Refinement of Coarse Indoor Models using Position Traces and a Low-Cost Range Camera

Michael Peter, Ali Mohammad Khosravani, Dieter Fritsch
Institute for Photogrammetry
University of Stuttgart
Stuttgart, Germany
[firstname.lastname]@ifp.uni-stuttgart.de

Abstract—**In our previous work, we described the automatic reconstruction of coarse models of building interiors using a single photograph of an evacuation plan and an available model of the building's external shell. Additionally, we presented the extraction of the initial position and orientation for a foot-mounted MEMS IMU positioning system from the plan as well as the use of the coarse model for alignment and map-matching purposes. In this paper, we propose the geometric and semantic refinement of the coarse model employing an analysis of position traces delivered by such a system. While the addition of semantic attributes requires user interaction (like taking a photograph of a door plate to be analyzed by Optical Character Recognition), the update of the model's geometry (i.e. reconstruction of doors) is delivered automatically by analyzing position traces and model conjointly. The model's geometry is further refined using an automatic approach for the reconstruction of detailed CAD models of individual rooms, from the point clouds collected by the low-cost RGB-D sensor Microsoft Kinect. The user's track derived from the MEMS IMU positioning approach will be used for the registration and fusion of the detailed models to the coarse model.**

Keywords-reconstruction, refinement, IMU, RGB-D, alignment

I. INTRODUCTION

With the rise of GNSS, virtual globes and navigation systems, maps and city models evolved from specialized tools used by planners to broadly available features for visualization and navigation support (e.g. landmarks in car navigation systems). The ubiquity of a sufficiently precise positioning system, detailed maps and mobile internet connections enable everybody to navigate and use Location Based Services – in outdoor scenarios. However, most people spend most of their time in building interiors, whether it is for work or leisure.

This results in an augmented interest in indoor positioning and navigation methods, which can be seen in the fact that all major commercial map publishers (e.g. Google, Bing, Nokia) as well as smaller companies like Aisle411, PointInside and Micello produce indoor maps and advertise their availability as unique features of their maps. From the scientific point of view, this augmented interest is reflected in publications on indoor positioning methods as well as reconstruction methods for building interiors.

Traditional use cases for indoor models were special interests like interior architecture, building management and disaster simulation, whereas, in the context of indoor positioning, they are needed e.g. for the installation of infrastructure-based methods (i.e. the placement of beacons) or the support of infrastructure-less methods (using map-matching or similar approaches). In the case of indoor navigation, their primary uses are route calculations as well as their visualization. As will be shown in section II, the reconstruction of indoor models is still mostly a tedious and expensive task. Thus, building on our previous work which presented the reconstruction of coarse indoor models employing the reverse-engineering of existing evacuation plans, we present means for their geometric and semantic refinement using position traces and a low-cost range camera.

The remainder of this paper is structured as follows: Section II gives a short overview over existing work in the field of indoor reconstruction. Section III summarizes our previously published approaches for the reconstruction of building interiors, whose results will be refined using the methods presented in the sections IV (using position traces and user interaction) and V (using a low-cost range camera).

II. RELATED WORK

The reconstruction of building interiors is subject to research in a multitude of fields ranging from image processing and photogrammetry to robotics and point cloud processing. It can be split into data acquisition and the actual reconstruction, each of which pose their respective challenges. Regarding the data acquisition, these can be found as the registration respectively the absolute geo-referencing of different data sets and disadvantageous conditions for passive sensing techniques (e.g. images). The model reconstruction, however, suffers from the low accuracy of some measurement principles (conflicting with the high detail density in indoor environments) as well as occlusions of the floor and walls caused by furniture or other objects.

Apart from semi-automatic methods basing on panoramic images [1, 2], approaches employing a model-constrained analysis of single images in order to derive boundary representation models have been presented [3, 4]. Instead of directly deriving a boundary representation model, most publications, however, cover the acquisition of point clouds as an intermediate or end product. In this category, passive approaches using statistical analysis of single images [5] or stereo-based methods [6] for the generation of point clouds as well as video-based SLAM approaches [7, 8] have to be mentioned. Due to the aforementioned problems encountered with passive acquisition methods, the majority of related publications are employing active sensors combined with SLAM processing. In this context, most combinations of sensor and carrying system can be found: human-operated [9, 10] as well as ground- [11] and air-borne [12] robotic systems using LiDAR, sonar [13] and time-of-flight cameras [14]. With the availability of the Microsoft Kinect, however, many research groups started to concentrate on this low-cost active sensing system with an acceptable accuracy for most indoor applications. While [15] is building on the complete RGB and depth data delivered by the Kinect, [16] and [17] only employ the depth data for the co-registration. Needless to say that its first applications in robotics [18] and unmanned aerial vehicles [19] have been described.

As for the majority of the use cases described in section I boundary representation models are preferable or indispensable, their reconstruction from the collected point clouds is an essential research task. While some publications concentrate on the pre-processing segmentation of planar patches [20], most classify the detected planes into walls, floor and ceiling e.g. by semantic nets [21] or graph analysis [22]. Due to the regularity of man-made structures like especially building interiors, this semantic classification often builds on the Manhattan World constraints [23]. These constraints are the base of the plane sweeping reconstruction approach presented by [24], while other publications present less constrained methods working on the wall detection in point density histograms [25, 26].

In contrast to these indoor reconstruction methods, our previously presented approach builds on models reconstructed from photographed evacuation plans and thus lives in the context of methods for the reverse-engineering of existing plans. In this field, [27] presented a system interpreting floor plans hand-sketched with constraints concerning the paper and pens used. However, most approaches build on scanned CAD plans. For example, [28] presents low-level analysis steps for such plans like the segmentation to graphics and text as well as thick and thin lines, vectorization and the detection of arcs and loops. The also presented higher-level approaches use this extracted information to construct a complete 3D model. In [29] this approach was developed into a complete system for the analysis of scanned CAD plans. Additional, more recent work in this field can be found in [31, 32]. In contrast to the high resolution scans employed by these approaches, we use lower-resolution photographs taken under the pre-existing lighting conditions. Additionally, the layouts to be expected are more heterogeneous and parts of the plans may be occluded by symbols. Finally, the scaling and geo-referencing of the reconstructed model is carried out automatically.

III. PREVIOUS WORK

In previous publications, we presented approaches for the derivation of coarse indoor models from photographed evacuation plans of known design [32] as well as of arbitrary designs [33]. The most important processing steps are depicted in Fig. 1. Both approaches build on an image enhancement step using automatic white balancing and Wallis filtering [34] in the LAB color space and binary image computation using adaptive thresholding. In order to derive geo-referenced, metric models, the outer contour of the photographed plan is matched to the building's external shell taken e.g. from OpenStreetMap.

In order to provide complete models, symbol areas have to be detected and bridged. In the case of plans of well-known design, cross-correlation template matching is employed for their detection. The generalization to arbitrary plans builds on the color information in the plans, as symbols are represented by pure signal colors (red, green, yellow and blue) in contrast to a light background and dark ground plan lines. The symbol areas identified using the Color Structure Code segmentation method [35] are bridged by prolonging end node edges identified in the skeletonized cleaned binary image.

Analyzing the contours computed from the bridged binary image, they may be categorized into rooms and stair candidates by their thresholds concerning their maximum width and minimum length. These stair candidates can be grouped to staircases. The overall number of stairs in a staircase combined with a standard stair height for public buildings provides an approximate floor height. Thus, the 2D contours can be extruded to a coarse 3D model (see Fig. 2).

In addition to their use for the reconstruction of building interiors, the photographed plans may provide the location and orientation of the user photographing it. As has been shown [32], these can be used as initial values for a foot-mounted MEMS IMU positioning method. Still existing misalignment errors have successfully been corrected by detection of straight lines in the traces and their alignment to the nearest building axis (computed from the external shell).

As the most prominent error source, however, remains the initial position. This is caused by two factors: firstly, the accuracy of the position in the plan (as well as the surrounding parts of the ground plan) is reduced by the displacement which is part of the generalization and secondly, the distance of the user taking the photograph to the plan cannot be recovered (unless the plan's dimensions are known). This error could be overcome by the use of additional known coordinate updates or map-matching.

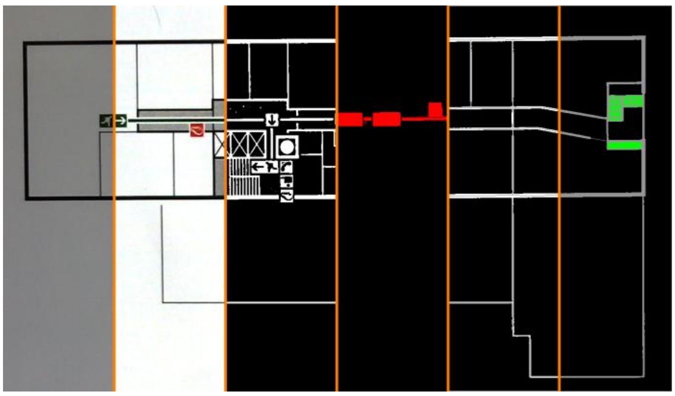

Fig. 1: Processing steps (from left to right): original image, enhancement, binarization, symbol detection, symbol bridging, stair detection (green)

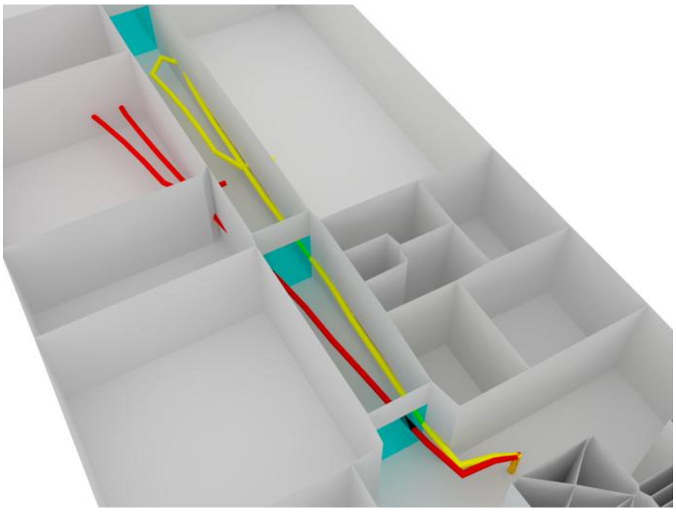

Fig. 2: User trace starting at the evacuation plan's position before (red) and after (yellow) alignment to building's principal axis

An efficient way to provide the geometric information for the update and refinement of the generated coarse models is to collect point clouds of the building interiors. In [36] we employed the RGB-D sensor Microsoft Kinect for this purpose. For the alignment of the collected point clouds we compared different approaches and analyzed their accuracy and flexibility in different scenarios [37]. In the first approach, the sensor poses were estimated using the Structure from Motion (SfM) approach employing point features detected and matched across corresponding RGB images. The accuracy of this approach depends on the amount and distribution of image features across the views. Therefore, this approach may not work properly in bad lighting conditions or in sparsely textured scenes.

To overcome this limitation, one may take the advantage of the 3D object space observations provided by the range sensor. In the second approach we used the KinectFusion software [16] to estimate the sensor pose purely using the geometric information. This software aligns the range images in real-time using the GPU implementation of a coarse-to-fine Iterative Closest Point (ICP) algorithm [38] and delivers a surface with low noise. This approach fails in the alignment of the range images along planar objects, as the sensor pose's 6 degrees of freedom cannot be fixed by the ICP algorithm.

Although the abovementioned approaches may solve the problem of point cloud alignment in many indoor scenes, still there are some scenarios where neither enough image features for a successful SfM can be found, nor can the ICP approach estimate the sensor pose. In such cases, external information (e.g. provided by an available Building Information Model or fused sensors like MEMS IMU) might solve the problem to some extent. In [39] we showed that the user's position track provided by the MEMS IMU positioning method can support the point clouds alignment in some scenes like hallways (poor texture and insufficient 3D information). The accuracy of this alignment approach is directly related to the precision of positioning per single step, which was estimated to be less than 10cm in our study case. Moreover, the information extracted from the available coarse model provided us with some geometric constraints to improve the alignment accuracy.

IV. Model Refinement using Position Traces

As visible in Fig. 2, apart from the detected stairs, the only additional information in the coarse model generated from the photographed sample evacuation plan are the doors which were identified as walls crossed by evacuation routes. The wall locations are sufficient for the model's use as support information for an infrastructure-less positioning method (see section IV.B). For the model's use in a navigation scenario, however, the door positions as well as semantic information like a room number are needed. In the following, approaches employing the user traces produced by the aforementioned foot-mounted MEMS IMU positioning method are used for the derivation of this information.

A. *Semantic annotation*

Due to the positioning method being available, users of the system are enabled to interactively collect any geo-referenced information and thus model the indoor environment in an OpenStreetMap-like fashion. In OpenStreetMap, such information is often collected using geo-tagged photos or audio files, which will be transferred to the map in a manual post-processing step. In building interiors, the locations of doors and windows are of interest as well as information about the rooms present in the coarse model (room number, people associated to the room, opening hours etc.).

Fig. 3: Semantic annotation (room number) by OCR analysis of a geo-referenced photograph of a door plate (overlay on Google Maps, Imagery ©2013 AeroWest, Landeshauptstadt Stuttgart, Map data ©2013 GeoBasis-DE/BKG)

In addition to a fully manual post-processing step, photographed door plates can be converted to text using state-of-the-art Optical Character Recognition (OCR) software (see Fig. 3). Analyzing the resulting text, room numbers can be extracted as well as people assigned to the room by comparing the text to a names database. The associated room can be detected employing the viewing direction of the user during the image capturing (extracted from the position traces). As additional information, the position of the door plate can be used as a good approximation for the location of a door.

B. Map matching and automatic door reconstruction

In addition to the aforementioned alignment, the availability of the coarse model enables the use of map-matching as a further correction method for the improvement of the position traces. Our simple map-matching approach employs constraints for the interaction between the user's movement vectors and the model which are inverse to the constraints used for the automatic reconstruction of door openings. The reconstruction of a door is only possible if the position trace hits a wall in the model a) in a reasonable angle and b) in a location where the wall features enough space for the reconstruction of a door. Using a standard door size (see [40]), the latter constraint is straightforward. For the angle threshold, this door size is combined with an average human shoulder width. This results in an angle interval of 40° to 140° in which an average human can pass a door comfortably and in which a trace crossing a wall will cause the reconstruction of a door. All other intersections between trace and wall and the translation needed for their correction contribute to a vector correcting the trace as a whole. A sample result for both map-matching and door reconstruction is depicted in Fig. 4.

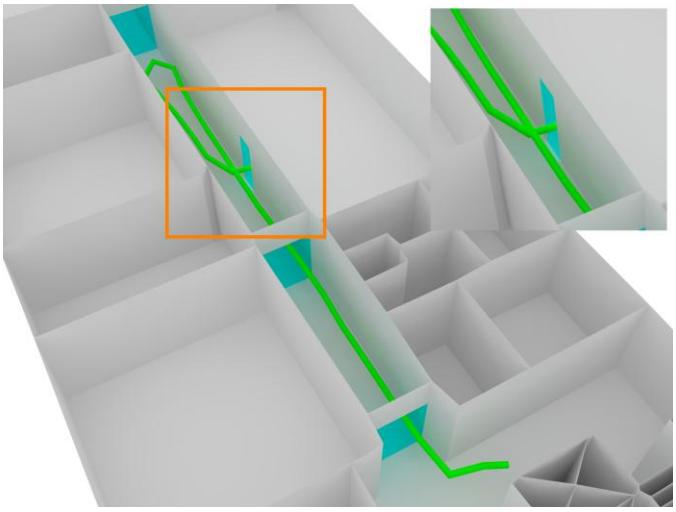

Fig. 4: Door automatically reconstructed from map-matched user trace

V. Model Update and Refinement using a low-cost Range Camera

The reconstructed coarse model can be further refined by adding the details missing in the evacuation plans due to the generalization process or recent changes in the building interiors (e.g. new walls and cupboards). For the geometry data acquisition, Microsoft Kinect is employed to actively sense the indoor scenes. This affordable and accessible sensor system captures the range images in real-time with few programming efforts. After the alignment of the collected point clouds, the CAD models of rooms and hallways are automatically reconstructed with a higher level of detail. The detailed models are then fused with the coarse model using the position traces derived by the positioning method described in section IV. The details are described in the following parts.

A. Reconstruction of individual entities

In this study case, the point clouds of the rooms are collected and aligned separately based on the SfM approach employing the point features detected and matched in the corresponding RGB images. The scale information is retrieved from the 3D coordinates of the image features, having the link between the RGB and range data provided by the stereo calibration of the RGB and IR cameras.

For the hallway scenario, as the amount of the point features is not sufficient for a robust estimation of the sensor pose, we use the position traces described in section IV. For the collection and alignment of the point clouds in this method, the user walks through the hallway and collects the MEMS IMU data, while automatically capturing the scene point clouds at the track points by a Kinect system. At each track point, the coordinates are used for the computation of the approximate sensor position, considering a constant shift between the MEMS IMU and Kinect. The approximate orientation in the horizontal plane is computed, assuming the sensor is oriented toward the direction of the next track point. The sensor tilt is estimated by the analysis of the normal vectors in an iterative process, assuming the ground points' normal vectors comprise smaller angles with the vertical axis. The estimated orientations are further improved by fulfilling some geometric constraints, e.g. parallelism of walls in the coarse model and the point clouds. More details about this approach can be found in [39].

To reconstruct the CAD models of the individual entities, we first reduce the problem from 3D to 2D space. This simplifies the algebraic relationships and the topological analysis (which is necessary for a consistent and robust reconstruction), while preserving the 2D structure of the building interiors. The 2D CAD models are reconstructed by estimation of the lines representing the walls in 3D and applying some topological corrections. The models can be reconverted to 3D by a simple extrusion, having the height of the rooms. The correct height of the rooms which is considered as an update for the coarse model can be derived by analyzing the height histogram of the point clouds. The following sections describe the procedure in more detail.

Pre-processing of the point clouds

To reduce the reconstruction task from 3D to 2D space, the point cloud has to be projected onto the horizontal plane after the compensation of the existing tilt of the point cloud with respect to this plane. To estimate the tilt, again, the orientations of the normal vectors are analyzed, clustering the points into groups of walls and floor/ceiling. After the compensation of the tilt and removing the points corresponding to the floor and ceiling, the furniture inside a room can be removed by filtering the points having heights less than a typical threshold (e.g. 1.5m). The final result of this step is the leveled point cloud of the inner shell of the room. Although some parts of the point cloud might be filtered out during the floor/ceiling and furniture removal process, the remainder can still deliver the information about the 2D structure of the walls. Fig. 5 shows the results of this process for an exemplary room.

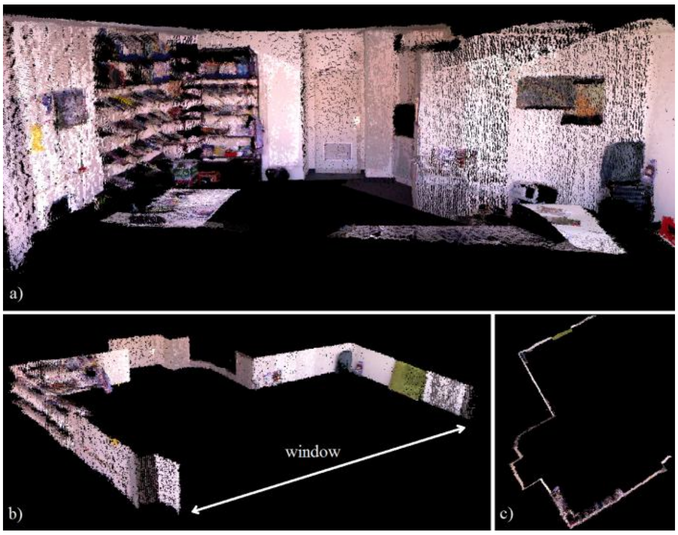

Fig. 5: Collected point cloud (a); removing the ceiling, floor and furniture by height filtering (b); projection of the points to the horizontal plane (c).

Estimation of the lines representing the walls in 2D

The resulting point cloud is additionally smoothed and converted to a 2D grayscale ortho projected image by counting the number of the points in each horizontal grid cell (Fig. 6-a). To estimate the lines in the projected image, the image first has to be binarized by setting a threshold on the gray values. The thresholding may remove the remaining horizontal surfaces and small objects while keeping the main vertical objects like walls.

In practice, due to the noise of the range data and errors in the alignment of the point clouds and tilt removal, the projections of the walls are represented by some shapes which are not necessarily line segments (depending on the grid size). Therefore, using morphological closing followed by a skeletonization process, the shapes are firstly thinned to 1 pixel width structures (Fig. 6-b and -c).

The skeletonized shapes are then generalized to straight lines using the Hough transformation. As it can be seen in Fig. 7-a, each wall segment might be approximated by several Hough line segments. Therefore, they are first clustered into the groups of individual walls, and then averaged within the groups to estimate an equivalent line segment per wall segment. Clustering of the line segments is performed hierarchically, first by their orientation and then by their distance to the centroid of the point cloud (Fig. 7-b). The thresholds for the angular and distance clustering are chosen regarding the noise of the point clouds and the level of detail required for the modeling step.

After the clustering process, for assigning a single line segment to each wall, the parameters of the resulting line segments are averaged within their groups (Fig. 7-c). The weights in this averaging correspond to the lengths of the segments.

Fig. 6: Grayscale ortho projected image (a); binarization and morphological closing (b); skeletonized image (c)

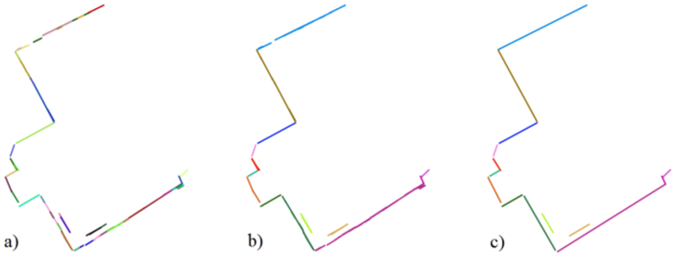

Fig. 7: Estimated Hough line segments (a); clustering the line segments (groups are distinguished with colors) (b); averaging the line segments within the groups (c)

Topological corrections

Considering the noise of the point clouds, imperfect point clouds alignment, occlusions in the point cloud and errors in the line estimation process, the resulting line segments do not fulfill a topologically correct geometry. Therefore, the following correction steps are applied to the resulting line segments to solve the topological issues.

In the first step, the orientations of the averaged line segments are refined by imposing a geometric constraint enforcing parallelism and rectangularity at the detection of small angular differences. For this purpose, the line orientations are analyzed and clustered to find the main two perpendicular directions. The main directions however may need to be slightly corrected to make a difference of 90 degrees. The orientations of the line segments are then compared to the two main directions, and are corrected if they make a difference of less than a given threshold (Fig. 8-a).

In the second step, the intersections of the line segments are analyzed. The end points of the resulting line segments do not necessarily meet each other at the intersection points (walls junctures). This issue might be solved by extension or trimming of the line segments within a specific threshold. The extension or trimming is performed first in the original direction and alternatively in the perpendicular direction, if no intersection is found. However, as it can be seen in Fig. 8-b and -c, the extension process may generate invalid line segments which will be detected and removed in the next step.

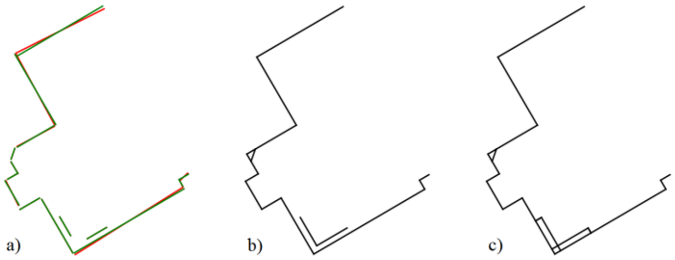

Fig. 8: Before (red) and after (green) parallelism and rectangularity constraining (a); extension and trimming in the original direction (b); extension and trimming in the perpendicular direction (c)

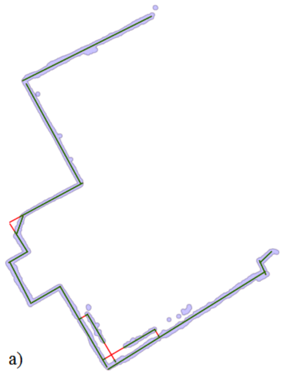

Fig. 9: Overlap with the point cloud data (a); marking as invalid (red), if the line segment is made by an extension and has less than 50% overlap (b)

In the third step, invalid line segments are detected by analyzing the overlap of the line extensions with the point cloud data. In other words, a line segment is marked as invalid, if it is made by an extension, and at the same time, it has insufficient (e.g. less than 50%) overlap with the point cloud data within a buffer (Fig. 9). For the correctness of the overlap analysis, it is necessary to consider the overlaps before imposing the parallelism and rectangularity constraining step. This means that the corrections made by the constraining step have to be temporarily applied in the opposite way.

As visible in Fig. 5, a part of the exemplary room corresponds to a window with open curtains, for which no range data from Kinect is available. To derive a closed polygon representing the room interiors, one may connect the free end points as the last correction step in simple cases. However, this solution may not be valid in a more general case, e.g. having windows in two perpendicular walls. To overcome this issue, information extracted from the coarse model will be used to correctly and consistently reconstruct the gaps, using the following fusion process.

B. *Fusion of detailed models and coarse model*

For the refinement of the coarse model, the detailed models of individual rooms (resulting from the previous step) have to be registered and merged into the 2D coarse model.

The initial position and orientation of the detailed models with respect to the coarse model can be inferred having the position traces derived from the described MEMS IMU positioning method. For this purpose, the user has to walk from the position of the evacuation plan into the room whose point cloud has to be acquired by the Kinect system. Then the centroids of the detailed model and the corresponding room in the coarse model will be coincided. Assuming the user starts capturing the point clouds while the sensor is oriented toward the door position, the initial orientation of the detailed model with respect to the coarse model is computed (Fig. 10-a).

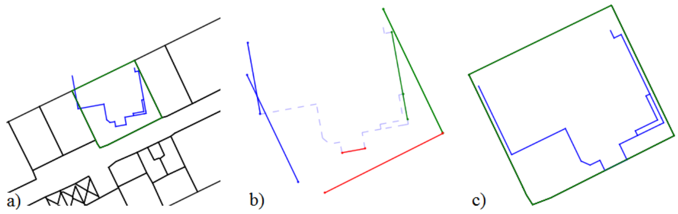

Fig. 10: Initial registration to the coarse model (a); corresponding lines in the detailed and coarse models (b); best fitting of the detailed coarse models (c)

The initial registration is further refined by the optimal fitting of the corresponding line segments in the detailed and the coarse models. To find the line correspondences, first the line segments corresponding to the outer shell of the detailed model have to be derived by analyzing the convex hull of the model. The correspondences in the coarse model are then found by searching for the closest line segments having closest orientations (Fig. 10-b). The optimal registration of the detailed model of the exemplary room to the coarse model is depicted in Fig. 10-c. As the door thickness in comparison to the wall thickness is usually negligible, the coincident of the line segments corresponding to the door are considered as a constraint in the registration process. Such line segments are already known in both models from the intersection of the user's track and the coarse model.

The wall thickness which is not considered in the coarse model is detectable after the registration process, by analyzing the registration residuals (distance between the corresponding line segments after the registration). In other words, residuals similar to a default value within a given threshold are considered as the actual wall thickness of the room. However, this value is affected by the accuracy of the detailed model which will be investigated later within the paper. The information about the wall thickness enables a correct fusion, and at the same time, comparison between the detailed and the coarse models for finding the changes in the building interiors and updates to the coarse model.

After the registration process, the remaining gaps in the detailed model can be reconstructed by the fusion of the detailed model and the parallel offset of the coarse model (to consider the wall thickness). For this purpose, the line segments are first converted to graph edges (Fig. 11-a), and those containing a vertex of degree 1 will be extended simultaneously in the same and perpendicular directions to find their first intersection with the offset shell (the edge with red color in Fig. 11-a). Assuming the gap parts are smaller than 50% of the model, edges generating the shortest path from the two most distant vertices of degree 1 will be merged to the detailed model to reconstruct the gaps (Fig. 11-b). Fig. 11-c shows the fusion of the detailed models of some exemplary rooms with the coarse model.

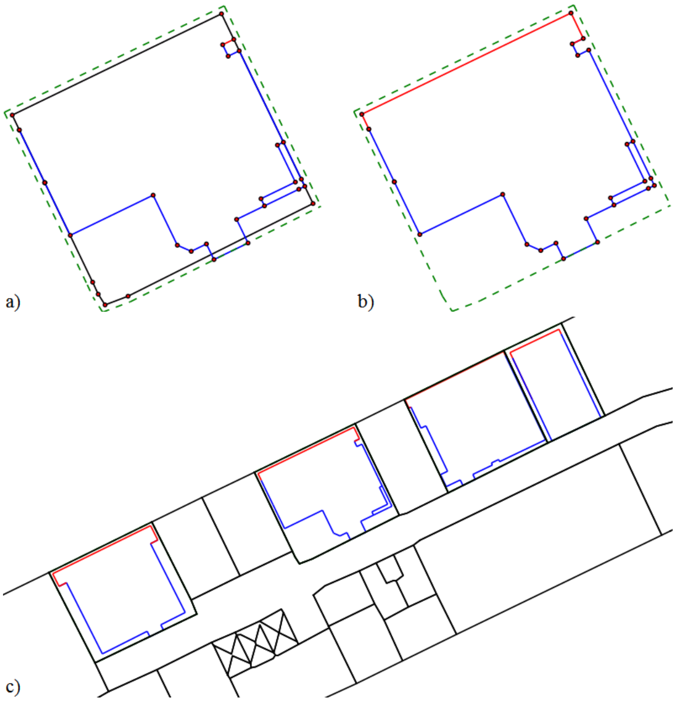

Fig. 11: Fusion of the detailed model to the coarse model. a): detailed model (blue), offset shell (black) and extensions connecting the vertices of degree 1 in the detailed model to the offset shell (red); b): detailed model (blue) and reconstructed gaps (red); c): fusion of the detailed models of some exemplary rooms and the coarse model

C. *Quality assessment of the reconstructed models*

Regarding the described procedure for the reconstruction of detailed models and their fusion with the coarse model, several error sources may affect the quality of such models. In more detail, one may expect an average noise of around 3cm for the collected point clouds by the Kinect system (at around 3m distance to the object). However, this value is increased by the errors in the alignment process of the point clouds. In contrast, the averaging concept which is the basis of the walls estimation together with the described topological corrections may reduce the effect of the errors on the final results.

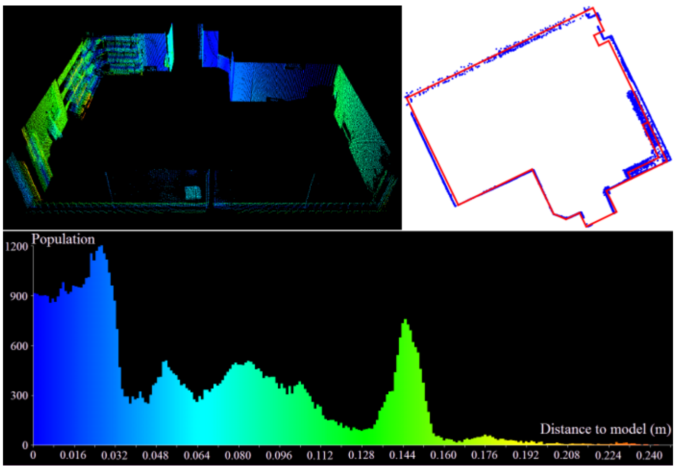

Fig. 12: Comparison of the laser point cloud to the reconstructed 3D model of the exemplary room: maximum distance: 0.259m, mean distance: 0.048m, standard deviation: 0.051m

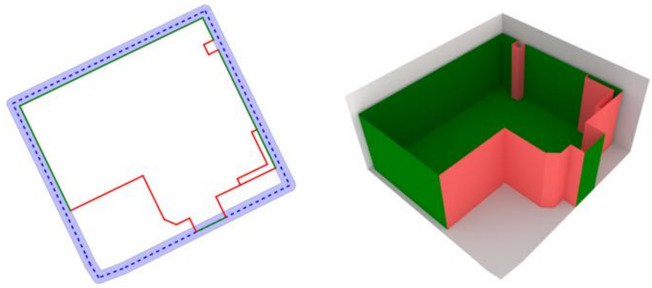

Fig. 13: Walls outside of the buffer (red) are considered as update to the coarse model

To assess the accuracy of the reconstruction process, the reconstructed 3D model of the exemplary room is compared to the point cloud collected by a terrestrial laser scanner as the ground truth (Fig. 12). As this TLS point cloud is registered in a different local coordinate system, the point cloud is registered to the 3D model using the ICP algorithm. The comparison shows a mean difference of 0.048m with a standard deviation of 0.051m.

The precision of the coarse models generated from the evacuation plans depends on the scale and drawing precision of the plans, and also the resolution of the camera capturing them. In this study case, the overall precision of the coarse model is around 0.10-0.15m. Therefore, the accuracy of the detailed model meets the requirements of this application.

The fusion and comparison of the detailed to the coarse models not only verifies the accuracy of the coarse model, but also enables the detection of changes and details missing in the coarse model. For the comparison, a buffer equivalent to the wall thickness (around 15cm in this example) is generated around the coarse model. Objects outside of the buffer are considered as changes to the building interiors (Fig. 13 and Fig. 14).

VI. Conclusions and Future Work

In this paper, we presented approaches for the geometric and semantic refinement and update of the coarse models generated based on the approaches presented in our previous work. In order to reach this goal, we employed a relative indoor positioning approach complemented by initial values extracted from the evacuation plan and supported by the coarse model. This enables the users to geo-reference semantic features in a way similar to the modeling approach used in OpenStreetMap. Moreover, it has been shown that a conjoint analysis of model and position traces may be used to refine the model's geometry by automatically reconstructing door openings.

Additionally, we showed that the geometry of the coarse models can be further refined by the fusion of the detailed model of individual rooms automatically reconstructed from the point clouds collected by the low-cost range camera Microsoft Kinect. The presented approach for the reconstruction of the detailed models delivers results with acceptable accuracy and robustness regarding the available noise in the Kinect point clouds. Imposing geometric constraints and topological corrections assures the correctness and consistency of the reconstruction process and the fusion with the coarse model.

As future work, the model shall be further enhanced by texture mapping using the RGB images captured by the Kinect system.

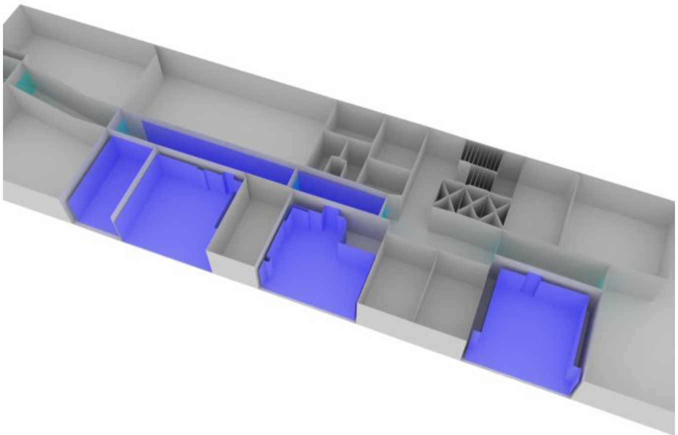

Fig. 14: Refinement of the coarse model by the fusion of detailed models of some exemplary rooms and a hallway (front walls are removed for visibility purposes)

ACKNOWLEDGMENT

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), grant FR 823/25-1 / RO 1086/17-1 and a grant from the Landesgraduiertenförderung Baden-Württemberg.

REFERENCES

[1] D. Farin, W. Effelsberg, and others, "Floor-plan reconstruction from panoramic images," in *Proceedings of the 15th international conference on Multimedia*, 2007, p. 826.

[2] T. K. Dang and M. Worring, "A semi-interactive panorama based 3D reconstruction framework for indoor scenes," *Comput. Vis. Image Underst.*, vol. 115, no. 11, pp. 1516–1524, 2011.

[3] E. Delage, H. Lee, and A. Y. 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*, 2006, vol. 2, pp. 2418 – 2428.

[4] 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*, 2009, pp. 140 –147.

[5] A. Saxena, S. H. Chung, and A. Y. Ng, "3-D Depth Reconstruction from a Single Still Image," *Int. J. Comput. Vis.*, vol. 76, no. 1, pp. 53–69, Jan. 2008.

[6] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, "Reconstructing Building Interiors from Images," in *Proceedings of ICCV 2009*, 2009.

[7] 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*, 2009, pp. 83–86.

[8] 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*, 2010, pp. 1498–1505.

[9] T. Liu, M. Carlberg, G. Chen, J. Chen, J. Kua, and A. Zakhor, "Indoor localization and visualization using a human-operated backpack system," in *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2010, pp. 1–10.

[10] M. Bosse, R. Zlot, and P. Flick, "Zebedee: Design of a Spring-Mounted 3-D Range Sensor with Application to Mobile Mapping," *Ieee Trans. Robot.*, vol. 28, no. 5, pp. 1104–1119, 2012.

[11] A. Fietz, S. M. Jackisch, B. A. Visel, and D. Fritsch, "Automated 2D Measuring of Interiors using a Mobile Platform," in *Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2010)*, Funchal, Madeira/Portugal, 2010.

[12] 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 *Association for Unmanned Vehicle Systems International (AUVSI) Symposium*, 2009.

[13] J. D. Tardós, J. Neira, P. M. Newman, and J. J. Leonard, "Robust Mapping and Localization in Indoor Environments Using Sonar Data," *Int. J. Robot. Res.*, vol. 21, no. 4, pp. 311–330, Apr. 2002.

[14] 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*, 2009, pp. 1673 –1678.

[15] 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.

[16] 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)*, 2011, pp. 127 –136.

[17] H. Roth and M. Vona, "Moving volume kinectfusion," in *British Machine Vision Conf.(BMVC),(Surrey, UK)*, 2012.

[18] C. Audras, A. Comport, M. Meilland, and P. Rives, "Real-time dense appearance-based SLAM for RGB-D sensors," in *Australasian Conf. on Robotics and Automation*, 2011.

[19] 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," *Int. J. Robot. Res.*, vol. 31, no. 11, pp. 1320–1343, 2012.

[20] D. Hähnel, W. Burgard, and S. Thrun, "Learning compact 3D models of indoor and outdoor environments with a mobile robot," *Robot. Auton. Syst.*, vol. 44, no. 1, pp. 15–27, Jul. 2003.

[21] A. Nüchter, H. Surmann, K. Lingemann, and J. Hertzberg, "Semantic scene analysis of scanned 3D indoor environments," in *Proc of the VMV Conference*, 2003.

[22] P. Jenke, B. Huhle, and W. Strasser, "Statistical Reconstruction of Indoor Scenes," in *Proceedings of WSCG 2009*, Plzen-Bory, Czech Republic, 2009.

[23] J. M. Coughlan and A. L. Yuille, "Manhattan world: Compass direction from a single image by bayesian inference," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, vol. 2, pp. 941–947.

[24] A. Budroni and J. Boehm, "Automated 3D Reconstruction of Interiors from Point Clouds," *Int. J. Arch. Comput.*, vol. 8, pp. 55–73, Jan. 2010.

[25] 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*, 2010, vol. 2.

[26] X. Xiong and D. Huber, "Using context to create semantic 3D models of indoor environments," in *Proc. BMVC*, 2010.

[27] 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, 1996*, 1996, vol. 3, pp. 747 –751 vol.3.

[28] 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*, 1997, pp. 347–351.

[29] P. Dosch, K. Tombre, C. Ah-Soon, and G. Masini, "A complete system for the analysis of architectural drawings," *Int. J. Doc. Anal. Recognit.*, vol. 3, no. 2, pp. 102–116, Dec. 2000.

[30] S. Or, K.-H. Wong, Y. Yu, M. M. Chang, and H. Kong, "Highly automatic approach to architectural floor plan image understanding and model generation," *Pattern Recognit.*, 2005.

[31] T. Lu, C. L. Tai, L. Bao, F. Su, and S. Cai, "3D Reconstruction of detailed buildings from architectural drawings," *Comput.-Aided Des. Appl.*, vol. 2, no. 1–4, pp. 527–536, 2005.

[32] M. Peter, N. Haala, and D. Fritsch, "Using Photographed Evacuation Plans to Support MEMS IMU Navigation," in *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN2011)*, Guimaraes, Portugal, 2011.

[33] M. Peter, D. Fritsch, B. Schäfer, A. Kleusberg, J. Á. Bitsch Link, and K. Wehrle, "Versatile Geo-referenced Maps for Indoor Navigation of Pedestrians," in *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation*, 2012.

[34] R. Wallis, "An approach to the space variant restoration and enhancement of images," in *Proceedings of the Symposium on Current Mathematical Problems in Image Science*, Naval Postgraduate School, Monterey CA, USA, 1976.

[35] "Color Structure Code (CSC)." [Online]. Available: http://www.uni-koblenz-landau.de/koblenz/fb4/institute/icv/agpriese/research/ColorImgSeg/download/csc. [Accessed: 19-Jun-2012].

[36] N. Haala, D. Fritsch, M. Peter, and A. M. Khosravani, "Pedestrian Mobile Mapping System for Indoor Environments based on MEMS IMU and Range Camera," *Arch. Photogramm. Cartogr. Remote Sens.*, vol. 22, pp. 159–172, 2012.

[37] A. M. Khosravani, M. Lingenfelder, K. Wenzel, and D. Fritsch, "Co-registration of Kinect point clouds based on image and object space observations," in *Proceedings of the LC3D Workshop*, Berlin, 2012.

[38] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *Ieee Trans. Pattern Anal. Mach. Intell.*, pp. 239–256, 1992.

[39] A. M. Khosravani, M. Peter, and D. Fritsch, "Alignment of Range Image Data Based on MEMS IMU and 3D Coarse Models Derived from Evacuation Plans," in *Proceedings of the 2013 SPIE Videometrics*, Munich, 2013.

[40] E. Neufert, P. Neufert, B. Baiche, and N. Walliman, *Architects' Data*, 3rd ed. Wiley-Blackwell, 2002.