

Accuracy Evaluation for Automated Optical Indoor Positioning Using a Camera Phone

Verena Händler and Volker Willert

Summary

In this paper, we focus on the accuracy of optical indoor positioning and the design of an automated and mobile positioning system 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 an object with four reference points with known image and world coordinates. 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 classification methods in terms of robustness and to present the accuracy of the image coordinate localization techniques.

Zusammenfassung

Im folgenden Beitrag wird die Genauigkeit optischer Indoor-Positionierungssysteme in Abhängigkeit von Bildmessdaten

analysiert und die Umsetzung automatisierter Klassifikations- und Detektionsverfahren für ein mobiles, bildbasiertes Indoor Positionierungssystem, realisiert durch eine Handykamera, vorgestellt. Die Positionsbestimmung erfolgt in diesem Fall anhand einer Einzelaufnahme, welche die Projektion eines Objektes mit vier Referenzpunkten enthält. Die Referenzpunkte besitzen dabei bekannte Objektkoordinaten. Um die Position der Kamera relativ zum Objekt zu erhalten, muss das Objekt zuerst detektiert und klassifiziert werden. Danach müssen die Bildkoordinaten extrahiert werden, um anschließend den räumlichen Rückwärtsschnitt zur Positionsbestimmung anwenden zu können. Wenn ein Objekt korrekt klassifiziert wird, hängt die Qualität der Position zu einem großen Teil von der Genauigkeit der extrahierten Bildkoordinaten und des verwendeten Algorithmus zur Lösung des Rückwärtsschnitts ab. Daher werden drei Ansätze zur Bestimmung des räumlichen Rückwärtsschnitts verglichen und zwei Verfahren zur Klassifikation georeferenzierter Türen und die Extraktion der Bildkoordinaten der Türeckpunkte vorgestellt. Anhand statistischer Untersuchungen werden die beiden Klassifikationsverfahren bezüglich ihrer Robustheit miteinander verglichen und die Genauigkeit der Extraktion der Bildkoordinaten aufgezeigt.

Keywords: indoor positioning, optical systems, camera phone, classification

1 Introduction

The field of indoor positioning systems 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 *wireless local area network (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. *radio frequency identification (rfid)* tags) before the positioning process can take place. Using approximate methods, a positioning accuracy of some meter can be achieved.

Precise methods (e.g. using *ultra wide band (uwb)* or *ultrasound* 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, see Fig. 1.

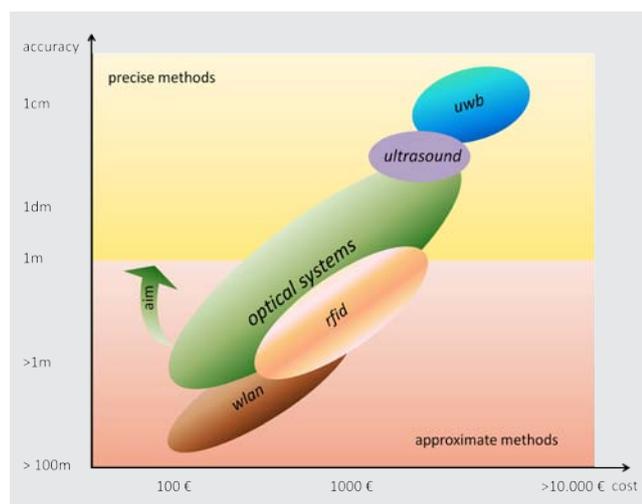


Fig. 1: Cost effectiveness of indoor positioning systems

Optical indoor positioning systems can be assigned to precise methods as well as to approximate methods, see Fig. 1. *Optical systems* provide precise positioning methods in the field of industrial surveying (Aicon 2011) but they also have been established as low cost approximate positioning systems. Here, they are widely spread in the field of robotic navigation (Jones and Soatto 2011). Also in the field of Augmented Reality, *optical systems* have already been presented in Hile and Boriello (2008) or Mulloni et al. (2009). The *optical systems* in Hile and Boriello (2008) and Mulloni et al. (2009) 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 com-

parable to those of other low cost indoor positioning systems. Because *optical systems* cover the whole spectrum from cheap and approximate up to expensive and precise systems, our aim is, to use rather low cost cameras and even so to achieve quite precise positioning results.

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.

An overall system is outlined in Fig. 2. The position of a person inside a public building is determined using a camera phone that includes wifi-functionality and an application for data exchange with an external server. First, a door known in the world coordinate system is photographed. The photo is sent to a server via wireless communication technologies. Next, a server application extracts the image coordinates of the door's corners as well as the code information. The world coordinates can be identified from the code information and the corresponding corners that are stored in a database. Instead of an explicit code, also a classifier based on visual cues can be used to identify a specific door with its corners. The extraction of the corners' 2D-coordinates enables the calculation of the 3D-position of the camera phone P_0 . The position information is marked on a map, which is sent back to the mobile phone and displayed on its screen.

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

Up to now different methods for evaluating the position using optical devices have been investigated and presented in Hile and Boriello (2008) and Mulloni et al. (2009). In Mulloni et al. (2009) fiducial markers themselves serve

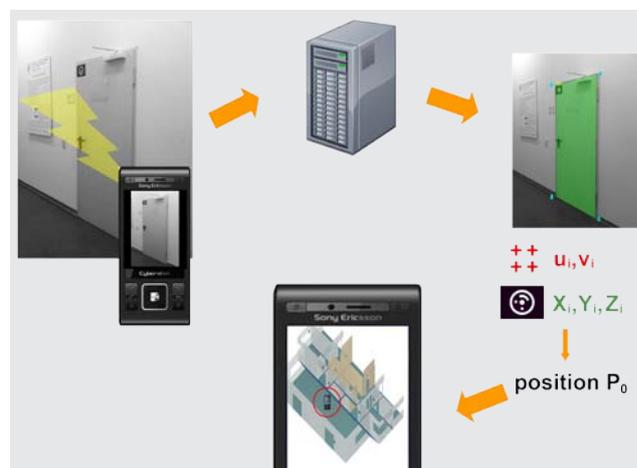


Fig. 2: Application flow of the positioning process

as references. The article focuses on the evaluation of the usability of their camera phone system under real world conditions. The application in Hile and Boriello (2008) uses a floor plan with relevant features marked on it for referring the image of a cell phone camera to the world coordinate system. The main focus here is the speed of the positioning process or rather the relation between speed and accuracy. In contrast, we focus on the influence of noisy measurements (image coordinates) on the quality of the positioning. In accordance to the accuracy estimations we present two automated classification methods and automated detection techniques.

2 Accuracy Estimation

The accuracy of an optical indoor positioning system depends on the geometrical configuration as well as on the uncertainties of the measurements and the reference stations. In Haralick et al. (1994) 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, v, X, Y, Z) 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 to be investigated separately for different geometrical configurations. In the following, it will be shown, that especially the accuracy of the image coordinates has a deep impact on the accuracy of the position estimate \tilde{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 are located on a regular grid, lying 3, 5 and 7 m away from the planar object and sideways up to 1 m. Adding Gaussian noise to the image coordinate data the position \tilde{P}_0 has been calculated with different approaches for solving the spatial resection and the results have been compared to the ground truth position P_0 . Three approaches for solving the spatial resection (Grunert 1841, Killian 1955, Rohrberg 2009) have been investigated according to their robustness against deviations in the measurements. The approaches in Grunert and Rohrberg use three corresponding points to calculate different solutions for P_0 and afterwards use a fourth point to get the unique solution whereas in Killian the fourth corresponding point is included in the calculation from the beginning assuming a planar object.

In Fig. 3 we show the impact of erroneous image coordinates on the positioning result with the image coordinates corrupted by additive zero mean Gaussian noise with increasing standard deviation. The ground truth position in this case is $P_0 = (3/5/1.50)$ m (also marked

in Fig. 5). The position estimate \tilde{P}_0 has been calculated 100 times, each time drawing a random sample from the same Gaussian distribution for all three algorithms. To judge the deviation and the uncertainty, we use the mean deviation of position d_p and the Helmertian error of position s_p^H (Niemeier 2001), respectively. The Helmertian error of position increases almost linearly for the 3-point-algorithms of Grunert and Rohrberg (except for very small noise with a nonlinear increase) as well as for the approach of Killian (which shows a linear increase also at the beginning). For the approach of Killian it is always smaller than that of the 3-point-algorithms. Hence, the 4-point-algorithm seems to be more robust against the configuration in combination with noisy measurements than the algorithms of Grunert and Rohrberg.

Looking at the mean d_p in Fig. 4, it can be seen, that it rises with increasing noise for all three algorithms. But the increase for the algorithms of Grunert and Rohrberg is much larger than for the approach of Killian. The occurrence of such an increasing offset is obvious, because the noisy 2D-coordinates are reprojected by the resection algorithms. Since reprojection is a nonlinear transformation, the resulting distribution of the 3D-coordinate of the estimated position is no longer Gaussian distributed. It keeps being unimodal but asymmetric. The asymmetry increases with increasing distance because the perspective transformation is proportional to the inverse of the distance (see also Fig. 5). Since the approach of Killian outperforms the Grunert and Rohrberg algorithms both, for s_p^H and the mean d_p , as from now, we decide to only evaluate the approach of Killian. As can be seen in Fig. 3 and Fig. 4, image coordinates with zero mean Gaussian noise and a standard deviation of 30px in an 2448×3264 px image lead to a mean d_p of 0.42 m and a s_p^H of 0.16 m. The image coordinates must be extracted with an accuracy of at least 10px to get results in the position with an accuracy of less than 1 dm but 5px should be strived to achieve a good solution for the position. A standard deviation of 5px results in a mean d_p of 1.6 cm with a s_p^H of 2.6 cm.

To examine the influence of the configuration in combination with noisy measurements on the position in space, we calculated the position with noisy image data for all grid points (100 times each), using the approach of Killian. The error ellipsoid as well as the deviation of position of every grid point is depicted in Fig. 5.

The deviation of position as well as the Helmertian error of position obviously depend on the configuration. Also the assumption is approved that a bias occurs that falsifies the position estimate systematically, depending on the distance between the reference object and the camera due to the nonlinear dependency between the 2D- and 3D-coordinates inherent in all resection algorithms. Therefore the position estimate \tilde{P}_0 is no longer Gaussian distributed.

In the following, different possibilities for extracting the image coordinates of certain characteristics have

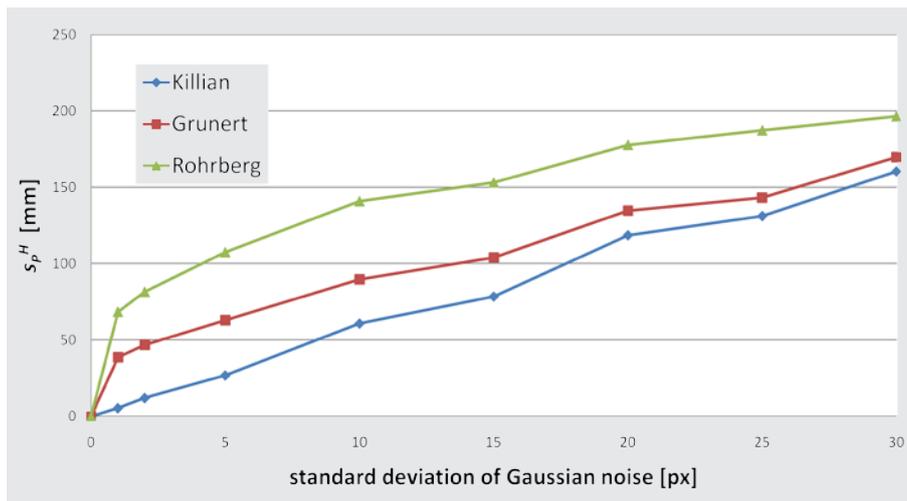
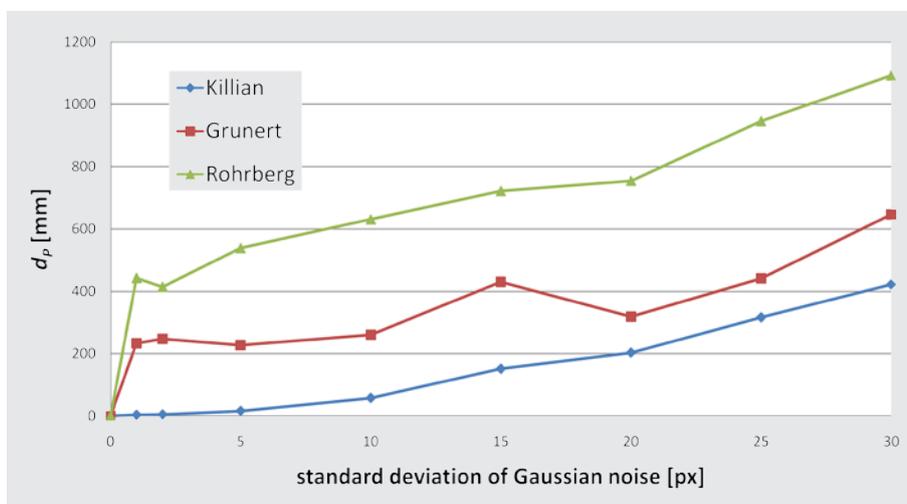
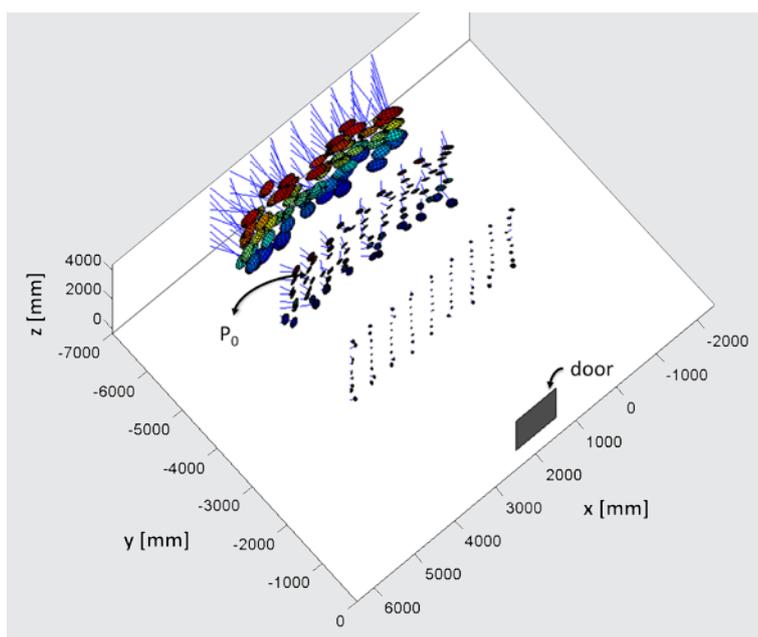
Fig. 3: s_p^H of spatial resection solutions for noisy dataFig. 4: d_p of spatial resection solutions for noisy data

Fig. 5: Error propagation of image coordinates with Gaussian noise of 20 px on the 3D position.

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 and positioning in the presented investigations.

3 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. In both methods 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.

3.1 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 (e.g. morphological operations (Jähne 2005)). There exist several tags in augmented reality communities like ARToolKit or ARTags (Fiala 2005). But these tags only work in close ranges up to 1.5m. Therefore we created our own fiducial marker, similar to markers used in Niederöst and Maas (1997).

Here, the visual target is a binary image characterized by a white circle (radius = 5 cm) on a black background (area = 252 cm²) surrounding a white visual code, see Fig. 7a.

To find the marker, the image is binarized and closed objects \mathcal{O}_c are localized using morphological operations (Jähne 2005) (combinations of *opening* and *closing*). These closed objects serve as candidates for markers. 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 $c(\mathcal{O}_c)$ has to be analyzed for all objects \mathcal{O}_c . 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

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

For a circle, c gets minimal with a minimum of $c=1$, see Jähne (2005). Calculating c and the area of all closed objects, the fiducial marker can be detected from the image. For an ellipse the minimum of c is $c \leq 1$ (Burger 2006). Therefore the threshold of the roundness has been set to 0.9, to detect all circular objects, even if they are distorted to an ellipse. Another criterion of a fiducial marker object is its area. For the fiducial markers used in the test-bed the area $A(\mathcal{O}_c)$ must have a value of at least 200px to be assumed as a visual target.

c) Classification via decoding the marker

To get the coded information, the detected marker first has to be normalized using an affine transformation

$$T = R \cdot S = \begin{bmatrix} \cos(\omega) & -\sin(\omega) & 0 \\ \sin(\omega) & \cos(\omega) & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{d}{b} & 0 & 0 \\ 0 & \frac{d}{a} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2)$$

The parameters of the transformation matrix are determined by an ellipse with its major axes a and b , that surrounds the object \mathcal{O}_c , Fig. 6a. The diameter d is given by a circle with an area equal to that of the ellipse, Fig. 6b. The angle ω between the major axis of the ellipse and the

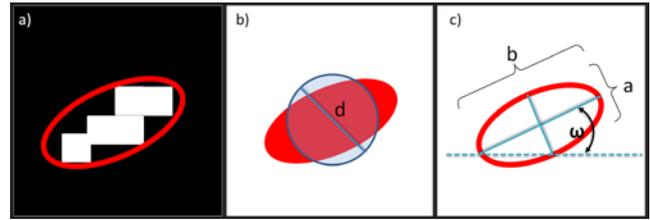


Fig. 6: a) Ellipse hull, b) Circle diameter d , c) Rotation angle ω and major axes a and b

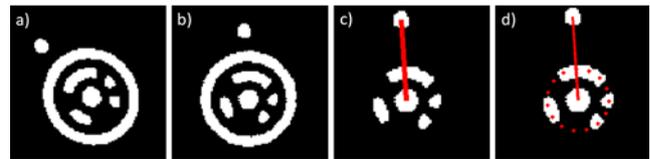


Fig. 7: Transforming and decoding the fiducial marker

horizontal axis of the image defines the rotation angle of the transformation, Fig. 6c. After the normalization, the information can be decoded, see Fig. 7b. The code exists of 15 segments representing a 15 bit code. The line between the center of the code and the center of the small circle outside the code circle in Fig. 7c defines the beginning of the marker code. The radius r of the marker is defined by the changes between black and white parts of the red line in Fig. 7c and 7d. Starting from the point on this red line with distance r from the code center, each 24 degrees a marker is set that defines one of the code segments, see Fig. 7d. 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.

d) Extracting the image coordinates of the reference points

First, we apply morphological operations on the binarized image, 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. 8a. Then, applying a *Hough Transformation* (Jähne 2005), the lines closest to the contour of the door segment are extracted, see Fig. 8b. The intersections of the extracted contour lines (red crosses) present the image coordinates of the projected known corners of the door.

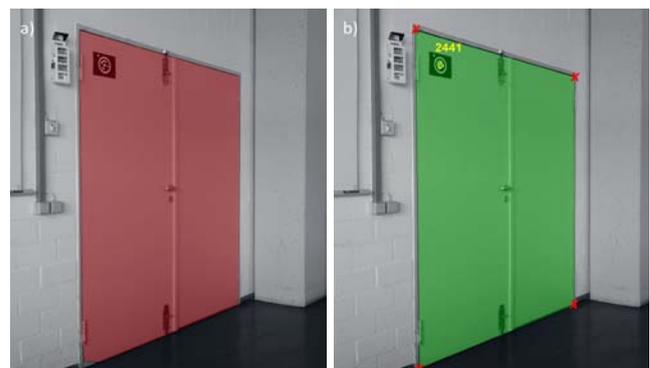


Fig. 8: a) Contour of a closed object, b) Door detection result

3.2 Classification using a reference image database

If there are no fiducial markers provided for detection and classification of the doors, only the images of the known doors 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. 9 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 question.

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 (see Fig. 10) are stored in a database together with the image and object coordinates of the labeled corners of the doors. Creating the database, we tried to take rather normalized reference images, where the doors fill most of the image and are only marginal



Fig. 9: Different views of the same door



Fig. 10: Reference images for the database

distorted. We also took more than one photo of a door from different views to augment the hit rate.

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 of a door are lines in the image that are extracted using the *Canny-Edge-Detection* (Jähne 2005, see Fig. 11a) and a subsequent *Hough Transformation* (Jähne 2005, see Fig. 11b). The green lines with yellow start- and end-points visualize the extracted contour features.

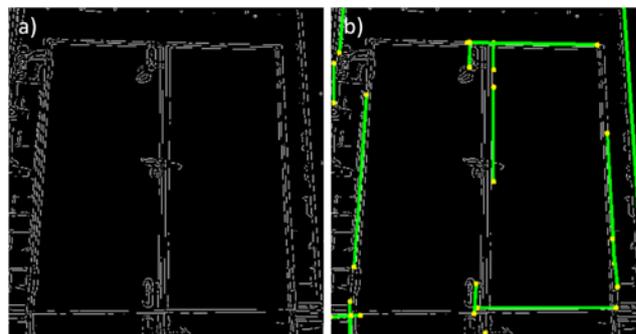


Fig. 11: Contour features extracted with a) Canny-Edge-Detection and b) Hough Transformation

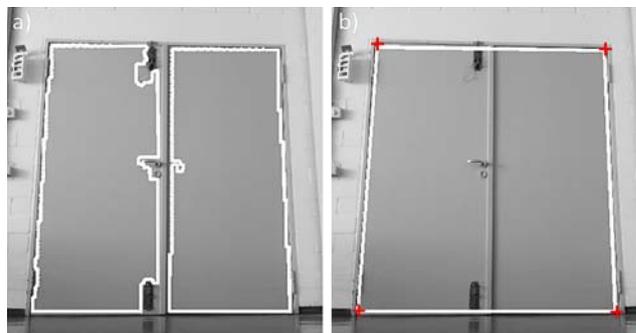


Fig. 12: a) Contour of quadtree segmentation, b) Quadrangle detection result

The surface features for the door are homogeneous gray value regions extracted by a *quadtree decomposition* (Jähne 2005). 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. 12a 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 C_H to the quadtree contour C_Q by minimizing the squared difference

$$C_H^* = \arg \min_{C_H} (A_Q(C_Q) - A_H(C_H))^2 \quad (3)$$

of the surface areas A_Q and A_H that are bounded by the contours C_Q and C_H with respect to C_H . The different values for $A_H(C_H)$ result from all possible quadrangular con-

tours \mathcal{C}_H that can be constructed out of the Hough lines nearby the quadtree contour. Fig. 12b shows the result \mathcal{C}_H^* of this quadrangle extraction process (white quadrangle).

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. 13.

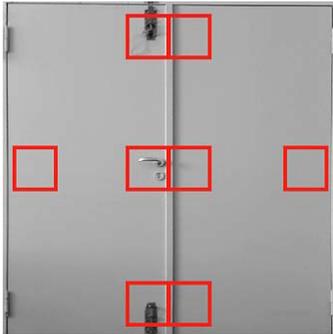


Fig. 13: Normalized image with ROIs

Then, for every quadrangle several *regions of interest* (ROIs) can be defined (see red boxes in Fig. 13), 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 I_R is perspectively warped $\mathcal{W}(I_R)$ such that the contours of the door within the reference image exactly overlay the quadrangular candidate door of the original (unnormalized) image 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 I_R

$$I_R^* = \arg \min_{I_R} \sqrt{(1 - \rho_{R,C})^2 + \left\| \begin{matrix} \mathbf{m}_C - \mathbf{m}_R \\ \mathbf{m}_R \end{matrix} \right\|^2} \quad (4)$$

where $\rho_{R,C}$ is the correlation coefficient (Luhmann 2010) between I_C and $\mathcal{W}(I_R)$ and the vectors \mathbf{m}_C and \mathbf{m}_R include the Hu-moments (Hu 1962) of I_C and $\mathcal{W}(I_R)$.

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. 12b shows one result of such intersections (red crosses).

4 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 each case, the effort for creating the database is more or less similar.

For practical investigations we used an environment including 11 different doors (3 two-wing-doors and 8 one-wing-doors). In case of similar door types we photographed the doors in a greater distance to store additional information of the surrounding of one door. 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 (Duda et al. 2000) of the binary classification problem (door/no door) are shown in Tab. 1 and Tab. 2.

Using visual tags for classification, the fiducial markers were detected 56 times by their roundness (Eq. 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).

In case of 30 reference images, the detection step extracted 84 door candidates out of 78 images including 72 doors applying the minimization (Eq. 3). 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 (Eq. 4).

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

Tab. 1: Confusion matrix using fiducial marker

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

Tab. 2: Confusion matrix using reference images

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

Tab. 3: Accuracy of extracted image coordinates

	height [px]	width [px]
min. dev.	1	0
max. dev.	269	248
av. dev.	31	23
stand.dev.	37	38

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 Tab. 3 we point out the results according to their accuracy.

The deviation is 1 to 2px at minimum. However, deviations appear up to 269px. The reason for the strong outliers are some false contour extractions during the detection phase.

To test the concept described in Section 1, the camera was positioned on certain control points, given in a 9×9m testbed (see Fig. 14a). The control points are marked with black crosses.

On these control points photos of one of the two doors have been taken (Fig. 14b) and the camera position has been evaluated, using the classification methods, described in Section 3.1 and 3.2 (green diamonds in Fig. 14a). In some cases only one of the two classification methods serves a solution. In general the d_p ranges from 2cm to 28cm. It varies according to the distance and to the configuration as well as to the accuracy of the detection of the door points in the image. Hence, d_p in position 13 (see Fig. 14a), is about 12.5cm whereas the greatest deviation of position can be found in position 10

with 28cm. So again it can be seen, that the accuracy of the detection of the image coordinates of a reference door has a deep impact on the positioning result.

5 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 corners of doors have been investigated especially according to their robustness and accuracy.

The analysis in Section 4 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 10px in an image with a resolution of 2448×3264 px, which cannot be achieved for all classified doors. In most cases, the reason for that are line detections that do not belong to the outer rim of the door frame but to other nearby edges or virtual edges caused by shadows of the door frame.

To improve the accuracy of the detection method described in Section 3.2 the creation of a suitable database of reference images has to be investigated more thoroughly in terms of viewpoint, distance and number of references from the same door. Looking at the results so far, it seems to be that the viewpoint is the critical parameter to optimize, since the distance only changes the resolution which is not that important for the recognition

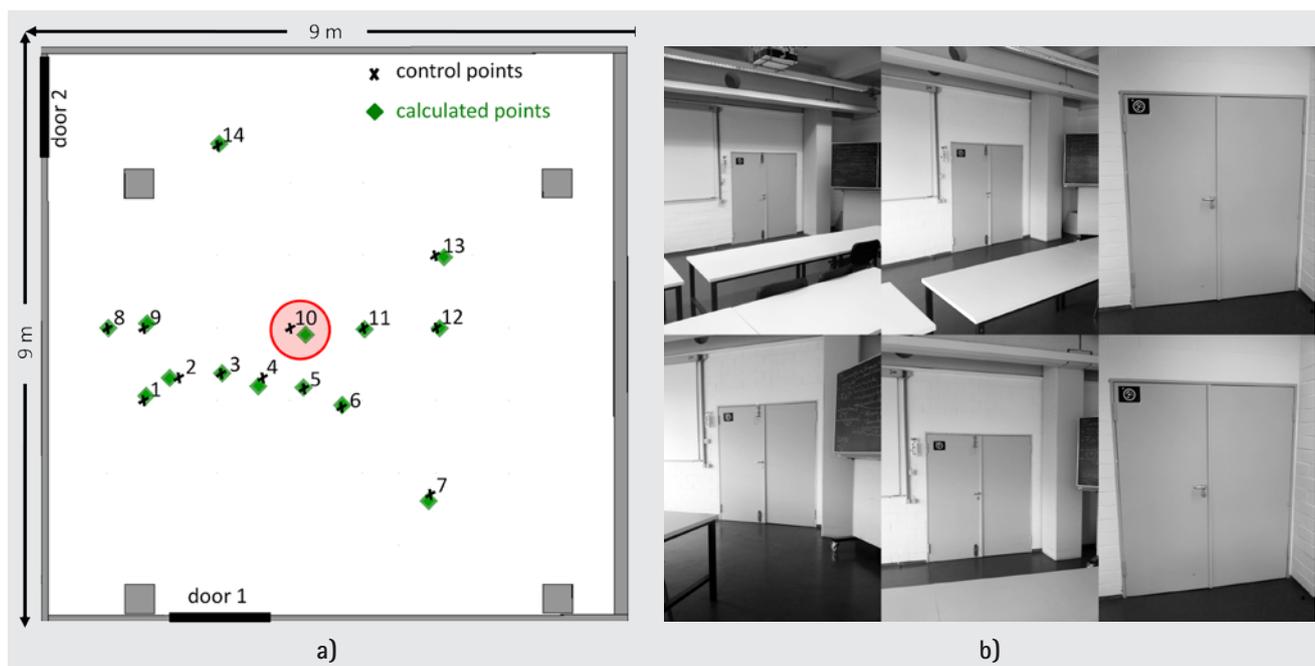


Fig. 14: a) Testbed with calculated positions using the two classification methods, b) Example images of the two doors in the testbed

of doors with a textureless surface. Enlarging the database by adding more reference images of the same door from different points of view would increase the computational costs and should therefore be avoided.

The comparison of both methods concerning their robustness indicates that positions nearby a door are sometimes not detectable with the method described in Section 3.2 because the comparison of the reference images with close-up pictures from a door does not work properly anymore. Vice versa, the method described in Section 3.1 sometimes breaks down for far away positions (≥ 8 m) because the projected size of the markers is too small ($A(\mathcal{O}_i) < 200$ px) to decode the marker depending on the resolution of the camera. Therefore, a combination of both methods could lead to a better performance for near (≤ 2.5 m) and far (≥ 7 m) away doors. Future research will focus on combining certain steps of the two presented methods to wipe out the disadvantages of the separate methods and to improve the accuracy of the position estimate and the classification performance.

Acknowledgment

We would like to thank Stefan Gering and Sebastian Raß for designing and implementing the classification algorithm using reference images and Johannes Etzel for implementing the fiducial marker detection algorithm. This work was gratefully supported by the German Research Foundation (DFG) within the GRK 1362 »Cooperative, Adaptive and Responsive Monitoring in Mixed Mode Environments«.

References

- Aicon: 2011. URL <http://www.aicon3d.de>. last access 05.12.2011.
- Burger, W., Burge, M.J.: *Digitale Bildverarbeitung – Eine Einführung mit Java und ImageJ*. Springer Verlag, 2 edition, 2006.
- Duda, R.O., Hart, E., Stork, D.G.: *Pattern Classification*. Wiley-Interscience, 2 edition, 2000.
- Fiala, M.: Comparing ARTag and ARToolKit Plus Fiducial Marker Systems. In: *IEEE International Workshop on Haptic Audio Visual Environments and their Applications*, pp. 148–153, 2005.
- Grunert, A.J.: *Das Pothenotische Problem in erweiterter Gestalt über seine Anwendungen in der Geodäsie*. Grunerts Archiv für Mathematik und Physik, 1: 238–248, 1841.
- Haralick, R.M., Lee, C., Ottenberg, K., Nolle, M.: Analysis and Solutions of the Three Point Perspective Pose Estimation Problem. *International Journal of Computer Vision*, 13/3: 331–356, 1994.
- Hile, H., Boriello, G.: Positioning and Orientation in Indoor Environments Using Camera Phones. *IEEE Computer Graphics and Applications*, 8: 32–39, 2008.
- Hu, M.-K.: Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8: 179–187, 1962.
- Jähne, B. *Digitale Bildverarbeitung*. Springer Verlag, 2005.
- Jones, E., Soatto, S.: Visual-Inertial Navigation, Mapping and Localization: A Scalable Real-Time Causal Approach. *International Journal of Robotics Research*, 30 no. 4: 407–430, January 2011.
- Killian, K.: Über das Rückwärtseinschneiden im Raum. *Österreichische Zeitschrift für Vermessungswesen*, 4: 97–104, 1955.
- Luhmann, Thomas. *Nahbereichsphotogrammetrie – Grundlagen, Methoden und Anwendungen*. Wichmann Verlag, 2010.
- Mulloni, A., Wagner, D., Schmalstieg, D.: Indoor Positioning and Navigation with Camera Phones. *Pervasive Computing*, 8: 22–31, 2009.
- Niederöst, M., Maas, H.-G.: Entwurf und Erkennung von codierten Zielmarken. In: *Tagungsband der 16. Wissenschaftlich-Technischen Jahrestagung der Deutschen Gesellschaft für Photogrammetrie und Fernerkundung 1996*, pp. 55–62, Oldenburg, 1997.
- Niemeier, W.: *Ausgleichsrechnung*. Walter de Gruyter Verlag, 2001.
- Rohrberg, K.: Geschlossene Lösung für den räumlichen Rückwärtsschnitt mit minimalen Objektinformationen. In: *Oldenburger 3D-Tage*, pp. 332–339. Wichmann Verlag, Heidelberg, 2009.

Authors' addresses

Verena Händler
TU Darmstadt, Institute of Geodesy
Petersenstraße 13, 64287 Darmstadt
haendler@geod.tu-darmstadt.de

Volker Willert
TU Darmstadt, Institute of Automatic Control
Landgraf-Georg-Straße 4, 64287 Darmstadt
vwillert@rtr.tu-darmstadt.de