Efficient probabilistic planar robot motion estimation given pairs of images

Paper-ID 225

$Id: planarPoseRSS10.tex 920 2010-01-22 02:27:4

***Abstract*— Estimating the relative pose between two camera positions given image point correspondences is a vital task in most view based SLAM and robot navigation approaches. In order to improve the robustness to noise and false point correspondences it is common to incorporate the constraint that the robot moves over a planar surface, as is the case for most indoor and outdoor mapping applications. We propose a novel estimation method that determines the full likelihood in the space of all possible planar relative poses. The likelihood function can be learned from existing data using standard Bayesian methods and is efficiently stored in a low dimensional look up table. Estimating the likelihood of a new pose given a set of correspondences boils down to a simple look up. As a result, the proposed method allows for very efficient creation of pose constraints for vision based SLAM applications, including a proper estimate of its uncertainty. It can handle ambiguous image data, such as acquired in long corridors, naturally. The method can be trained using either artificial or real data, and is applied on both controlled simulated data and challenging images taken in real home environments. By computing the maximum likelihood estimate we can compare our approach with state of the art estimators based on a combination of RANSAC and iterative reweighted least squares and show a significant increase in both the efficiency and accuracy.**

I. Introduction

Various vision based topological mapping [1, 2], geometrical mapping [3, 4] and robot navigation [5] approaches are based on the ability to compare pairs of images. A common way to do this is to automatically find similar looking image points [6], as done for two panoramic images in Figure 1. Because part of these point correspondences are the projections of the same 3D landmarks in the environment, they can be used to determine the relative camera pose up to an unknown scale [7]. A major challenge in determining the relative pose given point correspondences, is that a large percentage does not correspond to the same 3D landmark, but are so called *mismatches*. In addition the image point locations of correct matches are noisy, caused by for example noise of the imaging device and errors in the calibration.

To cope with this, so called *robust* algorithms are needed. There are three robust methods commonly used: RANSAC [8], M-Estimators [9] and the Hough Transform [10]. State of the art relative pose estimators first use RANSAC(RANdom SAmple Consensus) [8] combined with the closed form Eight [11] or Five point algorithm [12] for an initial estimate and then apply robust iterative reweighted least squares (IRLS) techniques such as M-Estimators (Maximum likelihood Estimators) [9] in combination with the Eight point algorithm [11] to improve it.

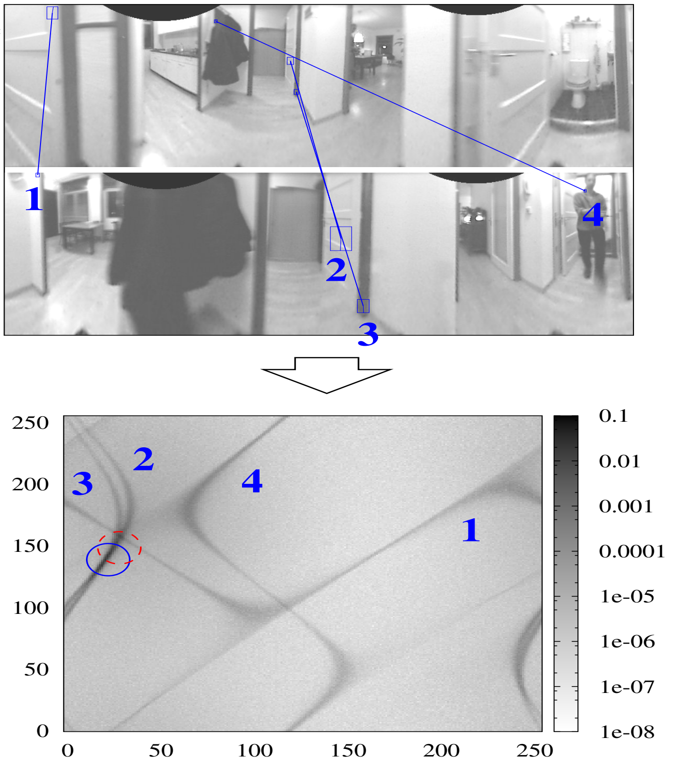


Fig. 1. An example. Applying standard SIFT matching on two panoramic images resulted in only four point correspondences including 2 mismatches and 2 almost degenerate correct matches. Still, the proposed method computes a full likelihood over the different possible relative robot poses, and the maximum likelihood (solid blue circle) is close to the ground truth (dashed red circle). The numbers relate the different correspondences in the image pair with the curves in the solution space.

Both RANSAC and the M-Estimator try to find a maximum likelihood solution by first rejecting mismatches and too noisy correspondences using an error threshold and base a least square solution on the remaining matches. Sophisticated techniques in Computer Vision try to determine this threshold from the image data itself [13, 14]. In the field of Robotics, where characteristics of the camera are available, this threshold is usually determined through some calibration procedure. The Hough Transform on the other hand, if seen probabilistically, computes the full likelihood on a discrete grid of poses. Because space requirements grow exponential with the number of parameters, it is in general not suited for pose estimation problems, although combinations of the Hough Transform with RANSAC [15] do exist, as well as methods that treat rotation and translation estimation separately [16, 17].

An approach to make pose estimation easier is to incorporate constraints on the possible relative poses. If the robot drives over a planar surface, then the camera can only rotate around a certain fixed axis which is perpendicular to the two dimensional translation direction. Given that the scale can not be determined, the number of degrees of freedom for this planar relative pose problem reduces to two. This constraint can be used to improve indoor mapping application using wheeled robots [5], but also vision based outdoor mapping using vehicles driving over planar roads [1, 18, 19].

Commonly this constraint is imposed rather heuristically, for example by taking only the horizontal displacement of image points into account [20, 4, 19]. A more proper solution is proposed by Brooks [21] which formulates a least square approach to the planar relative pose given noise free correspondences. This result was used in [22, 23, 2] for various robotics applications all in combination with RANSAC to make it robust against noise and mismatches. In [22, 24] it was shown that two correspondences are enough to solve the problem and both suggest algorithms, which are briefly evaluated in combination with RANSAC.

For practical reasons RANSAC based algorithms are used to randomly sample the solution space for many estimation problems. Discretizing and analyzing the whole solution space potentially leads to much more robust results. In this paper we show that this is feasible for the planar motion estimation case without the additional approximations as in [16, 17]. To the best of our knowledge this is the first time a Hough-like approach is directly used for motion estimation from images. Our results demonstrate the greatly increased robustness over the random sampling techniques. Furthermore, we present a probabilistic representation where we learn the needed conditional distributions from real or simulated data without any parameters, except the size of the discretization grid. This allows fast tuning of the approach for different robots and cameras since most standard approaches have a set of parameters that need to be tuned [4, 3]. Additionally, the learned conditional distributions properly capture other sources of uncertainty that are difficult to explicitly account for, e.g. shaking of the robot and inaccurate camera calibration. Therefore, more accurate and robust results are obtained as demonstrated in the experimental section. Finally, testing the whole solution space potentially can be computationally expensive. We present an efficient implementation using a precomputed LUT. The geometry of the problem is analyzed and a parameterization is proposed to reduce the dimensionality of the needed LUT to only three dimensions. The experimental results show that the whole scheme can be even faster than the common sampling based approach while at the same time we get more robust and accurate results.

The rest of the paper is organized as follows. First, in Section II we formalize the planar relative pose problem and describe how noise free correspondences relate to it. This relation is used in Section III to derive the novel estimator and how it can be trained using real image data. In Section IV we apply it on both simulated data and image data-sets taken in both an office and real home environments and compare the Maximum Likelihood estimates with the results from a planary constrained RANSAC combined with an M-Estimator. In Section V we discuss the qualitative advantage of having a full likelihood solution and propose some directions for improvement. Finally, in Section VI, conclusions are drawn.

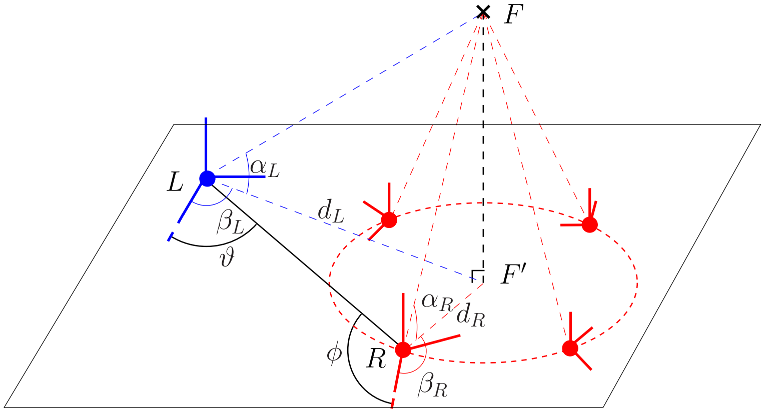


Fig. 2. 3D visualization of two cameras, L and R, positioned on the same plane both observing a landmark F. The dashed circle on the ground plane indicates the possible positions for robot R given the pose of L and observations of F.

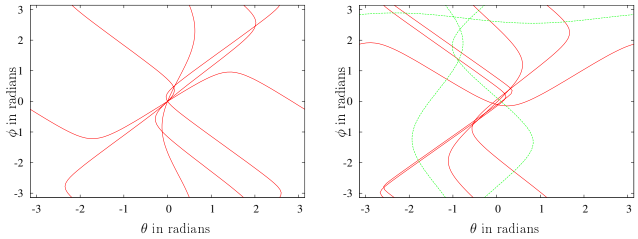
II. Relating correspondences to poses

The planar relative pose, which can be seen as a 2D translation and rotation, minus scale, can be parameterized in different ways. We choose to parameterize it using two angles ϑ and ϕ, see Figure 2. Angle ϑ denotes the direction of the translation, or heading, of robot R in the coordinate frame of robot L and angle ϕ denotes the heading of robot L in the coordinate frame of robot R. Another common parameterizations is using the heading and the rotation of robot R in the frame of robot L, such as in [21, 22]. However, our parameterization reflects the symmetry of the problem.

It is, for now, assumed that image point correspondences are obtained by a noise free projection of landmarks, without mismatches. An image point is usually denoted by a 3D vector of unit length $\mathbf{x} = [x, y, z]'$, where the z-axis is pointing in the direction of the camera axis and the y-axis is in the planar case perpendicular to the ground plane. Here we denote an image point by its horizontal angle $\beta = \text{atan2}(z, x)$ and its vertical angle $\alpha = \arcsin(y)$. This is similar to the azimuth and elevation used in the inverse depth parameterization by landmark based SLAM methods [25]. If two robots L and R observe a landmark F, the point correspondence is thus denoted by α_L, β_L and α_R, β_R.

The 3D problem as shown in Figure 2 can be reduced to a simpler 2D problem, by projecting F on the plane getting F'. We define d_L and d_R as the distances of L to F' and R to F' respectively. The length of F to F' can now be expressed by:

$$\overline{FF'} = \frac{\tan(\alpha_L)}{d_L} = \frac{\tan(\alpha_R)}{d_R}, \tag{1}$$



(a) (b)

Fig. 3. Visualization of the possible relative robot poses by plotting ϕ as a function of ϑ for multiple correspondences randomly picked using a relative pose with $\vartheta = \phi = 0$. (a) Generated using 5 noise free correspondences. (b) Generated using 5 noisy correspondences and 3 mismatches (dashed curves). Note that, by coincidence, one of the mismatches corresponds to a curve close to the actual pose.

which results in the following ratio r of d_L and d_R:

$$r = \frac{d_L}{d_R} = \frac{\tan(\alpha_R)}{\tan(\alpha_L)}. \qquad (2)$$

Angle $\angle F'RL$ can then be found using the Law of Sines:

$$\begin{aligned} \frac{\sin(\angle F'RL)}{d_L} &= \frac{\sin(\beta_L - \vartheta)}{d_R}, \\ \angle F'RL &= \arcsin\left(\frac{d_L}{d_R}\sin(\beta_L - \vartheta)\right). \qquad (3) \end{aligned}$$

Angle ϕ can now be expressed as a function of angle ϑ and a single point correspondence by adding the horizontal observation angle β_R, and using Equation (2):

$$\phi = \beta_R + \arcsin\left(\frac{\tan(\alpha_R)}{\tan(\alpha_L)}\sin(\beta_L - \vartheta)\right). \qquad (4)$$

We could also rewrite this formula into:

$$\vartheta = \beta_L + \arcsin\left(\frac{\tan(\alpha_L)}{\tan(\alpha_R)}\sin(\beta_R - \phi)\right), \qquad (5)$$

which is evident given the symmetry between ϑ and ϕ.

We now obtained functions that map ϑ to ϕ and vice versa given a single point correspondence. Figure 3(a) shows curves of possible relative robot poses generated using Equation 4 given some randomly picked point correspondences. Looking closely at this figure one can see that some pairs of curves intersect each other twice. This shows that, although it was assumed that two point correspondences can be used to solve the relative pose problem [22, 24], in some cases two different relative robot poses result in the same two point correspondences. Based on Equation 4 an algorithm was derived that computes both these solutions given two correspondences, which can be combined with hypothesize and test schemes such as RANSAC. This "Planar two point algorithm" is rather tedious and left out of this paper, but for completeness it is taken into account in the experiments.

III. Full likelihood estimator

We now use the relation between noise free correspondences and relative poses to develop an algorithm that can deal with noisy correspondences, including mismatches, see Figure 3(b). The problem can be formulated as determining the negative log likelihood of each pose (ϑ, ϕ) given n point correspondences $\{\xi_1, \ldots, \xi_n\}$, each parameterized by the angles $\xi_i = (\alpha_{Li}, \beta_{Li}, \alpha_{Ri}, \beta_{Ri})$ [26]:

$$\begin{aligned} \mathcal{L}(\vartheta, \phi) &= \sum_i \mathcal{L}_i(\vartheta, \phi) \qquad (6) \\ &= \sum_i -\log p(\xi_i|\vartheta, \phi), \qquad (7) \end{aligned}$$

where $\mathcal{L}_i(\vartheta, \phi) = -\log p(\xi_i|\vartheta, \phi)$ is the contribution of each point correspondence to the 2D log likelihood function. The 2D solution space is discretized into a 2D histogram. For each bin we need to sum the contribution of all point matches. Therefore for each point match we need to calculate the corresponding 2D histogram approximating its log likelihood contribution $\mathcal{L}_i(\vartheta, \phi)$. This is computational costly, and we show next how this can be efficiently performed using a precomputed look-up table to approximate the logarithm of the conditional probability $\mathcal{L}_i(\vartheta, \phi) = -\log p(\xi_i|\vartheta, \phi)$.

A. Look up table

The negative log likelihood of a single correspondence ξ is given by $-\log p(\alpha_L, \beta_L, \alpha_R, \beta_R|\vartheta, \phi)$. Thus, if we would naively construct such a look up table, it would have 6 dimensions, 4 for the point correspondence angles and 2 for the relative pose. In order to keep the size of the LUT comprehensible, the 6D space should be discretized in large bins, resulting in a large discretization error. Fortunately, the dimensionality of the LUT can be reduced to only 3 dimensions by using the one point mapping function introduced in Section II and some common assumptions about the noise characteristics of image points.

The one point mapping function given in Equation (4) can be written as

$$\phi - \beta_R + \arcsin\left(\frac{\tan(\alpha_R)}{\tan(\alpha_L)}\sin(\vartheta - \beta_L)\right) = 0. \qquad (8)$$

As can be seen the terms α_L and α_R are only used in the combination $\frac{\tan(\alpha_L)}{\tan(\alpha_R)}$. Also, variables ϑ and β_L, and variables ϕ and β_R are only used in the combinations $\vartheta - \beta_L$ and $\phi - \beta_R$ respectively, which describe the horizontal angles to the landmark relative to the heading of the cameras. The joint probability of observing a correspondence under a certain relative pose can thus be represented as:

$$p(\alpha_L, \beta_L, \alpha_R, \beta_R, \vartheta, \phi) = p\left(\frac{\tan(\alpha_L)}{\tan(\alpha_R)}, \vartheta - \beta_L, \phi - \beta_R\right). \qquad (9)$$

This holds if we assume that the noise on of the horizontal and vertical view angles do not depend on their value, which is similar to the common assumption that the noise of the pixel locations is homogeneous.

Because of the symmetry of the representation of the relative planar pose, the two points of the point correspondences and the heading angles can be swapped giving the same result:

$$p\left(r, \vartheta - \beta_L, \phi - \beta_R\right) = p\left(\frac{1}{r}, \phi - \beta_R, \vartheta - \beta_L\right), \qquad (10)$$

where we introduced $r = \frac{\tan(\alpha_L)}{\tan(\alpha_R)}$ for convenience. In practice this means that we only have to construct a LUT for $0 < r < 1$ and swap $\vartheta - \beta_L$ with $\phi - \beta_R$ and use $\frac{1}{r}$ instead of r if $r > 1$.

The likelihood can be determined from the joint probability by dividing by the probability of the pose $p(\vartheta, \phi)$:

$$\begin{aligned} p(\xi|\vartheta,\phi) &= p(\alpha_L, \beta_L, \alpha_R, \beta_R|\vartheta,\phi) & (11) \\ &= p(\alpha_L, \beta_L, \alpha_R, \beta_R, \vartheta,\phi)/p(\vartheta,\phi) & (12) \\ &= p\left(\frac{\tan(\alpha_{Li})}{\tan(\alpha_{Ri})}, \vartheta - \beta_{Li}, \phi - \beta_{Ri}\right)/p(\vartheta,\phi) & (13) \end{aligned}$$

Usually, during the construction the LUT, one takes care that the different relative poses are uniformly distributed, making the likelihood proportional to the joint:

$$p(\xi|\vartheta,\phi) \propto p\left(\frac{\tan(\alpha_L)}{\tan(\alpha_R)}, \vartheta - \beta_L, \phi - \beta_R\right). \quad (14)$$

Efficiently determining a full likelihood over the relative poses given a set of correspondences is now straightforward. For each correspondence ξ_i we compute the value of $\frac{\tan(\alpha_{Li})}{\tan(\alpha_{Ri})}$ and pick the corresponding 2D slice of the look up table. Then we shift it in the direction of β_{Li} and β_{Ri}, wrapping the values at the borders. This results in the negative log likelihood of each pose given a correspondence, which can be summed for the different correspondences, resulting in a full likelihood.

B. *Learning the conditional probability*

The LUT representing the negative log likelihood can be constructed from existing data. This data can be generated by a simulator modeling the planar pose problem including a vision system. Better is to use a representative image set for which ground truth robot pose data is available. The main problem of real data is that the relative poses are in general not uniformly distributed. This invalidates the simplification proposed in Equation (14) and results in a bias of the LUT towards certain poses which are overrepresented in the dataset. However, we can easily compensate for this problem.

When constructing the LUT, we explicitly take the probability of the relative pose into account (see Equation (13)). In a first step, a 2D discretized probability $p(\vartheta, \phi)$ is constructed by making a histogram for all poses in the dataset and normalizing it. Then, in a second step, the dataset is used to build the 3D LUT like for the simulator, with the difference that for each pose correspondence it adds $\frac{1}{p(\vartheta,\phi)}$ to the 3D histogram. Again, each value of the histogram is replaced by its negative log, resulting in a proper LUT.

A second problem, which can not be circumvented, is that the amount of data in an image dataset is limited. As a consequence we usually can not construct a LUT with a very high number of bins. In the next section we evaluate, among other things, the consequences of such a smaller LUT.

IV. Experiments

By determining the Maximum Likelihood from the estimated full likelihood, we can compare our method for planar relative pose estimation with the state of the art robust methods using RANSAC combined with an M-Estimator as described briefly in Section I. We combined these with the Planar two point, the Planar Three point and the general Eight point algorithm. We used simulated data and 4 datasets obtained from a robot with an omnidirectional vision system. The methods are compared on the basis of their robustness against mismatches and noise.

We evaluate the estimated heading and rotation angle as in [27] by taking the absolute difference with the ground truth values. Because these errors are not normally distributed we use the median to describe the error distribution. A robust way to describe the spread of these medians is to use the Median of Absolute Deviations (MAD). Another important evaluation criterion is the computational time used by the algorithms. For all experiments we report these times, as implemented in C++ and run on the single 2 Ghz CPU core of a Pentium PC.

A. *Experiments on simulated data*

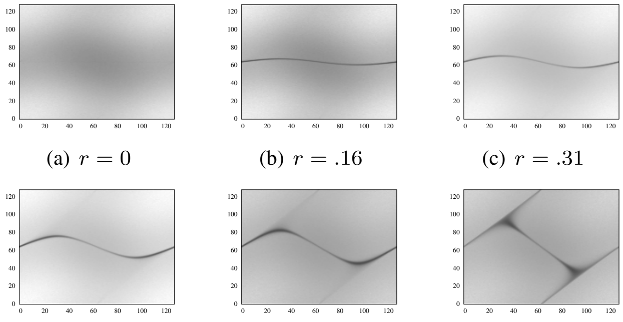
Using simulated data allows us to control the projection noise and number of mismatches. Also, it allows us to determine the percentage of correctly found mismatches.

1) Data: Data was simulated by randomly picking a uniformly distributed point cloud of 3D landmarks inside a sphere of size 2 around the origin. Two random camera poses on a circle with radius 1 in the x-y plane around the origin are chosen. From these the ground truth values for ϑ and ϕ are determined. Note that the distribution of ϑ and ϕ is approximately uniform.

A set of point correspondences is constructed by projecting the landmarks on a spherical shaped image surface with a radius of 1 around the camera pose. Thus, an ideal omnidirectional camera model is used with a full 360 degrees view angle in the horizontal and vertical direction. An amount of normally distributed noise with zero mean and standard deviation 0.01 is added to these projections. This value corresponds to an angular error of approximately .57 degrees, which corresponds to about 6 pixels for a typical conventional mega pixel camera with a focal length of 8 mm. This amount of projection noise seems quite large. However, it also accounts for the simplifications of the camera model, the calibration errors and errors of the image key point extractors. In addition, mismatches are added by creating false correspondences between projections of different landmarks. We use a mismatch rate of 90%.

2) Setup: A LUT was constructed using the same simulator Section III-B. The number of bins to represent ϕ, ϑ and r were all 128, which caused the computational time to be comparable with that of the RANSAC+M-Estimator combined with the the 3-point algorithm. The number of point correspondences used was $10^{1}0$, which took 3 hours to build. The error threshold for both RANSAC and the M-Estimator were set according to the projection noise of the simulator.

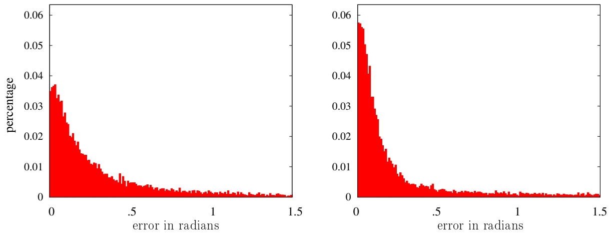
3) Resulting Look up table: Figure 4 shows the resulting look up table. In (a) one can see that point correspondences with a r value close to zero, do not tell that much about the data. This is due to the fact that there is a high chance that it resulted from a mismatch. Note also that the histogram corresponding to a r value close to 1 in (f) is almost symmetric.



(a) $r = 0$ (b) $r = .16$ (c) $r = .31$

(d) $r = .55$ (e) $r = .78$ (f) $r = .99$

Fig. 4. Look up table obtained from simulated data as denoted in Section III-B, which was also used throughout the experiments.



(a) Heading error 3pt. RANSAC + M-Estimator (b) Heading error LUT ML

Fig. 5. Comparison of the RANSAC+M-Estimator combined with the 3-point algorithm and the proposed LUT ML method for 90% mismatches. The distribution of rotation errors shows a similar pattern.

Indeed if r is exactly 1 then $r = \frac{1}{r}$ and we could swap ϕ and θ.

4) Resulting distributions: The distribution of errors is given in Figure 5. The LUT ML method results is superior for 90% percent mismatches. Also it is clearly visible that both distributions have long tails.

5) Sensitivity to mismatches: To test the robustness of the different methods to mismatches, we vary the number of mismatches, from 50% to 99%. In total 10^5 iterations were conducted. In Figures 6(a) and 6(d) the resulting estimation errors are plotted.

Both the heading and rotation results show the same trend. The error of the RANSAC+M-Estimator with the 8-point algorithm, which does not take planarity into account, increases fast if more than half of the correspondences are mismatches. The RANSAC+M-Estimator with the 2-point and 3-point algorithm behave similarly and start diverging at 65% mismatches. Note that the 3-point version is always slightly better than the 2-point version. The accuracy of the LUT based ML estimator, on the other hand, does not seem to influenced by high mismatch rates.

B. Experiments on real data

We compared the performance of our method with other methods on more than $3 * 10^6$.

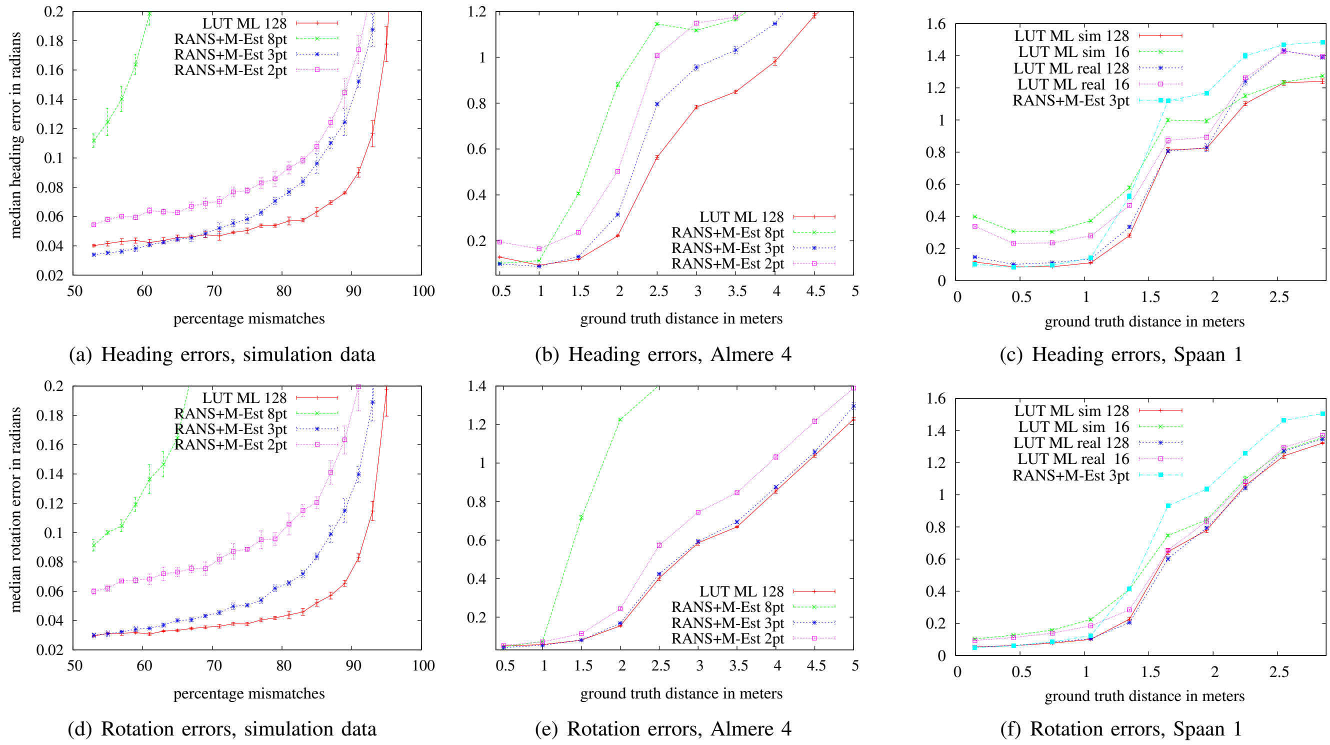
1) Data: We used four distinct image datasets. The first three were obtained using our Nomad Super Scout II and the fourth by the 'Biron' robot from the University of Bielefeld. On both an omnidirectional vision system was mounted consisting of a conventional Firewire camera pointing pointing upwards to a hyperbolic mirror. The 'Office' set was taken in a typical office environment. All other sets are taken in real home environments. In the 'Almere 4' set there are some people walking in the room, the 'Spaan 1' set is taken during evening hours, and the 'Biron 1' set is taken in a feature poor home. The ground truth robot poses for the home sets were obtained by applying the SLAM algorithm described in [28] on the laser scans and odometry. For the 'Office' set they were obtained by positioning the robot by hand using a small laser beam for accurate orientation.

2) Setup: From every dataset we use every pair of images. We discard the images taken at the same position. Also, if images are taken at more than 5 meters apart for the Almere 4 or more than 3 meters apart for the other sets, then the chance of finding point correspondences is small, so we also discard these pairs. Still, for each set there are around 10^6 image pairs left.

To extract point correspondences from the image pairs, the SIFT algorithm is used [6]. First omnidirectional images are mapped to panoramic images [30], from which the SIFT feature points are found. These features are described by the standard SIFT descriptor of 128 dimensions. If two features in the same image have a small distance in descriptor space then they are removed. A set of point correspondences between two images is determined by applying the standard matching scheme as described in [6]. This resulted in on average 25 matches per image pair. The groundtruth relative pose was computed from the groundtruth robot positions.

3) Sensitivity to mismatches, trained with simulated data: We first use a LUT constructed using the simulator described in Section IV-A.1 and see how well it compares to state of the art methods. In order to evaluate the performance of the methods we would like to vary the number of mismatches. This can not be controlled in real data, therefore we made subsets of the data on the basis of the distance between the poses. We assume that for larger distances it is more difficult to find matching features. In Figure 6(b) and 6(e) the heading and rotation error of the different methods is plot as a function of the distance between the images for dataset 'Almere 4'. It is clear that on a whole the errors are much larger than was the case for the simulation data. This is partly due to the fact that some of the views were obstructed by furniture, walls or people walking in the environment.

In the plot of the heading error (Figure 6(b)) one can see that the RANSAC+M-Estimator combined with the Two point algorithm is outperformed by the Three point algorithm version, which in turn is clearly outperformed by the novel LUT method for distances larger than 1.5 meters. The accuracy of the RANSAC+M-Estimator combined with the Eight point algorithm is not that bad. This could have been caused by the fact that the robot was leaning over when accelerating, slightly violating the planarity constraint. For the rotation error (Figure 6(e)), the improvement of the ML estimator over the



(a) Heading errors, simulation data

(b) Heading errors, Almere 4

(c) Heading errors, Spaan 1

(d) Rotation errors, simulation data

(e) Rotation errors, Almere 4

(f) Rotation errors, Spaan 1

Fig. 6. Comparison of LUT ML, trained using a simulator or real images and RANSAC+M-Estimator combined with different algorithms on the simulation dataset for different number of mismatches and Almere set 4 and Spaan 1 for different distances between the image pairs. The MAD is used to draw confidence intervals of the medians.

RANSAC+M-Estimator with Three Point is less clear.

4) Sensitivity to mismatches, trained with real data: Next we constructed a look up table using all the image pairs of the Almere 4 set that were within a 5 meter distance and applied it on the Spaan 1 set. We used two different binsizes, the first had 128 bins for all three dimensions and the other 16. The Maximum Likelihood solutions based on these two tables were compared to the RANSAC+M-Estimator combined with the Three point algorithm and solutions given two LUTs based on the simulator, also with dimensions 128 and 16.

In Figure 6(c) and 6(f) the results are shown. As can be seen the overall accuracy is less than for the Almere 4. A reason for this could be the motion blur, caused by the bad illumination of this dataset. The LUT ML with 128 bins based the simulator perform best, followed by the 128 bins LUT based on the real images. Probably this is due to the limited number of point correspondences in the real image set. For the much smaller LUTs with 16 bins this seems to be less problematic, visible by the improvement of the LUT based on real images over the one based on simulated data.

5) Averages over the data sets: Application on other datasets resulted in comparable errors. Table I summarizes these results.

6) Binsize vs CPU time: To evaluate the influence of different binsizes for the look up table, we tested the ML method for different numbers of bins. In Table II the average computational time in milliseconds is given per image pair for the 'Almere 4' set (other datasets showed similar trends). As can be seen small look up tables result in a large speed up, but this comes at the cost of more error (see Table I). The RANSAC+M-Estimator combined with the Planar three point algorithm is three times slower than the LUT method with 128^3 bins.

TABLE II

AVERAGE COMPUTATIONAL TIME USAGE PER RELATIVE POSE ESTIMATE IN MILLISECONDS FOR THE DIFFERENT METHODS.

ML				RANS+M-Est		
128	64	32	16	8pt	3pt	2pt
1.3	0.28	0.07	0.036	3.6	3.8	0.68

V. DISCUSSION

An important advantage of the LUT based pose estimator is that it provides a full likelihood over the descritized space of possible relative poses. Thus, appart from computing a Maximum Likelihood solution, as shown in the Experiments section, this could make the method useful for a range of other applications.

A nice illustration of the usefulness of a full likelihood can be seen in Figure 7. It shows an example likelihood for a typical situation that occurred in the Almere 4 set. In this case the robot did not move forward but rotated on the spot. Thus the heading of the translation of the robot, ϑ, is not determined. This is correctly reflected by the estimated

TABLE I

Performance of the tested pose estimation methods for the different data sets

		Office	Almere 4	Spaan 1	Biron 1
LUT ML sim 128	head err	0.022 (.0026)	0.548 (.0031)	0.538 (.0027)	0.813 (.0063)
	rot err	0.000 (.0000)	0.399 (.0017)	0.469 (.0063)	0.578 (.0029)
LUT ML sim 64	head err	0.000 (.0000)	0.597 (.0025)	0.585 (.0071)	0.850 (.0025)
	rot err	0.000 (.0000)	0.425 (.0044)	0.499 (.0029)	0.602 (.0039)
LUT ML sim 32	head err	0.000 (.0000)	0.669 (.0030)	0.664 (.0026)	0.912 (.0031)
	rot err	0.000 (.0000)	0.459 (.0027)	0.537 (.0019)	0.627 (.0030)
LUT ML sim 16	head err	0.071 (.0000)	0.786 (.0027)	0.771 (.0054)	0.975 (.0041)
	rot err	0.000 (.0000)	0.503 (.0013)	0.585 (.0040)	0.660 (.0049)
LUT ML real 128	head err	0.049 (.0000)	0.711 (.0035)	0.612 (.0059)	0.843 (.0031)
	rot err	0.000 (.0000)	0.401 (.0006)	0.449 (.0026)	0.502 (.0034)
LUT ML real 64	head err	0.027 (.0000)	0.717 (.0016)	0.616 (.0037)	0.844 (.0046)
	rot err	0.000 (.0000)	0.402 (.0011)	0.451 (.0058)	0.492 (.0047)
LUT ML real 32	head err	0.000 (.0177)	0.730 (.0036)	0.630 (.0060)	0.857 (.0032)
	rot err	0.000 (.0000)	0.410 (.0031)	0.469 (.0063)	0.500 (.0041)
LUT ML real 16	head err	0.071 (.0177)	0.766 (.0045)	0.684 (.0063)	0.881 (.0028)
	rot err	0.000 (.0000)	0.433 (.0021)	0.503 (.0032)	0.529 (.0020)
R+M 8pt	head err	0.031 (.0015)	0.957 (.0043)	0.925 (.0053)	1.244 (.0069)
	rot err	0.008 (.0013)	1.220 (.0027)	1.212 (.0026)	1.389 (.0100)
R+M 3pt	head err	0.028 (.0017)	0.677 (.0035)	0.585 (.0068)	1.101 (.0288)
	rot err	0.007 (.0010)	0.430 (.0032)	0.512 (.0051)	1.068 (.0089)
R+M 2pt	head err	0.040 (.0034)	0.903 (.0034)	0.778 (.0062)	1.371 (.0197)
	rot err	0.012 (.0014)	0.563 (.0041)	0.619 (.0049)	1.038 (.0181)

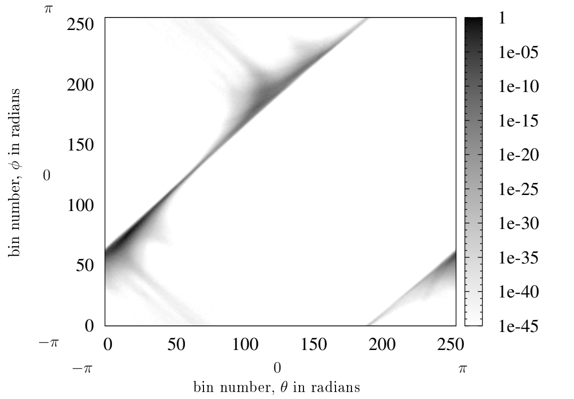


Fig. 7. The log posterior computed from the point correspondences between image 857 and 897 of Almere set 4. The distance between these camera positions was about 1 cm, while the rotation was 90 degrees.

posterior. The rotation, on the other hand, can be determined. This can be seen by the diagonal relationship between ϑ and ϕ in the posterior.

The proposed method could be readably applied in particle filter based robotA localization schemes [31] where each hypothesized robot pose can be weighted by the likelihood given newly acquired images. Also, geometric SLAM could benefit from the proposed method, because the uncertainty of the Maximum Likelihood can be estimated easily from the full likelihood. For example by fitting a Von Mises or mixture of Von Mises distributions on the descritized likelihood space [26].

Another task that is very much suited for the proposed example is that of topological mapping. State of the art topological mapping approaches use proper probabilistic data association techniques to compare pairs of images [1]. However, in addition they commonly apply ad hoc rules to check whether the matched point correspondence fit in a certain local geometry, by computing the relative pose. Because of the probabilistic nature of the proposed method, it is straightforward to combine it with these proper data association techniques, ending up in a fully probabilistic topological mapping method.

VI. Conclusion

In this paper we propose a novel approach to solve planar relative pose estimation from image point correspondences. We have shown the advantage of discretizing and analyzing the complete solution space, which is in the planar motion case 2 dimensional. Probabilistic methods were proposed that learn the likelihood over this space from a training set of representative images. Experiments on challenging image sets acquired in real homes showed a 20% increase in accuracy with respect to state of the art methods consisting of a planar constrained RANSAC and M-Estimators.

In addition an efficient technique was presented for building a concise look up table of the likelihood, reducing the estimation process to simple look ups. Computing a full likelihood given two images costs as little as 36 microsecond, as compared to the 3 milliseconds RANSAC uses. This could even be improved upon, for example, by using a multi-resolution approach as described in [32], in which a small look up table is used to isolate candidate areas for the ML solution, which are then investigated further using a bigger look up table. Another possibility is implementing the method on a GPU, which can much more quickly manipulate 2D histograms.

Continuing research will focus on taking advantage of the full likelihood that is estimated by the method. We foresee improvements in both geometric SLAM as in topological mapping, especially when it comes to uncertainty estimation.

ACKNOWLEDGMENTS

We would like to thank Gwenn Englebienne for fruitful discussions and contributions to this paper.

REFERENCES

[1] M. Cummins and P. Newman, “Highly scalable appearance-only SLAM - FAB-MAP 2.0,” in *Proceedings of Robotics: Science and Systems*, Seattle, USA, June 2009.

[2] C. Valgren and A. J. Lilienthal, “Incremental spectral clustering and seasons: Appearance-based localization in outdoor environments,” in *ICRA*. Pasadena, California, USA: IEEE, May 2008, pp. 1856–1861.

[3] R. Eustice, “Large-area visually augmented navigation for autonomous underwater vehicles,” Ph.D. dissertation, MIT - Woods Hole Oceanographic Institute, June 2005.

[4] H. Andreasson, , T. Duckett, and A. J. Lilienthal, “A minimalistic approach to appearance based visual slam,” *Transactions on Robotics*, vol. 24, no. 6, pp. 991–1001, October 2008.

[5] F. Fraundorfer, C. Engels, and D. Nister, “Topological mapping, localization and navigation using image collections,” in *IROS*. San Diego, USA: IEEE/RSJ, November 2007, pp. 3872–3877.

[6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[7] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision, second edition*. Cambridge University Press, 2003.

[8] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Com. of the ACM*, vol. 24, no. 6, 1981.

[9] P. H. S. Torr and D. W. Murray, “The development and comparison of robust methods for estimating the fundamental matrix,” *Int. Journal of Computer Vision*, vol. 24, no. 3, pp. 271–300, 1997.

[10] P. V. C. Hough, “Method and means for recognizing complex patterns,” 1962, u.S. Patent 3,069,654.

[11] H. C. Longuet-Higgins, “A computer algorithm for reconstructing a scene from two projections,” *Nature*, vol. 293, pp. 133–135, Sept. 1981.

[12] D. Nistér, “An efficient solution to the five-point relative pose problem,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 756–777, 2004.

[13] B. C. Matei, “Heteroscedastic errors-in-variables models in computer vision,” Ph.D. dissertation, Rutgers University, 2001.

[14] H. Wang and D. Suter, “Robust adaptive-scale parametric model estimation for computer vision,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1459–1474, 2004.

[15] R. den Hollander and A. Hanjalic, “A combined RANSAC-Hough transform algorithm for fundamental matrix estimation,” in *18th British Machine Vision Conference*. University of Warwick, UK, 2007.

[16] D. J. Heeger and A. D. Jepson, “Subspace methods for recovering rigid motion i: algorithm and implementation,” *Int. J. Comput. Vision*, vol. 7, no. 2, pp. 95–117, 1992.

[17] A. Censi and S. Carpin, “HSM3D: Feature-less global 6DOF scan-matching in the Hough/Radon domain,” in *ICRA*. Kobe, Japan: IEEE, May 2009.

[18] M. Pollefeys, D. Nistér, J. M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénius, R. Yang, G. Welch, and H. Towles, “Detailed real-time urban 3d reconstruction from video,” *Int. J. Comput. Vision*, vol. 78, no. 2-3, pp. 143–167, 2008.

[19] M. Milford and G. Wyeth, “Mapping a suburb with a single camera using a biologically inspired SLAM system.” *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1038–1053, 2008.

[20] P. Lamon, I. Nourbakhsh, B. Jensen, and R. Siegwart, “Deriving and matching image fingerprint sequences for mobile robot localization,” in *ICRA*, Seoul, Korea, May 2001.

[21] M. Brooks, L. de Agapito, D. Huynh, and L. Baumela, “Towards robust metric reconstruction via a dynamic uncalibrated stereo head,” *Image Vision Comput.*, vol. 16, no. 14, pp. 989–1002, 1998.

[22] D. Ortín and J. M. M. Montiel, “Indoor robot motion based on monocular images,” *Robotica*, vol. 19, no. 3, pp. 331–342, 2001.

[23] J. Kosecká, F. Li, and X. Yang, “Global localization and relative positioning based on scale-invariant keypoints.” *Robotics and Autonomous Systems*, vol. 52, no. 1, pp. 27–38, 2005.

[24] T. Goedemé, T. Tuytelaars, G. Vanacker, M. Nuttin, and L. V. Gool, “Feature based omnidirectional sparse visual path following,” in *IROS*. Edmonton, Canada: IEEE/RSJ, August 2005, pp. 1003–1008.

[25] A. J. D. J. Civera and J. M. M. Montiel, “Inverse depth parametrization for monocular slam,” *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 932–945, October 2008.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. Springer, 2006.

[27] D. H. Stewénius, C. Engels, and D. D. Nistér, “Recent developments on direct relative orientation,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 4, pp. 284–294, 2006.

[28] J. Folkesson, P. Jensfelt, and H. Christensen, “Vision SLAM in the measurement subspace,” in *ICRA*. Barcelona, Spain: IEEE, April 2005, pp. 30–35.

[29] Z. Zivkovic, O. Booij, B. Kröse, E.Topp, and H.I.Christensen, “From sensors to human spatial concepts: an annotated dataset,” *IEEE Transactions on Robotics*, vol. 24, no. 2, pp. 501–505, April 2008.

[30] R. Bunschoten, “Mapping and localization from a panoramic vision sensor,” Ph.D. dissertation, University of Amsterdam, November 2003.

[31] H.-M. Gross and A. Koenig, “Robust omniview-based probabilistic self-localization for mobile robots in large maze-like environments.” in *ICPR (3)*, 2004, pp. 266–269.

[32] E. B. Olson, “Real-time correlative scan matching,” in *ICRA*. Kobe, Japan: IEEE, May 2009.