electro-magnetic signals

The first published work on indoor positioning leveraging WiFi signal strengths describes the RADAR system Bahl and Padmanabhan (2000), which evaluated the potential of fingerprints as well as signal propagation modelling. Using three access points deployed for this purpose, signal strength readings were recorded in four orientations at 70 points in 1000m² during the offline phase. Concerning the fingerprinting-based positioning, the authors evaluate three different algorithms: the single nearest neighbour in signal space (error: 2.94 metres, 50%), k nearest neighbours (2.75 metres, 50%) and k nearest neighbours using the maximum signal strength across all orientations at each sample point (2.67 metres, 50%). Furthermore, the paper describes various experiments covering the number of data points, the number of samples per data point, the user orientation and a moving vs. a stationary user. Additionally, the applicability of signal propagation modelling is demonstrated using a CAD model and a wall attenuation factor derived experimentally. This model provides the expected signal strength which is computed for a set of grid points, resulting in a positioning error of 4.3 metres (50%, using the single nearest neighbour approach).

Lassabe et al. (2011) compare various signal strength map-based, wave propagation modeling (i.e. lateration) methods and hybrid approaches, where the signal strength maps are either collected or simulated. According to this comparison, Bahl and Padmanabhan (2000)'s RADAR delivers the best results with simulated signal strength and hybrid systems following close.

Signal propagation modeling approaches like the one described in the RADAR paper need accurate access point positions in order to function. While these positions often are assumed to be known or measured by manual surveying, there have been efforts to determine them from received signal strengths. Cho et al. (2012), for example, describe the estimation of access point positions using Gaussian process regression in addition to the weighed centroid localization as used in other related work.

In their COMPASS system, King et al. (2006) augment the fingerprint collection as well as the positioning by orientation information, thus reducing the search space. This method improves the accuracy to 1.65 metres on average, tested with fingerprints of one metre spacing. In the Horus paper by Youssef and Agrawala (2005), the authors describe experiments on temporal, large-scale and small-scale variations of the received signal strengths as well as methods for the compensation of those variations. One of those methods is a clustering operation as final step of the offline phase, building sets of locations which share a common set of visible access points. Using a training point spacing of between 1.52 and 2.13 metres, the Horus system is able to determine positions with errors between 0.86 and 1.32 metres (90%). There have been efforts to diminish the search space by hierarchical structuring of the data by buildings, floors and rooms (Marques et al., 2012).

The extensive amount of work that has to be invested for the fingerprint collection during the offline phase (and which has to be repeated each time the access point configuration changes) is the main downside of fingerprinting-based approaches. Thus, one current research direction aims at removing this necessity by enabling the crowd-sourced, smartphone-based fingerprint collection. Ledlie et al. (2012) use hierarchical structuring, furthermore accounting for temporal variations in the signal space as well as dynamically

increasing the scan rate on detected movement. Yang et al. (2012) use the smartphone's accelerometers to identify footsteps between consecutive fingerprint scans, thus building a floorplan in fingerprint space which is segmented into corridors, rooms and doors by clustering. This floorplan is matched to a CAD floorplan using the door information, establishing the connection between fingerprints and real-world coordinates. During the online phase the localization errors are similar to those found in state-of-the-art manual fingerprinting methods, as revealed by their evaluation. Wang et al. (2012) employ accelerometers/magnetometers as well as Wi-Fi found in a modern smart phone in order to identify landmarks usable for the position estimation. To this end, they first analyze an available floorplan, identifying so-called seed landmarks where the user has to behave in a predictable way, e.g. stairs, elevators and escalators (see also Alzantot and Youssef, 2012a). In the offline phase additional "organic" landmarks are detected like Wi-Fi landmarks (small areas with a unique set of visible access points) and accelerometer/magnetometer landmarks (unique disturbances in the magnetic field, elevator signatures in the accelerometer data). This method enables to achieve median errors of 1.69 metres during the online phase.

Apart from temporal and small-scale signal variations, the low dimension of Wi-Fi received signal strength fingerprints causes ambiguities. Sen et al. (2012) try to overcome this problem in their PinLoc system by using the PHY layer information instead, i.e. the multipath channel information between the access point and device. To this end, a Roomba robot is moving randomly in a "1m x 1m" spot during the offline phase, collecting the necessary fingerprints. During the online phase, this spot can be identified correctly in 89% of cases.

In addition to the WhereNet system already mentioned in the RFID section, the most prominent commercial system is Ekahau[29], providing room-level accuracy in at least 90% of cases. Additionally, systems like the NAO Campus system sold by PoleStar[30] or the service offered by Skyhook Wireless[31] are noted. Furthermore, the fingerprinting-based "awiloc" system developed by Fraunhofer IIS can be mentioned[32]. Commercial systems leveraging a mixture of electro-magnetic signals - among others - are available from InfSoft[33] as well as Roomaps[34]. According to InfSoft's web page, in addition to WiFi, BLE and cellular phone signals, the data from the complete range of sensor equipment found in a modern smartphone - magnetometers, barometers, accelerometers and gyroscopes - is included in the positioning system.

2.1.5 Maps

Some of the aforementioned approaches use maps consisting of features (fingerprints) which are not human-readable. However, to a certain extent human-readable maps or 3D models can not only serve for routing or visualization purposes, but also for as solution to the localization problem.

Despite the authors' claim of an infrastructure-free positioning system, Köhler et al. (2007)'s approach shall be named here. Their structured-light stereo system additionally contains a 3-axis accelerometer and mag-

[29] `http://www.ekahau.com/` (last visited 2nd February 2016)

[30] `http://www.polestar.eu/en/products/indoor-positioning.html` (last visited 2nd February 2016)

[31] `http://www.skyhookwireless.com/` (last visited 2nd February 2016)

[32] `http://www.iis.fraunhofer.de/de/ff/lok/tech/feldstaerke/rssi.html` (last visited 2nd February 2016)

[33] `http://www.infsoft.de/` (last visited 2nd February 2016)

[34] `http://www.roomaps.com/en/` (last visited 2nd February 2016)

netometer. In the point cloud delivered from this sensor, planes and corners are located and the relative 3D position is computed with an evaluated accuracy of 15.8 centimeters. For a position determination relative to a map the authors propose the additional use of a room-level-accurate positioning method. Together with the orientation provided by the magnetometer, this could disambiguate the location of the observed corner in a map, enabling the computation of coordinates relative to the map. Boehm and Pattinson (2010) present a related method, however, basing on an off-the-shelf range camera and a raw laser scan point cloud which represents the needed map. As is the case with Köhler et al. (2007)'s method, this approach would also need a coarse position and orientation estimate to be used as a fully functional indoor positioning method. The achievable accuracy of the computed camera position is evaluated as 26.5 centimetres. Kohoutek et al. (2010) describe a method which combines range camera measurements and an available CityGML indoor model. The coarse positioning is carried out by room identification using object detection in the range camera image and matching to the CityGML model, while centimetre-accurate positions can be computed by trilateration and spatial resection.

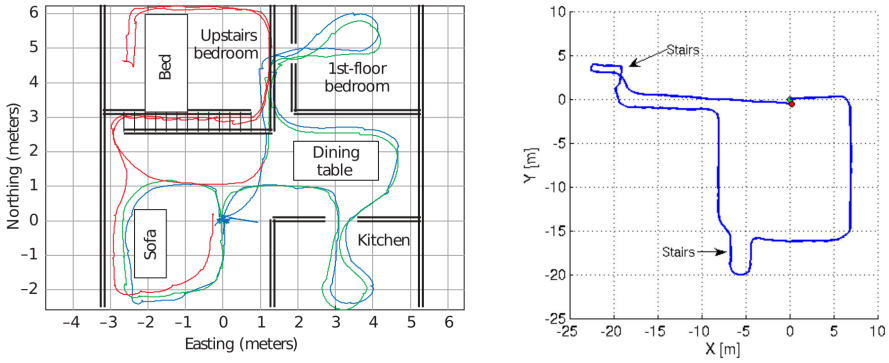
The FootPath system described in Link et al. (2011) uses smartphone sensor based step-detection and heading estimation. However, in contrast to dead-reckoning methods supported by map-matching as presented in the next section, FootPath depends completely on the availability of a map. The user requests a route which is computed using a navigation graph (modeled in OpenStreetMap). Using an approximate or pre-calibrated step length, this route is subdivided into steps with associated heading information. The steps detected in the accelerometer data as well as the magnetometer's noisy heading data are matched to this pre-computed string of expected steps using two presented path matching methods. In an indoor test, errors between 1.6 and 5.6 metres were achieved. In Bitsch Link et al. (2012), this approach is made usable for wheelchair users by estimating motion information from the visual flow of a smartphone camera, i.e. by extracting motion vectors from the hardware-accelerated H.263 compression of the video stream.

2.2 Infrastructure-independent

If no outdoor signals are receivable inside the building, no dedicated hardware is installed and possibly existing signals cannot be leveraged, e.g. because no fingerprint map was prepared, the only method to determine a position update in building interiors is dead-reckoning. In the outdoor domain, modern dead-reckoning systems as found aboard ships and airplanes use marine, navigation or tactical grade Inertial Measurement Units (IMUs) as their basis, resulting in errors as low as 1.8 kilometres *per day* after nave double integration of the accelerations to velocities and coordinate vectors[35]. In contrast, the accelerometer drift and heading errors produced by consumer grade sensors (which can be bought at reasonable prices and used for indoor positioning) will sum up to more than 2 kilometres *after one minute* if evaluated following the same principle.

To counteract this unfavourable error behaviour successful approaches for indoor dead-reckoning employ knowledge about human movement in order to distinguish stance and movement phases of the human gait cycle (Vaughan et al., 1992) in the IMU signals. The detection of stance phases subsequently allows for the correction of the drift and thus a reset of the errors. With the help of dedicated MEMS IMUs mounted to the user's foot, the full trajectory of the moving foot can be reconstructed (see next section), while methods

[35] according to `http://www.vectornav.com/support/library/inertial-sensor-market` (last visited 2nd February 2016)



(a) (b)

Figure 2.3: Results as presented in a) Foxlin (2005) and b) Ojeda and Borenstein (2006)

leveraging the sensor equipment found in smartphones mostly base on pre-calibrated or fixed step lengths (see section 2.2.2). As the resulting coordinates still contain small amounts of drift, the last section describes aiding methods such as heading correction and map matching.

2.2.1 Foot-mounted MEMS IMUs

By mounting the MEMS IMU to the user's foot, a strap-down system is realized which allows for the detection of zero velocity phases in the accelerometer data after integration to velocities. Assuming that these phases correspond to the stance phases of the human gait cycle, they can be integrated into a Kalman filtering operation as Zero Velocity Updates (ZUPTs) and the unavoidable drift in the accelerometers can be reset (Harle, 2013). The detected stance phase events together with the orientation derived from the gyrometers can either be used to apply a fixed, pre-calibrated step length or - more commonly - the drift-corrected measurements can be integrated twice in order to get a step vector or the full trajectory of the foot over one realization of the gait cycle. In both cases, the initial position and orientation have to be obtained externally in order to derive coordinates usable with respect to a map and its reference system. In most approaches, these values are assumed to be known (e.g. by user input) or derived from GNSS when entering the building (continuous positioning).

Harle (2013) presents a survey of Pedestrian Dead Reckoning systems which use inertial measurement units. Concentrating on foot-mounted systems, the author subdivides these into Inertial Navigation Systems (INS) which reconstruct the full trajectory of the foot and Step-and-Heading (SHS) systems delivering only step lengths and headings for every step. Systems of the latter type which use an INS to deliver this data the author calls SHS-INS. To this end, the primary phases of the human gait cycle (stance and swing) are described, completed by toe-off and heel-strike. The data sources for step cycle detection, i.e. accelerometer magnitude or angular velocity (or combinations), are presented as well as the methods for the actual detection: peak detection, zero crossings, auto-correlation, template matching or spectral analysis. The author also mentions the low reliability of magnetometers in indoor environments which can produce heading errors of up to 100° depending on materials and radiation in the sensors' vicinity.

The first publication describing the use of foot-mounted MEMS IMUs for indoor positioning is Foxlin (2005) mentioning earlier efforts which never published results, did not validate their method experimentally (Elwell, 1999), or show an error behaviour rendering them barely usable (Stirling et al., 2003). Foxlin (2005)

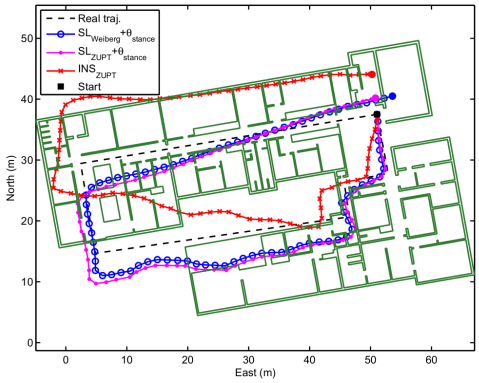


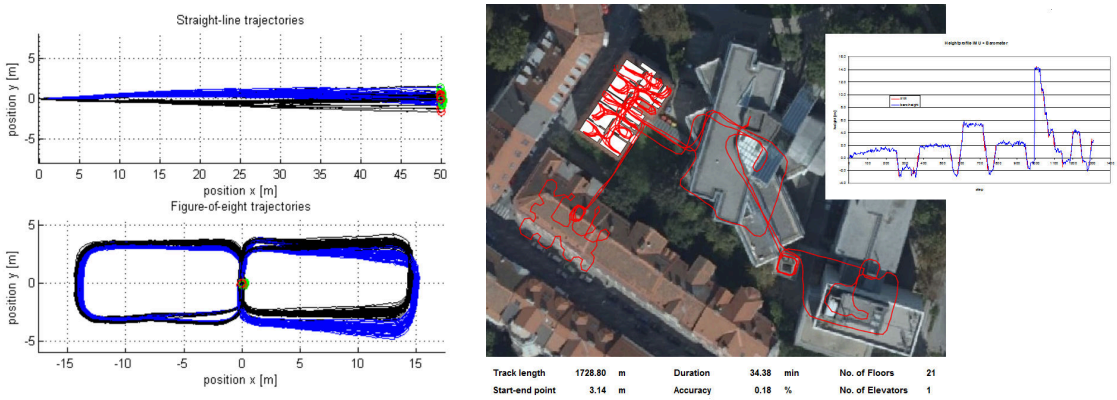
Figure 2.4: Ground truth (black), step length estimation derived using Weinberg (2002) (blue) and ZUPTs (pink) combined with heading, double-integration using ZUPTs (red) as presented in Jimenez et al. (2009)

presents the concept of using foot-mounted sensors in order to reduce drift by employing zero velocity updates (ZUPTs). The detection of stance phases in the human gait cycle and the application of ZUPTs as pseudomeasurements in an Extended Kalman Filter (EKF) are described. The result shown in the paper is depicted in figure 2.3a), the author reports 0.3% of error both in indoor and outdoor unaided use. These impressive numbers are also due to a special on-shoe magnetometer calibration reducing perturbances produced by metal parts of the shoe.

Another concept for the use of foot-mounted MEMS IMU is presented by Ojeda and Borenstein (2006). However, they do not employ the stance phase events in a Kalman Filter, but instead integrate the "dedrifted" acceleration values step-wise followed by a reset. Despite this simple approach they reach 0.8% of error in straight line experiments, up to 2.33% in a one kilometre 2D closed-loop experiment and 1.4% in a 3D closed-loop experiment (see figure 2.3b). A similar approach is described by Beauregard (2007), who also describe the applicability of the principle for other walking styles like side-stepping and diagonal stepping. Another quite early publication is Godha et al. (2006), where the GPS/INS integration using foot-mounted tactical-grade IMUs as well as MEMS IMUs is investigated. Thus, the experiments described were carried out mostly in outdoor signal degraded areas and only small parts in indoor areas. A follow-up paper shortly describes an indoor test run resulting in an error of 3.3 metres between start and end point (Godha and Lachapelle, 2008). Yun et al. (2007) follow the same principle, reporting travelled distance errors of up to 4.75% in a straight line running test, while the straight line walking tests and a test of the combination of forward walking, side stepping and backward walking show better results.

Jimenez et al. (2009) compare different methods for the detection of step events using either accelerations, gyroscope, or magnetometer measurements. They conclude that the magnetometer signal produces slightly less accurate results than the gyroscope signal whose performance is close to equal to the accelerations. For the stride length estimation, implementations from Weinberg (2002) and ZUPTs were evaluated. Here, Weinberg (2002) shows a better performance, yet it is rated less versatile than ZUPTs which (in theory) can be used to detect all kinds of human movement. Studies with more than 30 test trajectories showed errors between a supposedly equal start and end point of below 5% for all methods (see figure 2.4).

While the step lengths delivered by foot-mounted systems are very accurate, errors in the computed headings caused mainly by sensor misalignments deteriorate the overall results. In order to overcome this problem, Schäfer (2011) presents a method which bases on the different step lengths of the left and right foot when walking along a curved path. Thus, for the described experiments both feet were equipped with reduced IMUs, i.e. a single gyroscope and two accelerometers. The stance phase detection is carried out using



(a) (b)

Figure 2.5: Results as presented by a) OpenShoe (Nilsson et al., 2012) and b) AIONAV on XSens' web page

accelerations only. The accuracy evaluation for the step length which uses a GAITrite system[36] as reference, shows that those reported by the foot-mounted system are shorter by 10 centimetres, but have the same low standard deviation of 4 centimetres. The complete system is evaluated by walking a figure-of-eight. Still, errors remain due to the misalignment of the accelerometers' measurement axes, the "toe out angle" which was not taken into account and the results from the two feet not being combined, averaged or filtered.

Nilsson et al. (2012)'s OpenShoe system tries to facilitate the entry of research groups into the field of indoor positioning. As an open source project, board layouts, parts lists as well as software sources are freely available, ensuring that the reproduction costs can be kept as low as around 800$. The authors recommend the heel of a shoe as IMU mounting position, in their project this is done using special footwear. The software uses the accelerometer signal for stance phase detection, producing ±0.5 metres of error perpendicular to a straight line walk of 50 metres as well as ±0.15 metres along the line (see figure 2.5). The same error results from a figure-of-eight of 30x10m². In other publications, the OpenShoe group describes the initialization using inter-agent ranging as well as experiments covering the system's use in firefighter scenarios.

Due to the relatively new methods and the expensive hardware, few commercial systems are available. "Autonomous Indoor and Outdoor NAVigation (AIONAV)", a spin-off of the Technical University of Graz, advertises specialized systems for firefighters, police and special forces as well as the military. The patented positioning approach is outlined in (Walder and Bernoulli, 2010) and is able to cover all kinds of human movement such as walking, crawling and running with differing kinds of equipment and on varying surfaces. On the web page of its partner XSens[37] the company gives some information on the accuracy: In 150 test runs with an average length of over 600 meters the system produced an average error of 0.77% (or 3.75 metres). However, it remains unclear which additional sensors (GPS, barometer etc.) were used and which correction methods (alignment, map-matching etc.) were applied to reach this accuracy.

2.2.2 Smartphone sensor equipment

As presented before, systems comprising higher-grade MEMS IMUs and leveraging knowledge about the human gait cycle are able to fulfill the accuracy requirements of many indoor scenarios. For everyday use

[36] `http://www.gaitrite.com/` (last visited 2nd February 2016)

[37] `http://www.xsens.com/customer-cases/autonomous-indoor-positioning-navigation-inertial-measurement-unit/` (last visited 2nd February 2016)

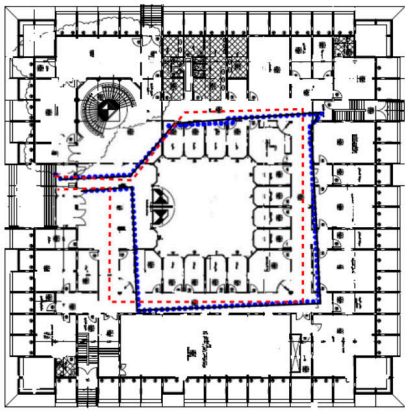


Figure 2.6: Smartphone sensor based dead-reckoning results as presented in (Alzantot and Youssef, 2012a)

two key system factors must change: 1) The systems have to be able to cope with the output of lower-quality sensors and 2) the need for a foot-mounted system in addition to hardware which the user is carrying anyway has to be dropped. Tests concerning only the latter requirement can be found in Renaudin et al. (2012). The authors describe the analysis of the signals of a MEMS IMU which is carried in the user's hand while walking and conducting typical smartphone tasks such as texting.

Jin et al. (2011) use two smart phones placed in the user's left and right trouser pockets in order to do step detection as well as stride length estimation. For the step detection the acceleration measurements are analyzed using a window size and a threshold which have to be pre-calibrated. The stride length estimation, on the other hand, uses only the vertical accelerations as well as a user-dependent calibrated constant. The authors describe position errors of around 2 metres as achievable.

A system fulfilling both aforementioned requirements is the UPTIME system described in (Alzantot and Youssef, 2012a), which leverages the sensor equipment found in a modern smartphone. The paper concentrates on the detection of the users' displacement, and leaves the accurate heading computation to one of the authors' future papers. Thus, the system is split into a step detection module and a stride length estimation module. The step detection module builds on a Finite State Machine (FSM) using fixed thresholds, which detects steps in the accelerometers of the phone carried in the user's hand or pocket. For this purpose, the length of the 3D acceleration vector is used as input signal for the FSM, which compares it to a series of thresholds in order to determine the state the user currently is in, e.g. "user has possibly started a step". The stride length estimation module does not estimate the stride length of the current step, but classifies the steps as one of the three gait types "walking", "jogging" or "running". Therefore, features like the duration of the step, the variance of the acceleration signal etc. are extracted and fed into a Support Vector Machine classifier. The constant stride length for the chosen gait type subsequently is employed in the estimate for the displacement of the user during the detected step. The authors describe resulting errors of 33 metres for an outdoor path of 623 metres and 4 metres for a 95 metres indoor path (see figure 2.6), which they seek to improve in future work using map matching and the accurate estimation of the user's heading. Some means to achieve these goals can be found in the CrowdInside (Alzantot and Youssef, 2012b) paper which is outlined in section 3.4, consisting among others of the detection of anchor nodes and the use of Wi-Fi signal strengths. In addition, section 3.4 describes more infrastructure-independent indoor positioning systems which are used in the context of indoor mapping.

2.2.3 Aiding methods

There are various areas in which infrastructure-independent positioning systems can be aided and improved by external measurements, or information. Some approaches have been presented which aim at improving the detection of the exact moment of the user's footfall. Others support the measurement of the orientation (changes) either observing the orientation (change) directly or through an alignment using external information. Finally, some map-matching methods to correct the full trace computed from the single steps were proposed. The following excerpts existing approaches in these fields.

In (Aubeck et al., 2011), the detection of the tips of the user's feet in the camera image of a handheld smartphone is leveraged to support step detection approaches. The detection is treated as a template matching problem, the needed template image is extracted manually and mirrored for the other foot. These templates subsequently are detected during the online phase using normalized cross-correlation. An example of approaches directly observing orientation changes is found in Ruotsalainen et al. (2011). The authors describe the detection of heading changes using a smartphone camera. To this end, the image is pre-processed, before being Canny filtered in order to detect the contained edges. The Hough transformation is employed to detect the lines in the resulting edge image. Using voting on all line pairs, the vanishing point can be computed efficiently by the weak smartphone processor. The heading change finally is computed from the movement of the vanishing point, resulting in mean errors of 1-2° (in optimal scenes). Another heading correction system using images can be found in Bernoulli et al. (2011). There a database of reference images and the according positions is built in an offline phase. This step is carried out by use of the aforementioned AIONAV positioning system and manual interaction providing ground truth positions on a map. During the online phase, the search space in the database is constrained by the approximate position from the positioning system, before the most probable image is searched using a hierarchy of image correlation algorithms. These consist of SURF features and perspective conditions as well as a Hough like transform and the final correlation based on edge orientation and spatial proximity. From the position of the retrieved image, only a heading correction is computed as the distances delivered by AIONAV are sufficiently correct.

In contrast to these methods which need additional measurements and sensors, alignment and map-matching methods only require a mobility model and/or a model of the environment the user is in. The latter is needed anyway for most applications e.g. navigation and for the trace visualization. An example of methods basing only on a mobility model is the Heuristic Drift Reduction (HDR) approach presented in Borenstein et al. (2009). There the fact that most man-made structures force pedestrian movement to follow straight lines is used to correct the drift of the gyroscopes. The authors describe means to counteract the effects of detractions from straight walk resulting from swaying, curving and turning. In an average of five equal walks they describe an improvement of the final and average heading error of nearly 10% respectively 7%. An update to this approach explicitly aiming at indoor positioning is the Heuristic Drift Elimination (HDE) described in Borenstein and Ojeda (2010). It takes into account information about the environment by using the dominant directions of the building the user is in. As the authors state, most corridors in man-made indoor environments are straight and either parallel or orthogonal to each other or the peripheral walls, with the rare exception of 45° angles. An exemplary result can be seen in figure 2.7. A similar approach is presented in Abdulrahim et al. (2010) where the dominant building directions are computed from aerial images using the Canny edge detector and the Hough transformation. The distance between the equal start and end point of a 2700 meters walk is as low as 2.3 metres, or roughly 0.1%. Jiménez et al. (2011a) enhance this approach in their improved HDE (iHDE) approach by the detection of the user walking along a straight

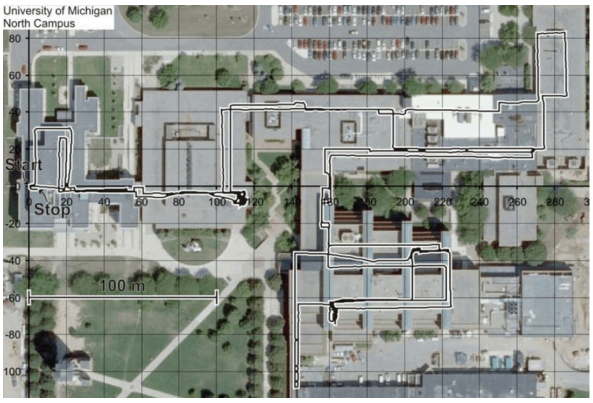


Figure 2.7: Heading-corrected indoor walk as presented in Borenstein and Ojeda (2010)

line. This approach is further improved using magnetometer data (Magnetically-aided improved heuristic drift elimination, MiHDE) as presented in Jiménez et al. (2012). While magnetometer data is disturbed in indoor environments e.g. by metal structures, the authors observed that a coarse heading estimation is possible when averaging the data over 100 measurements and correcting only after five consecutive steps with a similar heading.

These approaches base on very coarse environmental information found in aerial images or the external building shell which is available from city models or OpenStreetMap. Methods using more detailed indoor models for the correction of the positions resulting from dead-reckoning indoor positioning, shall be summarized below under the generic term map-matching.

Map-matching methods in edge-based maps - the most prominent example being street networks - build on the comparison of features such as curvature between the trajectory and the map (White et al., 2000; Czommer, 2001). Due to several reasons, the application of such methods for pedestrian indoor positioning systems is very limited. Indoor models, in contrast to street networks, represent the walkable space as faces instead of edges. Furthermore, while the assumption that a car stays on the street represented in the street network graph is true most of the time, pedestrians generally can move more freely.

Adopting the most popular method from robot positioning, Woodman and Harle (2008) present their approach for particle filter based map-matching. In their eyes map-matching is not only needed for the correction of errors occuring while the pedestrian travels through the building, but also for the determination of initial values for the relative positioning method. Particle filters implement belief propagation based on the current state, motion models and error characteristics of the used sensors. With the existence of new data, the propagated belief is corrected. Woodman and Harle (2008) model the particles using the x and y coordinate and the heading. The particle is constrained to the polygon from their 2.5D indoor model. If the particle hits a wall, its weight during the correction step is set to zero, otherwise it is set to the difference between the height change of the current step and the height change between the start and end polygons. This causes a very fast convergence of the filter when staircases are used. The remaining problems are the symmetry of man-made indoor environments, and the missing scalability of the filter resulting in the need for high particle numbers. To counteract, the authors propose the use of an altimeter to obtain an initial height value or an approximate initial position from WIFI fingerprints. While the latter reduces the number of needed particles to 1/30th, it requires extensive offline phase measurements. A comparison with ground truth measurements captured by the Active Bat system reveals an error of 0.73 metres (95%).

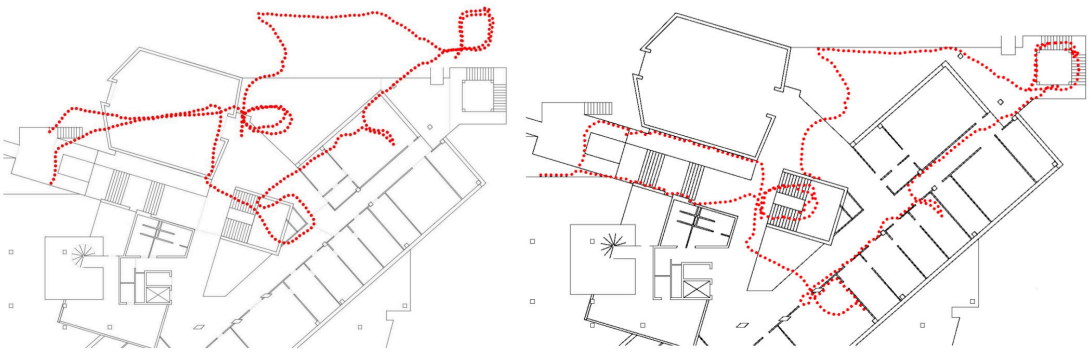


Figure 2.8: Raw trace (left hand side) and map-matching result as presented by (Walder and Bernoulli, 2010)

Walder and Bernoulli (2010) developed a map-matching method which employs heuristics of human movement in indoor environments. Their first assumption is the fact that pedestrians never walk through walls. Thus, if the trace hits a wall close to a door, it is shifted to the door. If far away from a door, the trace is rotated around the last correct position. The second heuristic states that pedestrians cannot pass doors in an acute angle. If this is violated, the trace is shifted "some distance away" from the door. Lastly, if the trace's z-values change and therefore indicate the usage of stairs, the position is moved to a nearby staircase. The map-matching approach developed as part of this thesis, which will be described in chapter 5, uses similar assumptions, however, does not require the door positions to be modeled. A follow-up paper (Bernoulli et al., 2013) describes ideas for using the magnetometer data as well as GNSS positions, however, merely giving hints on the methods used.

3 Mapping the interior structure of buildings

The most prominent, traditional uses for digital models of building interiors can be found in the field of disaster simulations, interior architecture and building information management. However, as shown by the last section of chapter 2, precise and complete models are of great importance in the context of pedestrian navigation. There are several factors motivating this.

Positioning components as described in the previous chapter represent only one subdomain of the navigation task - route planning is the other. In order to provide the user of a navigation system with a route to their desired destination, their position must be known with respect to a model of their environment ("map") which is modeled in the coordinate system also employed in the context of the positioning method. Furthermore, the map is needed to display the computed route to the user of the navigation system. Lastly, the map comprises constraints relating to the position and direction of movement of the navigating subject. Thus, the map can support the positioning method by balancing its limitations - e.g. remaining long-term drift errors - and thus improving the positioning accuracy. As already mentioned in the previous chapter, this role of the map is highly uprated in the context of indoor positioning compared to outdoor applications.

In contrast to outdoor features mapping of building interiors is hindered by various factors ranging from the size of the possible user group of the acquired map to the contrast between the required level of detail and the accuracy of the available data acquisition methods.

In an outdoor setting, professional data acquisition uses existing infrastructure like reference-point networks or GNSSs, as well as having access to established measurement principles like tachymeters or aerial, terrestrial or mobile laser scanning (Haala et al., 2008a,b). These measurement principles deliver data with an accuracy level which - in most scenarios - is at least one order of magnitude better than required by the desired level of detail of the resulting map. Additionally, with the availability of low-cost and pervasive GPS receivers and public-domain aerial images, the less accurate, but very complete and up-to-date results of crowd-sourced mapping have become an important factor in outdoor mapping. In addition to these technical reasons, mapping of outdoor features - being in the public domain - enjoys a much greater target audience in comparison to a map of a single building's indoor environment.

Even further, indoor mapping lacks the connection to reference-point networks and an absolute positioning method (like GPS in the outdoor world), thus complicating the registration of different data sets and low-cost manual modeling in the style of OpenStreetMap. Finally, the dimensions of features of interest and therefore the desired level of detail in indoor settings are much closer to the measurement accuracy of the available data acquisition methods.

This chapter is organized as follows: section 3.1 explains the existing data models which can be used to model 2D and 3D geo-data in general and the ones which were developed especially for models of building interiors. Sections 3.2, 3.3 and 3.4 describe the scientific and commercial efforts made in the past to solve

the indoor mapping problem. This overview is split into methods which do not (section 3.2) and which base on point clouds as intermediate products (section 3.3), followed by methods using pedestrian traces as input (section 3.4). Section 3.3 in turn is divided into acquisition methods for point clouds (sub-section 3.3.1) and segmentation and reconstruction methods (sub-section 3.3.2). As will be shown in section 4.1, indoor reconstruction can also be based on the automated reverse engineering of existing maps. Such methods are excluded from the overview of related work in indoor mapping presented in this chapter.

3.1 Data models

Indoor mapping consists of the two components *data acquisition* and *reconstruction*. The reconstruction result has to follow a data model in order to make it interchangeable between applications. The choice of the data model, as with other data in the context of Geographic Information Systems (GIS), is dependent on the (desired) dimensionality of the reconstruction result as well as its intended use.

2D

Two-dimensional geodata can be modeled as raster or vector data for the use in Geographic Information Systems (Bartelme, 2005). Raster data is represented by single or multi-band images with a geo-reference and scale. Thus, every cell of the raster refers to a patch in reality, which is well-defined in position and size. The advantages of raster data are found in the fact that all cell positions are implicitly represented and general image processing operations can be applied. However, this comes at the price of a higher memory consumption in comparison to vector data. In the outdoor domain, the most common raster data can be found as map tiles (e.g. when accessing OpenStreetMap or Google Maps). While these can only serve visualization purposes, robotic acquisition systems, for example, produce intermediate products like (raster) maps of accessible areas.

Geodata can also be modeled as vector data, i.e. as a hierarchy of faces, edges and nodes. Faces are bound by edges which in turn are bound by nodes which carry coordinates. Furthermore, attributes can be assigned to all elements, rendering vector data very flexible in terms of visualization. Other advantages include the higher memory efficiency in comparison to raster data, explicit storage of exact positions, and the implicitly present adjacency information for edges and faces. Vector data can be further subdivided into edge-based and polygon-based vector data. The edge-based vector data model proves particularly useful for the description of street networks. As the movement of cars is restricted to a network of interconnected streets and this model supports the efficient calculation of shortest routes as well as map-matching, data formats like Geographic Data Files (GDF) base on it. Polygon-based vector data, on the other hand, can be considered as a 2D version of the Boundary Representation (BREP) data model described below. Building ground plans and land use modeling (as in the German ATKIS/ALKIS systems) are two examples of the use of polygon-based vector modelling in the outdoor domain.

3D

Point clouds can be useful as data model in a number of scenarios, despite the missing explicit topology and the limited possibility to assign semantic information beyond per-point color or intensity. They are

often employed for visualization purposes[1], and their success used as positioning support information in automative scenarios has been evaluated (Schlichting and Brenner, 2014).

Apart from point clouds, a number of modeling concepts for 3D data exist. In his doctoral thesis, Kada (2007) gives a broad overview over the options available for 3D city models, identifying the top level categories *decomposition modeling*, *constructive modeling* and *boundary representation modeling*. Decomposition modeling, in turn, comprises cell decomposition modelling, spatial occupancy enumeration and binary space partitioning trees. From this list, spatial occupancy enumeration, i.e. building the model from voxel elements or voxels - can be seen as the 3D analog to raster data in 2D. Similar disadvantages apply, aggravated by the added dimension.

Boundary representation modelling (BREP), i.e. vector data in 3D, is the most versatile 3D data schema. BREP follows the same principle as the 2D vector data model with one difference - the nodes carry 3D coordinates. The vast majority of operations in the field of Computer Graphics is carried out using BREP in the form of triangular meshes. Polygonal modeling can be seen as a special case of triangular meshes used e.g. for city models as in the City Geography Markup Language (CityGML) (Kolbe et al., 2005). Advantages are the explicit definition of complete complex facades or ground plans as one polygon, the small memory footprint, and the assignment of semantic information per polygon. One disadvantage, the polygons have to be tesselated, i.e. split into triangles, for an efficient visualization.

Building interiors

In order to visualize the results of indoor positioning systems, raster data, e.g. scanned ground plans, is very commonly used (see also chapter 2). Furthermore, robotic acquisition systems use this data model as means to store intermediate products like maps of accessible areas in binary images. For most applications, however, the availability of vector data is advantageous or required. The same advantages of boundary modelling apply as before: the low memory footprint as well as the ease of semantic annotation and the simple, flexible visualization. However, additional applications can benefit from BREP modeling: building management, disaster simulations (Tang and Ren, 2012; Goetz and Zipf, 2012) as well as positioning and navigation support (map-matching, signal distribution calculations; see also chapter 2).

While the highly constrained movement of cars in the street network motivates the application of an edge- (i.e. graph)-based data model, pedestrian movement is much less restricted. Thus, the concept is not portable especially to indoor positioning and navigation and more often polygon-based BREP is applied. However, some research projects in this domain apply the graph-based concept nonetheless, e.g. FootPath Link et al. (2011). Obvious data formats for the storage of 2D vector-based indoor models are CAD formats e.g. AutoCAD's DWG/DXF or common GIS formats e.g. ESRI shapefiles. Furthermore, the IndoorOSM schema[2], proposed by the Institute of Geography of the University of Heidelberg (Goetz and Zipf, 2011), and other schemas[3] allow for the modeling of building interiors in OpenStreetMap. For such polygon-based vector models, however, the navigation graph, which is needed e.g. for route calculations, has to be derived by adequate approaches (Vanclooster et al., 2013). This is one of the disadvantages of this model.

[1] e.g. `http://matterport.com/` (last visited 2nd February 2016)

[2] at the time of writing broken due to changes to the OpenStreetMap data model

[3] `http://wiki.openstreetmap.org/wiki/Indoor_Mapping` (last visited 2nd February 2016)

Concerning 3D indoor modelling, Boehm and Pattinson (2010) showed that raw point clouds can serve as positioning support information in the same way they do outdoors (see also section 2.1.5). One of the most important data schemas in the context of 3D BREP modelling, the Industry Foundation Classes (IFC, ISO 16739:2013), the data schema used in Building Information Management (BIM) systems. City Geography Markup Language's Level-of-Detail 4 (LoD4) is designed for the storage of indoor models together with the surrounding building structure in other LoDs. CityGML is an application schema of the Geography Markup Language (GML, or ISO 19136) and is a realization of polygonal BREP modeling combined with semantic information. In addition to the aforementioned geometric and semantic data models, a specialized indoor data model - indoorGML[4] - was just accepted as an OGC standard. While it is possible to model the geometry in indoorGML, its main purpose is the modeling of topology, semantic information and navigation information (such as WiFi fingerprints or RFID beacon positions). This additional information is linked to geometry modeled e.g. in CityGML, IFC or ISO 19107.

3.2 Manual and image-based methods

This section gives an overview over existing manual (section 3.2.1) and semi-automatic methods (sections 3.2.2 and 3.2.3) for the mapping of building interiors.

3.2.1 Manual methods

The majority of authors describing indoor positioning and navigation approaches assume the general availability and thus do not explicitly state the sources of the maps they use for visualization or aiding methods. However, some others present the (manual) modeling strategies used in their research groups. Hub et al. (2005) describe the derivation of environment models for an indoor navigation system for visually impaired people, in Link et al. (2011) the data acquisition for the authors' aforementioned map-based smartphone positioning and navigation system is presented. In order to test their data structure for building interiors proposed to OpenStreetMap, Goetz and Zipf (2011) document a manual modelling strategy[5] based on photographed evacuation plans which was inspired by the approach developed in this thesis. Spindler et al. (2012) also base their manual modeling approach for the "Project Mobility" (navigation for visually impaired people) on evacuation plans.

Outside of academia, virtually all major map providers (e.g. Google Maps, Microsoft Bing, Nokia Here) are advertising the availability of indoor maps in their products, mainly concentrating on shopping malls, airports and museums. In addition to these global map providers, there are smaller map producers concentrating on the production of indoor maps, e.g. Aisle411[6], PointInside[7] and Micello[8]. The data acquisition and modeling methods of these private companies can only be guessed. However, Google Maps offers an

[4] `http://indoorgml.net/` (last visited 2nd February 2016)

[5] `http://wiki.openstreetmap.org/wiki/IndoorOSM` (last visited 2nd February 2016)

[6] `http://aisle411.com/` (last visited 2nd February 2016)

[7] `http://www.pointinside.com/` (last visited 2nd February 2016)

[8] `http://www.micello.com/` (last visited 2nd February 2016)

insight on the data acquisition process, providing a tool where users can upload and manually scale and geo-reference an image of a ground plan they possess[9], resulting in an image overlay which might be modelled manually in a follow-up process. A similar manual method is used by indoorAtlas[10].

Apart from these reverse-engineering approaches, in the field of manual data acquisition various measurement principles may be of use including measuring tapes, distance measurement devices, Flexijet[11], tachymetry (especially in combination with CAD software like TachyCAD[12]) or manual photogrammetric methods[13]. All these methods tap human reasoning and knowledge for the selection of the points important for the final model, either during the data acquisition or the final modelling process.

3.2.2 Semi-automatic methods using panoramic images

Apart from manual methods, a group of semi-automatic approaches basing on panoramic images can be identified in the literature. As a first example, (Shum et al., 1998) present a system for the interactive modelling of textured 3D indoor models from panoramas. The modelling process is split into the computation of camera poses with respect to the panoramic images and the estimation of both plane and line parameters. In their earlier paper, (Farin and de With, 2005) recover the camera positions using panoramic images of a single rectangular room, while generalizing the approach to arbitrary room layouts as well as multiple rooms in (Farin et al., 2007), as shown in figure 3.1a). However, their updated approach still needs a user-drawn sketch as input in addition to the panoramic images. This idea is adopted by (Dang and Worring, 2011), who explicitly constrain the user sketch to the Manhattan World. After the acquisition of this sketch, their system executes a module for planning the capturing process, informing the user which viewpoints to prefer. In the acquired and stitched panorama, the user identifies corners which, together with the coarse sketch, enable the reconstruction of the walls, floor and ceiling of the model. Additional model features can be added interactively using a "perspective extrusion" method also described in the paper. Exemplary results are depicted in figure 3.1b). Doubek and Svoboda (2002) produce one-shot panoramas using a camera and a hyperbolic mirror. Employing a few tie points hand-measured by the operator, ego-motion estimation and 3D reconstruction using Structure-from-Motion can be utilized.

In the absence of further information about the techniques used, the commercial smartphone software named MagicPlan by Sensopia[14] shall also be named in this context. While it is not working on a finally assembled panoramic image, this application presents the camera image of the smartphone to the user while he turns around and selects the corners of the room. It can only be guessed that a combination of optical flow and the accelerometer, gyroscopes and magnetometer sensors present in modern smartphones is used to produce a consistent model. Multiple rooms are captured individually and can be assembled to a floor plan interactively.

---

[9] `https://www.google.com/maps/about/partners/indoormaps/` (last visited 2nd February 2016)

[10] `http://www.indooratlas.com` (last visited 2nd February 2016)

[11] `http://www.flexijet.info/` (last visited 2nd February 2016)

[12] `http://kubit.de/CAD/Produkte/TachyCAD/index.php` (last visited 2nd February 2016)

[13] `http://www.photomodeler.com/index.html` (last visited 2nd February 2016)

[14] `http://www.sensopia.com/english/index.html` (last visited 2nd February 2016)

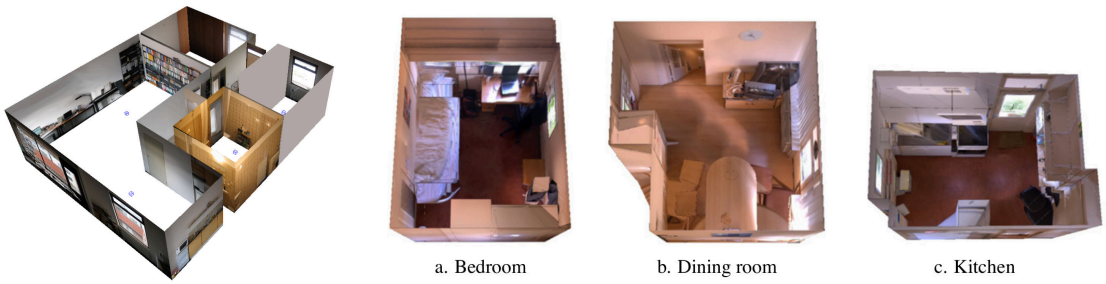


Figure 3.1: Methods employing panoramic images: a) (Farin et al., 2007), b) (Dang and Worring, 2011) (figures adapted)

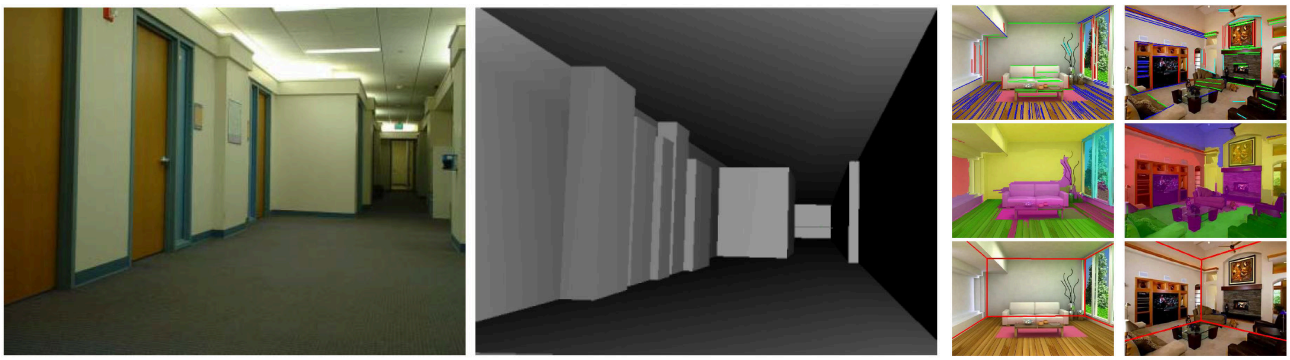
3.2.3 Automatic monocular methods

Patterns of man-made structures, especially buildings and their interiors, often follow the Manhattan world assumption. When used as a constraint, this assumption enables or strengthens reconstruction methods for those environments. Among others, it permits the interpretation of single images of indoor environments with the aim of directly deriving boundary representation models. Only monocular methods deriving boundary representation models are described below, for methods producing depth images, see the next section.

Very early work using the Manhattan world constraints in the field of monocular camera images can be found in Bellutta et al. (1989), Straforini et al. (1992) and Olivieri et al. (1992). Bellutta et al. (1989) (despite the rather general title "3D Visual Information from Vanishing Points") show the use of these constraints in the reconstruction of building interiors, calling them "Legoland". Using the Legoland constraints on images of indoor environments, the authors describe how vanishing points can be detected and used for the computation of the rotational component of viewer movement. In addition, a method for the reconstruction of so-called "simple polygons" is presented, i.e. polygons whose sides converge on only two vanishing points and therefore are likely to be the perspective projection of an entirely visible patch. Straforini et al. (1992) and Olivieri et al. (1992) extend Bellutta et al. (1989)'s approach in three ways: 1) by using the edge detection approach by Canny (1986), 2) by defining "maximal polygons" which are the union of the maximum number of adjacent simple polygons of the same orientation in space, and 3) by including low-level *a priori* knowledge in order to detect entities such as floors and walls.

(Delage et al., 2006) present the 3D reconstruction of indoor scenes using a calibrated camera with known viewpoint height and Manhattan world constraints. These constraints allow for the reduction of the 3D reconstruction problem to the detection of the floor boundary using a dynamic Bayesian network. As the accuracy of this approach is highly dependent of the accuracy of the viewpoint height's measurement, this approach is mostly applicable in robotic SLAM systems using a camera mounted at a fixed height.

Huang and Cowan (2009) extract the vanishing points of corridor scenes and use heuristics to identify ceiling and floor. They present algorithms to detect horizontal and vertical edges as well as to reconstruct missing edges basing on the detected junction types. Their method's unique feature is the scaling of the model using the known focal length and the assumption that the far and near end of the visible corridor have the same width.



(a) (b)

Figure 3.2: Automatic image-based methods: a) Huang and Cowan (2009), b) Hedau et al. (2009) (figures adapted)

Hedau et al. (2009) and Wang et al. (2010) use Support Vector Machines trained to find the best-fit box inside the image of a room based on texture as well as color, thus discriminating between walls, floor, ceiling and clutter. Flint et al. (2010) reduce the complexity of the problem after identifying the dominant surface orientations and detecting floor and ceiling by rectifying the vertical lines using a perspective transformation before estimating the final model using dynamic programming. The PhD thesis published by Lee (2011) as well as other ongoing research (Fouhey et al., 2014) indicate the ongoing interest in monocular methods.

3.3 Point cloud and depth map based methods

3.3.1 Acquisition methods

The methods described in the previous section merely solve sub-problems of the problem of acquisition and reconstruction of building interiors. This is due to the fact that the required manual effort, irresolvable scale problems and their limitation to the reconstruction of single rooms or areas bounded by the camera's field of view cannot be dismissed.

The related approaches described until now either employ the intelligence of a human operator and/or presumed model constraints in order to derive boundary representation models directly from the acquired raw data. Additionally, the vast majority of publications dealing with the reconstruction of building interiors decouple the two sub-problems *acquisition* and *reconstruction* and concentrate on one of them. As a common denominator, many approaches in the context of acquisition concentrate on depth images or point clouds as end products. The derivation of boundary representation indoor models using the point clouds' analysis is often out of the scope of the publications and left to an independent area of scientific interest.

The central issues of data acquisition in indoor environments are geo-referencing (respectively linking of different data sets), and the generally disadvantageous conditions for the application of image-based methods (lack of texture, poor lighting conditions, bad intersection angles). While the problem of geo-referencing is factored out or solved manually in the context of some publications, most available approaches consider it as a "Simultaneous Localization and Mapping (SLAM)" problem (Leonard and Durrant-Whyte, 1991). The unfavourable conditions for passive, image-based methods, however, cause a distinct bias of scientific interest towards active acquisition methods such as laser scanning and range cameras.

Thus, the related work in point cloud-based acquisition approaches is classifiable a) on the basis of being static (without a contained solution to the geo-referencing problem) or SLAM methods and b) depending on whether they employ active or passive sensors. The structure of this section follows this classification.

Image based, static

Analogous to the model-based single-image approaches, some methods aim at the derivation of depth images or respectively point clouds using the analysis of a single image. While the models underlying the methods described in section 3.2 facilitate the analysis of the image's content or serve for the improvement of the results, they concurrently constrain their application to the specific character of indoor environments. In contrast, the approaches to be described here deliver results for a broader range of input images (outdoor, indoor). The resulting point clouds or depth images, however, not only lack the higher level knowledge delivered by the boundary representation, but are also less accurate and reliable.

Battiato et al. (2004) describe a method for the depth map computation from single images which employs colour as well as vanishing point information. By use of a color-based segmentation, regions typical of landscape/outdoor images (e.g. sky) are detected and processed to a *qualitative depth map*. Additionally, this information is used to classify the image to the categories of *landscape, outdoor with geometric elements,* and *indoor*. The depth map generation uses vanishing lines in order to compute depth gradients which are fused with the qualitative depth map to the final result. Saxena et al. (2008) employ an analysis of both the global structure as well as local features of the image for the estimation of a depth image. Apart from certain a priori knowledge (blue color in the lower part of the image = water, blue color in the upper part = sky), they extract features for absolute as well as relative depth. Using the extracted features and ground-truth Laser scanning/stereo-vision data, a multi-scale Markov Random Field was trained and used for the online depth image generation. Saxena et al. (2008) conclude that the described single-image approaches should be considered as extensions to other approaches rather than independent solutions due to the lack of a solution to the scale problem.

Image based, SLAM

Doubek and Svoboda (2002) use a hyperbolic mirror to produce one-shot panoramas. Using these images in a Structure-from-Motion approach, the camera poses as well as the 3D coordinates of corresponding points are reconstructed. The use of stereo cameras is more common for the 3D data acquisition from images. Iocchi et al. (2001) present a robotic system carrying a stereo camera which is used - in combination with the odometer data - for the derivation of the robot pose. In contrast to most other methods which either employ the geometric information or image features, their approach adds the use of image texture by correlation. As will be presented later, Iocchi et al. (2001) also describe how to reconstruct planar structures from the point clouds. Furukawa et al. (2009) present the reconstruction of indoor environments from images by use of structure-from-motion (Bundler, Snavely et al. (2006)) and multi-view-stereo (PMVS, Furukawa and Ponce (2010)). In order to robustly produce simplified models, their stereo algorithm as well as the simplification method both incorporate the Manhattan world constraints.

A common use of visual SLAM approaches is augmented reality (AR). While AR approaches generally apply previously modeled knowledge about the user's environment, Klein and Murray (2007)'s parallel tracking

and matching (PTAM) approach acquires this model in real-time. This is achieved by splitting the process into feature detection and tracking, on the one hand, and mapping, on the other hand. After an initialization step using one image pair, the mapping bases on the video stream's keyframes which are processed using local and global bundle adjustment. The number of points is enlarged by applying an epipolar search. The method was proven to work on an off-the-shelf camera phone, as described in Klein and Murray (2009). Basing on PTAM and previous work in monocular SLAM (Davison, 2003), Newcombe and Davison (2010) collect only the highest quality points of its real-time SfM reconstruction and add a dense surface reconstruction.

Active, static

At the junction of passive and active sensing, methods replacing one of the cameras of a stereo system by a projector for structured light have to be mentioned. Bahmutov et al. (2006) describe a system employing a combination of a video camera, shaft encoders and a laser projector for an efficient manual indoor data acquisition. In order to acquire undisturbed texture images, their acquisition and the collection of the 3D points are split into two steps. The use of shaft encoders delivers angle readings and therefore provides real-time feedback to the operator of the system, which in turn enables a complete acquisition. Due to the limited resolution of the pattern projected by the laser projector, the authors call the result "dense color and sparse depth". The approaches for semi-automatic registration of the single point clouds as well as their reconstruction approach will be described later.

For obvious reasons, the collection of information about building interiors without entering them is of great interest to police and military applications. Baranoski (2006) describes efforts to acquire this information by sensing through the walls using RADAR. However, the measurement principle of choice for accurate large scale indoor data collection is "Light Detection and Ranging (LiDAR)", also called laser scanning. Apart from triangulating measurement principles using laser dots or lines monitored by a camera, laser scanning generally refers to distance measurement systems based on either time-of-flight measurement of laser impulses or phase comparison.

The application of laser scanning for building information modeling (BIM) is described in the U.S General Services Administration's guide called "GSA BIM Guide for 3D Imaging" (GSA, 2009). The guide caters to the selection of the correct sensor for required accuracies and levels of detail, project planning and execution as well as the registration of single scans. Laser scanner producers (e.g. Leica Geosystems, Riegl and Faro) bundle software for the (semi-)automatic registration using spherical or printed paper targets as well as modeling tools or interfaces to standard modeling software to their sensor systems.

In order to foster the development of indoor positioning approaches basing on Content-based Image Retrieval (see also chapter 2) in the TUMindoor project, Huitl et al. (2012) describe the acquisition of a dataset consisting of laser point clouds and images. This was done using a custom-built mapping trolley equipped with SICK profile scanners, a digital SLR and a panoramic camera. In contrast to the Trimble/Applanix system and the "Wägele" mentioned below, the laser point clouds were registered by hand.

In the commercial domain, Trimble/Applanix offer the "Trimble Indoor Mobile Mapping Solution (TIMMS)" which consists of a push-cart equipped with laser scanners, panoramic camera and an IMU[15]. The use of a

[15] `http://www.applanix.com/solutions/land/timms.html` (last visited 2nd February 2016)

(a) (b)

Figure 3.3: a) Results as presented by Bahmutov et al. (2006), b) the mapping trolley used in the acquisition described in Huitl et al. (2012) (figures adapted)

high-grade IMU to track the system's position frees the operator from the time-consuming manual or semi-automatic registration task, however, the drift errors influencing this relative positioning method require an update using a known position after a maximum of 100m. According to Trimble, TIMMS enables the user to capture an area of 14.000m² per hour with an accuracy of 3.5 centimetres per single point.

Matterport[16] sells a "scanning" system composed of PrimeSense range cameras together with capture software and the web hosting service of the produced 3D data. Sample applications of their system can be found in real estate, insurance, construction and retail. The precision is specified as to be in the range of 2.5 millimetres.

Active, SLAM

The term "Simultaneous Localization and Mapping (SLAM)" has its origin in the field of robot-based exploration of unknown areas. While robotic applications are still the main field of their application, the principles fundamental to it have been adopted successfully to the human-based acquisition of indoor environments.

Traditionally, sonar was the distance measurement method of choice for indoor environments, especially for robotic systems, before it lost its importance with the emergence or miniaturisation of more accurate methods e.g. laser scanning and range cameras. Burgard et al. (1999) used this fairly inaccurate measurement principle together with a SLAM implementation building on the expectation-maximization algorithm. In Leiva et al. (2001), a robotic system combining sonar ranging and images is described. The basic 2D geometry is reconstructed using image processing algorithms on the occupancy grid derived from the sonar measurements. The camera images, on the other hand, are used to texture the model. Instead of reducing the problem of reconstruction from noisy sonar data to image processing, Tardós et al. (2002) apply the Hough transform directly to the measurements, yielding impressive results which were evaluated using simultaneously collected LiDAR data.

Analogous to the Trimble/Applanix system described above, Biber et al. (2005) present the "Wägele", a push-cart carrying two 2D laser scanners mounted perpendicularly as well as a calibrated panoramic camera. In contrast to robotic systems, the cart's position is not needed during acquisition as the path planning is accomplished by the human operator (similar to their earlier publication presenting a remotely controlled

[16] `http://matterport.com/` (last visited 2nd February 2016)

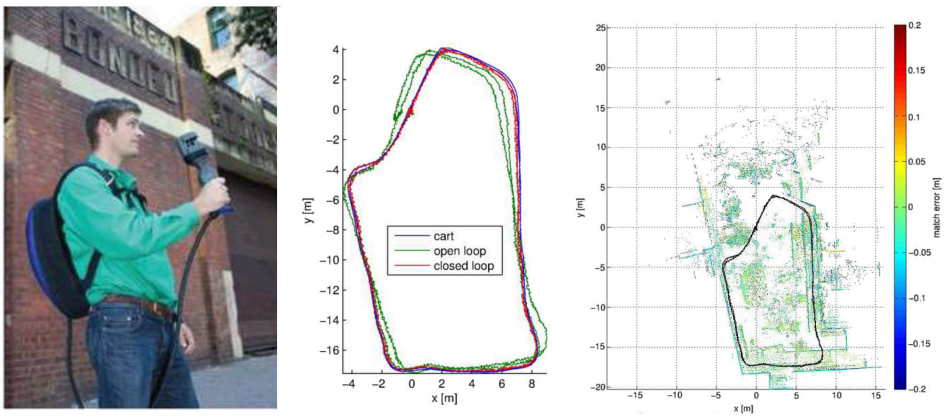


Figure 3.4: The Zebedee system and exemplary results (figures adapted)

robot (Biber et al., 2004)). Thus, the solution of the SLAM problem is done in a post-processing step using only the data of the horizontally mounted scanner and scan-matching as well as global optimization, followed by a second scan-matching operation between the single scans and the optimized model in order to derive the sensor positions. The final colored point cloud is obtained from a combination of the scans and the panoramic images using the lever arms determined in a system calibration. The paper describes the high flexibility using results from tests in outdoor and indoor scenarios, however, the design as a push-cart renders for instance the data collection in stair cases impossible.

This limitation is overcome by Liu et al. (2010)'s set-up of three laser scanners, three cameras and an IMU in a specially designed backpack. The measurements from the laser scanners and the IMU are jointly analysed by the SLAM algorithms presented in Chen et al. (2010), while loop-closure detection is carried out using features detected in the camera images. These also serve as color source for the final coloured point clouds. In order to evaluate the accuracy of the backpack system, a navigation-grade IMU is included, delivering ground truth measurements. The follow-up paper (Oreifej et al., 2014) shows the combination with thermal infra-red camera images.

An even more flexible LiDAR-based system called *Zebedee* is described by Bosse et al. (2012). Its components are a MEMS IMU and a 2D laser scanner as well as one (hand held system) or two (robotic system) springs. The operator's or robot's movement causes a stimulation of the spring(s) and the resulting irregular movement adds the third dimension to the point cloud acquisition. The post-processing SLAM approach bases on a special Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992) extended for the simultaneous estimation of time synchronisation and IMU errors. As depicted in figure 3.4, the *Zebedee* system delivers impressing results, especially when using loop closure detection and correction. A commercially available version is marketed by the British company 3D Laser Mapping.

While the aforementioned approaches left the path planning to human operators and thus enabled the post-processing solution of the SLAM problem, autonomous robotic systems depend on the online incremental map build-up and position computation for their automatic and independent path planning. The algorithms involve scan matching methods whose complexity may be greatly reduced by a reduction of the problem to two dimensions. When looking at robotic systems designed for the exploration of building interiors, this reduction to two dimensions makes sense for two reasons: First, the loss of information on interiour man-made structures is small, generally insignificant. Second, less complex algorithms put less demand on the limited computing and energy resources of mobile robot systems, especially in combination with the necessity of quasi real-time computations. In many approaches the reduction is realised in

hardware by the use of a single horizontally aligned profile scanner or the split into two scanners mounted perpendicularly.

Surmann et al. (2003) present an autonomous robotic system equipped with two laser scanners. As in most contemporary publications, the laser scanners are profilers. In order to achieve the collection of 3D data, one of them is mounted on a rotating platform. Apart from the description of the hardware, the paper outlines the software modules developed for scan registration (coarse registration using odometry data followed by ICP), best next view planning, as well as the collision avoidance with dynamic obstacles. Wulf et al. (2004) present a similar system. However, instead of registering the scans using the raw point cloud they construct a virtual 2D scan by choosing the most distant point in every vertical scan. This assumes that the walls are the most distant boundary of a room seen from the position of the robot. This virtual scan is used to derive linear features which, together with the odometry data, are then employed in a line-based registration step.

In his doctoral thesis, Fietz (2012) describes the development of a robotic system which can acquire point clouds using a profile laser scanner (2D) and photogrammetry (3D). Using Manhattan constraints to limit the search space for the next best view, the map is built using scan-matching and probabilistic filtering of the LiDAR point cloud. The sensor placement strategy for the photogrammetric 3D data acquisition uses the previously acquired 2D LiDAR base map. In contrast to other robotic SLAM approaches, the main focus of this work is on the accuracy of the resulting maps and their usability in human indoor navigation.

With recent developments resulting in the widespread availability of *Unmanned Aerial Vehicles (UAV)*, these systems are commonly applied to reduce the costs for small-scale aerial imaging projects in outdoor environments or augmenting the area possibly covered by close range photogrammetry scenarios, e.g. in the context of cultural heritage . However, as presented by Achtelik et al. (2009), researchers are also experimenting with the LiDAR and UAV based autonomous exploration of indoor environments. Achtelik et al. (2009) employ a laser scanner and a scan-matching algorithm together with a stereo camera system and feature tracking as a replacement for the odometry sensors common to ground-based robotic systems. Furthermore, this data in combination with measurements from a MEMS IMU is fed into the flight control computations. Pose estimation and path planning are covered by a SLAM algorithm analogous to ground-based robots, producing 2D point clouds.

Time-of-flight (ToF) cameras were considered as a very promising method for active 3D point cloud collection before the Microsoft Kinect entered the scene. Prusak et al. (2007) present a robot system carrying a Photonic Mixer Device (PMD) as well as a color fisheye camera. The system uses the colour images, Structure-from-Motion as well as a model-tracking approach in order to localize itself. Furthermore, 3D panoramic images are stitched using the PMD images. This 3D information, in turn, can be used to avoid obstacles during the robot's navigation. May et al. (2009), in contrast, use only the 3D point cloud information resulting from the sensor together with a registration approach basing on the Iterative Closest Point (ICP) algorithm and a loop-closure detection. This makes both a calibration and a filtering step of the usually noise ToF camera point clouds necessary. Hochdorfer and Schlegel (2010) present a third option by detecting SURF features in the depth image.

With the introduction of the Microsoft Kinect, ToF cameras were superseded by this structured light based system due to it being cheaper, more accurate and higher resolving at once (see below). The Kinect 2.0, presented together with the XBox One in the year 2013, however, is based on ToF technology. Apart from

this change in sensor technology, it brings a wealth of improvements over the original Kinect, among them a larger field of view, higher resolution and more instantaneous data availibility. In his Diploma Thesis, Schneider (2014) leverages this sensor by post-processing using KinectFusion (see below) and a special trajectory improvement step using ICP.

Since the emergence of the first generation Microsoft Kinect, an extensive number of new approaches building on this stable, relatively accurate and low-cost sensor device have been published. The Kinect, building on a sensor concept by PrimeSense, was designed as a controller-free interface for Microsoft's gaming console XBox, enabling the simultaneous tracking of multiple players' movements and their transfer to e.g. an avatar in the game. By the built-in structured light projector-camera system this is achieved detecting and tracking skeleton junctions in a depth image delivered with a temporal resolution of 30 frames per second. To this end, an included infra-red laser projector emits a pseudo-random point pattern to the scene which is detected and tracked by an infra-red camera. This enables the computation of disparity values and therefore VGA range images in a limited range of 0.7 metres to 6 metres. The point clouds derivable from these depth images furthermore may be colourised by the images captured by the equally included VGA RGB camera (if a stereo calibration of the RGB and IR sensors has been carried out beforehand). Despite characteristics like price, speed and size, this sensor system is very similar to Bahmutov et al. (2006)'s earlier system, and delivers results similar to those of ToF cameras, however, at a significantly lower price and a higher resolution.

Shortly after the emergence of the Kinect, Henry et al. (2010) presented their approach called "RGB-D Mapping", employing the Kinect for the acquisition of dense indoor point clouds. Exploiting the benefits of the real-time availability of depth as well as RGB imagery, their method solves the geo-referencing problem by using a combination of RANSAC-filtered SIFT/FAST feature matching (basing on the RGB images) and ICP alignment (basing on the point clouds' geometry). The drift problems encountered are tackled by a method for the detection of loop closures, followed by global optimization using bundle block adjustment. A follow-up paper adds an accuracy analysis stating mean errors of between 0.11 metres and 0.37 metres (Henry et al., 2012). A result can be seen in figure 3.5.

In contrast to the aforementioned integrated approach, Tykkälä (2013) uses only the RGB image information for the pose estimation step. To this end, direct methods for the pose estimation building on cost functions were investigated and developed. Thus, feature extraction and matching steps can be avoided. The result is delivered as a textured Poisson reconstruction instead of the raw point cloud.

Newcombe et al. (2011) present KinectFusion, the most successful RGBD reconstruction method to date. KinectFusion enables real-time dense volumetric reconstruction of room-sized scenes using a hand-held Kinect. In contrast to the approaches described above, Newcombe et al. (2011) only employ the geometric information stemming from the depth sensor part of the system. To this end, new incoming data is registered to the existing global model by use of a coarse-to-fine ICP algorithm which enables greatly improved results over a frame-to-frame tracking strategy. KinectFusion in turn is the base for many other projects, including KinFu Large Scale, which is part of the Pointcloud Library (PCL)[17], (Roth and Vona, 2012)'s Moving Volume KinFu, and Kintinous (Whelan et al., 2012). These approaches essentially aim at increasing the volume which can be covered by the system while keeping the hardware requirements low. Pagliari et al. (2014) furthermore describe the improvement of KinectFusion using a calibration of the depth camera while

[17] `http://www.pointclouds.org/` (last visited 2nd February 2016)

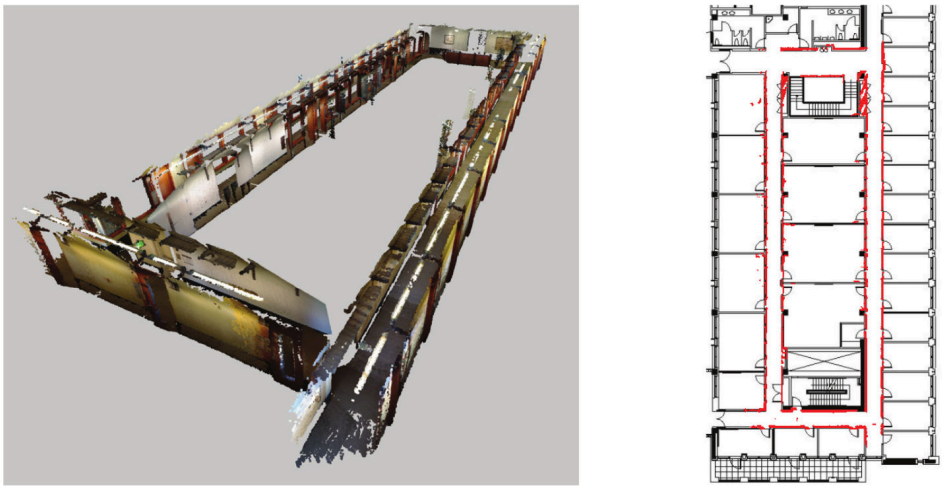


Figure 3.5: Results from RGB-D Mapping as presented in Henry et al. (2012) (figures adapted)

Peasley and Birchfield (2013) employ the RGB camera images for the pose estimation in addition to the geometric information the original KinectFusion approach relies on (similarly to the approach described by Henry et al., 2010).

In addition to these publications presenting general point cloud acquisition methods employing the Microsoft Kinect, first works using the sensor in a robotic system were presented in Audras et al. (2011). More recently, Chow et al. (2014) combined two Kinect sensors with a FARO Focus 3D laser scanner and a low-cost IMU in a robot system. To this end, they use the RGB and depth images as well as the IMU data for pose estimation in the context of a SLAM implementation, complemented by long-range laser scans recorded at discrete points. Finally, the Kinect was tested aboard a UAV (Bachrach et al., 2012) as well as combined with a thermal sensor with the goal of indoor point cloud acquisition in Vidas et al. (2013).

At least two commercial programs, ReconstructMe[18] and FARO's Scenect[19], exist which utilize the Microsoft Kinect for hand-held point cloud acquisition. The Kinect had a huge impact on the market of 3D capturing devices, despite its primary application being the control of video games. Understandably, its sensor concept is tried to be commercialized in products designed primarily towards 3D capturing. Two capturing systems aimed among others at the acquisition of building interiors in 3D are: The DPI-7 sensor[20], made by DOT product, and Google's "Project Tango". The DPI-7 combines a PrimeSense Carmine camera and an off-the-shelf Android tablet running a real-time point cloud registration program. Google's "Project Tango"[21] is a prototype device consisting of an Android smartphone or tablet containing a structured light sensor (PrimeSense Carmine) for depth measurements, an RGB camera, a fish-eye high-speed camera used for motion tracking, and co-processors for computer vision tasks.

[18] `http://reconstructme.net/` (last visited 2nd February 2016)

[19] `http://www.faro.com/scenect/scenect` (last visited 2nd February 2016)

[20] `http://www.dotproduct3d.com/pandt.html` (last visited 2nd February 2016)

[21] `https://www.google.com/atap/projecttango/` (last visited 2nd February 2016)

3.3.2 Segmentation and reconstruction methods using point clouds

The point clouds resulting from the aforementioned acquisition methods are themselves not usable as an end product, apart from visualization purposes and positioning methods as the one presented in Boehm and Pattinson (2010) (see also section 2.1). For the majority of applications for indoor models, boundary representation models have to be derived. To reach this goal Reconstruction methods must be robust against the inaccuracies of the provided point clouds. At the same time, occlusions of walls and the floor caused by furniture and other objects have to be detected and handled.

With the rise of low-cost sensors, the data collection process is clearly in the focus of researchers as well as sensor manufacturers. In contrast, the model reconstruction from these point clouds is still a mostly manual and thus expensive and time-consuming task. Laser scanner manufacturers bundle software with their sensor systems that enables the interactive visualization of point clouds in CAD modeling software, allowing users to profit from the point clouds' accuracy and point density while continuing the use of well-established software. Moreover, Leica Geosystems is offering Cyclone-MODEL which offers tools supporting semi-automatic reconstruction e.g. line and plane fitting to a chosen set of points and region growing of planar surfaces.

Automatic methods which are in the focus of current research range from the segmentation of point clouds into meaningful patches to mesh simplification to BREP reconstruction (with or without Manhattan constraints). Additionally, some of the published methods deal with the extraction of higher-level semantics and object recognition.

Martinez Mozos et al. (2007) present a 2D classification method which bases on the oriented distances of each non-wall point to the nearest wall point. These distances form a histogram descriptor which then is used to train a classifier. Subsequently, the classifier is used to identify rooms, corridors and door openings. In Huang et al. (2014) the segmentation of RGBD point clouds to arbitrarily shaped objects is described. To this end, the authors apply superpixel segmentation and so-called synthetic volume primitives (SVTs). By means of these SVTs, a common object is identified by coplanarity and colour coherence.

In addition to their robotic SLAM system described above, Hähnel et al. (2003) describe two specialized approaches for the reduction of the triangle number in meshes of indoor environments. While the results are not final BREP models, these approaches can be considered as pre-processing steps which are applicable in indoor and outdoor acquisition of man-made structures. The first approach presented in the paper uses a distance-based region growing method for the segmentation of the (very noisy) meshed point clouds, followed by a merging operation of neighbouring triangles. The second approach is based upon the idea of plane sweeping, i.e. reducing the search space to the one-dimensional movement of a specified plane along its normal vector and the collection of points most likely belonging to this plane. In Hähnel et al. (2003), the specification of the planes for the sweeps is realized by the computation of the histograms for the predominant angles found in the horizontal and vertical scans, respectively, and the selection of peaks in these histograms. The sweeping operation is split into a coarse (5cm step size) sweep and a fine (1cm step size) resolution second sweep around every peak found during the first sweep, allowing to distinguish doors from walls.

Wiemann et al. (2010) employ the Marching Cubes algorithm to build a triangular mesh from the point cloud, adapting the number of chosen nearest neighbours for the normal vector computation depending on

the local point density. Analyzing the mesh structure enables the detection of planar surfaces and results in the reconstruction of polygons. These are semantically annotated as floor, ceiling or walls. The point clouds in the paper's experiments were collected using a robot and a 6DOF SLAM method, the resulting models serve as map information for the localization of other robots based on LiDAR or ToF camera point clouds (see Wülfing et al., 2010). In their BREP reconstruction method for single rooms, Kotthauser et al. (2013) segment the point clouds collected by their autonomous robot platform in order to extract planar patches. These patches' extents are then leveraged to find the room's footprint and ceiling and subsequently the final 3D model.

Bahmutov et al. (2006) present an automatic modelling approach specialized to indoor scenarios and the "dense colour and sparse depth" point clouds resulting from their sensor system (described before). This approach comprises constraints for ceiling and floors as well as opposing walls being parallel, i.e. implements the Manhattan constraints. Building on these constraints, for each station the ceiling and floor planes are estimated by means of user-selected points and the remaining wall planes are searched automatically, based on edges drawn by the user in the RGB images. After intersecting the planes, the resulting model is categorized according to the template shapes Box (for rooms), I and L (for hallways) and T (for junctions). The authors state that L and T shaped part models can be registered automatically, while I shaped hallways have to be registered employing the operator. The registration approach for rooms, however, is not further explained.

In their robot system, Iocchi et al. (2001) use planes estimated from the collected point clouds to constrain the search space for the registration of image pairs and, ultimatively, for the computation of the robot's pose. To this end, a generalization of the line extracting 2D Hough transformation to the 3D space is used. Similarly, Kotthauser et al. (2013) apply a 3D Hough transformation in order to derive planar patches from the point clouds resulting from their LiDAR-equipped autonomous robot platform. Using these planar patches, the adjacency graph is derived as well as their footprints. This information is employed in the reconstruction of 3D BREP models and the identification of points inside the room.

Nüchter et al. (2003) present a segmentation method for planes in indoor scenes which were acquired by a robot mounted LiDAR system (see Surmann et al., 2003). For the detection of planes in the data and the computation of their geometry, they use a novel approach mixing elements from the RANSAC and ICP methods. Initially, local plane parameters are computed using a randomly chosen point and its two nearest neighbours. Then, all points closer to this plane than a distance threshold are counted. If their number exceeds a threshold, they are projected to the plane and used in an ICP to derive the final plane parameters. The semantic interpretation of the planes is carried out employing a predefined semantic net which comprises constraints for building interiors. As these constraints contain "walls are pair-wise parallel or orthogonal", this approach implements the Manhattan World constraints and vertical planes not fulfilling this constraint will be categorized as doors. In a final step, the planes are optimized by enforcing parallel and orthogonal planes.

Johnston and Zakhor (2008) concentrate on the reconstruction of building interiors from point clouds collected without access to the respective building. Using the Manhattan world constraints, they assume floors and ceilings to be perfectly horizontal and walls to be vertical. These constraints are implemented in a first step detecting floor and ceiling planes employing a vertical sweep of the xy plane and a rotational sweeps around the z axis using random non-floor/ceiling points. The resulting wall planes are further constrained to data areas, but no final CAD-like reconstruction process is carried out.

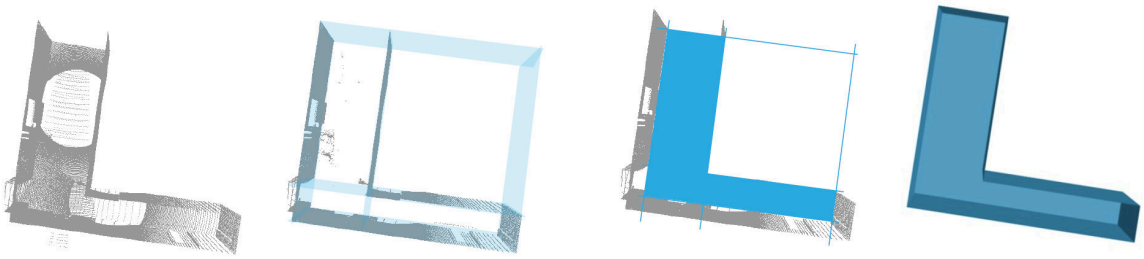


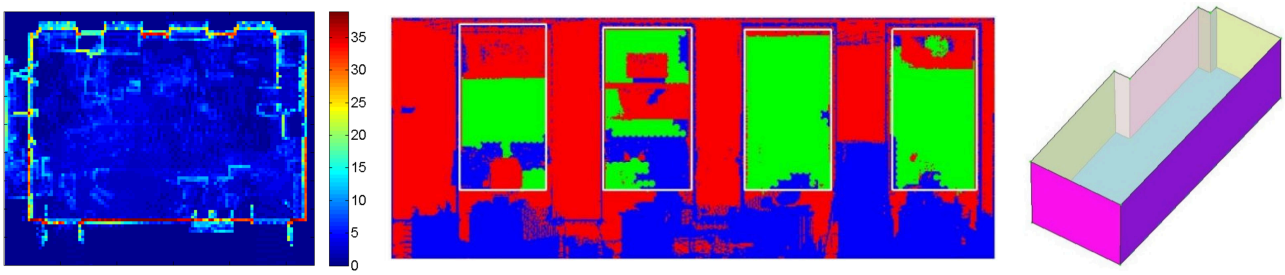
Figure 3.6: Modeling steps of the Manhattan World plane sweeping approach presented by Budroni (2013)

Budroni and Böhm (2009) present a Manhattan World constrained plane sweeping approach for indoor environments similar to the one presented by Johnston and Zakhor (2008). Assuming previously registered point clouds and a levelled laser scanner, their approach in an analogous manner employs a vertical sweep for the detection of floor and ceiling planes and a rotational sweep of a vertical plane around a random point from the remaining set in order to find the orientation of one wall plane. Implementing the full Manhattan world constraints, all other planes are identified by sweeping the plane identified by the rotational sweep and its orthogonal counterpart along their normal vectors. Using a half space modelling approach similar to the one implemented by Kada (2007) for 3D building generalization, the intersections of the thusly detected wall planes with the floor plane are computed and cells are established. In the final modelling step, these cells are categorized according to the ratio they are filled by ceiling points. Additionally, the boundary edges are extruded from the floor to the ceiling (see figure 3.6). In the corresponding doctoral thesis (Budroni, 2013), the approach is extended by a method for the detection of doors based on areas of missing points with a standard door size at a height of one meter. Its application to point clouds acquired using a robotic system and by a low-cost range camera (see also chapter 5) are also described. Previtali et al. (2014) present a similar half-space based reconstruction approach equally usable for outdoor and indoor environments.

Instead of reducing the problem to one-dimensional histograms by sweeping planes, Okorn et al. (2010) split it to a histogram analysis of the points' projection to the z-axis and the xy-plane respectively. In order to deliver BREP models built from the measurements of their robotic system, Leiva et al. (2001) analyse the thresholded occupancy grids using the Hough transform. In order to make the analysis independent from a varying point density, Okorn et al. (2010), employ a voxel discretisation step before identifying the floor and ceiling as the bottom-most and top-most maxima in the z-axis histogram. The 2D histogram resulting from a projection of the remaining voxels to the xy-plane is first thresholded to only retain high-density cells, followed by a Hough transform detecting the wall candidates as lines. These candidates are then corrected according to identified dominant wall orientations.

An approach presented by another member of the same group in Xiong and Huber (2010) uses a smoothness constrained region growing method similar to Rabbani et al. (2006) to segment planar patches from the voxel discretised point cloud. These are then classified to walls, floor, ceiling and clutter using a conditional random field classifier building on local features and context dependent relations between the patches. These two approaches can be categorized as point cloud segmentation methods as a final boundary representation modelling step is missing.

This limitation is overcome in another follow-up paper (Adan and Huber, 2011), which employs (Okorn et al., 2010)'s floor, ceiling and walls detection. Their aim is the reconstruction of boundary representation models with correctly detected and reconstructed windows and doors despite the presence of occlusion and clutter. Therefore, a ray-tracing occlusion labeling method is followed by the categorization of wall openings



(a) (b) (c)

Figure 3.7: a) 2D histogram for wall detection Okorn et al. (2010), b) wall opening detection from Adan and Huber (2011), c) resulting model from Valero et al. (2012) (figures adapted)

using a Support Vector Machine classifier. Possible perturbances of the detected wall openings stemming from occlusions are bridged by a suitable 3D inpainting method. The most recent approach expanding (Okorn et al., 2010)'s wall segmentation method is described in Valero et al. (2012). They improve the segmentation step by optimizing the voxel size according to the scanner accuracy and the flatness tolerance defined in DIN18202, and by extracting the wall planes as the outer boundary contour in the xy-projected voxels as well as deriving a final boundary representation.

Jenke et al. (2009) implement the Manhattan World constraints in their iterative optimization approach by the adjustment of cuboids to the point cloud. The plane segmentation step is done using RANSAC, followed by the organisation of the planes into a graph structure by means of the identification of pairwise perpendicular planes. By analysis of this graph structure cuboid patterns are detected which will be combined to form the BREP model. Also basing on cuboids fit to point cloud data, Khoshelham and Díaz-Vilariño (2014) developed a parametric shape grammar similar to the Palladian grammar presented by Stiny and Mitchell (1978). It consists of three rules which are applied to unit cuboids: one transformation rule using three translational and three scale parameters, one rule for the connection of neighbouring cuboids, and one for the merging of two adjacent cuboids. In contrast to Jenke et al. (2009)'s approach, the cuboids are directly fit to the point cloud without a preceding segmentation step. During this fitting operation, the grammar rules are detected from the data.

Nakagawa et al. (2014) build on panoramic range images generated from the point cloud using point-based rendering. This implies the need for a view point decision method which is described in the paper. Apart from edge detection carried out using the range image, it also serves for the computation of a normal vector image. The detected edges as well as the information about homogeneous surfaces gained by normal vector clustering deliver the final polygons. It has to be mentioned that these can also be curved surfaces as no Manhattan constraints are implied. The approach was tested with the help of data acquired using the Trimble Indoor Mobile Mapping Solution.

While many of the aforementioned methods enable the geometric reconstruction of various aspects of man-made indoor environments, the semantic information is mostly constrained to walls, ceiling, floor, and sometimes door and window openings. However, depending on the needs of the indoor model's application, the availability of additional semantic information may be needed or is at least favourable.

Furniture and similar items are mostly deemed to be clutter preventing the reconstruction of walls. Rusu et al. (2008) present a method to explicitly search for such objects and reconstruct them as semantically annotated cuboids. As their method is part of a project developing a service robot, they concentrate on the

detection of kitchen components such as cupboards, oven and dishwashers. Koppula et al. (2011), on the other hand, describe the segmentation of indoor point clouds into semantically consistent regions using a trained classifier and geometric relations like “on top of” and “in front of”. Having a similar goal, Shao et al. (2012) retrieve matching furniture models from a database using automatically detected and semantically annotated regions in the point clouds resulting from an RGBD camera. These regions can be improved by interactive guidance by an operator.

3.4 Methods using pedestrian traces

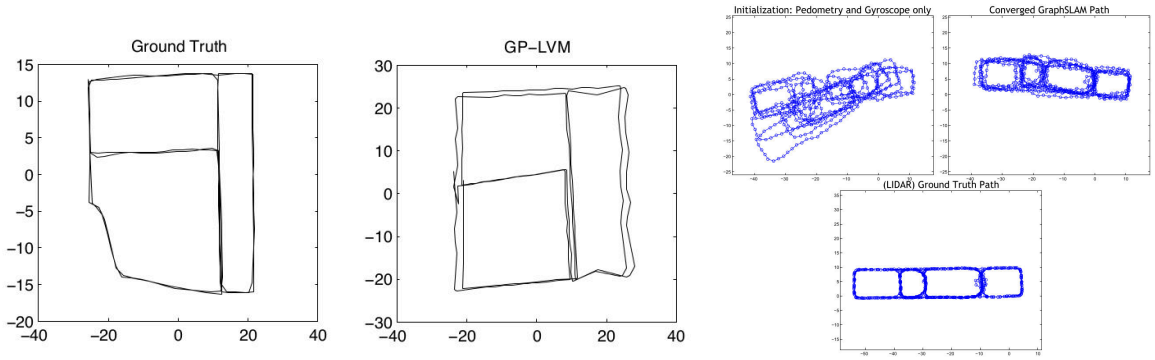
Leveraging data collected by pedestrians walking around in the building in order to derive floor plans without prior knowledge is of interest especially in the field of ubiquitous or pervasive computing. As these approaches present means to derive indoor positions without available infrastructure, they could likewise be mentioned in section 2.2.

One example for such approaches is the WiFi-SLAM system (Ferris et al., 2007), using only Wi-Fi signal strength readings from a mobile device for the estimation of the underlying floor plan. The method employs Gaussian process latent variable models (GP-LVM) for the mapping of the high-dimensional Wi-Fi signal strength data to a the low-dimensional 2D space representing the device’s positions. To this end, the GP-LVM is constrained as follows: 1) close locations refer to similar signal strengths, 2) the reverse constraint is used: similar signal strengths imply locations close to each other, 3) locations that follow each other in the data stream are spatially close to each other, representing a model of the user’s movement. As depicted in the paper, this approach results in a scaled, topologically correct corridor skeleton which is very similar to the ground truth (see figure 3.11). As a by-product, the resulting graph and Wi-Fi map can be used for positioning with an accuracy of around 4 metres.

Huang et al. (2011) combine Wi-Fi signal strength measurements with pedometry and gyro data in a state-of-the-art GraphSLAM as commonly used in robotics. In contrast to Gaussian process modelling, this allows for the removal of the restriction that similar signal strength fingerprints all have to correspond to a similar 2D location. Using the data captured by Quigley et al. (2010) this approach yields very strong results, as depicted in figure 3.8b) (upper right hand side). However, the result is produced in the unit “steps” and thus up to an unknown scale. For the accuracy assessment, it was scaled to match its total distance traveled to the ground truth’s total distance traveled, showing a mean localization error of 2.23 metres. The start-up company which evolved around this method was bought by Apple[22].

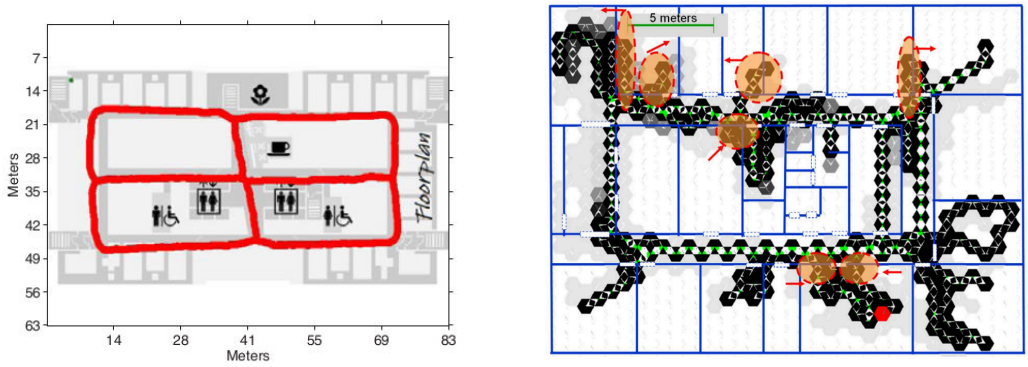
The Walkie-Markie system as published in Shen et al. (2013) employs the typical sensor infrastructure found in a modern smartphone. The trajectory acquisition is performed using step estimation from the smartphone accelerometers and a trained stride length per user in combination with the magnetometers. The paper’s central contribution is the definition of WiFi-Marks, i.e. landmarks derived from the received signal strength tipping point (RTTP) of one access point. These WiFi-Marks are augmented by use of the user’s orientation information when approaching and leaving the RTTP in order to ensure their uniqueness. Furthermore, the signal strength differences of neighbouring access points to the measurements of the defining access point are stored, which counteracts the fluctuations in Wi-Fi signal strengths. From the traces

[22] `http://techcrunch.com/2013/03/24/apple-acquires-indoor-gps-startup-wifislam-for-20m/` (last visited 2nd February 2016)



(a) (b)

Figure 3.8: Results as delivered by a) WiFi-SLAM (Ferris et al., 2007) and b) by WiFi GraphSLAM (Huang et al., 2011) and ground truth from Quigley et al. (2010) (figures adapted)



(a) (b)

Figure 3.9: Results from a) Walkie-Markie (Shen et al., 2013) and b) footSLAM (Robertson et al., 2009) (figures adapted)

merged by use of the WiFi-Marks the pathway map of the floor is built which, if no trained stride lengths are available, is up to scale (see figure 3.9a). The mapping accuracy is evaluated as 3 metres after 5-6 rounds of walking in an office environment with many access points. As a side product, a radio map of the environment is available, enabling the positioning of users with an average error of 1.65 metres together with the pathway map.

The Institute of Communications and Navigation at the German aeronautics and space research centre (DLR) has presented numerous indoor SLAM systems ranging from a purely foot-mounted MEMS IMU based method (Robertson et al., 2009) to a cooperative system (Puyol et al., 2012) and extensions using manual and automatic place annotations (Robertson et al., 2010) as well as incorporating Wi-Fi measurements (Bruno and Robertson, 2011). At the core of these approaches is the IMU data which is processed using a Kalman filter delivering coordinate updates. For the SLAM computation, a FastSLAM using a particle filter is applied which bases solely on the users' motion as a function of their location. As a result, an occupancy grid consisting of hexagons is available for the building interior. An exemplary result is depicted in figure 3.9b) in comparison to the ground truth floor plan (blue).

Shin et al. (2012) present the SmartSLAM system, also employing position estimation using smartphone sensors and employing the resulting traces for map construction purposes. The basis of their positioning approach is the combination of step detection using the accelerometers, a constant stride length and the heading information given by the magnetometers. For the analysis of these signals a particle filter is

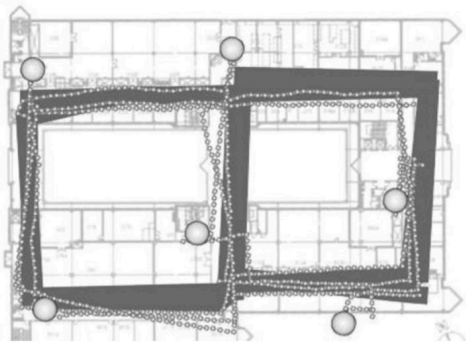


Figure 3.10: Results as shown in SmartSLAM (Shin et al., 2012)

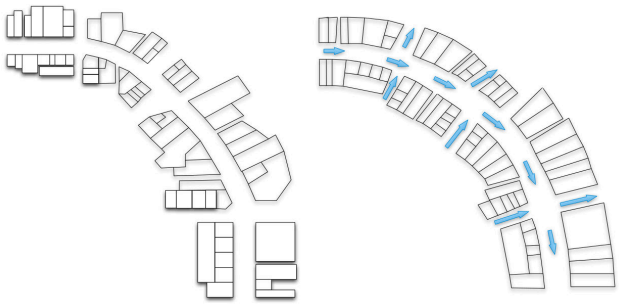


Figure 3.11: Results as delivered by Jiang et al. (2013) (left: ground truth)

employed and three different mobility models are presented: Gaussian noise, path-smoothing (using the assumption of users walking along straight lines), and "straight" (assuming all corridors are perpendicular). With this method, the authors claim to reach a step count accuracy of 95% at a standard deviation of 15 centimetres. In addition, the Wi-Fi signal strengths are used as map landmarks in SLAM processing supporting the positioning as well as delivering the map (see figure 3.10) which consists of the corridor layout and the Wi-Fi feature vectors.

Jiang et al. (2013)'s system builds on Wi-Fi fingerprints, however, combined with motion detection using the accelerometers and magnetometers found in modern smartphones. Assuming that a mapping of room IDs and signal strengths of Wi-Fi access points visible in the target room was established before (e.g. by a fingerprinting system presented by the same group, Jiang et al., 2012), the system estimates the floor plan layout. This includes room adjency graph construction, hallway layout learning and a finalizing step using force directed dilation.

The CrowdInside system, described in Alzantot and Youssef (2012b), builds on the UPTIME positioning approach (Alzantot and Youssef, 2012a) which was outlined in section 2.2 and the analysis of a large collection of user traces. Thus, it uses only the sensor infrastructure found in modern smartphones, i.e. accelerometers, magnetometers, GPS and Wi-Fi. The Finite State Machine used in UPTIME to derive user positions is extended in CrowdInside by the detection of GPS-based and intertial-based anchor nodes. GPS-based anchor nodes refer to the building's entrance which is detected by running the GPS receiver of the smartphone on a low duty cycle and detecting the signal loss. By averaging the positions of the GPS-based anchor nodes, the authors state that an accuracy of below 1 metre can be reached using at least 100 traces. Inertial-based anchor nodes are features detected in the accelerometer and magnetometer signals. The authors describe elevators which are detected by the overweight/weight-loss at the beginning and end of their use,

Figure 3.12: Acquired traces (left, blue: corridors, black: rooms) and floor plan (right) resulting from CrowdInside (Alzantot and Youssef, 2012b) (figures adapted)

escalators which are identified by the user being stationary while the magnetometer readings show a large variance, the use of stairs which show a larger correlation between the accelerometers' x- and z-readings and turning points detected by a threshold. In order to distinguish similar anchor points, the available Wi-Fi access points together with their signal readings are used. The floor plan is estimated by first segmenting and filtering the user traces and classifying the segments into corridor and room segments according to the features "average time spent per step in the segment", "segment length" and "neighbour traces density". Subsequently, the segments are clustered and the corridors respective rooms are derived as alpha shapes. Finally, the door positions are derived as transition points between corridor and room polygons. The authors claim that a relatively low number of 290 segments was sufficient for the full reconstruction of their sample floor plan. However, for a successful reconstruction of the room shapes, the traces have to follow the rooms' walls as closely as possible (as visible in figure 3.12).

Despite combining the inertial data collected using a smartphone with vision data and user interaction, Gao et al. (2014) compare their Jigsaw approach to CrowdInside. The main idea is the use of the expensive (in terms of energy consumption) vision technique for "landmark" areas where more accurate and detailed information has to be collected, while the cheap inertial data is leveraged for the placement of the landmarks. The vision-based landmark reconstruction is carried out using a combination of Structure-from-Motion and vanishing line detection, delivering the landmark's model in a local coordinate system. Involving the user by letting them fulfill so-called micro-tasks, the system determines values for the orientation differences and the distance between landmarks (and thus the scale). In a final optimization step, continuous boundaries are built between adjacent landmarks and the hallway connectivity is obtained. The authors claim to reach a fourfold improvement of their system over CrowdInside.

4 Reconstruction of indoor environments by reverse-engineering of existing maps

As shown in chapter 2, indoor LBS need complete and sufficiently accurate model information about the context the user is in. These models represent the map basis for route computation and visualization. Furthermore, the positioning methods currently available depend (to differing degrees) on map information, either for the improvement of their accuracy or as central part of the positioning method (see section 2.1.5). Additionally, as concluded from chapter 3, the acquisition of indoor models remains a tedious and/or expensive task.

In the context of this thesis, an approach for the automatic reconstruction of coarse indoor models from photographed evacuation plans in combination with an available model of the building's external shell was developed. As will be shown in chapter 5, the results are usable not only for the accuracy augmentation of existing infrastructure-less positioning methods, but also for the generation of semantically enhanced, navigatable models. The level of detail, accuracy and up-to-dateness of the models derived by this method depend on the designer and the age of the photographed plans. However, the efficient acquisition of huge areas (e.g. complete floors) using a single image and the automated analysis combined with automated geo-referencing represent the advantages of the approach.

4.1 Related approaches

The idea of locally augmenting the environment model by use of photographed maps was presented for outdoor scenarios in Cheverst et al. (2008) and Schöning et al. (2009). The authors describe how "you are here" maps of a person's vicinity can be referenced to the GPS position and the use of two GPS positions can deliver the information to scale and orient the map. These methods involve a certain degree of user interaction and the map is only used as overlay. Thus, it can serve for the purpose of visualization and provide information for the users to orient themselves. However, positioning correction methods like map-matching and route computation are not applicable. Finally, it is only applicable outdoors, because of the positioning method used.

Indoor reconstruction approaches related to the one to be presented here can be found in the field of automatic derivation of (3D) models from CAD files, scanned paper plans or hand-drawn sketches of floor plans. An early work in this field was described by Koutamanis and Mitossi (1992), describing first considerations concerning the design of a system for the recognition of elements in architectural drawings. The authors conclude that the emphasis in the system design should be on higher knowledge-based levels (the analysis of building elements and spatial articulation) instead of low-level machine vision analyzing e.g. geometric

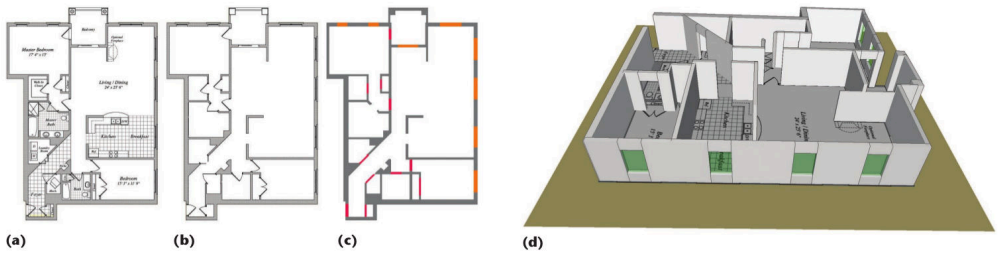


Figure 4.1: Processing steps and resulting 3D model as presented in Yin et al. (2009): a) original floor plan, b) denoising and text removal, c) symbol recognition and 2D geometry creation (figures adapted)

primitives. Ryall et al. (1995) present an approach usable to delineate different regions in scanned documents using the example of rooms in an image of a building floor plan.

Llados et al. (1996) and Aoki et al. (1996) present methods for the interpretation of hand-sketched floor plans. Llados et al. (1996) introduce the graph-based Hough transformation (GBHT) in order to detect hatched walls in the scanned and vectorized sketches. Constraining the walls to two dominant orientations, they appear in the GBHT as two peaks with an angular difference of 90°, a third peak denotes the orientation of the hatching pattern edges. Thus, the walls' widths can be derived from the length of the hatching edges and connected components be identified. After the removal of the walls, Llados et al. (1996) detect remaining elements using template-based graph matching. In Aoki et al. (1996)'s system, the drawing of the sketches is constrained by the use of a special rastered paper comprising fiducial marks as well as pens with specified widths. Their analysis of the scanned plans is split up into "line elements" (walls etc.) and "region elements" (doors etc.) and shows a high recognition rate independent of the designer and the drawn plan.

In their survey paper, Yin et al. (2009) give an overview over existing approaches in the field of analysis of architectural drawings as well as a general outline of the necessary processing steps of such systems (see figure 4.1). The authors' main differentiating factor is the use of digital CAD files or scanned paper plans in the works of Lewis and Séquin (1998), So et al. (1998), Lu et al. (2005), Dosch et al. (2000) and Or et al. (2005a). They are rated relative to the degree of automation and robustness, and advice for their improvement given.

By using existing CAD files, Lewis (1996) leaves out all the difficulties introduced by image processing on scanned plans. In the Master's Thesis (Lewis, 1996) and the follow-up paper (Lewis and Séquin, 1998) the steps which are necessary to process CAD files to topologically and geometrically correct 3D models are described.

Ah-Soon and Tombre (1997) categorize the construction planning process into the phases "sketches", "design drawings" and "detailed construction plans" and motivate their choice of scanned design drawings to be interpreted and reconstructed by the facts that sketches will be too different in style and construction plans too detailed for an automated analysis. Their approach comprises low-level processing steps such as segmentation into thick and thin lines, graphics and text, vectorization, detection of arcs and detection of loops. Two higher-level approaches are "structure analysis" consisting of an identification of walls (=thick lines), windows and doors, and "room analysis" which analyses detected loops, classifying them to rooms (which contain a door) and cupboards, and similar features.

This approach is the base for the complete system for the analysis of scanned CAD plans presented in Dosch et al. (2000). In addition to the segmentation steps already presented by Ah-Soon and Tombre (1997), they

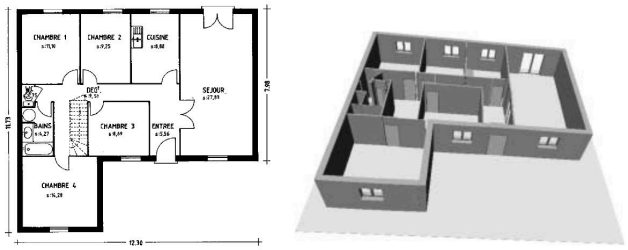


Figure 4.2: Input and result as presented in Dosch et al. (2000) (figures adapted)

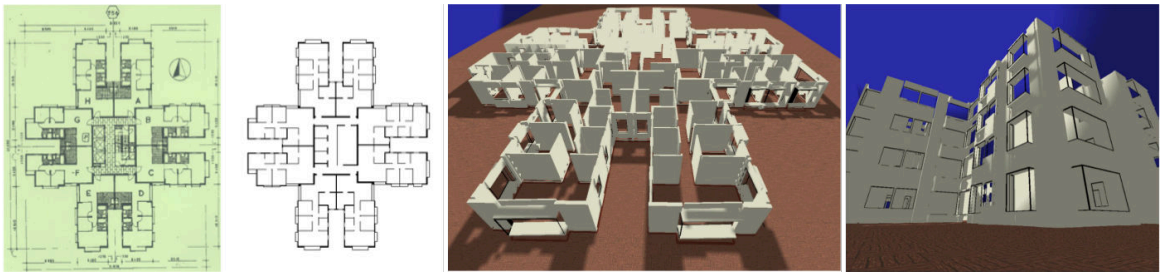


Figure 4.3: Input, preprocessed plan and 3D results as presented in Or et al. (2005b) (figures adapted)

employ the human operator for the interactive removal of misdetected dashed lines and detect staircases as well as complete symbols like doors. Their system comprises a graphical user interface and allows for the generation of 3D models using a floor height defined by the operator. Different floors are matched with the help of common features (corners, staircases, pipes). An exemplary result is depicted in figure 4.2.

Following the classification presented by Ah-Soon and Tombre (1997), Lu et al. (2005) concentrate on the analysis of detailed construction plans, aiming at the determination of material quantities necessary for the execution of the construction work.

Or et al. (2005b) use scanned images to derive a binary image and remove the contained text and symbols depicting e.g. furniture employing the human operator. The actual analysis is carried out on contours detected in the cleaned binary image. It includes wall classification by means of the analysis of opposite edges, door detection using the identification of curved polylines, window detection as narrow, adjacent boxes and room detection using (manual) scaling to metric scale. The detection of the outer contour of the model is employed in the differentiation between windows and other features.

More recent works around this topic can be found in Kashlev (2008), who is working on CAD files and thus is concentrating on the deduction of topologically and geometrically correct 3D models from the provided 2D ground plans (similar to Lewis, 1996). Using scanned paper plans as input, Ahmed et al. (2011, 2012) present a complete system for the detection and labeling of rooms. To this end, they implemented steps like graphics/text and thick/medium/thin line segmentation as well as wall detection and gaps closing. The room detection after gaps closing is carried out using connected component analysis supported by the analysis of text labels detected in the image.

A commercially available software for the semi-automatic analysis of scanned paper plans is represented by PlanTracer[1], a plug-in for AutoDesk's Architectural Desktop and AutoCAD. PlanTracer asks the operator to provide the scale for at least one element in the plan and allows for a check of the architectural symbols

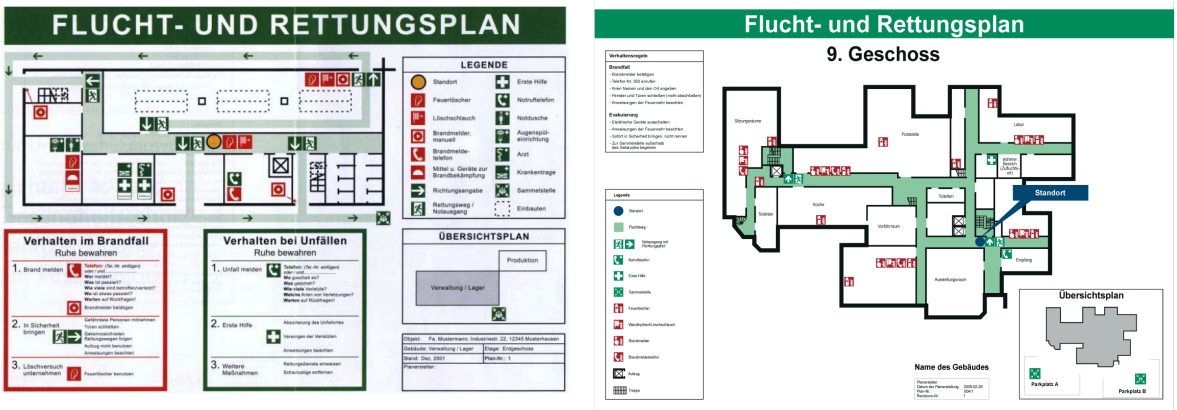
[1] `http://www.csoft.com/productcontent.php?prodid=6` (last visited 2nd February 2016)

which were automatically detected in the provided plan image. The final result is a CAD file consisting of layers containing architectural elements like windows, doors and furniture.

In contrast to these related approaches dealing with CAD files or scanned plans, the approach developed in the course of this thesis uses photographed evacuation plans together with a model of the building's external shell to deduce geo-referenced 2D and 3D models. Obviously, one of the main difficulties is the use of photographs in contrast to high resolution scans with complete control over the employed lighting conditions. This poses problems such as brightness differences throughout one image and between different images, resulting also in differing colour reproduction. While the colour reproduction is not an issue when dealing with CAD models and occlusions are minimal, the automatic analysis of evacuation plans suffers from occlusions caused by coloured symbols depicting evacuation paths or safety equipment. Many of the aforementioned approaches use symbol detection for the identification of door arcs and similar features. The symbols used in evacuation plans, however, can generally only be detected using the color information which generally is deteriorated by the lighting conditions. Finally, most approaches involve at least some kind of manual interaction, be it for the cleansing of the binary image, text removal or the solution of the scale problem. Here, the latter problem is solved automatically by matching the indoor model to a model of the building's external shell.

4.2 Evacuation plans

For the developed reconstruction method for building interiors, photographed evacuation plans are used as one of the data sources. The most important reasons to choose these plans consist 1) in their ubiquitousness, and 2) their design's degree of standardization. In fact, their ubiquitousness is part of their standardization as the International Organization for Standardization (ISO) defines evacuation plans as "essential part of the safety equipment of a physical structure" in DIN ISO 23601. DIN ISO 23601 is the international norm which aims at the harmonization of the various laws and bylaws regulating the design of such plans on national levels. In Germany, for example, the laws and bylaws ensuring the availability of evacuation plans in all publicly used buildings as well as at all workplaces comprise the „Berufsgenossenschaftliche Vorschrift für Sicherheit und Gesundheit bei der Arbeit (BGV-A8)" and the DIN 4844-3.



(a) (b)

Figure 4.4: Evacuation plans according to a) BGV A8 and b) DIN ISO 23601

Figure 4.4 shows the plan specimens - from BGV-A8 and DIN ISO 23601 respectively - which serve as illustration of the design principles. These design principles range from the definition of specific element sizes and compulsory elements to rules for the plan's location and materials to use for its production. Table 4.1 contains the most important design principles of both norms which are usable in the context of an automated analysis. This comparison exhibits that the norms agree in many standardized features supporting an automated analysis, as shown in the last column. For example, the colour specifications can be used in the context of image enhancement (white background reconstruction and colour correction, see section 4.4.3) and binarisation (facilitated differentiation between background and ground plan lines), or in the detection of coloured symbol regions. Minimum size constraints for the symbols and the plan itself as well as the compulsory use of standardised safety signs (square/rectangular) constrain the size of symbols in the image.

Feature	*DIN 4844-3:2003*	*DIN ISO 23601:2010*	*Usage in automated analysis*
use of colours	yes		detection of symbol areas, occlusions
background colour	white		color correction, facilitated binarisation
ground plan colour	black		facilitated binarisation
colour evacuation routes	light green/darker green (stairs)	arrows green; evacuation routes lighter green	detection of symbol areas, occlusions; information about the direction of staircases; information about the location of doors
minimum plan size	DIN A3		size constraints for the symbol detection
symbol size	preferable 10mm / "you are here" at least 10mm	minimum height 7mm	
standardised symbols	BGV A8 and DIN 4844-2	ISO 7010	templates for template matching based approach
correct positional arrangement	yes		computation of the initial orientation for positioning (see chapter 5)

Table 4.1: Overview over design regulations in DIN 4844-3:2003 and DIN ISO 23601:2010

Despite such detailed (and in theory) internationally applicable design rules, the reverse-engineering approach was designed to be sufficiently general instead of sticking too closely to the standardization documents. This results from, first, the fact that basing only on the German regulations would theoretically constrain the applicability of the approach to only German plans. Second, the regulations leave space for interpretation (stating in the DIN 4844 that at least "all plans inside one facility have to be in the same layout") and therefore plans differ to some extend. Third, the DIN ISO 23601, as a relatively young international standard, may not be followed everywhere yet. The plans found during data acquisition (in hotels and public buildings in various countries) support these assumptions. However, the data acquisition showed that the rules common to DIN 4844 and DIN ISO 23601 - i.e. the use of coloured symbols and black ground plans on white backgrounds - are used in the vast majority of plans.

The use of photographed evacuation plans for reconstruction can be seen in parallel to aerial images released by their owners being employed as base data in the OpenStreetMap modeling strategy. However, in contrast to those images, the legal situation is not as clear in the case of photographed evacuation plans. Thus, when proposing a manual modelling strategy for OpenStreetMap similar to the one presented in the

OpenStreetMap Wiki[2] in the OpenStreetMap mailing list[3], this caused a certain amount of discussion. Most importantly, references to the Berne convention for the Protection of Literary and Artistic Works were raised. This convention states - in short - that all works[4] should be copyrighted in all signing countries for at least 50 years after the author's death or, if the author is unknown, for 50 years after the first publication. Replies contained that a) not every country has signed the Berne convention and b) modeling from such a photograph instead of publishing the image itself is not covered by the convention. However, a building's interior design itself is creative work covered by the Architectural Works Copyright Act.

This shows that this topic seems to live in a legal grey zone, an opinion which is further supported by statements made among others by Marcus Götz, the author of the article in the OpenStreetMap Wiki (see above), during the Q&A session after his FOSSGIS 2012 talk[5]. Following the argumentation of a lawyer commissioned by him, the design of the evacuation plan is protected, however, the content is not. Similar opinions as mentioned above are expressed: the photograph might not be suitable for presentation, but the derived model might be. Furthermore, the evacuation plan is publicly available and thus not protected. However, conflicts with local regulations given by the building owner or "house rules" might arise.

4.3 Overview

The processing pipeline from the photographed evacuation plan to a final 2D or even 3D model is depicted in figure 4.5. A significant part of it consists of pre-processing steps needed to overcome flaws of the image e.g. perspective distortions and an adulterated colour reproduction (described in section 4.4). The pre-processed image is then binarised (section 4.7) and matched to an available model of the building's external shell in order to geo-reference and - more importantly - scale the model (section 4.5). The symbol areas detected by colour or shape constraints and template matching (section 4.6) are bridged, resulting in a final 2D model (section 4.8). If the plan contains stairs which can be detected, this 2D model can be extruded to a 3D model (section 4.9).

Some of the methods described here were developed in joint work with a Diploma Thesis candidate supervised by the author of this thesis (Otto, 2014). While the Diploma Thesis concentrated on plans with a well-known design, the other approaches aim at generalising the process in order to allow for the reconstruction of arbitrary evacuation plans.

[2] `http://wiki.openstreetmap.org/wiki/IndoorOSM` (last visited 2nd February 2016)

[3] `https://lists.openstreetmap.org/pipermail/talk/2011-September/060227.html` (last visited 2nd February 2016)

[4] except photographic and cinematographic works for which different regulations where stated

[5] a recording can be found under `http://db0smg.afug.uni-goettingen.de/~fossgis/FOSSGIS2012-468-de-indoor.mp4`

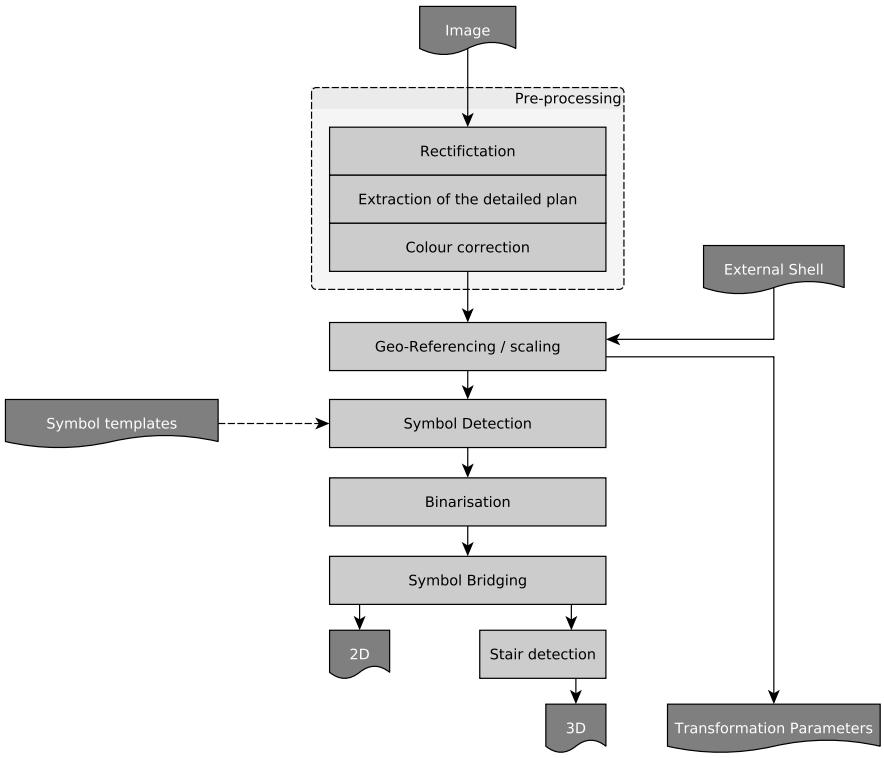


Figure 4.5: Flow chart of the processing pipeline from photographed evacuation plan to 2D/3D model

4.4 Image pre-processing

In contrast to the methods described in section 4.1 which base on scanned, high-resolution images, the photographs used here suffer from brightness and colour differences as well as reflections. Due to varying lighting conditions, this is not only true for images taken with different cameras, but also for different images taken with the same camera and for different regions inside a single image. Furthermore, in order to avoid reflections deteriorating the image quality, it may need to be taken at an oblique angle instead of perpendicularly. Section 4.4.1 covers a method to overcome the latter problem while sections 4.4.2 and 4.4.3 describe steps carried out to repair colour and brightness differences throughout the image.

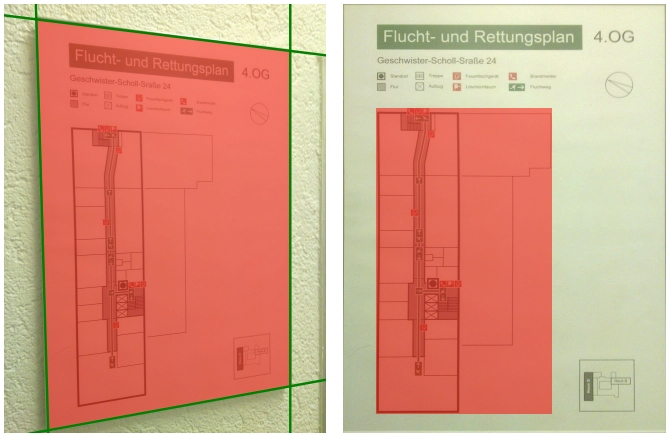
4.4.1 Image rectification and detection of the detailed plan

If the plan was photographed at an oblique angle, the image has to be rectified in order to be usable for the reconstruction. The perspective transformation parameters for the rectification can be recovered if parts of the frame of the paper the plan is printed on are visible in the image, similarly to the whiteboard scanning method presented by Zhang and He (2007).

To this end, the following steps are carried out:

1. In order to reduce image noise and remove other small structures disturbing the further analysis, the image is blurred heavily, and a bilateral filter is applied

2. edges in the blurred image are detected using the Canny operator (Canny, 1986)



(a) (b)

Figure 4.6: Rectification of the image and detection of the detailed plan

3. dominant lines are detected in the edge image using the Hough transformation and merged using the buffering operation presented by Kada (2007) in the context of cartographic building generalisation (described in more detail in section 4.5)

4. the topmost, lowermost, leftmost and rightmost dominant lines are selected and intersected

5. from the intersection points, the perspective transformation can be computed (see Appendix A on page 137), setting the image corners as target points. This will result in small distortions if the image's aspect ratio does not match the plan's aspect ratio. However, these distortions will be removed by the estimation of an affine transformation during the geo-referencing operation (see section 4.5).

In figure 4.6a) the detected and merged lines are shown, while figure 4.6b) depicts the rectified and cropped image. Furthermore, the image can be cropped to the detailed plan which will be used further in the processing pipeline. This is done using the following two steps: 1) binarise the rectified and cropped image using adaptive thresholding, 2) detect connected components and select the biggest contour (using Suzuki et al., 1985). Based on this contour's bounding box, the sought-after detailed plan is selected (see red rectangle in figure 4.6b).

4.4.2 Background normalisation

To overcome the brightness differences in a single image, Otto (2014) implemented a two-step process consisting of retroactive white balancing and the normalization of the image's background intensity. The idea for the white balancing method is taken from the image editing software "The GIMP"[6]. It consists of analysing both ends of the image's grey value histogram and discarding pixel colours which only a certain percentage of pixels use (representing dust or similar disturbances). While in GIMP a threshold of 0.05% is used, Otto (2014) sets it to 5%. This is followed by a histogram stretching operation which restores maximized contrast by ensuring the use of the full intensity interval of [0;255].

[6] `http://docs.gimp.org/2.8/en/gimp-layer-white-balance.html` (last visited 2nd February 2016)

The normalization of the image's background intensity bases on the idea of reconstructing the background brightness by removing all the foreground structures[7]. The foreground removal is implemented as a morphological opening of the grey scale image with a big structuring element[8]. Subsequently, the normalized image can be computed by dividing the grey scale image by the background image. The disadvantage of this approach is found in the fact that it enables only a correction of the grey scale image with the aim of allowing for a more robust binarisation. However, an overall correction of the image's colours is not feasible.

4.4.3 White background reconstruction and colour correction

In order to further improve the colour-based symbol detection an approach for the correction of the colour image was developed. This colour correction approach bases on the design guidelines presented in section 4.2, exploiting the white background which can be expected in a photograph of an evacuation plan. The colour correction is carried out after converting the image to the CIE L*a*b*[9] colour space. This colour space is based on the human colour perception, representing colours in a three-dimensional coordinate system whose axes are the colour's Lightness (L*) as well as a chromacity value with respect to the red/green axis (a* axis) and the yellow/blue axis (b*). In the original definition, the L* channel's domain is defined as [-100;100] and the a*/b* channels vary between -128 and +127. However, in order to enable their storage in 8 bit images, they are often scaled to [0;255], resulting in the white and black point being at (255,128,128) respectively (0,128,128). Due to these characteristics, a correction of the L* channel affects all colours present in the image.

Figure 4.7a) shows one of the expected problems of the image's quality. The cyan line represents the L* channel values of a single image column close to the border of the imagewhich is assumed to be completely white (see figure 4.8a)) before any correction. A brightness decrease towards the ends of the column (i.e. the image border) is clearly visible, caused by a lens limitation commonly called vignetting. Depending on the lighting conditions, this defect induced by the lens can be overlaid by a linear trend if the light is distributed inequally throughout the image (see dashed red line in figure 4.7a)).

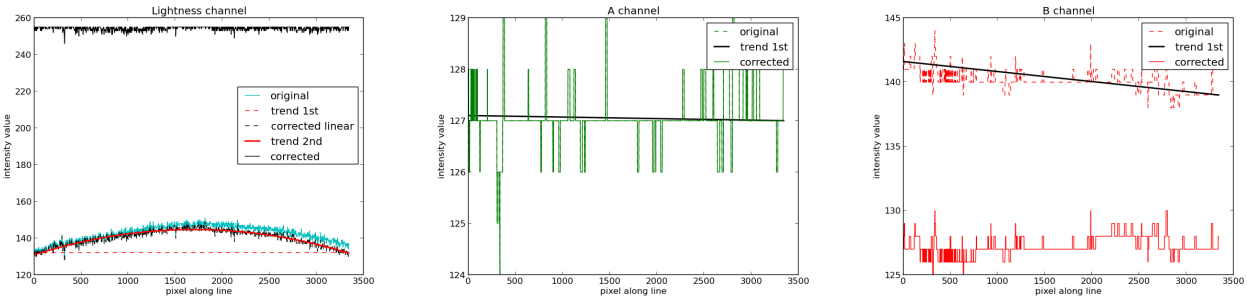
For the correction of the L* channel image, first the linear trend is estimated using the 10 first and 10 last pixels of the line. The result of the linear trend's removal from the original data (figure 4.7a), cyan line) is depicted as a black dashed line. In a second step, the offset parameter from the estimated linear polynomial is used together with a threshold of 25 to select L* channel values which allegedly represent the white background in the image. Finally, a second degree polynomial is estimated using the selected values (figure 4.7a), red solid line). The subtraction of this trend from the data and a simultaneous translation of the mean to 255 results in the final data, depicted in figure 4.7a) by the solid black line.

The partitioning into lightness and chromacity channels delivered by the conversion to the CIE L*a*b* colour space ensures that the brightness differences are completely corrected by the L* channel correction. However, the chromacity channels may contain slight colour changes throughout the image and, more importantly, tints which are reflected in a translation of the white point away from its nominal value of 128

[7] `http://dsp.stackexchange.com/questions/1932/what-are-the-best-algorithms-for-document-image-thresholding-in-this-example` (last visited 2nd February 2016)

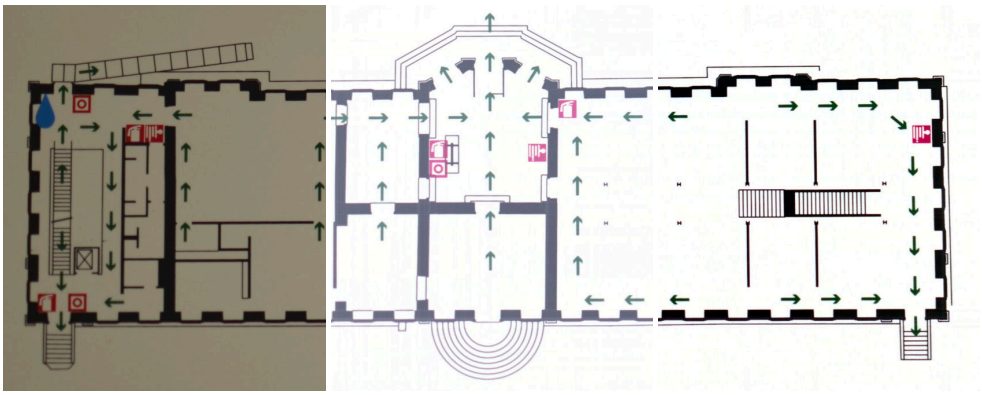
[8] with a size of e.g. 1/3 of the shorter image edge

[9] DIN EN ISO 11664-4



(a) (b) (c)

Figure 4.7: Corrections along one image row for white background reconstruction and colour correction



(a) (b) (c)

Figure 4.8: Original (a), colour corrected (b) and histogram stretched (c) images

(see figure 4.7c) in comparison to figure 4.7b)). Thus, for the a* and b* channels only the linear trend is estimated and corrected in combination with a translation of the mean value to 128. The corrected image is converted back to the RGB colour space.

As visible in figure 4.8b), the thusly enhanced image lacks contrast. To overcome this flaw, it is treated with the white balancing/histogram stretching method described before (using a threshold of 0.05% for the pixels to be discarded), resulting in the image depicted in 4.8c).

4.5 Geo-referencing and scaling

A common problem of the related approaches presented in section 4.1 is the definition of the coordinate transformation between the resulting image coordinates and a metric reference system. Without its definition in metric coordinates, the resulting model can serve visualization purposes at most. In the related work, the metric scale is mostly derived using human interaction (e.g. definition of one or more edge lengths). While scaling the model is sufficient for applications like indoor positioning and navigation as well as interior architecture, geo-referencing it in a world coordinate system allows for a seamless navigation between outdoor and indoor areas as well as the co-registration of different floors.

Additionally, the metric scale is needed for the approach presented here in order to define metric thresholds for the symbol bridging operation. Therefore, an automatic geo-referencing method was developed which

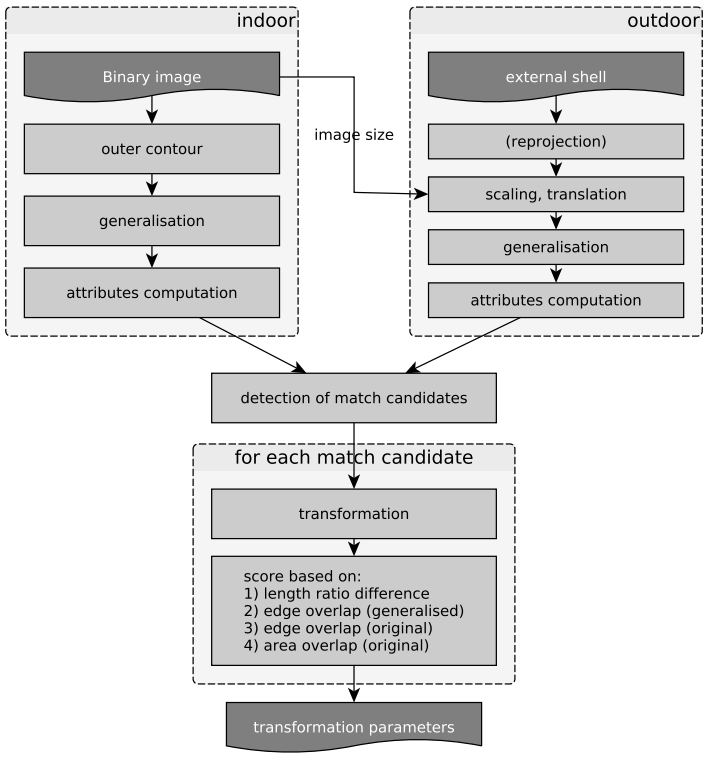


Figure 4.9: Flow chart depicting the geo-referencing process

makes use of contour matching between the model depicted in the evacuation plan and a model of the building's external shell available in a world coordinate system. The flow chart in figure 4.9 depicts the whole process which is described in depth in the following sections.

4.5.1 Preparation of the contours

The outer contour of the floor plan depicted in the evacuation plan - in the following called *indoor contour* - is detected in a binary image derived using adaptive thresholding on the pre-processed image. Small gaps are removed using morphological closing, followed by the detection of the contours contained in the image using the topological structural analysis approach described in Suzuki et al. (1985). If the image was cropped to show only the detailed plan, the sought-after indoor contour is found as the largest contour by area.

The external shell, on the other hand, may be given to the process as a ground plan or 3D model from an available city model, or, alternatively, can be taken from OpenStreetMap. Depending on the source, the first step is the projection of the polygons' coordinates to a projected coordinate system (e.g. Gauß-Krüger or UTM). Subsequently, the centroid of the model is subtracted and the model is scaled to fit into an image twice as large as the longest edge of the evacuation plan image by applying the following two equations:

$$c_{im} = \frac{X_{proj} - \overline{X_{proj}}}{s} + \max(r_{im}, c_{im}) \tag{4.5.1}$$

$$r_{im} = \frac{\max(Y_{proj} - \overline{Y_{proj}})}{s} - \frac{Y_{proj} - \overline{Y_{proj}}}{s} + \frac{\max(r_{im}, c_{im})}{2} \tag{4.5.2}$$

where X_{proj}, Y_{proj} are the coordinates of the external shell in a projected coordinate system, r_{im}, c_{im} are the row/column image coordinates, $\max(r_{im}, c_{im})$ is the maximum of the image dimensions, $\overline{X_{proj}}, \overline{Y_{proj}}$ are the coordinates of the external shell's centroid and $s = \max(\max(X_{proj}) - \min(X_{proj}), \max(Y_{proj}) - \min(Y_{proj}))$ $/\max(r_{im}, c_{im})$ is the approximate scale derived from the comparison of the longer edges of both contours' bounding boxes. This scaling ensures that the same parameters can be used in the following cartographic generalisation for both the indoor and the outdoor contours. The parameters of equations 4.5.1 and 4.5.2 together with the affine transformation estimated in the last step of the process define the reverse transformation from image coordinates to the world coordinate system.

4.5.2 Cartographic generalisation

Due to their differing scales, in most instances both contours will show quite different levels of detail, a fact which could prevent the applicability of the following contour matching approach. In order to overcome this problem, they are simplified using a cartographic generalisation method for building ground plans similar to the one presented in Kada (2007). Kada (2007)'s shape-preserving generalisation method for 3D building models bases on half-space modeling and cell decomposition. Here, only the ground plan simplification component is used, based on the ground plan edges and not the faces connected to them.

The essence of the approach is the aggregation of similarly oriented edges of the contour using a buffering operation. To this end, in a first stage the buffers are constructed from the edges using a normal vector based line representation (normal vector plus minimum and maximum distance) and the edges' weight deduced from their length. Then, the respective strongest buffer is determined by sorting the buffers by their weight and checking for other buffers which are mergeable (fulfilling the angle threshold and not causing the buffer to grow bigger than the distance threshold) or includeable (buffers not fulfilling the angle threshold and not causing the buffer to grow bigger than the distance threshold). The buffer merging operation causes the recomputation of the buffer's normal vector as well as its bounds, while the including operation only affects the bounds. This process is repeated in a loop and produces the respective strongest buffer for which the distance parameter is recomputed as the weighted mean distance of all buffers which were merged.

In a second stage, the lines derived from the found buffers are intersected, producing a cell decomposition representation of the sought-after simplified polygon. From the set of cells the ones which fulfil a threshold for the area overlap with the original polygon are selected and the final generalised model is constructed as the union of these cells.

While strongly simplifying the building model, this approach manages to retain the model's overall structure very well. The choice to compute the weighed mean distance results in simplified models which are optimized towards *area trueness* in comparison to the original models (Kada et al., 2009a). While this is a preferable feature in many cases, often the simplified models are requested to fulfil characteristics like a maintained common façade line or - more generally - *trueness to the nodes* of the original model. Thus, an extension to Kada (2007)'s approach was developed which aims at adjusting the generalised model to the original ground plan (Peter et al., 2008; Kada et al., 2009a; Peter, 2009). The extension consists of two steps: 1) the adjustment of the generalised ground plan to the strongest line constructible from each buffer and 2) the 3D model's adaption to the adjusted ground plan. The ground plan adjustment is done by employing two more strict thresholds for distances and angles in order to build the strongest constructible line instead of computing the line representing each buffer as the weighed mean.

4.5.3 Contour matching

With the availability of the generalised indoor and outdoor contours, the transformation parameters can be determined by matching the contours. Contour matching approaches for general polygons are available in related work, put to use for example in automatic puzzle solving or reconstruction of disrupted documents (Kong and Kimia, 2001). Veltkamp (2001) describes more applications for shape matching and the cores of various algorithms, their similarity measures, as well as the voting schemes which are used to compute the final similarity measure.

In the case of the generalised indoor and outdoor contours the problem is less general, as right angles will be most prominent and the polygons will supposedly be very similar after generalisation. Thus, a custom contour matching method was developed which leverages the per-node length ratio of the adjacent edges as well as the angle enclosed by them. Based on the length ratio difference and the angle difference, nodes fulfilling very loose thresholds[10] are chosen as match candidates.

For each of the match candidate nodes, the indoor and outdoor contour nodes themselves as well as their respective two neighbours are used to determine the parameters of an affine transformation (see Appendix A on page 137) which is used to transform the original and the generalised indoor contours[11]. Based on the transformed generalised indoor contour and the generalised outdoor contour, the following scoring parameters are computed: a) the mean length ratio difference of all nodes' adjacent edges and b) the edge overlap, determined using a buffer around the outdoor contour which has the size of the generalisation distance (as used in the quality assessment method presented in Filippovska (2012)). Additionally, c) the edge overlap and d) the area overlap between the original outdoor contour and the transformed original indoor contour are computed (see figure 4.10).

The best set of transformation parameters is chosen based on these scoring parameters, firstly scoring by a lower mean length ratio difference, followed by the decision based on a higher edge overlap. In the case of symmetric generalised models, however, these parameters may deliver an ambiguous result. Thus, the edge overlap as well as the area overlap between the original contours are also taken into account (see figure 4.10).

As an improvement, additional homologous points in the transformed indoor contour and the outdoor contour can be identified. These can be used either to improve the estimation of the affine transformation or to estimate the perspective transformation instead. However, this is generally not necessary, as the perspective distortions are removed during pre-processing already (see section 4.4).

The resulting transformation parameters may also be used to transform the photographed plan. Thus, it can serve as a backdrop to visualize positions and pedestrian traces without explicit reconstruction of a model.

[10] the threshold for the length ratio difference was chosen as 0.5, the angle difference threshold was set to 10° and kept for all data sets which were tested

[11] instead of a similarity transformation the affine transformation is needed, because the evacuation plan might be differently scaled along the two coordinate axes

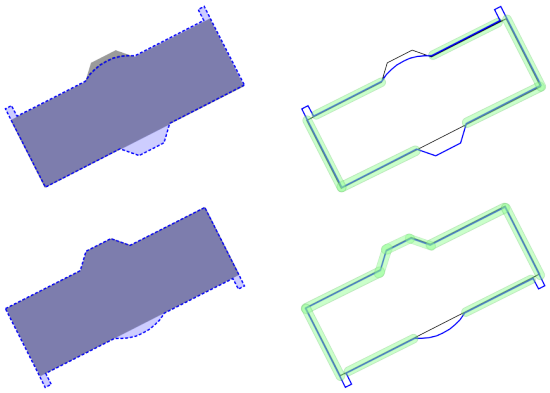


Figure 4.10: Symmetric generalisation result: decision based on area overlap (left) and edge overlap (right) between the original contours

4.6 Symbol detection

In contrast to the scanned architectural plans which serve as input to most of the methods described in section 4.1, photographed evacuation plans contain symbols depicting the location of emergency equipment (e.g. fire extinguishers, water hoses, communication facilities) or evacuation routes. As emergency equipment often is installed at walls, its symbols are likely to occlude parts of walls, while evacuation route symbols often are drawn over doors or stairs. Thus, in order to reconstruct a complete model, the symbols have to be detected and bridged. Apart from this most important reason, the symbols should be detected in order to represent the information as well as meta-information they contain in the final model. One example for meta-information is the down[12] direction of a flight of stairs which is covered by an evacuation route symbol, as evacuation routes are much more likely to lead to lower levels of the building than to the roof. Thus, the detection of symbols is a crucial part of the evacuation plan analysis.

4.6.1 Symbol detection using contour detection and template matching

The symbol detection method described in Otto (2014) bases on the facts a) that the input plan's layout is known, including the symbols which are available as templates, and b) that the symbols are surrounded by a white padding which separates them from the structures they occlude. The method consists of the following steps:

1. The first step of Otto (2014)'s method aims at the detection of the predominant symbol size in the image of the plan. As the layout of the plans used in the experiments closely follows the DIN 4844-3, the symbology consists of squares (emergency equipment, evacuation route arrows) and rectangles with an aspect ratio of 1:2 (evacuation route arrows with running person). The red signal colour of the emergency equipment symbols can be detected robustly by converting the image to the CIE L*a*b* colour space (see also section 4.4) and analysing the a* channel with a simple threshold. Using connected component analysis (Gonzales et al., 2004) with the resulting binary image, the red symbols' contours can be detected which are filtered by their squareness as well as by a size threshold derived from the size constraints stated in the norm. The predominant symbol size results as the median of all remaining bounding boxes' edge lenghts.

[12] if the floor is in an overground level

2. The background normalised image (see section 4.4) is binarised using adaptive thresholding. Using a second connected component analysis, all contours (rooms, stairs, symbols) can be extracted from the binary image. Here, the aforementioned white padding around symbols is leveraged, as it results in all symbols being found as separate contours. Symbol candidate contours are identified using the symbol size computed before by filtering the contours, only allowing those with sizes and aspect ratios *(symbol size):(symbol size)*, *(symbol size):(2 x symbol size)* or *(2 x symbol size):(symbol size)*. Further constraints include minimum length constraints for both the short and the long bounding box edges which depend on the symbol size, as well as a constraint excluding round contours.

3. As the templates for the symbols are available e.g. from the norm documents or extracted from one plan's legend, in theory, an identification of the symbol's meaning would be possible using template matching. In practice, however, the detection depends on thresholds which would have to be defined for each photographed plan separately, as Otto (2014) states. In the thesis, this approach is demonstrated for the "you are here" symbol only, which is very distinct from the other symbols. The fact that the regions of interest were detected before and their dimensions are known helps robustifying the method, as the template can be scaled to the correct size before matching as well as only the regions of interest have to be analysed.

This method, however, does not allow for a detection of the evacuation routes which are depicted as simple green lines between the evacuation route arrow symbols. Otto (2014) describes three different ways to detect this missing information: 1) detection of the remaining green areas and filtering using a size constraint, 2) detection of isolated contours between evacuation route arrow symbols in the direction of the arrows or 3) skeletonisation and vectorisation of the binary image cleaned by the symbol areas and detection of isolated line strings.

4.6.2 Symbol detection using colour segmentation

If the two prerequisites mentioned at the beginning of the previous sub-section are not met, the method developed in Otto (2014) will fail. For one, the symbols may not be rectangles but have arbitrary shapes (like arrows etc.). Secondly, they may be connected to wall structures, which renders their detection based on the contours' shapes in the binary image impossible.

Thus, a more general method based solely on the symbols' colours was developed. To this end, a colour segmentation method called Color Structure Code, presented in Priese and Sturm (2003) is used. Due to its hierarchical nature, this region growing approach uses local and global information to split the colour image into similarly coloured regions. Its result is represented by a label image and an image containing the similarly-coloured regions filled with their mean colour. The colour detection based on these regions was found to be more stable than based on the analysis of single pixels.

The colour classification is carried out on the CIE L*a*b* representation of the region image which results from the Color Structure Code as follows:

1. Detect the maximum and minimum values along the a* and b* axes, map the values to $[0;1]$ by subtracting 128 and dividing by *(max-128)* (positive values) or *(128-min)* (negative values).

2. Red and green are assumed to be existing in all plans in order to signalize emergency equipment and evacuation route symbols. The existence of blue and yellow is checked by $min(b^*) < min(a^*)$ (blue) and $max(b^*) > max(a^*)$ (yellow).

3. The final colour classification is done by means of the following thresholds (in this order): $a^* > 0.7$ and not $b^* < 0.4$ (red), $a^* < 0.4$ (green), $b^* < 0.4$ (blue) and $b > 0.6$ (yellow).

While this approach is more generally applicable than the template-based method, it is less robust (as building on less external information). Thus, coloured regions might be left undetected and be interpreted as structure to be included in the final model. Although these errors in the model must be corrected by manual interaction, the method nonetheless saves digitization time in all correctly vectorised parts of the model.

4.7 Binarisation

Using the images resulting from his aforementioned background normalisation method, Otto (2014) reports that the use of the normalised binarisation method presented by Sauvola and Pietikäinen (2000) delivers the best results for the layout of the plans used in his study. This method was developed for the binarisation of scanned documents and is optimised towards slight brightness differences in the image and black letters on white background. To this end, a threshold is selected per pixel based on the mean value as well as the standard deviation in a local neighbourhood.

In addition to Otto (2014)'s approach, a custom binarisation method was developed, motivated by two reasons: 1) evacuation plans may contain areas shaded in grey which depict hallways or other regions of interest, 2) light green areas like the evacuation routes which can be seen in figure 4.4b) are not detected by the symbol detection method described before. Depending on the neighbourhood size parameter chosen for the adaptive thresholding method, both of these may deteriorate the binarisation result by areas falsly categorized as foreground.

Therefore, a global thresholding method was developed which takes into account that - in addtion to white - a second colour may exist in the image which must be categorized as background. To this end, the following steps are carried out:

1. Set the detected symbol regions to white in the grey scale image.

2. Compute the histogram, smooth it.

3. Detect local extrema in the smoothed histogram, identify the first (black) and last (white) distinct maxima.

4. Search for a distinct maximum between the black and white values

    a) if such a maximum exists use the minimum before as the threshold for the binarisation, or

    b) use a fixed threshold of 200.

4.8 Symbol bridging and 2D modelling

In order to produce a complete final model, structural information that was occluded by the symbol regions has to be restored. To this end, the scale factor derived during the geo-referencing step is used to enable the definition of metric thresholds.

The following process was implemented in order to reach this goal:

1. The binary image is cleaned from the symbol regions and the cleaned binary image's skeleton is computed using the iterative thinning method presented by Zhang and Suen (1984). The iterative thinning method was chosen, as - opposed to distance transformation based methods (like Felzenszwalb and Huttenlocher, 2004) - it tends to produce less unwanted end point edges at sharp corners.
   The skeleton image represents the topological relations in the binary image as branch nodes (pixels with more than two filled neighbouring pixels) and end nodes (pixels with exactly one filled neighbouring pixel). Starting at the branch and end nodes and following the skeleton in between them, the skeleton is vectorized, resulting in an incomplete 2D vector model of the floor plan. The end nodes of this vector model serve as candidate starting points for the symbol bridging process.

2. Short edges with a length of less then 10 centimetres ending in two end points are very likely binarised and vectorised image noise and can be deleted.

3. While the vectorized image skeleton represents the topological relations very well, geometric precision may be lost during skeletonisation and vectorisation, especially at short end point edges. In order to restore perpendicularity, the adjacent edges' orientations of all end point edges shorter than 1 metre are analysed. Subsequently, the end point edge is rotated to be perpendicular to the predominant orientations of the neighbouring edges.

4. All end point edges which are longer than 10 centimetres are prolonged until either structure in the cleaned binary image or the image border is hit.

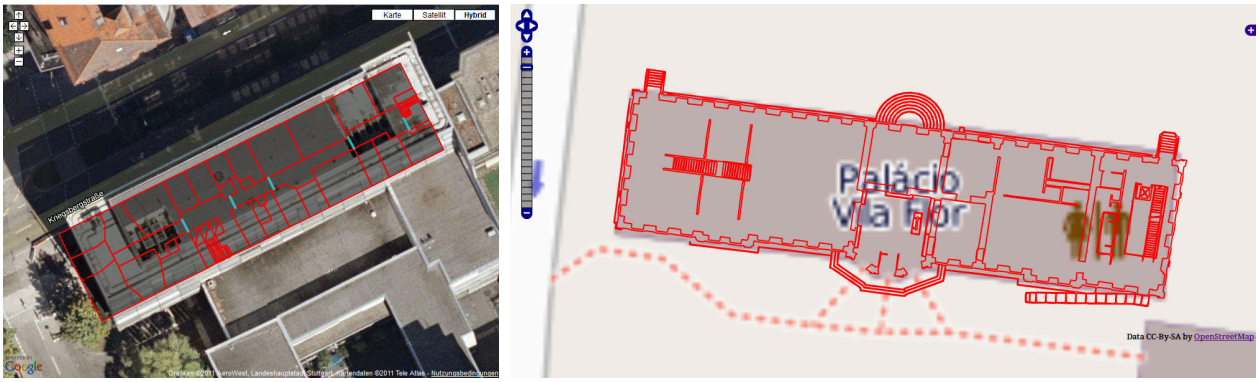
5. Intersections of the prolonged edges with the symbol contours are computed. Depending on the existence of intersections with symbol contours, the prolonged edges are treated differently:

   a) If the prolonged edge does not intersect with any symbol contour, it is selected to be included in the final model if the prolongation is shorter than 1 metre. This is motivated by the fact, that such edges can either close door openings (which are narrower than the standard door width of 0.8 metres) or small gaps in the model which are caused by binarisation errors.

   b) If the prolonged edge intersects with a symbol contour, its intersections with the other prolonged edges are computed. The parts which are entitled to be included into the final model are chosen based on the principle "longer edge stops shorter edge".

6. The final 2D model can be derived in two different ways:

   a) If a *one-sided walls* model is requested, the final 2D model is composed by the polygons which are computed as closed rings consisting of the vectorized and completed skeleton's edges. An exemplary result can be seen in figure 4.11a).



(a) (b)

Figure 4.11: Exemplary results: a) *One-sided walls* model overlaid on Google Maps, b) *Two-sided walls* model representing the wall widths as depicted in the photographed plan overlaid on OpenStreetMap

b) Furthermore, this method allows for the derivation of a *two-sided walls* model, i.e. a model representing the wall widths as drawn in the photographed evacuation plan. To this end, the width of each end point edge is computed as the median of the distance transformation's values extracted along the edge. This width is used to draw the symbol bridging edges into the cleaned binary image and the final polygons for the *two-sided walls* model can be detected using connected component analysis on the completed binary image. An example for a *two-sided walls* model is depicted in figure 4.11b).

4.9 Stair detection and 3D modeling

Evacuation plans are often produced by placing symbols and evacuation route information on a ground plan available e.g. as CAD model. Therefore, and because stairs are a very important part of the evacuation route information, it can commonly be observed that staircases are represented realistically in such plans. In combination with existing constraints for the height of stairs in public buildings (Neufert et al., 2002), this fact may be capitalised in order to derive an approximate floor height and extrude the 2D floor plan to a full 3D model. This involves the following steps:

1. Detection of stair candidates
   Candidates for valid stairs in public buildings generally fulfill constraints concerning their dimensions along their centre line and perpendicular to it. To be usable as stair, a minimum width (along the centre line) of 0.8 meters has to be fulfilled, while the platform's width (perpendicular to the centre line) must not fall below 0.15 meters and seldom exceeds 0.3 meters. The latter attribute is checked by computing the distance transformation for every polygon (Felzenszwalb and Huttenlocher, 2004) which enables the determination of each pixel's minimum distance to the boundary. Furthermore, the distance transformation can deliver the skeleton image which represents the centre line of the polygon. Figure 4.12 depicts an example for detected stair candidates.

2. Identification of staircases
   By searching for chains of stair candidates, outliers (like doors or windows) can be removed and staircases be identified. The search is conducted by computing the normal vector of each segment of the

vectorized centre line, constructing a line and intersecting it with all polygons. If the nearest intersection occurs with another stair candidate the two stair candidates are marked as being neighbours. Having identified the string of neighbouring stair candidates, the staircase can be built as a string of neighbouring stairs. Additionally, folded staircases are equalized in this step, reconstructing stairs which were previously occluded by evacuation route symbols.

3. Computation of the approximate floor height and construction of the 3D model
   If only the approximate floor height is of interest, the longest staircase is identified. Subsequently, the number of its stairs is multiplied by the common stair height found in public buildings which is 0.17 metres (Neufert et al., 2002), resulting in the approximate floor height[13]. Using the approximate floor height, the edges of the 2D model can be extruded to vertical rectangles, producing the final 3D model (an example is shown in figure 4.13). Optionally, if the direction of each part of the staircase could be identified, the stairs may be reconstructed explicitly as 3D structures.

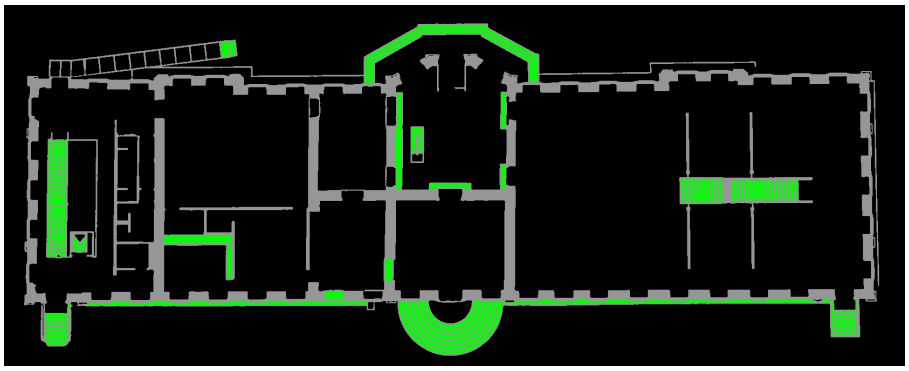


Figure 4.12: Example for detected stair candidates (green)

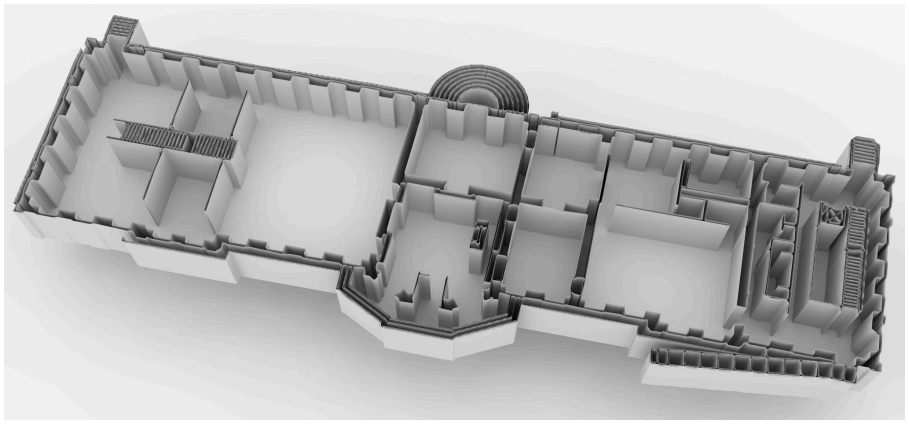


Figure 4.13: Rendered final 3D model

[13] including the height of the ceiling and/or floor

5 Reconstruction by use of pedestrian traces

Section 2.2 partly recapitulated the extensive research efforts being undertaken in the field of indoor positioning employing foot-mounted MEMS IMUs. It is obvious that, despite ongoing miniaturization efforts, a broad dissemination of foot-mounted sensors is unlikely to be the final solution to the problem of indoor positioning for the masses. However, the published works leveraging the sensor equipment of modern smartphones illustrate future possibilities to be expected, especially when combined with the projection that these sensors will improve in terms of accuracy. Thus, fairly expensive and bulky foot-mounted MEMS IMUs are used here as a "stand-in" to the performance of future indoor positioning technologies which will be available to many people.

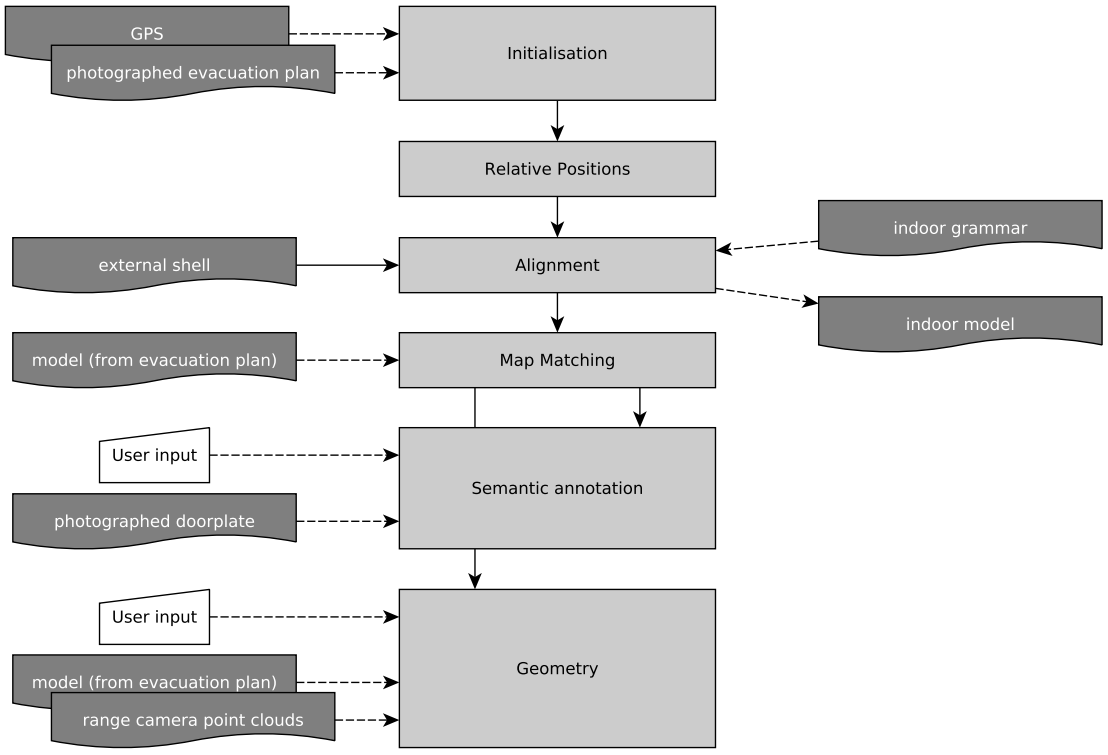


Figure 5.1: Flow chart of the pedestrian trace-based reconstruction and refinement

Figure 5.1 comprises the methods for the reconstruction of building interiors from pedestrian traces which will be presented in the following sections. Section 5.1 is dedicated to the description of the foot-mounted indoor positioning system delivering coordinates relative to the provided initial values. Furthermore, this section presents methods for the derivation of the initial values from a photographed evacuation plan, as well as supporting methods which improve the accuracy of the positioning system. These are: an alignment method using the building's external shell (analogously to the geo-referencing method described in section 4.5) and a map-matching method basing on the model resulting from the processing pipeline described

in the previous chapter. As these methods only provide a coarse model, the majority of the methods to be presented serve the refinement and update of the available model. This includes semantic information (like room numbers etc., section 5.3) as well as geometric features (e.g. doors, section 5.4) which can be acquired by pedestrians equipped with an indoor positioning system.

In addition to these refinement methods for existing models, an approach for the reconstruction of complete models from user traces without previously existing coarse model is presented which employs knowledge about the floor plan layout modelled in an indoor grammar (section 5.5).

5.1 Indoor positioning using a foot-mounted MEMS IMU

An indoor positioning system usable for the reconstruction or refinement of building interiors visited by users has to be infrastructure-independent. This is due to the fact that in an environment which is to be mapped based on the pedestrian traces the existence of external infrastructure as for example beacons emitting electro-magnetic signals or the knowledge of their locations cannot be assumed. Following the related work presented in section 2.2, the state of the art in infrastructure-less indoor positioning is the use of foot-mounted MEMS IMUs, delivering the most accurate infrastructure-independent coordinate measurements available when combined with signal processing approaches like the detection of stance phases and their use as zero-velocity updates (ZUPTs).

Figure 5.2: Foot-mounted XSens MTi-G

Initial tests of the positioning system as well as the alignment and height correction described in the next section were carried out in the context of a Study Thesis (Schenk, 2010) supervised by the author of this thesis. The implemented positioning system does not employ ZUPTs in a Kalman filter, but instead relies on a method similar to the one described by Ojeda and Borenstein (2006). Thus, the stance phase detection is implemented based on thresholding the angular rates' magnitude: If the norm of the three-dimensional gyro vector drops below 1.0 for more than half a second, a stance phase is detected (for further details of the method refer to Appendix B on page 141). In order to correct the drift that occurred since the end of the previous stance phase, the acceleration portion caused by gravity is removed from the data and the accelerations are integrated to velocities. The velocities during the stance phase are assumed to be zero, thus, their difference to this value can be used to estimate the drift. Subsequently, the swing phase velocities

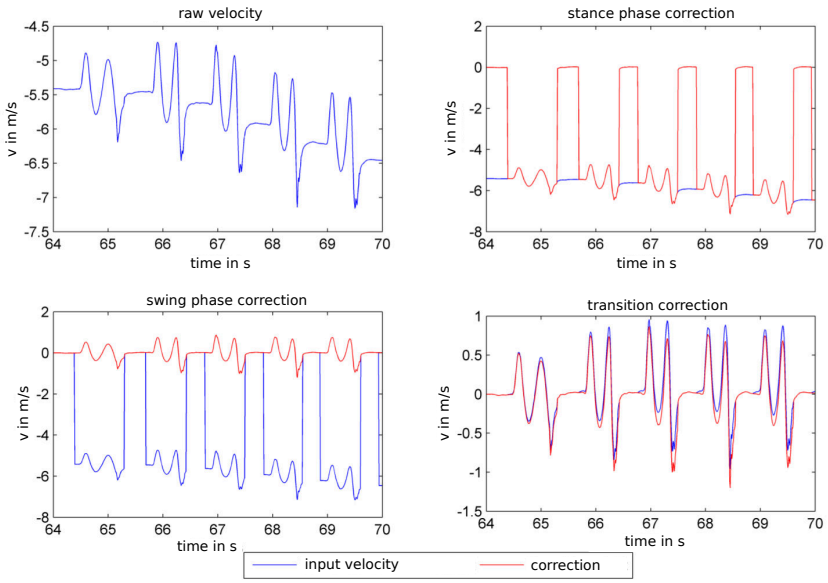


Figure 5.3: Principle of the applied velocity correction (adapted from Schenk (2010), translated)

are corrected by the drift and integrated to obtain a displacement vector, i.e. the step vector[1]. Schenk (2010), computing the full trace in post-processing, additionally describes a correction for the transitions between consecutive steps (see figure 5.3). However, in the implemented real-time positioning method, this transition correction is not needed, as the sensor can be reset after each detected step.

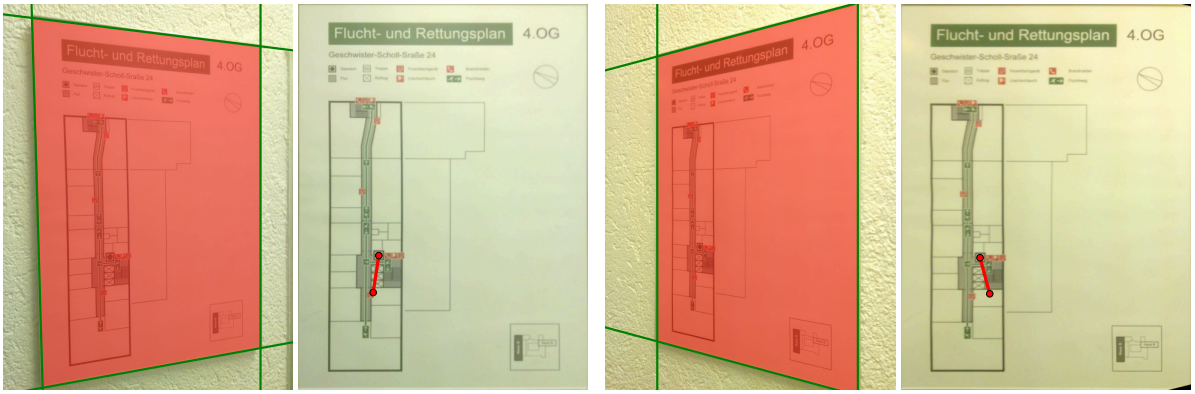
5.2 Positioning support methods

The aforementioned method greatly improves the results compared to naive double integration. Nonetheless, there still remain minor errors resulting from drift which was not completely corrected as well as from sensor misalignment. Moreover, the initial position and orientation are unknown, causing the system to only deliver coordinates without a reference to a superordinate coordinate system and thus without a reference to a map. Methods for the correction of these shortcomings will be presented in the following sections.

5.2.1 Deriving initial values from photographed evacuation plans

Positioning methods based on inertial measurement units provide local coordinates relative to an initial position and orientation. In order to describe the user's position in relation to a map, these initial values have to be defined in the map's coordinate system. Related scientific publications overcome this shortcoming either by picking both values in a map manually, scanning a QR code as in Sternberg et al. (2013) or by assuming a continuous positioning scenario and estimating both initial values from the last GPS coordinate measurements before entering the building (Alzantot and Youssef, 2012b). While the use of a possibly available magnetometer for the measurement of the initial orientation seems an obvious solution, magnetometer measurements suffer disturbances by electro-magnetic waves and metallic structures in building interiors. As will be shown, the list of possibilities for the derivation of the initial values can be extended by

[1] or, in the case of one sensor-equipped foot, the double step vector



(a) (b)

Figure 5.4: Derivation of initial orientation from evacuation plan for usage in foot-mounted IMU positioning

the possibility to derive the initial values from the photographed evacuation plan (Peter et al., 2010; Haala et al., 2011, 2012). In order to provide the viewer with his *position* in relation to the plan's information, one core element is the *you are here* symbol which is one compulsory element of such plans. In the case of plans with a known design, it may be detected using the template matching approach for symbols described in section 4.6.1. In all other plans, it should be unique in terms of colour, shape and/or size to allow for a fast differentiation from the symbols depicting emergency equipment or evacuation route information. Its uniqueness predestines this symbol for an automatic detection.

If the automatic detection is not successful, the user is asked to provide the location of the *you are here* symbol by clicking, similar to the outdoor approach presented by Cheverst et al. (2008). Using the transformation parameters between image and world coordinate system (see section 4.5), this position can be transformed to the coordinate system of the coarse model derived from the evacuation plan. However, mostly due to the fact that the distance between the photographer and the plan is not recoverable, these methods merely enable the automatic derivation of an approximate initial position. Additional factors adding uncertainty to the resulting initial position will be discussed in section 6.2.2.

For the computation of the user's *initial orientation,* one of the design principles for evacuation plans (see section 4.2) is put to use. This principle states that the detailed plan has to be oriented according to the viewer's orientation in order to improve the legibility. This results in the *up* direction in the plan's image equaling the viewing direction in reality. Like the initial position, this vector can be transformed to the coordinate system of the coarse model. Obviously, this initial orientation will be biased if the evacuation plan is not photographed perpendicularly. As described in section 4.4.1, the perspective distortions of the image can be computed if the frame of the plan is visible in the image. Apart from the rectification of the image, the availability of the perspective transformation also allows to recover an approximation of the initial orientation's unknown bias[2] as shown in Zhang and He (2007). The respective right hand side images in figure 5.4 show the resulting initial orientations.

5.2.2 Alignment

As stated in chapter 3, the arrangement of walls in indoor environments generally follows the Manhattan World constraints, resulting in strictly parallel and orthogonal walls in a coordinate system aligned to the

[2] under the premise that axis of the foot-mounted IMU and the camera viewing direction are approximately parallel

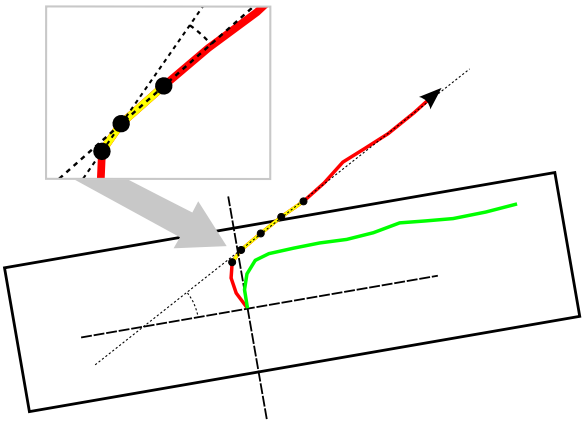


Figure 5.5: Straight line detection (inset) and alignment by rotation illustrated

external walls of the building. This is especially true for hallway structures, which - due to accessibility reasons and room size limitations in office environments or public buildings - follow the external structure of the building to a great extent. This fact is used by other alignment approaches similar to the one presented below, e.g. the HDR (Borenstein et al., 2009), HDE (Borenstein and Ojeda, 2010) or MiHDE (Jiménez et al., 2012) methods described already in section 2.2.3.

In combination with the detection of the user walking in a straight line, the assumption about corridors being parallel to elements of the building's external shell can be leveraged for the trace alignment (Peter et al., 2010). To this end, the building's orientation is computed from the building's external shell (which was used already in the geo-referencing step described in section 4.5) using the method presented in Duchêne et al. (2003).

In order to detect consecutive positions forming a straight line, the angle between their step vectors is compared to a threshold which was set to the empirically derived value of 7.5°. Consecutive step vectors enclosing a smaller angle are attributed as *line candidate*s. The correction of the misalignment angle between the detected straight line and the nearest principle axis of the building is initiated when the number of consecutive line candidates exceeds a value set by the user[3]. It is carried out as a rotation around the last point of the previous correction phase. These two operations are depicted in figure 5.5: In the red trace, four line candidate steps (yellow) are detected using the angle between two consecutive steps as shown in the inset. The angle between the line candidates' mean vector and the closest of the building's (black rectangle) main axes (dashed black lines) is then used to correct the trace by rotation. Here, the initial position is used as the rotation centre instead of the last point of the previous correction phase. The same method is also applicable if no building model is available in which case the trace is aligned to the coordinate axes. Figure 5.8 on page 89 depicts an original trace (red) as well as the trace corrected by alignment (yellow). As visible, the alignment corrects the slight drift errors occurring after both the first turn and the 180° turn at the end of the hallway, delivering a much more plausible trace.

5.2.3 Height correction using stair and elevator detection

While the support method presented in the previous section enables the correction of errors occurring in the xy plane, it cannot prevent drifting errors along the z axis. The use of a barometer is one option to

[3] three to four double steps have been determined as good values

overcome this flaw, however, barometers are prone to errors resulting from changes in the environment temperature which can happen in building interiors when changing rooms.

Alternatively, another fact can be capitalised: vertical movement of pedestrians in building interiors is highly constrained to features like ramps, escalators, stairs and elevators. Ramps are not very common and their detection was examined by Jiménez et al. (2011b), the detection of a user standing on an escalator is described in Alzantot and Youssef (2012b). Hence, only the detection of stairs and elevators and their use in height correction are described here.

Like the alignment, height correction by stair detection was tested in the context of the Study Thesis (Schenk, 2010). The initial height is set to a constant value which can either be zero, a known absolute height e.g. gathered from GPS or a height estimated from the floor number. Subsequently, the z coordinate difference between the start and end coordinates of a one step are compared to a threshold of 0.2m. This value is compiled from the standard stair height for office buildings (0.17m according to Neufert et al., 2002) and a small tolerance. A detected excess of this threshold causes the step to be marked as stair candidate. In order to robustify the process, a minimum number of consecutive stairs can be required. Each stair candidate accepted as stair produces the addition or subtraction of 0.17m to the height value before the step.

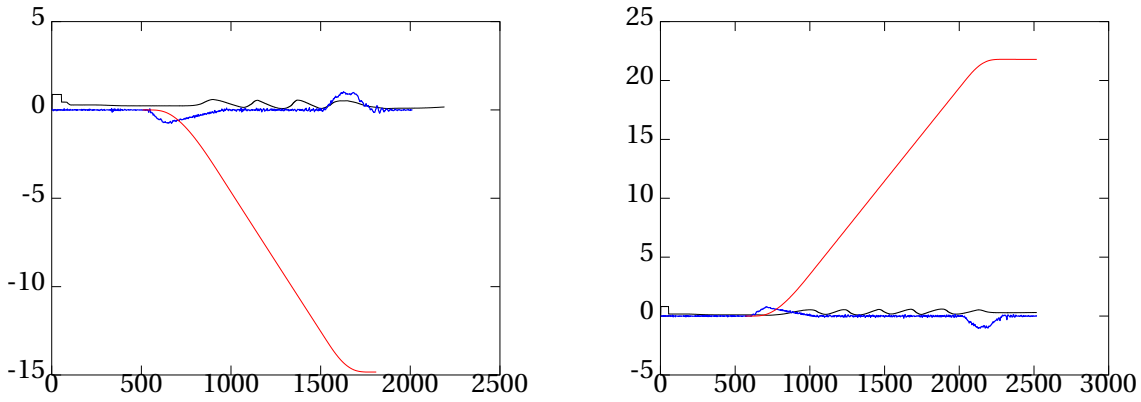
The described approach could be combined with available stair information in the model in two ways. Similar to the method described in Walder and Bernoulli (2010), the detected information “user is on stairs” could be employed to correct the 2D position towards a the position of a modelled staircase. On the other hand, stair candidates detected outside of modelled staircases may be discarded.

In order to detect elevators, on the other hand, the first step is the detection of a stationary user[4]. As also described in Alzantot and Youssef (2012b), the movement of an elevator shows a distinct pattern in the accelerometer data. This pattern consists of a weight-loss, a zero-acceleration and a weight-gain phase, with the phase's order depending on the direction of travel (see figure 5.6, blue line shows the z axis accelerations cleaned from the gravity part). The weight-loss and weight-gain phases can be detected as differences from zero. As a standing user is assumed before the first and after the second phase, the values in between can be drift corrected analogously to the correction described in section 5.1.

Additionally, the number of travelled floors can be detected using an available magnetometer. Variations in the magnetometer measurements caused by nearby metallic structures or strong electro-magnetic fields render them unusable for the trace's correct alignment. These variations, on the other hand, may be employed to distinguish between the elevator travelling inside a floor or past the structures between floors by detecting the local maxima in the data during the movement phase identified by use of the accelerometer measurements (see figure 5.6, black lines).

The remaining small drift as well as the fact that the sensor is not mounted on the foot in perfect alignment with the ground plane cause the result of the accelerometer-based method to not be very (metrically) accurate. However, in combination with the magnetometer-based approach, the number of travelled floors can be detected with high confidence.

[4] due to the great acceleration differences between walking and elevator movement, the detection of a user walking around in the elevator is virtually impossible



(a) (b)

Figure 5.6: Two exemplary elevator data sets: a) travelling from 4th floor to ground floor, b) travelling from ground floor to 6th floor (blue: accelerations along z axis minus mean (in $\frac{m}{s^s}$), black: norm of the magnetometer vector (in mGauss), red: drift-corrected, doubly integrated accelerations, i.e. z axis displacement (in m))

5.2.4 Map-matching

As depicted in figure 5.8 on page 89, the aligned trace fits reasonably well to the coarse model. However, it also becomes obvious that errors still remain in the data - i.e. the yellow trace crossing the walls in the upper part of the image - which necessitate further correction steps. These errors result from inevitable imperfections in the approach for the derivation of the initial position as well as from errors in the coarse model. A prime illustration, the *you are here* symbol, as the most important information contained in the plan, is often emphasized by its bigger size to facilitate its recognition by users of the plan. This results in errors in the location of the surrounding walls caused either by the use of a displacement operation during the generalisation or by more occlusions and adds a higher uncertainty of the reconstructed walls. In addition to errors stemming from this source, the distance between the camera and the plan is not recoverable from the image[5].

On a limited scale, these errors can be resolved by map-matching using the available coarse model. While map-matching methods in edge-based models - i.e. street networks - build on the comparison of the local curvature of the trajectory's and the map's edges (Czommer, 2001), the model at hand is face-based and this method is not applicable. An analogous approach using the edges representing the walls which are close to the trace for the comparison, or the centre line of the corridor was discarded. Building on the assumption of movement being parallel to the walls, it was too constrained to reflect the free movement of pedestrians. Map-matching approaches building on particle filters like the one found in Woodman and Harle (2008) are not applicable neither, as their main assumption is the "death" of particles which travel through walls. Thus, they require a complete model containing door openings.

Instead, a method similar to the one presented by Walder and Bernoulli (2010) was developed, which constrains the movement only regarding possible conflicts with the underlying environment model. Like the aforementioned particle filter methods, Walder and Bernoulli (2010)'s approach bases on a) a complete model of the building's interior structure, most importantly containing door openings, and b) the fact that pedestrians never walk through walls. Furthermore, the assumption is made c) that pedestrians cannot

[5]unless the exact size of one or more elements in the plan is known and these elements are detectable

pass doors in arbitrary angles. Expanding on these ideas, Walder and Bernoulli (2010) present heuristics for the correction of traces crossing walls, position correction for traces touching doors as well as repositioning if a wall was crossed in the neighbourhood of a door in an angle that allows passing doors.

However, in the case of the coarse models reconstructed from evacuation plans, the presence of door openings cannot be guaranteed. As a result, the constraint "pedestrians never walk through walls" does not hold true for the resulting models and has to be replaced by "pedestrians cannot walk through walls in arbitrary angles and at arbitrary positions". Thus, two constraints are necessary: an angle constraint and a constraint for valid door positions.

As depicted in figure 5.7a), the angle constraint depends on the door width as well as the user's shoulder span. In the experiments, a minimum value of 0.5 metres for the shoulder span (plus a small buffer)[6] was used together with a standard door width of 0.8 metres, resulting in an angle threshold of 40 respective 140 degrees. In reality, however, these extreme angles may only be reached at rare occasions, as their estimation bases on zero-width walls and does not take into account the door leaf which might further downsize the opening. Therefore, this threshold proved itself reasonable in the experiments carried out.

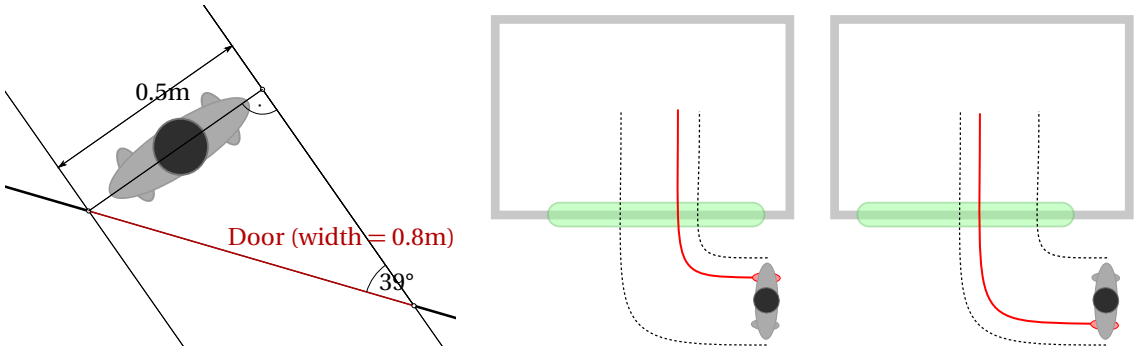


Figure 5.7: a) Angular threshold for map-matching and the automatic reconstruction of doors, b) valid door positions (green) for a sensor mounted to the right/left foot

As the second (valid door) constraint, the distance of the track's and wall's intersection point to nearby wall junctions is used. Depending on the sensor foot, this distance is compared asymmetrically against the standard door width plus a buffer of 0.2 metres on either side in order to determine if there exists sufficient space available for a door to be reconstructed. This concept is illustrated in figure 5.7b) showing that one of the method's input parameters is the foot the sensor is mounted to.

The actual correction is carried out in the following way:

1. Compute the vector between two consecutive trace positions.

2. Find the room in which the respectively first trace position is located using a point-in-polygon test.

3. Check for an intersection between the step vector and this room's walls.

[6] motivated by the average height of a male human in Germany (circa 1.8 metres according to the Federal Statistical Office of Germany, see `https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Gesundheit/GesundheitszustandRelevantesVerhalten/Tabellen/Koerpermasse.html`, last visited 2nd February 2016) and the height to shoulder span ratio of 4:1 from da Vinci's Vitruvian Man

4. If an intersection exists, compute the angle under which the intersection occurs, check against the angle threshold:

   a) If the angle threshold is fulfilled, but the intersection does not occur at a valid door position: move the full trace away from the wall junction until the intersection is at a valid door position,

   b) if the angle threshold is not fulfilled: move the full trace along the wall's normal direction until the step vector does not intersect with the wall any more.

In addition to the raw and the aligned trace, figure 5.8 shows the trace resulting from map-matching (green). It can clearly be seen that, after correction, the trace hits the wall only once at the end where the user obviously entered the room in a valid angle.

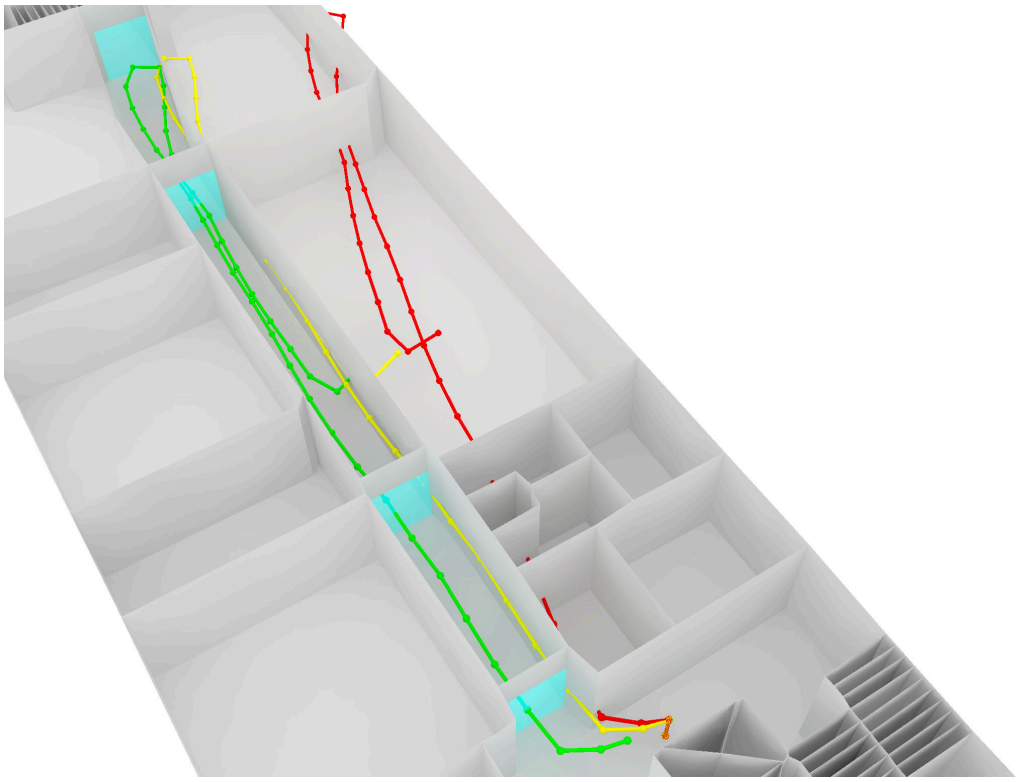


Figure 5.8: Raw (red), aligned (yellow) and map-matched (green) traces

5.3 Semantic model improvement

So far, the only semantic information present in the coarse model consists of possibly detected doors as well as staircases. As shown before, such a coarse model may serve as information supporting infrastructure-independent positioning methods, however, in order to be of use for indoor navigation and other indoor LBS at the minimum the door positions and semantic information like room numbers etc. are needed. In the following, approaches employing the user traces for the semantic annotation will be presented.

5.3.1 User interaction

As mentioned before, the modelling approach presented in this thesis has borrowed ideas from the mapping concepts behind OpenStreetMap. The OpenStreetMap wiki[7] gives an overview over ideas for mapping techniques OpenStreetMappers can use. Using a GPS logger as the geo-reference, mappers are advised to either write down notes with the ID of the current waypoint, geo-tag images that are used to record e.g. street signs or other map features or use voice/video recordings to collect data. Subsequently, this data is used in a manual map production step to add the features to the map.

In an analogous way, the user traces resulting from the foot-mounted MEMS IMU positioning system can be used to geo-reference features needed for the improvement of the coarse indoor model. In indoor environments, relevant information is found as the information associated with rooms like people, room numbers, public or private status, opening hours or the room's usage as a library, secretariat or restroom. Examples for a manually added semantic point feature ("coffee maker") and semantic information added to a room (selected room, "restroom, men") can be seen in figure 5.9b) on page 91.

5.3.2 OCR analysis of photographed door plates

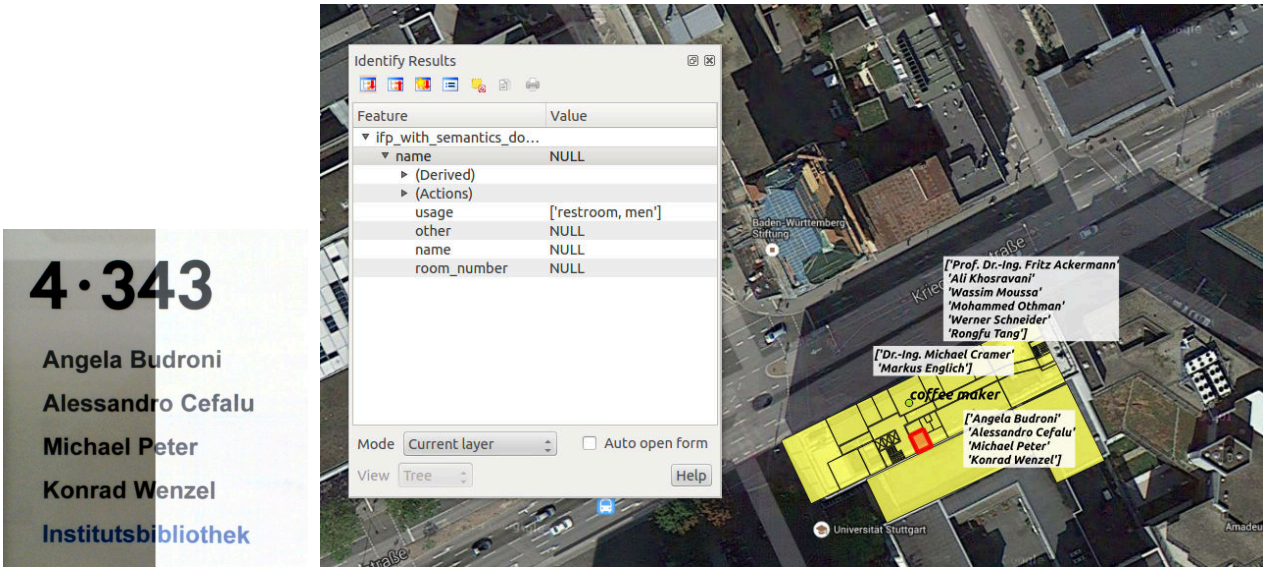
In addition to this manual model enrichment method enabling the operator of the data collection system to add virtually any attribute to the coarse model, interior environments contain information sources which are exploitable in an automated way. Door plates, as found on the surface of or next to most doors in public office buildings (Peter et al., 2013b), are one example for such information sources. While door plates may contain any information about the room behind the door, most commonly they contain room numbers as well as people assigned to the room and/or the room's use. Not to forget, they carry the approximate location of a door which can be helpful if the door is locked and thus its location may not be defined by use of the trace (see section 5.4.2). Additionally, it is not uncommon that the room number carries the floor number (see figure 5.3.2a). The automatic door plate analysis consists of the following steps:

1. The image pre-processing steps used during the evacuation plan analysis (section 4.4) are put to use in order to optimize the door plate image (see figure 5.3.2 a). The textual contents of the optimized image are converted to strings by applying Optical Character Recognition, i.e. the Tesseract engine (Smith, 2007).

2. The resulting string's content can generally be categorized as belonging to one of the following three categories: a) numbers, which mostly represent the room number or opening hours, b) names, i.e. the people assigned to the room and c) other text, most commonly some information about the use of the room. In addition, non-sense text stemming from OCR errors - e.g. "OCR'ed noise" - might be part of the result.

    a) Among this information, the routing in navigation scenarios profits most from the detection of the room number, which happens to be the most distinct information on the door plate (if present). Its detection is simplified to finding valid digits in the string provided by the OCR engine. In office buildings, the room number often is compiled using a pattern like "floor number

[7] `http://wiki.openstreetmap.org/wiki/Mapping_techniques` (last visited 2nd February 2016)



(a) (b)

Figure 5.9: a) Photographed door plate before and after correction, b) manually added information and information extracted using OCR (here: names) assigned to the model, overlaid on Google Maps

- delimiter - room number", which can be taken into account by identifying common delimiters like ".", "," or "-".

b) For the classification of the remaining text to the other categories, i.e. names and other information, a names database and a spell checker are used. With the help of the names database, lines containing valid names can be identified, while the spell checker enables the filtering of nonsense text. As the completeness of the names database cannot be guaranteed, lines containing at least one name found in the database will be categorized as a name string in total.

3. As the image is acquired outside of the room to which the attributes deduced should be assigned, the most likely room is identified as the one which is hit first by the viewing direction vector. This vector is derived from the current position and the one before, assuming that the user turned towards the doorplate in order to photograph it. Figure 5.3.2 shows the photographed door plate before and after correction as well as the name information extracted from three door plates assigned to the model.

5.4 Geometric model refinement

The need for a geometric refinement of the coarse model resulting from the evacuation plan analysis arises for various reasons. The two most significant are: 1) features like door openings or small rooms may have been omitted in order to improve the legibility of the plan, and 2) rearrangements of parts of the interior structure are not uncommon, however, they may not be represented in the evacuation plan. The following sections show how these flaws can be overcome either by user interaction, conjoint analysis of model and pedestrian trace or a combination with a low-cost range camera.

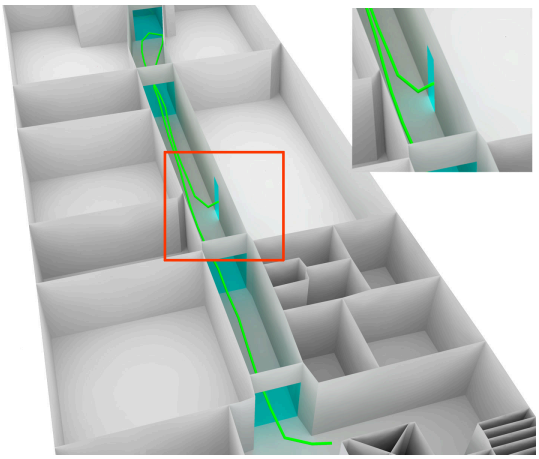


Figure 5.10: Exemplary result of the automatic door reconstruction using the map-matched track (green) as presented in Peter et al. (2013a)

5.4.1 User interaction

Similar to the semantic annotation of the existing coarse model by user interaction as described before, the user's current position can be employed to update the model's geometry.

The user can manually place geometric model templates like tables, chairs and other furniture at his current position. Alternatively, manual geometric updates of the coarse model can be performed. For example, the user can walk up to a wall, take manual measurements and update the coarse model in place by adding a window or door at the measured height and with the acquired dimensions.

5.4.2 Automatic door reconstruction

The availability of the coarse model and the user trace furthermore allows for the automatic reconstruction of door openings (Peter et al., 2013b). This method bases on the same assumptions as used in map-matching (see section 5.2.4), i.e. constraints for valid door locations and for the human movement when passing through a door. During map-matching, the violation of these constraints causes the trace to move. Here, their fulfillment causes the reconstruction of a door at the intersection point of trace and model edge/face.

If the threshold checks are passed at an intersection point between track and model edge, the model is updated by reconstructing a door with standard dimensions (0.8m width, 2.2m height) asymmetrically around the intersection point (depending on the foot which carries the sensor). An exemplary result can be seen in figure 5.10. The newly constructed door is annotated as being reconstructed using the track and the intersection point is stored. These attributes enable a later update of the door's position if additional hits occur within the door width.

5.4.3 Model refinement using a low-cost range camera

The aforementioned approach, while usable for the determination of an approximate door position, is only able to reconstruct doors with the dimensions provided to it, instead of the actually existing geometry. Addi-

tionally, further geometry missing in the coarse model due to map generalization reasons or modifications after the creation of the plan should be acquired in order to provide the most detailed model possible.

As detailed in chapter 3, the Microsoft Kinect became the preferred sensor system for the 3D acquisition of building interiors soon after its appearance. Most related approaches apply this low-cost range camera in the context of SLAM systems, either registering consecutive point clouds using the Iterative Closest Point (ICP) algorithm or exploiting the sensor's simultaneous supply of RGB and depth data in a combination of ICP and image feature matching. While these approaches produce impressive results in well-textured and/or geometry-rich scenarios, there are some characteristics of indoor scenes which may cause problems:

1. The combination of poor (artificial) lighting conditions and scarce texture (single-color walls) deliver an insufficient number of image features.

2. Geometric matching methods like ICP are doomed to failure in areas of uniform geometry (as found along hallways).

3. Doors offer very limited overlap between consecutive data sets, very likely causing both registration methods to fail.

While the first two characteristics are likely to be comprised in hallways, the third one hinders the detailed acquisition of multiple rooms or even complete floors, therefore leaving the data and resulting models incomplete or requiring manual work. In these cases, the external information provided by both the coarse model, and the positioning system may support the acquisition using the Microsoft Kinect as well as the reconstruction from the resulting point clouds.

For the experiments described in Khosravani et al. (2013) the data was acquired by simultaneous use of a hand-held Microsoft Kinect and the foot-mounted positioning system, recording a set of Kinect RGB and depth frames at each occurence of a step detected by the positioning system.

The alignment of corridor point clouds and the reconstruction of the corridor walls are carried out in three processing steps:

1. An approximate viewing direction is computed from the movement vector of the step which occurred after the point cloud acquisition. This viewing direction and the step's coordinates deliver the parameters for a first alignment.

2. The hand-held Kinect is not necessarily held horizontally as well as its height is unknown. Therefore, the aligned point cloud is segmented in order to identify possible ground points, i.e. points whose normal vectors - computed using their nearest neighbours - include an angle of 30 degrees or less with the vertical axis. The angle between the mean normal vector and the vertical axis is used to correct the sensor tilt, while the detected ground level is aligned with the coarse model's ground plane, delivering the Kinect's height over ground.

3. Still remaining misalignment errors can be overcome by aligning the point cloud's wall planes to nearby walls in the coarse model. To this end, following a similar approach as presented by Okorn et al. (2010), the floor and ceiling points are removed and the remaining point cloud is projected to the xy plane, resulting in an image which can be binarised using a threshold. The binary image is

treated by a morphological closing operator with a structure element sized according to the point cloud's expected precision. The thus merged regions are skeletonized and dominant lines are extracted using the Hough transformation, resulting in a 2D model of the corridor walls. By matching walls to their closest counterpart in the coarse model they can be aligned by a rotation around the z axis. Finally, the 2D walls can be extruded to 3D using the distance between the previously removed floor and ceiling points (see corridor in figure 5.11b)).

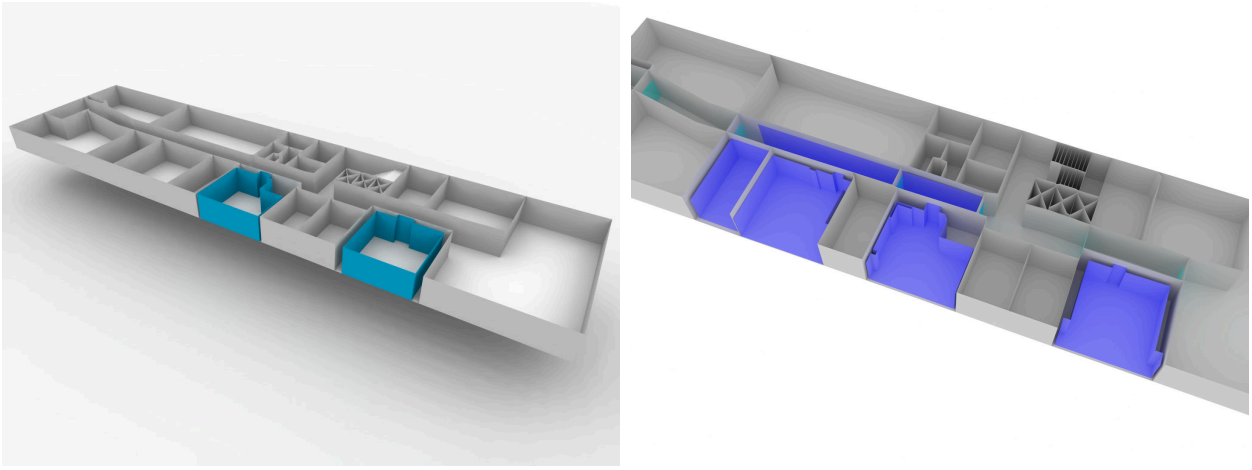
Building on this work, (Peter et al., 2013b) tackles the problems arising from the third characteristic described above in order to enable geometric updates to the coarse model's rooms. Here, the positioning system is put to use in the detection of the room in which the user starts the point cloud acquisition. The orientations of the point clouds are computed using Structure-from-Motion (SfM) based on the Kinect's RGB images, resulting in a combined point cloud covering the complete room.

In order to derive polygonal models from the combined point cloud, it may be evaluated by Budroni (2013)'s Manhattan world plane-sweeping approach (described in section 3.3.2), delivering more detailed replacements for the coarse model's rooms (see figure 5.11a)). In addition to using Budroni (2013)'s method, a less constrained reconstruction method for rooms (developed in joint work by Ali Mohammed Khosravani) was presented in Peter et al. (2013b). It is similar to the ones described by Leiva et al. (2001) and Okorn et al., 2010 (see also chapter 3) and follows the same principle as the hallway alignment and reconstruction described above:

1. The room's point cloud is cleaned by filtering out points below 1.5 meters, assuming they represent furniture.

2. Subsequently, it is converted to a 2D grey scale image by projecting the remaining points to the xy plane, counting the number of points per grid cell. Thresholding the resulting grey scale image produces a binary image which will be morphologically closed and skeletonized. Treating the skeleton image with the Hough transformation delivers a set of lines which are clustered hierarchically, first by orientation and subsequently by their distance to the point cloud's centroid.

3. The final model is built by enforcing rectangularity between edges which enclose an angle close to 90 degrees and removing dangling end point edges. An approximate alignment of the model is known from the last step provided by the positioning system before starting the data collection in the room. Thus, the updated model's alignment to the coarse model bases on the same principle as the final corridor alignment described before. It is carried out using close and similarly oriented edges in both models, followed by estimating a rotation angle around the z axis. Examples for refined room geometries can be seen in figure 5.9b).



(a) (b)

Figure 5.11: Model refined using low-cost range camera and a) Manhattan constraints plane sweeping (Budroni, 2013), b) method presented in Peter et al. (2013b)

5.5 Trace-based reconstruction supported by an indoor grammar

As stated in chapter 3, many reconstruction methods for man-made structures and especially building interiors include knowledge about the object to be reconstructed in the form of constraints like the Manhattan world assumption, i.e. parallelism and rectangularity. In contrast to these low-level general constraints, high-level object knowledge as e.g. the repetition of structures may be modeled using formal grammars. The concept of formal grammars was introduced to the field of geoinformatics by Parish and Müller (2001) and Müller et al. (2007), who documented their applicability to the procedural modeling of street networks using Lindenmayer-(L)-systems and the improvement of building façade reconstruction from images, respectively. Furthermore, Becker (2011) shows the grammar-based reconstruction and completion of façade structures from laser scanning point clouds.

However, the interest in the concept's adoption for indoor modeling has been limited to a small number of publications like Gröger and Plümer (2010) whose approach merely constructs synthetic models from known grammars and thus is not applicable for the reconstruction of building interiors from noisy data. In a similar manner, Hahn (2006), Marson and Musse (2010) as well as Mirahmadi and Shami (2012) report on the procedural generation of residential house layouts using tree maps. Yue et al. (2008) also describe the placement of rooms inside a known external building shell based on decision trees and shape rules. The completely procedural modeling of buildings as well as their interiors could be applied in use cases such as films and games as well as disaster and other simulations. However, as shown by Müller et al. (2007), Becker (2011) and Khoshelham and Díaz-Vilariño (2014) (see also section 3.3.2), grammars may also help to improve the robustness of data-driven reconstruction approaches by deducing higher-level constraints from the data and applying them.

As presented in section 3.4, indoor mapping using pedestrian traces collected by people walking around in the building is feasible. While the systems presented in section 3.4 yield impressive results considering the sensors used and their expected accuracies, they are mostly constrained to the mapping of the hallway structure found at a certain floor. This is due to the fact that hallways are much more frequented by pedestrians in the building than rooms and the statistical nature of the presented approaches, which need big amounts of traces in order to balance the sensors' unfavourable error characteristics. In the systems

which include room-level mapping, rooms are reconstructed as either simple (Jiang et al., 2013) or arbitrary shapes (Alzantot and Youssef, 2012b; Gao et al., 2014), not respecting architectural principles like rectangularity, parallelism or higher-level structures. Furthermore, only the floor plan's geometry is captured, leaving semantic information like room numbers as an open task, which prevents the resulting maps from being of use e.g. in pedestrian navigation.

In order to combine knowledge modelled in grammars with the aim of improving pedestrian trace-based indoor mapping, ComNSense[8], a joint project between the Institute for Photogrammetry and the Institute of Parallel and Distributed Systems at the University of Stuttgart, was founded in the year 2013. The greater goal of the project is to model the reconstruction of the indoor space as a closed loop, consisting of the generation of model hypotheses using the knowledge stored in the grammar, their check against newly acquired data as well as the update of the grammar from this data.

5.5.1 Grammar definition

For an indoor reconstruction approach using distributed trace acquisition to be applicable, the indoor space to be reconstructed has to be frequented by a large number of people. Thus, such a reconstruction approach aims at public buildings, whose floors are mostly assembled of a skeleton of straight corridors and non-corridor spaces containing the rooms. Analogous to Parish and Müller (2001)'s approach used for the description of street networks, the characteristics of the corridor skeleton can be described by a L-system. The subdivided nature of the non-corridor spaces into room elements, on the other hand, motivates the modelling of this part of the structures as a split grammar (Peter et al., 2013a; Becker et al., 2013).

Due to the similarity between street networks and indoor corridor structures and its inherent adaptability to observation data, the L-system described by Parish and Müller (2001) serves as base for the corridor L-system. Parish and Müller (2001)'s L-system decouples the generative part of the grammar from the actual functions for the setting of attributes, probabilities and constraints by moving them to external functions. The two external functions for the indoor corridor network are *LayoutSetting* which defines rules for the growth of the network based on probabilities, and *ConsistencyConstraints* which models self-sensitivity as well as sensitivity to the environment. Self-sensitivity ensures the avoidance of the occurrence of short dead-ends as well as the extension of slightly too short corridor segments. Sensitivity to the environment, on the other hand, prevents conflicts between the growing network and the existing external building shell.

The split grammar is designed under the premise that typical floors of public buildings contain hallways from which a linear sequence of rooms is entered, ensuring convenient access to rooms from the hallway as well to windows in the building's external shell. The grammar consists of an axiom, a set of non-terminals, a set of terminals and the production rules. The only non-terminal needed is *Space*, i.e. a 2D polygon[9], which is further sub-divisible. The set of terminals consists of the single element *space*, a polygon/solid which will not be further sub-divided. The production rules are modelled as replacement rules which either replace a *Space* by new *Spaces* by use of splits, multiple *Spaces* by one *Space* using a merger, or a *Space* by a *space* using instantiation (see figure 5.12). The split rules needed to describe arbitrary room layouts are

[8] `http://www.comnsense.de` (last visited 2nd February 2016)

[9] in general: a 3D solid, as the split grammar is also to be used for 3D reconstruction as well as the subdivision of the 3D building shell into floors

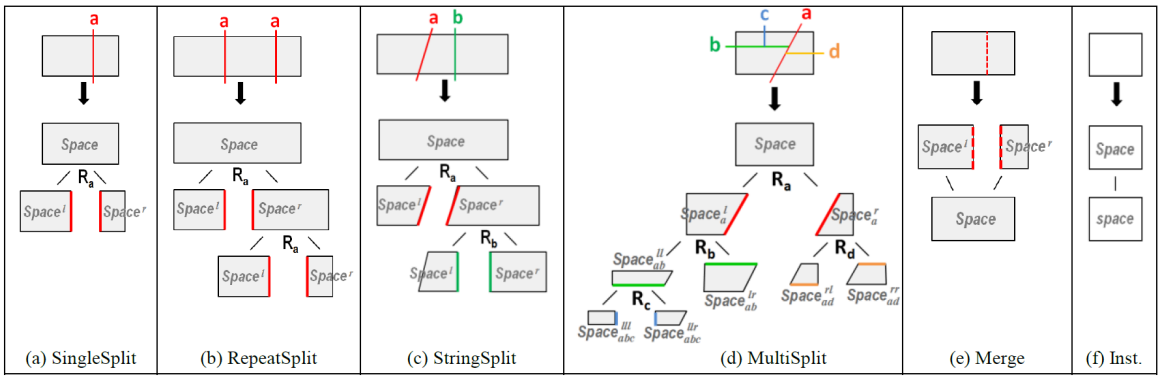


Figure 5.12: Overview over the rules of the split grammar used to describe room layouts as presented in Becker et al. (2013)

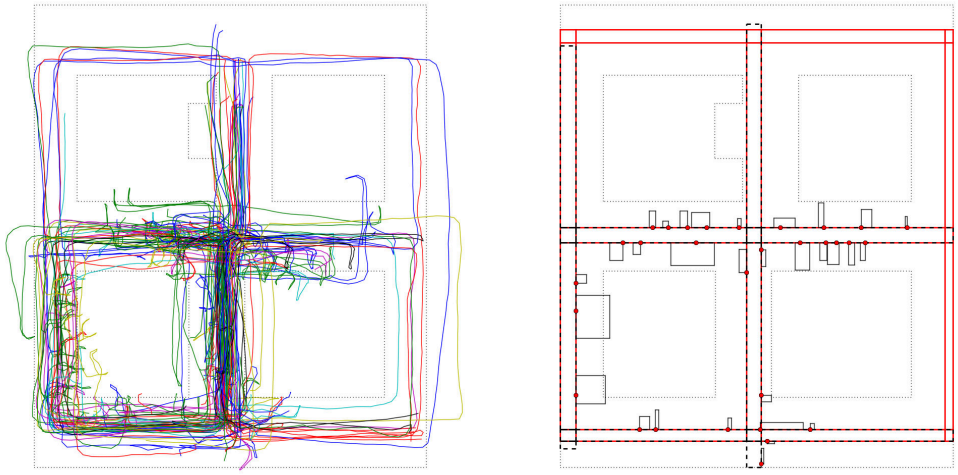
SingleSplit, *RepeatSplit*, *StringSplit*, and *MultiSplit*. The *SingleSplit* divides the *Space* into two *Spaces* along a partition plane which is defined by an orientation and a distance parameter. The other split rules can be used to group different room units: While the *RepeatSplit* generates a sequence of rooms with the same parameters, the *StringSplit* enables the reconstruction of rooms with different parameters. The *MultiSplit*, on the other hand, is used to represent non-linear layouts by the hierarchical combination of a suite of splits. The probability of a rule's occurrence is composed of the *a priori probability* which describes the rule's relative frequency of occurrence, and the *context aware probability* which is the conditional probability of neighbouring relationships.

5.5.2 Trace analysis

The distributed trace acquisition is implemented using the method described in section 5.1. In order to allow for increased mobility, the IMUs are connected to miniature PCs, type Raspberry Pi, which process the raw data to step vectors and send them to smartphones. The smartphones run an interface used to set the initial values by manual interaction. The pedestrian traces resulting from the combination of initial values and step vectors are uploaded to a server where they are stored in a database and can be accessed by the mapping application.

In contrast to the real-time alignment described in section 5.2.2, the mapping application developed in the context of a Diploma Thesis at the Institute for Parallel and Distributed Systems (Dibak, 2013) does the alignment in post-processing, before reconstructing the corridors and rooms from the traces. To this end, each trace is segmented into straight parts by testing consecutive step vectors against an angle threshold. Subsequently, the segments are categorized as corridors based on a length threshold and the correction is carried out by rotation (as shown in section 5.2.2). By analysis of a subset of the aligned corridor traces a base plan can be reconstructed as the axis-aligned bounding boxes of the respective segments. This subset is compiled from the full set of traces based on the rotation angle which was necessary during alignment, assuming that a smaller rotation angle indicates a higher-quality trace. Subsequently, the other traces can be further corrected towards the centres of the base plan's corridor rectangles and iteratively added in order to refine the base plan. For more details refer to Dibak (2013).

Once the corridor skeleton is reconstructed, all other step vectors can be analysed in order to find rooms. To this end, segments categorized as corridor, curves and too short segments are removed from the traces.



(a) (b)

Figure 5.13: a) Aligned traces and b) corridor skeleton and rooms (black) derived from the traces before (dotted black) and after (red) completion using L-system (as presented in Becker et al., 2013)

The remaining parts represent the rooms which are reconstructed as the axis-aligned minimum bounding rectangles.

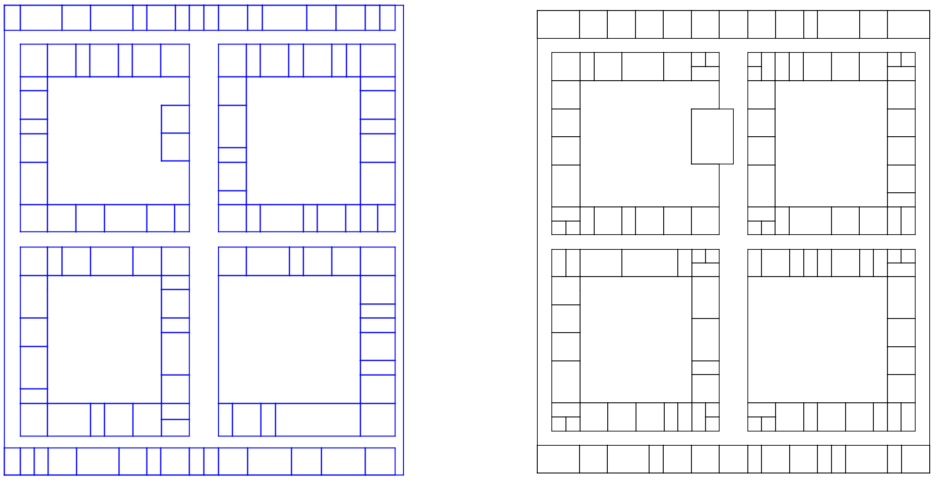
In order to evaluate this approach, a set of 227 traces covering a combined length of more than 23 kilometres was acquired by four individuals, following pre-defined scenarios like *visitor looking for a room, working in an office environment, taking part in a meeting* and *room cleaning/security personnel tour*. The aligned traces are depicted in figure 5.13a) together with the external shell. Figure 5.13b) shows the computed corridor skeleton (dotted black). In the test bed only 18 rooms were accessible, of which 13 could be found by the trace-based reconstruction method (see figure 5.13b), black rectangles).

5.5.3 Grammar-based completion and reconstruction

For the experiments described in Philipp et al. (2014) and Becker et al. (2013), both parts of the grammar were derived automatically from ground truth data of the first floor of the Computer Science building at University of Stuttgart, thus differing from the one used for the trace-based data acquisition (second floor). As the interior architecture of this floor follows the same style, the probabilities of the rules are slightly different, but similar. In future work, the detected corridors and rooms will serve for the automatic derivation and update of both the L-system and the rule parameters of the split grammar.

The first step in the grammar-based reconstruction is the completion of the possibly incomplete corridor skeleton by means of the corridor grammar. The axiom used for the L-system is found as the corridor segments derived by the trace-based reconstruction approach and the rules are processed sequentially.

The split grammar's axiom, on the other hand, consists of the empty non-corridor spaces which result from the combination of the corridor system with the existing model of the buidling's external shell, i.e. the polygon set difference of external shell and corridor system. Following the set difference rooms which are outside of the external shell are removed and a geometric generalization dragging the walls of corridor segments to nearby external walls is applied (visible when comparing the centre corridor area in figure 5.13b)



(a) (b)

Figure 5.14: a) Result of the applicaton of the split grammar from the first floor and b) ground truth model of the second floor (as presented in Becker et al., 2013)

and figure 5.14a)). The areas of the remaining rooms serve as constraints in the subsequent grammar application (Philipp et al., 2014). It is carried out as a constraint-augmented random walk using the room rectangles in constraints of the types “free” (“don’t place a wall inside an observed room”) and “wall” (“place a wall between observed rooms”).

Figure 5.14a) depicts the result of the application of the first floor grammar derived to the trace-based corridors and rooms acquired in the second floor. In comparison to the ground truth shown in figure 5.14b), it becomes apparent that the room layout style is reproduced very well, especially when considering the small number of accessible and therefore detectable rooms. The evaluation of the reconstruction’s quality is described in section 6.3.

6 Results and discussion

In the following sections, the results of all methods described before are evaluated. In section 6.1, the quality of the reconstruction from photographed evacuation plans is assessed. Section 6.2 presents the precision and accuracy evaluation of the positioning method and of the model refinements derived with its help. Section 6.3 is dedicated to the quality of the grammar-supported reconstruction solely from pedestrian traces.

6.1 Evaluation of reverse engineered plans

The main part of the accuracy evaluation of the reverse engineered evacuation plan is its comparison to existing ground truth data, which can be found in section 6.1.3. However, being an image-based method, some imaging parameters influence the accuracy which can be assessed as described in section 6.1.1. Furthermore, as the final model's scale is derived by matching it to an external model, the achievable accuracy is related to the accuracy of this model. Thus, the accuracy of the external model is described briefly in section 6.1.2.

6.1.1 Resolution

In order to assess the lower bound for the achievable precision of the elements modelled from the photographed evacuation plans, the data's resolution after geo-referencing using the method described in section 4.5 shall be examined. The transformation parameters available after matching the indoor model to the external building shell allow the computation of the scale factor, i.e. the number of pixels representing one meter in the image or the scaled size of one pixel. Of course, this scale factor depends on the precision and scale of the original plan, the camera's resolution and focal length, the distance between camera and plan as well as the size of the plan in the image. The latter two[1] factors will be adjusted by the user making sure that the plan is filling the image as complete as possible. Some examples can be found in table 6.1, showing that on average the sample distance is around three centimetres.

Elements smaller than the scaled size of a pixel cannot be reconstructed. Most plans will not contain such small features due to legibility improvements using generalisation during the production of the evacuation plan. However, this also results in features like stairs being reproduced in the image by areas only three pixels wide (rendering error-free reconstruction very difficult). This is the case for the Canon S90's 4.7 centimetres resolution and 15 centimetres wide stairs.

[1] or three, in the case of a zoom lens

Data set	*Camera and focal length (35mm equiv.)*	*Scaled size of a pixel*
Institute for Photogrammetry	Canon S90, 28mm	4.7cm
Institute for Photogrammetry	Nokia Lumia 925, 26mm	2.9cm
Palacio Vila Flor Guimaraes	Nikon D3100, 52mm	2.1cm
Institute of Parallel and Distributed Systems	Nikon D7000, 36mm	3.0cm
Institute of Navigation	Nokia N9, 28mm	2.9cm

Table 6.1: Scaled pixel sizes for exemplary data sets (georeferenced using OpenStreetMap)

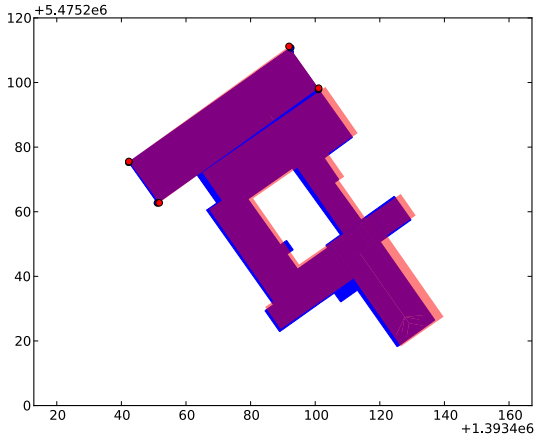


Figure 6.1: Comparison of the OpenStreetMap (red, transparent) and city model (blue) external shells

6.1.2 Accuracy of the geo-reference

For the majority of indoor use cases, the absolute accuracy of the model's coordinates in a world coordinate system is of subordinate importance. Thus, the precision of the external shell used for the geo-reference is much more important, as it limits the achievable precision of the derived indoor model via the scale factor. Being surveyed by tachymetry with centimetre accuracy, city models provided by surveying offices can be assumed as reference data. In order to make a statement about the usability of the OpenStreetMap building outlines for this purpose, they are compared to the Stuttgart city model as provided by the City Surveying Office.

Fan et al. (2014) gives an evaluation of OpenStreetMap's accuracy and completeness. A considerable average shift of 4.13 meters is reported between the OpenStreetMap building outlines and the city model ground truth. Most OSM ground plans are digitized from Bing or Yahoo aerial imagery and the geo-reference of these services' imagery must be accounted for this error in addition to the limited measurement accuracy of laymen and the possibly poor visibility of the building edge's points in aerial images. As depicted in figure 6.1, the shift which can be observed between both representations of the Institute for Photogrammetry's building is not as distinct as Fan et al. (2014)'s numbers, especially in the institute's building part (i.e. the northern-most part marked by the dots). The difference in the four marked points is on average 2.6 centimetres in x direction and 28.5 centimetres in y direction, totalling to an absolute distance of 28.6 centimetres.

While comparing the shape similarity of OpenStreetMap and city model ground plans by use of the Turning Angle function, Fan et al. (2014) does not examine the accuracy of OpenStreetMap buildings in terms of

OSM (m)	*city model (m)*	*error (m)*	*error (%)*
61.054	60.840	0.214	0.350
15.857	15.347	0.510	3.320
60.783	60.448	0.335	0.554
15.879	15.245	0.634	4.154
		average	2.095

Table 6.2: Edge lengths in OpenStreetMap outline compared to city model

edge lengths[2]. Figure 6.1 shows that this accuracy is considerably high, especially when considering the base data and the manual digitization method which was used in its creation. Table 6.2 confirms this visual impression by comparison to the city model's edge lengths. As visible in this table, the absolute edge length differences are between 21.4 centimetres (long edges) and 63.4 centimetres (short edges). These errors will be propagated throughout the precision of the resulting indoor model.

6.1.3 Evaluation of the resulting model's precision

As chapter 3 implies, the data acquisition aiming at the digital reconstruction of building interiors as well as the reconstruction process itself involve a lot of effort. In combination with the ease of rearranging furniture and reshaping by installing or removing walls etc., this results in a low probability that, for a given building and floor, a precise and up-to-date ground truth can be found. Furthermore, an accuracy evaluation of the presented methods needs, on the one hand, bare wall models for the result of the evacuation plan analysis. On the other hand, a current model including furniture is needed to assess the accuracy of the refinement using the low-cost range camera. Thus, only an exemplary accuracy analysis based on the Institute for Photogrammetry's interior structure is carried out.

Figure 6.2 shows the candidates for ground truth models available at the Institute for Photogrammetry. Figure 6.2a) shows a 1:100 paper plan drawn 1981, figure 6.2b) a CAD model provided by the administration of the University of Stuttgart, drawn from a scanned paper plan, figure 6.2c) a laser scanning point cloud, produced 2008 and thus the most recent data source and figure 6.2d) the institute's floor as part of the complete building's 3D CAD/CAFM model produced in the context of a study thesis (Beck, 2000). Figure 6.2e) shows an example for the models to be evaluated. Rooms which will be referenced in the following text are marked by letters.

It becomes obvious that reshaping work took place between the creation of the paper plan and the other representations. Thus, the paper plan, while presumably being the largest scale, could only serve for a comparison of distances between fixed, unchanged walls (drawn as thick lines). The CAD model provided by the university administration was drawn from scanned paper plans, supposedly resulting in lower accuracy, and does not contain furniture. The laser scanning point cloud, on the other hand, contains furniture, but in most cases not the walls hidden behind cupboards. (Beck, 2000)'s model contains the cupboards, but also the bare walls modelled as self-contained entities and thus can serve as a ground truth for both modelling methods which have to be evaluated. However, comparing TLS point cloud and (Beck, 2000)'s model, small differences in the positions of cupboards and even doors can be noticed (see e.g rooms A and D). Thus, both the TLS point and Beck (2000)'s model will be used for the evaluation.

[2]which, due to the differing levels of detail, would be very difficult to carry out on a large amount of models

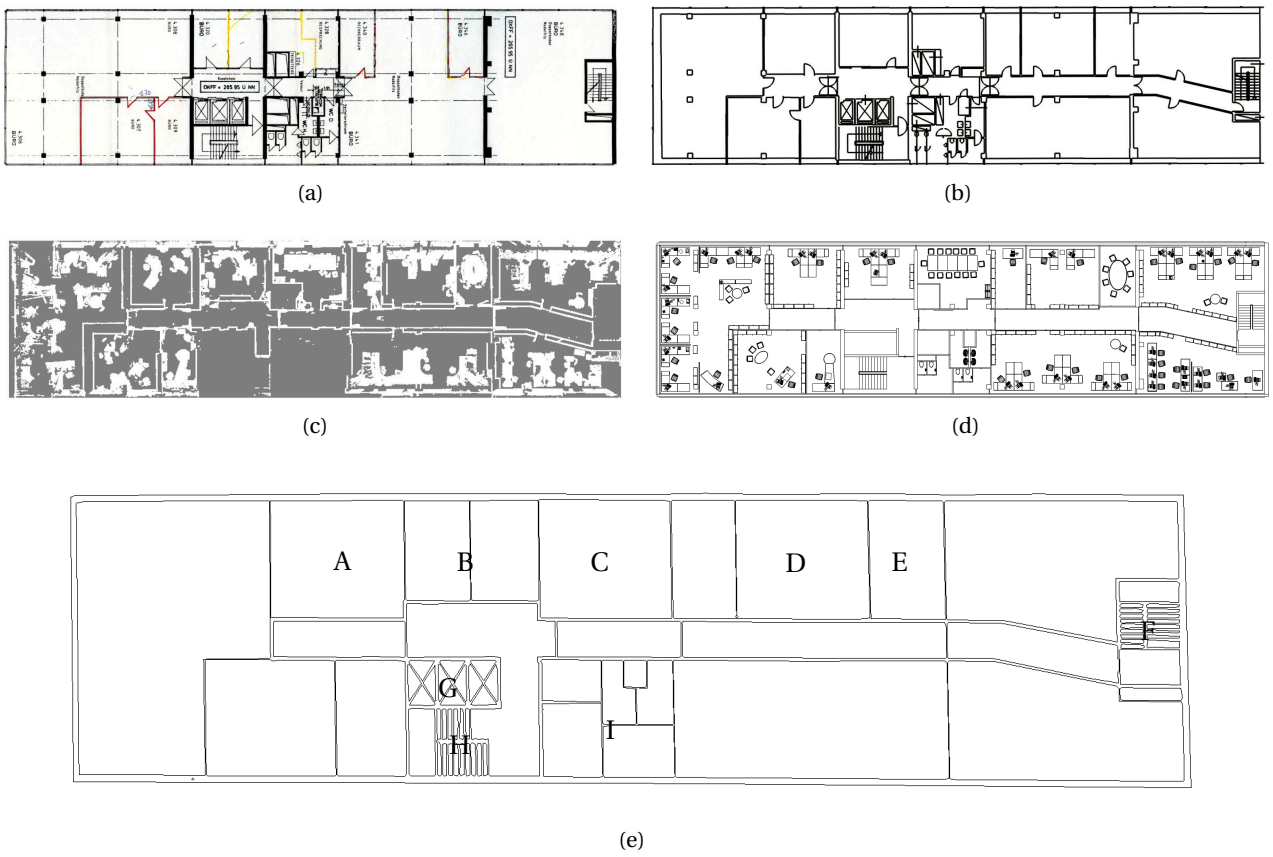


Figure 6.2: Ground truth candidates available at the Institute for Photogrammetry: a) scanned 1:100 scale paper plan, b) CAD model from the university administration, c) laser scanning point cloud, d) CAD model resulting from (Beck, 2000); e) example for a two-sided walls model to be evaluated

The evaluation of the 2D floor plan's precision is done by comparison to the chosen ground truth models. This comparison is carried out using the Open Source point cloud/mesh viewer and analysis software CloudCompare[3]. Cloud Compare allows for a comparison of two point clouds respectively one point cloud and one mesh. Therefore, the model to compare is sampled with 250 points per square meter and a one meter slice is cut out along the z axis[4]. The absolute distances of this slice's points to the reference model are computed and stored as a scalar field. The distance between two point clouds is computed by CloudCompare by searching the nearest neighbour in the reference data set and computing the euclidean distance. For the comparison between point clouds and meshes/BREP models the distance along the normal vector of the closest mesh triangle is computed.

The computed distances are used in the visualization as the points' colours. To this end, a green-yellow-red color map was chosen with a maximum distance of 5 metres (red) and a low yellow threshold, chosen as 0.5 metres. In addition to the scalar field, CloudCompare computes the mean, standard deviation and maximum value of the distances which can be found in table 6.3.

The analysis comprises comparisons between Beck (2000)'s model and models derived using the evacuation plan, using both models as reference respectively. Furthermore, the TLS point cloud was used as reference. As visible in the comparison figures, the biggest source of error results from geometry which is not available

[3] `http://www.danielgm.net/cc/` (last visited 2nd February 2016)

[4] in the case of available 3D models

Table 6.3: Comparison results between (Beck, 2000) (removed furniture and stairs) and results from the evacuation plan analysis

Reference	*Model*	*Geo-reference*	*Mean (m)*	*StdDev (m)*	*Max (m)*	*Figure*
(Beck, 2000)	one-sided walls	city model	0.146	0.217	2.663	6.3a)
one-sided walls	(Beck, 2000)	city model	0.257	0.299	3.564	6.3b)
(Beck, 2000)	one-sided walls	OSM	0.180	0.242	2.665	6.4a)
one-sided walls	(Beck, 2000)	OSM	0.328	0.299	3.453	6.4b)
(Beck, 2000)	two-sided walls	city model	0.199	0.231	2.680	6.5a)
two-sided walls	(Beck, 2000)	city model	0.229	0.341	3.801	6.5b)
(Beck, 2000)	two-sided walls	OSM	0.292	0.272	2.665	6.6a)
two-sided walls	(Beck, 2000)	OSM	0.315	0.348	3.584	6.6b)
TLS	one-sided walls	city model	0.242	0.215	1.748	6.7a)
TLS	one-sided walls	OSM	0.304	0.260	1.711	6.7b)
TLS	two-sided walls	city model	0.224	0.193	1.757	6.8a)
TLS	two-sided walls	OSM	0.252	0.202	1.753	6.8b)

in the respective other model. In figure 6.3a) this can be seen in the pillars which vary from yellow (0.5m) to red (4.1m) as well as the interior walls in the restrooms (room I in figure 6.2e) and kitchen (C), while in figure 6.3b) the removed wall in the lab (B) produces the maximum error. While this wall was physically removed and the change is not reflected in the evacuation plan, the former presumably were deleted as part of the plan's generalization improving its readability. Errors like the ones occurring around the elevators (G) and the northern staircase (F) can also be explained by generalization, however, here the geometry was altered by the *displacement* generalization operator in order to make space for the *you are here* and evacuation route symbols which are emphasized by enlargement.

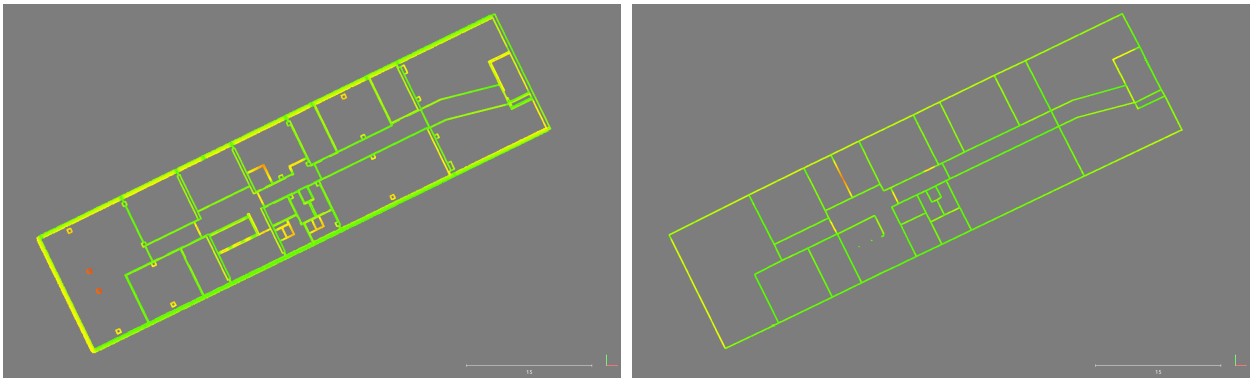
As mentioned in sub-section 6.1.2, the lower accuracy of the OpenStreetMap geo-reference affects the precision of the resulting indoor model. This fact can be seen in the numbers presented in table 6.3 as well as in the colours in the respective figures, especially at the long edges of the models, caused by the greater differences between the short edges of the external shell taken from the city model and OpenStreetMap.

Interestingly, the precision of the two-sided walls models is lower than the one found in the one-sided walls models. This allows the assumption that the wall widths as presented in the plans are not reliable, most probably caused by the plans' small scales.

The comparisons using the TLS point cloud as reference show a higher mean error, resulting mostly from the bigger amount of missing data. When looking at the colours' distribution in the respective figures, they confirm the results found in the comparisons with Beck (2000)'s model, however.

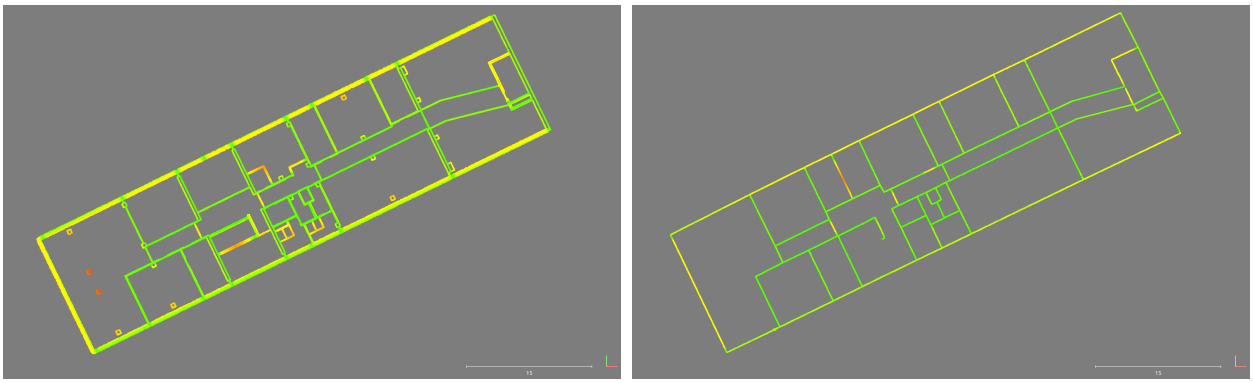
Apart from these differences, the models fit very well which is confirmed by the numbers in table 6.3. For example, using the CAD model as reference, the coarse model with one-sided walls, geo-referenced by the city model, delivers a mean error of only 15 centimetres. Taking into account the error resulting from the missing wall widths[5] as well as the maximum achievable precision (see section 6.1.1), this shows the potential of the method to deliver indoor models usable for a wide range of applications.

[5] around 12cm, measured in the terrestrial laser scan



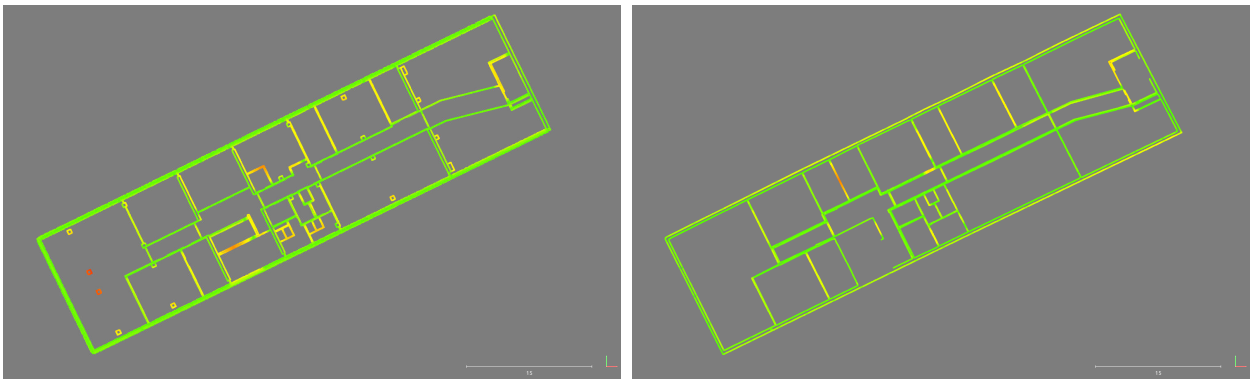
(a) (b)

Figure 6.3: a) Distances of Beck (2000)'s model to evacuation plan model (one-sided walls, geo-referenced using city model) , b) inverse analysis



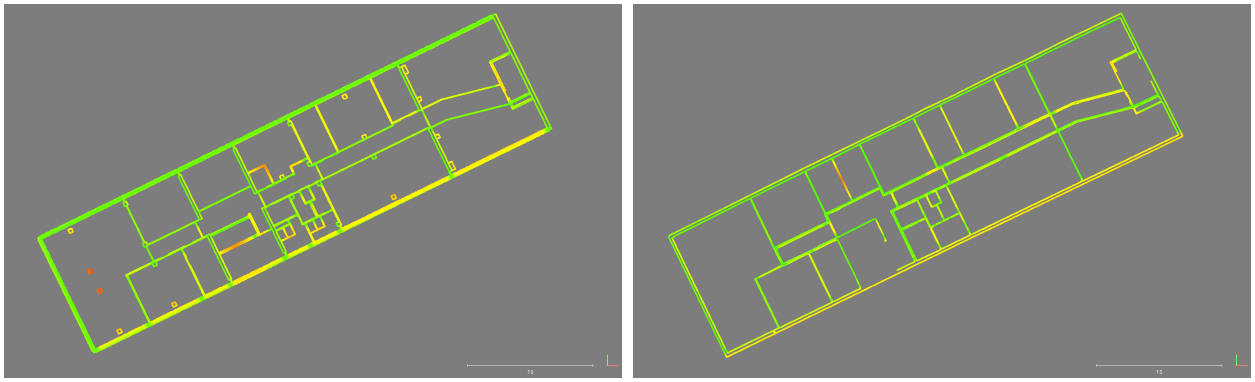
(a) (b)

Figure 6.4: a) Distances of Beck (2000)'s model to evacuation plan model (one-sided walls, geo-referenced using OSM) , b) inverse analysis



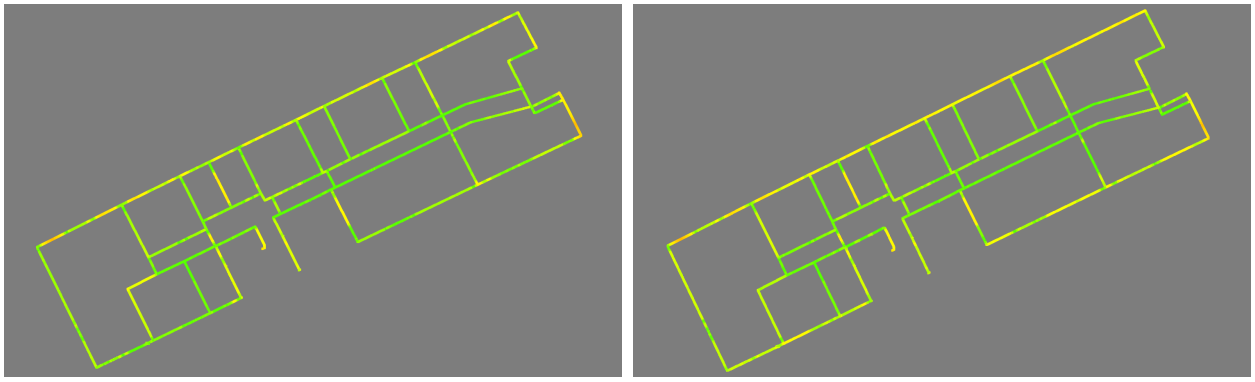
(a) (b)

Figure 6.5: a) Distances of Beck (2000)'s model to evacuation plan model (two-sided walls, geo-referenced using city model) , b) inverse analysis



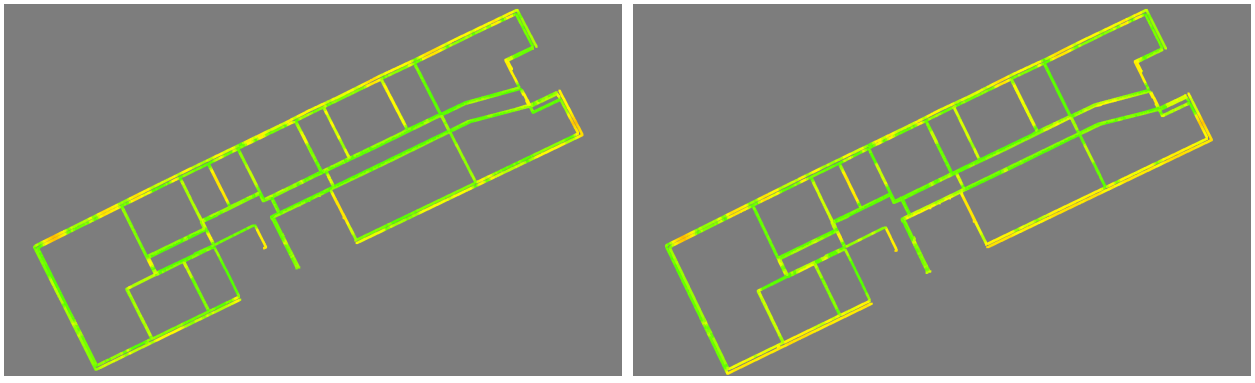
(a) (b)

Figure 6.6: a) Distances of Beck (2000)'s model to evacuation plan model (two-sided walls, geo-referenced using OSM) , b) inverse analysis



(a) (b)

Figure 6.7: Distances of the TLS point cloud to evacuation plan model: a) one-sided walls, geo-referenced using city model, b) one-sided walls, geo-referenced using OSM



(a) (b)

Figure 6.8: Distances of the TLS point cloud to evacuation plan model: a) two-sided walls, geo-referenced using city model, b) two-sided walls, geo-referenced using OSM

6.1.4 Height

The approximate floor height of the coarse model was computed as (20 stairs) x 0.17m = 3.4m. In Beck (2000)'s model, the height between the ceiling of the 3rd floor and the ceiling of the 4th floor is 3.46 metres, yielding a difference of only 6 centimetres. Thus, the height approximation is able to deliver very accurate results if non-generalized stairs can be detected in the evacuation plan. The missing information about the floor/ceiling thickness, however, distorts the height information towards a higher value, as the floor to ceiling height in the reference model is only 2.8 metres.

6.2 Accuracy evaluation of the refinement from pedestrian traces

As the pedestrian traces are composed of the initial values and the step vectors, an assessment of the accuracy of the positioning method has to be split into the evaluation of the step estimation's precision (see section 6.2.1) and the accuracy of the absolute positions (see section 6.2.2). In addition, the achievable precision of the refinement using a low-cost range camera is evaluated, described in section 6.2.3.

6.2.1 Precision of traces resulting from the positioning system

For the evaluation of the step estimation's precision, the related work (e.g. Nilsson et al. (2012)) proposes some strategies. For example, measuring the distance travelled by a user walking along a straight line and evaluating the distance between the equal start and end point in a closed circle walk (which additionally takes into account the geometry trueness) have been described.

Following these strategies, Schenk (2010) reports a straight-line test with a mean difference of 0.1 meters after a travelled distance of around 14 meters, in other words 0.7%. This value coincides with the evaluation results of similar positioning methods as presented in section 2.2. Additionally, a mean value for the precision of the geometry, i.e. the lateral deviation of the trace from a straight line, is given as 0.27 metres. However, due to swaying being part of the human gait (Borenstein et al., 2009), the significance of this value should not be overestimated.

Furthermore, Schenk (2010) describes a closed-loop test on an outdoor race course using the alignment method as presented in section 5.2.2. For one lap, which resulted in 421 metres of travelled distance[6], the difference between start and end point is 2.4 metres, or 0.5%. A walk of three laps resulted in 1261 metres and 3.6 metres (0.3%) difference.

6.2.2 Overall accuracy of the positioning method and the model refining features

While the step estimation yields good results in terms of precision, the accuracy of the absolute positions is highly dependent on the accuracy of the initial values, as already mentioned in section 5.2.4. In contrast to alternative methods for the acquisition of the initial values like GPS measurements, a concluding statement

[6]the 400 metres are defined at 30 centimetres from the inner border (IAAF, 2014). Thus, a lap results most certainly in more than 400 metres, depending on the sensor-equipped foot

about this accuracy is not possible for the method described in section 5.2.1. This is due to the following reasons:

- The accuracy depends on the accuracy of the *you are here* symbol's position in the plan. As described in section 6.1, due to its importance, this symbol may be subject to generalization operations like enlargement implicating displacement of the surrounding walls.
- furthermore, the accuracy depends on the accuracy of the symbol's detection.
- if the position derived using symbol detection is to be used in a world coordinate system, the accuracy of the transformation parameters has an influence on the initial position's accuracy.
- most importantly, the unknown distance between the photographed plan and the camera contributes to a great extent to the accuracy of the initial position. Unfortunately, this distance is not recoverable[7]. Thus, a final statement about the overall accuracy of the position cannot be given. By setting this distance to an arbitrary value like one metre, however, the results are expected to be at least equally accurate as clicking the position on a map displayed on a small smartphone screen or similar provisional methods when initializing the positioning inside the building. Furthermore, the accuracy may be improved by map-matching if the trace is sufficiently complex.

The positional accuracy of manually added geometric and semantic features and automatically derived door positions is directly correlated to the accuracy of the absolute coordinates resulting from the positioning system. Therefore, it suffers from the error sources described before and thus can only be estimated to be in the decimetre scale. However, for many purposes (e.g. routing in a navigation scenario) this accuracy is sufficient.

6.2.3 Accuracy of the geometric refinement using a low-cost range camera

In Peter et al. (2013b), a first comparison between the kitchen (C) model and the laser scanning point cloud of this room is described. The values shown in the first row of table 6.4 reveal a mean error of below 5 centimetres. It has to be noted that the model was aligned to the TLS point cloud using ICP before the comparison. Thus, these values represent the precision of the model rather than an absolute accuracy. Around 3 centimetres of the error can be accounted to the noise of the Kinect point cloud (Boehm, 2014).

Furthermore, the refined room models were compared to the CAD ground truth provided by Beck (2000). In contrast to the evaluation described in section 6.1, for this purpose the CAD model is only cleaned of furniture not connected to the walls. Other furniture items, like e.g. cupboards, are included in the comparison, as the employed reconstruction strategy is not capable of distinguishing between walls and such furniture.

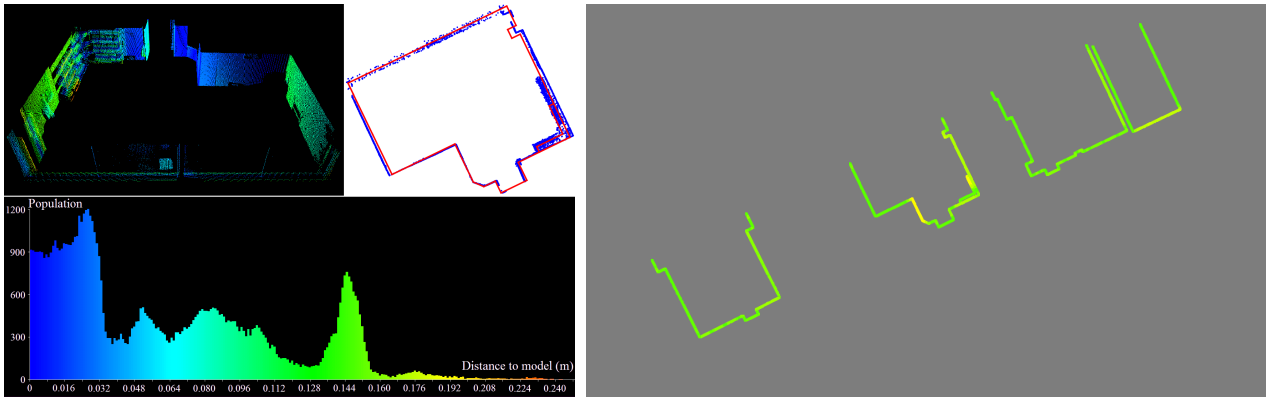
Table 6.4 and figure 6.9 show the results of this evaluation.

The height of this room's detailed model resulting from the refinement using the Kinect is reported as 2.65 meters. Comparing this to the ground truth value of 2.8 meters and considering that it reflects the real room height instead of the complete floor's height (including the height of the ceiling), this is a more useful number.

[7] if no element of known metric size is detectable in the image

Reference	*Model*	*Geo-reference*	*Mean (m)*	*StdDev (m)*	*Max (m)*	*Figure*
model from range camera	TLS point cloud	city model	0.048	0.051	0.259	6.9a)
(Beck, 2000) selected rooms	one-sided walls + range camera	city model	0.131	0.117	0.521	6.9b)

Table 6.4: Comparison results between Beck (2000) (including cupboards) respectively TLS and results from the refinement by means of the range camera



(a) (b)

Figure 6.9: a) Comparison between refined model and terrestrial laser scanning point cloud as presented in (Peter et al., 2013b), b) comparison between Beck (2000)'s model and the rooms refined using the low-cost range camera

6.3 Quality of the grammar-supported reconstruction

The data acquisition for the evaluation of this method (see figure 5.13a)) took place in the Computer Science building of the University of Stuttgart. Unfortunately, only rooms in the south-western quadrant could be accessed, a constraint which is revealed by the rooms recovered using the trace-based reconstruction depicted in figure 5.13b). Furthermore, 19% of the rooms in this quadrant are non-accessible maintenance rooms. Adding to that, the ground truth model could only be derived by digitizing it from a low-quality base map taken from a building information system, resulting in a not exceedingly accurate data source for comparison. These reasons motivate a different kind of quality assessment judging the number of reconstructed rooms by the applied approaches and just a rough idea about the absolute achievable accuracy (Philipp et al., 2014).

As the accuracy of the grammar-supported reconstruction process is highly dependent on *free* and *wall* constraints provided by the trace-based reconstruction, the model derived in areas without data is merely reproducing the room layout without sensor data base. As visible in the figures on page 99, using the purely trace-based reconstruction method 29 rooms out of the 131 rooms in the ground truth model could be derived. However, the shapes of the reconstructed rooms are directly related to the area the pedestrian traces cover, resulting mostly in very narrow rooms which do not reflect the ground truth very well. In contrast, the grammar-supported reconstruction method is able to reproduce a full model similar to the floor layout of the ground truth model, even using the grammar derived from a different floor. When applied to the non-corridor spaces derived from the incomplete corridor skeleton, the split grammar yields 92 rooms with an average room width error of around two metres. The number of rooms can be further improved if the split grammar is applied to the the non-corridor spaces derived from the corridors which were completed using the corridor grammar, resulting in 116 reconstructed rooms with a similar room width error (Becker et al., 2013). For an improvement of these results, two requirements must be fulfilled: a less constrained

Reference	*Model*	*Mean (m)*	*StdDev (m)*	*Max (m)*	*Figure*
ground truth	only rooms	1.043	1.303	9.662	6.10a)
only rooms	ground truth	1.000	1.086	4.818	6.10b)
ground truth	only corridors	0.424	0.600	4.857	6.11a)
only corridors	ground truth	0.364	0.415	2.452	6.11b)

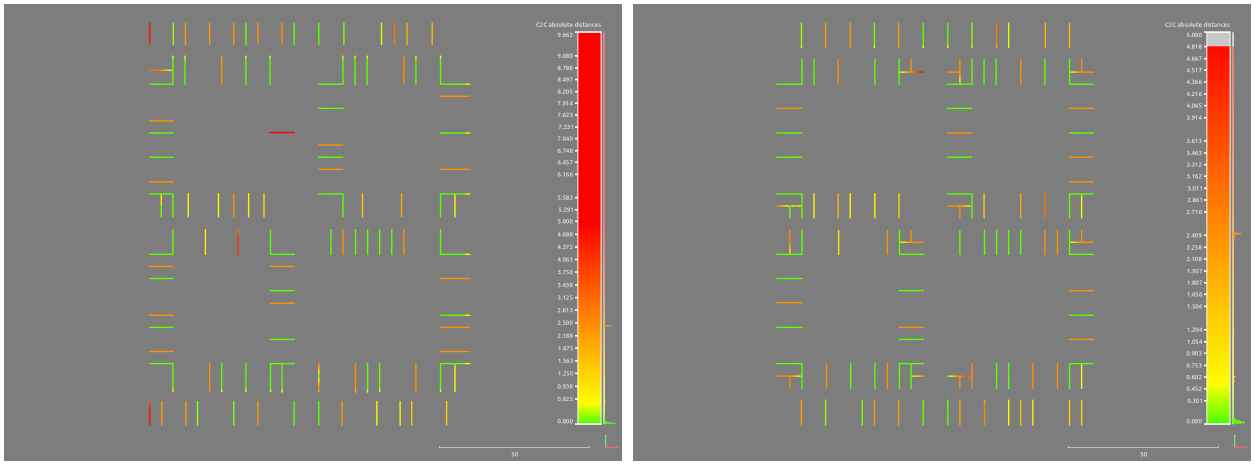
Table 6.5: Results of the comparison between trace and grammar based model and ground truth

data acquisition covering a higher percentage of the rooms to be reconstructed and traces covering more of the rooms' area.

In addition to this description, the results were analysed using CloudCompare as described in section 6.1. To this end, the external walls were removed to prevent them from biasing the results towards being overly positive. As the grammar-based reconstruction consists of two separate elements, i.e. the corridor reconstruction using the L-system and the room reconstruction using the split grammar, the analysis is differentiated accordingly. To this end, the corridor polygons are removed for the comparison of the rooms and the rooms are removed for the corridors' analysis. The results can be seen in the figures 6.10 and 6.11 as well as in table 6.5.

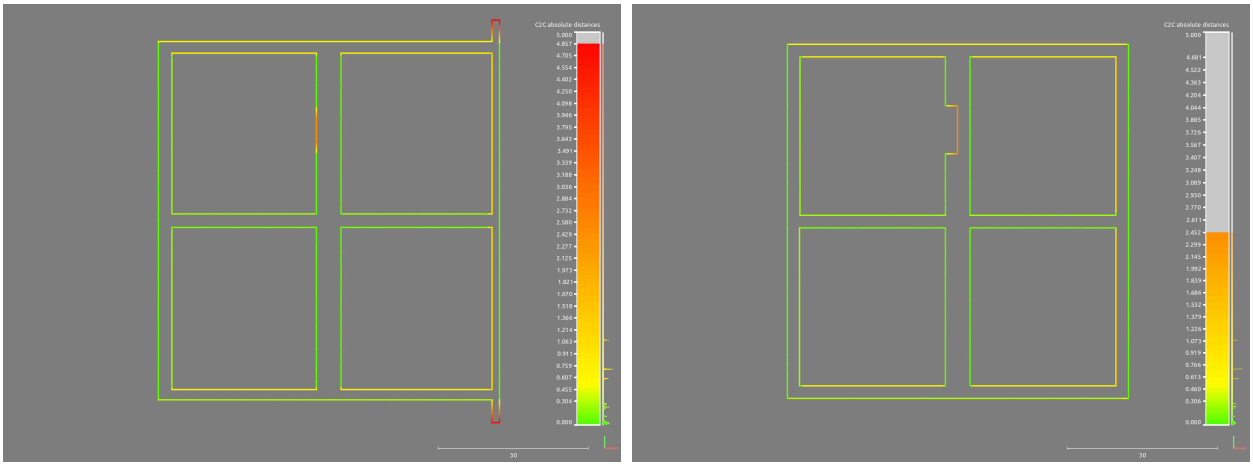
The analysis of the corridors shows that the result is fairly accurate, considering the noisy base data used, apart from two obvious reconstruction errors at the right hand end of figure 6.11.

The errors revealed by the comparison of the room walls alone, on the other hand, are generally bigger. This can be accounted to the fact that the data base for the room reconstruction is much smaller than the one which could be used for the corridor reconstruction. Most notably, however, a distinct peak can be seen around 2.4 metres in the histograms in figure 6.10. This is also due to the small data base, in combination with the "grid size" of the floor's layout, i.e. nearly all rooms' widths are integer multiples of 2.4 metres, a fact that is also reflected in the split grammar. Consequently, if a wall is placed wrongly during the grammar application without existing *free* or *wall* constraints, it is likely to be wrong by at least 2.4 metres.



(a) (b)

Figure 6.10: Room comparison results: a) ground truth used as reference model, b) inverse analysis



(a) (b)

Figure 6.11: Corridor comparison results: a) Ground truth used as reference model, b) inverse analysis

7 Conclusions and future work

In this thesis, various methods for the crowd-sourced reconstruction of building interiors have been presented. It has been shown that the automated reconstruction of building interiors by reverse-engineering of photographed evacuation plans is feasible. To this end, an image processing pipeline has been developed which enables the extraction of a 2D BREP model from such photographs using the following steps: image pre-processing, geo-referencing and scaling, symbol detection, binarisation and symbol bridging. Furthermore, a method for the automatic extrusion to a 3D model was described which is based on stair detection in the 2D model. The method has been proven to work well for plans with a known layout, however, methods for the generalisation to arbitrary plan layouts have also been developed, e.g. a purely colour-based symbol detection method. For the automatic geo-referencing and scaling of the resulting models, a contour matching method was developed which derives the sought-after parameters of the transformation between image coordinates and world coordinates by matching the indoor model to an available model of the external building shell. It was shown that this matching between models of differing LoD and scale is feasible if the models are pre-processed by a cartographic generalisation operator. As the accuracy evaluation in chapter 6 shows, the resulting models can be sufficiently accurate to serve a wide range of applications.

One of the applications is the support of indoor positioning methods using the models derived with the help of this approach as well as additional available information in the plans. To this end, methods for the derivation of the initial values - the position and orientation of the person photographing the plan with respect to the resulting model - needed for an infrastructure-independent positioning method based on a foot-mounted MEMS IMU were presented. Furthermore, the use of a model of the building's external shell for the traces' alignment as well as a map-matching method employing the coarse model were described. It was also shown that the coarse models derived from the photographed evacuation plans may be refined by analysis of these position traces. Presented methods for the refinement comprise manual semantic annotation and geometric update as well as automatic methods such as the extraction of semantic information from photographed door plates, and door reconstruction by a conjoint analysis of model and trace. A further evaluated geometric refinement method uses a combination of the available user position and the coarse model with a hand-held range camera. This approach can deliver point clouds whose analysis is able to produce more detailed room models. In a joint project the feasibility of indoor reconstruction using only pedestrian traces and an indoor grammar could be proved. Due to their different nature, the grammar is compiled from a Lindenmayer system for the corridors and a split grammar for the subdivision of non-corridor spaces into rooms or room units. In an evaluation using a great amount of distributedly collected pedestrian traces, the grammar proved suitable for the reproduction of the interior architecture's style modelled in the grammar from noisy data.

Obviously, each of the aforementioned methods implies future work and analyses, e.g. the reconstruction using photographed evacuation plans could benefit from an analysis of the plan's legend in various ways: extraction of the floor number, detection of the address in order to automatically determine the external

building shell's model to be downloaded, and extraction of symbol templates for the use in symbol detection.

As noted in section 2.1, Wi-Fi-based methods are among the most promising infrastructure-based positioning methods. These methods could benefit from the work described in this thesis in two ways: a) the availability of user positions while travelling through the building could facilitate the collection of Wi-Fi fingerprints in the offline phase and b) the coarse model derived from an evacuation plan could serve in the context of wave propagation modelling and the estimation of access point positions.

Considering the update and refinement of the coarse model using the Microsoft Kinect, the texturing of the resulting detailed room models using the Kinect's RGB image should be integrated into the processing pipeline.

The grammar-based reconstruction using only pedestrian traces needs more testing with a larger amount of data (i.e. more accessible rooms) in order to verify the approach concerning the reconstruction of rooms. Furthermore, in the experiments carried out so far, the grammar was produced from ground truth indoor models. Thus, the aim to install a closed loop consisting of learning and improving the grammar from incoming trace data and simultaneously reconstructing an indoor model was not met yet. To this end, the automatic derivation of the L-system as well as the split grammar from trace data has to be investigated. Lastly, the planned second funding period of the project is dedicated to the transition to 3D data which comprises the extension of the 2D grammar as well as the use of point clouds as base data. Following the spirit of the project, only opportunistically collected data - stemming from pervasive sensors like the emerging Google Tango Tablet or Google Glass shall be used. As a first test, Becker et al. (2015) presents the grammar-supported reconstruction from an existing LiDAR point cloud.

A general problem of indoor reconstruction which needs continued investigation is the protection of the building owner's and users' privacy and security. In previous work, the problem of privacy protection and possible solutions were explored in the context of city models and image collections like Google's Street View (Siemoneit et al., 2009; Kada et al., 2009b). Solutions included the obfuscation of parts of the building's interior which are visible through textured windows, the removal of textual information found on the façade or the adulteration of the building model's texture using sketch-like visualizations. Some of these methods might be applicable to (textured) 3D indoor models, too. However, while the exterior appearance of a building normally only reveals a small amount of details which could violate a resident's privacy, indoor spaces are much more likely to do so.

Concerning the security, there is a conflict between the necessity and obligation to mount evacuation plans in publicly accessible areas and the need to keep parts relevant to security undisclosed to the public. As shown in chapter 4, any person in the building with access to e.g. a smartphone camera can collect the full interior structure of a floor by only accessing the public areas. If the plan is visible through doors or windows this might even be possible from the outside. This poses a critical threat to security and should be taken into account when deciding about the design of the plan as well as its location.

List of Figures

List of Tables

Bibliography

K. Abdulrahim, C. Hide, T. Moore, and C. Hill. Aiding MEMS IMU with building heading for indoor pedestrian navigation. In *Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010*, page 6, October 2010. ISBN 978-1-4244-7880-4.

M. Achtelik, A. Bachrach, R. He, S. Prentice, and N. Roy. Autonomous navigation and exploration of a quadrotor helicopter in GPS-denied indoor environments. In *Proceedings of the First Symposium on Indoor Flight*, 2009.

A. Adan and D. Huber. 3d Reconstruction of Interior Wall Surfaces under Occlusion and Clutter. In *Proceedings of the 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, pages 275 –281, May 2011.

C. Ah-Soon and K. Tombre. Variations on the Analysis of Architectural Drawing. In *Proceedings of the 4th International Conference on Document Analysis and Recognition*, pages 347–351. IEEE Computer Society, 1997. ISBN 0-8186-7898-4.

S. Ahmed, M. Liwicki, M. Weber, and A. Dengel. Improved Automatic Analysis of Architectural Floor Plans. In *Proceedings of the 2011 International Conference on Document Analysis and Recognition (ICDAR)*, pages 864 –869, September 2011.

S. Ahmed, M. Liwicki, M. Weber, and A. Dengel. Automatic Room Detection and Room Labeling from Architectural Floor Plans. In *Proceedings of the 2012 10th IAPR International Workshop on Document Analysis Systems (DAS)*, pages 339 –343, March 2012.

E. Aitenbichler and M. Muhlhauser. An IR local positioning system for smart items and devices. In *Proceedings of the 23rd International Conference on Distributed Computing Systems*, pages 334–339, 2003.

M. Alzantot and M. Youssef. Uptime: Ubiquitous pedestrian tracking using mobile phones. In *Proceedings of the Wireless Communications and Networking Conference (WCNC)*, pages 3204–3209, 2012a.

M. Alzantot and M. Youssef. CrowdInside: automatic construction of indoor floorplans. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, pages 99–108, New York, NY, USA, 2012b. ISBN 978-1-4503-1691-0.

Y. Aoki, A. Shio, H. Arai, and K. Odaka. A prototype system for interpreting hand-sketched floor plans. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume 3, pages 747 –751 vol.3, 1996.

F. Aubeck, C. Isert, and D. Gusenbauer. Camera based step detection on mobile phones. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, page 7, 2011.

C. Audras, A. Comport, M. Meilland, and P. Rives. Real-time dense appearance-based SLAM for RGB-D sensors. In *Proceedings of the Australasian Conference on Robotics and Automation*, 2011.

A. Bachrach, S. Prentice, R. He, P. Henry, A. S. Huang, M. Krainin, D. Maturana, D. Fox, and N. Roy. Estimation, planning, and mapping for autonomous flight using an RGB-D camera in GPS-denied environments. *The International Journal of Robotics Research*, 31(11):1320–1343, 2012.

P. Bahl and V. N. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 775–784, 2000.

G. Bahmutov, V. Popescu, and M. Mudure. Efficient large scale acquisition of building interiors. In *Computer Graphics Forum*, volume 25, pages 655–662, 2006.

E. J. Baranoski. Visibuilding: Sensing through walls. In *Proceedings of the Fourth IEEE Workshop on Sensor Array and Multichannel Processing*, page 22, 2006.

N. Bartelme. *Geoinformatik: Modelle, Strukturen, Funktionen*. Springer, Berlin, 2005.

S. Battiato, S. Curti, M. La Cascia, M. Tortora, and E. Scordato. Depth map generation by image classification. In *Electronic Imaging 2004*, pages 95–104, 2004.

S. Beauregard. Omnidirectional pedestrian navigation for first responders. In *Proceedings of WPNC'07. 4th Workshop on Positioning, Navigation and Communication*, pages 33–36, 2007.

M. Beck. *Digitalisierung und Visualisierung eines 3D Gebäudemodells*. Study Thesis, University of Stuttgart, Stuttgart, Germany, 2000.

S. Becker. *Automatische Ableitung und Anwendung von Regeln für die Rekonstruktion von Fassaden aus heterogenen Sensordaten*. PhD Thesis, Universität Stuttgart, 2011.

S. Becker, M. Peter, D. Fritsch, D. Philipp, P. Baier, and C. Dibak. Combined Grammar for the Modeling of Building Interiors. In *ISPRS Acquisition and Modelling of Indoor and Enclosed Environments 2013 (Indoor 3D)*, 2013.

S. Becker, M. Peter, and D. Fritsch. Grammar-supported 3d Indoor Reconstruction from Point Clouds for "as-built" BIM. In *Proceedings of the 2015 conference on Photogrammetric Image Analysis (PIA2015)*, Munich, Germany, 2015.

P. Bellutta, G. Collini, A. Verri, and V. Torre. 3d visual information from vanishing points. In *Proceedings of the Workshop on Interpretation of 3D Scenes*, pages 41–49, 1989.

B. Ben-Moshe, H. Levi, P. Shamil, and A. Dvir. Indoor Positioning using Time of Flight Fingerprinting of Ultrasonic Signals. In *Proceedings of the 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN2013)*, 2013.

T. Bernoulli, M. Krammer, U. Walder, U. Dersch, and K. Zahn. Improvement of inertial sensor based indoor navigation by video content analysis. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, page 9, 2011.

T. Bernoulli, M. Krammer, S. Muhic, M. Schneeberger, and U. Walder. Context Dependent Heading Fixing Approaches in IMU-based Pedestrian Navigation. In *Proceedings of the European Navigation Conference*, Vienna, 2013.

P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on pattern analysis and machine intelligence*, pages 239–256, 1992.

P. Biber, H. Andreasson, T. Duckett, and A. Schilling. 3d modeling of indoor environments by a mobile robot with a laser scanner and panoramic camera. In *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004. (IROS 2004)*, volume 4, pages 3430 – 3435, 2004.

P. Biber, S. Fleck, D. Staneker, M. Wand, and W. Straßer. First experiences with a mobile platform for flexible 3d model acquisition in indoor and outdoor environments – the Waegele. *ISPRS Working Group V/4: 3D-ARCH*, 2005.

U. Birkel and M. Weber. Indoor localization with UMTS compared to WLAN. In *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN2012)*, page 6, 2012.

J. Bitsch Link, F. Gerdsmeier, P. Smith, and K. Wehrle. Indoor navigation on wheels (and on foot) using smartphones. In *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN2012)*, page 10, 2012.

J. Blankenbach and A. Norrdine. Position estimation using artificial generated magnetic fields. In *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN2010)*, page 5, 2010.

J. Blankenbach, A. Norrdine, H. Schlemmer, and V. Willert. Indoor-Positionierung auf Basis von Ultra Wide Band. *Zeitschriftenartikel aus: AVN Allgemeine Vermessungs-Nachrichten*, pages 169–178, 2007.

J. Boehm. Accuracy Investigation for Structured-light Based Consumer 3d Sensors. *Photogrammetrie - Fernerkundung - Geoinformation PFG*, 2:117–127, 2014.

J. Boehm and T. Pattinson. Accuracy of exterior orientation for a range camera. In *ISPRS Commission V Mid-Term Symposium: Close Range Image Measurement Techniques*, volume XXXVIII, part 5, pages 103–108, 2010.

J. Borenstein and L. Ojeda. Heuristic drift elimination for personnel tracking systems. *Journal of Navigation*, 63(04): 591–606, 2010.

J. Borenstein, L. Ojeda, and S. Kwanmuang. Heuristic reduction of gyro drift in IMU-based personnel tracking systems. In *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2009.

M. Bosse, R. Zlot, and P. Flick. Zebedee: Design of a Spring-Mounted 3-D Range Sensor with Application to Mobile Mapping. *IEEE Transactions on Robotics*, 28(5):1104–1119, 2012.

L. Bruno and P. Robertson. Wislam: Improving footslam with wifi. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, page 10, 2011.

A. Budroni. *Automatic Model Reconstruction of Manhattan-world Scenes from Dense Laser Range Data*. PhD Thesis, University of Stuttgart, Stuttgart, Germany, 2013.

A. Budroni and J. Böhm. Toward Automated Reconstruction of Interiors from Laser Data. In *Proceedings of 3D Virtual Reconstruction and Visualization of Complex Architectures (3D-ARCH'09)*, Trento, Italy, 2009.

W. Burgard, D. Fox, H. Jans, C. Matenar, and S. Thrun. Sonar-based mapping with mobile robots using EM. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 67–76. Morgan Kaufmann Publishers Inc., 1999.

J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):679–698, 1986.

G. Chen, J. Kua, S. Shum, N. Naikal, M. Carlberg, and A. Zakhor. Indoor Localization Algorithms for a Human-Operated Backpack System. *3D Data Processing, Visualization, and Transmission*, 2010.

Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha. Fm-based indoor localization. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 169–182, 2012.

K. Cheverst, J. Schöning, A. Krüger, and M. Rohs. Photomap: Snap, Grab and Walk away with a 'YOU ARE HERE' Map. In *Proceedings of MobileHCI'08: Workshop on Mobile Interaction with the Real World*, 2008.

Y. Cho, M. Ji, Y. Lee, J. Kim, and S. Park. Improved Wi-Fi AP position estimation using regression based approach. In *Proceedings of the 2012 Internation al Conference on Indoor Positioning and Indoor Navigation (IPIN2012)*, pages 13–15, 2012.

J. C. K. Chow, D. D. Lichti, J. D. Hol, G. Bellusci, and H. Luinge. IMU and Multiple RGB-D Camera Fusion for Assisting Indoor Stop-and-Go 3d Terrestrial Laser Scanning. *Robotics*, 3(3):247–280, July 2014.

J. Chung, M. Donahoe, C. Schmandt, I.-J. Kim, P. Razavai, and M. Wiseman. Indoor location sensing using geo-magnetism. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pages 141–154. ACM, 2011.

R. Czommer. *Leistungsfähigkeit fahrzeugautonomer Ortungsverfahren auf der Basis von Map-Matching-Techniken*. PhD Thesis, University of Stuttgart, 2001.

T. K. Dang and M. Worring. A semi-interactive panorama based 3d reconstruction framework for indoor scenes. *Computer Vision and Image Understanding*, 115(11):1516–1524, 2011.

A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1403–1410. IEEE, 2003.

G. Deak, K. Curran, and J. Condell. A survey of active and passive indoor localisation systems. *Computer Communications*, 35(16):1939–1954, 2012.

E. Delage, H. Lee, and A. Ng. A Dynamic Bayesian Network Model for Autonomous 3d Reconstruction from a Single Indoor Image. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2418 – 2428, 2006.

B. Denby, Y. Oussar, I. Ahriz, and G. Dreyfus. High-performance indoor localization with full-band GSM fingerprints. In *IEEE International Conference on Communications Workshops*, page 5, 2009.

C. Dibak. *Erfassung von Innenraummodellen mittels Smartphones*. Diploma Thesis, 2013.

P. Dosch, K. Tombre, C. Ah-Soon, and G. Masini. A complete system for the analysis of architectural drawings. *International Journal on Document Analysis and Recognition*, 3(2):102–116, December 2000.

P. Doubek and T. Svoboda. Reliable 3d reconstruction from a few catadioptric images. In *Omnidirectional Vision, 2002. Proceedings. Third Workshop on*, pages 71–78, 2002.

C. Duchêne, S. Bard, X. Barillot, A. Ruas, J. Trévisan, and F. Holzapfel. Quantitative and qualitative description of building orientation. In *Fifth Workshop on Progress in Automated Map Generalization*, 2003.

B. Eissfeller, A. Teuber, and P. Zucker. Indoor-GPS: ist der Satellitenempfang in Gebäuden möglich. *ZfV, Zeitschrift für Vermessung*, (4):130, 2005.

W. Elloumi, K. Guissous, A. Chetouani, R. Canals, R. Leconge, B. Emile, and S. Treuillet. Indoor navigation assistance with a Smartphone camera based on vanishing points. In *Proceedings of th 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN2013)*, page 9, 2013.

J. Elwell. Inertial navigation for the urban warrior. volume 3709, pages 196–204, 1999.

EPA. Report to Congress on indoor air quality: Volume 2. Technical Report EPA/400/1-89/001C, U.S. Environmental Protection Agency, Washington, D.C., 1989.

H. Fan, A. Zipf, Q. Fu, and P. Neis. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28(4):700–719, 2014.

D. Farin, W. Effelsberg, et al. Floor-plan reconstruction from panoramic images. In *Proceedings of MULTIMEDIA '07. The 15th international conference on Multimedia*, pages 823–826, 2007.

D. Farin and P. H. de With. Reconstructing Virtual Rooms from Panoramic Images. 2005.

FCC. First Report and Order for commercial use of UWB. Technical report, Federal Communications Commission (FCC), 2002.

FCC. Wireless E911 Location Accuracy Requirements, 2014.

P. Felzenszwalb and D. Huttenlocher. Distance transforms of sampled functions. *Theory of computing*, 8(1):415–428, 2004.

B. Ferris, D. Fox, and N. D. Lawrence. WiFi-SLAM Using Gaussian Process Latent Variable Models. In *IJCAI*, volume 7, pages 2480–2485, 2007.

A. Fietz. *Ableitung von Bewegungsstrategien zur automatisierten, vollständigen Vermessung von Innenraumszenen auf autonomnavigierender Plattform*. PhD Thesis, University of Stuttgart, August 2012.

Y. Filippovska. *Evaluierung generalisierter Gebäudegrundrisse in großen Maßstäben*. PhD Thesis, Universität Stuttgart, Stuttgart, Germany, 2012.

V. Filonenko, C. Cullen, and J. D. Carswell. Indoor positioning for smartphones using asynchronous ultrasound trilateration. *ISPRS International Journal of Geo-Information*, 2(3):598–620, 2013.

A. Flint, C. Mei, D. Murray, and I. Reid. A dynamic programming approach to reconstructing building interiors. *Computer Vision–ECCV 2010*, pages 394–407, 2010.

D. F. Fouhey, A. Gupta, and M. Hebert. Unfolding an Indoor Origami World. In *Computer Vision–ECCV 2014*, pages 687–702. Springer, 2014.

E. Foxlin. Pedestrian tracking with shoe-mounted inertial sensors. *IEEE Computer Graphics and Applications*, 25(6): 38–46, 2005.

Y. Fukuju, M. Minami, H. Morikawa, and T. Aoyama. DOLPHIN: An Autonomous Indoor Positioning System in Ubiquitous Computing Environment. In *Proceedings of the IEEE Workshop on Software Technologies for Future Embedded Systems*, pages 53–56, 2003.

Y. Furukawa and J. Ponce. Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010.

Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski. Reconstructing Building Interiors from Images. In *Proceedings of the 12th IEEE International Conference on Computer Vision (ICCV 2009)*, 2009.

R. Gao, M. Zhao, T. Ye, F. Ye, Y. Wang, K. Bian, T. Wang, and X. Li. Jigsaw: Indoor Floor Plan Reconstruction via Mobile Crowdsensing. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014.

S. Godha and G. Lachapelle. Foot mounted inertial system for pedestrian navigation. *Measurement Science and Technology*, 19(7):9, 2008.

S. Godha, G. Lachapelle, and M. E. Cannon. Integrated GPS/INS system for pedestrian navigation in a signal degraded environment. In *ION GNSS*, volume 2006, 2006.

M. Goetz and A. Zipf. Extending OpenStreetMap to Indoor Environments: Bringing Volunteered Geographic Information to the Next Level. *Urban and Regional Data Management: Udms Annual 2011*, pages 47–58, 2011.

M. Goetz and A. Zipf. Using Crowdsourced Geodata for Agent-Based Indoor Evacuation Simulations. *ISPRS International Journal of Geo-Information*, 1(2):186–208, 2012.

A. R. Golding and N. Lesh. Indoor navigation using a diverse set of cheap, wearable sensors. In *Wearable Computers, 1999. Digest of Papers. The Third International Symposium on*, pages 29–36. IEEE, 1999.

R. Gonzales, R. Woods, and S. Eddins. *Digital Image Processing Using Matlab. 2004*. Prentice Hall, 2004.

G. Gröger and L. Plümer. Derivation of 3d Indoor Models by Grammars for Route Planning. *Photogrammetrie-Fernerkundung-Geoinformation*, 2010(3):193–210, 2010.

GSA. U.S. General Services Administration BIM Guide for 3d Imaging version 1.0, January 2009.

Y. Gu, A. Lo, and I. Niemegeers. A survey of indoor positioning systems for wireless personal networks. *IEEE Communications Surveys & Tutorials*, 11(1):13–32, 2009.

N. Haala, M. Peter, J. Kremer, and G. Hunter. Mobile LIDAR Mapping for 3d Point Cloud Collection in Urban Areas - A Performance Test. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume Vol. XXXVII, Part B2, Commission 2, Beijing, China, 2008a.

N. Haala, D. Fritsch, M. Peter, and A. Khosravani. Pedestrian navigation and modeling for indoor environments. In *Proceedings of the 7th International Symposium on Mobile Mapping Technology*, 2011.

N. Haala, D. Fritsch, M. Peter, and A. Khosravani. Pedestrian Mobile Mapping System for Indoor Environmens based on MEMS IMU and Range Camera. *Archives of Photogrammetry, Cartography and Remote Sensing*, 22:159–172, 2012.

N. Haala, M. Peter, A. Cefalu, and J. Kremer. Mobile Lidar Mapping For Urban Data Capture. In *Digital Heritage - Proceedings of the 14th International Conference on Virtual Systems and Multimedia*, pages pp. 95–100, 2008b.

E. Hahn. *Persistent realtime building interior generation.* MSc Thesis, Carleton University, Ottawa, Canada, 2006.

D. Hähnel, W. Burgard, and S. Thrun. Learning compact 3d models of indoor and outdoor environments with a mobile robot. *Robotics and Autonomous Systems*, 44(1):15–27, July 2003.

R. Harle. A Survey of Indoor Inertial Positioning Systems for Pedestrians. *IEEE Communications Surveys Tutorials*, 15 (3):1281–1293, 2013.

A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster. The Anatomy of a Context-Aware Application. *Wireless Networks*, 8(2-3):187–197, March 2002.

V. Hedau, D. Hoiem, and D. Forsyth. Recovering the spatial layout of cluttered rooms. In *Computer vision, 2009 IEEE 12th international conference on*, pages 1849–1856. IEEE, 2009.

P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. RGB-D mapping: Using depth cameras for dense 3d modeling of indoor environments. In *the 12th International Symposium on Experimental Robotics (ISER)*, 2010.

P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. RGB-D mapping: Using Kinect-style depth cameras for dense 3d modeling of indoor environments. *The International Journal of Robotics Research 2012*, 31:647–664, 2012.

J. Hightower and G. Borriello. Location systems for ubiquitous computing. *Computer*, 34(8):57–66, 2001.

J. Hightower, R. Want, and G. Borriello. SpotON: An indoor 3d location sensing technology based on RF signal strength. *UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle, WA*, 1, 2000.

H. Hile and G. Borriello. Positioning and Orientation in Indoor Environments Using Camera Phones. *IEEE Computer Graphics and Applications*, 28(4):32–39, 2008.

S. Hochdorfer and C. Schlegel. 6 DoF SLAM using a ToF camera: The challenge of a continuously growing number of landmarks. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3981–3986, October 2010.

F. Hoflinger, R. Zhang, J. Hoppe, A. Bannoura, L. Reindl, J. Wendeberg, M. Buhrer, and C. Schindelhauer. Acoustic Self-calibrating System for Indoor Smartphone Tracking (ASSIST). In *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–9, November 2012.

S. Holm. Ultrasound positioning based on time-of-flight and signal strength. In *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN2012)*, page 6, 2012.

H. Huang, H. Jiang, C. Brenner, and H. Mayer. Object-level Segmentation of RGBD Data. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II(3):73–78, 2014.

J. Huang, D. Millman, M. Quigley, D. Stavens, S. Thrun, and A. Aggarwal. Efficient, generalized indoor WiFi GraphSLAM. In *2011 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1038–1043, 2011.

J. Huang and B. Cowan. Simple 3d Reconstruction of Single Indoor Image with Perspective Cues. In *Canadian Conference on Computer and Robot Vision, 2009. CRV '09*, pages 140–147, May 2009.

A. Hub, J. Diepstraten, and T. Ertl. Augmented indoor modeling for navigation support for the blind. In *Conference Proceedings: CPSN'05-The International Conference on Computers for People with Special Needs*, pages 54–59, 2005.

R. Huitl, G. Schroth, S. Hilsenbeck, F. Schweiger, and E. Steinbach. TUMindoor: An extensive image and point cloud dataset for visual indoor localization and mapping. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 1773–1776, 2012.

IAAF. International Association of Athletics Federations - Competition Rules 2014-2015, 2014.

L. Iocchi, K. Konolige, and M. Bajracharya. Visually realistic mapping of a planar environment with stereo. In *Experimental Robotics VII*, pages 521–532. Springer, 2001.

P. Jenke, B. Huhle, and W. Strasser. Statistical Reconstruction of Indoor Scenes. In *Proceedings of the 17th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Plzen-Bory, Czech Republic, 2009.

Y. Jiang, X. Pan, K. Li, Q. Lv, R. P. Dick, M. Hannigan, and L. Shang. Ariel: Automatic wi-fi based room fingerprinting for indoor localization. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 441–450, 2012.

Y. Jiang, Y. Xiang, X. Pan, K. Li, Q. Lv, R. P. Dick, L. Shang, and M. Hannigan. Hallway based automatic indoor floorplan construction using room fingerprints. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 315–324, 2013.

A. R. Jimenez, F. Seco, C. Prieto, and J. Guevara. A comparison of pedestrian dead-reckoning algorithms using a low-cost MEMS IMU. In *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*, pages 37–42, 2009.

A. R. Jiménez, F. Seco, F. Zampella, J. C. Prieto, and J. Guevara. Improved Heuristic Drift Elimination (iHDE) for pedestrian navigation in complex buildings. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, page 8, 2011a.

A. R. Jiménez, F. Seco, F. Zampella, J. C. Prieto, and J. Guevara. Ramp Detection with a Foot-Mounted IMU for a Drift-Free Pedestrian Position Estimation. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, 2011b.

A. R. Jiménez, F. Zampella, and F. Seco. Light-Matching: A new Signal of Opportunity for Pedestrian Indoor Navigation. In *Proceedings of the 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN2013)*, 2013.

A. R. Jiménez, F. Seco, F. Zampella, J. C. Prieto, and J. Guevara. Improved heuristic drift elimination with magnetically-aided dominant directions (MiHDE) for pedestrian navigation in complex buildings. *Journal of location based services*, 6(3):186–210, 2012.

Y. Jin, H.-S. Toh, W.-S. Soh, and W.-C. Wong. A robust dead-reckoning pedestrian tracking system with low cost sensors. In *Pervasive Computing and Communications (PerCom), 2011 IEEE International Conference on*, pages 222–230. IEEE, 2011.

M. Johnston and A. Zakhor. Estimating building floor-plans from exterior using laser scanners. In *Proceedings of the SPIE Electronic Imaging Conference, 3D Image Capture and Applications*, San Jose, California, 2008.

M. Kada, M. Peter, and Y. Filippovska. 3d Building Generalization under Ground Plan Preserving Quality Constraints. In *Proceedings of the 6th International Symposium on Spatial Data Quality*, St. John's, Canada, 2009a.

M. Kada. *Zur maßstabsabhängigen Erzeugung von 3D-Stadtmodellen*. PhD Thesis, Universität Stuttgart, 2007.

M. Kada, M. Peter, O. Siemoneit, D. Fritsch, and C. Hubig. Privacy-Enabling Abstraction and Obfuscation Techniques for 3d City Models. In *Proceedings of the 2nd SIGSPATIAL ACM GIS International Workshop on Security and Privacy in GIS and LBS*, Seattle, USA, 2009b.

D. Kashlev. *Efficient 3D building model generation from 2D floor plans*. Thesis, Massachusetts Institute of Technology, 2008. Thesis (M. Eng.)–Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 2008.

K. Khoshelham and L. Díaz-Vilariño. 3d Modelling of Interior Spaces: Learning the Language of Indoor Architecture. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL(5):321–326, 2014.

A. Khosravani, M. Peter, and D. Fritsch. Alignment of Range Image Data Based on MEMS IMU and 3d Coarse Models Derived from Evacuation Plans. In *SPIE Videometrics, Range Imaging, and Applications XII*, volume 8791A 1-6, Munich, May 2013.

T. King, S. Kopf, T. Haenselmann, C. Lubberger, and W. Effelsberg. Compass: A probabilistic indoor positioning system based on 802.11 and digital compasses. In *Proceedings of the 1st international workshop on Wireless network testbeds, experimental evaluation & characterization*, pages 34–40, 2006.

N. Kiryati, Y. Eldar, and A. M. Bruckstein. A probabilistic Hough transform. *Pattern recognition*, 24(4):303–316, 1991.

G. Klein and D. Murray. Parallel tracking and mapping on a camera phone. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 83–86, 2009.

G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234. IEEE, 2007.

M. Köhler, S. N. Patel, J. W. Summet, E. P. Stuntebeck, and G. D. Abowd. Tracksense: Infrastructure free precise indoor positioning using projected patterns. In *Pervasive Computing*, pages 334–350. Springer, 2007.

T. K. Kohoutek, R. Mautz, and A. Donaubauer. Real-time indoor positioning using range imaging sensors. In *SPIE Photonics Europe*, pages 77240K–77240K. International Society for Optics and Photonics, 2010.

T. H. Kolbe, G. Gröger, and L. Plümer. CityGML: Interoperable access to 3d city models. In *Geo-information for disaster management*, pages 883–899. Springer, 2005.

W. Kong and B. B. Kimia. On solving 2d and 3d puzzles using curve matching. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–583. IEEE, 2001.

H. S. Koppula, A. Anand, T. Joachims, and A. Saxena. Semantic labeling of 3d point clouds for indoor scenes. *Proceedings of the Advances in Neural Information Processing Systems*, 2011.

T. Kotthauser, M. Divband Soorati, and B. Mertsching. Automatic reconstruction of polygonal room models from 3d point clouds. In *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 661–667, December 2013.

A. Koutamanis and V. Mitossi. Automated recognition of architectural drawings. In *Proceedings of the 11th IAPR International Conference on Pattern Recognition, 1992. Vol.I. Conference A: Computer Vision and Applications*, pages 660–663, 1992.

H. Koyuncu and S. H. Yang. A survey of indoor positioning and object locating systems. *IJCSNS International Journal of Computer Science and Network Security*, 10(5):121–128, 2010.

J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer. Multi-camera multi-person tracking for easyliving. In *Visual Surveillance, 2000. Proceedings. Third IEEE International Workshop on*, pages 3–10. IEEE, 2000.

H. Kuusniemi, M. Z. H. Bhuiyan, M. Strom, S. Soderholm, T. Jokitalo, L. Chen, and R. Chen. Utilizing pulsed pseudolites and high-sensitivity GNSS for ubiquitous outdoor/indoor satellite navigation. In *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN2012) IPIN*, page 7, 2012.

G. Lachapelle. GNSS indoor location technologies. *Journal of Global Positioning Systems*, 3(1-2):2–11, 2004.

F. Lassabe, Y. Zheng, O. Baala, and A. Caminada. Comparison of measurement-based and simulation-based indoor Wi-Fi positioning algorithms. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, Guimaraes, Portugal, 2011.

J. Ledlie, J.-g. Park, D. Curtis, A. Cavalcante, L. Camara, A. Costa, and R. Vieira. Molé: a scalable, user-generated WiFi positioning engine. *Journal of Location Based Services*, 6(2):55–80, 2012.

D. C. Lee. *Three Dimensional Representation and Reasoning for Indoor Scene Understanding*. PhD Thesis, Carnegie Mellon University, 2011.

J. M. Leiva, P. Martinez, E. J. Perez, C. Urdiales, and F. Sandoval. 3d reconstruction of static indoor environment by fusion of sonar and video data. In *Proceedings of International Symposium on Intelligent Robotic Systems, Toulouse (France)*, 2001.

J. J. Leonard and H. F. Durrant-Whyte. Simultaneous map building and localization for an autonomous mobile robot. In *Intelligent Robots and Systems' 91.'Intelligence for Mechanical Systems, Proceedings IROS'91. IEEE/RSJ International Workshop on*, pages 1442–1447, 1991.

R. Lewis. *Generating three-dimensional building models from two- dimensional architectural plans*. MSc Thesis, Citeseer, 1996.

R. Lewis and C. Séquin. Generation of 3d building models from 2d architectural plans. *Computer-Aided Design*, 30 (10):765–779, 1998.

B. Li, T. Gallagher, A. G. Dempster, and C. Rizos. How feasible is the use of magnetic field alone for indoor positioning? In *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN2012)*, page 9, 2012.

J. Link, P. Smith, and K. Wehrle. FootPath: Accurate Map-based Indoor Navigation Using Smartphones. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, Guimaraes, Portugal, 2011.

T. Liu, M. Carlberg, G. Chen, J. Chen, J. Kua, and A. Zakhor. Indoor localization and visualization using a human-operated backpack system. In *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN2010)*, page 10, 2010.

J. Llados, J. Lopez-Krahe, and E. Marti. Hand drawn document understanding using the straight line Hough transform and graph matching. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 2, pages 497–501, 1996.

S. Lopes, J. Vieira, J. Reis, D. Albuquerque, and N. Carval. Accurate Smartphone Indoor Positioning Using Non-Invasive Audio. In *Proceedings of the 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN2013)*, pages 162–165, Montbéliard, France, 2013.

T. Lu, C. L. Tai, L. Bao, F. Su, and S. Cai. 3d Reconstruction of detailed buildings from architectural drawings. *Computer-Aided Design and Applications*, 2(1-4):527–536, 2005.

A. Madhavapeddy, D. Scott, and R. Sharp. Context-aware computing with sound. In *UbiComp 2003: Ubiquitous Computing*, pages 315–332. Springer, 2003.

A. Mandal, C. Lopes, T. Givargis, A. Haghighat, R. Jurdak, and P. Baldi. Beep: 3d indoor positioning using audible sound. In *2005 Second IEEE Consumer Communications and Networking Conference, 2005. CCNC*, pages 348–353, 2005.

N. Marques, F. Meneses, and A. Moreira. Combining similarity functions and majority rules for multi-building, multi-floor, WiFi positioning. In *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN2012)*, page 9, 2012.

F. Marson and S. R. Musse. Automatic real-time generation of floor plans based on squarified treemaps algorithm. *International Journal of Computer Games Technology*, 2010:7, 2010.

O. Martinez Mozos, R. Triebel, P. Jensfelt, A. Rottmann, and W. Burgard. Supervised semantic labeling of places using information extracted from sensor data. *Robotics and Autonomous Systems*, 55(5):391–402, 2007.

A. Matic, A. Papliatseyeu, V. Osmani, and O. Mayora-Ibarra. Tuning to your position: FM radio based indoor localization with spontaneous recalibration. In *Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on*, pages 153–161, 2010.

R. Mautz and S. Tilch. Survey of optical indoor positioning systems. In *2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, page 7, September 2011.

R. Mautz. Overview of current indoor positioning systems. *Geodezija ir kartografija*, 35(1):18–22, 2009.

S. May, D. Droeschel, S. Fuchs, D. Holz, and A. Nuchter. Robust 3d-mapping with time-of-flight cameras. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009. IROS 2009*, pages 1673–1678, October 2009.

C. Medina, J. C. Segura, and S. Holm. Feasibility of ultrasound positioning based on signal strength. In *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN2012)*, page 9, 2012.

M. Mirahmadi and A. Shami. A novel algorithm for real-time procedural generation of building floor plans. *arXiv preprint arXiv:1211.5842*, 2012.

V. Moghtadaiee, A. G. Dempster, and S. Lim. Indoor localization using fm radio signals: A fingerprinting approach. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, page 7, 2011.

P. Müller, G. Zeng, P. Wonka, and L. Van Gool. Image-based procedural modeling of facades. *ACM Transactions on Graphics*, 26(3):85, 2007.

M. Nakagawa, K. Kataoka, T. Yamamoto, M. Shiozaki, and T. Ohhashi. Panoramic rendering-based polygon extraction from indoor mobile LiDAR data. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1:181–186, 2014.

M. Nakajima and S. Haruyama. New indoor navigation system for visually impaired people using visible light communication. *EURASIP Journal on Wireless Communications and Networking*, 2013(1):1–10, 2013.

E. Neufert, P. Neufert, B. Baiche, and N. Walliman. *Architects' Data*. Wiley-Blackwell, 3rd edition, August 2002. ISBN 0632057718.

R. A. Newcombe and A. J. Davison. Live dense reconstruction with a single moving camera. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1498–1505, 2010.

R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 127 –136, October 2011.

L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil. LANDMARC: indoor location sensing using active RFID. *Wireless networks*, 10 (6):701–710, 2004.

J.-O. Nilsson, I. Skog, P. Handel, and K. V. S. Hari. Foot-mounted INS for everybody-an open-source embedded implementation. In *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION*, pages 140–145. IEEE, 2012.

A. Norrdine and J. Blankenbach. Indoor-Positionierung mit künstlichen Magnetfeldern. Von der Innenraumpositionierung zu standortbezogenen Diensten in Gebäuden. *Zeitschrift für Vermessungswesen*, (1), 2013.

A. Nüchter, H. Surmann, K. Lingemann, and J. Hertzberg. Semantic scene analysis of scanned 3d indoor environments. In *Proc of the VMV Conference*, 2003.

L. Ojeda and J. Borenstein. Non-GPS navigation for emergency responders. In *2006 International Joint Topical Meeting: Sharing Solutions for Emergencies and Hazardous Environments*, pages 12–15. Citeseer, 2006.

B. Okorn, X. Xiong, B. Akinci, and D. Huber. Toward automated modeling of floor plans. In *Proceedings of the Symposium on 3D Data Processing, Visualization and Transmission*, volume 2, 2010.

P. Olivieri, M. Gatti, M. Straforini, and V. Torre. A method for the 3d reconstruction of indoor scenes from monocular images. In G. Sandini, editor, *Computer Vision — ECCV'92*, number 588 in Lecture Notes in Computer Science, pages 696–700. Springer Berlin Heidelberg, January 1992.

S. Or, K. H. Wong, Y. Yu, and M. M. Chang. Highly Automatic Approach to Architectural Floorplan Image Understanding & Model Generation. *Proc. Vision, Modeling, and Visualization, IOS Press*, pages 25–32, 2005a.

S.-h. Or, K.-H. Wong, Y.-k. Yu, M. M.-y. Chang, and H. Kong. Highly automatic approach to architectural floor plan image understanding and model generation. *Proceedings of Vision, Modeling and Visualization 2005*, pages 25–32, 2005b.

O. Oreifej, J. Cramer, and A. Zakhor. Automatic Generation of 3d Thermal Maps of Building Interiors. In *Proceedings of the ASHRAE Annual Conference*, Seattle, USA, 2014.

R. J. Orr and G. D. Abowd. The smart floor: a mechanism for natural user identification and tracking. In *CHI'00 extended abstracts on Human factors in computing systems*, pages 275–276, 2000.

V. Otsason, A. Varshavsky, A. LaMarca, and E. De Lara. Accurate gsm indoor localization. In *UbiComp 2005: Ubiquitous Computing*, pages 141–158. Springer, 2005.

T. Otto. *Automatische Innenraum-Modellierung auf Basis von Fluchtplänen*. Diploma Thesis, University of Stuttgart, Stuttgart, Germany, 2014.

D. Pagliari, F. Menna, R. Roncella, F. Remondino, and L. Pinto. Kinect Fusion Improvement using Depth Camera Calibration. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5):479–485, 2014.

Y. I. H. Parish and P. Müller. Procedural modeling of cities. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '01, pages 301–308, New York, NY, USA, 2001. ACM. ISBN 1-58113-374-X.

B. Peasley and S. Birchfield. Replacing Projective Data Association with Lucas-Kanade for KinectFusion. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 638–645. IEEE, 2013.

M. Peter, S. Becker, and D. Fritsch. Grammar Supported Indoor Mapping. In *Proceedings of the 26th International Cartographic Conference*, page 18, Dresden, August 2013a.

M. Peter, A. M. Khosravani, and D. Fritsch. Refinement of Coarse Indoor Models using Position Traces and a Low-Cost Range Camera. In *Proceedings of the 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN2013)*, pages 787–795, Montbéliard, France, October 2013b.

M. Peter. Presentation and Evaluation of Inconsistencies in Multiply Represented 3d Building Models. In *Proceedings of QuaCon 2009 - First International Workshop on Quality of Context Data*, Stuttgart, Germany, 2009.

M. Peter, N. Haala, and D. Fritsch. Preserving Ground Plan and Facade Lines for 3d Building Generalization. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XXXVII, Part B2, Commission 2, page 481, Beijing, China, 2008. ISBN ISSN 1682-1750.

M. Peter, N. Haala, M. Schenk, and T. Otto. Indoor Navigation and Modeling using Photographed Evacuation Plans and MEMS IMU. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII, Part 4 Commission IV Symposium, Orlando, USA, November 15-19, 2010.

D. Philipp, P. Baier, C. Dibak, F. Dürr, K. Rothermel, S. Becker, M. Peter, and D. Fritsch. MapGENIE: Grammar-enhanced Indoor Map Construction from Crowd-sourced Data. In *Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2014.

A. Popleteev. *Indoor positioning using FM radio signals*. PhD Thesis, University of Trento, 2011.

M. Previtali, M. Scaioni, L. Barazzetti, and R. Brumana. A flexible Methodology for Outdoor/Indoor Building Reconstruction from Occluded Point Clouds. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II(3):119–126, 2014.

L. Priese and P. Sturm. Introduction to the color structure code and its implementation. Technical Report, Fachbereich Informatik, Universität Koblenz-Landau, 2003.

N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The cricket location-support system. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 32–43, 2000.

A. Prusak, O. Melnychuk, H. Roth, I. Schiller, and R. Koch. Pose estimation and map building with a pmd-camera for robot navigation. In *Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D)*, volume 1, 2007.

M. G. Puyol, M. Frassl, and P. Robertson. Collaborative mapping for pedestrian navigation in security applications. In *Future Security*, pages 49–60. Springer, 2012.

M. Quigley, D. Stavens, A. Coates, and S. Thrun. Sub-meter indoor localization in unmodified environments with inexpensive sensors. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 2039–2046. IEEE, 2010.

F. H. Raab, E. B. Blood, T. O. Steiner, and H. R. Jones. Magnetic position and orientation tracking system. *Aerospace and Electronic Systems, IEEE Transactions on*, (5):709–718, 1979.

T. Rabbani, F. van den Heuvel, and G. Vosselmann. Segmentation of point clouds using smoothness constraint. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5):248–53, 2006.

M. Rabinowitz and J. Spilker, J.J. A new positioning system using television synchronization signals. *IEEE Transactions on Broadcasting*, 51(1):51–61, March 2005.

N. Ravi and L. Iftode. Fiatlux: Fingerprinting rooms using light intensity. In *Proceedings of the 2007 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2007.

V. Renaudin, M. Susi, and G. Lachapelle. Step length estimation using handheld inertial sensors. *Sensors*, 12(7):8507–8525, 2012.

I. Rishabh, D. Kimber, and J. Adcock. Indoor localization using controlled ambient sounds. In *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN2012)*, page 10, 2012.

C. Rizos, Y. Li, and W. Jiang. Locata for Indoor Positioning using Beam-Forming Technology, 2013.

P. Robertson, M. Angermann, and B. Krach. Simultaneous localization and mapping for pedestrians using only foot-mounted inertial sensors. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 93–96, 2009.

P. Robertson, M. Angermann, and M. Khider. Improving simultaneous localization and mapping for pedestrian navigation and automatic mapping of buildings by using online human-based feature labeling. In *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION*, pages 365–374. IEEE, 2010.

H. Roth and M. Vona. Moving volume kinectfusion. In *British Machine Vision Conf.(BMVC),(Surrey, UK)*, 2012.

L. Ruotsalainen, H. Kuusniemi, and R. Chen. Heading change detection for indoor navigation with a Smartphone camera. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, page 7, 2011.

R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz. Towards 3d Point cloud based object maps for household environments. *Robotics and Autonomous Systems*, 56(11):927–941, November 2008.

K. Ryall, S. Shieber, J. Marks, and M. Mazer. Semi-automatic delineation of regions in floor plans. In *Proceedings of the Third International Conference on Document Analysis and Recognition*, volume 2, pages 964–969, 1995.

A. C. Salas. *Indoor Positioning System based on Bluetooth Low Energy*. PhD Thesis, Universitat Politècnica de Catalunya, 2014.

T. Sato, S. Nakamura, K. Terabayashi, M. Sugimoto, and H. Hashizume. Design and implementation of a robust and real-time ultrasonic motion-capture system. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, page 6, 2011.

J. Sauvola and M. Pietikäinen. Adaptive document image binarization. *Pattern recognition*, 33(2):225–236, 2000.

A. Saxena, S. H. Chung, and A. Y. Ng. 3-D Depth Reconstruction from a Single Still Image. *International Journal of Computer Vision*, 76(1):53–69, January 2008.

B. Schäfer. Determination of heading information from gait cycle pattern using stride length estimation with reduced IMUs on right and left foot. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011), Guimaraes, Portugal*, 2011.

M. Schenk. *Innenraumnavigation mit einer MEMS-Inertialmesseinheit*. Study Thesis, University of Stuttgart, Stuttgart, Germany, 2010.

C. K. Schindhelm and A. MacWilliams. Overview of Indoor Positioning Technologies for Context Aware AAL Applications. In R. Wichert and B. Eberhardt, editors, *Ambient Assisted Living*, pages 273–291. Springer Berlin Heidelberg, January 2011. ISBN 978-3-642-18166-5, 978-3-642-18167-2.

A. Schlichting and C. Brenner. Localization using automotive laser scanners and local pattern matching. In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pages 414–419. IEEE, 2014.

M. Schneider. *Erstellung von Innenraum 3D Modellen mit dem Kinect for Windows v2 Sensor*. Diploma Thesis, University of Stuttgart, 2014.

S. Schon and O. Bielenberg. On the capability of high sensitivity GPS for precise indoor positioning. In *5th Workshop on Positioning, Navigation and Communication, 2008. WPNC 2008*, pages 121–127, 2008.

J. Schöning, A. Krüger, K. Cheverst, M. Rohs, M. Löchtefeld, and F. Taher. PhotoMap: using spontaneously taken images of public maps for pedestrian navigation tasks on mobile devices. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, page 14, 2009.

V. Schwieger. Positioning within the GSM Network. In *Proceedings of FIG Regional Conference*, volume 350, page 100, 2007.

S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka. Spot localization using phy layer information. In *Proceedings of ACM MOBISYS*, 2012.

T. Shao, W. Xu, K. Zhou, J. Wang, D. Li, and B. Guo. An interactive approach to semantic modeling of indoor scenes with an rgbd camera. *ACM Transactions on Graphics (TOG)*, 31(6):136, 2012.

G. Shen, Z. Chen, P. Zhang, T. Moscibroda, and Y. Zhang. Walkie-Markie: indoor pathway mapping made easy. In *Proceedings of the 10th USENIX conference on Networked Systems Design and Implementation*, pages 85–98, 2013.

H. Shin, Y. Chon, and H. Cha. Unsupervised Construction of an Indoor Floor Plan Using a Smartphone. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 42(6):889–898, 2012.

H.-Y. Shum, M. Han, and R. Szeliski. Interactive construction of 3d models from panoramic mosaics. In *1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1998. Proceedings*, pages 427–433, 1998.

O. Siemoneit, C. Hubig, M. Kada, M. Peter, and D. Fritsch. Google StreetView and Privacy. In *Proceedings of the Fifth Asia-Pacific Computing and Philosophy Conference*, Tokyo, Japan, 2009.

R. Smith. An overview of the Tesseract OCR engine. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 2, pages 629–633, 2007.

N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. *ACM Transactions on Graphics (SIGGRAPH Proceedings)*, 25(3):835–846, 2006.

C. So, G. Baciu, and H. Sun. Reconstruction of 3d virtual buildings from 2d architectural floor plans. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 17–23, 1998.

J. Song, S. Hur, Y. Kim, K. Yoo, and Y. Park. Comparison of WLAN and Geomagnetic Fields for Indoor Positioning. In *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN2012)*, page 3, 2012.

M. Spindler, M. Weber, D. Prescher, M. Miao, G. Weber, and G. Ioannidis. Translating Floor Plans into Directions. In K. Miesenberger, A. Karshmer, P. Penaz, and W. Zagler, editors, *Computers Helping People with Special Needs*, volume 7383 of *Lecture Notes in Computer Science*, pages 59–66. Springer Berlin / Heidelberg, 2012. ISBN 978-3-642-31533-6.

H. Sternberg, F. Keller, and T. Willemsen. Precise indoor mapping as a basis for coarse indoor navigation. *Journal of Applied Geodesy*, 7(4):231–246, 2013.

G. Stiny and W. J. Mitchell. The palladian grammar. *Environment and Planning B*, 5(1):5–18, 1978.

R. Stirling, J. Collin, K. Fyfe, and G. Lachapelle. An innovative shoe-mounted pedestrian navigation system. In *Proceedings of European navigation conference GNSS*, page 15, 2003.

D. Stojanović and N. Stojanović. Indoor Localization and Tracking: Methods, Technologies and Research Challenges. *Facta Universitatis, Series: Automatic Control and Robotics*, 13(1):57–72, 2014.

M. Straforini, C. Coelho, M. Campani, and V. Torre. The recovery and understanding of a line drawing from indoor scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):298 –303, February 1992.

H. Surmann, A. Nüchter, and J. Hertzberg. An autonomous mobile robot with a 3d laser range finder for 3d exploration and digitalization of indoor environments. *Robotics and Autonomous Systems*, 45(3):181–198, 2003.

S. Suzuki et al. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985.

F. Tang and A. Ren. GIS-based 3d evacuation simulation for indoor fire. *Building and Environment*, 49:193–202, March 2012.

J. D. Tardós, J. Neira, P. M. Newman, and J. J. Leonard. Robust Mapping and Localization in Indoor Environments Using Sonar Data. *The International Journal of Robotics Research*, 21(4):311–330, April 2002.

S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik. Indoor localization without infrastructure using the acoustic background spectrum. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pages 155–168. ACM, 2011.

T. M. Tykkälä. *Real-time Image-based RGB-D Camera Motion Tracking and Environment Mapping*. PhD Thesis, Lappeenranta University of Technology, Lappeenranta, Finland, 2013.

E. Valero, A. Adán, and C. Cerrada. Automatic Method for Building Indoor Boundary Models from Dense Point Clouds Collected by Laser Scanners. *Sensors*, 12(12):16099–16115, November 2012.

A. Vanclooster, P. Viaene, N. Van de Weghe, V. Fack, and P. De Maeyer. Analyzing the applicability of the least risk path algorithm in indoor space. In *ISPRS Acquisition and Modelling of Indoor and Enclosed Environments 2013 (Indoor 3D)*, pages 19–26. International Society for Photogrammetry and Remote Sensing (ISPRS), 2013.

C. L. Vaughan, B. L. Davis, and J. C. O'connor. *Dynamics of human gait*. Human Kinetics Publishers Champaign, Illinois, 1992.

R. C. Veltkamp. Shape matching: Similarity measures and algorithms. In *Shape Modeling and Applications, SMI 2001 International Conference on.*, pages 188–197. IEEE, 2001.

S. Vidas, P. Moghadam, and M. Bosse. 3d thermal mapping of building interiors using an RGB-D and thermal camera. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 2311–2318. IEEE, 2013.

U. Walder and T. Bernoulli. Context-adaptive algorithms to improve indoor positioning with inertial sensors. In *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 965–970, 2010.

H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury. No Need to War-drive: Unsupervised Indoor Localization. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, pages 197–210, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1301-8.

H. Wang, S. Gould, and D. Koller. Discriminative learning with latent variables for cluttered indoor scene understanding. In *Computer Vision–ECCV 2010*, pages 497–510. Springer, 2010.

R. Want, A. Hopper, V. Falcão, and J. Gibbons. The active badge location system. *ACM Transactions on Information Systems (TOIS)*, 10(1):91–102, 1992.

H. Weinberg. Using the ADXL202 in pedometer and personal navigation applications. Analog Devices AN-602 Application Note, 2002.

T. Whelan, M. Kaess, M. Fallon, H. Johannsson, J. Leonard, and J. McDonald. Kintinuous: Spatially Extended KinectFusion. In *Proceedings of the RSS Workshop on RGB-D: Advanced Reasoning with Depth Cameras*, 2012.

C. E. White, D. Bernstein, and A. L. Kornhauser. Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies*, 8(1):91–108, 2000.

T. Wiemann, A. Nuchter, K. Lingemann, S. Stiene, and J. Hertzberg. Automatic construction of polygonal maps from point cloud data. In *2010 IEEE International Workshop on Safety Security and Rescue Robotics (SSRR)*, page 6, 2010.

O. Woodman and R. Harle. Pedestrian localisation for indoor environments. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 114–123, 2008.

O. Wulf, K. O. Arras, H. I. Christensen, and B. Wagner. 2d mapping of cluttered indoor environments by means of 3d perception. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 4, pages 4204–4209, 2004.

J. Wülfing, J. Hertzberg, K. Lingemann, A. Nüchter, S. Stiene, and T. Wiemann. Towards real time robot 6d localization in a polygonal indoor map based on 3d tof camera data. In *Symposium on Intelligent Autonomous Vehicles*, 2010.

X. Xiong and D. Huber. Using context to create semantic 3d models of indoor environments. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2010.

Z. Yang, C. Wu, and Y. Liu. Locating in fingerprint space: wireless indoor localization with little human intervention. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 269–280, 2012.

X. Yin, P. Wonka, and A. Razdan. Generating 3d Building Models from Architectural Drawings: A Survey. *IEEE Computer Graphics and Applications*, 29(1):20–30, 2009.

M. Youssef and A. Agrawala. The Horus WLAN location determination system. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, pages 205–218. ACM, 2005.

K. Yue, C. Hickerson, and R. Krishnamurti. Determining the interior layout of buildings describable by shape grammars. In *CAADRIA '08: Proceedings of the 13th International Conference on Computer Aided Design Research in Asia*, pages 117–127, 2008.

X. Yun, E. R. Bachmann, H. Moore, and J. Calusdian. Self-contained position tracking of human movement using small inertial/magnetic sensor modules. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 2526–2533. IEEE, 2007.

T. Y. Zhang and C. Y. Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3): 236–239, 1984.

Z. Zhang and L.-W. He. Whiteboard scanning and image enhancement. *Digital Signal Processing*, 17(2):414–432, March 2007.

Z. Zhou, M. Kavehrad, and P. Deng. Indoor positioning algorithm using light-emitting diode visible light communications. *Optical Engineering*, 51(8):085009–1, 2012.

D. Zuehlke. SmartFactory—Towards a factory-of-things. *Annual Reviews in Control*, 34(1):129–138, April 2010.

Appendix A: Affine and perspective transformation parameters estimation

Affine transformation

A linear transformation describing the mapping between two sets of 2D points is called affine transformation if it maintains collinearity, parallelism and distance ratios. It is defined by the following equation:

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

where X, Y and x, y, are the coordinates in the source and target system, respectively. c_{ij} are the matrix coefficients, while t_x, t_y depict translation parameters. Alternatively, it can be written in homogenous coordinates as follows, in order to compute the complete transformation using a single matrix operation:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} c_{00} & c_{01} & t_x \\ c_{10} & c_{11} & t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

The parameters can contain all needed superimposed operations. These are:

- two translations: $\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$
- one rotation: $\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} cos(\alpha) & sin(\alpha) \\ -sin(\alpha) & cos(\alpha) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$
- two scalings: $\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$
- two shears: $\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & m_x \\ m_y & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$

In order to compute the six transformation parameters from a set of homologous points, at least three point pairs have to be defined in the source and the target coordinate systems, which are not allowed to be collinear. This results in a linear system with a minimum of six equations:

$$\underline{A} \cdot \underline{x} = \underline{b} + \underline{v}$$

with

$$\underline{A}=\begin{bmatrix} x_0 & y_0 & 1 & 0 & 0 & 0 \\ x_1 & y_1 & 1 & 0 & 0 & 0 \\ x_2 & y_2 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & x_0 & y_0 & 1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \underline{x}=\begin{bmatrix} c_{00} \\ c_{01} \\ t_x \\ c_{10} \\ c_{11} \\ t_y \end{bmatrix}, \underline{b}=\begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ Y_0 \\ Y_1 \\ Y_2 \\ \vdots \end{bmatrix}$$

Parameter estimation in the case of an overdetermined system can be achieved using least squares adjustment following:

$$\underline{A}^T \cdot \underline{A} \cdot \underline{x} = \underline{A}^T \cdot \underline{b}$$

$$\underline{x} = (\underline{A}^T \cdot \underline{A})^{-1} \cdot \underline{A}^T \cdot \underline{b}$$

Perspective transformation

The perspective transformation enabling the removal of image distortions caused by an image capturing perspective other than perpendicular, is defined by the following equations:

$$X = \frac{c_{00}x+c_{01}y+c_{02}}{c_{20}x+c_{21}y+1}$$

$$Y = \frac{c_{10}x+c_{11}y+c_{12}}{c_{20}x+c_{21}y+1}$$

In the same way as the affine transformation, it can be written in homogenous coordinates:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} c_{00} & c_{01} & c_{02} \\ c_{10} & c_{11} & c_{12} \\ c_{20} & c_{21} & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

In order to compute the eight transformation parameters from a set of homologous points, at least four point pairs have to be defined in the source and the target coordinate systems, resulting in a linear system with a minimum of eight equations:

$$\underline{A} \cdot \underline{x} = \underline{b} + \underline{v}$$

with

$$\underline{\underline{A}} = \begin{bmatrix} x_0 & y_0 & 1 & 0 & 0 & 0 & -x_0X_0 & -y_0X_0 \\ x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X_1 & -y_1X_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2X_2 & -y_2X_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3X_3 & -y_3X_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & x_0 & y_0 & 1 & -x_0Y_0 & -y_0Y_0 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1Y_1 & -y_1Y_1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2Y_2 & -y_2Y_2 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3Y_3 & -y_3Y_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \underline{x} = \begin{bmatrix} c_{00} \\ c_{01} \\ c_{02} \\ c_{10} \\ c_{11} \\ c_{12} \\ c_{20} \\ c_{21} \end{bmatrix}, \underline{b} = \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \\ \vdots \\ Y_0 \\ Y_1 \\ Y_2 \\ Y_3 \\ \vdots \end{bmatrix}$$

Parameter estimation in the case of an overdetermined system can be achieved using least squares adjustment following:

$$\underline{x} = (\underline{\underline{A}}^T \cdot \underline{\underline{A}})^{-1} \cdot \underline{\underline{A}}^T \cdot \underline{b}$$

Appendix B: Foot-mounted MEMS IMU positioning

Stance phase detection

Section 2.2 already covered some of the fundamentals of indoor positioning using MEMS IMUs attached to a person's foot(s). This mounting location allows for a reset of the inevitable drift caused by double integration of small errors in the raw accelerations by detection of features in the data for which true known true values exist. In the case of an IMU mounted to a human's foot, a true value is found as the resting foot during the human gait cycle's stance phase. An overview over existing approaches for the detection of these phases without movement either using the accelerometers or the gyroscopes of an IMU can be found in Harle (2013).

In this thesis, the stance phase detection is realised by an analysis of the gyroscopes' measurements, i.e. detecting phases in the raw data where the norm of the three-dimensional angular rates vector

$$|\underline{g(t)}| = \left| \begin{bmatrix} g_x(t) \\ g_y(t) \\ g_z(t) \end{bmatrix} \right| = \sqrt{g_x(t)^2 + g_y(t)^2 + g_z(t)^2}$$

is less or equal an experimentally derived value[1] of one radians for at least 500 milliseconds:

$$|\underline{g(t)}| \leqq 1.$$

Drift correction

The drift correction bases on the fact that the velocities in all three axes are supposed to be zero during the detected stance phases[2]. Thus, the raw accelerations have to be integrated once in order to derive velocities. The raw accelerations, however, contain influences caused by earth's gravitation, which first have to be removed with the help of the gyroscopes measurements:

$$\underline{a_g(t)} = \underline{a_r(t)} + \underline{R(t)} \cdot \begin{bmatrix} 0 \\ 0 \\ -9.81 \frac{m}{s^2} \end{bmatrix}$$

[1] using a XSens MTi-G MEMS IMU

[2] this is the reason for the term zero-velocity update (ZUPT) which is used for this method if the zero-velocity phases are employed as updates for a Kalman filter

with $\underline{a_g(t)}, \underline{a_r(t)}$ being the corrected and raw accelerations, respectively, and $\underline{R(t)}$ being the rotation matrix constructed from the gyroscope measurements. The corrected accelerations are defined in the sensor's coordinate frame and have be transformed to the local tangential coordinate system by

$$\underline{a_c(t)} = \underline{R(t)^{-1}} \cdot \underline{a_g(t)}$$

For the actual drift correction, at first the corrected accelerations collected during the stance phase have to be integrated to obtain velocity values. This can be realised using the cumulative sum (for each axis separately):

$$v(t) = \int_{begin}^{end} a_c(t) dt \approx \frac{1}{T} \sum_{begin}^{end} a_c(t)$$

with $begin, end$ being the begin and the end of the stance phase, respectively, and T the sampling frequency. Subsequently, the offset of these velocity values is computed as their mean value:

$$\overline{v} = \frac{1}{t_{end} - t_{begin}} \sum_{begin}^{end} v(t_i)$$

As this value should equal to zero in a drift-free scenario, its difference from zero serves as an estimate for the drift which occurred during the swing phase before the current stance phase. Therefore, assuming a linear development of the drift during the swing phase, the drift-corrected velocities can be obtained by

$$v_c(t) = v(t) - \overline{v} \cdot \frac{t - t_{begin}}{\Delta t}$$

with t_{begin} being the begin of the swing phase and Δt the sampling interval.

Displacement vector calculation

Using the drift-corrected velocities, the final displacement vector describing the displacement of the sensor-equipped foot is computed by another integration:

$$p(t) = \int_{begin}^{end} v_c(t) dt \approx \frac{1}{T} \sum_{begin}^{end} v_c(t)$$

with $begin$ being the end of the last stance phase and end being the begin of the current stance phase.

Acknowledgements

Herewith, I would like to thank everybody who had an influence on the formation of this thesis. I am very grateful for the funding provided by the German Science Foundation (Deutsche Forschungsgemeinschaft, DFG) who provided the financing for most of my time at the Institute for Photogrammetry. This enabled me to work on the ComNSense project, together with our project partners at the Institute for Parallel and Distributed Systems (IPVS) at the University of Stuttgart, Patrick Baier, Damian Philipp, Dr. rer. nat. Frank Dürr, and Prof. Dr. rer. nat Kurt Rothermel. My special thanks goes to the committee members of my defence, Prof. Rothermel and Prof. Dr.-Ing. Nico Sneeuw, and especially Prof. Dr.-Ing. Dieter Fritsch who gave me the opportunity to work at the Institute for Photogrammetry and to compose this thesis.

This thesis would not have been possible without the help of my colleagues at the Institute for Photogrammetry, most importantly Dr.-Ing. Susanne Becker, Ali Mohammed Khosravani, Alessandro Cefalu, and apl. Prof. Dr.-Ing. Norbert Haala. I would like to thank Susanne Becker for the excellent co-work in projects, the collaboration in the grammar-based reconstruction, and the co-authoring of several papers published during the research for this thesis. My thanks goes to Ali Khosravani for the collaboration in low-cost point cloud based reconstruction and the co-authoring of papers, to Alessandro Cefalu for many fruitful discussions during the time we spent as office neighbours, and to Norbert Haala for his support, especially in the beginning phase of my work in indoor reconstruction and positioning.

Very warm thanks go to Sabina Kannegießer and Brenda Roggeveen who spared no pains while proofreading, correcting and improving this thesis in terms of the English language.

Most importantly, however, I would like to thank my parents and family with all my heart for all their support and encouragement during my studies as well as during the time spent for the making of this thesis.

Curriculum Vitae

Personal facts

Michael Sebastian Peter
Born 5th of August 1981 in Alzenau (Germany)

School education

1988 – 1992	Primary school in Krombach
1992 – 2001	Spessart-Gymnasium Alzenau
29th of June 2001	Abitur

University education

2002 – 2007	Diploma course Geodesy and Geoinformatics at the University of Stuttgart
August 13th 2007	Diploma degree at the University of Stuttgart

Profession

September 2007 - December 2014	Research assistant at the Institute for Photogrammetry of the University of Stuttgart