

# Hough-transform based map stitching after localization uncertainty

Okke Formsma  
University of Amsterdam

August something, 2012

# Abstract

Abstract

# Acknowledgements

Thanks Arnoud, etc. Thanks Julian? Thanks Magda?

# Contents

<b>Abstract</b>	<b>1</b>
<b>1 Introduction</b>	<b>2</b>
1.1 Background . . . . .	2
1.1.1 USAR Sim . . . . .	2
1.1.2 State . . . . .	2
1.1.3 Gaussian state representation . . . . .	4
1.1.4 Scanmatching . . . . .	4
1.1.5 ManifoldSLAM . . . . .	6
1.1.6 2D slam in 3D environment . . . . .	6
1.1.7 Uncertainty measure . . . . .	6
<b>2 Map Stitching with the Hough Transform</b>	<b>7</b>
2.1 Hough transform . . . . .	7
2.2 Hough spectrum . . . . .	8
2.3 Finding rotation $\theta$ . . . . .	9
2.3.1 Periodicity . . . . .	10
2.4 Finding translation $t$ . . . . .	11
2.5 Evaluating stitching results . . . . .	11
2.6 Variations on the Hough-transform based stitching . . . . .	11
<b>3 Method</b>	<b>12</b>
3.1 Map Stitching . . . . .	12
3.2 SLAM confidence measure . . . . .	12
3.3 Inertial sensor data . . . . .	12
3.4 Confidence measures . . . . .	12
<b>4 Future work</b>	<b>15</b>

# Chapter 1

## Introduction

The Simultaneous Localization And Mapping (SLAM) method employed by the AJOFF (Amsterdam-Oxford Joint Rescue Forces team [cite team description paper]) is not robust against sensor failures, which occur for example when the laser scanner is severely tilted.

In this report, we propose a novel method which splits a map when the localization uncertainty exceeds a certain threshold. The different pieces of the map are then stitched together using a Hough transform based map stitching method.

### 1.1 Background

#### 1.1.1 USAR Sim

The Urban search and rescue challenge

What's wrong with SLAM

Why Map Stitching through hough transform One of the goals of the USAR challenge was to create a useable map of the environment.

The problem an agent faces when it needs to find it's way in an unknown environment is called SLAM, for Simultaneous Localization and Mapping. The agent has no map of the environment, and also no way of knowing it's exact location. It needs to infer the map and it's position on the map from the information it gathers from it's sensors throughout time.

Thus, SLAM is a chicken-and-egg problem: without a map it is hard to estimate the agent's position, and without the agent's position it's hard to estimate the map!

In this section we will first examine how the robot keeps track of the state of the world, and then how the SLAM problem can be solved with Extended Kalman Filters.

#### 1.1.2 State

The state of the world as it is known to the agent is denoted  $x$ . The value of this set of variables may change over time, as the agent collects sensor data from the environment. The state at time  $t$  is denoted  $x_t$ .

Many variables may be included in the state variable  $x$ . For the purpose of this work, we will assume that it contains at least the agent's *pose and a map* of the environment.

In most SLAM approaches, two types of sensor data is collected: *environmental measurement data and control data*. Environmental measurement data collects information about the environment, while control data collects information about the robot within the environment. Environmental measurement data is denoted  $z$  (at a specific point in time  $z_t$ ). Examples of environmental sensors available to an agent are laser scanners, sonars and video cameras. Control data and GPS signals. Control data sensors collect information intrinsic to the agent: it's velocity and position. Examples of control sensors are odometry sensors, inertia sensors and global positioning systems.

For the purpose of this paper we are mainly interested in the online SLAM problem, in contrast to the full SLAM problem. Online (or incremental) SLAM seeks to estimate the current pose  $x_t$  and map  $m$ :

$$p(x_t, m | z_{1:t}, u_{1:t}) \quad (1.1)$$

In contrast, the full SLAM problem estimates all poses  $x$  and map  $m$ :

$$p(x_{1:t}, m | z_{1:t}, u_{1:t}) \quad (1.2)$$

In practice, the difference is that the full SLAM problem reconsiders the location of all previous poses in estimating the map, while online SLAM treats these as given. The full SLAM problem is computationally much more expensive. Thus for time sensitive applications, such as real time robot control, usually only online SLAM is performed.

The robot's knowledge of it's current state is reflected in the *belief* function. The belief function gives a posterior probability over the state variable. The belief at time  $x_t$  is defined as follows:

$$bel(x_t) = p(x_t | z_{1:t}, u_{1:t}) \quad (1.3)$$

This belief is usually updated in two steps: once when the control data ( $u_t$ ) is received, and again when new measurement data ( $z_t$ ) is received. The belief of the robot after the control update is called the prediction, denoted  $\overline{bel}(x_t)$ :

$$\overline{bel}(x_t) = p(x_t | z_{1:t-1}, u_{1:t}) \quad (1.4)$$

A general algorithm for calculating beliefs is the *Bayes filter*. It first calculates the prediction  $\overline{bel}(x_t)$  by integrating out the possible values of the previous belief  $x_{t-1}$  from previous belief  $bel(x_{t-1})$  and the expectation of the current state given the previous state and control data  $p(x_t | u_t, x_{t-1})$ .

$$\overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1} \quad (1.5)$$

The final step is the measurement update, in which the measurement data  $z_t$  is incorporated:

$$bel(x_t) = \eta p(z_t | x_t) \overline{bel}(x_t) \quad (1.6)$$

In the previous formula,  $\eta$  is a normalization constant which ensures that  $bel(x_t)$  is a proper probability.

### 1.1.3 Gaussian state representation

It is often convenient to represent the belief of the current state as a multivariate Gaussian. A Gaussian is defined by its mean  $\mu$  and covariance  $\Sigma$ :

$$p(x) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \right\} \quad (1.7)$$

Gaussian state representations have the advantage of being easily understood and straightforwardly manipulated. This comes with a cost: a Gaussian is unimodal – there is only one maximum (the mean) possible. This makes Gaussian representations ill suited when there are many different hypothesis distributed across a map.

In UsarCommander, the state is represented by a 3-dimensional vector:  $(x, y, \theta)$  and the corresponding 3-by-3 covariance matrix  $\Sigma$ .  $x$  and  $y$  are the respective coordinates of the robot in a global frame, while  $\theta$  is the rotation angle of the robot with respect to the  $y$  axis.

When the state is represented by a Gaussian, the formulas in the previous section (1.1.2) can be relatively easily calculated.

### 1.1.4 Scanmatching

A robot has access to a variety of sensors, as discussed in section 1.1.2. Many sensors are not very accurate and reliable: odometry sensors yield erroneous values when the robot’s wheels spin on a slippery slope, GPS sensors are unreliable indoors, and inertia sensors experience ‘drift’ – they need to be calibrated every once in a while. Laser range sensors provide an alternative. It measures with high precision the distance to the first object in a certain direction.

Laser range sensors are often put in a rotating device, which measures the distance to objects on a (often horizontal) plane. A field of view of  $180^\circ$  with a  $1^\circ$  resolution is common, which yields 181 distance scans. The 181 scans form a point cloud of 181 points. See image ??.

The ‘scans’ that the laser range sensor returns are processed through a scan-matching algorithm. These algorithms compare two scans (or a scan and a map) to see how they relate to each other. The scanmatcher needs to find out the relative change in orientation and position between the two scans.

In the context of this paper, we will focus on the Weighted Scan Matcher, but various other approaches exist. Slamet and Pfingsthorn[?] provide an excellent overview of various scanmatching algorithms such as Normal Distribution Transform (NDT), Iterative Dual Correspondence (IDC), Metric-based Iterative Closest Point (MbICP) and ICP.

#### Weighted Scan Matcher

The Weighted Scan Matcher (WSM) was introduced by Pfister et al [?]. In contrast to other scan matchers, the influence of a correspondence between two measurements is weighted with the certainty of how well the two points match. When two point clouds align perfectly, the uncertainty of these matches is zero. This will not generally be the case. Points that have less associated uncertainty will weigh more in the final scan matching result. The Weighted Scan Matcher explicitly models three error sources: the correspondence error, measurement error, and bias. See figure 1.1.

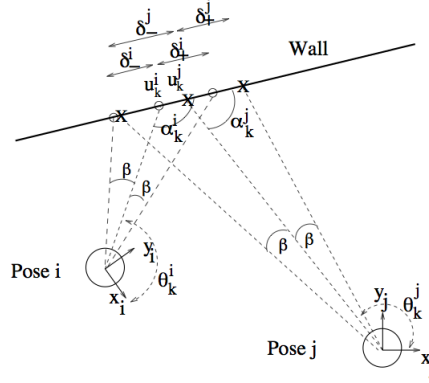


Figure 1.1: Geometry of the scan matching process. Courtesy of [?].

The correspondence error results from the limited resolution available to laser scanners. The resolution in this case is not the resolution in distance measurement, but the angular resolution of the scanner. At a distance of 5 meter and an angular resolution of  $1^\circ$ , the two measurements lie almost 9 centimeters apart ( $500 * \tan(1^\circ) \approx 8.7 \dots$ ). When the robot scans again after rotating  $0.5^\circ$ , all measurements will lie in between two previous measurements, and the best match will be off a few centimeters, depending on the distance from the scanner. Even worse: a thin object, such as a iron pole, can be easily missed by the scanning beams as well. All of these errors are examples of the correspondence error.

The measurement error models the error in the measurement process itself. This can be because of any external factors which make the measurement erroneous.

Finally, the bias incorporates the difference between measurement methods. For example, laser scanners and sonar have different performance characteristics. The bias term compensates for this.

In general, the error between two scans  $a_i$  and  $b_i$  can be expressed as a rotation matrix  $R$  and translation  $t$ :

$$\epsilon = a_i - R \cdot b_i - t \quad (1.8)$$

With the assumption that all noise is Gaussian, measurement  $a_i$  can be expressed as the addition of the measurement error  $\delta r_{a,i}$  and bias error  $d_{a,i}$  to the actual distance  $r_i$  (similar for  $b_i$ ):

$$a_i = r_{a,i} + \delta r_{a,i} + d_{a,i} \quad (1.9)$$

By substituting equation 1.9 into equation 1.8, the full error term is obtained:

$$\begin{aligned} \epsilon &= (r_{a,i} + \delta r_{a,i} + d_{a,i}) - R \cdot (r_{b,i} + \delta r_{b,i} + d_{b,i}) - t \\ &= \underbrace{(r_{a,i} - R \cdot r_{b,i} - t)}_{\text{correspondence error}} + \underbrace{(\delta r_{a,i} - R \cdot \delta r_{b,i})}_{\text{measurement error}} + \underbrace{(d_{a,i} - R \cdot d_{b,i})}_{\text{bias}} \end{aligned} \quad (1.10)$$

In an application where only one range measurement system is employed (such as a specific type of laser range scanner), the bias error term can be



dropped. (Remember, the bias only models differences between measurement methods). In addition, the measurement error with modern laser scanners is so small that the measurements can be taken as a good approximation of the real measurement ([?]: “For practical computation, we can use [measurements]  $\theta_k^i$  and  $l_k^i$  as a good estimates for the quantities [actual values]  $\Theta_k^i$  and  $L_k^i$ .”). Thus, when the sensor error is small enough, we can safely ignore this source of error as well.

This simplifies equation 1.10 further, so it contains only the correspondence error:

$$\epsilon = r_{a,i} - R \cdot r_{b,i} - t \quad (1.11)$$

This error term  $\epsilon$  is minimized through an iterative algorithm in which the rotation and translation are estimated in turn. When the algorithm is suitably converged, it terminates. The full covariance matrix for the scan match is computed from the individual correspondence errors. We will see how this covariance matrix comes in handy for the SLAM algorithm and error estimation procedures in later chapters.

### 1.1.5 ManifoldSLAM

ManifoldSLAM was developed by Bayu Slamet and Max Pfingsthorn in 2006 [?] to compete in the RoboCup Urban Search and Rescue Simulation and won first place. It is based on the Manifold data structure by Howard et al. [?]. The world/map representation is

Many SLAM algorithms use a bayesian filter to refine the state by incorporate the current knowledge into a new assessment of the current state of the robot. The SLAM algorithm in use by the AJORF team takes the map  $m_{t-1}$  and previous position  $x_{t-1}$  as given.

### 1.1.6 2D slam in 3D environment

Only a slice of the world is returned - so we need to make sure we are horizontal. Otherwise, the scanner returns wrong readings.

### 1.1.7 Uncertainty measure

UNCERTAINTY MEASURE BY ARNOUD: To see if the reduced correlation distance also improves the robustness of the scan matcher the uncertainty measures returned by the algorithms were analyzed. This uncertainty measure is the full 3-by-3 covariance matrix of the Gaussian distribution over the displacement estimate and does not lend itself well for charting in its original form. Therefore we took the on-diagonal elements that describe the independent uncertainty in x and y direction and combined these into a Euclidean distance measure. The trace of the covariance matrix acquired with Q-WSM is in most cases (94average uncertainty of the incremental version.

## Chapter 2

# Map Stitching with the Hough Transform

Why hough based? Alternatives?

[what is map stitching, why is it necessary again?]

Stefano Carpin presented a novel map stitching approach based on the Hough Transform in [?]. It is an extension of the Hough transform which Censi et al. introduced as the *Hough spectrum* [?]. Censi et al. used the hough spectrum for a novel scan matcher.

Remember from the SLAM section (??) that aligning two maps takes two parameters: the rotation angle  $\theta$  and translation vector  $t = [x, y]^T$ . The Hough spectrum is used for finding the rotation  $\theta$  between a scan and reference map. The implementations of Carpin and Censi et al. diverge in their method of finding the translation vector  $t$  between two scans. First, we'll look at what the Hough transform and Hough spectrum are. Then, we'll take a look at how the best matching rotation  $\hat{\theta}$  is found, and finally we'll examine translation  $t$  further.

### 2.1 Hough transform

The Hough transform is in origin a line detection technique, patented by Paul Hough in 1962[?]. An improved version, which is presented here, was introduced by Duda and Hart[?]. The Hough transform is also extended to detect circles, ellipsis and parabolas, and also arbitrary 2D shapes[?].

The Hough transform detects lines in the following manner. A line in Cartesian  $(x, y)$  space can be defined by it's angle ( $\theta$ ) and distance from the origin  $\rho$ :

$$x \cos \theta + y \sin \theta = \rho \quad (2.1)$$

Every point in  $(x, y)$  space lies on infinitely many lines satisfying equation 2.1. Every angle  $\theta$  has a single accompanying distance from the origin  $\rho$ . Thus, a point in  $(x, y)$  space can be represented in the *Hough domain*  $(\theta, \rho)$  as a curve. The locations where many curves intersect in the Hough domain

correspond to the lines in the original image; remember that lines correspond to points in the Hough domain.

Finding the intersection of many curves in the continuous domain is computationally expensive. The discrete Hough transform mitigates this problem by discretizing the Hough domain into bins. Every point in the original image ‘votes’ for each bin it belongs to in  $(\theta, \rho)$  representation. The strongest lines stand out as bins with a high number of votes.

There are  $n_\theta$  angle-bins, and  $n_\rho$  distance bins in discrete Hough transform  $\mathcal{HT}$ .

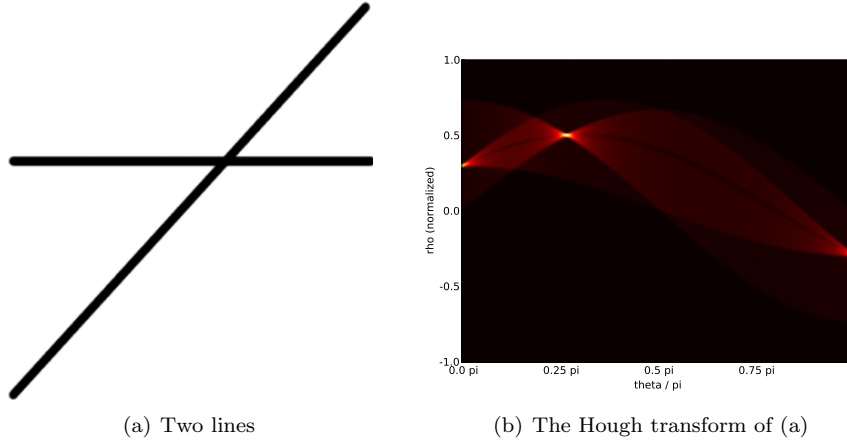


Figure 2.1: Example of the discrete Hough transform. Notice the two maxima in (b) which correspond to the two lines in (a). There seem to be three maxima (one at 0, one at  $\frac{\pi}{4}$  and one at  $\pi$ ). However, the maxima at 0 and  $\pi$  are the same maximum: see section 2.3.1

## 2.2 Hough spectrum

The major contribution of Censi et al.[?] is the introduction of the Hough spectrum. The Hough spectrum ( $\mathcal{HS}$ ) is a measure for the direction of lines in an image. The discrete Hough transform finds the specific lines in the image, while the Hough spectrum finds the most pronounced *directions* of lines in the image:

$$\mathcal{HS}(k) = \eta \sum_{i=1}^{n_\rho} \mathcal{HT}(i, k)^2 \quad 1 \leq k \leq n_\theta \quad (2.2)$$

$\eta$  is a normalization value to limit  $\mathcal{HS}(k)$  to the domain  $[0, 1]$ .

As we’ve seen in the previous section, each bin in the discrete Hough transform ( $\mathcal{HT}(i, k)$ ) contains the number of pixels that lie on a line defined by  $(\theta_i, \rho_k)$ . The hough spectrum is a measure for how strong the lines are on a specific angle.

Thus, the Hough spectrum yields the highest value in the direction in which the most pronounced lines run. In an indoor environment, there will usually be two peaks, separated by a  $90^\circ$  angle, corresponding to grid-like structure most

buildings are created (see figure 2.2). Incidentally, in a bee-hive, with hexagonal chambers, we would see 3 peaks each separated by  $60^\circ$  angles. See figure 2.3.

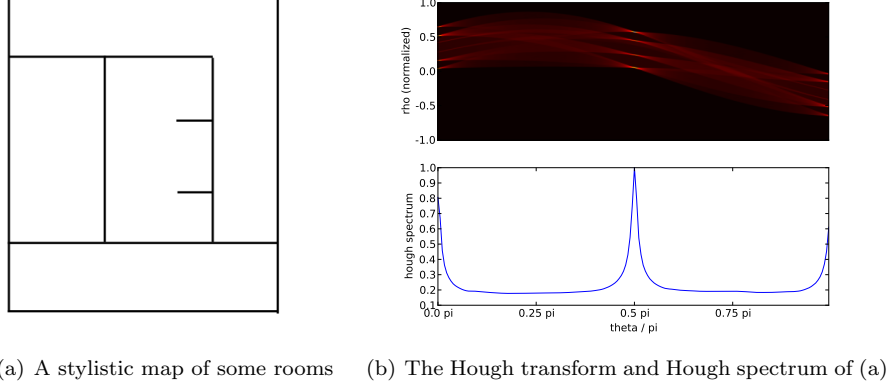


Figure 2.2: Example of a hough spectrum in a human-made environment.

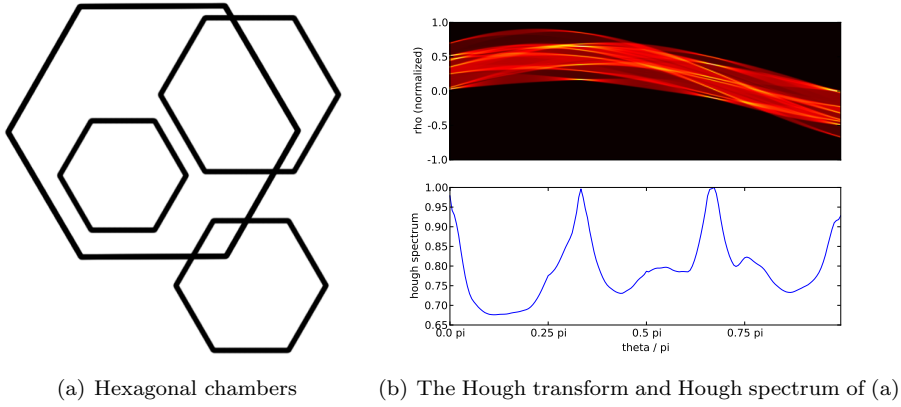


Figure 2.3: Example of the Hough spectrum in a hexagonal environment. Notice the three maxima in the Hough Spectrum.

## 2.3 Finding rotation $\theta$

Now we can calculate the Hough spectra  $\mathcal{HS}_{M_1}$  and  $\mathcal{HS}_{M_2}$  for maps  $M_1$  and  $M_2$ .

Notice that the spectra have a periodicity of  $180^\circ$ , and have similar peaks. However, the peaks are at different angles. The cross-correlation of the two signals shows the similarity of the two spectra as one of them is shifted over the x-axis. The cross-correlation  $\mathcal{CC}_{M_1, M_2}$  is calculated as follows:

$$\mathcal{CC}_{M_1, M_2}(k) = \sum_{i=1}^{n_\theta} \mathcal{HS}_{M_1}(i) \times \mathcal{HS}_{M_2}(i+k) \quad 1 \leq k \leq n_\theta \quad (2.3)$$

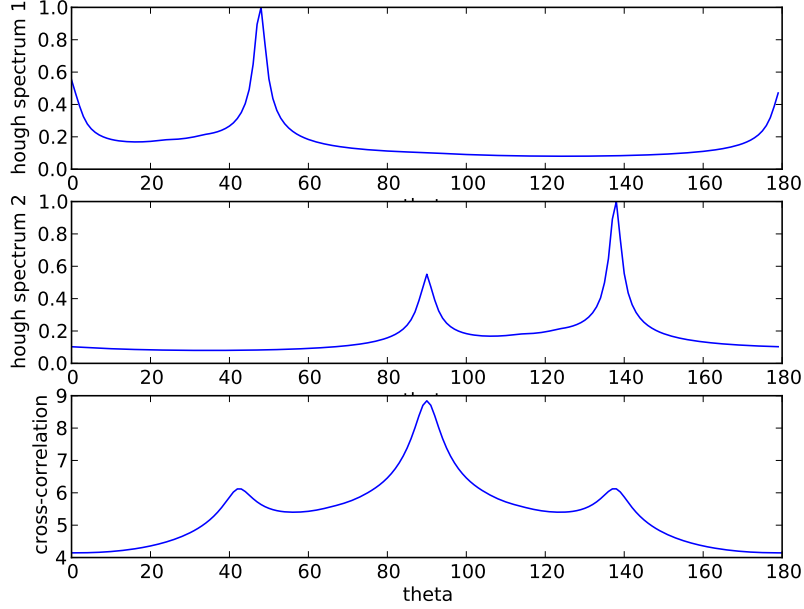


Figure 2.4: The top two plots are the hough spectra of image 2.1(a). For the second hough spectrum, it was rotated  $90^\circ$  counter-clockwise. The cross-correlation shows that the most probable rotation is indeed  $90^\circ$ , with smaller peaks for  $45^\circ$  and  $135^\circ$ .

### 2.3.1 Periodicity

Why are do the Hough spectra have a periodicity of  $180^\circ$ ? The answer lies in the following equations:

$$x \cos \theta + y \sin \theta = \rho \quad (2.4)$$

$$x \cos(\theta + \pi) + y \sin(\theta + \pi) = -\rho \quad (2.5)$$

Thus, the following equation holds:  $\mathcal{HT}(\theta_\alpha, \rho_\alpha) = \mathcal{HT}(\theta_\alpha + \pi, -\rho_\alpha)$ . In turn, the Hough spectrum of  $\theta_\alpha$  and  $\theta_\alpha + \pi$  are the same, because  $\mathcal{HS}$  is summed over all  $\rho$ .

Thus, the domain of  $\theta$  only needs to be  $[0, \pi)$ , whereafter it wraps back to itself. Prior work by Jankowska used the domain  $[0, 2\pi)$  [?], which is thus superfluous.

One must take care when using this representation to remember the possible alternative hypothesis for  $\theta$ . Every solution  $\hat{\theta}$  found by the cross-correlation method has an additional solution  $\hat{\theta} + \pi$ !

## 2.4 Finding translation $t$

The translation  $t$

## 2.5 Evaluating stitching results

## 2.6 Variations on the Hough-transform based stitching

\* Assumption: indoor environment \* Can be used for re-initialization of the manifold method when position is lost! Awesome if a full scan matcher is used after the hough transform based method.

# Chapter 3

## Method

### 3.1 Map Stitching

Map stitching prior work, magda etc

### 3.2 SLAM confidence measure

The confidence measure based on the determinant of the covariance matrix of the pose uncertainty. The hypothesis is that if this measure is larger, then the confidence is lower (the uncertainty is bigger).

The determinant of the covariance matrix is used as metric:

is the sample variance-covariance matrix for observations of a multivariate vector of  $p$  elements. The determinant of  $D$ , in this case, is sometimes called the generalized variance. (<http://www.itl.nist.gov/div898/handbook/pmc/section5/pmc532.htm>)

### 3.3 Inertial sensor data

Gives the rotations on three axis, enables you to detect

### 3.4 Confidence measures

cite bayu and max: Good examples of single confidence values that can be derived from a covariance matrix are the determinant  $\det()$  or the trace  $\text{trace}()$ . The relations lend themselves very well for use in path-planning and search algorithms. In addition, heuristic algorithms can take advantage of the uncertainty information stored on the links. (P61)

(paper not found): L. Carlone, M. Kaouk Ng, J. Du, B. Bona, and M. Indri, Reverse KLD-Sampling for measuring uncertainty in Rao-Blackwellized Particle Filters SLAM, in Proc. of the 2009 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Workshop on Performance Evaluation and Benchmarking for Next Intelligent Robots and Systems, 2009.

Carlone et al. use a particle filter based representation of their belief.

In this work we examine a few different confidence measures: 1) attack angle 2) determinant of match 3) trace of match

From SlametPfungstorn:

To asses the quality of fit of a new range scan, the same notion of being well-fitted is used as during the construction of local sub-maps (refer to Section 5.3.3). Recall that we already obtained the covariance matrix  $\Sigma$  of the Gaussian distribution that was estimated by the scan matcher. The Fisher information matrix  $I$  is computed by taking the inverse of this covariance matrix  $I = \Sigma^{-1}$  and we compute the 95



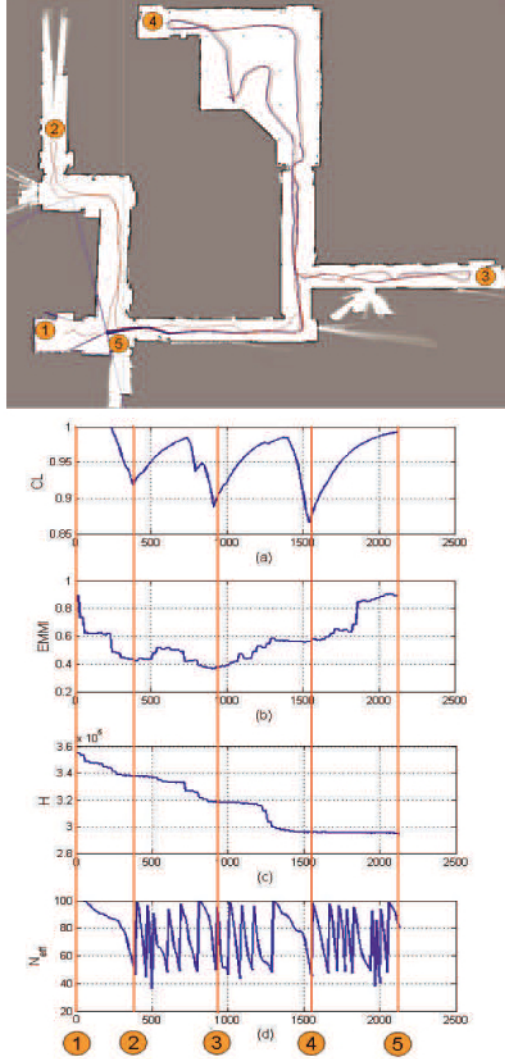


Fig. 4: Time-history of the different uncertainty estimators obtained in real experiment performed in Politecnico di Torino (100 particles): (a) CL, (b) *EMMI*, (c) Map Entropy  $H$  and (d) The effective sample size  $N_{eff}$ .

Figure 3.1:

## Chapter 4

### Future work

Use the quad-tree based scanmatcher with a guessed initial position over the entire map.

# Bibliography