Alignment of Range Image Data Based on MEMS IMU and Coarse 3D Models Derived from Evacuation Plans

Ali M. Khosravani*[a], Michael Peter[a], Dieter Fritsch[a]
[a] Institute for Photogrammetry (ifp), University of Stuttgart, Germany
Geschwister-Scholl-Straße 24D, D-70174 Stuttgart

ABSTRACT

Within the paper, we present an approach for the alignment of point clouds collected by the RGB-D sensor *Microsoft Kinect*, using a MEMS IMU and a coarse 3D model derived from a photographed evacuation plan. In this approach, the alignment of the point clouds is based on the sensor pose, which is computed from the analysis of the user's track, normal vectors of the ground points, and the information extracted from the coarse 3D model. The user's positions are derived from a foot mounted MEMS IMU, based on zero velocity updates, and also the information extracted from a coarse 3D model. We will then estimate the accuracy of point cloud alignment using this approach, and discuss about the applications of this method in indoor modeling of buildings.

Keywords: Multisensor, IMU, Point Cloud, Registration, Evacuation Plan, Modeling

1. INTRODUCTION

This paper deals with the alignment of range images using a MEMS IMU and coarse 3D models derived from an evacuation plan, in an indoor application. The range sensor used in this study is Microsoft Kinect, which is a low-cost multisensor system. This system consists of an RGB and a monochrome IR CMOS sensor, both working in 30Hz with VGA resolution (640 by 480 pixels), and an IR laser projector. Kinect measurements are disparity values, resulting from a comparison between the reference and collected IR patterns. Having disparity values, range images are then computed in a limited range of 0.7m to 6m. Stereo calibration of the RGB and IR sensors creates a link between the color image space and the 3D space.

In a previous work, we have aligned the Kinect range images, making use of color image and object space observations [1]. In other words, such range images were aligned first by the estimation of the sensor's pose by applying the structure from motion (SfM) method on color images. Then, as a complementary method, the range images were aligned using their geometric information, i.e. iterative closest point (ICP) approach. The shortcomings of these methods were in scenarios where neither enough image features for a successful SfM can be found nor the ICP approach could fix the sensor's pose 6 degrees of freedom. This is especially the case when dealing with corridors in indoor scenarios.

Therefore, an extension to the previous work is the integration of positioning solutions (e.g. MEMS IMU) to the system, to support the sensor's pose estimation. In recent years, there has been an increasing interest in the use of such sensors for indoor navigation. IMUs are often used as foot-mounted systems and are combined with algorithms like zero velocity updates (ZUPT). Moreover, due to the existence of drift errors or inaccurate initial values, other methods have to be employed, e.g. map matching algorithms using available indoor models. In our previous work [2] we used a given external building shell and photographed evacuation plans as support methods for foot mounted MEMS IMU navigation.

In this study, we use the same positioning method to support the sensor's pose estimation. In more details, we can acquire the sensor's position from the described positioning method, and the initial orientation from the user's track analysis in the horizontal plane. The orientations can be further improved by analysis of the normal vectors, and the information provided by the coarse 3D model derived from the evacuation plan [2]. The estimated poses are then directly used for the alignment of the range images collected at different points. This enables our future works, which will be

* ali.khosravani@ifp.uni-stuttgart.de

texturing, updating and refinement of the coarse 3D models. Such refinement includes adding windows and features which were newly constructed or were neglected in the evacuation plans due to the generalization process.

2. INDOOR RECONSTRUCTION USING PHOTOGRAPHED EVACUATION PLANS

In our previous work [2], the general feasibility of the reconstruction of indoor environments from photographed evacuation plans has been presented. Within this work, image pre-processing, binarization, symbol detection and bridging as well as geo-referencing were identified as the main steps of this image processing task.

The approach presented there is working on the well-known layout of our building's plans using cross-correlation template matching in order to detect symbols. However, the generalization of the approach to arbitrary plan layouts relies on the colors found in evacuation plans. Thus, after pre-processing and binarizing the image, we use color segmentation (i.e. the Color Structure Code [3]) for the detection of the most common signal colors which represent the sought-after symbol areas. In order to distinguish rooms from stairs, detect door openings and to deliver the final model in metric coordinates, the transformation parameters between image coordinates and world coordinates have to be computed by matching the floor plan's outer contour to an available external model (e.g. from OpenStreetMap). Due to the differences concerning scale and level-of-detail, this matching step is carried out on versions of both polygons which were generalized by a 2D variation of the generalization approach presented in [4]. After skeletonizing and vectorizing the binary image, the symbol areas are bridged by prolonging the edges ending in end nodes until a metric length threshold is exceeded, not taking the distance traveled on symbol areas into account and stopping at structures in the cleaned binary image. The final 2D model is derived as the contours detected in the binary image containing the bridging edges (see figure 1).

A further analysis of the room polygons of this 2D model in terms of maximum width and minimum length delivers a set of stair candidates. These stair candidates can be grouped into staircases and their number combined with a standard stair height for public buildings (see [5]) provides an approximate floor height. This floor height is used to extrude the room polygons to a 3D model (see figure 1).

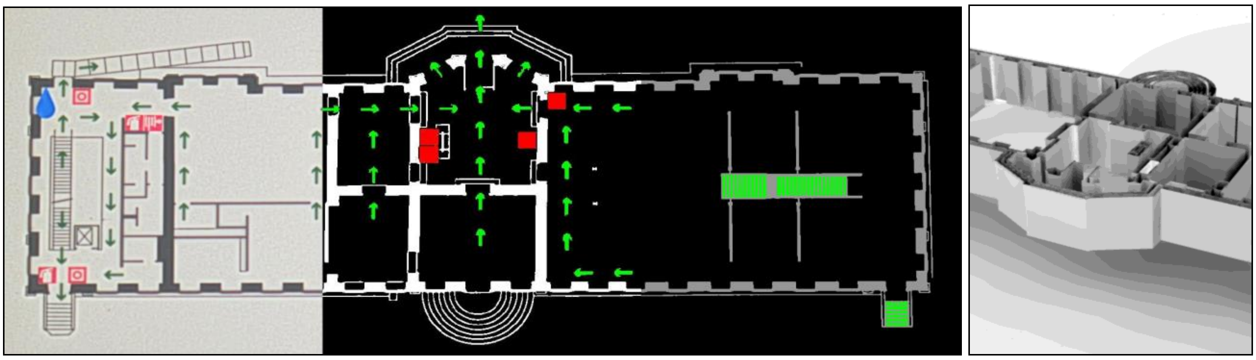

Figure 1. From left to right: pre-processed image; wall structures (white) and detected symbol areas (colored); symbol bridging edges and detected stair candidates (green); reconstructed 3D model

3. MODEL REFINEMENT EMPLOYING USER TRACKS

For the refinement of the coarse models, the user's initial position and orientation are extracted from the same photograph and are put to use in a foot-mounted MEMS IMU positioning approach as presented in [2]. The initial position for this relative positioning approach can be extracted from the photographed plan either by template matching or as the position of a symbol unique in shape or color. For the transformation to world coordinates the transformation parameters computed during the reconstruction process are used. The initial orientation is acquired as the viewing angle of the user towards the plan (using the approach presented by [6]) combined with the fact that the plan is oriented according to the local environment. The positioning is carried out using the well-known zero velocity update approach [7], supported by the detection of traveled straight lines and their alignment to one of the external building shell’s main axes (see also [2]).

The user's positions derived by employing this positioning approach may be used to geo-reference semantic information acquired by user interaction. As an example, we have investigated the derivation of room numbers and people assigned to a room from photographed door plates using Optical Character Recognition (OCR). Moreover, they may also be used for an update of the model's geometry by a fully automated derivation of door positions. This employs the fact that the user is not able to pass through a door in arbitrary angles. Thus, if an average person's position track hits a wall in the model in an angle between 40° and 140°, a door will be reconstructed. Implicitly, this constraint provides us with a simple map-matching solution, correcting the track whenever a wall is hit at angles different from that (see figure 2).

Finally, we used this positioning approach to geo-reference the range images in the following sections, to align the point clouds with the 3D model.

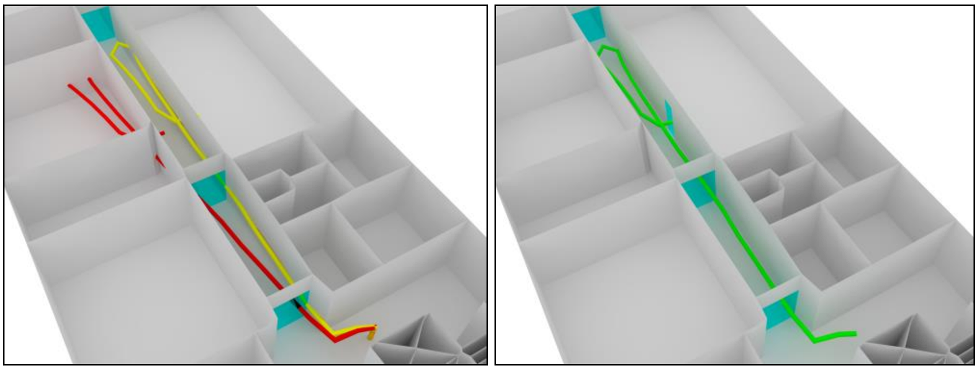

Figure 2. LHS: Coarse model, uncorrected (red) and aligned (yellow) tracks; RHS: automatic door reconstruction by user track analysis, track corrected by map-matching (green)

4. ALIGNMENT OF RANGE IMAGES

In this section, we present an approach for the alignment of the Kinect range images, using the mentioned positioning approach and the information included in the coarse 3D model. For this purpose, the user holds a Kinect sensor while walking through a corridor, and capturing the MEMS IMU data, simultaneously.

The case of this study is a corridor to show the performance of this approach in a scenario, in which the approaches presented in our previous work [1] are not successful. This is due to the fact that this approach neither uses visual features (for the relative orientation of the corresponding color images [8]), nor geometrical features (for sequential ICP approach [9]).

The alignment procedure includes the pre-processing of the individual range images, and then applying the transformation parameters derived from the analysis of user's track and the coarse 3D model, which is discussed in more details in the following sections.

4.1 Pre-processing of the range images

In the pre-processing step, the normal vectors of the range images are analyzed to find the ground points, in order to level the point clouds in the horizontal plane, and also to correct the heights of the point clouds. Points are segmented as ground points if the angle between their normal vector and the vertical axis is less than a threshold (e.g. 30°). This step will then correct the tilt of the sensor while capturing data. The point cloud of the walls can also be segmented using a similar analysis, which is necessary for the next steps.

4.2 Extraction of 3D rigid transformation parameters

In section 3, the user's track was successfully integrated to the coarse model, and was geo-referenced by matching the floor plan's outer contour to an available external model. In this step, the transformation parameters for the geo-referencing of the point clouds are computed. The coordinates of the track points are used to directly compute the translations. The orientations in the horizontal plane can be approximately computed at each point, assuming the sensor is oriented according to the direction of the next tracking point.

4.3 Improvements of the results using the coarse 3D model

The orientation of the range images can be further refined by setting some constrains. An example of such constraining in this scenario is parallelism of the corresponding walls in the range images and in the 3D model. This solution can be reduced into two dimensions by projecting the walls onto the horizontal plane. In this case, the corresponding line segments (representing the walls) in both 2D image spaces shall be parallel.

The trace of walls of the range images in the horizontal plane contains noise, which is inappropriate for direct fitting of straight lines. For this reason, the range images and the resulted 2D images shall be further processed. Therefore, firstly the outliers from the walls data are removed in order to avoid extra noise in the 2D projected image. Afterwards, by morphological closing (dilation followed by an erosion), the trace of the walls will be converted to closed structures by filling the holes and removing remaining noise. The traces are then skeletonized for a following line fitting using the Hough transformation (figure 3).

In the next step, a similar projected image of walls from the 3D model is required. This image is already available, which is the processed evacuation plan before extrusion to a 3D model. The image can be directly used to extract line segments using the Hough transformation.

As the range images are already approximately aligned with the coarse 3D model using the user's track analysis, the corresponding walls in the two image spaces are assumed to be the nearest lines in the overlay of the two images. Figure 4 depicts an example of line matching using this method. Having found the orientation of the corresponding lines in the two image spaces, the alignment of each range image can be further improved. Figure 5 shows the final results of the alignment after this improvement. As depicted in this figure (top view), the walls point clouds are now parallel to the walls of the 3D model. However, it shows that the width of the corridor in the 3D model is larger than reality (measured from the range images), which can be considered as an update to the 3D model.

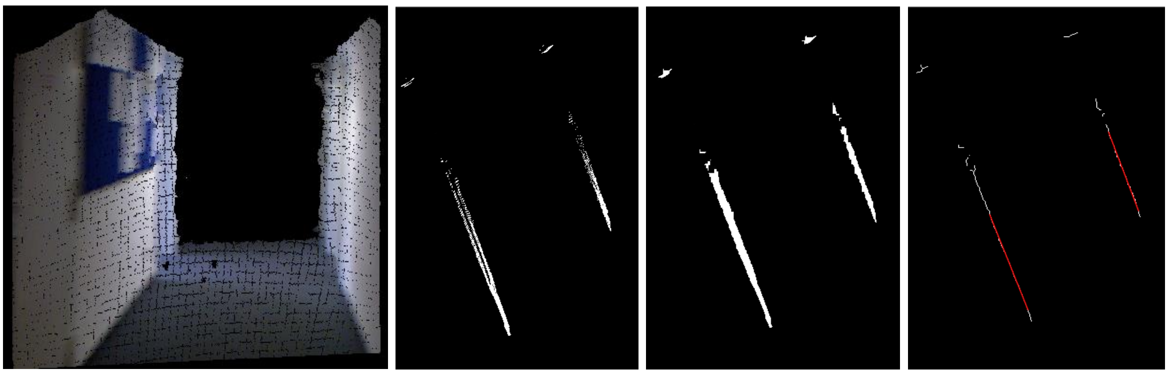

Figure 3 . From left to right: a single point cloud; projection of walls onto the horizontal plane; morphological closing; and skeletonization (white) followed by a Hough transformation (red)

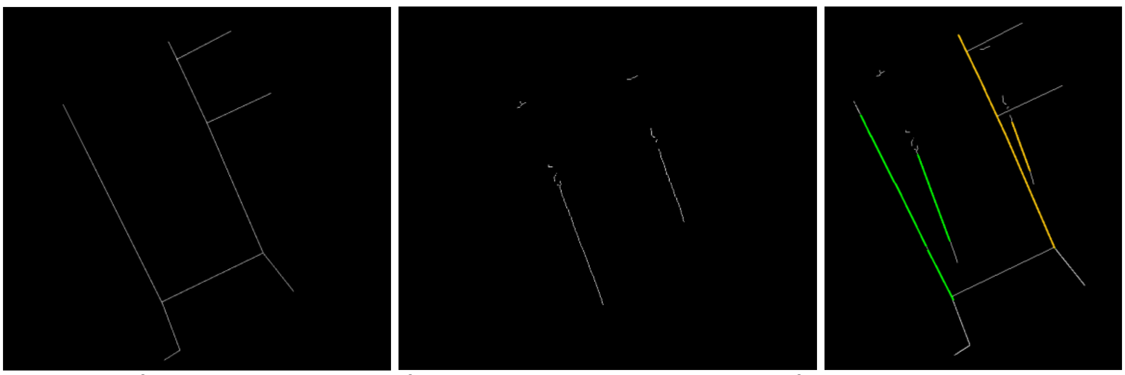

Figure 4 . From left to right: projected walls of the coarse 3D model; projected walls of a single range image; and the overlay of the two images (similar colors identify the corresponding walls)

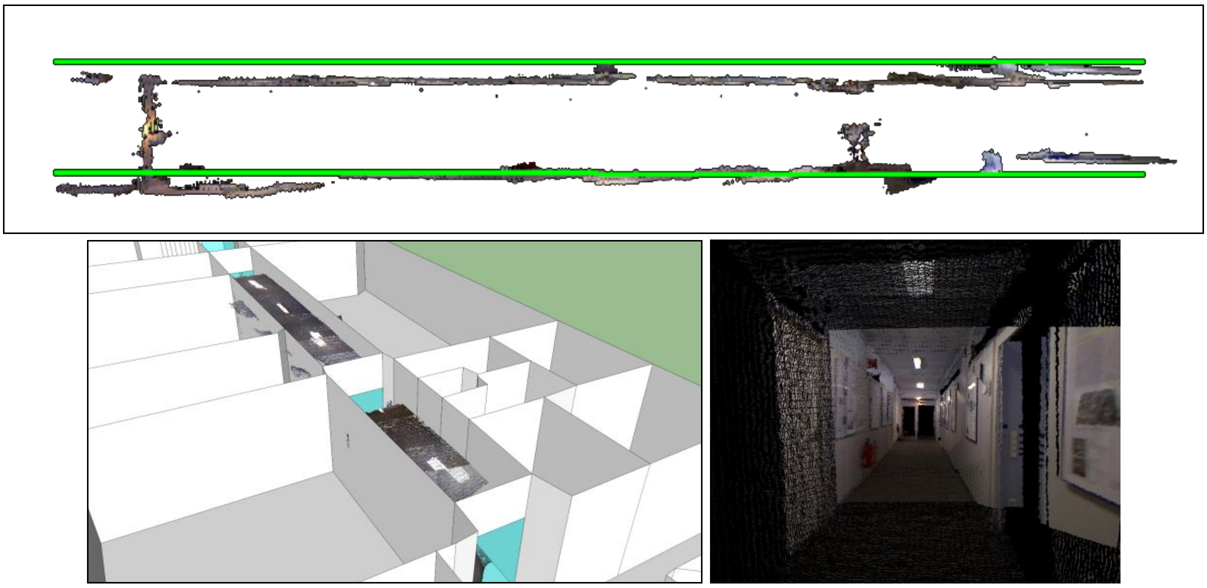

Figure 5 . Final alignment of the range images (top and inside views). The top figure shows the alignment of the walls range data with the walls of the 3D model (green) (other structures than the hallway walls were removed for visibility reasons).

Of course, the described method for the refinement of the orientations is only valid for Manhattan-World scenarios, in which the line matching can be carried out by finding the closest line. For more sophisticated scenarios (e.g. curved walls and corridors) other methods shall be employed for finding and matching the corresponding curves.

5. ACCURACY ANALYSIS

The accuracy of the alignment of the range images using the described method is directly related to the accuracy of the positioning method. To have a rough view about this accuracy, the 3D coordinates of some identical features in the consecutive point clouds were measured and compared. The differences of the coordinates are considered as the internal accuracy of the positioning, and therefore the range image alignment (figure 6). This contains the error of measuring the coordinates of the features (due to the noise of the range images), and errors due to changing the configuration of the hand-held range sensor relative to the foot-mounted MEMS IMU at different measurement epochs.

The estimated accuracies are almost less than 10cm, which seem to be rational regarding the mentioned sources of errors and also the accuracy which one may expect from the evacuation plans. Therefore, this approach can be considered as an approximate solution for the pose estimation, which might be sometimes the only available solution for the alignment of the range images.

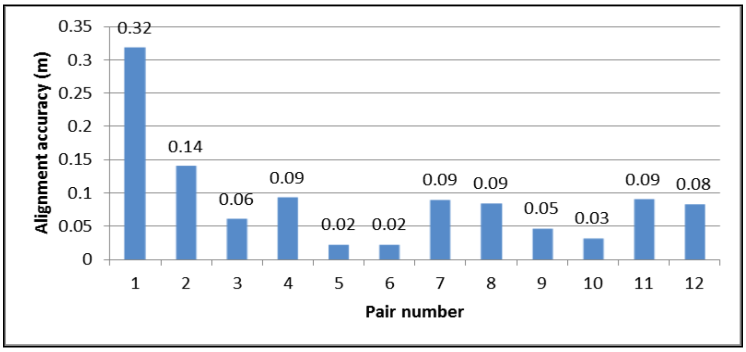

Figure 6 . Estimated internal accuracy of positioning and alignment of the range images

6. CONCLUSIONS AND FUTURE WORKS

Within the paper, an automatic approach for the alignment of range images was presented. This approach analyzes the user's track derived from a foot mounted MEMS IMU and information extracted from a coarse 3D model for an approximate alignment of the range images. The alignment was further improved by constraining the orientation of the range images using the 3D model. This approach was successfully tested for a corridor dataset, where other approaches like relative orientation of the corresponding color images or sequential ICP face difficulties.

The alignment of the range images and the coarse 3D model is a necessary step for the update and refinement of the coarse 3D models. Using the described approach, we aim at an approximate alignment of the point clouds of individual rooms with the coarse 3D model. However, the point clouds of the individual rooms can be aligned separately by other approaches. The alignment of the single rooms point clouds with the 3D model can be further improved by the ICP approach.

Therefore, a future extension to this work will be merging all the described steps into an automatic pipeline for the update and refinement of available coarse 3D models, including reconstruction and update of geometric features, texturing from the Kinect's color images implicitly oriented.

7. ACKNOWLEDGEMENTS

Parts of this work are supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), Grant FR 823/25-1 / RO 1086/17-1.

REFERNCES

[1] Khosravani, A., Lingenfelder, M., Wenzel, K. and Fritsch, D., "Co-Registration of Kinect Point Clouds Based on Image and Object Space Observations," Proc. LC3D Workshop, Berlin (2012).

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

[3] Priese, L., "Color Structure Code," Institute of Computational Visualistics, University Koblenz-Landau. <http://www.uni-koblenz-landau.de/koblenz/fb4/institute/icv/agpriese/research/ColorImgSeg/download/csc> (19 June 2012).

[4] Kada, M., Peter, M. and Filippovska, Y., "3D Building Generalization under Ground Plan Preserving Quality Constraints," Proc. 6th International Symposium on Spatial Data Quality, St. John's, Canada (2009).

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

[6] Zhang, Z., and He, L.-W., "Whiteboard Scanning and Image Enhancement," Digital Signal Processing, 17(2), 414-432 (2007).

[7] Foxlin, E., "Pedestrian Tracking with Shoe-Mounted Inertial Sensors," IEEE Computer Graphics and Applications, 25(6), 38-46 (2005).

[8] Henry, P., Krainin, M., Herbst, E., Renand, X. and Fox, D., "RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments," 12th International Symposium on Experimental Robotics (ISER), 20, (2010).

[9] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A. and Fitzgibbon, A., "KinectFusion: Real-Time 3D Reconstruction and Interaction using a Moving Depth Camera," Proc. UIST '11 the 24th annual ACM symposium on User Interface Software and Technology, 559-568 (2011).