Published in final edited form as:

Auton Robots. 2009 November ; 27(4): . doi:10.1007/s10514-009-9149-4.

Using virtual scans for improved mapping and evaluation

Rolf Lakaemper and

Temple University, Philadelphia, PA, USA

Nagesh Adluru

University of Wisconsin, Madison, WI, USA

Rolf Lakaemper: lakemper@temple.edu

Abstract

In this paper we present a system to enhance the performance of feature correspondence based alignment algorithms for laser scan data. We show how this system can be utilized as a new approach for evaluation of mapping algorithms. Assuming a certain a priori knowledge, our system augments the sensor data with hypotheses ('Virtual Scans') about ideal models of objects in the robot's environment. These hypotheses are generated by analysis of the current aligned map estimated by an underlying iterative alignment algorithm. The augmented data is used to improve the alignment process. Feedback between data alignment and data analysis confirms, modifies, or discards the Virtual Scans in each iteration. Experiments with a simulated scenario and real world data from a rescue robot scenario show the applicability and advantages of the approach. By replacing the estimated 'Virtual Scans' with ground truth maps our system can provide a flexible way for evaluating different mapping algorithms in different settings.

Keywords

Robot Mapping; Evaluation of Mapping Algorithms; Sparse scan alignment; Spatial cognition in mapping; Force fields

1 Introduction

Robot mapping based on laser range scans has become a major field of research in robotics in the recent years. The basic task of mapping is to consistently combine spatial data, usually obtained from laser range devices, called 'scans', to a single data set, the 'global map'. The global map forms the basis for the task of robot localization. The map represents the environment scanned from different locations, even possibly scanned by different robots in the case of 'multi robot mapping'. The main problem in mapping is to correct for pose (position and heading direction) errors in scanning the environment. One class of solutions estimates rigid transformations (translations and rotations) of scans based on feature correspondences between the individual scans to correct for the pose errors. Techniques like ICP (Iterative Closest Point, e.g. Besl and McKay 1992; Nüchter et al. 2005 and Lu and Milios 1997) or FFS (Force Field Simulation based alignment, Lakaemper et al. 2007b)

belong to this class. They show impressive results, but are naturally restricted: (1) Since they are based on feature correspondence, they require the presence of overlapping features among scans. (2) These algorithms depend on sufficiently good initialization to avoid local minima in the estimation of transformations.

In this paper, we suggest a solution to the first problem: correct alignment in the absence of sufficient feature correspondences. This problem for example can arise in search and rescue environments (these environments typically show a small number of landmarks) or when a team of multiple robots builds a global map jointly. In this situation scans acquired from different views do not necessarily reveal the entire structure of the scanned object. The motivation to our approach is that even if the optimal relations between structures of scans is not known, it is possible to infer hypotheses of underlying structures from the sub-optimal alignment of single scans based on the assumption of certain real world knowledge. The hypotheses can then be used to improve the alignment. Figure 1 illustrates the basic idea. The top row shows a situation where the relations between features of scans can not reveal the real world structure, and therefore leads to misalignment. In the bottom row, analysis from a global view estimates the underlying structure. This hypothesis then augments the real world data set, to achieve a correct result.

The motivational example shows the ideal case; it doesn't assume any error in the *analysis* of global map in detecting the underlying structure. Our system also handles the non ideal situation where detecting structures and alignment becomes an iterative process. It utilizes a feedback structure between hypothesis generation and the scan alignment. The hypotheses are improved from the current alignment which are in turn used to improve the alignment. This will be discussed in more detail below. We first want to explain our approach in a more general framework.

Feature correspondence algorithms such as ICP or FFS can be seen as “Low Level Spatial Cognition” processes (LLSC), since they operate based on low level geometric information. The feature analysis of the global map, which is suggested in this paper, can be described as “Mid Level Spatial Cognition” process (MLSC), since we aim at analysis of mid-level features like lines, rectangles, etc. Augmenting real world data with expected models can be seen as an example of integration of LLSC and MLSC processes to improve the performance of spatial cognition tasks in robotics. We are using the area of robot perception for mobile rescue robots, specifically alignment of 2D laser scans, as a showcase to demonstrate the advantages of these processes.

In robot cognition, MLSC processes infer the mid level from low level data based on *global* properties of the data. In our case, we detect the presence of simple mid level objects viz. line segments and rectangles. The MLSC processes model world knowledge, or assumptions about the environment. In our setting for search and rescue environments, we assume the presence of (collapsed) walls and other man made structures. If possible wall-like elements or elements resembling rectangular structures are detected, our system generates the most likely model as a hypothesis, called a ‘Virtual Scan’. Virtual Scans are represented using the same data format as the raw sensor data. Hence the low level alignment process can align the original scan data augmented by Virtual Scans.

In robot cognition, LLSC processes usually describe feature extraction based on *local* properties of the data (e.g. orientation of data point). In our system laser scans (virtual and real) are aligned into a global map using ‘Force Field Simulation’ (FFS) as the LLSC core process. FFS was recently introduced to robotics (Lakaemper et al. 2007b). In FFS, data points and correspondences are assigned weights, or values of certainty in an adaptive way. Hence FFS is a *natural* choice over its main competitor, ICP (Besl and McKay 1992; Nüchter et al. 2005), for using the Virtual Scans. The certainty values of the virtual data points can be used to indicate the strength of hypotheses.

FFS is an iterative alignment algorithm. The two levels (LLSC: data alignment by FFS, MLSC: data augmentation) are combined into a feedback structure as shown in Fig. 2. The following steps are repeated in each iteration:

- The FFS-low-level-instance pre-processes the data. Correspondences are based on low level features. The low level processing builds an estimate of the global map, which provides input for the mid level cognition module.
- The mid level cognition module analyzes the current global map, detects possible mid level objects present in the real world. These can be seen as suggestions which are fed back into the low level module as Virtual Scans. The low level system then aligns the augmented data for re-evaluation by the mid level system.

The following example will illustrate the feedback: Fig. 3 shows two scans from two different poses (see also Fig. 1). An MLSC process detects a rectangular structure (the assumed world knowledge) and adds a generating model (rectangle) to the data set. The LLSC module aligns the augmented data. The hypothesis now aligns the scans in an improved way. In each iteration, the aligned scans are analyzed to adjust the MLSC hypothesis. Thus LLSC and MLSC assist each other in a feedback loop. An earlier version of this paper appeared in IROS (Lakaemper and Adluru 2008).

The rest of the paper is organized as follows: Sect. 2 discusses related work, Sect. 3 reviews the basic Force Field Simulation algorithm for mapping, Sect. 4.3 presents details about the mid-level analysis of the maps for creating virtual scans, Sect. 5 presents the improved mapping algorithm and extensive experimental validation. Section 7 demonstrates how our system can be extended for evaluation of mapping algorithms under different scenarios and Sect. 8 summarizes the conclusions of the paper.

2 Related work in spatial cognition and mapping

The potential of MLSC has been largely unexplored in robotics, since recent research mainly addressed LLSC systems. They show an astonishing performance: especially advances in statistical inferences (Doucet et al. 2001; Grisetti et al. 2005; Knill and Richards 1996) combined with geometric modeling of human perception (Field et al. 1993; Eric and Grimson 1990; Pentland 1986) and the usage of laser range scanners contributed breakthroughs in robot applications, with spectacular results achieved in the DARPA Grand Challenges (2005 and 2007) where several autonomous vehicles were able to successfully complete the race (Thrun et al. 2006). Although sophisticated statistical and geometrical

models like Extended Kalman Filters (EKF) (Huang and Dissanayake 2006), Particle Filters (Grisetti et al. 2005) and ICP (Iterative Closest Point) (Besl and McKay 1992; Nüchter et al. 2005) utilized in mapping approaches show impressive results, their limits are apparently clear in the aforementioned scenarios like search and rescue. These systems are still based on low level cognitive features, since they construct metric maps using correspondences between sensor data points. Having these well-engineered low level systems at hand, it is natural to combine them with MLSC processes to mutually assist each other and obtain improved performance in broader applications.

The study of MLSC in humans, in particular in spatial intelligence and learning, is advancing rapidly (Freksa et al. 2004; Kruijff et al. 2007; Uttal 2000). Such results are used to model generic representations of space for mobile robots using both symbolic (Kuipers 2000) and non symbolic (Ghiselli-Crippa et al. 1996) approaches. Naturally, SLAM (Simultaneous Localization and Mapping) which is a spatial cognitive process (Dissanayake et al. 2000) is often used as an example application (Martinelli et al. 2003). In Vasudevan et al. (2006), a spatial cognition based map is generated based on High Level Objects. Representation of space is mostly based on the notion of a hierarchical representation of space. Kuipers (2000) suggests a general framework for a Spatial Semantic Hierarchy (SSH), which organizes spatial knowledge representations into different levels according to ontology from sensory to metrical information. SSH is an attempt to understand and conceptualize the cognitive map (Kuipers 1983), the way we believe humans understand space. More recently Yeap and Jefferies (2000) trace the theories of early cognitive mapping. They classify representations as being space-based and object-based. Comparing to our framework, these classifications could be described being related to LLSC and High Level Spatial Cognition (HLSC), hence the proposed LLSC-MLSC system would relate closer to space-based systems.

In Bertel et al. (2006), the importance of ‘Mental Imagery’ in spatial cognition is emphasized and basic requirements of modeling are stated. Mental Images invent or recreate experiences to resemble actually observed events or objects. This is closely related to the “Virtual Scans” described in this proposal. Recently, Chang et al. (2007) presented a predictive mapping approach (P-SLAM), which analyzes the environment for repetitive structures on the LLSC level (lines and corners) to generate a “virtual map”. This map is either used as a hypothesis in unexplored regions to speed up the mapping process or as an initialization help for the particle filter when a region is first explored. In the second case the approach has principles similar to those in using Virtual Scans. The impressive results of P-SLAM can also be seen as proof of concept of integrating prediction into robot perception.

Geometric robot mapping is defined as the problem of aligning a set of scans (views) of the environment to build the map of the scene. The objective function for the alignment process is defined in terms of consistency of the geometric structures identified using different sensors in the different scans. Thus the goal of geometric robot mapping is to find the maximum likelihood of the fusion of sensor information from different scans. On the LLSC level the problem of simultaneous aligning of scans has been treated as estimating sets of poses (Lu and Milios 1997). There are numerous image registration techniques, the most famous being Iterative Closest Point (ICP) (Besl and McKay 1992), and its numerous

variants to improve speed and convergence basins. All these techniques essentially perform a search in transformation space (translations and rotations) to find the set of optimal pairwise transformations of scans. They optimize some objective function defined on transformation spaces. The techniques vary in defining the optimization functions that range from being error metrics like “sum of least square distances” to quality metrics like “image distance” (Birk and Carpin 2006). Force Field Simulation (FFS) (Lakaemper et al. 2007b) minimizes a potential derived from forces between corresponding data points in all the scans. The Virtual Scan technique presented in this paper will be integrated with FFS as underlying alignment technique.

3 Scan alignment using FFS

Understanding some aspects of FFS is important for understanding the proposed extension using Virtual Scans. We will give an overview of FFS here. FFS aligns scans obtained by robots, typically from different poses. We assume the scans to be roughly pre-aligned (see Fig. 11(b)), using odometry or shape based pre-alignment (Adluru et al. 2006). This is in accord with the performance comparison between FFS and ICP described in Lakaemper et al. (2007c). In FFS, each scan is seen as a non-deformable entity, a ‘rigid body’. In each iteration, a translation and rotation is computed for all scans *simultaneously* (Lakaemper et al. 2007a). This process minimizes an objective function defined on the set of all data points (real and Virtual Scans). The optimization is performed using a gradient descent approach motivated by simulation of dynamics of rigid bodies (the scans) in gravitational fields, but “*replaces laws of physics with constraints derived from human perception*” (Lakaemper et al. 2007b). The gravitational field is based on strength of correspondences (forces) between all pairs of data points. The forces are designed in such a way that a low overlaying potential corresponds to a visually good appearance of the global map. FFS minimizes the potential induced by the forces and converges towards a local minimum of the potential, representing a locally optimal configuration of scans. The scans are translated and rotated according to the laws of motion of rigid bodies in a force field. Figure 4 shows the basic principle: forces (red arrows) are computed among four scans. FFS simultaneously transforms all scans until a stable configuration is reached.

If S_1, S_2 are two different scans, the force between two data points $p_i \in S_1$ and $p_j \in S_2$ is defined as a vector

$$V(p_i, p_j) = M(p_i, p_j) \frac{p_i - p_j}{\|p_i - p_j\|} \quad (1)$$

Its magnitude is defined as:

$$M(p_i, p_j) = \frac{1}{\sigma_t \sqrt{2\pi}} e^{\left(-\frac{\|p_i - p_j\|^2}{2\sigma_t^2}\right)} w_i w_j \cos(\angle(p_i, p_j)) \quad (2)$$

where σ_b , w_i , w_j , $\angle(p_i, p_j)$ defined as follows: $\angle(p_i, p_j)$ denotes the angle between the orientations of points, which is defined as the angle between orientations of the underlying locally linear structures. See Fig. 5(a) for an example, which especially shows the influence of the cosine term in (2): forces are stronger between parallel structures. The computation of force between two linear structures is similar (but not identical) to the way the tangent directions of the scan points are used in computing the amount of displacement in each iteration of Lu/Milios' simultaneous ICP (Lu and Milios 1997). σ_t is a parameter steering the radius of influence of force exerted by a data point on other data points. σ_t is decreased during the iterative process as the alignment improves, thus changing the influence of each data point to be more local. The weights w_i , w_j express the certainty of the points p_i , p_j . They can model the "feature importance" of a data point. We utilize this feature of FFS to model the strength of hypothesis in the Virtual Scans. Hence in (2) the interaction between LLSC and MLSC can be seen directly: the distance ($\|p_i - p_j\|$) and cosine terms correspond to LLSC, while the weights are derived from MLSC (in our case the Virtual Scans).

The computation of the transformation of each scan is determined assuming movement of rigid bodies in the given force field. The force field experienced by a point p_i in a configuration of scans with points $\mathcal{P} = \{p | p \in \cup_{i=1..n} S_i\}$ is given by summing up all the forces as:

$$F(p_i) = \sum_{p_j \in \mathcal{P} \setminus p_i} V(p_i, p_j) \quad (3)$$

All the points $p_i^j \in \mathcal{P}$ in a single scan S_j share the same transformation, consisting of rotation and translation. The basic laws of dynamics of rigid bodies in force fields accumulate the translation of all masses of a single scan into a single translation and a single rotation around a defined center. For each scan S_j , the translational and rotational accelerations have to be determined. The translational acceleration $a_T(S_j)$ of a scan S_j is given by:

$$a_T(S_i) = \frac{\sum_{p \in S_i} F(p)}{\sum_{p \in S_i} w_p} \quad (4)$$

The rotational acceleration a_R is computed using torque and moment of inertia. Torque and inertia play the role of force and mass respectively, but take into account the distance to the rotational center c_R .

$$\begin{aligned} inertia &= \sum_{p \in S_i} m w_i \|p_i - c_R\|^2 \\ torque &= \sum_{p \in S_i} \|p_i - c_R\| \times F(p) \end{aligned}$$

a_R is defined as

$$a_R = \frac{\text{torque}}{\text{inertia}} \quad (5)$$

The rotational center c_R is either defined as the robot's position, or by the center of mass. Experiments show, that in the first iteration steps it is useful to set the rotational center to the center of mass, while in later steps the robot's position is preferable. The first choice enables easier rotation, the second is modeling the actual scan setting more precisely. Hence, the closer the global map is to the solution, the more preferable is the robot's position as rotational center.

With a_T and a_R the transformation $t_k = (x_k, y_k, \theta_k)$ for scan S_k is defined by:

$$(x_k, y_k) = \frac{1}{2} a_T \Delta_t^2 \quad (6)$$

$$\theta_k = \frac{1}{2} a_R \Delta_t^2 \quad (7)$$

Δ_t being the step width of the gradient descent. Computing the total force in a naïve way has a time complexity of $O(m^2)$ where m is the number of data points in a particular configuration. But using 'neighborhood approximations' and 'Fast Gauss Transforms' the complexity can be reduced to $O(m \log m)$ or $O(m)$. Please see Sect. IV of Lakaemper et al. (2007a) for details on complexity analysis.

The motion updates of the scans are computed using Newton's laws of motion by applying the forces for Δ_t units of time. Then the forces are re-computed by analyzing the new configuration of the scans. For a single transformation step see Fig. 5(b). Δ_t can be seen as the step width in the gradient descent process. It is monotonically decreased, allowing the system in early iterations to jump out of local minima, yet to be attracted by local features in later steps for convergence. The interplay between σ_t and Δ_t is an important feature of FFS. See Figs. 11 and 12 for an example of the performance of FFS on a laser range data set.

FFS is closely related to *simultaneous* ICP and a performance evaluation of both algorithms showed similar results (Lakaemper et al. 2007c). In general, FFS can be seen as more robust with respect to global convergence with poor initialization, since the point correspondences are built not in a hard (nearest neighbor) way but in a soft(sum of forces) way. Also the inclusion of weight parameters makes it a natural decision for our purpose of extension using Virtual Scans.

4 Creating virtual scans: mid level analysis

The laser range data goes through two pre-processing steps before the scans are aligned. First, the underlying local linear structures (line segments) are detected in each scan. Classic

approaches like Hough line detection rely on local linearity of the underlying data points. Hence such techniques are very sensitive to noise. We therefore use a recently published technique (Latecki et al. 2006) which is specifically tailored to model laser scan data with line segments, using extended Expectation Maximization. Second, data points are equidistantly sampled along these line segments. The original data is discarded in favor of the newly sampled points. This solves certain problems of noisy scans. It also reduces the number of points drastically. See Fig. 6 for an example. In each iteration the mid-level analysis is performed on the current estimate of the global map (without the Virtual Scans). It involves detecting linear and rectangular structures as explained below.

4.1 Lines

The usage of linear structures for our Virtual Scan approach is motivated by the world knowledge assumption of scanning a man made environment (e.g. a collapsed house): although these environments often locally don't show major linear elements any longer, a global view often reveals an underlying global linear scheme, which we try to capture using a global line detection. Here (in contrast to our *local* line segment detection in the pre-processing step) we use the classic line detection approach of Hough transform (Hough 1962), since it detects *globally* present linear structures. Hough transform does not only show location and direction of a line, but also the number of participating data points. We use this value to compute a certainty-of-presence measure, i.e. the strength of the line hypothesis. The data points on these structures inherit this certainty as their weights (e.g. w_i , w_j in (2)). We only use lines above a certain threshold of certainty. We will specify how the detected lines are utilized to create the Virtual Scan in Sect. 4.3. The Hough detection is performed on the entire set of (re-sampled) data points. We do not use the local linear information of the line segment representation of the data here, since we aim at global linearity. This is more robustly detected by Hough transform.

4.2 Rectangles

The rectangles are detected using the entire set of local line segments of all scans, obtained after the first pre-processing step. We use the rectangle detection approach described in Lagunovsky and Ablameyko (1997): each line segment (of each scan) is translated into 'S, L, D space' (Slope, Length, Distance), which simplifies the detection of appropriate (rectangular like) configurations of four near parallel and near perpendicular segments. For details see Lagunovsky and Ablameyko (1997). Before detecting rectangles we merge spatially close lines in a cluster to a single prototype line using a line merge approach described in Lakaemper et al. (2005) as shown in Fig. 7(b). The rectangle detection module then predicts location, dimension and certainty-of-presence of hypothetical, rectangles present in the data set of merged line segments. The certainty, or strength of the hypothesis is derived from properties (segment length, perpendicularity) of line segments participating in generating rectangles. This value is used to create the weight of the rectangle (and data points on it) in the Virtual Scan.

4.3 Creating a virtual scan

A Virtual Scan is a set of virtual laser scan points. The detected line segments and rectangles are represented as point sets in a Virtual Scan as if detected by a virtual laser scanner. All

detected MLSC elements (lines, rectangles) are represented in a single Virtual Scan. A sample Virtual Scan is shown in Fig. 8(b).

An important feature of the Virtual Scan is that each of its point is assigned a weight, representing the strength of the hypothesis of the underlying virtual structure. Utilizing this feature, we benefit from the weights that steer the FFS alignment. As defined in (2), the weights w_i , w_j directly influence the alignment process: stronger points, i.e. points with higher value of w , have a stronger attraction. Hence, a strong hypothesis translates into a locally strongly attractive structure. The strength of a hypothesis reflects the belief of the hypothesis relative to the real data. All data points of the real scan are assigned a ‘normal/uniform’ weight of 1.

5 Improved mapping using virtual scans: algorithm

The Algorithm 1 describes the interaction between LLSC (FFS) and the MLSC processes. S_i , $i = 1..n$, denotes the real scan data, consisting of n scans. $V^{[t]}$ denotes the Virtual Scan in iteration t .

Algorithm 1

Force field based mapping using Virtual Scans.

-
- 1: Create set of line segments L_i for each scan S_i using the technique in Latecki et al. (2006).
 - 2: Init: $t = 1$, $V^{[0]} = \emptyset$
 - 3: **repeat**
 - 4: Perform FFS on $\cup_{i=1..n} S_i \cup V^{[t-1]}$, resulting in transformations (translation, rotation) $T_i^{[t]}$ for each scan $S_i \in 1..n$ (Sect. 3).
 - 5: Construct a global map G and a set of global GL of line segments by transforming the scans and their line segments into a common global frame: $G = \cup_{i=1..n} T_i^{[t]}(S_i)$; $GL = \cup_{i=1..n} T_i^{[t]}(L_i)$.
 - 6: Detect set of lines $\mathcal{L}$ in G , set of rectangles $\mathcal{R}$ in GL (Sect. 4.1, Sect. 4.2).
 - 7: Create Virtual Scan $V^{[t]}$ using the scan points representing the elements of $\mathcal{L}$ and $\mathcal{R}$ (Sect. 4.3).
 - 8: Update σ_t and ρ_t for the FFS process (Sect. 3).
 - 9: **until** FFS converges to a stable global map
-

The algorithm is greedy in the sense that it uses the strongest hypotheses. Like in every greedy approach, the problem of being trapped in a local minimum is given. The current algorithm tends to work against the local minimum problem in the following way: each hypothesis is valid for a single iteration only. Single wrong hypotheses, which are not consistent with the global appearance of the map will not be recreated. FFS itself is designed to escape local minima using the step-width parameter, which follows a cooling strategy. FFS therefore does not greedily follow the suggestion of a (probably wrong) hypothesis, but uses it as a tendency towards which to rearrange the scans. If this new arrangement favors the hypothesis, it will be recreated, which is more likely for ‘correct’, i.e. globally consistent hypotheses. Surely, the approach is not guaranteed to escape local minima, therefore it is important to use the virtual scans carefully. This means, only significantly strong hypotheses

should be used, additionally, the weights of the features in the virtual scans have to be adjusted carefully.

6 Experiments and results

6.1 Sparse scanning (simulated data)

This experiment shows the effect of Virtual Scans in a sparsely scanned environment. It features a simple environment to highlight the principle of Virtual Scans and to show the improvement in the alignment process. Please compare this to the motivational example introduced in Sect. 1, as well as to Figs. 1 and 3. Figure 9 shows a simulated arena consisting of four rectangular rooms scanned from five different positions (shown as ‘x’ marks). Each scan is preprocessed and randomly translated and rotated to simulate pose errors. We first try to align this data set using FFS without Virtual Scans. The performance of FFS depends on the initial value of $\sigma_t(2)$ i.e. σ_0 . We tried multiple initial values: results of $\sigma_0 = 30$ and $\sigma_0 = 80$ are shown in the bottom row of Fig. 9 left and right respectively. σ is measured in units of the data set. The width of the simulated arena is 400 units. In bottom left, with a lower σ_0 , local structures are captured and aligned correctly, but global correspondences can not be detected (the ‘hallway’ between the rooms is offset incorrectly). As can be seen in bottom right, increasing σ_0 and thereby strengthening the influence of global structures leads to wrong results since local correspondences become relatively unimportant: FFS optimizes correspondences of major structures (although they are far from each other in the initial map). The disability of balancing the influence of local and global structures is an inherent drawback of alignment processes which are based on point correspondences (e.g. ICP, FFS), and not a special flaw of FFS only (other values of σ did not improve the alignment either).

Figure 10 shows the improvement when Virtual Scans are used. Here we use $\sigma_0 = 30$. FFS is able to improve the local structures, and the global structures are captured through augmentation by Virtual Scans. Also the effect of the hypothesis adjustment by feedback is clearly visible in Fig. 10: Top left shows an early hypothesis, which contains a wrong rectangle and misplaced lines. This early hypothesis is improved by the feedback process between FFS and the rectangle/line detector. Top right shows a later iteration, the line position is adjusted (though not perfect yet), two rectangle hypotheses compete (lower right corner). The final result is shown in the bottom row. The detected lines significantly improved the expected global structures (the walls of the ‘hallway’), the winning rectangle hypothesis ‘glued together’ the corners of the bottom right room (the slight incorrect overlap of the bottom right room is due to sparse features in the simulated environment). Please notice that this room is a structure that is not entirely present in any individual scan, but only detectable in the global map. Hence only the Virtual Scan enhanced FFS could perform correctly.

6.2 NIST disaster area

This experiment shows the improved performance of the alignment process on a real world data set. The data, provided by National Institute of Standards and Technology (NIST) represents a simulated indoor disaster scenario. The data set consists of 60 single laser scans,

taken from 15 different positions in 4 directions (N,W,S,E) with 20° overlap. It can be interpreted as a scene scanned by 15 robots, 4 scans each. No particular order of scans is assumed. The scans have little overlap and have no distinct landmarks. The initial global map was computed using a shape based approach described in Adluru et al. (2006). See Fig. 11 for sample individual scans and the initial global map. We used the initial global map for two different runs of FFS, one using Virtual Scans, one without Virtual Scans. The increase in performance is qualitatively evaluated by visual inspection, since the ground truth data is not available. As can be seen in Fig. 12 the utilization of Virtual Scans leads to distinct improvement in overall appearance and mapping details. Overall, the map is more ‘straight’ (see for e.g. the top wall), since the detection of globally present linear structures (top and left wall in Fig. 12) adjusts all participating individual segments to be collinear. These corrections advance into the entire map. The improvements can be better seen in certain details: the most distinct improvements are encircled in Figs. 12(d) and (f). The rectangle in the center of the global map is an excellent example for a situation where correct alignment is not achievable with only low level knowledge. Only the rectangular structures from the Virtual Scan (see Fig. 12(c)), can force the low level process to transform the scan correctly. Without the assumed rectangle the low level optimization process necessarily tries to superimpose the two parallel sides of the rectangle to falsely appear as one. Magnification of both situations can be seen in Fig. 12(e).

6.3 Low cost low effort scanning: NIST response robot evaluation exercise

The data for this experiment was collected during the organized Response Robot Evaluation Exercise organized by NIST (National Institute of Standards and Technology) in Texas, November 2008 (NIST 2008). These exercises for US Rescue teams ‘*introduce emerging robotic capabilities to emergency responders within their own training facilities, while educating robot developers regarding the necessary performance requirements and operational constraints to be effective*’. The exercise was located in ‘Disaster City’, ‘*a 52-acre training facility designed to deliver the full array of skills and techniques needed by urban search and rescue professionals*’. Disaster City is a TEEX (Texas Engineering Extension Service, a part of Texas A&M University, College Station) facility. The facility offers multiple sites, for example collapsed buildings, trains etc. for simulated disaster environment training. During the exercise, different sites were scanned with different kinds of equipment (2D scanners, 3D scanners, cameras etc.) to evaluate the performance of sensors and algorithms and to demonstrate their usability to first responders. The purpose of the specific experiment described here was to gain experiences with extremely cost efficient equipment. The equipment consisted of a single Hokuyo URG-04LX 2D laser scanner (<2500 US\$) mounted on a stick, see Fig. 14. This device was carried through the environment by a human first-responder to retrieve scans in a ‘step and shoot’ manner. No additional sensors were used. The executing responder was instructed to keep a constant step-length, and to trigger three single horizontal scans (90° left, walking direction, 90° right) at each step from knee height. A photo of the scanned area (Disaster city location ‘House of Pancakes’) can be seen in Fig. 14, right. The result of the superimposed 27 raw scans is shown in Fig. 15. The superimposition assumes a predetermined constant step length (93 cm) of the responder. Expectedly, simple superimposition is insufficient. Due to imprecise odometry (changing step length) and angular errors (human estimated, error prone

pitch, roll and yaw of the laser) the measurements lead to a blurry representation of the environment.

The data shown in Fig. 15(a), was used as an input to our Virtual Scan mapping system. Since the environment is an indoor scenery, we enabled the system to assume straight walls. An intermediate step (iteration 10) is shown in Fig. 15(b). The system detects a single straight wall, which is correct. This hypothesis is kept during the following iterations. Hence the virtual scan attracts all scans containing this wall to align in a collinear fashion, see final result, Fig. 15(c). The high density of single scan points along the (virtual) wall in the final result appears as a dense line. The correct alignment extends into the global map, the overall appearance (compare Figs. 15(a) and (c)) improved significantly. Note that the data itself is imprecise, since the stick was not held perfectly horizontal in each scan. Therefore it is not possible to align the data perfectly; only an estimation of the real environment can be given. However, significant structures of the environment are captured, e.g. the wall and the desk (with missing right drawer), compare Figs. 14, 15 and 16.

Figure 16 shows the result of a post processing algorithm, based on the underlying linear structures (line segments) used in FFS. Figure 16, left, shows the same alignment of scans as in Fig. 15, bottom, but shows estimated line segments instead of laser reflection points. We post-process these line segments using a line merging algorithm. It merges sets of similar segments to single representative segments. (The algorithm and its possible applications in robot mapping are part of a forthcoming publication.) After the merging process, certain line segments of the merging result were erased based on length and density of the corresponding scan points. The final result is shown in Fig. 16, right. The result consists of <50 line segments, which are stored using their endpoints. Line segments achieve a very compact representation of the environment, which is desirable in disaster scenarios: data like this can be reliably transmitted to hand held devices for the responders since high redundancy can be added due to the low data volume.

7 Using virtual scans for map evaluation

7.1 Motivation

Mapping, in general, is spatial analysis of environmental features of interest. Inherent to this process is its task dependency, hence there is no ‘optimal general mapping’. Mapping of spatial features can be divided into two classes, topographic and topological mapping. While topographic mapping is concerned with detailed, correct geometry of the spatial features, topological mapping aims for correct spatial relation between features only; it favors topological correctness over geometric accuracy, also often referred to as ‘global correctness’ vs ‘local accuracy’. In robot mapping, two approaches relate to topographic and topological mapping. They are *grid based* and *pose based* approaches (pose based approaches are still geometric, and not real topological mapping. However, looking at low geometric feature level approaches only, pose based evaluation can be utilized to relate closer to topological mapping).

We first illustrate grid and pose based approaches, followed by an example of FFS/VS pose based evaluation. We close the section with an outlook how to extend the FFS/VS pose based evaluation to a hybrid grid/pose approach.

7.2 Grid based approaches

In *grid based* approaches, the map to be evaluated (target map) and the ground truth map are both embedded into an occupancy grid. The grid cells are labeled using properties like ‘object’, ‘empty space’ or ‘hidden’ (or assigned an occupancy probability value). An example for such an approach is the *Jacobs Map Analysis Toolkit* (Varsadan et al. 2008), as used in the 2008 RoboCup Rescue competition. In RoboCup Rescue 2008, the target maps were submitted to the system either as GeoTiff formatted image-files, or as a printout; one advantage of a grid based approach is that it can deal with such basic map representations. In case of the Jacobs Toolkit, similarity is computed as sum of nearest distances between target and ground truth cells of same label. The underlying similarity function is an efficiently implemented version of the distance transform. The Jacobs Toolkit therefor measures the *local* geometric accuracy of the map. Since only low level features (object/empty space) are incorporated, the target map must be close to the ground truth map: it is assumed that low level correspondences infer higher level correspondences (e.g. object-object ‘door’ $\Rightarrow$ ‘door’). Larger errors in the global appearance of maps can not be quantified, globally erroneous maps are classified as ‘wrong’—even if they are locally correct, see the example in Fig. 17. The example illustrates a case which is typical for non-autonomous mapping, where first responders need a map to get an overview of the environment to rescue a victim. Global geometric correctness might be of minor interest compared to only locally geometric, yet global topological accuracy. Figure 17, left, shows the ground truth map. Figure 17, center, illustrates a mapping result with high global geometric correctness, although the bottom part is wrong in details. The right example is an example for a map with a high global geometric error. However, all details (obstacles, victim’s position in bottom room) are mapped correctly, the map is also topologically correct (two rooms are connected by hallways). A grid based approach will prefer the center map to the right one. However, if the map is a navigation assistance for first responders, the right map is of higher quality: it shows correctly that the victim (red dot) is reachable from the current position (black dot) using the right hallway. The center map misleads the first responders to take the left hallway, a probably fatal mistake.

Grid based approaches like the Jacobs Toolkit aim to measure the global topographic quality of a robot map, they can not quantify the topological qualities of a map.

7.3 Pose based approaches

A different approach to mapping evaluation is *pose based map quality estimation* (Lakaemper et al. 2007c). Pose Based fitness exploits the fact that precise robot localization is dual to robot mapping: if the robot pose is precisely known in the ground truth map, the scans can be registered based on the pose estimates. Since robot pose measurements are imprecise, the scan data itself has to be taken into account to register the scans in a common coordinate system. Successful registration of scans adjusts the robot poses defined by the target map into the ground truth coordinate system.

Evaluation based on pose information compares the adjustment of robot poses, i.e. an error e is computed for every pose as

$$E(x, y, \theta) = \gamma |(x, y) - (x_G, y_G)| + (1 - \gamma) |\theta - \theta_G| \quad (8)$$

with a (scale dependent) weight parameter $\gamma \in [0, 1]$. $(x, y) - (x_G, y_G)$ is the translational offset between scan and ground truth, $\theta - \theta_G$ denotes the rotational correction. The sum of all pose errors yields the overall error (Lakaemper et al. 2007c). The main advantage of pose based evaluation is its applicability in higher dimensions (e.g. 6D-SLAM). The number of poses to be evaluated is dimensionality independent, whereas the memory consumption of a grid based evaluation approach increases for 3D applications to a prohibitive cubic behavior. Hence there is a high interest in gaining knowledge about pose based evaluation.

Although still a geometric measure, pose based map estimation relates closer to evaluation of topographical maps, see the example in Fig. 18. The global topological correctness is captured by the fact that only a few rotations are needed to achieve the optimal result.

Due to different local and global influence of rotational errors, it is hard in pose based estimation to precisely quantify local geometric correctness, a major drawback compared to grid based approaches. Also, pose based evaluation requires the less intuitive single scans along with their aligned poses as input. In practice this is not a drawback, since mapping approaches are naturally based on single scans. Once a pose based evaluation standard is established, mapping algorithms can be required to save single poses before merging the scans to a global map. Another critical factor is the parameter γ in (8). Angular errors translate radius dependent to absolute errors, which is hard to model in a single parameter. Additionally, simple summing of errors does in general not reflect the influence of pose errors in real world settings.

7.4 Utilizing FFS/VS for pose based map evaluation

To use Virtual Scans for evaluation of mapping results, we assume the presence of a ground truth map G . To evaluate a given mapping result R with respect to G , both are fed into the FFS/VS system. In order to be able to use R as input, it has to be translated into scan-format. This is a straightforward task, if for example R is given as an image (edge pixels resemble reflection points), or as a set of aligned scan data. G is assumed to be in scan format. G is used as Virtual Scan. Therefore, instead of *creating* a Virtual Scan using mid-level analysis (Sect. 4.3), the ground truth data G is *inserted* into the system, see Fig. 19 left. Assigning a high confidence weight to G will force the evaluated data R to align to the ground truth Virtual Scan. The alignment energy for this process is directly readable from the FFS module. The energy is a measure for visual closeness, see Fig. 19, right. This evaluation procedure is adjustable to evaluate different regions of the ground truth map, since the adjustment energy in regions of interest can be emphasized as explained below.

The energy computed in FFS is a symmetric measure, i.e. aligning R to G leads to the same measure as aligning G to R . This can be used for ‘inverse evaluation’ (evaluating the ground truth G with R) in the following manner: G can be manually split into local regions of

interest (rooms, hallways, etc.). These regions are represented as single scans and used as input for the FFS system, while the map R is used as Virtual Scan. Such a setting has huge advantages, since, on one hand, it is more independent of the actual data representation of R . On the other hand, and more important, the manual split of G defines regions which can be assigned independent evaluation scores.

Observe again that we transform the ground truth map, not the target map, for the two reasons: first, it makes the evaluation independent of the target map's data format. Since the target map is not transformed, it can be given in any format, e.g. GeoTiff. Second, and more important, the part-decomposition of the ground truth map can reflect the task specific requirements of the mapping approach. An example is given in Fig. 22(a), where the ground truth map is decomposed into top room, hallways and bottom room.

The following experiment will illustrate the approach. In this example, we want to evaluate different mapping results R_1 , R_2 , see Fig. 20. The data set 'apartment' was taken from the IROS 2006 test data <http://staff.science.uva.nl/~zivkovic/FS2HSC/dataset.html>. R_1 and R_2 are mapping both results of FFS without Virtual Scans (in general these could be results from different mapping algorithms). For R_1 , Fig. 20(a), we deliberately mistuned the parameters to achieve a suboptimal result.

R_1 , R_2 differ mainly in the upper segment of the apartment: in R_1 it was erroneously tilted. Figures 20(c,d), show the maps with superimposed ground truth map G (in red), which was obtained manually for this example. The difference between R_1 and R_2 in the top segment is clearly visible. This local difference can be captured with our approach if the ground truth map G is manually decomposed into G_d which consists of two scans, representing the upper and lower apartment segment, see Figs. 21(c,d). Such a manual decomposition determines the local regions of interest, which can be left to the evaluators' discretion. The possibility to determine these regions of interest shows the flexibility of the evaluation approach.

In experiment one, R_1 serves as Virtual Scan, while the two scans of G_d are aligned to it using FFS. In experiment two, R_2 takes the role of the Virtual Scan. The quality of R_1 and R_2 is evaluated using the relative transformation of the scans of G_d to a reference scan. We denote the index of the reference scan with $\hat{i}$. In our example, the lower part (green part in Fig. 21) is the reference scan. FFS results in translation $t(i)$ and rotation $r(i)$, $i = 1 \dots n$ for all n scans in G_d . For this example, (8) yields:

$$\begin{aligned} E &= \sum_{i=1}^n \gamma \|t(i) - t(\hat{i})\| + (1 - \gamma)(r(i) - r(\hat{i})) \\ &= \sum_{i=1}^n \gamma \|\Delta t(i)\| + (1 - \gamma)\Delta r(i) \end{aligned} \quad (9)$$

In this experiment, we set $\gamma = 0.9$, which means that a translation of 1 meter is about as influential as a rotation about 10 degrees. Equation (9) leads to the following results:

Map	$\parallel$	$t \parallel$	(m)	r (degree)	E
-----	-------------	---------------	-----	--------------	-----

$R1$	0.76	11.1	1.80
$R2$	0.75	5.1	1.18

This experiment was designed with focus on the global correctness of the upper segment of the apartment. For this setting, we can infer that the mapping result $R2$ is better than $R1$. Quantitatively the result $R2$ is 1.5 ($\approx 1.80/1.18$) times better than $R1$.

7.5 Outlook: hybrid grid/pose based evaluation

In a different setting, we might be interested in a more detailed view without focussing on global configuration. For example, let us assume a rescue setting where a first responder is more interested in local structures (in the apartment: sofas, tables), yet less in the correct global geometry. In such a setting, $R1$ and $R2$ would be equally helpful. This can be quantified if the Virtual Scan approach is utilized in a different way: the maps $R1$, $R2$ can be partitioned using the corresponding aligned scans of G_d as a partitioning scheme, i.e. the scans in G_d define (not necessarily disjunct) areas in $R1$, $R2$ which can be evaluated individually with grid based evaluation (Lakaemper et al. 2007c). This means, the amount of FFS adjustment is, in this extreme case, disregarded. In a real hybrid approach, Pose and Grid based errors can be combined. However, the underlying principle stays the same:

1. Substitute the alignment of real scans to virtual scans with an alignment between target map and ground truth map.
2. Split the ground truth map into parts with required high geometric accuracy. These parts are the ‘single scans’, they will be aligned by FFS/VS to the target map (= the virtual scan).
3. Instead of aligning the target map to the ground truth map, use reverse alignment: align the ground truth map to the target map
4. After alignment, use the part-alignment parameters for a pose based evaluation. Additionally, evaluate the geometric preciseness of each part (after FFS/VS alignment) using a grid based approach. The weight of each part-transformation as well as the relative weight of pose-based and grid based evaluation scores are predetermined in relation to the task the map was created for.

Items (1), (2) and (3) define the mapping evaluation in the framework of the FFS/VS mapping approach. Again, using the target map as a fixed virtual scan with high confidence weight, these steps align the single scans of the decomposed ground truth map to the target map, see Fig. 22. Item (4) quantifies the map quality, using pose based parameters from FFS/VS, and grid based parameters from additional evaluations on the single parts. Their relative pose, defines the global appearance of the map. It will be captured by the transformation parameters. The importance weight of the transformation parameters can be individually determined. In the example of mapping-assistance for first responders, a rotational error among a certain threshold could be omitted, leading to a preference of the right map in Fig. 17.

Note that in the Virtual Scan based evaluation approach, ground truth data representation is not limited to *physical objects*, but can consist of *geometric properties* (e.g. evaluate how well a map represents lines/rectangles). In this case, the original Virtual Scan approach is utilized, instead of insertion. The properties are then defined by means of MLSC-analysis modules. This becomes especially interesting if no ground truth data is available. Thus our FFS/VS can be used as a very flexible evaluation tool for different settings and mapping results.

7.6 Extendability to different data sources

The question might arise, if the principles of FFS and Virtual Scans can be extended to non-laser range data, e.g. camera images. The main difference between range and camera images (apart from the obvious range vs reflectivity) is, that a horizontally taken 2D range image yields a (limited) image representing the environment from a bird's view angle; the camera image can be interpreted (intuitively) as orthogonal to that. However, camera images taken from the bird's view, e.g. satellite images, are quite similar to ground based laser range images, if they are converted to *edge images*. Edges are pixels with significant change in brightness relative to a certain spatial neighborhood. Detection of edges in general is a complicated problem in computer vision. The reason is, that in real world scenarios, edges belonging to single objects, are rarely seen as single, continuous and closed boundaries. Due to noise (e.g. shadows) but especially occlusion, only partial edges can be detected. Noise adds extra edges, which can hardly be filtered without higher knowledge about the objects to be detected. In the case of satellite images and special cases, like buildings, it is however relatively simple to detect edges. In these cases, the edge images are very similar to laser range scans. They even have the advantage, that they do not suffer from the problem of hidden objects (objects in the shadow of other objects). Experiments with such data sets have been performed; we did not include the results in this paper.

8 Conclusions

The presented implementation of an extension to the FFS alignment process using Virtual Scans could significantly improve the results for the alignment task. The Virtual Scans are composed of hypothetical mid level real world structures. The implementation provides a proof of concept of the applicability of the presented concept for the combination of LLSC and MLSC processes. The detection of simple elements (lines, rectangles) based on weak real world assumptions could improve the performance. With a straightforward extension, we demonstrated how Virtual Scans can be utilized for mapping evaluation in different settings.

Acknowledgments

We would like to thank Raj Madhavan, Oak Ridge National Labs, Chris Scrapper, Adam Jacoff and Elena Messina (all NIST, Gaithersburg, MD) for their great and hard work of organizing and building the 2008 Disaster-City-Exercise experience. We also want to thank Steve Richards, Acroname Inc., for collecting the data set 'dwelling' (Fig. 15) using the Acroname Laser Stick (Fig. 14). N. Adluru is supported by Computation and Informatics in Biology and Medicine (CIBM) program and Morgridge Institute for Research at the UW-Madison.

This work was supported by the Department Of Energy under Grant No. DE-FG52-06NA27508. N. Adluru is supported by Computation and Informatics in Biology and Medicine (CIBM) and Morgridge Institute for Research (MIR) at the UW – Madison.

Biographies

Rolf Lakaemper is an Assistant Professor in the Department of Computer and Information Sciences, Temple University. His primary research interests are in computer vision and robotics. He has done significant work in shape based object recognition and is currently focussed on developing robust spatial cognition based mapping algorithms and evaluation systems.

Nagesh Adluru is a postdoctoral research associate at University of Wisconsin, Madison. He graduated with a PhD in Computer and Information Sciences at Temple University. His PhD work mainly focussed on Contour Grouping and Robot Mapping where he developed novel techniques using shape models and particle filters. His primary research interests are in computer vision, robotics, medical imaging. He is currently working on brain imaging at the Waisman Center in UW – Madison.

References

- Adluru, N.; Latecki, L.J.; Lakämper, R.; Madhavan, R. Proc of the IEEE Int workshop on safety, security and rescue robotics (SSRR). Gaithersburg, Maryland, USA: 2006. Robot mapping for rescue robots.
- Bertel S, Barkowsky T, Engel D, Freksa C. Computational modeling of reasoning with mental images: Basic requirements. Proceedings of the 7th international conference on cognitive modeling. 2006:50–55.
- Besl P, McKay N. A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1992; 14(2):239–256.
- Birk A, Carpin S. Merging occupancy grid maps from multiple robots. Proceedings of the IEEE. 2006; 94:1384–1397.
- Chang HJ, Lee CSG, Lu YH, Hu YC. P-SLAM: Simultaneous localization and mapping with environmental-structure prediction. IEEE Transactions on Robotics. 2007; 23(2):281–293.

- Dissanayake G, Durrant-Whyte H, Bailey T. A computationally efficient solution to the simultaneous localization and map building (slam) problem. *Proceedings of the IEEE int. conference on robotics & automation (ICRA)*. 2000:1009–1014.
- Doucet, A.; de Freitas, N.; Gordon, N. *Sequential Monte Carlo methods in practice*. Berlin: Springer; 2001.
- Eric, W.; Grimson, L. *Object recognition by computer: the role of geometric constraints*. Cambridge: MIT Press; 1990.
- Field DJ, Hayes A, Hess R. Contour integration by the human visual system: evidence for a local association field. *Vision Research*. 1993; 33:173–193. [PubMed: 8447091]
- Freksa, C.; Knauff, M.; Krieg-Brckner, B.; Nebel, B.; Barkowsky, T. *Spatial cognition IV, reasoning, action, interaction*. Berlin: Springer; 2004.
- Ghiselli-Crippa, T.; Hirtle, SC.; Munro, P. *Connectionist models in spatial cognition*. Vol. 32. Dordrecht: Springer; 1996. p. 87–104.
- Grisetti G, Stachniss C, Burgard W. Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling. *Proceedings of the IEEE int. conference on robotics & automation (ICRA)*. 2005:2443–2448.
- Hough, PVC. *Methods and means for recognizing complex patterns*. United States Patent 3069654. 1962.
- Huang S, Dissanayake G. Convergence analysis for extended Kalman filter based slam. *Proceedings of the IEEE int. conference on robotics & automation (ICRA)*. 2006:412–417.
- Knill, DC.; Richards, W., editors. *Perception as Bayesian inference*. New York: Cambridge University Press; 1996.
- Kruijff GJM, Zender H, Jensfelt P, Christensen HI. Situated dialogue and spatial organization: what, where and why? *International Journal of Advanced Robotic Systems*. 2007; 4(1):125–138.
- Kuipers B. The spatial semantic hierarchy. *Artificial Intelligence*. 2000; 119:191–233.
- Kuipers, BJ. The cognitive map: could it have been any other way?. In: Pick, HL.; Acredolo, LP., editors. *Spatial orientation: theory, research, and application*. New York: Plenum; 1983. p. 345–360.
- Lagunovsky, D.; Ablameyko, S. *CAIP'97: Proceedings of the 7th international conference on computer analysis of images and patterns*. London: Springer; 1997. Fast line and rectangle detection by clustering and grouping; p. 503–510.
- Lakaemper R, Adluru N. Map merging for distributed robot navigation. *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2008:2915–2921.
- Lakaemper, R.; Latecki, LJ.; Sun, X.; Wolter, D. *Geometric robot mapping*. Vol. 3429. Berlin: Springer; 2005. p. 11–22.
- Lakaemper, R.; Adluru, N.; Latecki, LJ. *European conference on mobile robots*. Freiburg, Germany: 2007a. Force field based n -scan alignment.
- Lakaemper R, Adluru N, Latecki LJ, Madhavan R. Multi robot mapping using force field simulation: research articles. *Journal of Field Robot*. 2007b; 24(8–9):747–762. DOI: 10.1002/rob.v24:8/9
- Lakaemper, R.; Nüchter, A.; Adluru, N.; Latecki, LJ. *Workshop on performance metrics and intelligent systems (PerMIS)*. Gaithersburg, MD: 2007c. Performance of 6d LuM and FFS SLAM: an example for comparison using grid and pose based evaluation methods.
- Latecki LJ, Sobel M, Lakaemper R. New EM derived from Kullback-Leibler divergence. *ACM SIGKDD int conf on knowledge discovery and data mining*. 2006
- Lu F, Milios E. Globally consistent range scan alignment for environment mapping. *Autonomous Robots*. 1997; 4(4):333–349. DOI: 10.1023/A:1008854305733
- Martinelli A, Tapus A, Arras K, Siegwart R. Multi-resolution slam for real world navigation. *Proceedings of the 11th international symposium of robotics research*. 2003
- NIST. NIST response robot evaluation exercise. Search and rescue: Texas Engineering Extension Service (TEEX). 2008. <http://www.teex.com/teex.cfm?pageid=USARprog&area=usar&templateid=1538>

- Nüchter, A.; Lingemann, K.; Hertzberg, J.; Surmann, H.; Pervözl, K.; Hennig, M.; Tiruchinapalli, KR.; Worst, R.; Christaller, T. Proc of the IEEE int workshop on safety, security and rescue robotics (SSRR). Kobe, Japan: 2005. Mapping of rescue environments with kurt3d.
- Pentland AP. Perceptual organization and the representation of natural form. *Artificial Intelligence*. 1986; 28(3):293–331. DOI: 10.1016/0004-3702(86)90052-4
- Thrun S, Montemerlo M, Dahlkamp H, Stavens D, Aron A, Diebel J, Fong P, Gale J, Halpenny M, Hoffmann G, Lau K, Oakley C, Palatucci M, Pratt V, Stang P, Strohband S, Dupont C, Jendrossek LE, Koelen C, Markey C, Rummel C, van Niekerk J, Jensen E, Alessandrini P, Bradski G, Davies B, Ettinger S, Kaehler A, Nefian A, Mahoney P. Stanley: The robot that won the darpa grand challenge: Research articles. *Journal of Robotic Systems*. 2006; 23(9):661–692. DOI: 10.1002/rob.v23:9
- Uttal DH. Seeing the big picture: Map use and the development of spatial cognition. *Developmental Science*. 2000; 3(3):247–264.
- Varsadan I, Birk A, Pfingsthorn M. Determining map quality through an image similarity metric. *Proceedings of the RoboCup symposium*. 2008
- Vasudevan S, Nguyen V, Siegwart R. Towards a cognitive probabilistic representation of space for mobile robots. *IEEE international conference on information acquisition*. 2006:353–359.
- Yeap W, Jefferies M. On early cognitive mapping. *Spatial Cognition and Computation*. 2000; 2(2):85–116.

Fig. 1.

(Color online) Motivation of the Virtual Scan approach. Top row shows alignment without Virtual Scans. Bottom row shows it with Virtual Scans. **(a)** Rectangular object scanned from two positions (*red/blue* robots). **(b)** Correspondences (*block arrows*) between the two scans (*red/blue*) does not reveal the scanned structure. **(c)** Misalignment due to wrong correspondences. **(d)** Analysis of estimated global map detects the rectangular structure (*dotted rectangle*). **(e)** The structure is added as a Virtual Scan. **(f)** Correct alignment achieved due to improved correspondences (*curved arrows* in **(e)**) between real world scans and Virtual Scans

Fig. 2.

LLSC-MLSC feedback structure. The LLSC module works on the union of real scans and the Virtual Scan. The MLSC module in turn re-creates a new Virtual Scan based on the output of the LLSC module

Fig. 3.

(Color online) Feedback between Virtual Scans (VS) and FFS. **(a)** Initial state of real sensor data. **(b)** Real data augmented by VS (*red rectangle*). **(c)** After one iteration of FFS using real and virtual scans. **(d)** New hypothesis (*red square*) based on the map in **(c)**. **(e)** Next iteration of FFS. Since this result resembles an ideal rectangle, adding an additional VS would not relocate the scans. Thus the system converged

Fig. 4. (Color online) Basic principle of FFS. Forces are computed between four scans. *Red arrows* illustrate the principle of forces. The scans are iteratively (here: two iterations) transformed by translation and rotation until a stable configuration is achieved

Fig. 5.

(Color online) FFS example: Showing forces and update. **(a)** Forces (*green*) between two rigid structures in two different scans (*brown, black*). The *black and brown lines* connect the actual data set for display reasons only. The figure shows a magnification of the upper left corner of Fig. 11(b). **(b)** Example of force and movement. *Dotted lines* show two scans (*black, brown*) and their forces (*green*) in iteration t . *Solid lines* show the resulting transformed scans at iteration $t + 1$

Fig. 6. (Color online) Pre-processing steps in FFS. *Top*: Original scan. *Bottom*: Line segments (*black*) and re-sampled data points (*red*). The scale is in centimeters

Fig. 7.
 (Color online) Rectangle detection. **(a)** Global map built by line segments of all individual scans. **(b)** Result of global line merging. **(c)** Magnification of area encircled in **(b)**. The detected rectangles are shown in *red*

Fig. 8. Virtual Scans in an early stage of FFS. **(a)** Global map. **(b)** The Virtual Scan consisting of points representing detected linear and rectangular structures. **(c)** Superimposition of real scans and the Virtual Scan. This is the data used in the following iteration of FFS

Fig. 9.

Top row. Left: Simulated arena consisting of four rooms scanned from five different positions (*crosses*). Points of same color belong to a single scan. *Right:* Pre-processed scans after adding random pose errors. *Bottom row.* Results of FFS without Virtual Scans. Initialized with configuration in top right. *Left:* $\sigma_0 = 30$. *Right:* $\sigma_0 = 80$

Fig. 10.

Same experiment as in Fig. 9, but using FFS with Virtual Scans and $\sigma_0 = 30$. Elements of Virtual Scans (*lines and rectangles*) are shown in *black*. *Top row. Left:* After iteration 10. *Right:* After iteration 20. *Bottom row. Left:* After iteration 30 (final result). *Right:* Same as that in left, but the Virtual Scan is not shown

Fig. 11.

The NIST disaster area data set. **(a)** 6 example scans (from a total of 60). *Crosses* show each robot's position. **(b)** 60 scans superimposed using a rough pre-estimation using shape similarity among scans (Adluru et al. 2006). This is the initial global map for the experimental results shown in Fig. 12

Fig. 12.

(Color online) Alignment of NIST data set (initial alignment is shown in Fig. 11(b)) after (a) Iteration 4. (b) Iteration 8. (c) Iteration 10. In (a–c) *red lines* show the data of the Virtual Scan (VS). Please notice the mutual adjustment of hypotheses and real data. (d) Final result using Virtual Scans, after 100 iterations. The VS is not shown for clarity of display. (f) Final result of alignment without VS. Encircled areas show examples of improvement of (d) over (f). (e) The rectangle in the center could only be aligned correctly using the Virtual Scan. Please compare (e) to motivational examples in Figs. 1 and 3

Fig. 13.
Disaster City, the Texas Engineering Extension Service (TEEX) disaster simulation facility

Fig. 14.

Retrieving the scan in Disaster City. Photo of the scanned environment (the wall to the left appears at the top in Fig. 15, the drawer (*right*) appears on the bottom of Fig. 15). The foreground shows the simple scan-tool, a stick-mounted laser

Fig. 15. (Color online) Result of scanning the environment shown in photo (Fig. 14(c)). **(a)** 27 scans superimposed using the estimated odometry (human step length). **(b)** Iteration 10 of FFS/VS. *Red line* shows Virtual Scan. **(c)** Final result

Fig. 16. Postprocessing of FFS/VS result (Fig. 15(c)). *Left:* line segment representation. *Right:* simplified line segments

Fig. 17.

(Color online) Grid based evaluation. *Left:* Ground truth map. Two rooms (*green, top and bottom*) with obstacles (*gray*) are connected by two hallways (*green, center*). The victim (*red dot*) can only be reached from the current position (*black dot*) using the right hallway. *Center:* mapping example with high global and low local correctness. *Right:* mapping example with low global and high local correctness. A grid based map evaluation will prefer the center map, although e.g. for first responders the right map is of higher use

Fig. 18.

Pose based evaluation. The target map (*right*) is transformed to match the ground truth (*left*). The transformation parameters (here: *rotation, arrows*) are used to quantify the map quality. The topological correctness of the target map is reflected by the fact that only two rotations are needed to achieve the optimal map

Fig. 19.

Evaluation using FFS/VS. *Left:* Instead of creating a Virtual Scan, ground truth is used.

Right: the alignment energy gained from FFS is a measure for visual fitness of the evaluated global map

Fig. 20. Two maps for evaluation. (a) $R1$, (b) $R2$, (c), (d) show $R1$, $R2$ with superimposed ground truth map G

Fig. 21.

Alignment of the partitioned map G_d (a) to $R1$ and (b) to $R2$. (c) and (d): like (a), (b), but without underlying $R1$, $R2$

Fig. 22.

FFS/VS based hybrid evaluation. (a) Decomposed ground truth map, 3 parts. (b) Target map. (c) FFS/VS transforms (a) to (b). (d) Grid based evaluation on transformed parts. The final score is computed using task adjusted weights for transformation parameters and grid evaluation results