

Automated extraction of image coordinates for Optical Indoor Positioning

V.E. Willert¹ and V.V. Willert² and S. Gering² and S. Raß² and J. Etzel¹

¹ TU Darmstadt, Institute of Geodesy, Petersenstr. 13, DE-64287 Darmstadt. Email: willert@geod.tu-darmstadt.de

² TU Darmstadt, Institute of Automatic Control, Landgraf-Georg-Str. 4, DE-64287 Darmstadt. Email: vwillert@rtr.tu-darmstadt.de

Abstract—In this paper, we focus on the accuracy of optical indoor positioning based on pictures taken by a cell phone camera. We restrict ourselves to automated relative pose estimation given only one image including the projection of a known object with at least three known reference points with known absolute position. To infer the relative pose from the image coordinates of the reference points, we first have to detect and classify an object, afterwards localize the image coordinates, and finally apply spatial resection. We show that if an object is correctly classified, then the quality of the positioning heavily depends on the accuracy of the localization of the image coordinates and on the choice of the spatial resection algorithm. To this end, we compare three different spatial resection algorithms and present two combinations of object classification and image coordinate localization techniques using doors as known objects. Statistical evaluations are provided to judge the different methods in terms of accuracy and robustness.

Index Terms—optical indoor positioning, camera phone

I. INTRODUCTION

The field of Indoor Local Positioning Systems (ILPS) has widened up enormously during the last years. Apart from optical systems all other positioning systems depend on external signals to receive enough information for the positioning process. The positioning process of such systems can be divided into approximate and precise methods.

Approximate methods (e.g. based on WLAN) either use the existent infrastructure (for example WLAN access points) or the infrastructure has to be set up with low cost references (e.g. RFID tags) before the positioning process. Using approximate methods, a positioning accuracy of some meter can be achieved.

Precise methods (e.g. using UWB technology) enable more accurate results that are within one decimeter. But the more accurate the result of the positioning method, the higher is the effort for setting up the infrastructure. Besides, the costs of such precise positioning systems are quite high.

Optical indoor positioning systems are widely spread in the field of robotic navigation [1]. Also in the field of Augmented Reality, optical systems have already been presented in [2] or [3]. They capture the environment by taking photos and then process their position by evaluating the images and linking the image informations (for example certain landmarks that occur in the image) with the world coordinate system. The effort of setting up the infrastructure in an optical system matches the effort of positioning systems based on approximate methods. The expenses of an optical system are also comparable to

those of other low cost ILPS systems. Just one static image taken by a digital camera and including an object referred to the world coordinate system (e.g. a door, that could be signed with a code) can be sufficient to calculate the camera position. The image coordinates (u_i, v_i) of the object e.g. the door's corners as well as site-specific information of the object are extracted by image processing algorithms. The world coordinates (X_i, Y_i, Z_i) of the object can be identified from the code or by the object itself. With at least three corresponding points, the position (P_0) and the orientation of the camera can be calculated by spatial resection.

If the accuracy of such an optical system would exceed the accuracy of ILPS systems based on approximate methods, optical systems would be a competitive alternative.

II. ACCURACY ESTIMATION

The accuracy of an optical ILPS system depends on the geometrical configuration as well as on the uncertainties of the measurements and the reference stations. In [4] different approaches of spatial resection have been investigated according to their numerical stability of the positioning result for different geometric configurations. In the case of accurate image- and world coordinates (u_i, v_i, X_i, Y_i, Z_i) the investigated algorithms offer quite stable solutions. But in practical use deviations always occur in the measurements ($\tilde{u}_i, \tilde{v}_i$) as well as in the object points ($\tilde{X}_i, \tilde{Y}_i, \tilde{Z}_i$). Therefore, the influence of all parameters have been investigated separately for different geometrical configurations in [5]. It has been shown that especially the accuracy of the image coordinates has a deep impact on the accuracy of the position P_0 .

We provide a statistical evaluation generating a data set of corresponding image- and object points simulated by given positions for the corners of a rectangular planar object as well as for the camera positions. The positions of the camera were located on a regular grid lying 5 m away from the planar object and are located sideways up to 1 m. Adding Gaussian noise to the image coordinate data the position has been calculated with different approaches for solving the spatial resection and the results have been compared to the ground truth positions. Three approaches for solving the spatial resection (Grunert [6], Killian [7] and Rohrberg [8]) have been investigated according to their robustness against deviations in the measurements. The approaches in [6] and [8] use three corresponding points to calculate different solutions for P_0 and afterwards use a

fourth point to get the unique solution whereas in [7] the fourth corresponding point is included in the calculation from the beginning assuming a planar object.

In Fig. 1 we show the impact of noisy image coordinates on the positioning result. The mean error of position (*EOP*) increases almost linearly in the case of Killian in contrast to the 3-point-algorithms of Grunert and Rohrberg. So the Killian approach seems to be more robust against the configuration in combination with noisy measurements. In general the 4-point-algorithm achieves the best results according to its mean *EOP*.

But even using the 4 point-algorithm of [7] the solution worsens strongly with increasing noise on the image points. Image coordinates with zero mean Gaussian noise and a standard deviation of 30 px in an 3264 px × 2448 px image lead to a mean error of position (*EOP*) of 0.3 m. The image coordinates must be extracted with an accuracy of at least 15 px to get results in the position with an accuracy of ≤ 1 dm.

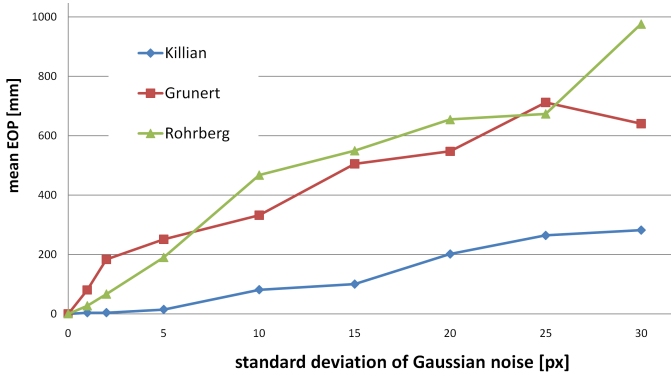


Fig. 1. Mean EOP of spatial resection solutions for noisy data.

Therefore, different possibilities for extracting the image coordinates of certain characteristics have been investigated and compared according to their accuracy and robustness. In this case, doors served as significant objects. They will be the central object of detection in the presented investigations.

III. CLASSIFICATION METHODS

We present two different classification methods for finding known objects in an image. The first one tries to detect visual targets featuring a prominent contour that includes a visual code that identifies the object and its position. The second method also tries to detect a candidate contour that is characteristic for the known object but afterwards classifies the image region assigned to the contour by comparing them to reference images from an image data base including images of all known objects. Here, different doors are chosen as objects with known coordinates of some reference points. The process of finding a door in an image, classifying what kind of door it is, and finding the coordinates of the reference points can be divided into four parts: a) Generation of a reference database, b) detection, c) classification, and d) extraction of the image coordinates.

A. Classification using fiducial markers

a) *Generation of the fiducial marker database* Fiducial markers are visual targets including a visual code that stores information referring to the corresponding world coordinates of the edges of the door that is equipped with this target. For each known door a fiducial marker has to be created with a unique code. All the marker codes are stored in a database.

b) *Detection of a marker* For the detection of a visual target a characteristic feature is needed that can easily be found in the image using state-of-the-art image processing techniques. Here, the visual target is a binary image characterized by a white circle on a black background surrounding a white visual code, see Fig. 2 a).

To find the marker, the image is binarized and closed objects $\mathcal{O}_c$ are localized using morphological operations [9] (combinations of opening and closing). The area $A(\mathcal{O}_c)$ and the circumference $U(\mathcal{O}_c)$ of all closed forms can be determined evaluating their pixels. To distinguish the circular fiducial marker from other closed objects, the roundness c has to be analyzed for all objects. The roundness is a dimensionless value. We assume a closed object to be a marker, if the roundness falls below a certain threshold t . The roundness is defined by the form of the object

$$c = 4\pi \cdot \frac{A(\mathcal{O}_c)}{U(\mathcal{O}_c)^2} < t, \quad t \geq 1. \quad (1)$$

For a circle, c gets minimal with a minimum of $c = 1$, see [9]. Calculating c and the area of all closed objects, the fiducial marker can be detected from the image.

c) *Classification via decoding the marker* To get the coded information, the detected marker first has to be normalized using an affine transformation. After that, the information can be decoded, see Fig. 2 b). The code exists of 15 segments representing a 15 bit code. The pixel of all marker elements with the smallest y -coordinate defines the beginning of the marker code. The radius of the marker in Fig. 2 c) is defined by the beginning and the centre of the code. Starting from this point, each 24 degrees a marker is set that defines one of the code segments, see Fig. 2 d). Retrieving the values of every segment (0 or 1), the 15 bit marker can be decoded. The resulting number can then be synchronized with the database to receive all necessary informations of the door.

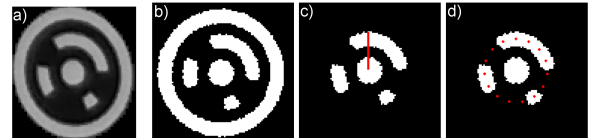


Fig. 2. Transforming and decoding the fiducial marker.

d) *Extracting the image coordinates of the reference points* First, we apply morphological operations (combinations of opening and closing) to extract the contour of the closed object that surrounds the fiducial marker. The result is assumed to be an approximate segmentation of the door, see Fig. 3 a). Then, applying a *Hough Transformation* [9], the lines closest

to the contour of the door segment are extracted, see Fig. 3 b). The intersections of the extracted contour lines (red crosses) present the image coordinates of the projected known corners of the door.

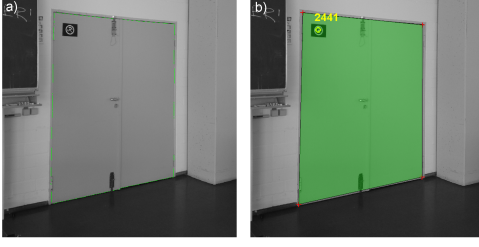


Fig. 3. a) Contour of a closed object, b) Door detection result.

B. Classification using a reference image database

If there are no fiducial markers provided for detection and classification of the doors, the images of the known doors alone have to be used instead. Having a look at the variability in relative positioning with respect to a door, the appearance of the objects in the resulting images heavily varies in size and viewpoint. Fig. 4 shows the variability of one door taken from different viewpoints. Besides the problem of strong variability, the question of what the unique features of one specific door are that differentiates it from all the other doors poses another problem.



Fig. 4. Different views of the same door

a) *Generation of the reference image database* Pictures of all doors in the environment have to be taken before the first positioning process can be started. These reference images are stored in a database together with the image and object coordinates of the labeled corners of the doors.

b) *Detection of a door* The detection process is a combination of 1) object contour features and 2) object surface features that characterizes a door best to detect good candidate doors for classification. The contour features for a door are lines in the image that are extracted using the *Canny-Edge-Detection* [9] and *Hough Transformation* [9]. The surface features for the door are homogeneous gray value regions extracted by a *quadtree decomposition* [9]. The criterion to split a leaf of the quadtree is a threshold on the difference between the maximum and minimum gray value within each square (leaf). Fig. 5 a) shows the resulting surfaces of such a quadtree decomposition (white dashed contour). Since we know that projected doors always hold a quadrangular contour, we fit a quadrangular contour $\mathcal{C}_H$ to the quadtree contour $\mathcal{C}_Q$ by minimizing the squared difference

$$\mathcal{C}_H^* = \underset{\mathcal{C}_H}{\operatorname{argmin}} (A_Q(\mathcal{C}_Q) - A_H(\mathcal{C}_H))^2 \quad (2)$$

of the surface areas A_Q and A_H that are bounded by the contours $\mathcal{C}_Q$ and $\mathcal{C}_H$ with respect to $\mathcal{C}_H$. The different values for $A_H(\mathcal{C}_H)$ result from all possible quadrangular contours $\mathcal{C}_H$ that can be constructed out of the Hough lines nearby the quadtree contour. Fig. 5 b) shows the result $\mathcal{C}_H^*$ of this quadrangle extraction process (white quadrangle).

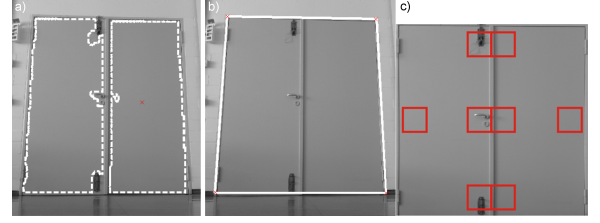


Fig. 5. a) Contour of quadtree segmentation, b) Quadrangle detection result, c) Normalized image with ROIs.

c) *Classification of a door* First, the image of every detected quadrangle (door candidate) will be transformed and reduced to a detail image that only shows a normalized door. Here, normalization means all angles of the door are 90 degrees and all edges have the same length, which is shown in Fig. 5 c). Then, for every quadrangle several *regions of interest* (ROIs) can be defined (see red boxes in Fig. 5 c)), where specific door features are expected to appear, e.g. a door handle. This leads to a pre-classification into three different subclasses of door types, namely two-wing-doors with or without door bolts and one-wing-doors. After that, the pre-classified candidate doors are compared to all reference images of the database that belong to the same subclass. For comparison, each reference image $\mathbf{I}_R$ is perspectively warped $\mathcal{W}(\mathbf{I}_R)$ such that the contours of the door within the reference image exactly overlay the quadrangular candidate door of the original (unnormalized) image $\mathbf{I}_C$ including a candidate door. Thereby, we simulate a reference image taken from the same spatial position as the image to be classified. To decide which reference door corresponds best to the candidate door, we minimize the following error measure with respect to the reference image $\mathbf{I}_R$

$$\mathbf{I}_R^* = \underset{\mathbf{I}_R}{\operatorname{argmin}} \sqrt{(1 - \rho_{R,C})^2 + \left\| \frac{\mathbf{m}_C - \mathbf{m}_R}{\mathbf{m}_R} \right\|^2}, \quad (3)$$

where $\rho_{R,C}$ is the correlation coefficient [9] between $\mathbf{I}_C$ and $\mathcal{W}(\mathbf{I}_R)$ and the vectors $\mathbf{m}_C$ and $\mathbf{m}_R$ include the Hu-moments [10] of $\mathbf{I}_C$ and $\mathcal{W}(\mathbf{I}_R)$.

d) *Extracting the image coordinates of the reference points* The image coordinates have already been extracted during the detection process and are given by the intersections of the Hough lines that form a quadrangle closest to the quadtree contour. Fig. 5 b) shows one result (red crosses).

IV. RESULTS AND COMPARISON

Both methods depend on a certain database that has to be produced a priori and stores the information of the object coordinates of the doors. One drawback of the marker based method is the additional effort of tagging the doors with

certain fiducial markers, leading to a modified environment. Additionally, the code has to be defined beforehand and the number of known objects is limited depending on the used code. Using reference images of the doors, these images have to be taken and the edge points of the door in these images have to be extracted before the positioning process. So in either case, the effort for creating the database is more or less similar.

To get information about the robustness, we first evaluated 71 images using fiducial markers and 78 images using reference images. In the first case, all images included a door. In the second case, 6 images did not include a door. The confusion matrices [11] of the binary classification problem (door/ no door) are shown in Table I and II.

Using visual tags for classification, the fiducial markers were detected 56 times by its roundness (1), that means 79% of all markers were detected and all of the detected ones were decoded correctly. The detection fails in images that have been shot too far away from the door (≥ 8 m).

TABLE I
CONFUSION MATRIX USING FIDUCIAL MARKER

	assumed positive	assumed negative
positive	56 true positives	15 false negatives
negative	0 false positives	0 true negatives

In case of 30 reference images, the detection step extracted 84 door candidates out of 78 images including 72 doors applying the minimization (2). From these candidates, 66 were truly doors, so the detection step missed 6 doors. 63 doors were detected correctly which corresponds to a detection rate of 87.5%. All of these correct detections were assigned to the correct reference image applying the minimization (3).

For both methods the classification performance was reduced drastically or even failed completely if the doors were partly occluded or the doors were opened.

TABLE II
CONFUSION MATRIX USING REFERENCE IMAGES

	assumed positive	assumed negative
positive	63 true positives	3 false negatives
negative	6 false positives	12 true negatives

Secondly, we investigated the accuracy of the automated extraction of the image coordinates. We compared the automated extracted points with the manually gripped edges of the doors. In Table III we point out the results according to their accuracy. The visual target method leads to better results. Nevertheless, we have to mention that we used different datasets for each method and the dataset for the visual target method was chosen with much less variability in camera position, because the read out of the fiducial marker does not work for large affine transformations, since it is not rotational invariant so far (the starting point of the code is assumed to be on top of the fiducial marker). The deviation of both methods is 1-2 px at minimum. However, using only natural characteristics,

TABLE III
ACCURACY OF EXTRACTED IMAGE COORDINATES

	class.	visual tag	class.	ref. image
	[height]	[width]	[height]	[width]
min. dev. [px]	0	0	1	0
max. dev. [px]	22	63	269	248
av. dev. [px]	5	4	56	40
stand.dev. [px]	9	26	62	48

deviations appear up to 269 px whereas the deviations reach up to 63 px taking fiducial markers into account. The reason for the strong outliers in case of reference images are some false contour extractions during the detection phase.

V. CONCLUSION

In this paper, we pointed out the need of robust and accurate extraction of image coordinates in optical indoor positioning. It has been shown that especially the image coordinate measurements have a big influence on the accuracy of positioning.

Hence, two methods for automated extraction of image coordinates of the edges of doors have been investigated especially according to their robustness and accuracy.

The analysis in section IV shows that it is possible to achieve a robust positioning with state of the art image processing techniques. To achieve a positioning accuracy of 1 dm at most, the image coordinates have to be detected with an accuracy of at least 15 px in an image with a resolution of 8 Mpx, which could not be achieved for all classified doors up to now. So the attempt is combining certain steps of the two presented methods to wipe out the disadvantages of the separate methods to improve the accuracy.

REFERENCES

- [1] E. Jones and S. Soatto, "Visual-Inertial Navigation, Mapping and Localization: A Scalable Real-Time Causal Approach," *International Journal of Robotics Research*, vol. 30, no.4, pp. 407–430, 2011.
- [2] H. Hile and G. Boriello, "Positioning and Orientation in Indoor Environments Using Camera Phones," *IEEE Computer Graphics and Applications*, vol. 8, pp. 32–39, 2008.
- [3] A. Mulloni, D. Wagner, and D. Schmalstieg, "Indoor Positioning and Navigation with Camera Phones," *Pervasive Computing*, vol. 8, pp. 22–31, 2009.
- [4] R. M. Haralick, C. Lee, K. Ottenberg, and M. Nolle, "Analysis and Solutions of the Three Point Perspective Pose Estimation Problem," *International Journal of Computer Vision*, vol. 13/3, pp. 331–356, 1994.
- [5] V. E. Willert, "Optical Indoor Positioning using a Camera Phone," in *IPIN, Abstract Volume*, 2010.
- [6] A. J. Grunert, "Das Pothenotische Problem in erweiterter Gestalt über seine Anwendungen in der Geodäsie," *Grunerts Archiv für Mathematik und Physik*, vol. 1, pp. 238–248, 1841.
- [7] K. Killian, "Über das Rückwärtseinschneiden im Raum," *ÖZfV*, vol. 4, pp. 97–104, 1955.
- [8] K. Rohrberg, "Geschlossene Lösung für den räumlichen Rückwärtsschnitt mit minimalen Objektinformationen," in *Oldenburger 3D Tage*, 2009.
- [9] B. Jähne, *Digitale Bildverarbeitung*. Springer Verlag, 2005.
- [10] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, pp. 179 – 187, 1962.
- [11] R. Duda, E. Hart, and D. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, 2000.