



12-2008

Uncertainty Minimization in Robotic 3D Mapping Systems Operating in Dynamic Large-Scale Environments

Sreenivas Rangan Sukumar
University of Tennessee - Knoxville

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss



Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Sukumar, Sreenivas Rangan, "Uncertainty Minimization in Robotic 3D Mapping Systems Operating in Dynamic Large-Scale Environments. " PhD diss., University of Tennessee, 2008.
https://trace.tennessee.edu/utk_graddiss/527

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Sreenivas Rangan Sukumar entitled "Uncertainty Minimization in Robotic 3D Mapping Systems Operating in Dynamic Large-Scale Environments." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Electrical Engineering.

Mongi A. Abidi, Major Professor

We have read this dissertation and recommend its acceptance:

Michael J. Roberts, Hamparsum Bozdogan, Hairong Qi, Andreas F. Koschan

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Sreenivas Rangan Sukumar entitled "Uncertainty Minimization in Robotic 3D Mapping Systems Operating in Dynamic Large-Scale Environments." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Electrical Engineering.

Mongi A. Abidi, Major Professor

We have read this thesis
and recommend its acceptance:

Michael J. Roberts

Hamparsum Bozdogan

Hairong Qi

Andreas F. Koschan

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Uncertainty Minimization in Robotic 3D Mapping Systems Operating in Dynamic Large-Scale Environments

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Sreenivas Rangan Sukumar
December 2008

Acknowledgements

Before this document starts dwelling on technical details, I would like to express my gratefulness to the supporting pillars of this dissertation work. I have to begin from where I began; my parents Mrs. Malathi Sukumar and Mr. Chellappa Sukumar. They have always been there for me with all their love and affection whenever I needed them emotionally, financially and every possible situation I imposed upon them. I owe them indefinite gratitude for their contribution to my career and for stimulating the quest for knowledge from a very young age. They knew I dreamt of being a scientist as a child, and today with the PhD degree, if I am closer to becoming one, it is because, I dreamed and they made sure I had everything necessary to realize my dream.

Dr. Abidi took over as the parent in graduate school, fuelled my quest for knowledge, supporting me technically and financially with the resources in the Imaging, Robotics and Intelligent Systems Lab (IRIS). Dr. Page, Dr. Koschan, Dr. Gribok and Dr. Besma were invaluable expertise that Dr. Abidi introduced me to and I attribute my technical development in the area of focus in this dissertation to them. In particular, I can never thank Dr. Page enough for the book-referrals, the brainstorming white board sessions, the personal career discussions and the weekly meetings inspiring the march towards “great” research; the impact of which, I am confident is dispersed in the content presented in this dissertation. I am also indebted to thank Dr. Bozdogan, Dr. Roberts and Dr. Qi, who in addition to serving on my committee, have been of significant influence in the development of the work presented. The foundation of this dissertation work builds over the concepts and tools learnt from their inspiring lectures.

My peers in the IRIS lab have been a great learning source in this research journey. From 2002 until the day this document becomes official, the healthy competition and synergy in the lab in the form of “Have you read this interesting paper?”, “Have you heard of this algorithm?”, “Here is a smart programming hack!”, and “Here is what we can do? How about working together on this idea?” “Did your paper make it to that journal?” and the umpteen “Eureka’s...!” have to be credited to fellow lab mates Brad Grinstead, Yohan, Faysal, Jessica (Yao Yi), Sijie, Cheng, Balaji, Santosh, Nikhil, Umayal and several others who I plead to kindly forgive me for not mentioning by name. It was a pleasure working with you all. Tak Motoyama, Doug Warren and Justin Acuff deserve a special mention. Thanks Tak and Doug for sharing your system instrumentation expertise in building our imaging prototypes and thanks Justin for meticulously solving one software system issue after another.

Sincere thanks to you all.

Abstract

This dissertation research is motivated by the potential and promise of 3D sensing technologies in safety and security applications. With specific focus on unmanned robotic mapping to aid clean-up of hazardous environments, under-vehicle inspection, automatic runway/pavement inspection and modeling of urban environments, we develop modular, multi-sensor, multi-modality robotic 3D imaging prototypes using localization/navigation hardware, laser range scanners and video cameras.

While deploying our multi-modality complementary approach to pose and structure recovery in dynamic real-world operating conditions, we observe several data fusion issues that state-of-the-art methodologies are not able to handle. Different bounds on the noise model of heterogeneous sensors, the dynamism of the operating conditions and the interaction of the sensing mechanisms with the environment introduce situations where sensors can intermittently degenerate to accuracy levels lower than their design specification. This observation necessitates the derivation of methods to integrate multi-sensor data considering sensor conflict, performance degradation and potential failure during operation.

Our work in this dissertation contributes the derivation of a fault-diagnosis framework inspired by information complexity theory to the data fusion literature. We implement the framework as opportunistic sensing intelligence that is able to evolve a belief policy on the sensors within the multi-agent 3D mapping systems to survive and counter concerns of failure in challenging operating conditions. The implementation of the information-theoretic framework, in addition to eliminating failed/non-functional sensors and avoiding catastrophic fusion, is able to minimize uncertainty during autonomous operation by adaptively deciding to fuse or choose believable sensors. We demonstrate our framework through experiments in multi-sensor robot state localization in large scale dynamic environments and vision-based 3D inference. Our modular hardware and software design of robotic imaging prototypes along with the opportunistic sensing intelligence provides significant improvements towards autonomous accurate photo-realistic 3D mapping and remote visualization of scenes for the motivating applications.

Publication History

Accepted/ Published as of October 2008

Book chapter

S.R. Sukumar, D.L. Page, A. Koschan, and M.A. Abidi, “Under vehicle inspection with 3D Imaging,” in *3D Imaging for Safety and Security*, A. Koschan, M. Pollefeys, and M. Abidi (Eds.), Springer, Dordrecht, The Netherlands, pp. 249-278, 2007.

Journals

S.-J. Yu, **S. R. Sukumar**, A. Koschan, D. L. Page, and M.A. Abidi, “3D reconstruction of road surface using an integrated multi-sensory approach,” *Optics and Lasers in Engineering*, Vol. 45, No. 7, pp. 808-818, 2007.

B. Grinstead, **S. R. Sukumar**, D. L. Page, A. F. Koschan, and M. A. Abidi, D. Gorsich, “Mobile scanning system for the fast digitization of existing roadways and structures,” *Sensor Review Journal*, Vol. 26, No. 4, pp. 283-289, 2006.

S. R. Sukumar, D. L. Page, A. V. Gribok, A. F. Koschan, M.A. Abidi, D. J. Gorsich, and G. R. Gerhart, “Robotic 3D imaging system for under vehicle inspection,” *Journal of Electronic Imaging*, Vol. 15, No. 3, 033008, 2006.

S. R. Sukumar, P. Govindasamy, A. Koschan, D. L. Page, M.A. Abidi, “Thermal modeling and simulation of as-built automotive parts,” accepted to the *Virtual and Physical Prototyping Journal*.

Refereed Technical Conferences

S. R. Sukumar, H. Bozdogan, D. L. Page, A. Koschan, B. Abidi, M. A. Abidi, “Uncertainty minimization in multi-sensor localization systems using model selection theory,” *to appear in the Proc. of the IEEE International Conference on Pattern Recognition (ICPR)*, December 2008.

S. R. Sukumar, H. Bozdogan, D. L. Page, A. F. Koschan, and M. A. Abidi, “Learning structurally discriminant features in 3D faces,” *in the Proc. of the IEEE International Conference on Image Processing (ICIP)*, pp. 1912-1915, October 2008.

S. R. Sukumar, H. Bozdogan, D. L. Page, A. F. Koschan, and M. A. Abidi, “On handling uncertainty in the fundamental matrix for scene and motion adaptive pose recovery,” *in the Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, June 2008.

Y. Yao, **S. R. Sukumar**, B. Abidi, D. Page, A. Koschan, M.A. Abidi, “Automated scene specific selection of feature detectors for 3D face reconstruction,” *in the Proc. of the 3rd International Symposium on Visual Computing*, Lake Tahoe, November 2007, Proceedings also appears as Lecture Notes in Computer Science, Advances in Visual Computing, Vol. 4841, pp. 476-487, 2007.

- S. R. Sukumar**, D. L. Page, A. Koschan, M.A. Abidi, “MuFeSaC: Learning when to use which feature detector”, in the *Proc. of the IEEE International Conference on Image Processing (ICIP)*, San Antonio, Vol. VI, pp. 149-152, September 2007.
- S. R. Sukumar**, H. Bozdogan, D. L. Page, A. F. Koschan, and M. A. Abidi, “Sensor selection using information complexity for multi-sensor mobile robot localization,” in the *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, Roma, Italy, pp. 4158-4163, April 2007.
- S. R. Sukumar**, D. L. Page, A. V. Gribok, A. F. Koschan, and M. A. Abidi, “Shape measure for identifying perceptually informative parts of 3D objects,” in the *Proc. of the IEEE 3rd International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, Chapel Hill, North Carolina, pp. 679-686, June 2006.
- D. L. Page, A. Koschan, **S. R. Sukumar**, B. R. Abidi, and M. A. Abidi, “Shape analysis algorithm based on information theory,” in the *Proc. of the International Conference on Image Processing (ICIP)*, Vol. I, pp. 229-232, September 2003.

Workshops/Topical meetings/ Conferences

- S. R. Sukumar**, D. L. Page, A. F. Koschan, and M. A. Abidi, “Towards understanding what makes objects appear simple or complex,” in the *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Perceptual Organization in Computer Vision Workshop (POCV-CVPR)*, Anchorage, June 2008.
- S. R. Sukumar**, S.-J. Yu, D. L. Page, A. F. Koschan, and M. A. Abidi, “Multi-sensor integration for unmanned terrain modeling,” in the *Proc. SPIE Unmanned Systems Technology VIII*, Vol. 6230, Orlando, FL, pp. 65-74, April 2006.
- A. F. Koschan, P. Govindasamy, **S. R. Sukumar**, D. L. Page, M.A. Abidi, and D. Gorsich, “Thermal modeling and imaging of as-built vehicle components,” in the *Proc. of the SAE World Congress 2006: Driving Innovation through Partnerships, Cobo Center, Detroit, Michigan, April 2006*, SAE Technical Paper 2006-01-1167 (also appears in the book SP-2040 of the SAE).
- S. R. Sukumar**, D. L. Page, A.V. Gribok, A. F. Koschan, and M. A. Abidi, “Intelligent 3D sensing for robotic inspection of hazardous facilities,” *Transactions of the American Nuclear Society*, Vol. 92., San Diego CA, USA, pp. 54-55, June 2005.
- S. R. Sukumar**, D. L. Page, A.V. Gribok, A. F. Koschan, M. A. Abidi, D. Gorsich, and G. Gerhart, “Surface shape description of 3D data from the under-vehicle inspection Robot,” in the *Proc. of SPIE Unmanned Ground Vehicle Technology VII*, Vol. 5804, Orlando, FL, USA, pp. 621-629, March 2005.

Note: Text and images appearing in this document may appear in copyrighted publications listed above.

Contents

1. Introduction	1
1.1 Motivation.....	1
1.2 Problem statement.....	4
1.3 Areas of contribution.....	9
1.4 Document organization.....	10
2. Background : Robotic Mapping	11
2.1 3D sensing technologies.....	11
2.2 Robotic 3D mapping.....	14
2.3 Handling uncertainty in robotic mapping.....	25
3. System : Modular Design	33
3.1 Modular architecture in system design.....	33
3.2 Hardware and software.....	38
3.3 Multi-modality multi-sensor data integration.....	40
3.4 Belief propagation theory for modeling and handling uncertainty.....	42
4. Theory : Selfish-Altruist Fusion	49
4.1. Philosophical inspiration for statistical inference.....	49
4.2. Our proposed approach.....	50
4.3. Quantifying sensor self confidence.....	57
4.4. Quantifying co-operative confidence.....	61
4.5. Resolving the fusion versus selection dilemma.....	69
5. Experiments : Handling Uncertainty	71
5.1. Simulation study	71
5.2. Handling localization uncertainty in instrumentation-based mapping systems.....	75
5.3. Handling localization uncertainty in image-based mapping systems.....	81
5.4. Summary.....	95
6. Applications : Research Impact	96
6.1. Under vehicle inspection.....	96
6.2. Airport runway/Pavement crack inspection.....	104
6.3. Large scale terrain mapping.....	108
7. Conclusions	112
7.1. Dissertation key points.....	112
7.2. Future directions.....	113
Bibliography.....	115
Vita.....	125

List of Tables

Table 3.1: <i>Kalman filter algorithm adapted from (Thrun et al., 2005).</i>	45
Table 4.1: <i>Model selection criteria derived out of minimizing Equation 4.11.</i>	58
Table 4.2: <i>Co-operative team of sensor clusters in the validity inference for a 3-sensor system.</i>	62
Table 4.3: <i>Possible hypotheses to evaluate for sensor reliability.</i>	63
Table 4.4: <i>Computing the log-likelihood term in Equation 4.25 for the sensor validity score.</i>	66
Table 4.5: <i>Computing the $C_1(\hat{F}^{-1})$ term in Equation 4.25 for sensor the validity score.</i>	67
Table 5.1: <i>Parameter parsimony penalty in the model consensus score.</i>	89

List of Figures

Figure 1.1: <i>The motivating application illustrated in this figure is under-vehicle inspection.</i>	2
Figure 1.2: <i>The motivating application illustrated in this figure is autonomous localization and mapping in dynamic unstructured environments.</i>	3
Figure 1.3: <i>The motivating application illustrated in this figure is road surface inspection for airport runways and distressed pavement highways.</i>	5
Figure 1.4: <i>The motivating application illustrated in this figure is real-world large-scale terrain mapping of environments for finite-element based vehicle-terrain simulators and driver-behavior analysis simulators.</i>	6
Figure 1.5: <i>Block diagram summarizing the efforts in this dissertation emphasizing contributions towards reconstructing 3D models of real world environments using mobile robotic platforms with autonomous capability.</i>	8
Figure 2.1: <i>The classification of popular 3D imaging methods based on the physics of range sensing.</i>	11
Figure 2.2: <i>Scan line (range profile) matching to estimate the motion of the robot.</i>	17
Figure 2.3: <i>Block diagram for image-based motion and structure estimation following (Pollefeys et al., 1999).</i>	18
Figure 2.4: <i>Pictorial description of structure and motion estimation algorithm on video frames.</i>	19
Figure 2.5: <i>Inlier classification using RANSAC.</i>	20
Figure 2.6: <i>Area of interest to evaluate different techniques for 3D mapping.</i>	23
Figure 2.7: <i>Comparison of 3D imaging techniques.</i>	24
Figure 2.8: <i>GPS localization and the uncertainty about the position estimate in a test dataset in Knoxville, Tennessee.</i>	26
Figure 2.9: <i>Image-based pose recovery in a challenging environment.</i>	27
Figure 2.10: <i>The line fitting example to show inlier-outlier distributions that introduce uncertainty into the image-based geometric inference.</i>	28
Figure 2.11: <i>Summary of uncertainty minimization methodologies used in robotic mapping and navigation.</i>	29
Figure 3.1: <i>The sensor bricks fit into mobility platforms without having to reconfigure hardware or software in communicating information.</i>	34
Figure 3.2: <i>The “sensor brick” architecture lends to different levels of modularity.</i>	36

Figure 3.3: <i>The “sensor brick” architecture lends to different levels of modularity with large-scale mapping also.</i>	37
Figure 3.4: <i>Specification of the components in our modular approach along with some design notes towards reproducing our system.</i>	39
Figure 3.5: <i>The block diagram for integrating multi-sensor data into a 3D model.</i>	40
Figure 3.6: <i>Spatial integration of multi-sensor data requires a global reference frame and interpolation to consider different sampling rates of sensors.</i>	41
Figure 3.7: <i>Uncertainty in the 3D map is a combination of error in the localization and the measurement uncertainty in the structure recovery sensors.</i>	42
Figure 3.8: <i>The simplified Bayes network representing localization as a recursive state estimation problem.</i>	43
Figure 3.9: <i>An example for formulating the motion model for a robot with three degrees of freedom in motion.</i>	46
Figure 4.1: <i>Several sensors measure the state of our system with an uncertain belief.</i>	51
Figure 4.2: <i>Block diagram of the proposed selfish-altruist fusion algorithm.</i>	53
Figure 4.3: <i>Propagation of uncertainty in multi-sensor belief propagation systems.</i>	54
Figure 4.4: <i>Perceiving fusion as a model selection problem.</i>	56
Figure 4.5: <i>Understanding the measurement uncertainty score using a simple example of covariance conflict with several sensors.</i>	60
Figure 4.6: <i>Understanding sensor validity estimation.</i>	61
Figure 4.7: <i>Perceiving sensor validity computation as a multi-sample clustering problem.</i>	65
Figure 4.8: <i>Our formulation of M_i and V_i is able to identify the separability in the mean and the covariance.</i>	68
Figure 5.1: <i>Localization in a simulation environment.</i>	72
Figure 5.2: <i>The study on how the indirect parameters might affect inference.</i>	74
Figure 5.3: <i>Localization in urban environments.</i>	76
Figure 5.4: <i>Reducing the uncertainty about localization improves the integrated map.</i>	77
Figure 5.5: <i>Localization of a mobile robot with multiple cameras in an indoor environment.</i>	78
Figure 5.6: <i>Localization and mapping of a mobile robot with a camera and a range sensor in an indoor environment.</i>	79

Figure 5.7: <i>There are several options available in the flow process for estimating the fundamental matrix.</i>	81
Figure 5.8: <i>Revisiting epipolar geometry.</i>	82
Figure 5.9: <i>Different feature detectors identify different monocular depth cue pixels (shown in red).</i>	83
Figure 5.10: <i>Different forms of the motion model \mathbf{M} for the fundamental matrix relating image frames.</i>	85
Figure 5.11: <i>Our approach to improve the reliability in the fundamental matrix by generating convergence uncertainty statistics.</i>	87
Figure 5.12: <i>Our mobile robotic platform with vision-based pose recovery navigates near the Math building in our University.</i>	92
Figure 5.13: <i>Autonomous robot localization with a laser scanner and a vision sensor.</i>	93
Figure 5.14: <i>Vision-based localization in dynamic environments.</i>	94
Figure 5.15: <i>Timing analysis for learning competing models.</i>	95
Figure 6.1: <i>The traditional method for inspecting a vehicle undercarriage is to use a mirror attached to the end of stick.</i>	97
Figure 6.2: <i>An alternative to a mirror-based inspection is a low-profile mobile robotic platform (called SafeBot) that can navigate under a vehicle while inspection personnel remain at a safe distance.</i>	97
Figure 6.3: <i>The images on the right are high resolution mosaics of visual and thermal data from the robot scanning the carriage.</i>	98
Figure 6.4: <i>3D under-vehicle inspection using our prototypes.</i>	100
Figure 6.5: <i>Visualization of multi-sensor data on the 3D geometry helps isolate radioactive sources in the scene.</i>	101
Figure 6.6: <i>Both the 3D scanners lend to easy texture mapping.</i>	102
Figure 6.7: <i>Scene verification approach for detecting changes made to the vehicle by comparing with archived scans.</i>	103
Figure 6.8: <i>Summary of technologies demonstrated for road surface inspection.</i>	105
Figure 6.9: <i>Multi-modal integrated 3D data of an area of interest with three small zoomed in sections of areas with different roughness and depth of cracks.</i>	106
Figure 6.10: <i>Large areas digitized at very high resolution.</i>	107
Figure 6.11: <i>Large areas of urban environments digitized at very high geometric resolution with high fidelity texture.</i>	110

Figure 7.1: <i>Preliminary results implementing our framework to guide camera hand over for target localization.</i>	114
---	-----

1 Introduction

Targeting military safety and civilian security, we have witnessed the implementation of imaging and automation solutions in several public access points. Some that we notice in our daily life include X-ray baggage inspection and breach detection at airports, under-vehicle inspection at check-points, multi-camera surveillance for asset monitoring, and multi-modality biometric recognition among several others. This dissertation research is motivated by four such safety and security related projects. In Section 1.1, we detail the motivation behind this dissertation work along with the research objectives. Then, we formulate these objectives into a problem statement and summarize the solution this dissertation provides in Section 1.2 before finally claiming system level and technical contributions for each one of these applications in Section 1.3.

1.1 Motivation

The four applications motivating this dissertation are (1) under-vehicle inspection, (2) robotic mapping in hazardous large-scale environments, (3) runway/pavement inspection and (4) terrain mapping for simulators in virtual proving grounds. We discuss the research objectives in each of these applications in the following paragraphs and establish a basis to showcase our research efforts.

- ***Under-vehicle inspection:*** The goal of this research project under the United States (U.S) Army's Security Automation for Future Electromotive Robotics (SAFER) program is towards the deployment of intelligent robots for under-vehicle inspection at check-points in war-torn countries, gate-entry terminals and parking lots (Figure 1.1). Using multi-modality measurements of temperature, range, color, radioactivity and with future potential for chemical and biological sensors, our work is a small part of the modular robotic "sensor brick" architecture (Sukumar et al., 2007a) that integrates multi-sensor data into scene intelligence in remote 3D virtual reality environments. In transforming the inspection task into an unmanned robotic mission, the focus related to this dissertation work is on the design, development and deployment of the 3D range "sensor brick" as a vital autonomous component in this multi-sensor robotics framework and also to demonstrate the potential of automatic threat detection using the geometric information from the 3D sensors (Sukumar et al., 2006a).
- ***Robotic mapping in hazardous environments:*** The U.S. Department of Energy under the University Research Program in Robotics (URPR) requires autonomously navigating robots for mapping large-scale radioactive hazardous environments for future cleanup and maintenance (Figure 1.2). Considering the mature field of mobile robotics, the challenge we address is to extend beyond structured environments that present day robots can handle to unstructured dynamic environments (Sukumar et al., 2007b). The goal is to develop systems that can navigate, map and communicate in challenging environments without having to worry about sensor failure and insufficiencies. This dissertation presents a multi-modality multi-sensor data fusion solution for such a goal by implementing a statistical framework generic for fusing uncertain, imprecise and conflicting data demonstrating improved reliability in the autonomous operation.

Areas for deployment: Gate-entry terminals, war zone check-points, parking lots



State-of-the-art: Mirror on the stick, mosaic from video acquired by a robot

View in the mirror



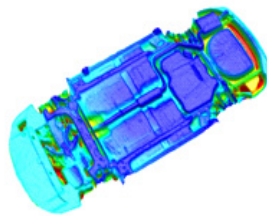
Mosaic generated in a single pass.



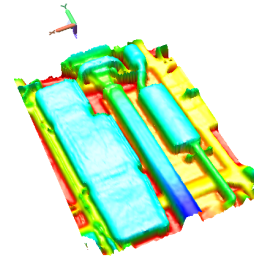
Image from (Ng, 2003)

Our solution: Autonomous mobile robot maps the entire under carriage

Mapping robot



Underside of a SATURN car



Underside of a DODGE Van

Figure 1.1: The motivating application illustrated in this figure is under-vehicle inspection. We have shown the potential areas for the deployment of the technology in gate entry terminals, war-zone check points and parking lots. The contemporary approach for under-vehicle inspection is using a mirror on a stick or a mobile robot mounted with a visual camera. The mirror-on-the stick lacks the resolution and clarity for detecting threats while the camera approach has field of view and illumination issues. We overcome these limitations with several other enhancements with our robotic 3D mapping solution.

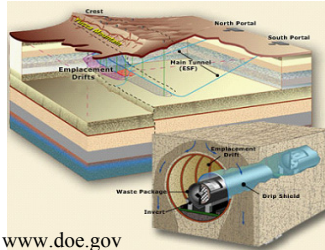
Areas for deployment: Robotic surveillance of hazardous environments

High radioactivity zones



Images source: www.doe.gov

Nuclear waste dump-sites



State-of-the-art: Instrumentation approach, video-based systems

Ayres Hall, aerial view



Instrumentation-based approach

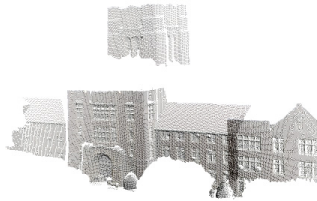


Image-based approach



Our solution: System with autonomous navigation and mapping capability



Navigation and mapping in dynamic environments

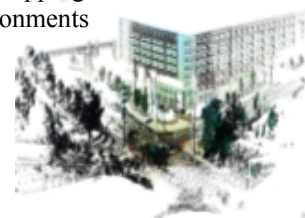


Figure 1.2: The motivating application illustrated in this figure is autonomous localization and mapping in dynamic unstructured environments. With focus on environments where we cannot assume satellite reception of global position systems, we combine the state-of-the-art instrumentation-based approach with passive image-based localization and mapping techniques. The instrumentation hardware and the cameras along with the algorithms for smart integration of data from complementary sensors increase 3D mapping capability in dynamic operating conditions.

- **Airfield/Pavement distress:** In a more relaxed form of robotic mapping, another application we address is airfield runway/pavement distress inspection (Figure 1.3). Traditionally, general aviation airfield pavements are maintained based on the inspection staff's judgment and experience. The inspection personnel walk or drive slowly through asphalt and concrete pavements observing surface defects and degradation to make recommendations for immediate and long term maintenance. The manual inspection procedure is not only cumbersome, time consuming and expensive but is also susceptible to human error and inefficiency. With safety of aircrafts and passengers in mind, this functional and important process of inspection can be significantly improved using a formalized imaging system that will ease the effort required to inventory airfield conditions through periodic evaluations and subsequent surface distress management. We have demonstrated a 3D imaging system (Sukumar et al., 2006b and Yu et al., 2007) that can measure the depth of cracks as a significant illumination independent improvement in automatically detecting and analyzing cracks compared to camera-based systems that are still under development. The important constraint with this application is to map large-scale environments with micro-scale accuracy.
- **Terrain mapping for simulators/ Virtual proving grounds (VPG):** The U.S Army has constructed several proving grounds scattered all over the United States. The idea being that army vehicles (tanks and carrier equipment) can be tested for wear and tear on rough and uncertain terrain before deployment in real-world scenarios. The testing is performed by driving assembled and fully equipped vehicles across different types of surfaces taking measurements for fatigue, damage, and deformation of different components along with the driver behavior in the automobile. Recently, the U.S Army concluded that experiments using real vehicles is very expensive, placing the testing soldier safety under concern. The solution we proposed (Sukumar et al., 2006b) was to help the move towards virtual reality testing (Zhang et al., 1999) using vehicle dynamics models of automobiles on 3D models of real terrain. The terrain models for these vehicle-terrain simulators are presently generated using 1D profilometers that capture limited data and do not include the real-world uncertainty into the simulation. We address this need for generating immersive 3D models of large-scale dynamic environments such as hilly terrain, a speed breaker, and a gravel road etc. from the real world with sufficient accuracy for finite-element based simulations.

Figures 1.1 through 1.4 illustrate the motivating application, the state of the art and the improvement that we provide as solutions for each of these applications.

1.2 Problem Statement

Initially, though the applications described in Section 1.1 appear unrelated, we note that the solutions we have demonstrated in (Sukumar et al., 2006a), (Sukumar et al., 2006b) and (Grinstead et al., 2006) for these applications involves the design of a manned/unmanned/tele-operated mobile platform, with a suite of sensors collecting data in real time, that on further processing, can deliver geometrically accurate, geographically meaningful photo-realistic 3D models of the scenes of interest. In studying the state-of-the-art and identifying the potential enhancements with 3D imaging, we realize the need for the deployment of multiple multi-modality sensors like inertial navigation systems (INS), global positioning systems (GPS) in addition to vision-based systems for high accuracy self-localization, mapping and photo-realistic visualization. In this dissertation, we build prototypes for real time acquisition and processing towards 3D mapping and take on the challenge of implementing intelligence into these robots to operate autonomously. The intelligence that we desire is that of opportunistic sensing with the ability to evolve a belief and data fusion policy on sensors within a multi-agent framework to survive and counter concerns of failure in challenging operating conditions.

Areas for deployment: Airport runways and road pavements



Airport runway

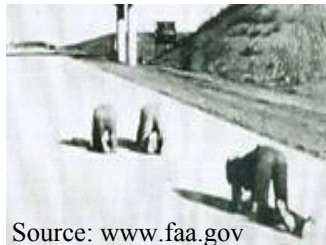


Different types of pavement distress

Source: (Walker, 2004)

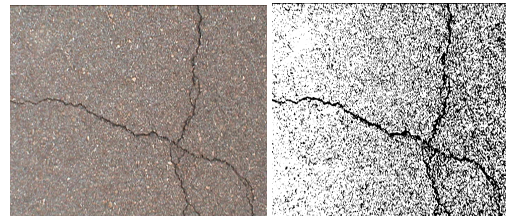


State-of-the-art: Manual inspection, video-based inspection



Source: www.faa.gov

Image processing approach to crack detection



Our solution: Mobile van maps the asphalt surface to detect cracks in 3D.



Depth information is used for detection and classification.

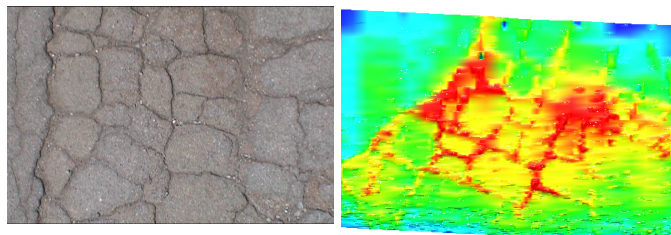


Figure 1.3: The motivating application illustrated in this figure is road surface inspection for airport runways and distressed pavement highways. The aim is to automatically identify different types of cracks and recommend necessary action. Manual inspection by an expert appears to be the state-of-the-art. Video-based image processing systems, still under development, are capable of detecting cracks but cannot recover the depth of cracks accurately. Our mobile mapping system that can scan asphalt surfaces at normal driving speeds avoids the difficulty and susceptibility with manual inspection and improves over video-based techniques by detecting even 2mm deep cracks. Crack classification over large areas is also possible with our approach.

Areas for deployment: Terrain for virtual proving grounds and driving simulators

Source: <http://www.apg.army.mil>



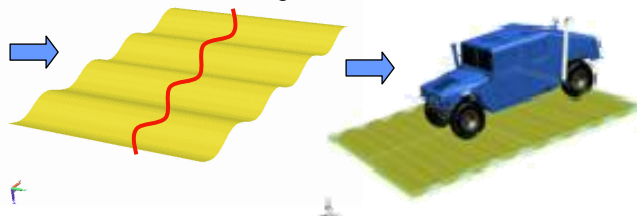
Real environments in simulators



State-of-the-art: Profilometers generate statistical models of terrain



Terrain generated from the statistical models of profilometer data



Our solution: Our system generates real terrain for immersive environments

Real 3D terrain for vehicle-terrain interaction analysis

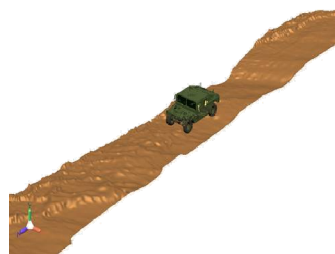


Figure 1.4: The motivating application illustrated in this figure is real-world large-scale terrain mapping of environments for finite-element based vehicle-terrain simulators and driver-behavior analysis simulators. We have shown a picture of an army vehicle tested in the Aberdeen proving grounds in Maryland, USA along with the virtual proving grounds as a safe and inexpensive approach to vehicle testing and analysis. The bottom insets show our effort in mapping terrain from the real-world and delivering immersive 3D models for such simulators.

The recent trend with building autonomous robotic navigation and mapping systems is to use multi-modality multi-sensor systems (DARPA Urban Challenge, 2007). We follow the trend and design multi-sensor systems that execute motion commands on the robot towards being in a required or pre-planned state exploiting its sensors for both gaining new information in an unknown environment and also as feedback to confirm the desired state. For example, with the 3D mapping systems that we have built, the mobile robotic platform would be loaded with pre-defined position landmarks that we intend the robot to visit sequentially. GPS sensors would indicate the global position of the system, a video camera in addition to details of the scene in combination with the ego-motion estimation algorithm would also provide estimates of relative position and motion. The robot, based on the feedback from the vision algorithm and the availability of GPS signals, would then decide what confident action to take to reach the desired next state. Such a scenario is popularly known as the recursive state-space estimation problem of global localization.

In an ideal noiseless world, if all systems were perfect, there would be no need to handle uncertainty. But there is uncertainty about how the system changes its state and how the sensors measure state. Sensors are subject to situations that perturb their measurements in unpredictable ways. In addition, sensors fail in uncertain dynamically-changing environmental conditions and detecting such a fault is extremely difficult (Thrun et al., 2005). For example, the GPS depends on the number of satellites available at the time of operation and is known to operate with lesser accuracy in areas with tall buildings, while INS is notorious for the drift over time and temperature. Vision and range-based mechanisms depend on the environment also, but can be a reliable complementary pose recovery method in real world situations.

The performance of these sensors being environment dependent, we have to realize that not all sensors can be believed at a particular instant. This necessitates the development of a procedure that can infer believable sensors on the go, minimizing global uncertainty introduced by the dynamism of the environment in real-world situations. The probabilistic Bayes network framework with Markov assumptions (Thrun et al., 2005) for noise-variance minimization includes methods for dealing with such situations and in maximizing belief about the state of the system. Proven to be robust in most controlled structured environments, our experience with implementing the probabilistic algorithms discussed in (Thrun et al., 2005) to the large-scale real-world applications exposes the assumption about bounded effect of environment dynamism. Several researchers also agree with our observation. (Urmson et al., 2008; Bacha et al., 2008). Hence, we have developed the integration pipeline illustrated in Figure 1.5 that builds on the robust probabilistic reasoning methods.

Our goal in this dissertation is towards the design of an algorithm for including uncertainty in sensors along with the inference about uncertainty introduced by the dynamic nature of the environment. We assume that we have several data sources (sensors) that measure the state of a system of which some sensors (at the least one) is/are operating with believable accuracy. We also allow sensors to operate with intermittent high belief with short-term failure and recovery. Our objective then is to find the state of the system with maximum belief while operating even in constantly changing environments.

Mathematically, the problem that we are trying to solve can be formulated as shown in Equation 1.1. We only present the problem statement to introduce our solution in this Chapter but explain the procedure for implementation in detail in Chapter 4.

$$\text{Given } \mathbf{S} = \{ S_i \}, i = 1, 2, \dots, N \text{ where } S_i(x) = \hat{x}_i \quad (1.1)$$

$$\text{Find } \arg \max_{\tilde{x}, f(S_i, \{B_i\})} \text{Belief}(\tilde{x}) \quad (1.2)$$

where f is the fusion functional, B_i is the belief about sensor measurements of state $x = \hat{x}_i$ from sensor S_i and \tilde{x} is the best estimate from sensor feedback after multi-sensor fusion.

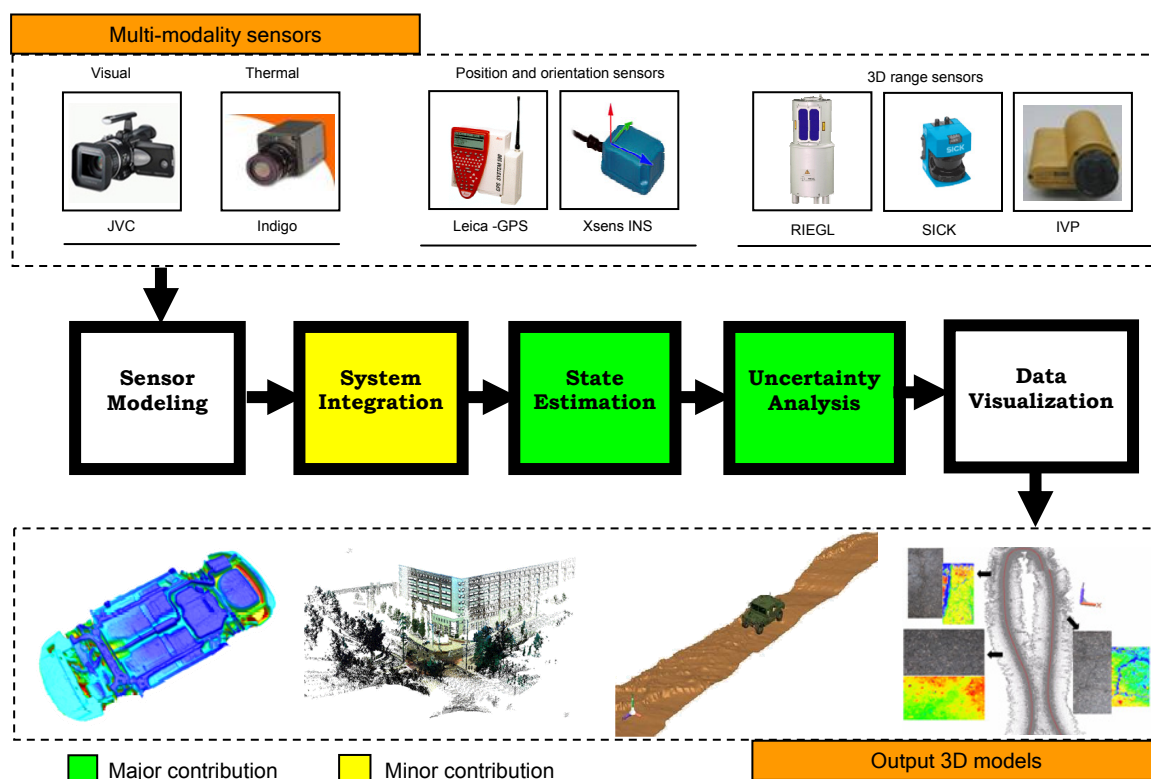


Figure 1.5: Block diagram summarizing the efforts in this dissertation emphasizing contributions towards reconstructing 3D models of real world environments using mobile robotic platforms with autonomous capability. The data from multi-modality sensors like cameras, GPS, INS and laser range scanners are used for state and structure estimation. The contributions of this dissertation are in using the data from these sensors along with an uncertainty analysis framework to make these systems operational in dynamic unstructured environments with high degree of accuracy. Each of these contributions will be explained in detail in individual chapters. Chapter 3 covers the system architecture. Chapter 4 derives the selfish-altruist uncertainty handling scheme for multi-sensor localization systems. Chapters 5 and 6 introduce the theory into the real world documenting our experimental success.

Towards the goal formulated in Equations 1.1 and 1.2 to choose believable sensors performing consistently at a given instant in time in an unknown environment, we derive scores for self confidence and co-operative confidence to evaluate success in fusion. By including the best case apriori bounded-noise models to propagate belief as the system state changes, we provide definitions to estimate sensor measurement uncertainty and sensor validity in such state-space systems. Being able to infer sensor validity from the bounded noise belief is one of our major contributions with this work. For quantifying measurement uncertainty we score the Bayesian belief probability density using a model selection criterion, and for sensor validity, we evaluate belief on pose estimates from different sensors as a multi-sample clustering problem. The minimization of the combined uncertainty allows us to intelligently choose a subset of sensors that contribute to improved state estimation. Our theory for determining the success in fusion is based on information complexity with an additional inference mechanism to resolve the dilemma of fusion versus selection. We demonstrate our approach in robot localization using GPS, INS, and video and range sensors in large-scale dynamic environments, robust feature selection for geometric inference and finally propose to apply the statistical framework for target localization in multi-camera networks.

1.3 Areas of contribution

Motivated towards addressing the needs in the applications mentioned in Section 1.1, our research demonstrates the convergence of relatively new 3D sensing technology, modular robotics, 3D computer vision algorithms and probabilistic reasoning. In making such a claim with this dissertation, we list the following contributions addressing key technical issues in 3D reconstruction using autonomous robotic systems.

- **3D imaging and automation solutions:** The system-level contribution that this dissertation claims is the readily deployable prototypes built as automation solutions to address safety and security applications described in Section 1.1. The evaluation efforts toward hardware design involving the choice of equipment mountable on mobile robotic platforms considering repeatability, reliability and robustness to structured and unstructured real world environments combines with the supporting software development for data acquisition, processing, data fusion and visualization in outperforming state-of-the-art and commercial systems with significant improvements.
- **Uncertainty management in multi-sensor state localization:** Belief propagation methods are the state-of-the-art with multi-sensor state localization problems. However, when localization applications have to deal with multi-modality sensors whose functionality depends on the environment of operation, we understand the need for an inference framework to identify confident and reliable sensors. We derive an information-theoretic statistical framework to resolve sensors in conflict during autonomous operation in (Sukumar et al., 2007b). Such a framework helps eliminate failed/non-functional sensors from the fusion process minimizing uncertainty while propagating belief in real world multi-sensor state localization.
- **Reliability in image-based state estimation:** Vision sensors being an indispensable component in our system, we study the key step of estimating the fundamental matrix in feature-based camera ego-motion computation for scene modeling and vehicle navigation. We present a new method of analyzing and further reducing the risk in the fundamental matrix due to the choice of a particular feature detector, matching algorithm, the motion model, and iterative hypothesis generation/verification paradigms (Sukumar et al., 2008). Our scheme makes use of model-selection theory guiding the switch to optimal methods within the hypothesis-and-test architecture. We demonstrate our method for vision-based robot localization in large-scale environments where the target scene is dynamic and navigation within the environment becomes challenging.

1.4 Document organization

The remainder of this document is arranged as follows:

- **Chapter 2** presents the background of how ideas in the literature help us design multi-modality multi-sensor 3D robotic mapping systems and emphasizes the need to handle uncertainty in real world situations for these robots to operate autonomously.
- **Chapter 3** discusses the system architecture governing our prototypes. We also document the characterization of the sensors, uncertainty propagation in how data from different sensors are aligned and integrated into a 3D virtual environment.
- **Chapter 4** takes over from the issues and shortcomings with implementing methods in Chapter 3 amidst sensor conflict and derives the theory to handle information fusion in such situations.
- **Chapter 5** documents the implementation of the theory derived in Chapter 4 in simulated and real world situations. We present promising results in hybrid pose recovery for unmanned mobile robots in large-scale dynamic environments and improved image-based geometric inference. We show the improvement in global localization of mobile robots by demonstrating the ability to detect faulty sensors and also guide the switch to the available next best sensor.
- **Chapter 6** showcases the research impact of the work presented by deploying and demonstrating our methods in real-world situations directly relating to the motivating applications introduced in Chapter 1.
- **Chapter 7** concludes with the dissertation key points identifying avenues for extending our generic theory beyond multi-sensor state localization of mobile robots to other applications.

2 Background: Robotic Mapping

In the previous chapter, we introduced the applications that motivate this dissertation and emphasized the need for imaging-based robotic automation solutions. In this chapter, we will briefly discuss the contemporary 3D sensing methods and dwell on the state-of-the-art with robotic mapping. The study helps us identify the gaps in the literature motivating our work towards claiming the convergence of 3D imaging techniques and mobile robotics for high accuracy mapping of large-scale environments. We begin with a survey of 3D sensing techniques in Section 2.1, brief 3D mapping methods using mobile robots in Section 2.2 and document the different sources of uncertainty while mapping the real world in Section 2.3. Finally, we summarize methods to handle such uncertainty in large-scale environments in Section 2.4.

2.1 3D sensing technologies

In this subsection, we discuss 3D sensing methods and weigh their suitability for our mobile mapping robots. We begin by presenting a classification of different techniques in Figure 2.1. The classification is based on a study similar to Blais’s review (2004) on 3D range sensing in which different methods of 3D sensing fall into two categories as passive and active.

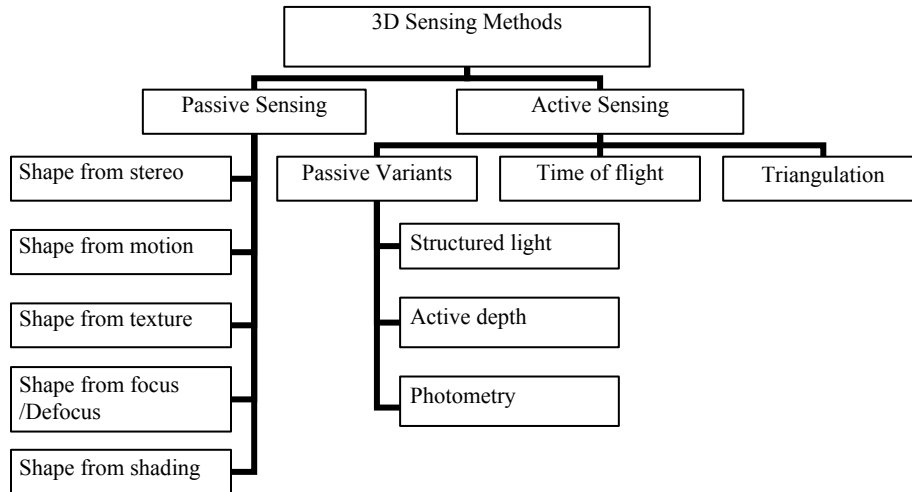


Figure 2.1: The classification of popular 3D imaging methods based on the physics of range sensing. Of these methods, we find the time-of-flight approach and the triangulation approach meeting accuracy, resolution, and real-time acquisition requirements for our large-scale mapping applications. The passive image-based pose and 3D structure recovery, particularly shape from motion, shows immense potential as a relatively inexpensive solution for mobile robots.

Let us begin our discussion with passive techniques first. Passive triangulation by stereo is similar to the way humans perceive depth and involves two cameras taking a picture of the same scene from two different locations at the same time. Just like our eyes, image-based passive 3D reconstruction methods take 2D pictures as projective inputs of the 3D world and recover depth using computational substitutes of human perception. One computational approach is to estimate depth information by matching correspondences in the images from the two cameras and applying epipolar geometry. An extension of passive stereo that uses only a single camera is the shape from motion method. Shape from motion algorithm is also based on epipolar geometry, the difference being frames in a video are considered as data of the same scene taken from different viewpoints.

The other computational approach avoids matching individual pixels and instead models disparity between stereo pairs into a regularized global energy function that is iteratively optimized for a tradeoff between intensity disparity and smoothness support from neighboring pixels. With additional knowledge of camera parameters and focal length the disparity at each pixel estimated using the energy function is converted to range measurements. Passive triangulation algorithms, both shape from stereo and shape from motion, are challenged by the ill-posed problem of correspondence in stereo matching. We will be discussing shape from motion algorithms and epipolar geometry in detail in the later sections of the chapter. We will attempt to address the reliability issues in recovering geometry amidst ill-posedness when we revisit the topic in Chapter 5.

Another idea to extract 3D shape is by using the principle of focusing and defocusing. The method infers range from two or more images of the same scene, acquired under varying focus settings. By varying the focus of a motorized lens continuously and estimating the amount of blur for each focus value; the best focused image is determined. A model linking focus values and distance is then used to approximate distance. The decision model makes use of the law of thin lenses and computes range based on the focal length of the camera and the image plane distance from center of the lens. However, this method has its limitation on the fact that blur estimation influences the focal length computation and the derived range. The system required for the imaging process best suits microscopy applications but not mobile robots.

While shape from stereo, shape from motion, shape from focus/defocus infer 3D geometry from two or more images, there exist methods for shape recovery from a single image. Shape from shading, for example, uses the patterns of light and shading for establishing a fundamental equation from a single image relating the image intensity and surface slope. The fundamental equation, the idea of the reflectance map and a Lambertian assumption about the surface helps approximate underlying shape by solving a set of differential equations (Trucco and Verri, 1999). In the real world on mobile robots, the physics and the mathematics required to solve for structure gets complicated. Also, shape from shading will not meet the accuracy requirements for our applications.

Another method that tries to infer shape from a single image uses the distortion in texture created by the imaging process, when a 3D point in space is projected into a 2D plane. The method selects a representation scheme adequate for the texture cues in the image, computes the chosen distortion parameters in the representation scheme and combines the distortion with texture gradients to estimate local orientation of the surface at each pixel. Recently, interesting methods documented in (Saxena et al., 2008), (Hoeim et al., 2005) and (Criminisi et al., 2000) propose the estimation of 3D structure from a single image. Criminisi et al. (2000) recover 3D structure by computing the vanishing point and vanishing line using line segments in the 2D image. Hoeim et al. (2005) use spatial features to define superpixels and classify pixels in the image into different 3D planes. Saxena et al. (2008) implement a machine learning approach to estimating 3D structure by supervised learning of monocular depth cues using ground truth range data. The passive variants for shape recovery listed under the active sensing category in Figure 2.1 improve upon the passive methods discussed thus far by introducing an additional source of illumination.

For example, the structured lighting approach projects a pre-designed pattern of pixels usually in the form of grids and bars and observes the deformation of the pattern on the surface of the object to learn about the 3D shape.

Photometric methods use additional hardware in the form of an optical receiver that includes a photo sensor configured to detect spatio-temporal modulated optical signals directed at the scene from a set of spatially dispersed optical transmitters. The receiver also converts the optical signals from each of the optical transmitters to a corresponding electronic signal that is further analyzed to determine geometric properties of the scene borrowing principles of interferometry. Active depth estimation using holography is another idea that uses a special interference pattern created in a photosensitive medium like photographic film. Depth is inferred from the combined beams of the interference pattern projected and reflected off the surface of interest. Spatial interferometry based sensors provide high accuracy for applications requiring short range (in the order of a few meters) but can have issues with dynamically changing scenes or when the scene is imaged using a mobile platform.

We are left with two methods from the classification chart namely the active triangulation and time-of-flight systems. Both these systems are laser-based. With the active triangulation scheme, a laser in the visible spectrum (usually a line laser) illuminates the scene. The laser line profiles a line of the surface in the scene that is imaged using a high speed camera. By using a special calibration procedure to estimate depth, the surface profiles can be accumulated into a metric 3D structure by moving the camera and laser arrangement over the scene of interest. This approach can be configured to a high degree of accuracy and readily lends to our application, where the scene is static and the mobile robot with the sensor can map the static scene real-time. Again, being camera-based, such a system will have the same field-of-view restrictions as the passive methods. On the other hand, the time-of-flight systems are based on physical principles of estimating distance from a scene by shooting out a laser and sensing the reflection. With the knowledge of the speed of the laser, the observed time taken for the laser to travel, reflect and return is then used to compute the distance from the laser source. The time-of-flight approach does not provide high accuracy as the laser triangulation methods but usually spans a larger field of view.

With each acquisition method having its own advantages and disadvantages, we had to make several considerations in building a 3D imaging system for our applications of interest. We identified that the two most significant factors that determine the choice of the sensor for robotic data collection are the field of view and accuracy. For the under-vehicle inspection application, we realized that the field of view is limited by the ground clearance and the large variation in the size of the components that make up the scene. Hence, we require that our system accommodate the variance in the ground clearance of a variety of cars and automobiles from different manufacturers (typically varying from a minimum of 10 centimeters in compact cars to 90 centimeters in large trucks). We also require that the 3D data provide us high fidelity shape information from the scene for us to be able to perform automated shape analysis. Considering these factors, we narrow down our choice to the time-of-flight and laser-triangulation based scanners over other methods for under-vehicle inspection.

Also, with the limited field of view with a single camera, multiple passes under the vehicle are required to reconstruct the entire undercarriage in 3D. Most of the passive methods discussed thus far, can be extended for better accuracy using an active source as an additional component with the camera. However, the only problem is that the passive variants of shape from images, may not serve well as a real-time acquisition approach for our application. For example, the depth estimation using structured light requires a stationary scene that a camera will have to image using different patterns of coded light. Imaging the same stationary scene with different patterns of structured illumination takes about 20 seconds for a relatively small area of interest (less than 1 square meter). Given the low clearance and limited field of view, in our application, mapping the entire undercarriage would become a tedious time-consuming task.

With the large-scale terrain mapping applications where centimeter accuracy is sufficient, time-of-flight-based scanners appear to be the ideal choice. Vision-based systems, especially those using shape from motion algorithms, sound very promising for centimeter level accuracy also. But, for crack inspection and detection, millimeter accuracy is desired and only laser triangulation systems are able to digitize high accuracy high fidelity 3D geometry at the rate of a few thousand profiles in one second. But, we already know that triangulation systems have a limited field of view. In such situations when large areas need to be digitized, our recommendation is to use an array of laser triangulation sensors. On the one hand, laser scanners are very expensive costing a few thousand dollars at the least, vision-based 3D sensing using shape from stereo or shape from motion only costs as much as the camera, priced within a few hundred dollars. However, our survey study reveals that the accuracy and precision of active sensing methods is much more reliable than image-based passive 3D sensing methods.

2.2 Robotic 3D mapping

Robotic mapping has been an active research area in the last two decades bridging the mechanical mobility and artificial intelligence concepts in real world applications. The success with mobile robots towards reducing human efforts in performing hard tasks and also avoiding humans working in hazardous conditions has only encouraged ambitious ventures over the years. Some applications of that kind are mine mapping (Morris et al., 2005), industrial inspection using snake robots (Granosik et al., 2005), indoor mapping (Lu and Milios, 1997) and cleanup in nuclear facilities (Grinstead et al., 2004). As noted by the authors in each of these papers, the robotic systems still suffer from certain insufficiencies. The issues caused by such shortcomings in acquiring the spatial model of a robot's environment are not only from the mechanical limitations of the robot, but also from sensors that perceive the outside world. Sensors used in mobile robots including vision cameras, range finders, sonar, lasers and infra-red sensors and even GPS and inertial navigation systems are prone to errors. The data from these sensors and the motion commands issued to the robot in navigating the environment are critical to the integrated map. With neither robot motion nor the sensors perfect; uncertainty is introduced in the pose (location and orientation) of a robot. The error in localization propagates as error in the integrated map.

Most of the methods in existing robots for mapping real environments always address the self-localization problem in conjunction with the mapping problem. Localization is the procedure in estimating the position and orientation of the robot. The reason for this conjunction is that robotic mapping and localization are like the chicken and the egg. Focusing and believing in robot motion alone would introduce systematic noise in the integrated map and believing in the sensory data alone without dealing with measurement uncertainty will also begin to guide the robot towards executing motion commands drifting from the desired goal. Looking at the same problem, from a different perspective, if we know the robot's pose or path accurately, map building is a simple and trivial task. On the contrary, if we have an accurate map of the environment, then we can construct robust and elegant solutions to determine robot's pose. But, the problem that we are trying to solve is a tough combination of simultaneously estimating the pose and the map. Also, with 3D mapping, the accuracy of the final integrated 3D model depends heavily on the relative motion between the sensors on the mobile platform and the scene.

We will now very briefly trace through the history of robotic mapping beginning in the late 1980's and through the early 1990's to the mobile robots of today. The late 1980's focused more on the construction and design of mechanical systems for mobile platforms capable of carrying a suite of sensors. Then, the mapping methodologies like occupancy grid mapping (Elfes, 1987), topological map building (Kuipers and Byun, 1991) and metric map building were derived. These algorithms assumed that the robot's path is known with sufficient confidence and then balanced between world-centric versus a robot-centric metric representation of the map. Maps in these cases were usually in two dimensions (2D) considering motion only in 3 degrees of freedom for robot motion. After the success in building 2D maps, robot mapping

literature has been flooded with probabilistic methods that can deal with uncertainty in sensor measurements and robot motion relaxing confidence limits on robot path determination. Seminal papers from Smith, Self and Cheeseman (Smith et. al, 1990) introduced a powerful statistical framework for simultaneously solving the mapping problem and the induced problem of localizing the robot relative to its growing map. Since then, robotic mapping has commonly been referred to as SLAM, short for simultaneous localization and mapping. A robot that implements SLAM possesses sensors that detect spatial landmarks and algorithms that track the spatial landmarks to determine the position of the robot relative to the environment simultaneously constructing the map also. We note that our applications of interest require SLAM-based solutions with no apriori information about the environment.

Recently, with the commercial availability of laser range sensors and relatively cheap vision sensors, 3D mapping has become feasible (Cole and Newman, 2006). Furthermore, we have seen more efforts towards mapping large-scale urban environments (Zhao et. al, 2005; Zhao and Shibasaki, 2001b). The mapping motivation has led to the development of improved image-based methods for autonomous navigation of ground vehicles. In the success of the DARPA Grand Challenge Competition (DARPA Challenge, 2005), several cars such as the Volkswagen and Hummer were modified to operate autonomously on a 132 mile course in the Mojave Desert. The development of the system involved advanced control mechanisms along with 3D laser and vision sensors providing spatial feedback in following a sequence of GPS points of a pre-defined path, avoiding obstacles and yet managing an average speed of 18 mph without any manual intervention from start to finish. Such intelligence was demonstrated by integrating data from an array of range and vision sensors using probabilistic methods (Thrun et al., 2006) of reasoning. However, the same autonomous mobility was presented as a bigger challenge in urban environments (DARPA Challenge, 2007) with the possibility of GPS failing and the need for better mechanisms for robot self localization in such situations. This necessitates the mobile robot to have sensors and methods that can be used for simultaneous structure and state estimation. The systems that we will use to support this dissertation work possess sensors with such capability that we present in detail in Chapter 3. We will now explain different state and structure estimation methods to operate on data from such sensors. From here on, we will use the terms state estimation synonymous with localization and structure estimation synonymous with mapping.

2.2.1 Instrumentation-based localization and mapping

Typically, instrumented approaches for localization and 3D mapping in large-scale environments rely on GPS data for position information, and INS data for determining the orientation (Bretz, 2000). Fusion of the two datasets is usually performed using Kalman filters assuming that sensors are characterized apriori. Kalman filtering takes advantage of the similar characteristics of INS and GPS data to provide an integrated ego-motion estimate, with performance superior to that of either individual system. Kalman filter in this case is used to track the drifting parameters of the system to provide accurate estimates of the system's position, orientation, and velocity and compensates based on the apriori noise model (Grewal et al., 2001).

Several researchers have used instrumented approaches to define the scanner's pose over the years. One of the early efforts is the airborne scanning applications that use GPS/INS packages exclusively to estimate the pose of the cameras during the scanning process (Bossler & Schmidley, 1997; Frere, et al., 1998; Moons, et al., 1998). These efforts were improved with further processing to correct for inter-scanline registration errors in (Fricker et al., 1999; Crombaghs et al., 2000). Nygård, et al. have implemented a different fusion scheme for integrating GPS, INS, and digital compass information for their Unmanned Aerial Vehicle (UAV) project (Nygård et al., 2004). The air-borne instrumented approach for localization has been extended to ground-based vehicles also. One such extension is Cui and Ge (2003) who propose the use of Extended Kalman Filter (EKF) in combination with linear path approximation to estimate the ego-motion of a vehicle in a cluttered urban environment. The method described by Julier and Durrant-

Whyte (2003) is another extension that integrates a vehicle model into their EKF-based pose estimator, to show improved performance in areas where GPS signal quality was low. Masson, et al. (2003) also integrate GPS, INS, bearing and laser range data using an EKF to provide accurate positioning information, even in the cases where one or more of the sensors is not reliable. Direct pose estimation has also been used in ground-level urban scanning systems such as those of Zhao, Shibasaki, and Manandhar (Zhao & Shibasaki, 2001a; Manandhar & Shibasaki, 2000) for improved accuracy with 3D maps even though they have special schemes for pose recovery implemented to support their mapping scheme.

This idea to measure global pose using specific hardware instruments is known as “direct” pose estimation. With the structure sampled as profiles by the laser range scanners, the alignment based on the localization data from the direct state measurement instruments simplifies the mapping problem. However, the 3D map becomes too dependent on the hardware which may be influenced by several environmental factors during operation. We will discuss challenges with this approach later in Section 2.4. We classify all these efforts into instrumentation-dependent localization in spite of these methods using specific fusion algorithms because the fusion scheme is still on data from sensors that directly measure the required “state” of the mobile platform and not infer state from passive or indirect means that we explain in the next few paragraphs.

In 1991, Leonard and Durrant-Whyte introduced the idea of using a 2D range sensor to localize a robot and at the same time avoid obstacles and build a floor plan of indoor environments (Leonard and Durrant-Whyte, 1991). His approach is now very common in robotic systems, where repeated horizontal range scans of an environment are matched together to estimate the motion of the sensing platform between successive scans. The salient features in the range scan, are also archived as the robot explores the environment and later used to build a comprehensive “map” of the environment to aid in matching further scans. This process of matching successive scans to the online model can be seen in Figure 2.2. Matching the newest scan line to the model involves finding the transformation that aligns the new data with the model. The rotation and translation represented by this transformation is the motion undergone by the scanning package.

Many extensions have been made of the original SLAM theory based on laser scans to different environments and purposes. Simon, et al. developed an extension of the SLAM algorithm to a single line range scanner (Simon et al., 1994), where the individual scans are matched to the developing model using the Iterative Closest Point algorithm (Besl & McKay, 1992). The ICP algorithm was first introduced by Besl and MacKay (1992). Its basic version aligns a set $S = \{s_1, s_2, \dots, s_N\}$ of 3D scene points with a geometric model $M = \{m_1, m_2, \dots, m_N\}$, by minimizing the sum of the squared distances between the scene points and the model. For every point s_i in S the distance to M is defined as: $d(s_i, M) = \min \|s_i - m_i\|$. In our case M refers to the points in the map.

The algorithm is best summarized as the following procedure:

Step 1: Start with an initial transformation (R_0, t_0) .

Step 2: For $k = 1, \dots, k_{\max}$ or until stopping criteria met do:

2.1. Compute $s_i^{k-1} = R_{k-1}s_i + t_{k-1}$.

2.2. Build the correspondence set $C^{k-1} = \bigcup_{s_i \in S} \{(s_i^{k-1}, \arg \min_{m \in M} \|s_i^{k-1} - m\|)\}$.

2.3. Using the pairs C compute transformation (R, t) minimizing the sum of squared distances (Horn, 1987).

Step 3: Use the optimized transformation (R, t) to estimate motion in the global map co-ordinate system.

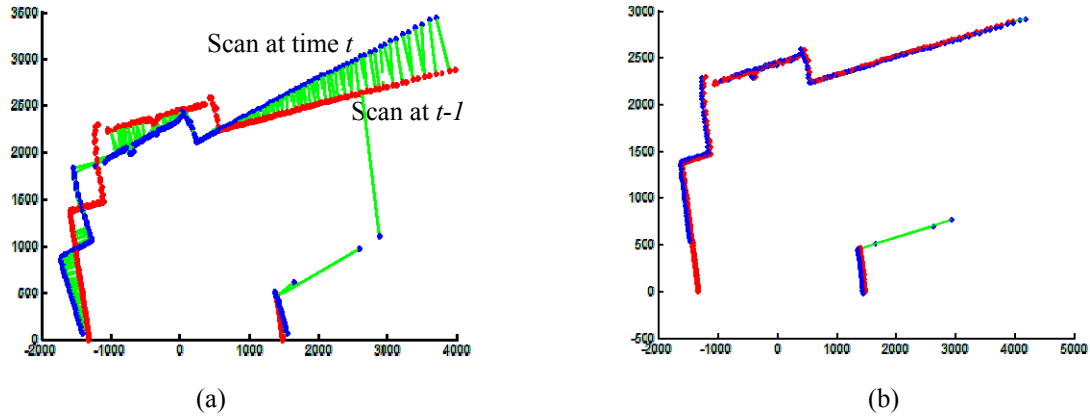


Figure 2.2: Scan line (range profile) matching to estimate the motion of the robot. (a) Successive profiles from a robot navigating indoors. (b) Result of aligning scans indirectly estimates the motion of the robotic platform.

The urban scanning projects of Zhao and Früh (Früh & Zakhor, 2001a; 2001b; Zhao & Shibasaki, 2001b; Früh & Zakhor, 2002) also use orthogonally-mounted laser range scanners and the SLAM methodology to simultaneously capture urban geometry and perform pose estimation. Some recent developments in scan line matching is documented in the paper (Martinez et al., 2006) that proposes speed up of pose recovery using genetic algorithm instead of the iterative search using gradient descent initially proposed by Durrant-Whyte (1991). Früh and Zakhor (2002) have used their vertical laser scanner and the assumption that the buildings being scanned should be orthogonal to the plane of motion to some degree of success. The laser-based SLAM formulation has even been extended to any form of mobile mapping and self-localization, such as using full 3D corridor scans to perform the SLAM duties (Surmann et al., 2003). Hähnel, et al. use 2 different configurations for their laser SLAM approach (Hähnel, et al., 2003). The first is the orthogonal system similar to those above, and the second is a single scanner mounted on a pan-tilt unit. Histogram analysis and planar approximations help their system efficiently store and process the 3D map. Unfortunately, due to the fact that this scan line matching process is restricted to the plane of the laser scanner, this method at best, can only provide positioning within a single plane, and orientation only about the plane's normal, restricting its use to a planar environment such as hallways indoors, or a relatively flat outdoors environment such as a parking lot.

2.2.2 Image-based localization and mapping

Indirect methods of pose estimation include the scan line matching methods from laser range data and pose estimation from video methods. In the next few paragraphs, we will explain the pose from video approach that we will use for our work and then provide extensions such as the auto-calibration approach, and the factorization approach for structure estimation. Our inspiration to study video-based pose estimation for mobile robots in a variety of indoor and outdoor environments is after observing the implementation of visual servoing in papers (Shapiro et al., 1994; Johnson and Matthies, 2000; Pollyfeys et al., 2002; and Armangue et al., 2003).

We will explain a two-frame method similar to that used by Johnson and Matthies (2000) for simplicity sake to obtain the 6 degree of freedom (DOF) position and orientation estimates. The idea on two frames extends to multiple frames also. We note that our pose from video algorithm incorporates an absolute

distance measurement from the laser scanner present on the mobile scanning system to recover scale. The block diagram in estimating ego-motion from video is illustrated in Figure 2.3. For each image pair in the sequence, discrete features are detected in the images, sifted to find the corresponding matches between the successive image frames, and then used to determine the motion estimate of the platform between the views. We will use the calibration approach where we estimate the intrinsic parameter of the camera used in the system using Zhang’s method (Zhang, 2000). We like the calibrated approach because of the dependable accuracy with 3D structure and the computations we are able to save in real time localization compared to the uncalibrated approaches such as (Pollefeys, 1999).

After offline calibration, the online motion and structure estimation from images begins with feature detection. There are a number of feature detectors available for this task, and the standard Harris corner detector (Harris and Stephens, 1988) appears to be the most common in the literature for its robustness to noise, stability and performance (Schmid et al., 1998; Roberts, 1994). In Figure 2.4, we show the Harris corners as red and green markers on two successive frames on one of our experimental datasets in the downtown area of Knoxville, Tennessee. The Harris features are used as the starting locations for a window-based intensity correlation matching. This matching process is typically an $O(N^2)$ operation, but it can be improved by reducing the search space. We do this by restricting the range of search to those features in the second image that lie within R pixels of the same feature in the first image. This radius is determined based on the assumed range of velocities of the mobile platform, as compared to the acquisition rate of the camera. The resulting correspondences are then filtered using an algorithm based on common behavior of the correct matches, similar to that developed by Adam, et al. (2001).

The bottom insets in Figure 2.4 show the image frame with the motion tracks of observed features superimposed. Notice that while the majority of the feature tracks indicate motion along the same direction, some of them exhibit anomalous behavior. Such anomalous motion tracks are called outliers. Experience has taught us that outliers surface mainly because of noise – either through the estimated motion of features in the scene or by false matches from the correlation stage. Removal of these noisy feature tracks increases the video localization system’s robustness to noise, and provides a more accurate estimate of the platform’s pose.

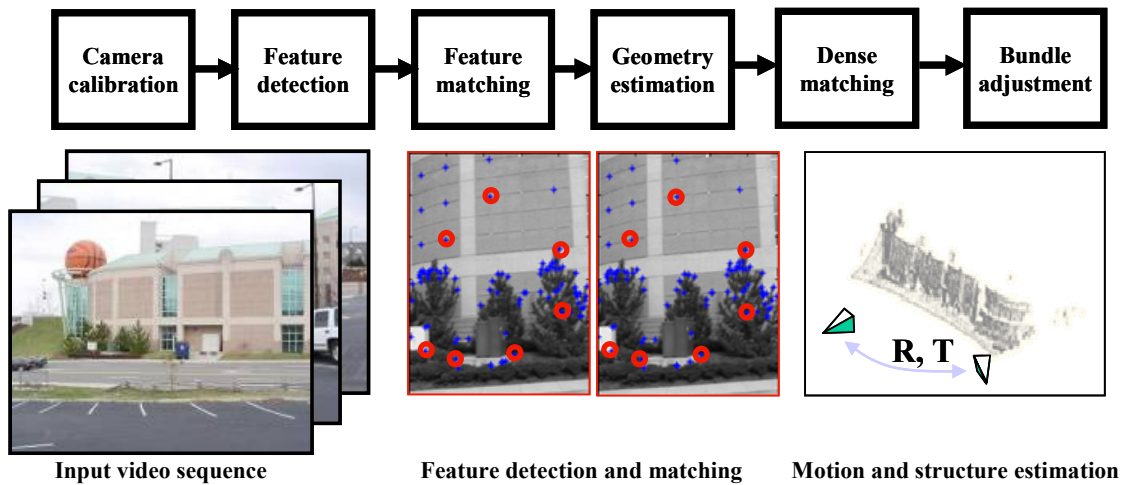


Figure 2.3: Block diagram for image-based motion and structure estimation following (Pollefeys et al., 1999).

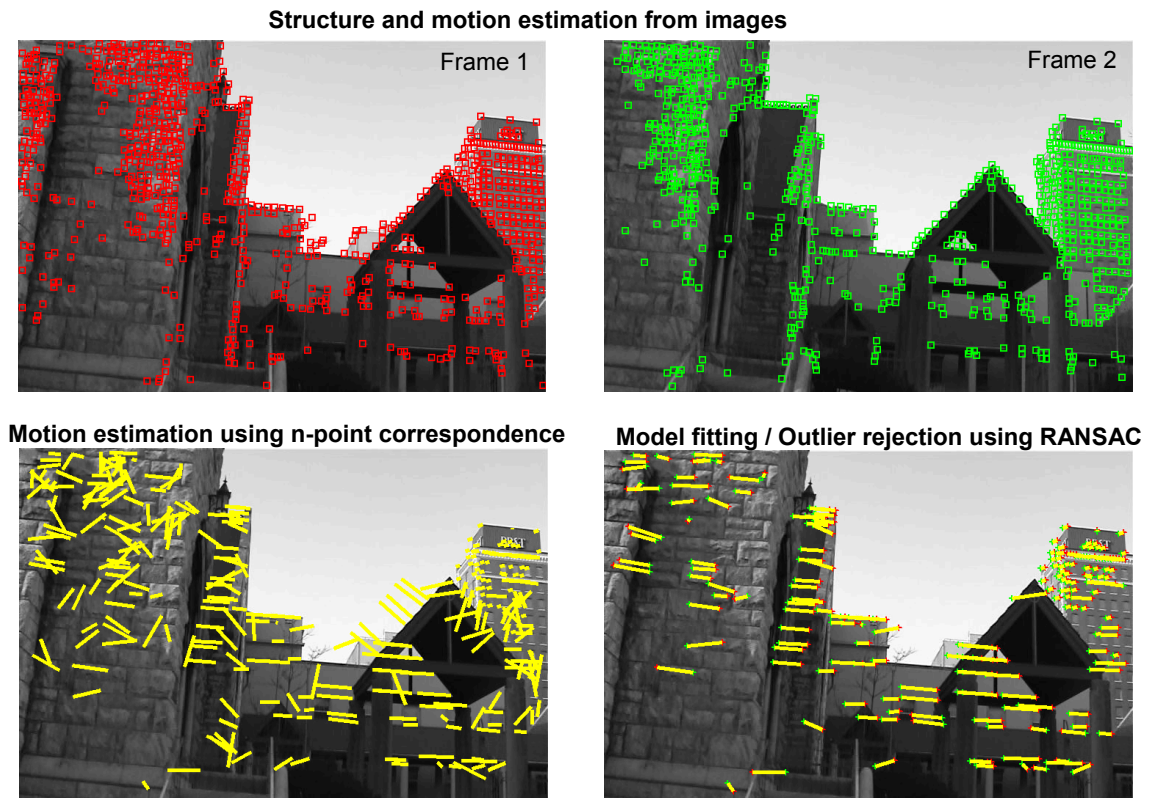


Figure 2.4: Pictorial description of structure and motion estimation algorithm on video frames. The top images represent two successive frames collected while experimenting in the downtown area of Knoxville, Tennessee. The bottom left image shows the motion vectors estimated from the image data and the bottom right image shows the result of outlier rejection. The inliers of the motion matches are then used to compute the 3D motion of the camera that generated these images.

The most common procedure to remove such outliers and make pose estimation robust to noise is through a random sample consensus algorithm (RANSAC). The RANSAC approach is a probabilistic solution, introduced by Fischler and Bolles (1981). A small subset of feature correspondences are randomly selected from the set of all feature tracks estimated after correlation matching. This subset of correspondences defines a fundamental matrix F for the image pair. F is a rank 2, 3 by 3 matrix. For simple projection models, the minimum cardinality of the feature track subset is three. But based on the motion assumptions and the projection model, there are several options such as the 5-point (Nister, 2004b), 6-point (Hartley and Zisserman, 2000) and 8-point (Hartley, 1997) matching algorithms that will recover the matrix F as a linear system relating correspondences between successive image frames. Next, the epipolar error e_i is computed for all feature tracks, measuring the distance of each feature from its corresponding epipolar line, defined by the computed F_i . Since, the subset of correspondences can contain errors, we will have to evaluate a sufficiently large number of such feature subsets. Each evaluation will provide a hypothesis about the state of the camera system and the structure. The RANSAC procedure iterates through all these hypotheses to choose a set of feature tracks that have maximal support. The cost function in RANSAC is usually the mean epipolar error. If the mean epipolar error for a subset is less than that from previous iterations, the current fundamental matrix and its associated mean epipolar error become the best estimate for this two-frame motion. The process is then iterated until convergence within a threshold. We illustrate this procedure in Figure 2.5 in a simplified line fitting example.

The green markers in Figure 2.5 represent the inliers and the red the outliers. The problem of estimating F then in this simplified line-fitting example is to randomly select two points from the data of matches and then seek the support from other matches iteratively. The linear model F that relates the image features in successive frames is evaluated based on the distance between each feature track and the linear motion model. Through the iterative procedure, several hypotheses are evaluated and the one that converges to maximal support is chosen. The number of minimal subsets M_h (Equation 2.1) to evaluate depends on the feature detector. In Equation 2.1, p refers to the probability that a pixel in the image is a feature, ε is the error associated with the location of features detected by the feature detector and s is the choice of the n -point matching algorithm. If using the 5-point algorithm, $s = 5$. When the threshold for support search is set appropriately in RANSAC, the algorithm has been proved to be robust in rejecting outliers as shown in Figure 2.5b.

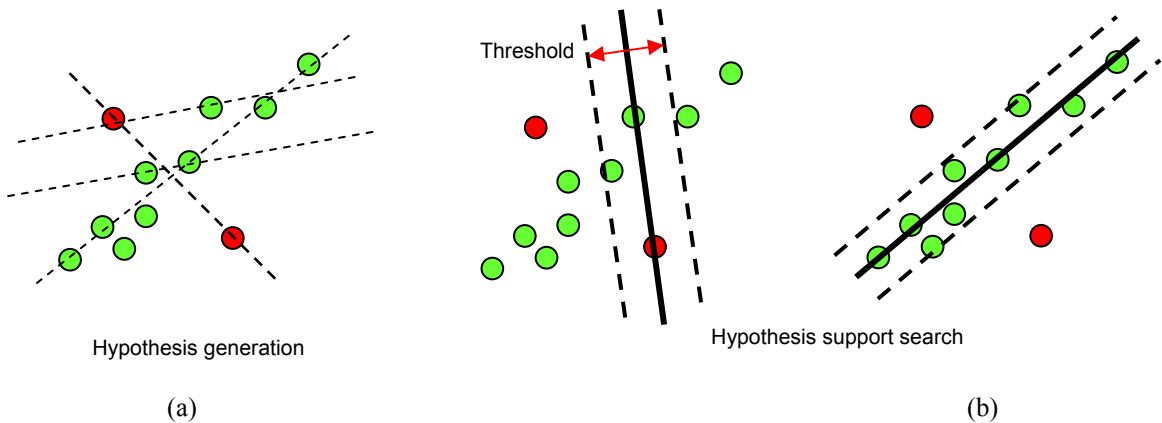


Figure 2.5: Inlier classification using RANSAC. This n -point matching algorithm generates different hypotheses by randomly sampling the motion matches and fits a model based on the minimal subset. Competing hypotheses are iteratively scored based on a threshold to choose the one with maximal support.

$$M_h = \log(1 - p) / \log(1 - (1 - \varepsilon)^s). \quad (2.1)$$

When the iterative procedure is complete, the best estimate of F is computed by removing all features that were considered outliers with epipolar error greater than 2 pixels and re-computing F using all the inlying feature matches. Then, the camera's pre-computed calibration matrix K is used to calculate the *essential matrix* E via

$$E = K^T F K. \quad (2.2)$$

The next stage of motion calculation is to extract the translation and rotation parameters from E . It can be shown that the translation vector T_s is the solution to $\min \|E^T T_s\|$ - the unit eigenvector with the smallest Eigen value of the matrix EE^T . The sign of the translation vector can be determined using the constraint that the imaged scene must lie in front of the camera.

Determining the solution to the rotation matrix R involves solving

$$\min \|(R^T [-T_s]_x - E^T)\|, \quad (2.3)$$

which can be efficiently solved using the quaternion form. The output of this motion estimation system is a 5 DOF motion state, with an unknown scale factor γ . We then use an absolute distance measurement from the onboard laser range scanner to provide the scale factor γ .

So far, we have only described state estimation from images. We now describe the structure estimation process. Though several methods exist for 3D reconstruction from images (Ma et al., 2003), the fast factorization approach for projective reconstruction appears to be the most suited for our application. Please note that the structure estimation algorithm that we use is not a two-frame method but multi-frame method to counter the effect of vibrations in the robotic platform. The geometry estimation was explained based on the two frame method in earlier paragraphs for the sake of simplicity and is easily extendable to multiple frames.

Let us now consider recovering the projective structure from matched features in a video frame. Suppose the j^{th} point in the i^{th} frame, \mathbf{x}_{ij} is projected from the scene point X_j by $\lambda_{ij}\mathbf{x}_{ij} = P_i X_j$, where λ_{ij} and P_i denote the projective depths and projection matrices, respectively. Given N_p matched points in N_f frames we have:

$$\begin{bmatrix} \lambda_{11}\mathbf{x}_{11} & \lambda_{12}\mathbf{x}_{12} & \cdots & \lambda_{1N_p}\mathbf{x}_{1N_p} \\ \lambda_{21}\mathbf{x}_{21} & \lambda_{22}\mathbf{x}_{22} & \cdots & \lambda_{2N_p}\mathbf{x}_{2N_p} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{N_f1}\mathbf{x}_{N_f1} & \lambda_{N_f2}\mathbf{x}_{N_f2} & \cdots & \lambda_{N_fN_p}\mathbf{x}_{N_fN_p} \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_{N_f} \end{bmatrix} \begin{bmatrix} X_1 & X_2 & \cdots & X_{N_p} \end{bmatrix}, \quad (2.4)$$

where the matrix on the left hand side is the measurement matrix.

The initial depths can be set to unity or obtained using Sturm and Triggs' method (Strum and Triggs, 1996). After the depths are normalized, we find the nearest rank-4 approximation of the measurement matrix using singular value decomposition (SVD), based on which the camera matrices and 3D reconstructed points are derived. These reconstructed points are re-projected into each image to obtain new estimates of the depths. The process is repeated until the variations in the projective depths are negligible.

The step-by-step iterative projective reconstruction algorithm can be summarized as shown below.

Step 1: Normalize the image data into homogenous co-ordinates using isotropic scaling.

Step 2: Start with an initial estimate of the projective depths.

Step 3: Normalize the depths

$$(3.1) \text{ Rescale each column of the measurement matrix so that } \sum_{i=1}^{N_f} \lambda_{ij}^2 \mathbf{x}_{ij}^T \mathbf{x}_{ij} = 1.$$

$$(3.2) \text{ Rescale each triplet of rows of the measurement matrix so that } \sum_{j=1}^{N_p} \lambda_{ij}^2 \mathbf{x}_{ij}^T \mathbf{x}_{ij} = 1.$$

$$(3.3) \text{ Repeat until there is no significant change in the measurement matrix.}$$

Step 4: Form the measurement matrix, find its nearest rank-4 approximation using SVD and find the camera matrices and 3D points.

Step 5: Re-project the points into each image to obtain new estimates of the depths.

Step 6: If the variations in the projective depths are small enough, stop. Otherwise repeat Steps (3)-(6). The initial depths can be set to ones or obtained using Sturm and Triggs' method or a method from (Mahmud et al., 2001).

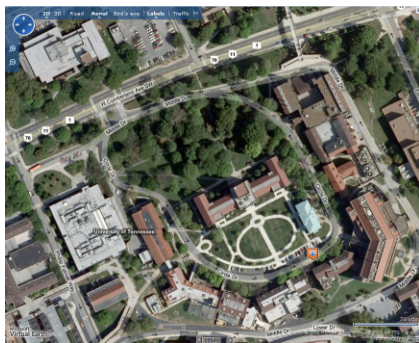
The output of the procedure is a projective reconstruction that we need to transform into a metric reconstruction. Towards that goal, we need to find a projective transformation matrix H and update the projective reconstruction by HX_j . Using the dual absolute quadric Ω^* we have $\omega_i^* \sim P_i \Omega_i^* P_i^T$ where $\omega_i^* = K_i K_i^T$ with K_i as the camera's intrinsic matrices (Strum and Triggs, 1996). A linear solution of Ω^* can be obtained by imposing additional constraints on the camera's intrinsic parameters, such as zero skew, unit aspect ratio, and zero principal point, and the rank-3 property is applied for improved accuracy. Please note that we already have the K matrix from apriori calibration. The projective transformation matrix is then obtained by forcing $H \Omega_i^* H^T = \text{diag}(1, 1, 1, 0)$ and projective reconstruction is elevated to metric reconstruction by $P_{E,i} = P_i H^{-1}$ and $X_{E,j} = HX_j$. Finally, bundle adjustment is carried out to minimize the projection errors over several frames by computing the $\min \sum_{i,j} \|\mathbf{x}_{ij} - P_{E,i} X_{E,j}\|^2$. Once the sparse points are reconstructed, dense matching (Strecha et al., 2003) is carried out to transform each pixel in the 2D image into a 3D point. Based on the sampling density requirement on the final 3D model, interpolation is also carried out.

Our experience with implementing image-based motion estimation encourages the implementation of vision-based navigation in structured environments with buildings where GPS sensors can fail. However, we also realize that motion estimation from video can be perturbed by several factors also. Illumination change, wind, weather, type of motion (as in rotation only or translation only), moving objects in the scene, multiple layers of objects in the camera's field of view can affect pose recovery. Further, we also note that there is error that propagates at each stage illustrated in the Figure 2.3. At the feature detection stage, there is error about the pixels that are mistakenly classified as features, and in environments with lot of vegetation, several features in the image could appear as potential matches in the correlation matching phase creating confusion with several hypothesis with support in the RANSAC stage. The structure estimation is completely dependent on the initial sparse reconstruction and the estimation of the fundamental matrix, and our observation is that if at all we can improve the accuracy in localization and the map, it has to be performed via uncertainty analysis at the geometric estimation stage.

2.2.3 Evaluation and summary

In two broad categories as instrumentation-based and image-based localization and mapping, we presented a variety of ideas from different researchers in the previous sections. We were inspired by some of these ideas and decided to evaluate them for our applications of interest. The goal behind such an evaluation was to gain experience on the 3D sensing methods and develop the expertise to design robust, efficient and cost effective 3D imaging systems given the requirements for a specific application. We document our experience in the following paragraphs in this section. We show the test area that we chose for our rigorous evaluation in Figure 2.6. The maps show the Ayres Hall building at the University of Tennessee, Knoxville.

The techniques that piqued our interest with promise for robotic 3D imaging were (1) to use off-the-shelf laser scanners on a stop-scan-move strategy (2) use the scanners in a continuous profiling mode as the robot traverses an intended path (3) recover 3D structure from a single image for localization purposes and (4) recovering state and structure from continuous video. We show the results from each one of these techniques in Figure 2.7. The left insets of each figure 2.7 (a)-(d) shows the relative depth sensed and the right insets are screenshots of the rendered 3D model. The color code on the relative depth is used to enable emphasis on the geometric fidelity of the scene. We are able to see that the 3D scanners provide high accuracy. But, the stop-scan-move faces the problem of view occlusions that become visible in the 3D rendering. Also, the point cloud from the static scan mode is scattered and does not guarantee uniform sampling density suggesting that stop-move-scan mapping idea requires view planning to strategize our locations for optimal stopovers. Robotic mapping using a stop and scan strategy will require some apriori knowledge of the scene for autonomous mapping, which none of our applications can assume. Figure 2.7 (b) is the result after integrating the localization data from GPS and inertial units with the 3D scanners operating in continuous profiling mode. This approach compared to the stop-scan-move has reduced view occlusion issues and promises uniform sampling throughout the scene. We are able to see even the geometry in the window panes in this dataset that was not obvious in the stop-move-scan result. Figure 2.7 (c) and (d) are result of recovering 3D structure from a video sequence following (Pollefeys, 1999) and a single frame following (Saxena et al., 2008) respectively. The visual rendering is more photorealistic compared to the instrumentation approach, but we are able to perceive the geometric inaccuracies in the color-coded images.



(a)



(b)

Figure 2.6: Area of interest to evaluate different techniques for 3D mapping. (a) Aerial view of the Math department at the University of Tennessee, Knoxville. (b) The bird's eye view of the building and the park in front of the building. (Original images courtesy of www.maps.live.com)

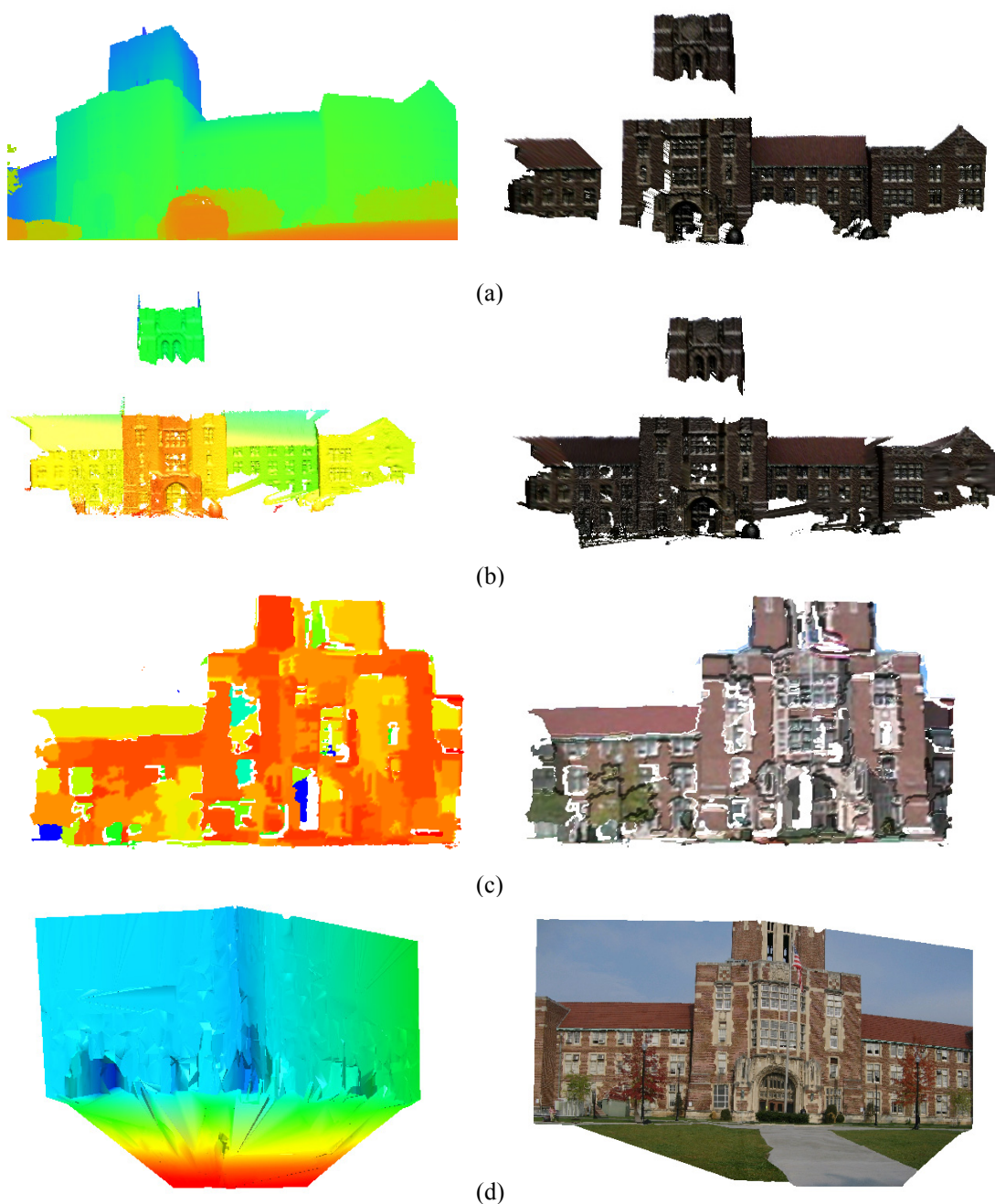


Figure 2.7: Comparison of 3D imaging techniques. (a) Static scanning using a laser scanner. (b) Mobile mapping using a laser scanner and instrumentation-based localization. (c) State and structure recovery from video frames only. (d) 3D structure from a single image. Laser scanners provide better geometric fidelity while image-based methods provide better texture fidelity. The left insets on figures (a)-(d) encode relative depth of the scene in color and the right insets are the screen captures of the rendered 3D model.

We note that the results from the image-based recovery methods are up to a scale and not in global world co-ordinates as the instrumentation results. Image-based methods being several thousand dollars cheaper than the instrumentation approach can not guarantee dense sampling even after interpolation. Also, their accuracy is not always reliable and repeatable. Saxena's method (Saxena et al., 2008) and Hoeim's approach (Hoeim et al., 2005) to recover 3D structure from a single image succeeds only in 35-40% of the real world scenes. Though the shape from motion approach can do better, we still see the geometric discontinuities in Figure 2.7. Though the method is able to differentiate structure on the roof slope, the window panes and the trees in front of the building, we are also able to see that, many pixels are not consistent with the geometry expected from the scene.

As any evaluation, we extended our experiments to several real world environments and concluded that the instrumentation approach to robotic mapping is best suited for application with high geometric accuracy requirements. We also learnt that the error in 3D models integrated using the instrumentation approach relies heavily on the state estimation hardware. Uncertainty in pose propagates as error in the 3D points of the integrated map. Image-based methods on the other hand can localize and map environments without expensive equipment, but error in the geometric optimization can lead to both localization and depth uncertainty distorting the metric structure of the scene. The error in the 3D model in the instrumentation approach can be as good as the pose hardware and the 3D scanner but the noise characteristics for image-based methods are not predictable. With high accuracy 3D scanners and high precision pose estimation hardware we see the potential in being able to design systems to operate at a specified accuracy. It is not that simple with image-based methods. Though we can compute a worst case limit on the accuracy of the 3D model, we observe recovered 3D points do not adhere to a bounded noise model. There is also the problem with the depth discontinuity and occlusions in the image-based approach. The models recovered from images will also be up to a scale factor, the determination of which requires additional information about the scene. Though these are some issues with the methods for large-scale mapping, the bigger challenges arise in situations from the dynamic nature of the environment. In the next section, we enlist such issues and also summarize the literature associated with potential solutions.

2.3 Handling uncertainty in robotic mapping

2.3.1 Sources of uncertainty

In Section 2.2, we had presented several methods used in the localization and mapping. Here, we document some real world challenges that we face while implementing these solutions. We begin with the instrumentation approach and in particular the GPS and the inertial systems that we use with our prototypes. Both these systems though aid global self localization, depend on several environmental factors. GPS systems function based on signal reception from satellites orbiting the earth. To be able to localize based on GPS alone, we require the availability of signals from at least four satellites. Our GPS in the differential mode is a 2 cm accurate system when the system receives signals from at least 6 satellites. Lesser number of satellites implies lesser accuracy. The signal availability is dependent on several factors, for example at any point in the day or night; satellites might not be in line of sight for communication. Sometimes, when navigating in areas with tall buildings, GPS signals can be attenuated. We show one such example in Figure 2.8. When our system receives signals from 6 satellites, the error of the system is as little as 3 centimeter, but the moment we get closer to tall trees in the test area as shown in the aerial image inset, the error unexpectedly blows ten-fold up to 3 m. While navigating in a highway, 3 m of drift can mean a completely different lane and can result in accidents during autonomous operation. This type of intermittent failure creates the need for building systems that can operate well amidst tall buildings and also handle GPS failures.

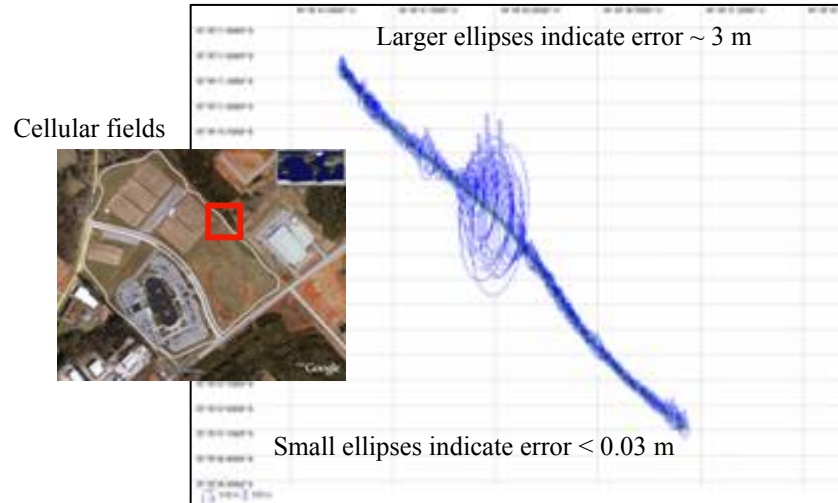


Figure 2.8: GPS localization and the uncertainty about the position estimate in a test dataset in Knoxville, Tennessee. The aerial picture of the ‘U.S Cellular fields’ in Knoxville is a 2.2 mile long stretch we tried mapping using our system.

The immediate hardware replacement to localization using GPS is the inertial instruments. But the inertial systems get perturbed by the vibrations in the robot’s motion and since the position and orientation are actually measured by integrating previous measurements over time, the small errors due to vibration accumulate and grow to large proportions over long periods of operation. Also, the very design of the gyroscopes in the inertial instruments on long usage get heated up and the temperature exponentially affects the accuracy of the measurements. Hence we are forced to look into vision-based systems as a dependable backup.

The vision system that we described in Section 2.2 is a good candidate for localization in urban environments. But they also have environmental factors that influence its accuracy. Illumination change, dynamic moving objects in scene, and lack of traceable features are situations that arise while operating in real world environments. We show one such situation in Figure 2.9. The Harris corner based scheme that we described earlier in Figure 2.4 can lead to a different geometric inference. Image-based localization appears to have issues with vegetation because, green leaves and brown twigs that are the major features in the scene are detected as similar features creating confusion in the correlation based matching stage leading to a high percentage of outliers in feature tracks. We are able to see two major motion directions in the bottom left image, where the outlier tracks representing top-down motion appears to have as much support as the actual translation motion from left to right. Such situations introduce uncertainty in the geometric inference and hence error in the motion and structure estimation. Falling back to the line fitting example again, the type of uncertainty that competing motion hypothesis can create is shown in Figure 2.10. We show the ideal case with structured environments and other possibilities. The degeneracy in model fitting could happen when the motion is strictly rotational without any translation. The vegetation case, the multiple motion layers or the dynamic objects in scene could potentially result in the multiple hypotheses as shown in the figure.

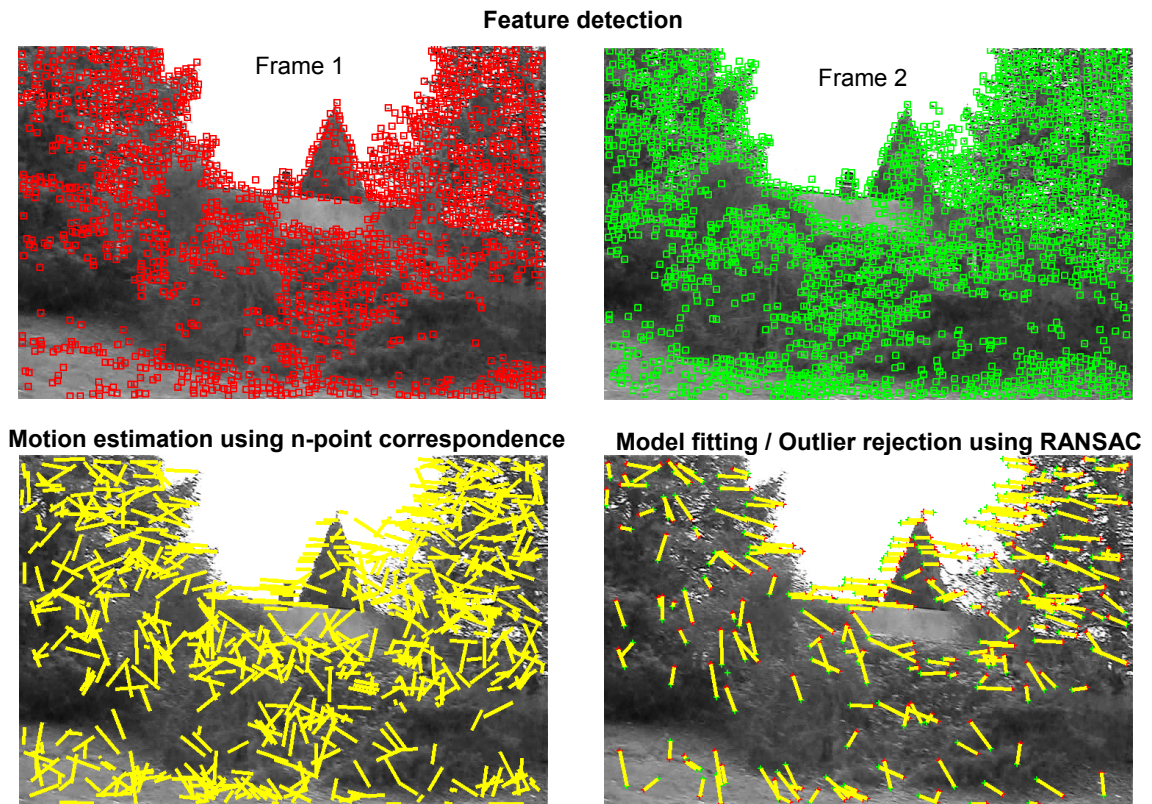


Figure 2.9: Image-based pose recovery in a challenging environment. We saw that image-based pose recovery was very reliable with structure environments in Figure 2.4. In this figure, we observe that the exact same method might not work as efficiently as it did with buildings on vegetation, indicating the need for an improved state estimation procedure to handle such situations.

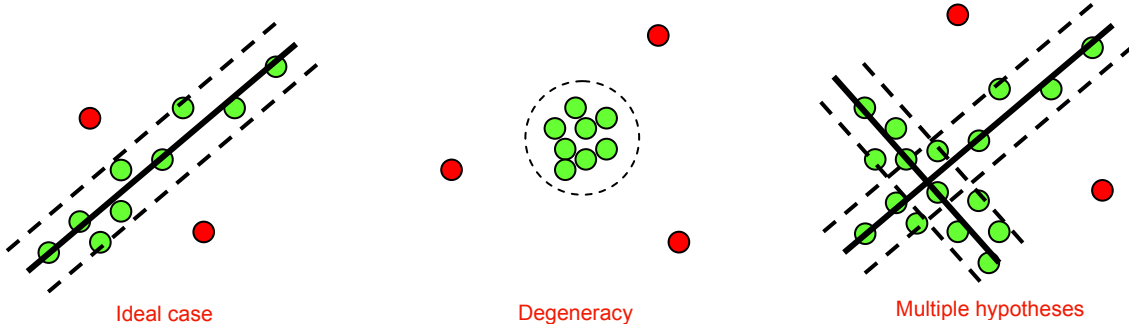


Figure 2.10: The line fitting example to show inlier-outlier distributions that introduce uncertainty into the image-based geometric inference. In the ideal case with structured environments, only the inliers are used to compute the 3D motion, but because of the dynamic nature of the scene, we realize the necessity to handle degeneracy and multiple motion model hypothesis of 3D motion in the model fitting.

We cannot depend on the other indirect pose estimation method from range data completely also because, the range data can utmost recover three degree of freedom motion with sufficient confidence. Also, range data while navigating in large-scale environments at high speeds can become unreliable with no correlation between successive profiles. In some cases, there may be too few features in the environment to sufficiently localize robot pose. Also, in large-scale environments, the scattered non-uniformly sampled 3D profile does not help either. The uncertainty analysis of scan line matching using ICP for localization (Censi, 2007) explores the statistical concern and supports our observation.

2.3.2 Uncertainty minimization methods

So far, we presented the options available for state and structure estimation towards mapping large-scale environments. We realized that sensors and sensing methods are noisy and can be perturbed by the unpredictable real world situations. The uncertainty in the sensor operation and its measurements propagate into the integrated 3D map. Our motivation being able to generate photo-realistic geometrically accurate 3D models of environments for safety and security applications, we have to handle the uncertainty in the sensor measurements and make the mapping mission as stable and independent as possible to the environment of interest, before claiming break-through capability of our 3D mapping systems.

Depending on the application and the environment to be digitized, we have argued that we require data from various types of sensors to correctly model the target scene. We justified the use of GPS and inertial systems as an easy but an expensive solution and suggested that image-based localization and mapping as a suitable inexpensive approach to operate when hardware solutions are drifting and/or are not functional. We also identified a method to recover pose and map structure from range sensors also. The next significant step is to register the data from the various sensors into a common coordinate frame for processing. Data fusion is then necessary to combine the multi-modal data into a 3D model, which is ready for visualization or further processing. A good fusion scheme should work synergistically with all these sensors combining aspects of redundancy, complementary, timeliness and cost for efficiency of operation. We summarize such smart integration methods for uncertainty minimization through multi-sensor localization and mapping into a classification chart as shown in Figure 2.11. The methods that we will discuss in the later paragraphs, decide the fusion of data based on the uncertainty that each sensor is propagating. These methods guide the convergence of different technologies for 3D sensing onto mobile robots.

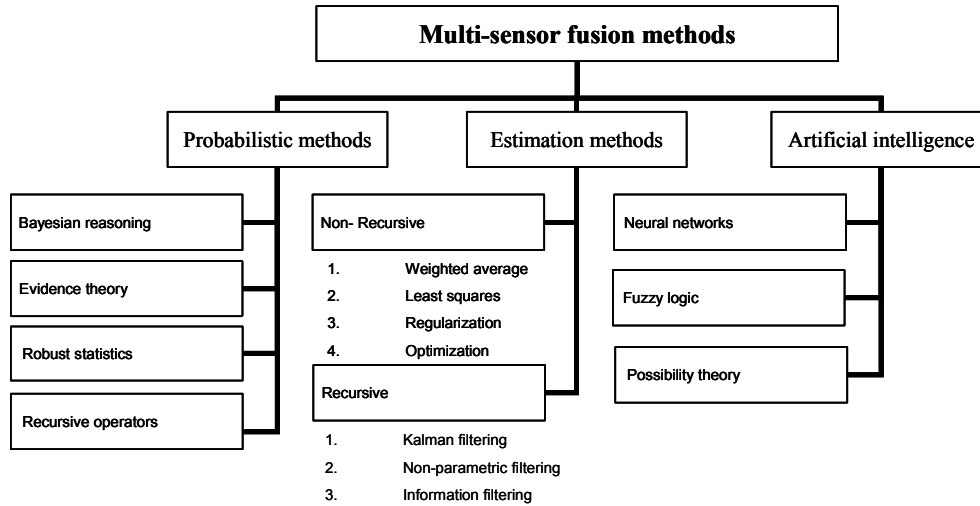


Figure 2.11: Summary of uncertainty minimization methodologies used in robotic mapping and navigation.

We should be beginning this discussion by describing the belief propagation framework for robot state localization. But, we will not discuss belief propagation until the next chapter. For now, we will assume that we have understood and implemented one of several methods such as Kalman filters (Kalman, 1960), extended Kalman filters (Cui and Ge, 2003), particle filters (Fox et al., 1999), sparse information filters (Liu and Thrun, 2003) operating on the state measured from hardware or estimated from range and vision sensors. The function of these algorithms is to provide an estimate of the most probable state as the robot moves and propagating uncertainty about the state using the apriori sensor's noise characteristics and the robots motion model. Such methods are turning out to be ubiquitous on robotic systems to include the sensor measurement uncertainty. These mathematical methods help robotic systems decide when to use the complementary nature of different sensors and when to treat them as redundant information. In our application, with probable states from several sensors like the GPS, INS, vision and range, we now have to decide how to combine these states in such a way that it will minimize the error and the uncertainty about the 3D points in our integrated 3D models. Towards that goal, we discuss different fusion techniques in the following paragraphs.

We will treat uncertainty minimization on multiple estimates of state and motion from several sensors as a special case of the algorithms defined for data fusion (Abidi, 1992) in multi-sensor systems. Based on a thorough study of the literature, we find that most methods are built around a probabilistic framework and broadly fall into one of three categories as reasoning methods that are primarily based on Bayesian or Dempster-Schafer evidence reasoning, estimation methods that include Kalman filters, particle filters and least squares methods and the artificial intelligence-inspired methods that use neural networks and fuzzy logic (Maskell, 2008). All these methods operate on a sensor model that represents the error in the data from each sensor. Usually, this is modeled as a Gaussian distribution and for most electronic instruments; the Gaussian assumption appears to be valid. The variance of the distribution provides the bound on the error for each sensor. Fusion methods operate on this sensor model only. We will survey methods in multi-sensor fusion and describe the general idea behind each of these methods without going into implementation details.

Bayesian estimation is the most common formalism for multi-sensor fusion that helps combine data from multiple sensors based on the rules of probability (Durrant-Whyte, 1987). Within this framework, uncertainty is represented as conditional probabilities and the Bayes theorem is used to propagate belief by computing posterior uncertainty distributions. The redundant information from a group of sensors is then used using the odds and the likelihood ratio formulation of the Bayes rule. Durrant-Whyte describes his method based on Bayesian principles for a multi-sensor robotic application (Durrant-Whyte, 1988). But one of the drawbacks that Garvey (Garvey et al., 1995) improves over Bayesian reasoning in robotics is the inability to include level of detail from each sensor. In the Bayesian formulation, all propositions and hypotheses for which there is no priori available are considered equally likely. When additional information is available from a particular sensor and we are still evaluating different propositions without using the extra information, Bayesian estimates result in an intuitive undesired fusion result. Dempster-Schafer theory overcomes this aspect of Bayesian reasoning by assigning probabilities for ignorance that is updated as and when more data is available. Bogler (1987), Waltz (1995) and Buede (1988) have applied this theory to target identification applications.

Recursive estimation methods for state estimation in robotics using the Kalman filter (Kalman, 1960) and its extensions (Fox et al., 2003) combine sensor data by weighing their posterior covariance uncertainty estimates. We note that the Kalman filter in this approach is used for fusion and not belief propagation. The sparse information filter (Liu and Thrun, 2003) is also a recent trend in robotics but is still facing problems in interpreting initial sensor modeling in the information theoretic sense. One of the issues with both the Bayesian reasoning and the evidential reasoning approach is the Gaussian assumption. Robust statistical methods (Borges & Aldon, 2003) like M-estimators, EM algorithms (McLachlan & Krishnan, 1997) are also used in multi-sensor robotics. They outperform Bayesian methods when some of the outliers that a Gaussian assumption would eliminate are actually new hypothesis and also need to be considered for fusion. Some methods that use this idea are (Thrun et al., 2003; and Burgard et al., 1999).

Uncertainty ellipsoids as proposed by (Nakamura & Zu, 1989) are also an interesting class of methods that aim to infer about reliability of data for fusion by computing the Jacobians of state estimate and computing the high-dimensional overlap as the decision criterion for fusion. The next class of methods is the artificial intelligence-based methods that try to bring in machine learning techniques into robotics. The idea to use neural networks (Capriglione, 2003) to learn about sensors though appears to be logical, faces several real-time implementation challenges. Real-time implementation issues and a generic repeatable working methodology limits methods based on the fuzzy approach for localization also. There are several methods for multi-sensor fusion, of which we have only presented a few suited to our application. Hence, we refer the reader to the following seminal surveys on multi-sensor fusion (Luo et al., 2002; Castellanos & Tardos, 2000; Xiong and Svensson, 2002; Mitchell, 2007). In our case with large-scale mobile robot localization and mapping (Bailey, 2002), our experience indicates that fault detection instead of fusion as the promising direction. We make such a statement because; we are trying to find a solution to the problem of intermittently operating sensors. Very few methods in the fusion literature fall into this category and are called fault detection methods. In the following sub-section, we present the literature on fault detection.

2.3.3 Uncertainty minimization by fault detection

We are able to find key sensor selection methods amidst the fusion papers addressing each one of the three categories. For the probabilistic approach, Isler and Bajcsy (2007) use entropy as the criterion to select a sensor for target localization using visual camera networks. This paper essentially follows the recent trend in using information theoretic formulation for sensor selection as published in (Denzler and Brown, 2002) and (Thrun et al., 2004). Luo and Lin (Luo and Lin, 1988) also previously used probability distribution functions to build a cost matrix whose coefficients are proportional to the area of overlap of the distribution functions in choosing functional and believable sensors. Addressing least squares-based methods Scheduling et al. (Scheduling et al., 1998) utilize the frequency domain form of the Kalman filter to derive a metric for

detecting faults. (Koshizen, 2000a) proposes the expectation maximization algorithm for sensor selection and further include fusion techniques (Koshizen, 2000b) to prove that fusion of selected sensors can perform better than selection of a particular sensor. Kobayashi et al. (Kobayashi, 1999) has also formulated a possibility measure based on fuzzy logic that can automatically learn reliable sensors.

A more recent paper that proposes artificial neural networks for sensor selection is (Capriglione, 2003). Novel methods that do not fall into any of these categories also exist. Some examples are the geometric approach for sensor selection proposed by (Giraud and Jouvencel, 1995) that considers the interaction of the sensor and the environment and the stochastic dynamic programming approach of Hovland and McCarragher (1997). The common aspect with all the aforementioned fault detection methods is their ability to evaluate and characterize sensor performance and establish confidence factors for the individual sensors. They present different methodologies for adaptively determining sensor confidence factors based upon evidential reasoning and statistical clustering which utilize the confidence factors. The advantage that fault detection methods provide over fusion methods is that the sensor confidence measures in fault detection include uncertainties due to the sensor characteristics and performance degradation due to environmental factors.

2.3.3 Gaps to fill in the literature

So far in this chapter, we studied different technologies for robotic 3D mapping. We started with 3D sensing methods and their deployment on mobile robots and enlisted the different sources of uncertainty that influence the accuracy achievable with these sensors. We realized that the error in the integrated map is not only because of measurement errors from the sensor, but also the localization error in the mobile platform and how these two errors combine in real world situations. Further, we discussed sensor fusion methods for combining pose estimates from several complimentary sensors towards better localization and mapping. We understood that current sensor fusion methods assume bounded nature of the sensor measurement uncertainty.

In our example with the GPS, we illustrated that though the normal expected noise variance on a measurement is 3 cm, when non-functional it can be greater than 3 m. The error in a functional sensor is an order different in a situation when it is not functioning at the desired and expected accuracy. Also, adding to the problem, systems like the GPS can acquire more satellites and spring back functional as the state of the robot changes, rejecting the idea of fault isolation and thereby necessitating an inference framework to identify such intermittent failures. In addition, we have multi-path scenarios, where the system is very confident about a wrong state. Also, with the alternate vision-based localization and mapping methods, the performance degradation in unknown and unstructured environments is quite unpredictable.

These situations pose several interesting questions. How should we design a 3D imaging system that can operate autonomously in a wide variety of environments? What sensors and how many to use accounting for the inevitability that sensors can degrade or fail? Does using several sensors guarantee better localization and better map accuracy? How to integrate uncertain, imprecise and conflicting data from several sensors imaging the same scene? How to integrate localization and map data measurements from sensors with different noise models as is the case with GPS, INS, vision and range? How to identify reliability of a sensor online during autonomous operation in unknown unstructured environments? How can we make these multi-sensor agents robust to the operating uncertainties of the individual sensors? At any instant in time, how do we identify believable sensors and isolate failed/failing sensors? How to implement opportunistic intelligence to adaptively decide between believing the best sensor versus fusing conflicting/complementary data from uncertain sensing agents while maintaining desired confidence levels? Is it possible to decide when to fuse sensor data and when to selectively reject sensor data? Will it be possible to identify sensors springing back to normal behavior accommodating for intermittent failures?

Having studied the design of 3D imaging systems, we do realize that contemporary systems are specifically designed to operate in controlled environments. State-of-the-art systems have concentrated more on demonstrating feasibility with little attention to accuracy, precision, realism, capability and robustness to different environments. With global localization being the challenge in mapping, we recognize the need to develop methods that minimize localization uncertainty and also increase the reliability of existing localization methodologies. Furthermore, in handling situations of sensor performance degradation, current fusion methods are not able to decide on sensor functionality online during autonomous operation and adapt accordingly. They are not able to infer reliability of the sensor measurements when sensors operate beyond expected noise models. Though we did study a few methods in Section 2.3 that tried to evaluate sensor reliability based on consensus, the measurement uncertainty and the sensor consensus have always been treated and evaluated as two different entities, raising the question of statistical concern, when to fuse data and when to discard data and avoid catastrophic fusion.

This dissertation attempts to fill these gaps through the design and development of a robust system for 3D imaging. We present the system design in the Chapter 3 justifying the choice of sensors and the communication/interaction architecture for each of our motivating applications. In Chapters 4 and 5, we present a statistical framework for reliable localization using multi-modality complimentary state estimation techniques and demonstrate our approach in several real world situations.

3 System: Modular Design

This dissertation work was motivated by the need to build 3D models of dynamic large-scale environments using mobile robotic platforms. Towards that goal, we summarized the state of the art in 3D mapping in Chapter 2 and discussed methods for handling uncertainty in dynamic environments. In this chapter, we emphasize the need for a robotic architecture towards extracting scene characteristics using multi-modality sensors and provide the definition of a “sensor brick” architecture that showcases the potential of our robotic platforms as being robust, compact, modular and also autonomous independent units with high degree of interoperability. The deployment of multi-modality visual, range, position and orientation sensors, that can both act independently and in combination with one another, is a significant first step in achieving our goals listed in Chapter 1. We present our prototypes adhering to such a robust architecture that takes off-the-shelf hardware and through the efficient system design and requisite software development enables the integration of reconfigurable mobile robotic systems for the application in hand.

3.1 Modular architecture in system design

Our modular architecture in system design uses the idea of bricks as basic entities. A “sensor brick” is an autonomous platform promoting the notion of a three-module concept with mobility, sensing and communication capabilities. The sense-fuse-communicate (SFC) modules of each “sensor brick” have well defined functionalities. That is, the sensor module contains one or more sensors to collect data about the target environment, the fusion module processes this data and incorporates reasoning and analysis and the communication module transmits this information to appropriate end users. Thus, each “sensor brick” built on this SFC structure “sees”, “thinks” and “reports” as an independent, self-contained unit to a remote end user. The construction of such sensor bricks begins by first making the sensor component a “plug and play” device that when powered will automatically load in the information about the individual sensors in the brick and determine its functionality. A “mobility brick” is another basic entity that is a robotic platform capable of programmable locomotion at least by tele-operation if not completely autonomous. By such a construction, the task of improving functionality by combining the bricks is simplified because each sensor module is conscious about its composition and functionality, readily lending to interaction among sensor bricks within the architecture. We have pictorially depicted this concept in Figure 3.1.

Each “sensor brick” utilizes primary subsystems to acquire data, interact with a “mobility brick”, as well as the artificial intelligence and computing resources that are necessary for navigation, data processing, and communication. Additionally, each brick also has an independent power supply system which consists of any batteries, converters, or other devices pertaining to the operation of other components in the brick. The specific choice of each component within this brick architecture is driven by an important design consideration to use readily available, reliable, time-tested, rugged and inexpensive technology that will facilitate maintenance of the system components and also minimize adverse effects due to a component failure. In our implementation, this design consideration explains the use of off-the-shelf components from commercial vendors allowing integration and improvement of our bricks with advanced technologies in the future. The architecture definition and the implementation details are documented elaborately by Wilson (2005) and Naik (2006).

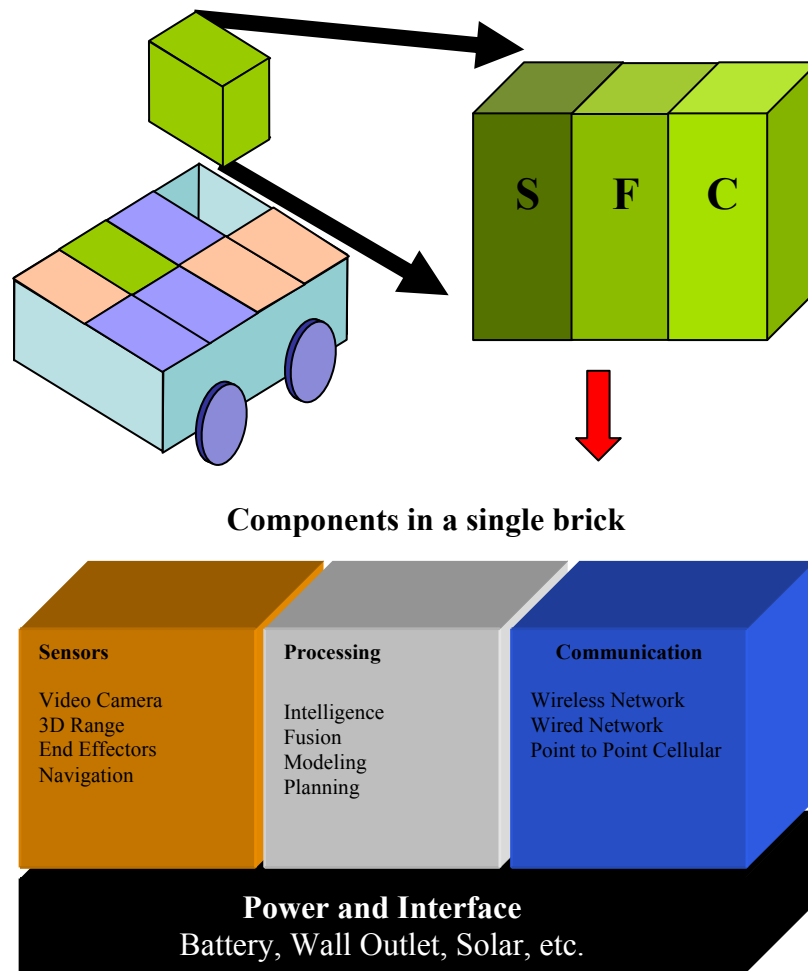


Figure 3.1: The sensor bricks fit into mobility platforms without having to reconfigure hardware or software in communicating information. They are able to do because each one of the bricks has its own power, communication, processing and sensing module with delineated tasks and task definitions.

Based on our study of the state-of-the-art robotic systems for inspection purposes, we observe that vision-based sensors provide useful information for humans to discern and make a decision about a threat in the scene. However, the threat may be in the form of explosives under the car or foreign object debris on the runway. In reducing the susceptibility to human carelessness or the failure to detect bombs that could cause destruction to property and life, we require “perception intelligence” from the mobile robot. By perceptual scene intelligence, we are referring to aspects of sensing abnormal “activity” in the scene.

For example, with under-vehicle inspection, the activity in the scene could be in the form of an additional object that is not supposed to be part of the undercarriage but attached to the vehicle, or could be neutron radiations from a nuclear source, or free radicals from a chemical explosive. Such scene information can only be inferred using specific sensors. An anomalous component in the undercarriage could be detected by visualizing the geometry from a range sensor, or by searching for abnormal temperatures of the components in the undercarriage using a thermal imager. Nuclear radiations, chemical explosives and biological agents can be also detected using specific sensor packages. With such advanced sensing technologies available, we are now confronted by the challenge of how to use these sensors in a real-world environment for automatically localizing potential threat objects in the under-vehicle scene. Such automation calls for interaction and fusion of spatial data from vision-based 2D and 3D sensors with 1D measurements from the nuclear, chemical and biological detectors. The sensor brick architecture achieves synergy using several sensors by taking hardware off the market shelf and programming the functionality.

The idea is illustrated for better understanding in Figure 3.2. As explained earlier, the modularity level determines the functionality of the system. Each brick can work on its own as an individual sensor and can co-operate with one another for automation requiring no new software. Operating individually as a “sensor brick” on a “mobility brick”, the mobility platform transforms into a visual inspection robot, or a thermal sensing robot with night vision or a path planning robot. When two or more of such bricks come together, in addition to increased functionality there is new software-based intelligence where the system automatically learns its capability to plan a course of action, map the underside in 3D and also take measurements of chemical activity to localize a threat. The robotic architecture monitors and ensures such synergistic sensor interaction, data fusion and communication. When a localization sensor like the SICK (SICK Laser measurement systems, 2006) scanner is included into the inspection robot, the configuration now realizes the potential to plan its own path, starts looking for the nearest car, gets close to the car based on the tire profile acquired and plans the navigation steps for mapping. Just before mapping, the architecture takes care of orienting the sensors and synchronizing the data collection from other sensors present in that level of modularity archiving multi-sensor data for threat localization.

The architecture also extends to systems for mapping large environments as shown in Figure 3.3. Irrespective of the application in being able to measure depth of cracks along the road, or being able to map hazardous environments, or capture a road surface for a vehicle-terrain simulator the modular architecture in integrating hardware components with intelligent software brings together different types of 3D sensors for localization and mapping along with pose measurement hardware like the GPS and the INS. We have included three different 3D acquisition methods (triangulation-based, time-of-flight and structured lighting) into our architecture. The reason behind including these methods was through initial experiments, where we concluded that the triangulation-based 3D sensing matched our requirements for high speed and high accuracy crack inspection (though we realized we might have to use an array of sensors for larger fields of view); for urban terrain like buildings the RIEGL (RIEGL Laser measurement systems, 2000) time-of-flight-based scanner was the better option, and for mapping terrain for simulators, the SICK scanner’s accuracy and resolution seemed to be sufficient. The SICK time-of-flight approach also is dependable for localization purposes. Hence, our software design accommodates all the three types of scanners and the architecture separates the data acquisition hardware and decides on their functionality making sure our integration methods accommodate several sensors.

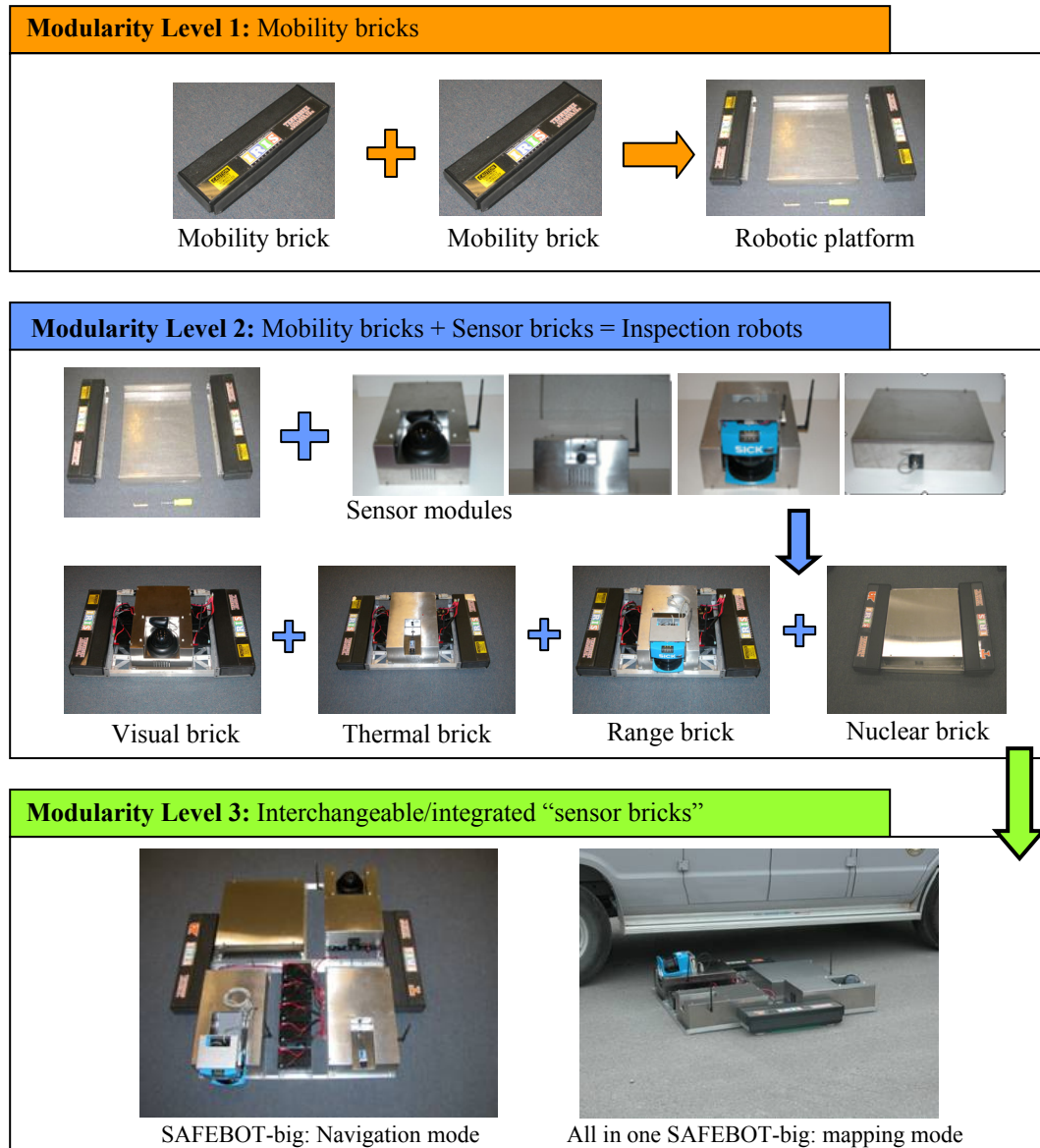


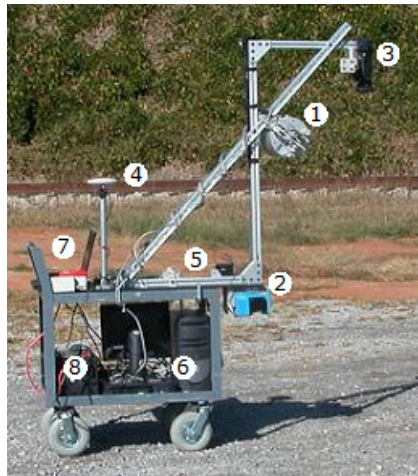


Figure 3.2: The “sensor brick” architecture lends to different levels of modularity. The base level begins with mobility platforms capable of housing self-contained wireless-enabled sensor modules that can both function independently on its own as a mobile robot and also in combination with other sensor bricks.

Sensors implemented into our architecture

Position and orientation sensors		Visual	3D range sensors		
					
Leica -GPS	Xsens IMU	JVC	RIEGL	SICK	IVP

Development prototype with several sensors



System prototype = Hardware + Software

- 1 – Riegl LMS-Z210 Laser Range Scanner
- 2 – SICK LMS 220 Laser Range Scanner
- 3 – JVC GR-HD1 High Definition Camcorder
- 4 – Leica GPS500 D-RTK Global Positioning System
- 5 – XSens MT9 Inertial Measurement Unit
- 6 – Processing unit for acquiring SICK, GPS, and IMU data and the architecture.
- 7 – Processing unit implementing integration and communication algorithms.
- 8 – Power system

Our prototypes exhibiting application-specific modularity

Urban-mapping	Terrain for simulators	Runway inspection
		

Figure 3.3: The “sensor brick” architecture lends to different levels of modularity with large-scale mapping also. We have shown the sensors implemented into the “brick” architecture where the sensors can act synergistically and in a synchronized fashion. After testing a development prototype on a push cart, we have improved towards application-specific professional packaging.




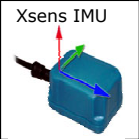


3.2 Hardware and software

In Section 3.1, we presented the implementation of different sensors into a common modular architecture simulating the device plug-and-play concept for robotic 3D mapping. In this section, we describe the hardware components used to digitize 3D environments as accurately and quickly as possible with available technology and processing equipment. Our focus is more on the image acquisition system than the processing. Earlier in Chapter 2, we had argued that the map of any environment can only be as good as the 3D sensing method and the localization hardware. The modular architecture was hence formulated to include different types of sensors making the 3D mapping process less dependent on the application in hand.

There are different versions of the 3D mapping system that we have developed (Figure 3.3). We call the right inset runway inspection system in Figure 3.3 as a “micro-scale” system that is capable of measuring the dense geometry (1mm between data points along a profile) with an accuracy of 2 mm, at the cost of having a limited field-of-view but potentially exorbitant size of raw data to process. The other system uses laser range scanners designed to scan larger objects, similar to the RIEGL or the SICK scanner. The “micro-scale” system uses the IVP Ranger SC-386 (Integrated Vision Products, 2000) laser profiling system to acquire high-resolution 3D models of road surfaces. This level of resolution is useful for tasks such as pavement inspection. The macro-scale terrain scanning system uses a SICK LMS-200 and the RIEGL LMS-Z210 scanners to acquire large-scale terrain models. We use all three scanners in the continuous profiling mode while mapping. The sheet-of-light triangulation-based 3D sensor that we use is capable of acquiring 2000 profiles/second that corresponds to 6 mm separation between successive profiles driving at 30 miles/hour. In terms of accuracy, our system that was placed 70 cm above the road surface and configured for a baseline of 70 cm and a triangulation angle of 45 degrees gives 1 mm accurate depth information. The price that we however pay in using such a system is the field-of-view. We are able to scan 0.6 m wide section of the road using a single sensor. We believe using an array of such sensors as a possible solution to large area micro-scale data collection. For other applications such as terrain and urban mapping, we list the specifications of the sensors we recommend in Figure 3.4.

For our integrated maps to have global reference, we collect physical location information by setting up a GPS base station and placing a receiver on the mobile platform. The GPS data is accurate up to 3 cm in the motion direction and gives us 10 samples of position information in one second. The GPS can be thought of as sampling the 3D motion of the mobile platform that houses the sensors. In the prototypes shown in Figure 3.3, we see a video camera mounted on a rod, whose image axis is orthogonal to the surface of interest. We prefer the orthogonal field-of-view for generating texture because it makes the registration of range and intensity profiles trivial and considerably improves integration time without having to consider CCD calibration and rectification of images. In addition to using video for texture, we also use additional cameras to help estimate the motion of the sensor platform as back up to GPS satellite signals that may not be intermittently available during certain time intervals of the day. Also, as a backup localization system, and for compensating the vibrations and the resulting oscillations on the mobile platform caused by the suspension system in unstructured terrain, we have used the inertial measurement unit (IMU) for measuring the orientation Euler angles (roll, pitch and yaw) of the sensor mount during data collection. Our IMU unit is manufactured by Xsens whose recent product is up to a-tenth of a degree accurate with its orientation measurements.

We use two high performance multi-core computers with Pentium 4 processors supporting hyper threading with 1 gigabyte of memory along with special high speed serial interface cards as processing equipment. One computer hosts our acquisition programs (and the architecture) as multi-threaded, multi-document graphical user interfaces written in C++ capable of real-time buffer memory management and storage while the other computer takes care of the processing and integration we describe in the Section 3.3.

 <p>RIEGL</p>	<p>The time-of-flight range scanner is a large scale mapping instrument with an accuracy of 5 cm that can digitize 3m – 300 m far objects at 21 profiles/sec. This scanner is used for mapping large scale urban environments centimeter accuracy is sufficient.</p>	 <p>Leica -GPS</p>	<p>Our differential GPS system consists a base station and the rover. The base station is placed in an area with no obstructions and the rover with one GPS radio receiver and antenna is mounted on the mobile platform (with a common axis and in line with 3D sensor) to communicate with the base station. With clear weather and satellite availability positional accuracy of up to 2 cm is achievable.</p>
 <p>IVP</p>	<p>The triangulation-based laser scanner can provide depth accuracy of 1 mm configured for a baseline of 70 cm and stand off 70 cm and a triangulation angle of 45 degrees gives 1m width of view at 2000 profiles/sec. We recommend the use of this system for applications like crack inspection that require detail and accuracy.</p>	 <p>Xsens IMU</p>	<p>Is placed as close as possible to the 3D range sensor and measures orientation angles accurate up to a tenth of a degree. Inertial measurements are generally noisy and have to be characterized before we can use the data for 3D profile alignment.</p>
 <p>SICK</p>	<p>The range accuracy of the high-speed, time-of-flight range scanner is up to 5 mm in mapping 16 m wide profiles of the scene at 30 profiles/sec. This scanner can be noisy and might require prior noise characterization. We use this scanner in mapping road terrain where cm accuracy is sufficient.</p>	 <p>JVC</p>	<p>Placed on a rod for increased field-of-view orthogonal to the road surface and also pre-registered with the range profile. This high resolution video camera can provide as 1024 x 720 image frames at a frequency of 30 Hz.</p>

Sensor Performance

	RIEGL	SICK	Leica -GPS	Xsens IMU	Video camera
	LMS-Z210	LMS200	GS500	MT9	JVC-HD1
Measurement Range	300 m	8 m	worldwide	300 deg/sec	visible
Acquisition Rate	10K points/sec	35 Hz	10 Hz	100 Hz	30 Hz
Resolution	1 cm	1 cm	1 cm	0.5 deg	1280 x 720i
Accuracy	5 cm	1.5 cm	2 cm	< 1 deg	n/a
Computer Interface	RS232/Parallel	RS422	RS232	RS232	IEEE 1394
Operating Voltage	12 VDC	24 VDC	10.5-32 VDC	3.4-12 VDC	7.2 VDC
Dimensions (W x L x H)	21x21x44 cm	16x19x16 cm	13x13x9 cm	58x58x22 mm	12x27x10 cm
Weight	13 kg	4.5 kg	600 g	50 g	1.3 kg

Figure 3.4: Specification of the components in our modular approach along with some design notes towards reproducing our system. We have included the size, weight factors to emphasize the portability and robustness. We have also provided the sensor characteristics and their expected accuracy that will later be used as a bound in the noise model for the sensors.

3.3 Multi-modality multi-sensor data integration

We have built our system by choosing sensors based on the application requirements. These sensors fall into two categories as pose recovery sensors (GPS, INS, cameras) and structure recovery sensors (cameras and laser scanners) with the potential of using some structure recovery sensors to infer pose also. Though the idea of mapping appears trivial once the GPS provides global location and IMU provides relative orientation information to align the 3D profiles from the laser scanners into a global co-ordinate system, we have to discuss several issues before we actually deliver an integrated 3D model. In this section, we describe the procedure for integrating the data collected from multiple sensors, into one complete single multi-modal 3D dataset. The processing steps that we implement are shown in a block diagram in Figure 3.5. The task of spatial alignment is not trivial because; each of the measurement systems has their own reference co-ordinate system differently oriented in free space. Hence as a first step towards integration and fusion of the data, we need to choose a global reference co-ordinate system to represent our data. We use the GPS co-ordinate frame as our reference frame and transform the range and intensity profiles to that frame without losing geographic location information of the scene.

We need to deal with another important issue before transforming the data to the real-world co-ordinates. We attribute this issue to different acquisition rates from different sensors. The GPS supplies data at a frequency of 10 Hz, video camera at 30 Hz, the IMU at 100 Hz, while the range profiles are acquired at nearly 2000 Hz. We have two ways of resolving this issue (1) to discard the range data and use the profiles that are time synchronized with the GPS data or (2) use all the points of the range data and align the profiles based on interpolated GPS path at time instants that we have acquired the range data. We lose more information in discarding acquired data by choosing the former solution. We hence suggest cubic spline interpolation of the GPS path as a 3D curve at time stamps recorded by the range sensor. The IMU orientation data also needs to be interpolated. But before that, having characterized our IMU sensor, we apply moving average smoothing techniques to reduce the noise in its measurements before interpolation. For pose recovery from images we have implemented (Nister, 2004a) and for pose recovery from range we have implemented (Martinez, 2006). We implement passive pose recovery from cameras and laser scanners for complementary state estimation. However, the implementation of (Nister, 2004a) and (Martinez, 2006) can result in conflicting states differing from the GPS and INS measurements. The uncertainty analysis block takes care of the belief propagation on sensor data before spatial alignment in such situations.

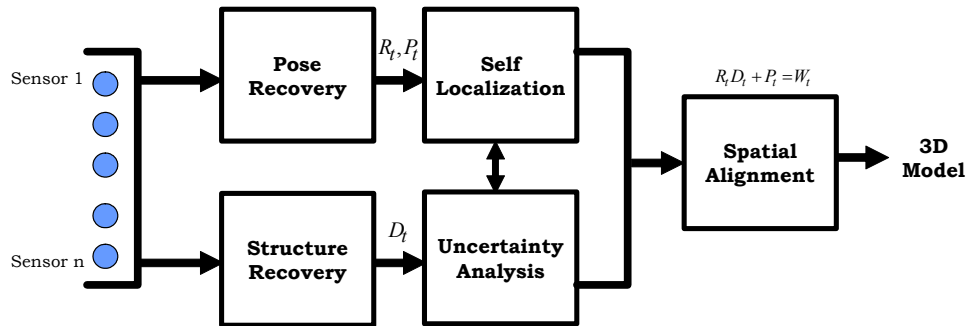


Figure 3.5: The block diagram for integrating multi-sensor data into a 3D model. The sensors provide localization and structure information which is fused and aligned into a 3D model. We have included an uncertainty analysis step before the alignment to handle dynamic situations in the real world.

Towards the spatial alignment of data, if we can denote the Euler angles of roll, pitch and yaw from the IMU by (ω, ϕ, κ) and 3D range measurements by $D_t = (x_r^t, y_r^t, z_r^t)$, at a particular time t , we assume that we have already interpolated localization sensor data to synchronize in time with the range profiles and let the GPS measurements, after considering the moment arm distance along each dimension of the range sensor from the GPS receiver be (x_g, y_g, z_g) , the mapping to the real-world co-ordinate system W_t of the profile acquired for that instantaneous time t can be computed using Equation 3.1.

$$R_t D_t + P_t = W_t, \text{ where} \quad (3.1)$$

$$R_t = \begin{bmatrix} \cos \phi \cos \kappa & \sin \omega \sin \phi \cos \kappa + \cos \omega \sin \kappa & -\cos \omega \sin \phi \cos \kappa + \sin \omega \sin \kappa \\ -\cos \phi \sin \kappa & -\sin \omega \sin \phi \sin \kappa + \cos \omega \cos \kappa & \cos \omega \sin \phi \sin \kappa + \sin \omega \cos \kappa \\ \sin \phi & -\sin \omega \cos \phi & \cos \omega \cos \phi \end{bmatrix} \quad (3.2)$$

$D_t = [x_r^t, y_r^t, z_r^t]^T$ is the measurement from the 3D range sensor and
 $P_t = [x_g^t, y_g^t, z_g^t]^T$ is the position of the range sensor through GPS measurements.

The transformation and alignment based on multi-sensor data (collected over a time period) gives us an unorganized point cloud of data that, for visualization purposes, we triangulate using the method described by Hoppe in (Hoppe, 1992). Our experience indicates that triangulation should be performed on smaller patches as the data is acquired and later merged into a large 3D dataset. We represent the final triangulated mesh as a VRML model that can be rendered in most computers that support OpenGL or DirectX. The dense point cloud is then converted into a mesh that can then be textured using the color images from the video. By the design of our setup and initial hardware registration step, we can actually map the color pixels in the CCD to the range profile as a quick method for multi-modal visualization. The process of digitizing a real world scene by sampling the geometry as points and profiles D_t , sampling color using cameras and aligning geometry and color in a global co-ordinate frame can be better understood from illustration in Figure 3.6.

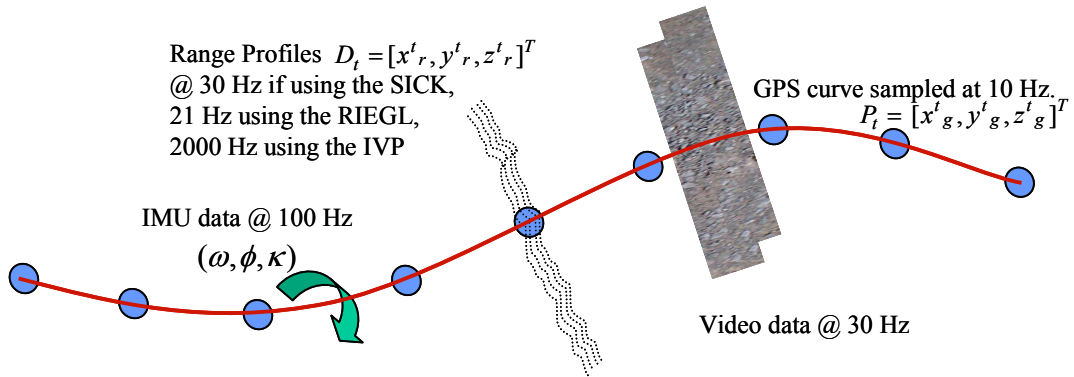


Figure 3.6: Spatial integration of multi-sensor data requires a global reference frame and interpolation to consider different sampling rates of sensors. The range profiles are in a local co-ordinate frame that is transformed into the GPS co-ordinate frame based on the self-localization data and integrated as a textured 3D model

If the sensors are all functional and perfect, there would be no error in the integrated 3D map after spatial alignment. But the sensors are noisy and can fail. The noise manifests in the localization measurements and also in the structure measurements as shown in Figure 3.7. Uncertainty in the state recovered during self localization propagates as uncertainty into the integrated map. Hence, before we claim robustness to noise and a bound on the error in the integrated 3D map using our systems, we have to handle uncertainty from both of these sources. Our modular design, to include different sensors, minimizes the measurement uncertainty in the geometric samples of the scene. But, we still have to deal with the uncertainty in localization. Also, if the mapping has to be performed autonomously, localization appears to be a much more significant and tougher problem requiring models for predicting expected uncertainty from the sensors. We present the uncertainty models for localization for a mobile robot with vision and range sensors in the following section.

3.4 Belief propagation theory for modeling and handling uncertainty

We have realized both from the literature and through our experience that there is uncertainty about how the system changes its state and how the sensors measure state, particularly, when we have to deal with sensors that may be subject to situations that perturb their measurements in unpredictable ways. Sensors can fail in uncertain dynamically changing environmental conditions that cannot be modeled apriori and detecting such a fault is the tough task in our hands. In this section, we present methods to formulate the autonomous localization and mapping problem as a recursive state estimation problem to deal with expected sensor noise. Then, we will discuss methods to model the noise characteristics of the sensors and the actuators to generate what is called belief distributions about the localized pose or state of the mobile platform in an environment. Following (Thrun et al., 2005), we introduce a model for the state transition that will include the uncertainty of the robot's actuators and a sensor model that will include the noise characteristics for the measurements of robot pose. The problem of handling uncertainty and making decisions in dynamic situations gets simplified within the probabilistic framework. We build over the idea of state transition/sensor measurement models and belief propagation towards learning the performance of a particular sensor to handle situations beyond the expected noise levels in Chapter 4.

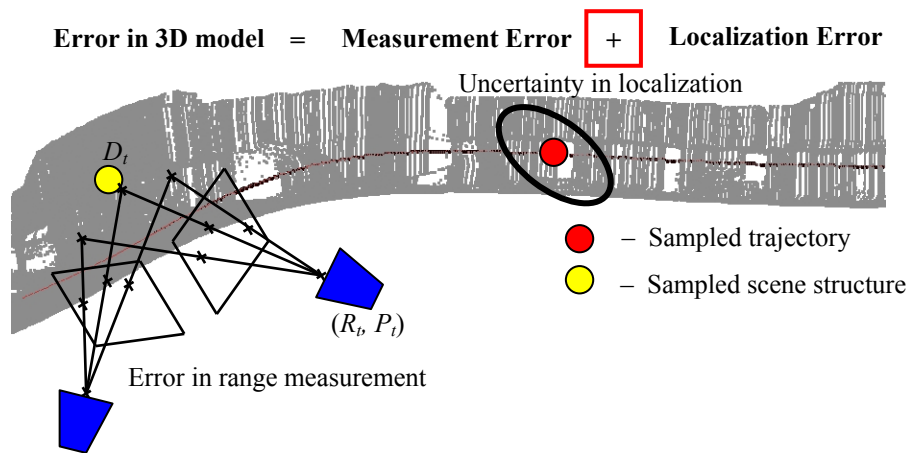


Figure 3.7: Uncertainty in the 3D map is a combination of error in the localization and the measurement uncertainty in the structure recovery sensors. The measurement uncertainty is only as good as the sensor used and the best we can do is by choosing and implementing the right sensors into our architecture.

With that motivation, let us begin with a sensor S that measures state x as $\hat{x} = x + \varepsilon$ with error ε within a bounded noise model $p(\hat{x} | x)$ in the form of a conditional probability distribution. If the sensor is a direct state measurement device like the GPS or the INS that measures pose the bounded noise model $p(\hat{x} | x)$ can be empirically estimated through different characterization techniques before mounting on the robot or obtained from the manufacturer's specifications. The typical state vector \hat{x} with robot localization is usually position and orientation with 3 to 6 degrees of freedom in motion. Sometimes, derivatives of pose are also included in the state vector for maintaining continuity and smoothness.

We will assume that we are able to approximate the uncertainty Σ_0 about the initial state and associate an initial belief distribution about the initial state $x_0 = \text{Belief}(x_0)$. The initial belief distribution could be a uniform distribution if there is no prior knowledge or a tightly bound Gaussian based on sensor characterization. With most of our engineering equipment, we can assume that $\text{Belief}(x_0)$ is normally distributed with parameters (x_0, Σ_0) . Starting from initial state x_0 , we associate control commands the robot executes at time t as u_t to propagate the uncertainty belief in the initial state. We also assume that the robot control systems are calibrated apriori and we have an approximation to $p(x_{t+1} | x_t; u_t)$, the state transition probability. The state transition probability $p(x_{t+1} | x_t; u_t)$ represents the probability that the system state is x_{t+1} given that the system executed the command u_t while at state x_t .

Every time, the mobile robotic platform executes a motion command, we have to include the noise characteristics of the actuators before we can compute the uncertainty in the form of a belief distribution about the new state of the robot and $p(x_{t+1} | x_t; u_t)$ takes care of that. The sensor S is trying to help us avoid drifting from the goal due to the actuator noise by updating us with measurements of the state of the robot. The sensor measures the current state as \hat{x}_t with the uncertainty $p(\hat{x}_t | x_t)$ as feedback to localize the robot. Then, the problem that we have in our hands is a recursive state estimation problem that we illustrate in Figure 3.8, where the state of the system changes from state x_t by executing control signals u_t to a new state x_{t+1} that is measured by sensors as \hat{x}_{t+1} . Since the measurements \hat{x}_{t+1} are noisy too, we have to determine the new state and the uncertainty about the state by compensating for the noisy sensor feedback and the noisy actuator command executed. We describe the methods to propagate the belief and the uncertainty about the state in the following paragraphs.

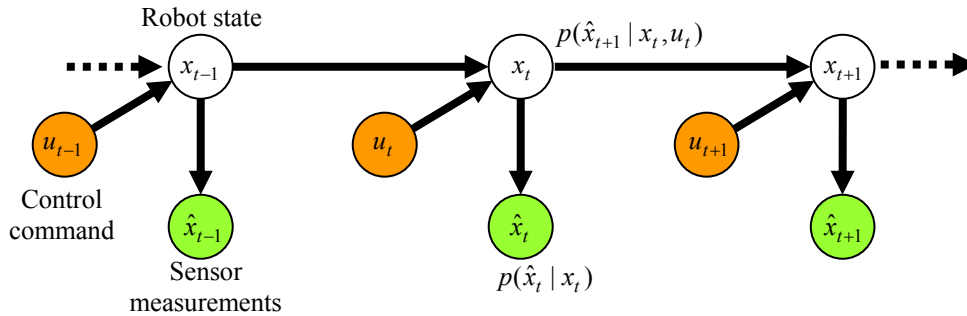


Figure 3.8: The simplified Bayes network representing localization as a recursive state estimation problem. The dotted arrows are symbolic of the Markov assumption that the current state of the system depends only on the previous state and the most recent control command lessening the burden of having to integrate a history of measurements.

We have introduced two probabilities, the state transition probability $p(x_{t+1} | x_t; u_t)$ and the sensor bounded noise model probability $p(\hat{x}_t | x_t)$. Belief propagation via Bayesian filtering techniques (Thrun et al., 2005) provide a powerful statistical tool to help manage uncertainty using these probabilities. Bayes filters represent the state at time t , x_t by random variables, and the uncertainty at each state by a probability distribution over all possible x_t called $Belief(x_t)$. The key idea being able to sequentially estimate such belief distributions based on the information contained in the sensor data. If the sensor measurements of state are indexed as $\hat{x}_{1:t}$, then the belief distribution is defined as the posterior density over the random variable x_t , conditioned on all available sensor data up until time t by Equation 3.3.

$$Belief(x_t) = p(x_t | \hat{x}_{1:t}) . \quad (3.3)$$

Intuitively, the belief distribution can be considered to answer the question “What is the probability that the robot is at location given the history of sensor measurements for all possible locations of x ?” Because of the exponential nature of the problem to include history in computing the posterior distributions, we will make the assumption that the robot dynamics is sufficiently modeled as a Markov process, that is, all information about previous states is encapsulated in the current state x_t . This assumption reduces the belief propagation considerably to a much familiar form of the fundamental Bayes theorem. Since in our case, we have two systems to monitor, namely the controls and the sensor feedback, belief propagation is carried out in two steps (i) prediction and (ii) correction. The prediction follows the rule in Equation 3.4 and the correction includes sensor models as shown in Equation 3.5.

$$Belief_{predictor}(x_{t+1}) = \int p(x_{t+1} | x_t; u_t) Belief(x_t) dx_t . \quad (3.4)$$

As mentioned earlier, $p(x_{t+1} | x_t; u_t)$ describes how the state of the robot changes over time. With the state transition probability encapsulating the uncertainty in system dynamics, the predictor belief distribution is the conditional probability that answers where the robot might be given that the robot was previously in x_t and it executed a motion command u_t . But, since the robot motion controllers are not as accurate as one would expect, the robot will not reach the desired state x_{t+1} . Over a period of time, these errors can accumulate taking us completely away from the desired goal. Hence we need to correct the posterior belief of the state by using sensor data as feedback using Equation 3.5.

$$Belief_{correction}(x_{t+1}) = \eta p(\hat{x}_{t+1} | x_{t+1}) Belief_{predictor}(x_{t+1}) . \quad (3.5)$$

Recollect that $p(\hat{x}_t | x_t)$ is the perceptual model determined apriori for each sensor used in the robot that describes the likelihood of making an observation \hat{x}_t if the robot was actually at state x_t . The term η in Equation 3.5 is a normalizing constant which ensures that the posterior over the entire space sums to 1. With $Belief(x_0)$ known apriori, (or at least modeled as a uniform distribution if we don't have the required information), the performance of the belief propagation procedure depends on the accuracy of the perceptual model, the dynamics model $p(x_{t+1} | x_t; u_t)$ and the representation used for $Belief(x_t)$. With $Belief(x_0)$ known apriori, (or at least modeled as a bound uniform distribution if we don't have enough apriori information) the performance of the belief propagation depends on the accuracy of the perceptual model, the dynamics model $p(x_{t+1} | x_t; u_t)$ and the representation used for $Belief(x_t)$. That is, the success of belief propagation not only depends on the statistical parameters that describe distributions $p(x_{t+1} | x_t; u_t)$ and $p(\hat{x}_{t+1} | x_{t+1})$ but also on the type of the functional form defined as Gaussian, Rayleigh, uniform, exponential etc.

The most commonly used form of the Bayes filter is the Kalman filter (Kalman, 1960). Kalman filter approximates belief as unimodal Gaussian distributions, parameterized by the mean μ_t and the covariance Σ_t . The mean of the posterior belief estimated using the Kalman filter is the expected location and the covariance representing the uncertainty in the mean estimate. In spite of the Gaussian assumption, Kalman filters have been applied successfully in robot localization tasks, the main advantage being their computational efficiency. Multi-hypothesis tracking (Jensfelt and Kristensen, 2001) extends the Kalman filter to include multi-modal belief, by representing the belief distribution as a mixture of Gaussians, where each mixture is then tracked using a separate Kalman filter. Particle filters (Doucet et al., 2001) on the other hand represent beliefs by sets of weighted samples distributed according to the previous belief. The method based on sequential importance sampling can represent arbitrary probability densities, which makes them extremely robust compared to Kalman filter. However, the worst case complexity of the sampling-based particle filter is exponentially cumbersome as the dimensionality of the state increases.

We will use Kalman filters in our implementation because the control system and the sensors that we use in our system prototype follow Gaussian distributions for $p(x_{t+1} | x_t; u_t)$ and $p(\hat{x}_t | x_t)$. Further, the conclusions in Fox et al. (2003) supports our decision that Kalman filters are best suited for vision cameras, laser-range finders and GPS systems. We explain the equations that govern belief propagation using Kalman filters in Table 3.1. Kalman filters in addition to the state transition probability $p(x_{t+1} | x_t; u_t)$ with covariance R and the bounded noise model $p(\hat{x}_t | x_t)$ with covariance Q , requires also the motion model that decides how the state changes for an input command u_t , and the sensor model of how a sensor would measure the new state x_{t+1} . An example of a generic linear form for a motion model and a sensor model is Equation 3.6 and 3.7 respectively.

$$\text{Motion model: } x_{t+1} = F_t x_t + B_t u_t + G_t w_t; w_t \sim p(x_{t+1} | x_t; u_t) \quad (3.6)$$

$$\text{Sensor model: } \hat{x}_{t+1} = H_{t+1} x_{t+1} + \varepsilon; \varepsilon \sim p(\hat{x}_t | x_t) \quad (3.7)$$

We explain the formulation of the motion model and the sensor model with the help of Figure 3.9. But before that, we note that the procedure listed in Table 3.1 does not necessarily indicate that for every action (motion command) there is a sensor measurement update. There can be situations where there are several motion commands before one sensor update, or more sensor updates compared to the motion. Factors that decide the number of prediction cycles (Step 2) and sensor update cycle (Steps 3-5) to propagate belief are the velocity expected out of the robot and the sampling frequency of the sensors used.

Table 3.1: Kalman filter algorithm adapted from (Thrun et al., 2005).

<i>Kalman filter algorithm in equations</i>	<i>What happens at each step?</i>
Input: (μ_t, Σ_t, u_t)	Current belief and the motion command.
Step 1: $\bar{\mu}_{t+1} = F_t \mu_t + B_t u_t$	Naïve state update based on system dynamics.
Step 2: $\bar{\Sigma}_{t+1} = F_t \Sigma_t F_t^T + R_t$	Motion model prediction: Uncertainty grows.
Step 3: $K_{t+1} = \bar{\Sigma}_{t+1} H_{t+1}^T (H_{t+1} \bar{\Sigma}_{t+1} H_{t+1}^T + Q_t)^{-1}$	Kalman gain K estimates how much to correct.
Step 4: $\mu_{t+1} = \bar{\mu}_{t+1} + K_{t+1} (\hat{x}_{t+1} - H_{t+1} \bar{\mu}_{t+1})$	Includes Kalman gain with residual error.
Step 5: $\Sigma_{t+1} = (I - K_{t+1} H_{t+1}) \bar{\Sigma}_{t+1}$	Uncertainty estimate shrinks based on measurement.
Result: $(\mu_{t+1}, \Sigma_{t+1})$	Parameters of the belief distribution at $t+1$.

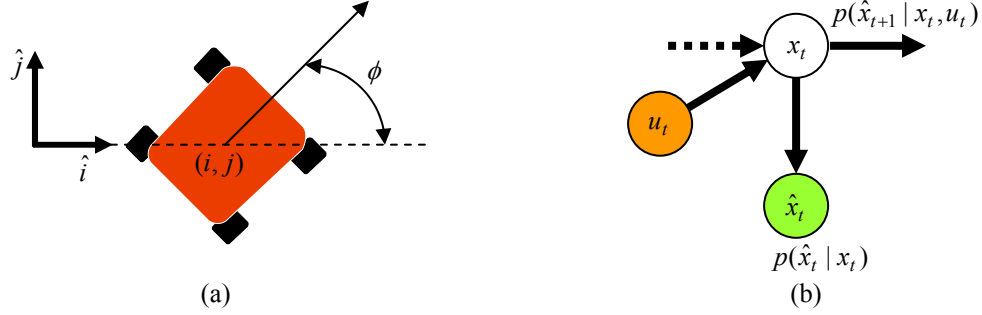


Figure 3.9: An example for formulating the motion model for a robot with three degrees of freedom in motion. (a) A robot and the reference axis. (b) The recursive state-space estimation framework that we would like to use for localization and mapping.

Let us consider an example to understand the derivation of a state-space model for the robot locomotion. Suppose, we had a robot like the one shown in Figure 3.9 in a global reference axis, the state vector x_t of a robot considering three degrees of freedom can be assumed to have the following form.

$$x_t = (i_t, j_t, \phi_t, v_t) \quad (3.8)$$

where (i_t, j_t) refer to the spatial position in a global co-ordinate frame, ϕ_t is the instantaneous orientation of the sensor and v_t is the magnitude of velocity at time t .

The motion dynamics for this robot can be modeled in the physical world as the following set of equations 3.9, 3.10 and 3.11.

$$i_{t+1} = i_t + v_t \cos \phi_t + \varepsilon_i; j_{t+1} = j_t + v_t \sin \phi_t + \varepsilon_j \quad (3.9)$$

$$\phi_{t+1} = \phi_t + \Delta \phi_t + \varepsilon_\phi; v_{t+1} = v_t + \Delta v_t + \varepsilon_v \quad (3.10)$$

$$\varepsilon_i \sim N(0, Q_i); \varepsilon_j \sim N(0, Q_j); \varepsilon_v \sim N(0, Q_v); \varepsilon_\phi \sim N(0, Q_\phi) \quad (3.11)$$

where $\Delta \phi_t$ and Δv_t are motion commands to the actuators and noise models ε follow a Gaussian distribution.

Then the linear state-space motion model $x_{t+1} = Ax_t + Bu_t + \varepsilon_t$ translates to Equation 3.12.

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ \phi_{t+1} \\ v_{t+1} \end{bmatrix} = \begin{bmatrix} x_t + V_t \cos \phi_t \\ y_t + V_t \sin \phi_t \\ \phi_t \\ V_t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \Delta \phi_t \\ \Delta v_t \end{bmatrix} + \begin{bmatrix} \varepsilon_i \\ \varepsilon_j \\ \varepsilon_\phi \\ \varepsilon_v \end{bmatrix} \quad (3.12)$$

We note that the state transition probability $p(\hat{x}_{t+1} | x_t, u_t)$ is incorporated into the motion model in the form Q_i, Q_j, Q_v and Q_ϕ variances.

We use a similar model as the one presented with our robots for small environments with three degrees of motion. For large-scale mapping and navigation, we allow six degrees of freedom for motion and derive the motion model in a complicated form of Equation 3.12. With that example of deriving the state model for robot localization the next model that we discuss is the sensor model. The sensor model defines the relationship and interaction of how the state is estimated by the sensor as a measurement. The sensors in our application that provide estimates of the state vector in the form $\hat{x} = x + \varepsilon$ are the GPS and INS instruments. The sensor model $\hat{x} = x + \varepsilon$ is very straightforward and simple for direct pose measurement GPS and INS devices. But indirect pose estimation sensors like the range sensor and the camera, do not measure state directly.

For the vision-based SLAM, the measurements are typically landmarks l_i identified as pixels in an image sequence. In addition to the robot pose with six degrees of freedom (3D position and orientation), we consider landmarks observed in the environment as part of the state vector also and model the relationship between the 3D co-ordinate of the landmark and the pixel measurement as $l_i = H_{vision}(x_t) + \varepsilon$ in a similar form as Equation 3.7. If $X_c = [x_c, y_c, z_c]$ be a 3D landmark co-ordinate in the camera reference frame (not global latitude and longitude) and the landmark l_i identified as pixel (u, v) in the image, the relationship between the measured landmark and the state of the system in the camera reference frame is modeled as Equation 3.13.

$$H_{vision}(x_t) = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix} + (1 + k_1 r^2) \begin{pmatrix} f_x \frac{x_c}{z_c} \\ f_y \frac{y_c}{z_c} \end{pmatrix} \quad (3.13)$$

where $r^2 = \left(\frac{x_c}{z_c}\right)^2 + \left(\frac{y_c}{z_c}\right)^2$ and f_x and f_y are focal lengths in the horizontal and vertical directions, (θ_x, θ_y)

is the principal point of the camera and k_1 is the radial distortion co-efficient estimated through apriori calibration. ε is normally distributed sensor noise whose parameters are estimated through sensor characterization by imaging ground truth data at the same distance (range) as the scene. Davison et al. (2007) provides the details for monocular vision-based SLAM. We use Davison's (2007) implementation in our systems.

Thus far, we presented the idea to model the uncertainty in locomotion of the mobile robot and also model the uncertainty in the sensors taking measurements from the real world and relating the sensor measurements to the scene for improved state and structure estimation. These models dictate the prediction correction stages in the algorithm summarized in Table 3.1.

We implement the extended Kalman filters (Thrun et al., 2005) as our Bayesian filtering technique into our integration software for each sensor individually. The extended Kalman filter allows differentiable functions for the state transition and sensor observation models allowing non-linearity in Equations 3.6 and 3.7 presented earlier. The filter operating on each sensor will output the most likely state and the uncertainty about the state considering the state transition noise and the sensor noise while compensating for either. We choose a sufficient time interval as a parameter in the architecture for deciding to fuse the state measurements after Bayesian filtering from the GPS, INS, vision and range sensors. The time interval tries to accommodate for the fact that sensors are operating at different sampling rates in our mobile mapping system by adjusting the number of prediction and correction cycles within the filter for each sensor. Within our modular approach each sensor brick is programmed for localization and mapping using an implementation of the SLAM procedure. The vision brick has an implementation similar to (Davison, 2007) and the range brick implements (Martinez, 2006) without the genetic algorithms. The models for uncertainty within the SLAM modules are derived through extensive sensor characterization.

We would like to discuss a few issues about using the Kalman filter and the Bayesian filters in general before concluding this chapter. The first one is about the error that is propagated. Theoretically speaking, we observe that Kalman filters eliminate the worst case performance of noisy sensors. They are able to achieve such performance because of the fact that, if the actuator uncertainty is large, sensor updates dominate the propagation, and if the sensor uncertainty is large, the motion process propagation dominates the state estimate. However, there is an inherent problem as well, which is called the “health” of the Bayes filter when the residual in Step 4 of the algorithm in Table 3.1 blows above the 3Σ bound of the apriori uncertainty estimate. Such a case caused by sensor failure or performance degradation, results in erroneous estimates of the state and the uncertainty resulting in a rapidly diverging estimate of belief. This is the major concern with our mapping systems in large-scale environments. As emphasized in the concluding section of Chapter 2, the GPS system error can become 10-fold when not completely functional; the pose from video may also not be completely reliable in certain areas and such instances of performance degradation can contribute to the increased uncertainty in localization and the integrated map.

Based on our experience, we are able to conclude that there are two types of sensor problems associated with position and orientation (pose) uncertainty in localizing a mobile robotic platform: (i) the sensor noise and (ii) the validity of sensor measurements. Leonard and Durrant-Whyte (1992) also support our conclusion. In the robotics literature, the uncertainty due to the sensor noise is well understood and is efficiently handled by using one of several Bayes filters by representing uncertainty using probability density functions under bounded noise models. However, the second problem of sensor validity attributed to the dynamic nature of environments poses a greater challenge because uncertainty due to sensor validity extends beyond the boundaries of modeling noise. In the following chapter, we describe a potential solution to robot localization by modeling and minimizing both types of uncertainty in dynamic environments.

4 Theory: Selfish-Altruist Fusion

The problem we have in our hands is that of data fusion amidst sensor conflict and performance degradation in multi-sensor state estimation, with focus particularly on the state measurements meant for localization of 3D mapping mobile robots. We motivate our statistical solution to this problem from philosophical observations of survival and adaptation in Section 4.1 and derive a mathematical procedure for quantifying a sensor's self and co-operative confidence. We discuss the idea to guide the fusion amidst uncertainty in Section 4.2. Further, we explain our approach for quantifying self confidence in Section 4.3 and co-operative confidence in Section 4.4. Our formulation helps derive a criterion to resolve the dilemma between fusion and selection. We conclude this Chapter by discussing the criterion in Section 4.5.

4.1 Philosophical inspiration for statistical inference

We have already drawn inspiration from adaptive biological organization towards the design of our mobile robotic mapping systems. Particularly, the idea behind creating flexible subsystems in the form of sensor bricks is that the modular design enables the collective function of subsystems that is much more efficient and resilient, even in face of near-catastrophic failures. The inspiration is from the tiny ants, a colony of relatively simple individuals with delineated tasks that coordinate actions efficiently to locate food sources, relocate and survive in hostile dynamic environments. Though our sensor bricks (vision and range) try to simulate the ants as autonomous agents with SLAM capability, locating and memorizing spatial landmarks and relationships, we have a problem in dynamic situations exposing the need for implementing methods for intelligent behavior.

The evolutionary stable solution (Krebs and Davies, 1997; Parker, 1984) studied in the analogous problem of intelligence among biological species is that of adaptation. Adaptation as Lerman (2004) notes is an essential requirement for systems composed of autonomous agents functioning in dynamic environments. By dynamic environments, we are referring to places like urban canyons, tunnels and mines of which we have no apriori judgment and are explored for the first time. Adaptation allows the system of agents, in the form of multi-modality sensors on a robot or nodes implemented into a sensor network to change their behavior in response to changes in the environment and actions/inactions of fellow agents. Such a behavior guarantees improvements on the overall performance of the system as a whole. In our case, the autonomous agents are the sensor bricks with sensors for pose and structure recovery that aims towards a higher autonomous behavior of robot localization in unknown environments. While localizing in such environments, autonomous survival implies constantly providing feedback in the form of a spatial map in dynamic environments while functioning at a higher level of self-realization than what is expected out of the sensor systems individually.

The economics of survival with continuous resource management works best when the entities of the system or society act independently with a selfish attitude and evolve to make philanthropic decisions to co-operate for higher success (Axelrod, 1984; Chavas 1993). Fortunately for us humans, we have the survival machines in the form of the genes working "over time" (Dawkins, 1982) dictating the survival of the fittest through natural selection. How can we get computers to do this task for our mobile robots?

We begin by relating the survival of our mobile mapping system in an environment to its success in fusion of multi-sensor data. The sensors are the autonomous agents working towards better self-realization that in robotic terms is better state localization. We treat the sensors in a system like entities of a species allowing them the freewill to act on their ego, dispense their duty to localize, and also permit them to compete aggressively for leadership with fellow sensors in the system. In such a scenario, we would ideally want each sensor to diligently succeed individually in state recovery and in the face of trouble or changing environmental conditions evolve to co-operate with other sensors placing the survival of the imaging system ahead of them as individual bricks. This evolution is exactly how the human race has succeeded over the years by acting as individuals for most of their life time with sporadic acts of philanthropy and altruism towards bigger goals. The bigger goal could be in the interest as big as the human presence on the planet or as trivial as the success of an organization. Targeting such success, we derive our selfish-altruist data fusion algorithm.

The success principle behind companies that hire members capable of independent efficient work to fit a team (Collins, 2001) translates as the utility and importance analogue of how sensor agents are expected to behave in a robot. The utility is decided by how confident the sensor can act on its own and how willing the sensor is towards co-operating for the success of the robot. These analogies establish an interesting avenue of thought to introduce a statistical perspective for success in fusion. We see the potential of an information theoretic approach, implementing intelligent adaptive behavior by associating a measure to quantify how much information a sensor provides about the state and another measure quantifying the information gain or synergy when the same sensor is working in a co-operative team with other sensors. Inspired by those thoughts, we derive mathematical formulae for evaluating the self confidence of a sensor about its state measurement and co-operative confidence the team associates with the sensor.

4.2 Our proposed approach

We have thus far discussed the philosophical inspiration to proceed towards implementing the inspiration as statistical inference for our mobile mapping applications. The key to achieving adaptive behavior using statistical data fusion is the representation of uncertainty towards extracting information from multiple sources. A number of different paradigms using fuzzy logic, Bayesian reasoning, Dempster-Shafer theory and transferable belief models have been developed for modeling and handling uncertainty. We note that most if not all of these theories assume uncertainty models and representations of belief similar to the content presented in Chapter 3. We would like to briefly discuss and inspire the derivation of our approach from some of those ideas. The idea to use fuzzy logic approach represents belief by defining a mapping between state variables of interest and belief functions (Zadeh, 1975).

Bayesian reasoning methods on the other hand, (Elfes, 1992; Mitchell 2007) assign probability mass functions to mutually exclusive hypotheses as belief distributions and propagate posteriors by incorporating apriori with evidence. The Dempster-Shafer theory (Zhang, 1994) which can be considered a generalization of Bayesian reasoning combines different sources of information and evidence by calculating the probability with upper and lower bounds on the variables of interest. The transferable model theory is a further generalization on the Dempster-Shafer theory by assigning probabilities to a power set of mutually exclusive hypotheses over the existing assignments of probability using the Dempster-Shafer theory.

There are several arguments in the literature with regards to the application of these theories in real world situations (Maskell, 2008). Bayesian statisticians consider Bayes approach as the only consistent reasoning methodology to manipulate belief whereas advocates of evidence reasoning (French, 1980; Zhang, 1994) vehemently point out the inability of the Bayesian approach to satisfactorily manipulate belief amidst uncertain, imprecise and conflicting data. Our approach leverages the Bayesian approach for uncertainty propagation and power set evidence reasoning as the statistical tool for implementing adaptive and stable behavior in unknown unstructured dynamic situations.

Let us consider a generic N -sensor system ($N > 2$) with several sensors S_i providing state x_t (position and orientation estimates in our case) of d -state dimensions at time t . By feeding in apriori noise model $p_i(\hat{x}_t | x_t)$ associated with the pose measurement using sensor S_i to a Bayes filter (Extended Kalman filter in our case), we are able to associate a belief distribution B_t^i , and hence a likely pose μ_t^i and the uncertainty Σ_t^i about that pose for each sensor S_i . Each one of these estimates μ_t^i contributes to the most likely pose of the robot with the Σ_t^i quantifying the doubt in that state. If not using the Kalman filter, both μ_t^i and Σ_t^i can be computed as the first two moments of the belief probability density function B_t^i . We estimate the belief distributions using the procedure explained towards the concluding section of Chapter 4.

We have N distributions one for each sensor at regular instants of time as our input to learn and implement intelligent adaptive behavior. Among those N distributions, the sensor conflict and performance degradation appear in two forms illustrated as a simple example in Figure 4.1, the conflict in the maximum likelihood state that we refer to as mean conflict and the conflict in confidence about the measured state that we call the covariance conflict. The Gaussian curves in Figure 4.1 are belief distributions of the state as measured by each sensor in a simple one-dimensional case. When one sensor is operating with high belief but not in agreement with other sensors on the system, we need to decide which sensor(s) to believe before executing the next motion command. How do we decide which sensor or set of sensors are believable? When should we fuse multi-sensor data and when should we perform sensor selection? How do we adapt and learn if the sensor is springing back to normal operation? Towards the goal of answering such questions, we formulate our objective function as shown below.

$$\text{Given } \mathbf{S} = \{ S_i \}, i = 1, 2, \dots, N \text{ where } S_i(x) = \hat{x}_i$$

$$\text{Find } \arg \max_{\tilde{x}, f(S_i, \{B_i\})} \text{Belief}(\tilde{x}) \quad (4.1)$$

where f is the fusion function (like a covariance weighted summation), B_i is the belief about sensor measurements of state $x = \hat{x}_i$ from sensor S_i and \tilde{x} is the best estimate from sensor feedback after multi-sensor fusion in the N -sensor system.

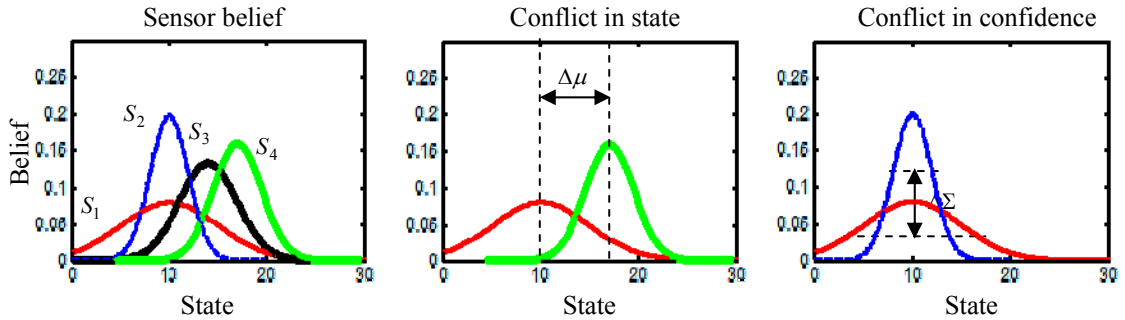


Figure 4.1: Several sensors measure the state of our system with an uncertain belief. Using the middle and the right figures, we are able to visualize sensor conflict manifesting in two forms as the separability of the mean in the belief distributions and as the separability based on the variance. While trying to fuse sensor data based on sensor belief distribution during the operation of an unmanned vehicle, we would like to derive an algorithm that will guide us towards the choice of sensors that when integrated leads us as close and confident as possible to the actual state.

The functional f could be as simple as a weighted average of maximum likelihood state measurements or fusion using advanced probabilistic methods. In our implementation, we will use the weighted average as the fusion function because of the linearity in fusion and the theoretical strength that guarantees performance better than the best sensor (Rao, 2001). With that problem statement, we now derive the framework from first principles restating Equation 4.1 in words as: “Success in fusing state information using several sensors can be achieved by maximizing the information about the state variables of interest using data from each sensor individually and also maximizing the information gain from fusing data from different sensors while penalizing for the information corruption.”

Success in fusion = Maximize(Information about state measurement from sensors individually
+ Information gain when sensor data is fused – Information corruption during the fusion)

Mathematically, we evaluate the potential of each sensor during online operation by modeling success as shown in Equation 4.2.

$$\text{Sensor Success} = \text{Belief}(\tilde{x} | S_i(x)) + \text{Belief}(S_i(x) | S_j(x))_{i \neq j, j=1,2,\dots,N} \quad (4.2)$$

With that formulation of sensor success, success in multi-sensor localization follows as Equation 4.3.

$$\text{Objective for success in fusion towards localization} = \max(\text{Sensor Success}) \quad (4.3)$$

We note that both $\text{Belief}(\tilde{x} | S_i(\hat{x}_i))$ and $\text{Belief}(S_i(\hat{x}_i) | S_j(\hat{x}_j))_{i \neq j, j=1,2,\dots,N}$ are conditional probability distributions

that have already accounted for the bounded noise characteristics of sensors. Our approach reduces Equation 4.2 and 4.3 whose components are random variables and probability distributions to numerical scores. We achieve that by defining an uncertainty term U_i for each sensor S_i such that this term is a function of the measurement uncertainty or self-confidence M_i and the sensor validity or co-operative confidence V_i . We explain the derivation and implementation specification for M_i and V_i in the later sections of this chapter.

$$\begin{aligned} U_i = M_i + V_i &= \text{Measurement uncertainty} + \text{Validity uncertainty} \\ &= -(\text{Self-confidence} + \text{Co-operative confidence}) \end{aligned} \quad (4.4)$$

With such a formulation of U_i the hope is to select the most reliable sensor(s) as the one with the minimal uncertainty

$$S_m \text{ s.t. } U_m = \min\{U_1, U_2, \dots, U_N\} \quad (4.5)$$

With those notations, we introduce our method as a block diagram in Figure 4.2. Our approach evaluates the confidence of each sensor (self confidence) along with the reliability of sensors (altruist confidence or co-operative confidence) based on statistical clustering. The self-confidence is a Bayesian inspired measure and the reliability is a power set reasoning measure. The method we use for the statistical clustering and the confidence estimation is derived based on information theory helping us make the fusion decision physically meaningful unlike state of the art methods where reliability and confidence are treated as different entities. The method that we propose for computing scores M_i and V_i are the result of statistical insight on uncertainty from methods in the literature. Before we get into the specifics of implementation; we discuss the insight for modeling information corruption and information gain and perceive the similarities of the inference required in the fusion problem with the statistical notion of information complexity.

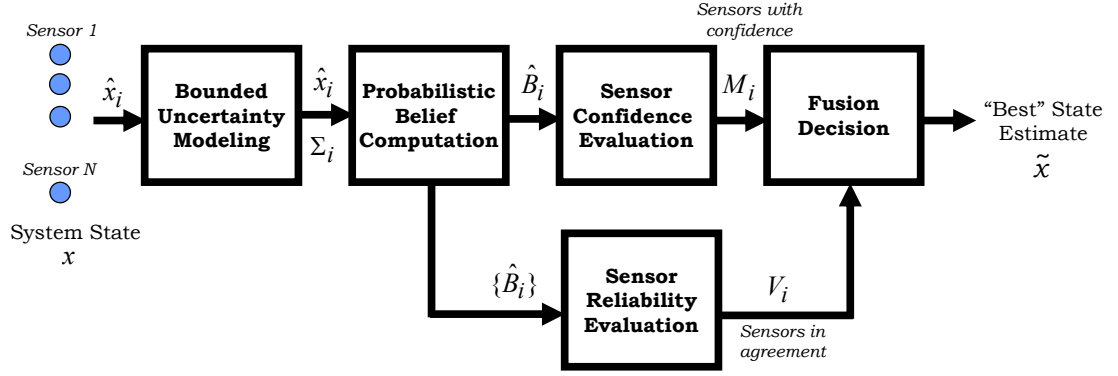


Figure 4.2: Block diagram of the proposed selfish-altruist fusion algorithm. We presented the models of uncertainty modeling along with the idea of belief propagation using Kalman filters in Chapter 3. We discuss the procedure for sensor confidence evaluation in Section 4.3, sensor reliability evaluation in Section 4.4 and the criterion resolving fusion versus selection in Section 4.5.

We borrow definitions of information gain and information corruption while integrating data from several sensors from (Gardner and Uhlmann, 2005). We illustrate their covariance intersection approach as a tool to model information gain and the covariance union definition as a tool to approximate the information corruption in Figure 4.3. Figure 4.3 is actually an extension of Figure 4.1 to two dimensional state with three sensors. We have illustrated typical situations that arise during state estimation from several sensors in this figure.

The ellipses in Figure 4.3 are the 2Σ contours of the belief distributions of sensors S_1 , S_2 and S_3 . We show three situations that are very typical of unmanned vehicle localization (1) sensors are all in agreement operating at similar levels of uncertainty (2) Sensors are in conflict on the state at same level of confidence and (3) when sensors are in conflict with different levels of confidence on the state. In real world situations, Case 1 can be interpreted as the ideal case, Case 2 the worst case and Case 3 the erratic behavior like multi-path scenarios with GPS.

The filled ellipses in Figure 4.3 (b) are the result of covariance intersection rule used as an optimistic state integration method for propagating belief by integrating the sensor state from several sensors. The rule is best summarized using the Equations 4.6 and 4.7.

$$\Sigma_f^{CI} = (w_1 \Sigma_1^{-1} + w_2 \Sigma_2^{-1} + \dots + w_N \Sigma_N^{-1})^{-1}. \quad (4.6)$$

$$\mu_f^{CI} = \Sigma_f^{CI} (w_1 \Sigma_1^{-1} \mu_1 + w_2 \Sigma_2^{-1} \mu_2 + \dots + w_N \Sigma_N^{-1} \mu_N) \quad (4.7)$$

where w_1, w_2, \dots, w_N are weights that can be associated with sensors individually for the fusion. If all sensors are to be treated equal $w_i = \frac{1}{N}$. The estimate of the fused state μ_f^{CI} and the uncertainty about that state

Σ_f^{CI} using this approach is an approximation to the optimistic estimate of the gain possible given the data and their uncertainty in the form of ellipses. We also note that this is the fusion scheme f we want to implement in Equation 4.1. This scheme is optimal as long as sensors do not violate apriori bounds.

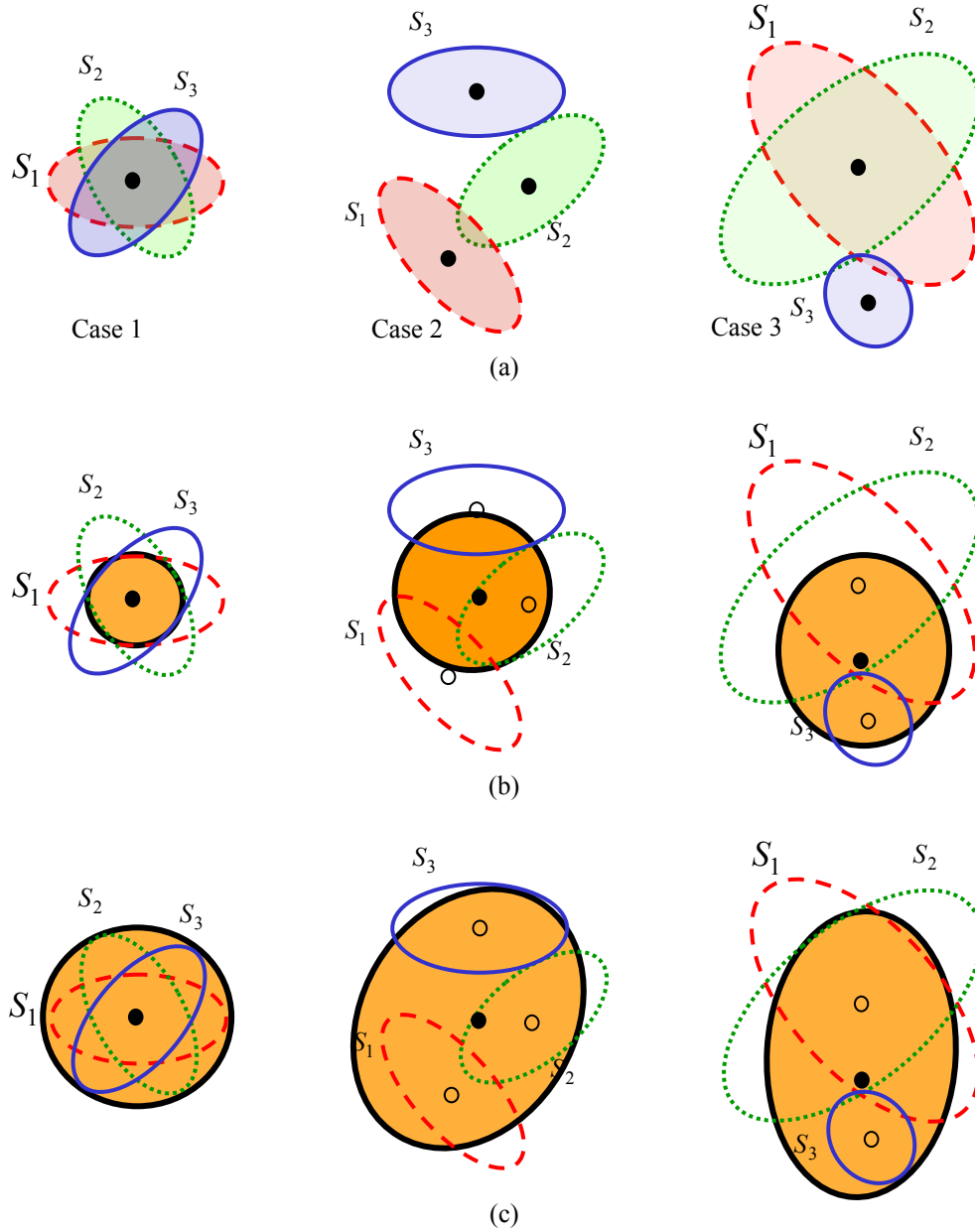


Figure 4.3: Propagation of uncertainty in multi-sensor belief propagation systems. (a) Sensor situations on a two dimensional state illustrated as ellipses. (b) Multi-sensor state fusion using the covariance intersection rule. (c) Multi-sensor state fusion using the covariance union rule. Both the rules suggested by (Gardner and Uhlmann, 2005) are typical decentralized data fusion tools for propagating belief on state information from several sensors. We are able to see that the intersection rule is an optimistic estimation of state while the union rule appears to be a worst case pessimistic estimate.

But, for our applications, we have to allow sensor failure and erratic behavior with performance degradation. For such situations, covariance union is used to estimate the worst case corruption estimate. The ellipses in Figure 4.3 (c) illustrate this notion of information corruption. The filled ellipse even when sensors agreed on the state provides a worst case of belief to be propagated. The equations used to estimate the covariance union state and uncertainty are presented below. We note that initialization of the fused state with covariance union uses the covariance intersection estimate presented earlier.

$$\Sigma_f^{CU} \geq \Sigma_i + (\mu_f^{CI} - \mu_i)(\mu_f^{CI} - \mu_i)^T; i = 1, 2, 3, \dots, N \quad (4.8)$$

$$\mu_f^{CU} = \Sigma_f^{CU} (w_1 \Sigma_1^{-1} \mu_1 + w_2 \Sigma_1^{-1} \mu_2 + \dots + w_N \Sigma_1^{-1} \mu_N) \quad (4.9)$$

The inequality is resolved using an optimized measure of size such as the determinant or trace of the estimated covariance in Equation 4.8. We are able to see the contrast between the union and intersection operations in Equation 4.6 and 4.8 in Figure 4.4. The focus of the intersection operation in Equation 4.6 was on reducing the uncertainty and improving the information gain when sensors are operating at different covariance but converging and consistent on the state. The covariance union approach on the other hand tries to accommodate for the mean conflict and estimates the smallest covariance that is large enough to guarantee consistency regardless of which of the sensors is close to the truth. Similar to these scores, we would also like to estimate a state with unimodal belief (Denzler and Brown, 2002) and with as less variance as possible. The path that we are interested in taking is that of eliminating the sensors that do not contribute to information gain through unimodality.

Several implementations of the idea of covariance intersection and union in belief propagation systems (Collins and Nicholson, 1998; Uhlmann, 2003) have proved successful. However, for efficient belief propagation amidst sensor conflict with our systems, we would ideally want to choose covariance intersection in Case 1 and covariance union for Cases 2 and 3. But such a goal requires the resolution of the mean conflict and covariance conflict from the belief distributions. In Figure 4.4, we have traced the 2Σ probability space with individual sensor support and the probability space of the fused estimate using the intersection rule. We would like to use this figure to establish the basis of our approach. We have separated the data from sensors and perceive the fusion problem as an information complexity problem. By quantifying the size or area of the ellipses, taken one sensor at a time, we see that we are able to identify the sensor confidence conflict. By looking at the power set of combinations of sensors, we are able to visualize the increase in complexity of those the 2Σ curves indicating state conflict in the presence of covariance conflict. We try to use these cues to quantify the sensor self confidence and co-operative confidence.

For quantifying measurement uncertainty as a measure of self confidence we score the Bayesian belief probability density using a model selection criterion. We use model selection theory that guides the choice of competing distribution models for given data towards quantifying the confidence among competing sensor belief distributions in our case. For sensor validity or the co-operative confidence, we evaluate belief on pose estimates from different sensors as a multi-sample clustering problem. Multi-sample clustering procedure provides the inference to the question, how many of the competing distributions are in agreement and how many are not? If they do differ, do they differ in the state or because of the conflict in confidence? We will be using the multi-sample clustering problem as a means to resolve the mean-variance conflict by statistically differentiating and grouping belief distributions. The minimization of the combined uncertainty (measurement uncertainty score + sensor validity score) allows us to intelligently choose a subset of sensors that contribute to accurate state estimation. We explain the implementation and theoretical specifications in the following paragraphs. We present the score of self-confidence quantifying the uncertainty in each sensor first.

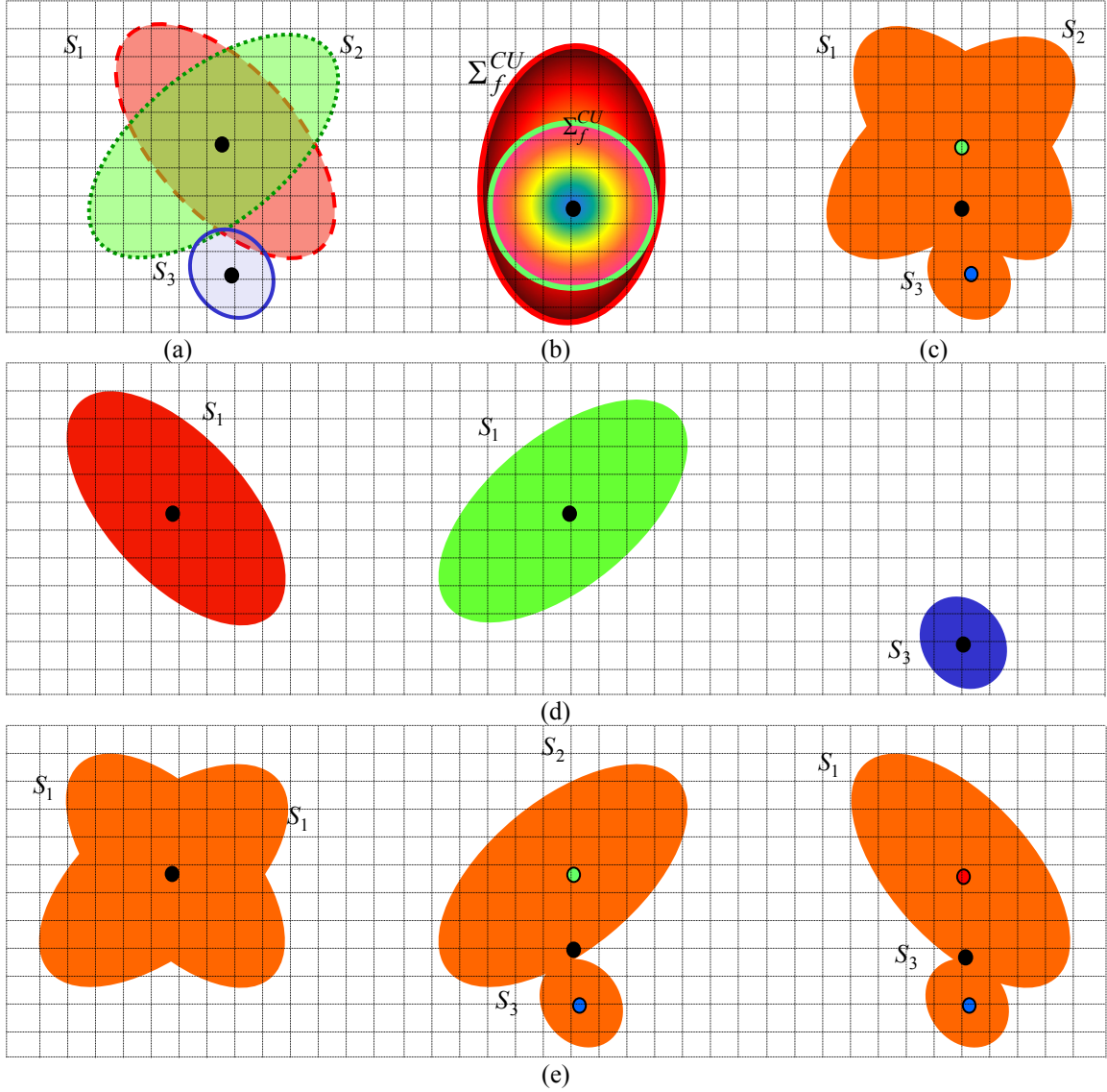


Figure 4.4: Perceiving fusion as a model selection problem. (a) We take the Case 3 example from Figure 4.3 to elaborate. (b) The uncertainty ellipses indicate the probability space after integration using covariance intersection and union methods. (c) Probability state space spanned with 2Σ confidence at least from one sensor. (d) Uncertainty from each sensor taken individually. (e) Uncertainty when set of sensors are considered. The notion that we have illustrated here is that uncertainty propagates in different forms as the area of the ellipses, the shift in fused estimate, the belief strength and the shape of the contour. The shape of the contour becomes more and more complex with state and confidence conflict. Our approach will try to learn and infer the conflict by considering the sensor confidence individually and in combination with one another.

4.3 Quantifying sensor self confidence

Based on the belief estimates alone, if we were to choose the best sensor in the system, we would pick the sensor that is indicative of maximum likelihood with minimum uncertainty. Hence, when sensors with measurements are indicating belief B_i^i within in an update cycle of the Bayes filter, the confidence can be measured as the information distance (Kullback, and Leibler, 1951), between the expected belief distribution and the estimated belief distribution as shown in Equation 4.10 where \hat{B}_i is the estimated belief based on state estimates \hat{x}_i and the actual distribution B_i^* on state x . G in Equation 4.10 is the belief representation model (Gaussian in the case of Kalman filter).

$$KL(B_i^*; \hat{B}_i) = \int G(\hat{x}; B_i^*) \left[\log \frac{G(x; B_i^*)}{G(\hat{x}; \hat{B}_i)} \right] d\hat{x} . \quad (4.10)$$

Equation 4.10 in the discrete finite sampling implementation form transforms to Equation 4.11.

$$KL(B_i^*; \hat{B}_i) = \sum_{j=1}^c G(y_j; B_i^*) \left[\log G(y_j; B_i^*) \right] dy_i - \sum_{j=1}^c G(y_j; B_i^*) \left[\log G(y_j; \hat{B}_i) \right] . \quad (4.11)$$

where c is the sampling used for the belief estimation or can be the measurements for every cycle of the Kalman filter. y_j refers to the actual measurements or the sampling from the belief distribution depending on the application and the sensor. We note in scoring the confidence of a belief estimate our interest by minimizing the information distance is only in the second term (the first term being the neg-entropy independent of the variable y) of Equation 4.11 that on simplification becomes

$$- \sum_{j=1}^c G(y_j; B_i^*) \left[\log G(y_j; \hat{B}_i) \right] = -E(\log G(y_j; \hat{B}_i)) . \quad (4.12)$$

where E is the expectation operator.

Now let us consider a simple 1D case, where the true belief of a particular state is $B^* \sim No(\mu^*, \sigma^{*2})$ and the belief from Sensor #1 $\hat{B}_1 \sim No(\mu_1, \sigma_1^2)$ is calculated using c_1 samples in the update cycle in the belief propagation, $\hat{B}_2 \sim No(\mu_2, \sigma_2^2)$ be the belief calculated for Sensor #2 using c_2 samples, we now have to choose the sensor that is the closest to the actual underlying belief. This can be achieved by computing the Kullback-Liebler (KL) distance in Equation 4.13.

$$KL(\hat{B}_1, B^*) = \int_{-\infty}^{\infty} B^*(x) \ln B^*(x) dx - \int_{-\infty}^{\infty} B^*(x) \ln \hat{B}_1(x) dx . \quad (4.13)$$

For the simple 1D case, when B is Gaussian, Equation 4.13 reduces to the equation below.

$$= \frac{1}{2} \ln \left(\frac{\sigma^2}{\sigma^{*2}} \right) - \frac{1}{2} + \frac{1}{2} \left[\frac{\sigma^{*2}}{\sigma^2} + \frac{(\mu^* - \mu)^2}{\sigma^2} \right] . \quad (4.14)$$

For example, if $B^* \sim \text{Normal}(0,1)$, $\hat{B}_1 \sim \text{Normal}(0.5,1)$ and $\hat{B}_2 \sim \text{Normal}(0,1.5)$, both the sensors with state (mean) conflict and variance conflict, we are able to calculate $KL(\hat{B}_1, B^*) = 0.125 < KL(\hat{B}_2, B^*) = 0.036$ and infer that the sensor with belief \hat{B}_2 is more confident in itself than sensor \hat{B}_1 . We extend this idea to multi-dimensional state vector to quantify self confidence.

Towards that goal of trying to extend Equation 4.13 to include multi-dimensional state vector in the finite sampling discrete form, we realize that Equation 4.12 is the objective function for several statistical model selection criteria (Burnham and Anderson, 2002) such as the Akaike information criterion (AIC), Information complexity (ICOMP) and the Bayesian Information criterion (BIC) etc. summarized in Table 4.1. L in Table 4.1 refers to the likelihood function, k the dimensionality of the parameter vector $\hat{\theta}$, c is the number of samples used to estimate the parameters of belief B , F^{-1} the inverse Fisher information matrix and C_I is the covariance complexity.

Each one of these criteria fall into a general form for modeling risk as shown in Equation 4.15.

$$\text{Overall Risk} = \text{Risk in modeling} + \text{Risk in estimation.} \quad (4.15)$$

Risk in modeling refers to the choice of the model B and the risk in estimation refers to limiting ourselves in the number of computable parameters for the model B using maximum likelihood estimation. For example, if B was Gaussian, we have two parameters for the model representing the belief, and if we were using multi-hypothesis tracking for belief propagation, we will have several Gaussian mixture parameters that contribute to evaluating more parameters for better accuracy. We studied the derivation of several methods such as AIC (Akaike, 1974), minimum description length (Rissanen, 1978), BIC (Schwarz, 1978), CAIC (Bozdogan, 1987) and its more recent form ICOMP (Bozdogan, 2000) to minimize Equation 4.12.

We observe that these criteria have a form as shown below.

$$\text{Model selection criteria} = \text{Lack of fit} + \text{Lack of parameter parsimony.} \quad (4.16)$$

$$= -\log(\text{Likelihood of model } G) + \text{Parameter penalty term} \quad (4.17)$$

The lack of fit part quantifies the strength about the state (mean) from the belief distributions, while the parameter penalty term takes care of the shape of the 2Σ curve. The more complicated the shape, the higher the value on the parameter penalty term.

Table 4.1: Model selection criteria derived out of minimizing Equation 4.11.

Name	Formula
AIC (Akaike, 1974)	$-2 \log L(\hat{\theta}; B) + 2k$
MDL (Rissanen, 1978)	$-2 \log L(\hat{\theta}; B) + \frac{1}{2} k \log c$
BIC (Schwarz, 1978)	$-2 \log L(\hat{\theta}; B) + 2k \log c$
CAIC (Bozdogan, 1987)	$-2 \log L(\hat{\theta}; B) + k(\log c + 1)$
ICOMP (Bozdogan, 2000)	$-2 \log L(\hat{\theta}; B) + 2C_I(F^{-1}(\hat{\theta}))$

Following Bozdogan (2000), in a generic derivation without assuming anything about the parametric nature of the model G used in the estimation of B , the criterion we will use is ICOMP. ICOMP in its finite element derivation includes sensor bias, in addition to hinting us about the number of parameters by penalizing based on the curvature of the log likelihood function. We are particularly interested in this definition based on the Fisher information matrix because, in our case the parameter that we would like to penalize is the covariance indicating uncertainty and not the number of parameters as in AIC and its variants. Hence, in extending Equation 4.14 to a multi-dimensional state vector, we will use Equation 4.18.

$$ICOMP(IFIM) = -2 \log L(\hat{\theta}; B) + 2C_1(F^{-1}(\hat{\theta})) \quad (4.18)$$

We reformulate the derivation to define our measurement uncertainty term. Model selection criteria AIC, ICOMP, BIC in the statistics literature are used to select a model based on the goodness of fit and with minimal parameter complexity, minimizing KL distance between the estimated samples and the underlying probability density. Quantifying the uncertainty in robot self localization from competing belief distributions is also an analogous model selection problem, that ICOMP can be modified and applied. We hence derive the measurement uncertainty M_i as the score of confidence or the figure of merit on the belief distribution of each sensor as shown in Equation 4.19. Lower the magnitude of M_i , implies lesser the measurement uncertainty and greater self-confidence of the sensor.

$$M_i = -2 \log L(\mu_i^t; B_i) + 2C_1(F^{-1}(\Sigma_i^t)) \quad (4.19)$$

where F^{-1} is the inverse Fisher information matrix and,

$$C_1(F^{-1}(\Sigma_i^t)) = \frac{s}{2} \log \left[\frac{\text{tr}(F^{-1}(\Sigma_i^t))}{s} \right] - \frac{1}{2} \log |F^{-1}(\Sigma_i^t)| \quad (4.20)$$

with s being the rank of F^{-1} , $|\cdot|$ refers to the determinant and tr refers to the trace of the matrix. F^{-1} is computed as

$$F^{-1}(\Sigma_i^t) = \begin{bmatrix} \Sigma_i^t & 0 \\ 0 & \frac{2}{c_i} D_p^+(\Sigma_i^t \otimes \Sigma_i^t) D_p^+ \end{bmatrix} \quad (4.21)$$

with D_p^+ being the Moore-Penrose inverse of the duplication matrix D_p , \otimes represents the Kronecker product and c_i the number of samples used to compute the covariance uncertainty of S_i . D_p is a unique $d^2 \times \frac{d}{2}(d+1)$ matrix which will transform $v(\Sigma_i^t)$ into $\text{vec}(\Sigma_i^t)$. $\text{vec}(\Sigma_i^t)$ denotes the vectorization operator

that transforms a matrix into a vector by stacking columns, $v(\Sigma_i^t)$ denotes the $\frac{d}{2}(d+1) \times 1$ vector obtained

by eliminating the supra-diagonal elements of Σ_i^t . F^{-1} gives the lower bound for the covariance matrix of the parameter estimates and measures the accuracy of the maximum likelihood estimators μ_i^t and Σ_i^t . The C_1 measure for penalizing uncertainty is obtained by maximizing mutual information in d -dimensions. The derivation behind Equations 4.20 and 4.21 can be obtained from Bozdogan (2000) and Van Emden (1971).

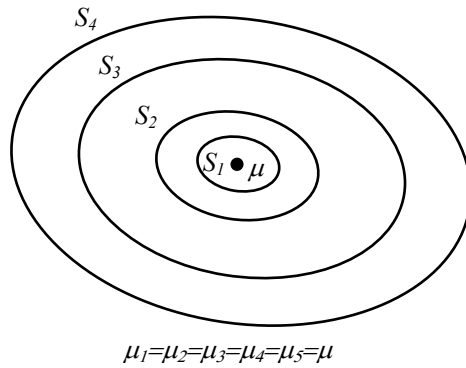
Though these equations appear complex, a distributional assumption, such as a Gaussian, reduces Equation 4.19 to a much simpler finite sampling distributional form as shown in Equation 4.22. We do note that Equation 4.19 does not make assumptions on the functional form of the density and can be used on belief estimates from multi-hypothesis trackers or even particle filters.

$$M_i = 2\left[\frac{c_i d}{2} \log(2\pi) + \frac{c_i}{2} \ln |\Sigma_i^t| + \frac{1}{2} \text{tr}(\Sigma_i^{t-1}) \left(\sum_{j=1}^{c_i} (y_j - \mu_i^t)(y_j - \mu_i^t)' \right) + 2\left[\frac{d}{2} \ln\left(\frac{\text{tr}(\Sigma_i^t)}{d}\right) - \frac{1}{2} \ln |\Sigma_i^t| \right] \right] \quad (4.22)$$

where d is the dimensionality of the state vector and y_j are the c_i measurements of sensor S_i used for estimating belief.

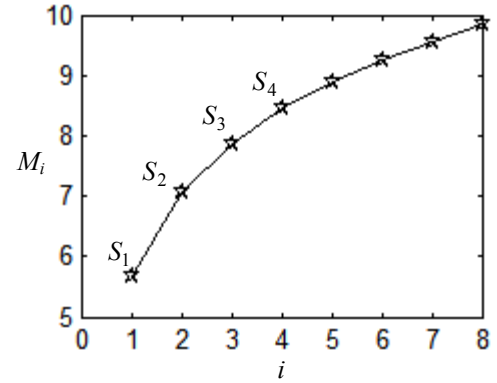
We illustrate the function of the measurement uncertainty score in Figure 4.5. We consider covariance conflict alone and assume that there is no conflict in the state estimate. Further, we have assigned each sensor's S_i covariance equal to $i\Sigma$. The reason behind this study is to prove that M_i can identify sensors operating at different levels and help eliminate them from the belief propagation. A typical real world example is the GPS outage or attenuation, when the sensor operating at 3 cm bound suddenly loses signal strength to operate at 3 m uncertainty. We propose to use the measurement uncertainty score to differentiate sensors operating in a desired bound from the ones that are not even though they may be in agreement with the state estimate from other sensors. Referring back to the covariance intersection and union rules, elimination of such sensors is significant because, they will unnecessarily increase the size of the uncertainty ellipse during belief propagation.

We note that we do not require that all sensors provide the same number of samples for one propagation cycle. For example, suppose GPS sensors operate at 10 Hz and the vision at 20 Hz and a motion command is executed every second, 10 samples from GPS can be used in the belief propagation for the hardware sensor and 20 samples for the video in computing and representing the belief distribution. The other alternative is to adjust the prediction and correction stages within the Bayes filter based on the number of motion commands executed for every sensor update and estimate the parameters of the belief distribution. The samples of the underlying distribution can then be generated using pseudo random generators for the inference. We suggest and recommend the generation of equal number of samples from the belief distributions for consistency in the inference. Though the criterion we use is a bias correcting criterion and can include sensor bias in its computation, using a fixed number of samples for the inference provides better results.



$$\Sigma_1 = \Sigma; \Sigma_2 = 2\Sigma; \Sigma_3 = 3\Sigma; \dots; \Sigma_N = N\Sigma$$

(a)



(b)

Figure 4.5: Understanding the measurement uncertainty score using a simple example of covariance conflict with several sensors. (a) Each sensor S_i is measuring the state μ with different covariance. Particularly, we force the covariance for sensor S_i to be of i times Σ and observe the measurement uncertainty score M_i . (b) The plot of M_i for each sensor S_i . The lesser the value of M_i the greater the self confidence of the sensor. We are able to show that the score M_i can be used to identify covariance conflict.

4.4 Quantifying co-operative confidence

Our idea to approximate sensor validity as the cooperative confidence is based on the argument that the best that we can learn from conflicting multi-sensor measurements is by grouping sensors that work towards confident state estimation. The more sensors that provide the same information, the higher the sensor validity we can attach to sensor's in agreement and its measurements. Though the logical argument sounds very easy, converting the argument into a mathematical form involves more work. The problem in our hand is that of comparative inference that has to resolve and quantify the conflict in the state and the covariance not only for sensors but all possible combinations of sensors. In the statistics literature, the commonly used technique is the analysis of variance (ANOVA) along with the f -statistics for univariate distributions. The multivariate extension of ANOVA, MANOVA uses the Wilks criterion in a multiple comparison procedure appears to be a potential solution. But there are several issues; the multiple comparison procedure using MANOVA (Montgomery, 1991) is a pair wise comparison procedure requiring apriori significance level determination. Saiki et al. (2008) have tried to use similar ideas with chi-squared statistics. But for our systems, we desire that the significance level be adaptive to the sensor performance.

We need a method that can parsimoniously group these distributions associated with the robots pose from different sensors to quantify a measure that indicates optimal clustering in the probability space. For the N -sensor system, this unwinds to a computationally prohibitive hypothesis testing problem that requires a fuzzy estimator as demonstrated in (Kobayashi et al., 1999). Our approach stays away from fuzzy logic and is inspired by methods described in the survey in (Rogova and Nimier, 2004) and information theoretic methods in (Denzler and Brown, 2002; Thrun et al., 2004) to formulate sensor validity in a novel information theoretic sense.

Recollecting from Figure 4.4, we note the lack of co-operation gets exposed in the fused distributions as increased area of the probability space of interest and increased complexity of the shape of 2Σ contour. We illustrate a similar example in Figure 4.4 but show the fused distribution instead of the contour in Figure 4.6. Our idea is to estimate the information gain after fusion and compute the information corruption as the shape deformation from unimodality and combine them as a score of co-operative confidence. The hope is that by doing so, we will be able to identify the conflict among sensors and easily infer that sensors 2 and 3 converge on the state estimate and hence can work together as a team for better state localization.

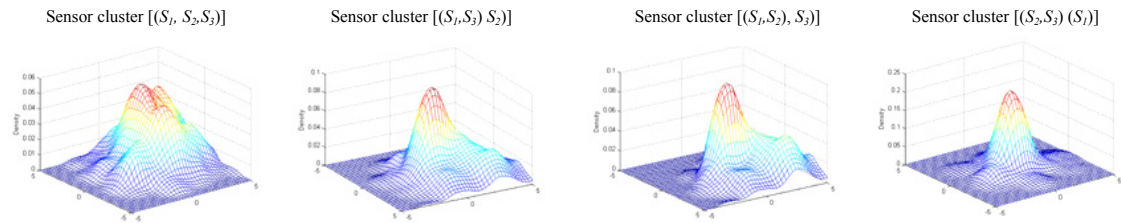


Figure 4.6: Understanding sensor validity estimation. We have shown the belief distributions when two sensors' data are combined at one time to estimate a new belief. We see that when disparate sensors are integrated, several modes appear in the distribution indicating uncertainty, increasing the number of parameters required to represent the distribution while sensors in agreement increase the confidence in estimated state. Fusing sensors not co-operating with another not only spans a larger probability space but also has several modes.

The problem of quantifying validity as co-operative confidence can be coined as follows: Given N sensors, and each of their belief distributions B_i and the fusion functional f on these belief distributions, we need to infer how many sensors boost the confidence of a particular state. We perceive the problem in our hands as a multi-sample clustering problem. We treat the belief distributions from different sensors as datasets from different sources and try to group sensors with similar localization result. For example, if c_1 samples of Sensor 1 and c_2 samples from Sensor 2 are considered for construction of B_1 and B_2 , we try to measure the information distance in probability space by considering c_1+c_2 samples to construct another belief distribution B_{12} for localization. We measure the information distance between B_{12} and B_1 / B_{12} and B_2 as the significant new information gained after fusing data from both the sensors. That is co-operative confidence V_1 of sensor S_1 will be the information distance between B_{12} and B_1 and V_2 will be information distance between B_{12} and B_2 . Extending this simple example to an N -sensor system, we are able to understand that the information gain after evaluating all the combinatorial clusters of sensors will lead us to the maximal group of sensors that essentially converge on the state vector. We will use information gain as the yardstick to quantify co-operation.

We get into the implementation specifics of our idea. Since the problem we have in our hands is that of classifying sensors into homogenous groups, let us begin the procedure by constructing a validity distribution \hat{V} , whose samples are the elements of matrix V . The validity matrix V is an inspiration from idea previously demonstrated in (Luo and Lin, 1988). As shown below in Equation 4.23, V includes samples $Y_i = \{y_i^j, j = 1, 2, 3, \dots, c_i\}$ per update cycle from all sensors S_i . Note that Y_i is a c_i by d sensor measurement matrix, d being the dimensionality of the state vector and y_i^j is the d by 1 state measurement.

$$V = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} \begin{matrix} c_1 \times d \\ c_2 \times d \\ \\ c_N \times d \end{matrix} \quad (4.23)$$

The goal of clustering is then to classify the N sensors (as samples used in the distribution B_i) into k - groups, where k is unknown and variable. The procedure to infer about the minimal k and hence the maximal group of homogenous sensors should evaluate all possible clustering alternatives of sensor combinations and further determine the optimal partitioning of the N sensors into k groups. One among the k -groups in the optimal partition will have the maximum sensor cardinality κ indicating homogenous sensors. Our approach will make use of an information criterion to perform the clustering, find the κ homogenous sensors and finally associate a validity score to them. The motivation being, if we can opportunistically maximize the information gain while simultaneously penalizing the information corruption during the fusing process, fusion of data from one of the sensors clusters will perform better than each sensor and possible sensor teams. Such an inference using the criterion will also indicate that the κ homogenous sensors will work best as a team. In Table 4.2 below, we have shown a simple example using a 3 sensor system and explain the procedure in the following paragraphs.

Table 4.2: Co-operative team of sensor clusters in the validity inference for a 3-sensor system.

k – How many groups?	κ – Maximum sensors in group	Sensor clusters ($N=3$)
3	1	$[(S_1) (S_2) (S_3)]$
2	2	$[(S_1) (S_2, S_3)] [(S_1, S_2) (S_3)] [(S_1, S_3) (S_2)]$
1	3	$[(S_1, S_2, S_3)]$

Matrix V is a mixture of N distributions, that can be partitioned into g ($g \leq N$), $X_g(c_g \times d)$ such that $c_{total} = c_1 + c_2 + \dots + c_N = \sum_{g=1}^N c_g$. Our goal then is to exhaustively search for all possible sensor clusters to determine the optimal k and κ .

At this point, we recollect that the validity matrix V has samples of $p(f(S_i, B_i))$, the belief probability distribution after fusion that we are trying to represent using minimal number of parameters Θ . Then, essentially by evaluating the clustering alternatives, we are seeking $\Theta = [\Theta_1, \Theta_2, \dots, \Theta_k]$ for V such that X_g is completely described by parameters Θ_g when k is optimal and each sensor in the group X_g with maximum number of κ sensors (highest cardinality) are homogenous. For example, in the 3 sensor case if sensor cluster $[(S_1, S_2, S_3)]$ is evaluated, to be optimal, k will be equal to 1, X_g will equal V and will require $\Theta = [\mu_g, \Sigma_g]$, with only 2 parameters if normality can be assumed. However, for the cluster $[(S_1) (S_2) (S_3)]$, we will have $X_g = [X_1, X_2, X_3]$, $X_1 = Y_1$, $X_2 = Y_2$, $X_3 = Y_3$ and require $\Theta = [\mu_1, \Sigma_1, \mu_2, \Sigma_2, \mu_3, \Sigma_3]$ with 6 parameters. For the cluster $[(S_1), (S_2, S_3)]$; $X_1 = Y_1$, but $X_2 = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ with dimensionality $c_1 + c_2$ by d , requiring $\Theta = [\mu_1, \Sigma_1, \mu_{23}, \Sigma_{23}]$ with 4 parameters.

Generalizing the 3-sensor example to the N -sensor case, the number of ways of clustering N sensors into k -cluster groups where $k \leq N$ such that none of the k -clusters are an empty set is given by Equation 4.24. The quantity w is also called the Stirling's number of the second kind (Duran and Odell, 1974).

$$w = \frac{1}{k!} \sum_{g=0}^k (-1)^g \binom{k}{g} (k-g)^N \quad (4.24)$$

Equation 4.24 generates w clustering alternatives that can be classified into one of the following three cases as listed in Table 4.3.

Table 4.3: Possible hypotheses to evaluate for sensor reliability.

Case	Hypothesis : Parameter vector form	No. of parameters
1. Sensors in agreement	$\Theta = \left[\underbrace{\mu, \mu, \mu, \dots, \mu}_{N \text{ times}}, \underbrace{\Sigma, \Sigma, \Sigma, \dots, \Sigma}_{N \text{ times}} \right]$	$d + \frac{d(d+1)}{2}$
2. Minimal belief	$\Theta = \left[\mu_1, \mu_2, \mu_3, \dots, \mu_N, \underbrace{\Sigma, \Sigma, \Sigma, \dots, \Sigma}_{N \text{ times}} \right]$	$Nd + \frac{d(d+1)}{2}$
4. Sensors confused	$\Theta = [\mu_1, \mu_2, \mu_3, \dots, \mu_N, \Sigma_1, \Sigma_2, \Sigma_3, \dots, \Sigma_N]$	$Nd + \frac{Nd(d+1)}{2}$
4. k -Cluster ability	$\Theta = \left[\underbrace{(\underbrace{\mu, \mu, \dots, \mu}_{\kappa \text{ times}}, \underbrace{(\mu_g, \mu_g)}_{k \text{ groups}}) \dots, \mu_N}_{N \text{ sensors}}, \Sigma_1, \Sigma_2, \Sigma_3, \dots, \Sigma_N \right]$	$kd + \frac{kd(d+1)}{2}$

Please note that when Case 1 occurs in real time, μ and Σ are enough to describe the fused distribution $p(f(S_i, B_i))$ indicating that fusion of the sensors is only increasing the confidence in the localized estimate μ . In Case 2, each sensor is localizing the platform at a different location and fusion of data from such sensors is only going to increase the uncertainty of the fused estimate. Case 3 helps identify the necessity to evaluate w alternative clusters κ sensors (Case 4) through the combinatorial search that when fused will increase the confidence in localization. We illustrated this idea using a simple example in Figure 4.6. By inferring the occurrence of Case 1, by evaluating the first three hypotheses in Table 4.3, can save a lot of time while operating in the real world by not having to evaluate the combinatorial case alternatives.

We illustrate a simple example with uncertainty ellipses in Figure 4.7 for a three sensor system to understand the three cases better. We see that Case 1 refers to the possibility when all sensors are essentially indicating the same localization result. Case 2 points to ambiguity in the localization as sensors are indicating different robot pose with the same belief. In Case 3, when a smaller group of sensors are considered, we notice that a particular pose estimate appears more likely. We evaluate these three hypotheses first using the information criterion in Equation 4.25. The minimizer of the criterion identifies the situation and guides us towards Case 4, if necessary, to find group of sensors that are maximising a particular likelihood after the combinatorial analysis. Once the optimal sensor clustering is scored using the criterion, we are able to easily find k , and the κ sensors in agreement.

$$SV(X_g) = ICOMP(\Theta; X_g) = -2 \log \text{Likelihood}(\Theta | Y_i; i = 1, 2, \dots, N) + 2C_1(F^{-1}(\Theta)) \quad (4.25)$$

where X_g refers to the partitioning of the validity matrix V into various clustering alternatives and Θ is the parameter vector for X_g . We provide the formulae to compute the log-likelihood and $C_1(F^{-1}(\Theta))$ for the different cases but direct the reader to (Bozdogan, 1986) for implementation details and the derivation of the formulae presented.

In Equations 4.26, 4.27, 4.28 we provide the generic form of the likelihood term and the model complexity term from criterion in Equation 4.28 to evaluate arbitrary X_g , and then the case specific simplifications in Table 4.4 and 4.5. These simplified equations assume all distributions B_i used in V to be normally distributed.

$$\begin{aligned} \log L(\Theta = \{\hat{\mu}_g\}, \{\hat{\Sigma}_g\} | V) = \\ -\frac{c_{total}d}{2} \log(2\pi) - \frac{c_{total}}{2} \log |\hat{\Sigma}| - \frac{1}{2} \text{tr}(\Sigma^{-1}) \sum_{g=1}^N A_g - \frac{1}{2} \text{tr}(\Sigma^{-1}) \sum_{g=1}^N c_g (\bar{Y}_g - \mu_g)(\bar{Y}_g - \mu_g)' \end{aligned} \quad (4.26)$$

where $\bar{Y}_g = \text{mean}(Y_i \text{'s in each } X_g; g = 1, 2, 4, \dots, k)$

$$\hat{F}_g^{-1}(\text{Cov}\{\hat{\mu}_g\}, \{\hat{\Sigma}_g\}) = \begin{bmatrix} Q^{-1} \otimes \Sigma_g & 0 \\ 0 & \frac{2}{c_{total}} D_p^+(\Sigma_g \otimes \Sigma_g) D_p^+ \end{bmatrix} \quad (4.27)$$

where $Q = \text{diag}(\frac{c_1}{c_{total}}, \frac{c_2}{c_{total}}, \dots, \frac{c_N}{c_{total}})$; and other notations are similar to the ones described in Equation 4.21.

$$C_1(\hat{F}_g^{-1}) = \frac{\text{rank}(\hat{F}_g^{-1})}{2} \log \left[\frac{\text{tr}(\hat{F}_g^{-1})}{\text{rank}(\hat{F}_g^{-1})} \right] - \frac{1}{2} \log |\hat{F}_g^{-1}| \quad (4.28)$$

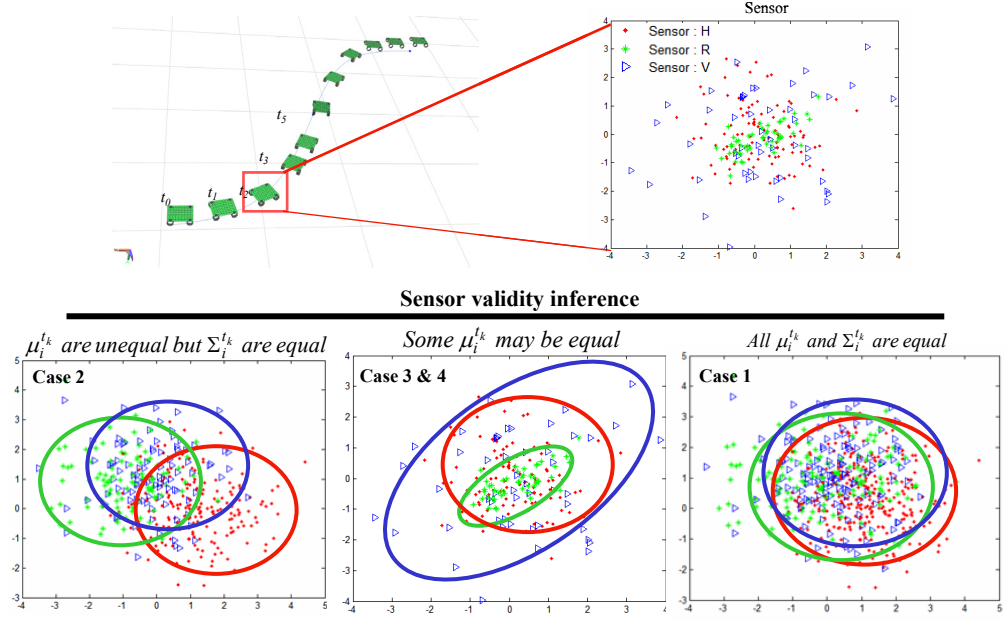


Figure 4.7: Perceiving sensor validity computation as a multi-sample clustering problem. We have shown three possible cases of mean and covariance conflict that can occur during localization. The ellipses indicate the uncertainty from each sensor. By using multi-sample clustering we are able to infer and identify the case that best fits the sensor belief distributions.

In our implementation, as soon as we infer one of the first three cases to minimize ICOMP, we assign all the sensors the ICOMP value of case hypothesis to be the sensor validity score. Then, for Case 2 and Case 3 alone, we perform the sensor clustering and evaluate all cluster combinations. The minimizer of the ICOMP points us to the optimal clustering of κ valid sensors. We assign this minimized ICOMP value for only the sensors within the maximal sensor cluster as their corresponding sensor validity scores V_i according to the rules in Equation 4.30. Since ICOMP is an estimate of the KL distance between competing model distributions, our sensor clustering can be considered as based on information distances between belief distributions.

$$V_i = V(S_i) = \min_{Hypothesis} SV_{Hypothesis}(S_i) \quad (4.29)$$

$$SV_{Hypothesis}(S_i) = \begin{cases} SV_{Case\ 1}(S_i) & \text{All } \mu_i^{t_k} \text{ are equal and } \Sigma_i^{t_k} \text{ are equal} \\ SV_{Case\ 2}(S_i) & \text{All } \mu_i^{t_k} \text{ are equal but } \Sigma_i^{t_k} \text{ are not} \\ SV_{Case\ 3}(S_i) & \text{All } \mu_i^{t_k} \text{ and } \Sigma_i^{t_k} \text{ are not equal} \\ SV_{Subset\ \kappa}(S_i) & \text{Maximal } \kappa \text{ sensors converge on } \mu_i^{t_k} \text{ and} \\ & S_i \in \kappa \text{ selected sensors} \end{cases} \quad (4.30)$$

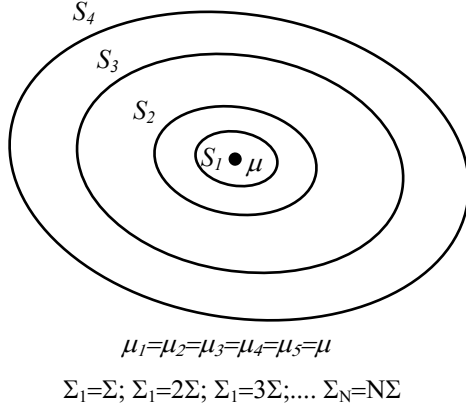
By assigning the validity score in the nested fashion, we have made sure that we are able to accommodate for sensor tendency and preference to succeed while working as a team preserving the robustness expected out of power set reasoning. We illustrate the impact of this approach in Figure 4.8. The red crosses indicate the V_i scores assigned to each sensor.

Table 4.4: Computing the log-likelihood term in Equation 4.25 for the sensor validity score.

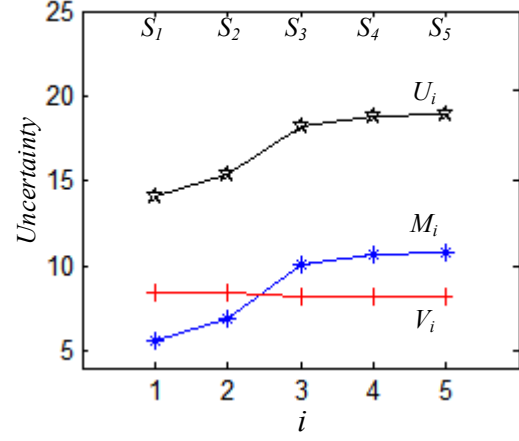
Case	Reduced likelihood term (L)
Case 1: $\Theta = \left[\underbrace{\mu, \mu, \mu, \dots, \mu}_{N \text{ times}}, \underbrace{\Sigma, \Sigma, \Sigma, \dots, \Sigma}_{N \text{ times}} \right]$	$L = -\frac{c_{total}}{2} \log c_{total}^{-1} T ; \quad T = V^* V$
Case 2: $\Theta = \left[\mu_1, \mu_2, \mu_3, \dots, \mu_N, \underbrace{\Sigma, \Sigma, \Sigma, \dots, \Sigma}_{N \text{ times}} \right]$	$L = -\frac{c_{total}}{2} \log c_{total}^{-1} W ; \quad W = \sum_{g=1}^N A_g$ $A_g = \sum_{l=1}^{c_g} (y_l^g - \bar{Y}_g)(y_l^g - \bar{Y}_g)'$
Case 3: $\Theta = [\mu_1, \mu_2, \mu_3, \dots, \mu_N, \Sigma_1, \Sigma_2, \Sigma_3, \dots, \Sigma_N]$	$L = -\frac{1}{2} \sum_{g=1}^N c_g \log c_g^{-1} A_g ;$
Case 4: $\Theta = \left[\underbrace{(\underbrace{\mu, \mu, \dots, \mu}_{\kappa \text{ times}}, \underbrace{\mu_g, \mu_g}_{k \text{ groups}}), \dots, \mu_N}_{N \text{ sensors}}, \Sigma_1, \Sigma_2, \Sigma_3, \dots, \Sigma_N \right]$	<i>Nested form of Case 3</i>

Table 4.5: Computing the $C_1(\hat{F}^{-1})$ term in Equation 4.25 for the sensor validity score.

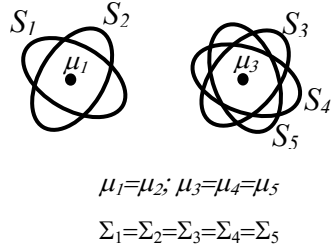
Case	Simplified $C_1(\hat{F}^{-1})$
<p>Case 1:</p> $\Theta = \left[\underbrace{\mu, \mu, \mu, \dots, \mu}_{N \text{ times}}, \underbrace{\Sigma, \Sigma, \Sigma, \dots, \Sigma}_{N \text{ times}} \right]$	$C = \frac{d + d(d+1)/2}{2} \log \left[\frac{\text{tr}(Q^{-1})\text{tr}(\hat{\Sigma}) + \frac{1}{2}(\text{tr}\hat{\Sigma})^2 + \sum_j \sigma^2_{jj}}{d + d(d+1)/2} \right]$ $\hat{\Sigma} = \frac{1}{c_{total}} T$
<p>Case 2:</p> $\Theta = \left[\mu_1, \mu_2, \mu_3, \dots, \mu_N, \underbrace{\Sigma, \Sigma, \Sigma, \dots, \Sigma}_{N \text{ times}} \right]$	$C = \frac{Nd + d(d+1)/2}{2} \log \left[\frac{\text{tr}(Q^{-1})\text{tr}(\hat{\Sigma}) + \frac{1}{2}(\text{tr}\hat{\Sigma})^2 + \sum_j \sigma^2_{jj}}{Nd + d(d+1)/2} \right]$ $\hat{\Sigma} = \frac{1}{c_{total}} W$
<p>Case 3:</p> $\Theta = [\mu_1, \mu_2, \mu_3, \dots, \mu_N, \Sigma_1, \Sigma_2, \Sigma_3, \dots, \Sigma_N]$	$C = \frac{Nd + Nd(d+1)/2}{2} \log \left[\frac{\text{tr}(Q^{-1})\text{tr}(\hat{\Sigma}_g) + \frac{1}{2}(\text{tr}\hat{\Sigma}_g)^2 + \sum_j \sigma^2_{jj}}{Nd + Nd(d+1)/2} \right]$ $\hat{\Sigma}_g = \frac{1}{c_g} A_g$
<p>Case 4:</p> $\Theta = \left[\underbrace{\underbrace{(\mu, \mu, \dots, \mu)}_{\kappa \text{ times}}, (\mu_g, \mu_g), \dots, \mu_N}_{N \text{ sensors}}, \Sigma_1, \Sigma_2, \Sigma_3, \dots, \Sigma_N \right]$	$C = \frac{kd + kd(d+1)/2}{2} \log \left[\frac{\text{tr}(Q^{-1})\text{tr}(\hat{\Sigma}_g) + \frac{1}{2}(\text{tr}\hat{\Sigma}_g)^2 + \sum_j \sigma^2_{jj}}{kd + kd(d+1)/2} \right]$ $\hat{\Sigma}_g = \frac{1}{c_g} A_g$



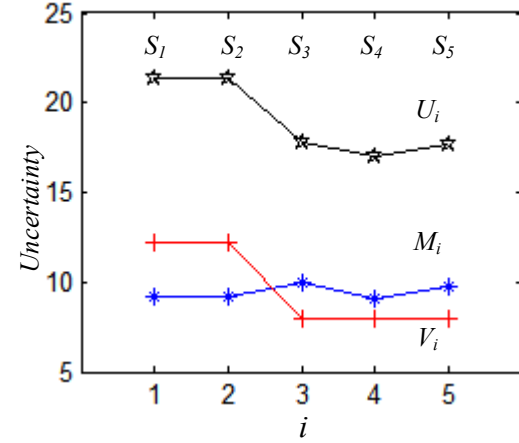
(a)



(b)



(c)



(d)

Figure 4.8: Our formulation of M_i and V_i is able to identify the separability in the mean and the covariance. (a) V_i being nearly constant identifies sensor agreement on state μ while the M_i monitors the confidence levels. Our framework decides to use only Sensor 1 in the belief propagation. (b) V_i identifies two clusters of sensors agreeing on the same state, while M_i indicates that sensors are operating at the same confidence. We decide to fuse sensors 3, 4 and 5 as $V(S^*) < M(S^*)$ for $S^* = \text{Sensor 4}$.

4.5 Resolving the fusion versus selection dilemma

We presented graphs simulating two situations in Figure 4.8. In Figure 4.8(a), we considered five sensors, and the corresponding uncertainty scores when each sensor agrees on the mean but the covariance on the Sensor S_i is $i\Sigma$. For example, if Sensor 1 operates at Σ , Sensor 2 operates at 2Σ , Sensor 5 at 5Σ . We are able to see that our validity score (score of co-operative confidence) for the sensors plotted as a red cross in Figure 4.8 (b) identifies sensors that are in agreement about the state. The measurement uncertainty score (score of self confidence) plotted as blue asterisk marks increases monotonically as expected. In Figure 4.8 (c), we showed another case where some sensors agree and some do not and interpret the results of our uncertainty handling framework in Figure 4.8 (d). In both the graphs, please note that the total uncertainty curve in black follows the trend as either validity or measurement certainty, clearly separating and handling both the kinds of sensor conflict, the separation in the mean and also the separation in covariance whichever one is significant.

Both the measurement uncertainty score and the sensor validity scores being normalized information measures of complexity in our implementation, we use the sum of the two measures of sensor measurement uncertainty $M_i = M(S_i)$ and sensor validity $V_i = V(S_i)$ to choose the sensors with minimum total uncertainty. The sum operator on the uncertainties is justified because both Equation 4.19 and 4.24 are log likelihood functions. Adding two log likelihood functions is actually multiplying their joint likelihood assuming mutual independence. This formulation of M_i and V_i , helps us decide on the dilemma between fusion and selection. Following the rules in building fusers that perform better than the best sensor (Rao, 2001), we are able to derive the following condition.

Condition: If C^* represents the optimal cluster with maximal k reliable sensors identified in Section 4.2, and V_i and M_i the respective reliability and confidence scores of the sensors, find sensor S^* with $\min(V_i + M_i)$. If $M(S^*) < V(S^*)$ decide sensor selection and choose S^* , else decide to fuse ' k ' reliable sensors in cluster C^* of which S^* is a member for belief propagation.

While formulating our approach we had argued that success in fusing state information is achieved if we maximize the information about the state variables of interest using data from each sensor individually and also maximizing the information gain from fusing data from different sensors while penalizing for the information corruption. The self-confidence score M_i is a measure of how much information does the sensor provide about the state individually. V_i monitors both the information gain in the fusion penalizing for the corruption in the form of number of modes and increased area of the fused distribution. Since the fused distribution is always expected to be more confident than the sensors taken individually, we expect that V_i scores be always less than the corresponding M_i scores except in the case of catastrophic fusion and a trivial case of a 1-sensor system where M_i and V_i would be equal. Our condition to resolve fusion versus selection makes sure, catastrophic fusion never occurs by eliminating sensors in state conflict and by avoiding sensors with uncertainty beyond expected levels from the fusion.

We are able to demonstrate how our approach answers the selection versus fusion dilemma avoiding "catastrophic fusion" (Mitchell, 2007) in the same example as Figure 4.8. Starting with Figure 4.8(a), when sensors all agree on the state with different levels of confidence, our method is able to eliminate less confident sensors. Our criterion selects Sensor S_1 , keeping sensors with confidence less than 2Σ away from the belief propagation stage. In Figure 4.8 (b), we show 5 sensors, out of which 2 clusters agree on the state, but with the same level of confidence from all the sensors. Our framework is able to identify this situation in the reliability score and decides to fuse sensors S_3 , S_4 and S_5 . This result is significant because, when we started, we desired that each sensing mechanism on our mobile platforms be productive to the best of its ability as an individual unit while making sure that the collective operation and success of individual sensing mechanisms will avoid the failure of the multi-agent system. The resolution of the dilemma is critical to the success of the initial desire.

In summary, we have derived an uncertainty minimization framework for robot self-localization that can simultaneously quantify the uncertainty due to noise associated with each sensor measurement and also infer evidence about sensor validity using belief estimates from multiple sensors. We have explained a procedure that uses the theory of information complexity in constructing scores for both the sensor validity and the measurement uncertainty from the Bayes belief distributions towards choosing a reliable subset of the multi-sensor data. Furthermore, by resolving the fusion versus selection dilemma, we are able to implement a self-correcting stability into the multi-agent scenarios guiding sensor adaptation in unseen unexpected operating conditions.

Having explained the theory in this Chapter, we demonstrate promising results of implementing this framework in several real world situations in Chapter 5 to claim the following contributions:

- An information theoretic framework to simultaneously score sensor measurement uncertainty as *self confidence* and sensor validation uncertainty as *co-operative confidence* in the context of data fusion. Thus far, though both these uncertainties have received attention, they have been considered as distinct and independent problems.
- A new algorithm that can automatically detect performance degradation and sensor conflict during operation to eliminate failed sensors and bad pose recovery due to data association problems in a dynamic environment, guiding the switch to the next available good sensor.

5 Experiments: Handling Uncertainty

In this chapter, we demonstrate the implementation of the uncertainty minimization theory developed in Chapter 4. We begin with Section 5.1 describing the simulation experiments that encourage the real world robot localization experiments. Then, we present real world deployment results emphasizing the improvements on the instrumentation-based localization in Section 5.2 and state estimation from images using the multi-feature consensus idea in Section 5.3.

5.1 Simulation study

We conduct the following experiments to demonstrate the capability of the statistical framework explained in Chapter 4. Our simulation results provide the proof of concept in handling dynamic uncertainty towards localization and mapping. We implemented a simulation environment by considering a three sensor system with sensors H , R and V measuring a d -dimensional state vector. We tested several motion paths under different noise models within the simulator to evaluate our method for efficiency and robustness. In Figure 5.1, we have shown three different synthetically generated cases on 2-dimensional state vector that help visualize, understand and appreciate our algorithm. The ellipses simulate the uncertainty estimate output from a Kalman filter that we use for belief propagation. In some cases, these ellipses are not visible because of the high degree of certainty they encapsulate. We interpret the output of our data fusion algorithm in two forms: (i) which sensor to believe (bottom inset) and (ii) the number of sensors contributing to that belief (top inset) in the figure.

Figure 5.1 (a) shows the case where the robot tries to stay on a sigmoid path with sensor H being more accurate at the first few samples (the red ellipses are not visible because of the high belief), while the other sensors converge on the localization information over time. Our sensor selection algorithm correctly picks up sensor H , switches to the next available sensor R , and also is able to infer that all three sensors are converging after the first few samples. In Figure 5.1 (b), sensors H , R and V are all in agreement on the localization data. Our algorithm successfully infers the agreement. The third example, in Figure 5.1 (c), is the closest to reality where some sensors can fail suddenly forcing the need to switch to a better sensor. We observe the likely pose estimate from the sensor H consistent and believable to begin with but deteriorating over time. Our sensor selection result automatically detects the deteriorating belief on sensor H , guiding the switch to sensor V as a better option.

We compare our adaptive approach with covariance-weighted fusion for localizing the mobile robot in Figure 5.1 (d). We show the sum-squared error of the pose vector from the intended path in each of the three examples considered in Figure 5.1(a) (b) and (c). The time required for each of these approaches is also plotted in Figure 5.1 (d). We observe that the sensor selection performs better by minimizing localization error at each instant. However, the accuracy appears to come at the expense of a few extra computations compared to covariance weighted-fusion (Hackett and Shah, 1990). The extra computations that consume a few milliseconds more do not seem to impose a problem for real-time operation. Another interesting aspect to note from the error analysis graph is that our approach performs as well as the best sensor if not better. This means that our method can deal with possibly invalid data without compromising the best case performance.

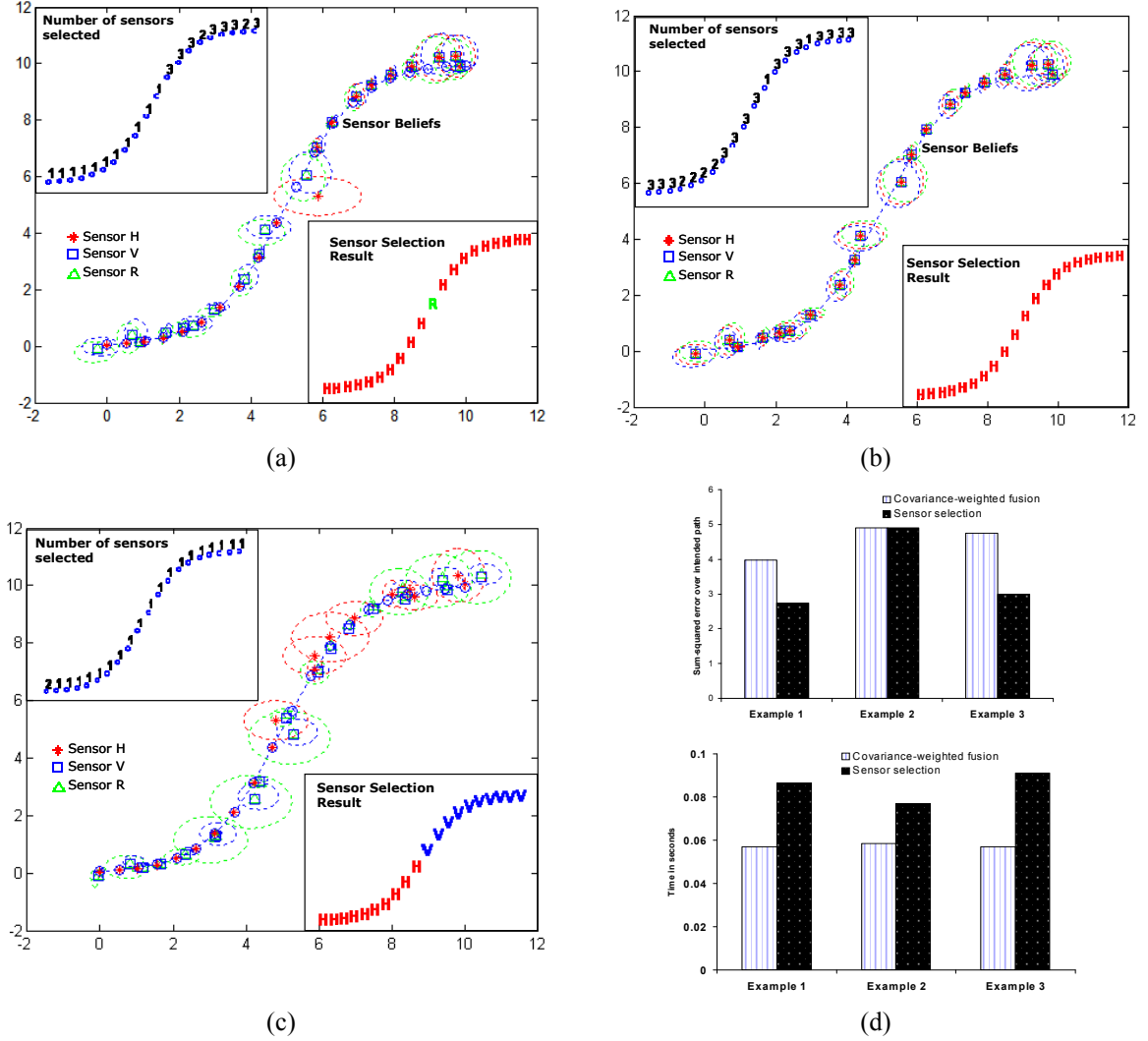


Figure 5.1: Localization in a simulation environment. (a) Example 1: Detecting intermediate failure. (b) Example 2: Self-localization when sensors are converging on a believable pose. (c) Example 3: Guiding the sensor switch when the belief in particular sensor pose deteriorates. (d) Error and timing analysis for sensor selection for the three examples presented. The top inset is the plot of sum-squared error over the entire intended path localized using the covariance weighted fusion method and our sensor selection method. The bottom inset is the timing analysis comparing fusion with selection.

On Example 2, when all sensors were essentially telling the same information, we see that sensor selection is not able to perform better than the covariance-weighted fusion while on cases with varying beliefs over time, selection leads to better localization. Our framework being able to decide fusion and selection is an added feature. Though these experiments provide the proof of concept, several questions arise in bridging the gap between the theory and practical implementation. How many samples does our framework need for a confident decision? Will some sensors weigh more than others because of their sampling rates? What is the minimum sampling required to separate the sensors based on their uncertainty variance? Is our method capable of real-time operation? How many sensors can we accommodate for real time inference? What is the ideal dimensionality for the state vector? The next set of results in Figure 5.2 based on the simulation study tries to provide answers to those questions and identify the empirical bounds of the framework.

Figure 5.2 (a) is a plot to reveal the minimal samples of the belief distributions required to separate sensors beyond a particular uncertainty level. We have plotted the sensor confidence score (the self confidence) for different sampling frequencies for a two dimensional state vector. The plot reveals that we require at least 100 samples of belief distributions to separate a sensor with a preferred confidence Σ and eliminate the rest with confidence uncertainty greater than 2Σ . Once we decide the level of confidence to allow in the system, we study this plot, and empirically choose a value on the sensor confidence score, that can be used to normalize the confidence score when different sensors are operating at different sampling frequencies. Figure 5.2(b) takes care of resolving the sampling issue for different size of the state vector providing bounds for a particular system model. The magnitude of the confidence score from plots in Figure 5.2 (a) and (b) is the lower bound on the uncertainty achievable with a sensor operating at a predefined sampling frequency measuring p -dimensional state.

Figure 5.2 (c) and (d) are timing results. Figure 5.2 (c) plots the time required to compute the total uncertainty for one sensor measuring states with varying dimensionality. We see that up to 15 variables in the state vector for belief propagation, our inference mechanism can keep up with real-time operation requirements. For systems that have to deal with larger state dimensions, we recommend that the model be broken down into smaller sub-models before using our framework. The next experiment we conduct is more significant and reveals a caveat. We evaluate the number of sensors the inference framework can handle and perform within real-time requirements. Since the clustering for the altruist confidence is a combinatorial process, we see that our framework can only handle up to 6 sensors within the real-time requirement of making a decision in less than a second. We tried to avoid the combinatorial evaluations using smarter search strategies like genetic algorithms and randomized search. Genetic algorithms on average reduce the search time for the optimal clustering by 15% and the randomized search with special stopping constraints was slightly better at 20%. However, for systems with more than 8 sensors, we recommend a binary search approach by first finding the cluster of half or more sensors in agreement and then directing the search based on the previous result.

The next question that we try to answer is regarding the Gaussian apriori assumption. Since our motive with uncertainty minimization with state-space systems is to prefer unimodality, and the assumptions about all the noise models is Gaussian, we cannot be absolutely sure about the performance of the method in cases, where we cannot assume Gaussian apriori. However, the information complexity criterion and its newer model misspecification tolerant version (Bozdogan, 2000) can help us deal with non-Gaussian prior. We propose to include this newer version into our framework and be able to at least quantify the error due to misspecification in belief propagation with the Cramer-Rao bound as the worst case limit. However, for the results presented in this dissertation, the misspecification implementation was not necessary because most of the noise models with our sensors indeed follow the Gaussian distribution. We have shown the impact of misspecification of the noise model when Gaussian assumption is made on uniformly distributed noise models in Figure 5.2 (e). We observed that the misspecification resistant implementation matched the ground truth expected result 95% of the time in 100 simulated trials.

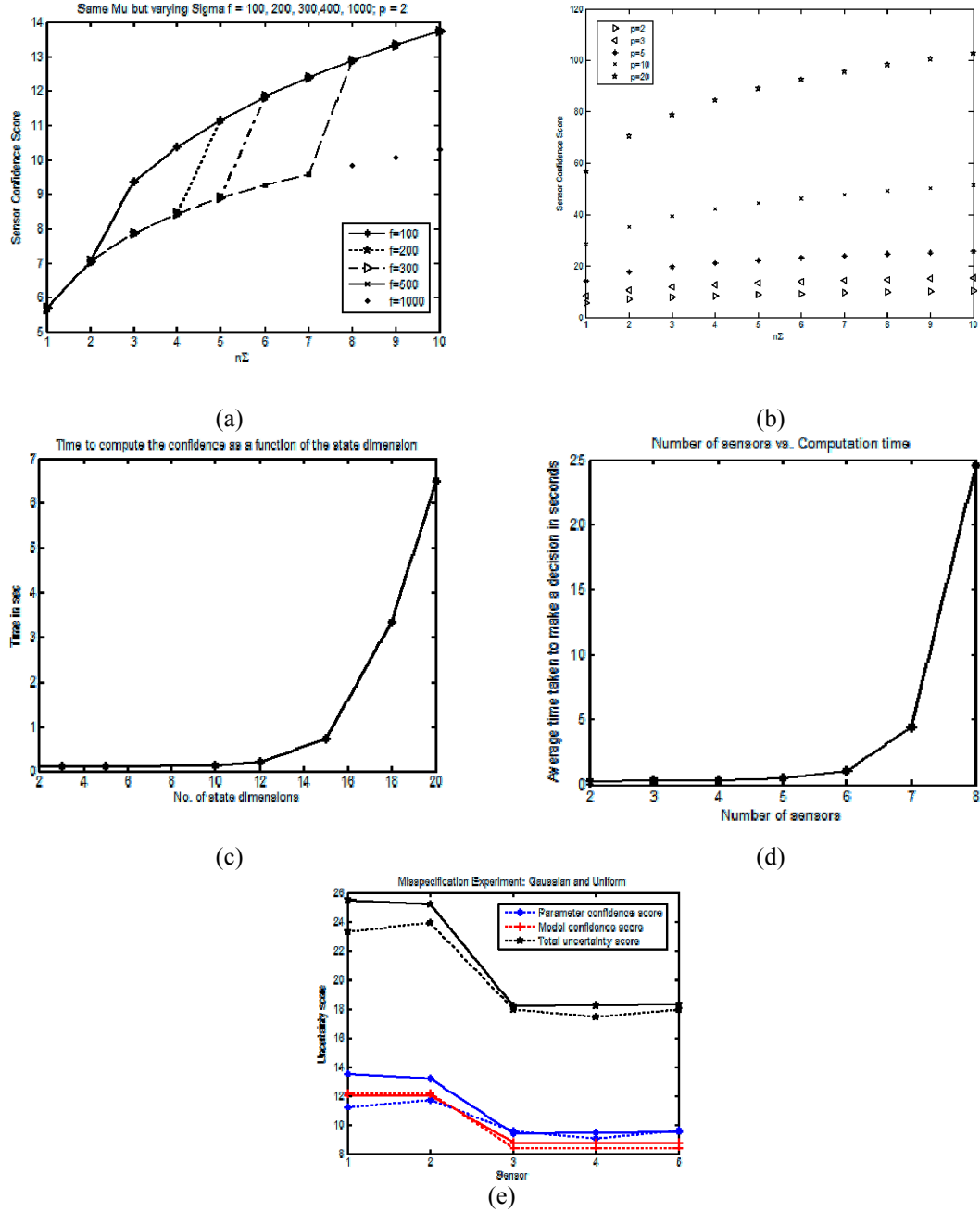


Figure 5.2: The study on how the indirect parameters might affect inference. (a) Effect of sampling frequency (b) Effect of dimensionality on the state vector. (c) Timing analysis for state models of different dimensions (d) Timing analysis for as the number of sensors in the system increases. (e) Inference when noise models are not Gaussian. The dotted line is the ground truth generated from uniform distribution and the thick lines are estimate from the implementing the recent form of the information criterion (Bozdogan, 2004).

5.2 Handling localization uncertainty in instrumentation-based mapping systems

Moving away from simulations, the first real-world application that we discuss is on a mobile platform with laser scanners, video cameras, global positioning systems (GPS) and inertial equipment mounted on a van as shown in Fig. 5.3a. The noise models for these systems were built through extensive apriori characterization of each of the sensors. With the van, though not completely autonomous in its operation in urban environments, we use the position information from the GPS, orientation from the inertial measurement unit, recover pose from the 3D profile data (Martinez, 2006) and video data (Nister, 2004a), and use these datasets to automatically detect a GPS outage which is common in urban environments and switch to the next reliable sensor.

We show our mobile scanning system along with the intended path overlaid on the aerial map of the city. The test course in Figure 5.3 is approximately 300 meters long. Figure 5.3 (c) shows the result of pose recovery from several sensors indicating areas that have error in the order of a few centimeters that later builds to the order of a few meters. Figure 5.3 (d) is the result of our data fusion algorithm indicating reliable pose recovery methods along the total path. We are able to see the areas in which pose from video was not a reliable method compared to pose recovery using hardware and range sensors. In Figure 5.3 (e), we compare the result from our approach to using a Kalman filter-based fusion (Gao and Harris, 2002). For a 300 m path, the fusion without sensor rejection produces a drift of 2 m which can be serious while operating autonomously in real world environments. That being the improvement in localization error, we further emphasize our improvement in the map in Figure 5.4. We show the bird's eye view of the building and render the profiles after integrating by Kalman fusion and our proposed approach. Visually, we are able to see that our approach has eliminated several crisscrossing profiles in resolving the geometry at higher precision.

The next set of experiments we conduct uses a small mobile robot navigating a corridor. The robot in the corridor shown in Figure 5.5 has five cameras. The idea is that one camera looks ahead in the path and avoids obstacles, while the other four look at different fields of view for localization. Our test environment has doors, windows, and objects like chairs, book shelves etc. in the path on either side. We use multiple cameras to accommodate for lack of features on painted walls and also for a larger coverage in the corridor looking for traceable features. A pose from video algorithm similar to (Nister, 2004a) and (Davison, 2007) provides the robot's pose for localization using apriori calibration information. The uncertainty in pose is determined by estimating the confusion in converging to optimal relative pose as discussed in (Zhu et al., 2005). Our sensor selection algorithm operated on these values in successfully localizing the robot through the entire intended path that we show in Figure 5.5 (b). We use this environment as a test bed for localization where we know that there may not be enough traceable features on all cameras (the field of views of the camera are shown in Figure 5.5 (c)) over the entire intended path at a given point of time. We expect that when one sensor is tracking features on doors and windows, the others might be struggling to locate interesting features on plain walls.

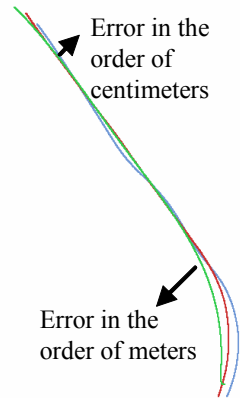
We plot the number of oriented matches between successive frames in pose recovery using the images from these cameras in Fig. 5.5 (d). We are able to see that our sensor selection algorithm automatically switched to the available next best sensor, when the pose recovery was not within acceptable levels minimizing the overall uncertainty about the state. This emphasizes the capability of our method in automatically switching to a good sensor while navigating in a dynamic environment where pose recovery methods can fail due to data association problems or lack of features. We show results of another experiment in Figure 5.6. Here, laser range data and the vision data were used to generate the floor plan of the corridor. Our method performed the switch between video and range to localize and align the model range profiles in autonomously generating the map just like most animals do in unknown environments.



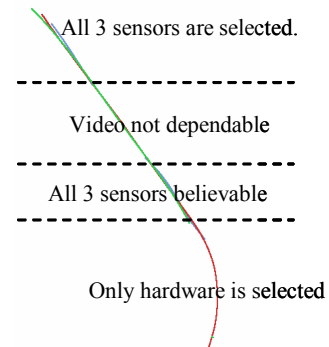
(a)



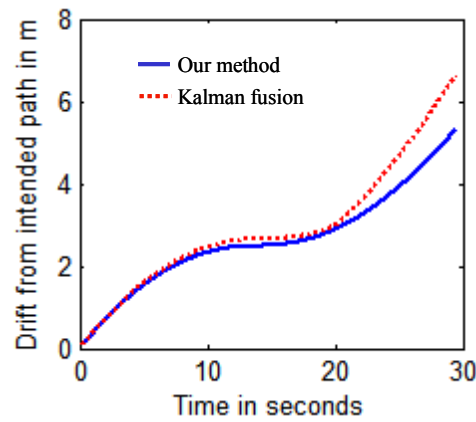
(b)



(c)



(d)



(e)

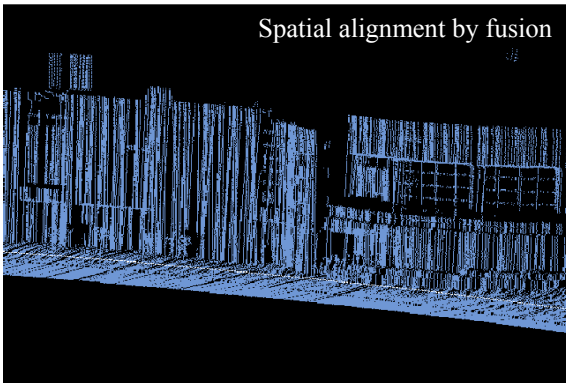
Figure 5.3: Localization in urban environments. (a) Our mobile platform with laser range scanners, video cameras, GPS and inertial measurement units. (b) Intended path (blue curve) in the city. (c) Pose recovery from hardware instrumentation (red curve), range scanner (green curve) and the video camera (blue curve). (d) Sensor selection result. (e) Drift produced in conflicting sensor data is not discarded.



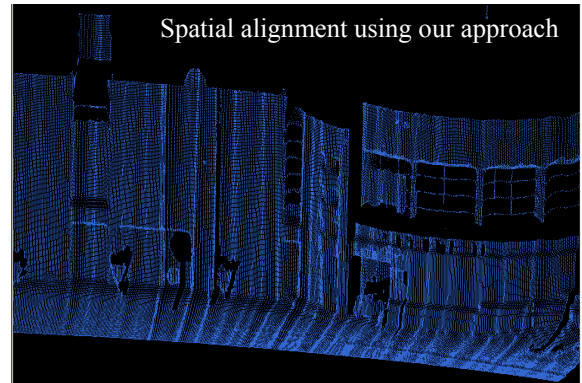
(a)



(b)



(c)



(d)

Figure 5.4: Reducing the uncertainty about localization improves the integrated map. (a) The bird's eye view of the Women's Basketball Hall of Fame building, University of Tennessee, Knoxville from www.maps.live.com. (b) Screen capture after rendering our textured 3D result. (c) Profile alignment if data from all the sensors are fused together. (d) We see that profile alignment using our approach as compared with the image in 5.5 (c) has minimized the crisscrossing of the range data.

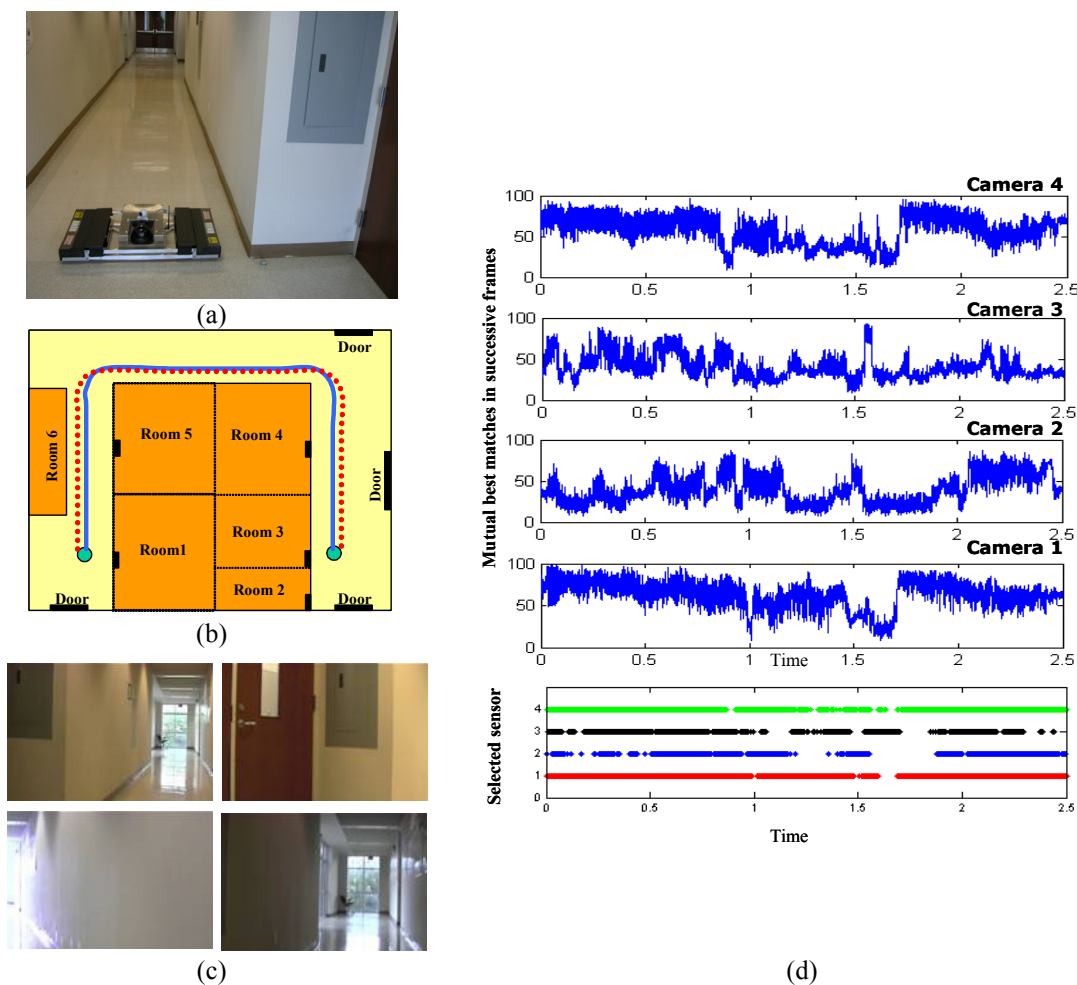


Figure 5.5: Localization of a mobile robot with multiple cameras in an indoor environment. (a) Our mobile platform with five cameras (b) Intended path (blue curve) in the corridor and the localized path (red dotted curve). (c) Field of view of four cameras localizing the robot. As the robot navigates in the area, we expect that some cameras do not have sufficient feature correspondences for localization. (d) Number of mutual matches between successive frames in the video that can be loosely related to the confidence with pose recovery along with the sensor selection result.

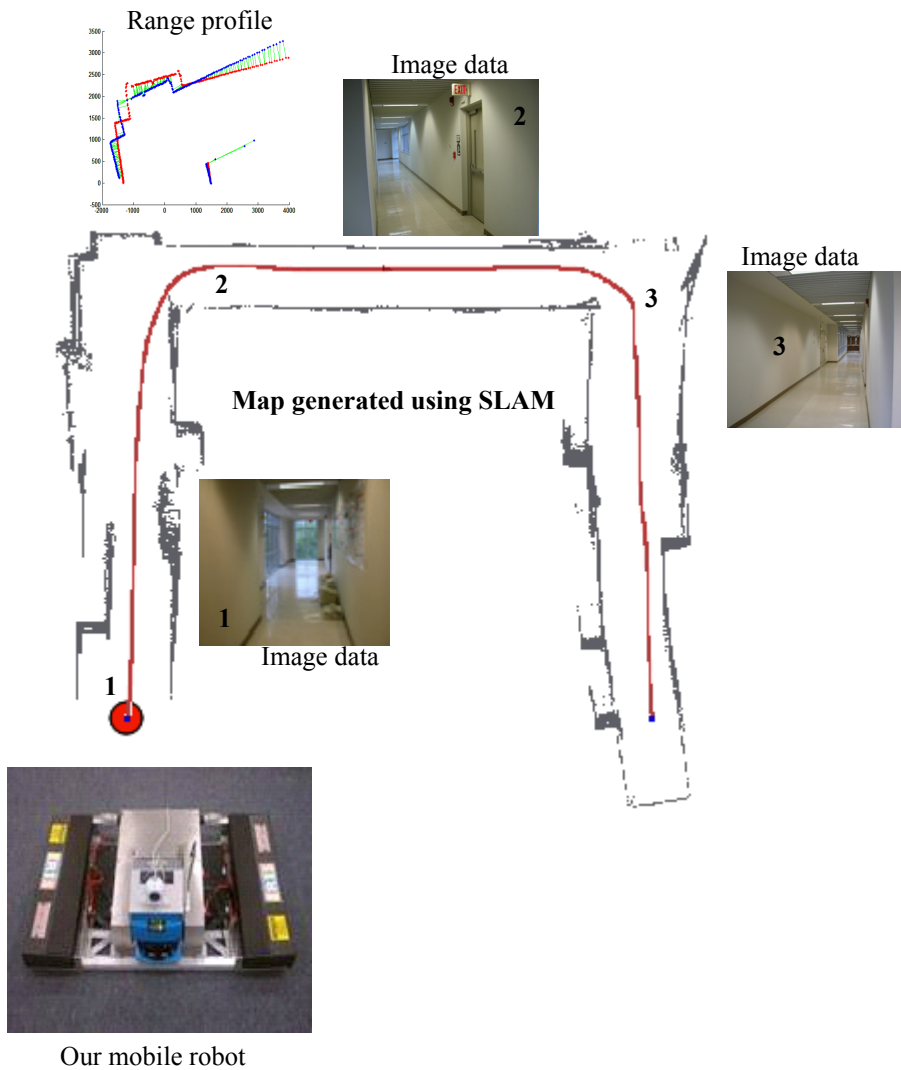


Figure 5.6: Localization and mapping of a mobile robot with a camera and a range sensor in an indoor environment. We have presented the output 2D floor plan of the corridor after simultaneous localization and mapping.

Animals have the survival instinct to adapt in any unknown environment. For example, a bee though does not possess several sensors like the humans, remembers and retraces the path traversed to localize, navigate, explore and learn an unseen environment in the absence of a global landmark like the sun (Dyer et al, 1993; Menzel, 1996). A rat uses a photographic technique to remember places of interest and recalls the iconic images from memory for surviving in the new environment searching for new food sources. Monkeys and humans remember informative features, combine several sensory mechanisms to navigate through an environment (Wang and Spelke, 2002; Touretzky et al., 1994). These ideas inspire a modern day robot that is built with range, vision and odometer sensors (Trullier, 1997). The mobile robot in Figure 5.6 with a laser scanner and a camera tries to emulate the capability of such animals and the goal with our algorithm is to help the robot make intelligent decisions while surviving to explore in an unknown environment.

Our approach tries to make the robot behave like intelligent bees, rats and humans switching characteristics as the robot moves around. Animals while exploring for food use a variety of tools for landmark selection, archival and retrieval. A robot trying to map an area autonomously also implements the capability to select features in image and range data to use them in localizing itself. But the tougher question that we have tried to address in this dissertation is how a robot should behave in an unknown environment, where we are not sure that vision-based SLAM or range-based SLAM would work best. The implementation of our statistical approach to minimize uncertainty in localization imparts the intelligence by guiding the robot to behave like a bee when there are dependable range features in the scene, behave like a rat when image features are dominant and use a combination of both like the humans, when both are reliable maintaining high levels of certainty in the self localization of the robot. The map integrated in Figure 5.6 is by using such intelligence. The map of the floor plan generated in Figure 5.6 automatically switched sensors during localization along the trajectory (red curve) shown in the figure. The robot was programmed to dead reckon and move forward with a motivation to explore the corridor.

We have tried to illustrate the situations faced by the robot where our inference framework guided the sensor switch. Close to the starting point marked as a red circle in Figure 5.6, the computer equipment on the floor provided reliable features in the range and vision sensors. The belief in the range sensor peaked around the corner marked “2” with interesting spatial corners. But immediately after the turn in the corner, range features were not distinct and distinguishable and the robot switched to vision-based navigation. The belief on the range sensor springs back again as the robot gets closer to corner marked “3”.

So far, we presented a sensor selection algorithm based on information theoretic model selection criteria for robotic localization. Our approach to bring together measurement uncertainty and reliability using information measures for uncertainty minimization in localization is able to efficiently work on real and synthetic robot environments. Our results further encourage implementation on unmanned ground vehicles in following a well defined path where localization is necessary feedback due to the dynamic nature of the environment. An example of such a scenario is urban traffic, where an unmanned vehicle in addition to maneuvering amidst traffic should be able to switch sensors managing sudden unexpected GPS outage to stay on course towards the intended destination.

Our method is particularly suited for such applications involving multi-modality sensors for navigation and localization. Furthermore, we also note that our method can also perform as good as the best sensor in the system and sometimes better than the best sensor as expected in an ideal case of information fusion discussed in (Rao, 2001). Such a success can be attributed to decision fusion criterion explained in the previous chapter. Our framework is generic to any multi-sensor state localization application and thus far we have discussed our implementation of multi-modality instrumentation-based localization. With a slightly different implementation, our framework can also improve the reliability in localization for image-based systems also. We present the results on handling uncertainty in image-based systems in the following section.

5.3 Handling localization uncertainty in image-based mapping systems

5.3.1 Revisiting image-based localization

We introduced the algorithms for 3D inference from images in Chapter 2 and identified the suitability of shape from motion algorithms in vision-based navigation and mapping. We revisit the procedure for estimating pose and structure from image sequences described in Chapter 2 in the form of a block diagram in Figure 5.7. The volume of literature available in (Ma et al., 2006) and (Forsyth and Ponce, 2003) documents several methods based on the epipolar geometry (illustrated in Figure 5.8) for navigation and mapping purposes. We find several methods for feature detection, different methods for feature matching, different motion models for the features extracted from the scene and several minimal methods for inferring the 3D relationship between image frames in the form of a fundamental or essential matrix.

The fundamental matrix \mathbf{F} that relates two perspective images of a single rigid object/scene is estimated by solving the epipolar constraint in Equation 5.1, where $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$ are corresponding points in two images \mathbf{I} and \mathbf{I}' respectively (Figure 5.8). Assuming that the calibration matrix (\mathbf{K}) that includes CCD parameters including the principal point (θ_x, θ_y) , the sensor size parameters (s_x, s_y, s_θ) along with the focal length (f) of the camera acquiring images of the scene from different viewpoints is available through apriori calibration; \mathbf{F} is instrumental in the estimation of the relative rotation (\mathbf{R}) and translation (\mathbf{t}) of the camera relating the two images. The notation $[\mathbf{t}]_x$ refers to the anti-symmetric matrix form of the translation vector \mathbf{t} in Equation 5.1.

$$\tilde{\mathbf{m}}_i \mathbf{F} \tilde{\mathbf{m}}'_i = 0 \quad ; \mathbf{F} = \mathbf{K}' [\mathbf{t}]_x \mathbf{R} \mathbf{K}^{-1} \quad (5.1)$$

In vision-based vehicle navigation/localization applications, we note that the uncertainty about the pose (\mathbf{R} , \mathbf{t}) is directly related to the uncertainty on \mathbf{F} . Unfortunately, the accuracy and the uncertainty about \mathbf{F} depends on the quality of feature correspondences $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$. Though, theoretically all pixels in \mathbf{I} and \mathbf{I}' are eligible to contribute to $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$, for computational reasons and also for improved performance, we choose special interest points to build the list of correspondences $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$.

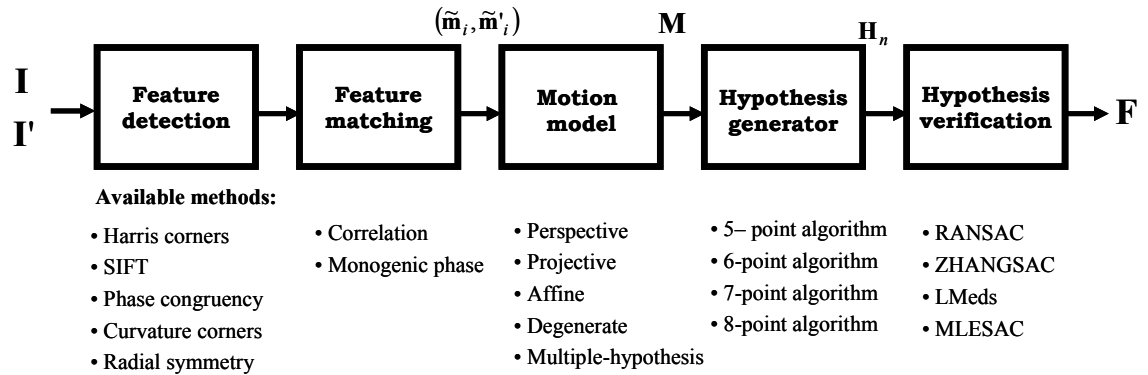


Figure 5.7: There are several options available in the flow process for estimating the fundamental matrix. However, these options are scene and motion specific and the choice of these methods (pipeline) in a real world situation has to be adaptive to the environment. In this section, we propose a statistical decision procedure to increase the confidence and reliability of the fundamental matrix.

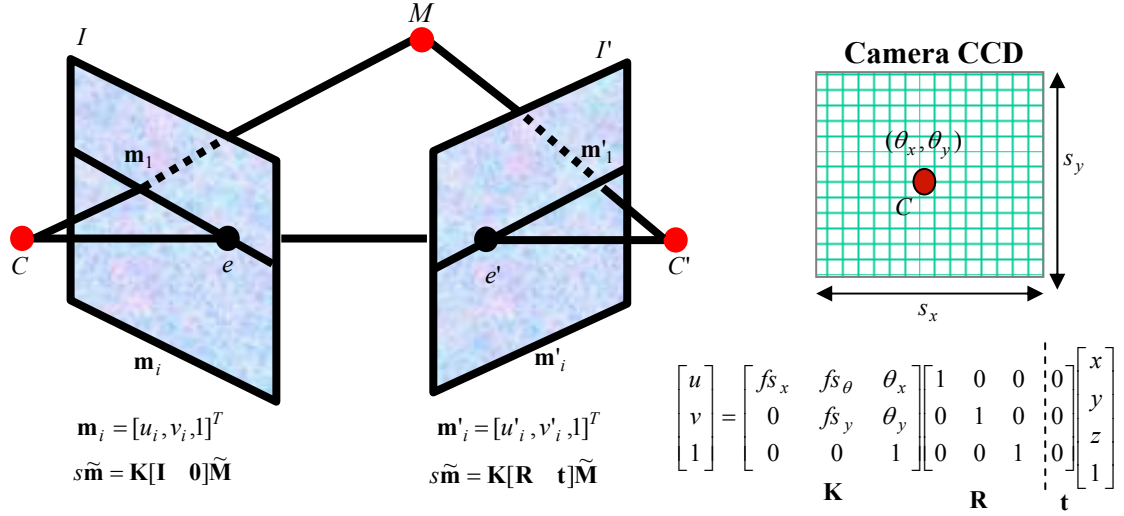
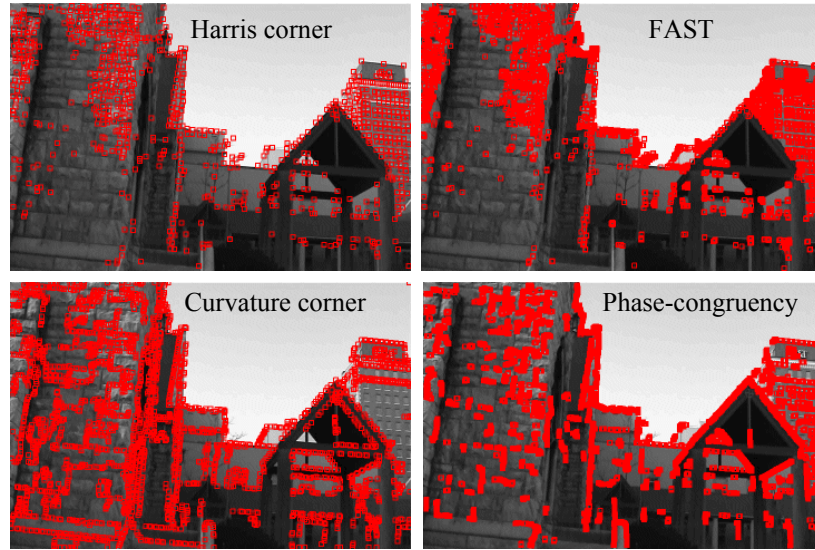


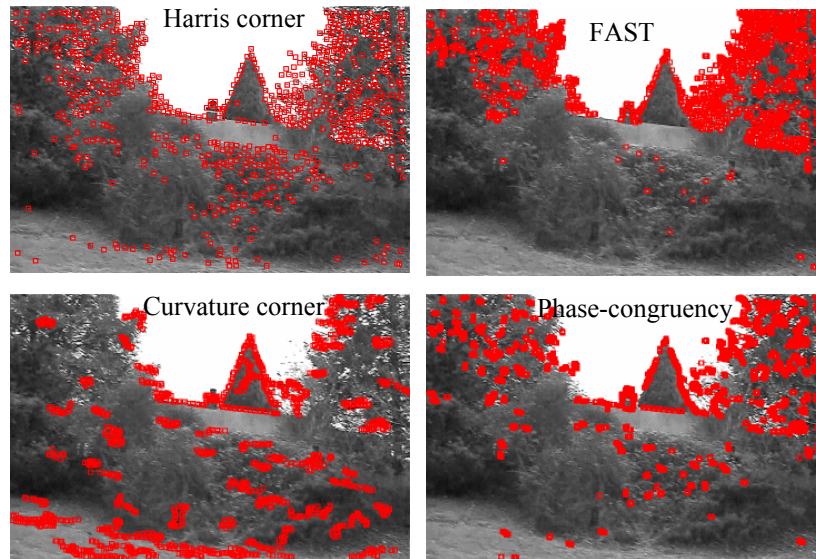
Figure 5.8: Revisiting epipolar geometry. The figure shows the geometry relating the 3D point in the scene (M) based on the putative matches from the images and the equations explain how the 3D point is projected into the physical CCD camera. We estimate \mathbf{K} using the calibration procedure in (Zhang, 2000).

The special interest points $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$ extracted are inspired by perceptual heuristics from human vision theory. For example, the Harris corners (Harris and Stephens, 1988) are intensity-gradient-based interest points, the curvature corner (He et al., 2004) is edge-based, the phase congruency corners (Kovesi, 1999) are spatial frequency inspired while SIFT (Lowe, 2004) is a multi-resolution feature. If all these feature point detectors extracted points with 3D characteristics that can be matched across images, the estimate of \mathbf{F} would be independent of the interest point detector. But, in reality we observe that feature point detectors perform better in some situations while not meeting the expectations in some others. The reason behind such an observation is that each of these interest point detectors extract monocular depth cues from the image. From studies of human perception, the linear perspective which is the most mathematically favored depth cue looks for lines and orientation of the lines in the image to infer about the 3D projection. The aerial perspective depth cue is observed through the color differentiation particularly between the blue and green. Other monocular depth cues in addition to the ones discussed in Chapter 2 include relative size, interposition, lighting, shading and the monocular parallax. Different feature detectors are inspired by one of these monocular cues. We show a simple example in Figure 5.9. We show feature points extracted using (Harris and Stephens, 1988), (Kovesi, 1999), (Lowe, 2004) and (Rosten and Drummond, 2006) on a structured building scene with linear perspective and another scene with a lot of vegetation. With this example, we illustrate that feature detectors that extract pixels for perspective inference in certain scenes may not be best suited in other situations. So, how do we choose the right feature detector for the scene of interest? Is it possible to choose the feature detector as the scene changes on a mobile platform adaptively?

Furthermore, there are two types of errors associated with image interest points from low-level feature detectors: (a) classification errors and (b) measurement errors. Classification errors occur when a feature detector incorrectly identifies a portion of the image as an interest point, while measurement errors occur when the feature is identified correctly but not at a precise repeatable location. The repeatability is a significant requirement while recovering pose from video frames. The measurement errors are usually smoothed by modeling the error as a normal distribution, but classification errors are gross deviations and cannot be averaged out.



(a)



(b)

Figure 5.9: Different feature detectors identify different monocular depth cue pixels (shown in red). (a) Feature points detected when the linear perspective monocular depth cue is available in structured environments. (b) Feature detectors in a challenging scene with vegetation. We are able to see that while some detectors are able to extract pixels with 3D characteristics in structured scenes, they are not as successful with vegetation and vice versa.

The error in the feature detection is significant for pose recovery in many ways. If ε is the probability that a feature point is in error, and p is the probability indicative of a feature match error, the number of minimal hypothesis n of s sub-samples taken at a time to fit a motion model \mathbf{M} of the fundamental matrix is shown in Equation 5.2. The confidence on the estimated geometry is based on an inlier bound condition modeled as the variance of a probability distribution (Equation 5.3).

$$n = \log(1 - p) / \log(1 - (1 - \varepsilon)^s) \quad (5.2)$$

$$\Pr(\{\tilde{\mathbf{m}}_i, \tilde{\mathbf{m}}'_i\}) = \prod_i \left(\frac{1}{2\pi\sigma^2} \right) e^{-\frac{d(\tilde{\mathbf{m}}_i, \hat{\mathbf{m}}_i)^2 + d(\tilde{\mathbf{m}}'_i, \hat{\mathbf{m}}'_i)^2}{2\sigma^2}} \quad (5.3)$$

Equations 5.2 and 5.3 hint at how the uncertainty on the corresponding points propagates as outliers in the subsequent matching stage corrupting the model parameters during the iterative hypothesis-verify scheme involved in solving the epipolar constraint.

Over the years of computer vision research for feature-based 3D geometric inference summarized in Figure 5.7, we have seen several interest point detection methods (Harris and Stephens, 1988; He and Yung, 2004; Koveti, 1999; Lowe, 2004; Loy and Zelinsky, 2003) claiming the reduction of pixel localization error ε . We have seen improved matching methods based on correlation (Huang and Netravali, 1994) and monogenic phase (Kunio, 2007) to operate on those interest points trying to produce matches with minimal probability of error p . Furthermore, there is also a recent body of literature on the s -point ($s = 8, 7, 6, 5$) algorithms (Nistér et al., 2004b; Stewenius et al., 2006; Hartley, 1997) for generating hypothesis model parameters \mathbf{M} , along with several iterative verification improvements over RANSAC (CVPR Workshop, 2006) to choose the best \mathbf{M} and its parameters.

The vast literature has forced performance evaluation publications comparing feature detectors (Mikolajczyk and Schmid, 2004; Zuliani et al., 2004; Rodehorst and Koschan, 2006), feature matching strategies (Kunio, 2007), hypothesis generators and the verification schemes (Hartley and Zisserman, 2000). Incongruously, comparisons from different authors conclude differently. For example, SIFT features performed the best in (Mikolajczyk and Schmid, 2004), Harris in (Zuliani et al., 2004), and Forstner in (Rodehorst and Koschan, 2006). Another example is the comparison of hypothesis generators (Nister, 2004b) that shows Nister's 5-point solver performing better in sideways motion while the traditional 8-point algorithms (Hartley, 1997) perform better for forward motion. There is another issue at the hypothesis generation phase, namely the choice of the motion model relating the two images. In Figure 5.10, we present the structure of the fundamental matrix for different cases that we face while trying to localize in an unknown unstructured environment using image data.

So, for an arbitrary video sequence, in addition to the risk in the choice of the feature detector, how do we decide which minimal hypothesis generator (as in 5 point, 6 point, 7 point or 8 point solver) to use for the varying and intangible quality of feature correspondence data? Which model best fits the correspondence for a particular feature detector? Do all the feature detectors agree on the structure of the fundamental matrix epipolar relationship?

The observation that the performance of feature detectors, matching algorithms and hypothesis generators are data/scene and motion specific motivates the implementation of our theory developed in Chapter 4. In the following paragraphs, we discuss related work in handling the uncertainty in the fundamental matrix. In Section 5.3.3, we explain the theoretical inspiration for the inference engine and detail the implementation procedure in Section 5.3.4. We demonstrate our work in real world situations both in indoor and outdoor navigation applications and finally summarize our efforts.

Generic	Affine	Perspective	Multiple hypothesis
$F = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}$	$F_A = \begin{bmatrix} 0 & 0 & f_3 \\ 0 & 0 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}$	$F_P = \begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & b_9 \end{bmatrix}$	$F_1 = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \quad F_2 = \begin{bmatrix} g_1 & g_2 & g_3 \\ g_4 & g_5 & g_6 \\ g_7 & g_8 & g_9 \end{bmatrix}$
$\tilde{\mathbf{m}}_i \mathbf{F} \tilde{\mathbf{m}}'_i = 0$	$\tilde{\mathbf{m}}_i \mathbf{F}_A \tilde{\mathbf{m}}'_i = 0$	$\tilde{\mathbf{m}}_i = \mathbf{F}_P \tilde{\mathbf{m}}'_i$	$(\tilde{\mathbf{m}}_i \mathbf{F}_1 \tilde{\mathbf{m}}'_i)(\tilde{\mathbf{m}}_i \mathbf{F}_2 \tilde{\mathbf{m}}'_i) = 0$
General case	Features are far away	Camera rotation/Planar scene	Multiple motion

Figure 5.10: Different forms of the motion model \mathbf{M} for the fundamental matrix relating image frames.

5.3.2 Related work on handling uncertainty in image-based pose recovery

In the well researched area of camera ego-motion estimation, we begin by understanding methods discussed in (Hartley and Zisserman, 2000). Out of the several methods in that book, our focus is on the feature-based 3D geometric inference using a single camera. Mining the literature further, we realize that we are not the first to raise the concern about the uncertainty in the fundamental matrix estimation for pose recovery. Researchers over the last decade (Csurka, 1995; Zhang, 1998; Torr, 1998; Kanatani, 2004) have published ideas for dealing and quantifying uncertainty in the fundamental matrix from a statistical perspective.

The seminal effort in (Zhang, 1998) addresses our specific concern with attention to uncertainty of the fundamental matrix explaining different methods to estimate the epipolar geometry. We use the methods described to estimate \mathbf{F} and its uncertainty along with the normalization procedure *FDiff* to normalize fundamental matrices from the implementation in (Zhang, 1998). Kanatani (2004) proposes the idea of asymptotic analysis for uncertainty modeling and explains how to handle uncertainty in geometric fitting using model selection criterion. In the same paper, Kanatani modifies the Akaike information criterion for geometric inference and argues the need for a geometric extension from statistics. Using a similar model selection tool, Torr (1998) clusters feature matches for motion model determination detecting degeneracy, affine motion and even multiple motions between successive frames. Kanatani's remarks in (Kanatani, 2004) and Torr's future directions in (Torr, 1998) act as inspiration for our work in considering the fundamental matrix estimation process as a random process of sub-systems where the choice of the model \mathbf{M} is as important as its estimated model parameters that make \mathbf{F} .

5.3.3 Theoretical inspiration behind our approach

The significant improvement that we provide over Torr (1998) is that we perceive the fundamental matrix solver as a stochastic process and the correspondence data as the excitation signal to the process. Now, the sub-samples drawn from the large set of noisy correspondence data for model fitting is analogous to the time-shifted signal input for the model-fitting random process to produce the same or equivalent fundamental matrix as output. The n different hypotheses that the solver generates before making a decision is indicating uncertainty on the estimated fundamental matrix in a time-stationary sense.

Our proposed idea to increase the reliability on the fundamental matrix is by generating an uncertainty model in a space-ergodic sense. We excite the random solver with spatially different correspondence data generated by different interest point detectors to generate the ensemble uncertainty samples. Our method then tries to compensate the deficiencies (uncertainty) in time-stationary sense with space-ergodicity (and vice versa) with the objective of reducing the total uncertainty on the fundamental matrix.

This idea underlies our approach to uncertainty management on the fundamental matrix. We recollect that, in estimating \mathbf{F} with noisy $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$ using the hypothesis-test-verify paradigm, there is two inherent embedded problems. The first one is the determination of the model \mathbf{M} and the second, the parameters of \mathbf{M} . While there has been considerable statistical attention on the estimation accuracy of the parameters of \mathbf{M} , the validity for the model choice \mathbf{M} is significantly ignored in the literature. Our approach generates statistics for the model validity using several interest point detectors and combines the model validity with the model fitting accuracy in choosing optimal minimal algorithms based on the scene and motion.

Suppose correspondences $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$ are ideal as in a synthetically generated case, all the n different hypotheses would lead us to the same or equivalent \mathbf{F} . But, in the noisy real world, robust model-fitting or hypotheses verification paradigms (inspired by RANSAC) can iterate to an estimate of \mathbf{F} that is sub-optimal or iterate to a non-acceptable result because of the outliers. If the quality of feature matches is good, the distribution of the model parameters during the iterative procedure will be confident and tightly bound. We model this convergence of the fundamental matrix parameters towards learning the optimal choice of methods that will lead to increased confidence in \mathbf{F} .

We illustrate this idea in Figure 5.11. We analyze the statistics of different hypotheses (\mathbf{f}) that minimal sub-samples generate given a set of correspondence data and compare the performance with the statistics generated during the convergence from correspondence data generated by different spatial feature point detectors. The convergence in Figure 5.11 (a) encapsulates uncertainty in the parameters of \mathbf{f} through competing hypotheses and the convergence in Figure 5.11 (b) emphasizes the reliability in the model for \mathbf{f} . To the extent possible, we desire the estimate of \mathbf{F} to be both confident and reliable. We describe the statistical procedure and the implementation details in the following section.

5.3.4 Learning the optimal minimal algorithm

The quality of feature correspondences $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$ contains two types of errors, localization errors and gross classification errors. The bad matches also called outliers infiltrates uncertainty into the model estimation and fitting process. Our experience with real world environments is that the variance bound in Equation 5.3, needs to be adjusted drastically in images with nice structural features and ones with poor structure. Even robust outlier rejection methods like MLESAC (Torr and Zisserman, 2000) are not able to handle such situations mostly because of the infinite possibilities to consider in real world dynamic situations. We will let the commonly implemented the model fitting method to consider modeling the error between re-projected interest points, while we consider the iterations of the consensus indicative of different hypothesis on vector form \mathbf{f} of \mathbf{F} as our uncertainty sample.

A good feature detector, matching algorithm, and a robust hypothesis generation algorithm would produce equivalent fundamental matrices \mathbf{F} with a majority of the n minimal hypotheses evaluated within the inliers bound. In other words, the confidence in the distribution of \mathbf{f} measures the uncertainty over generating the model using the corrupt correspondence data. Then, we look at the parameters generated by different pipelines (choice of different feature detectors, hypothesis generation schemes and motion models) and evaluate the different estimates of \mathbf{F} for model support or reliability. This situation draws direct analogy with the selfish-altruist data fusion scheme described in Chapter 4.

The vector form \mathbf{f} is the state vector and different pipelines from Figure 5.7 are the sensors measuring \mathbf{f} at different uncertainty levels. However, the difference is the fact that the dimensionality of \mathbf{f} , (which depends on the choice of the model for the perspective relation between image frames) is not known and has to be learnt from the uncertainty samples. By observing the convergence to the parameters using different feature detectors and by also observing the statistics of the models generated by different hypothesis generators, we implement a variant of the theory developed in Chapter 4 to learn and minimize the uncertainty about the fundamental matrix.

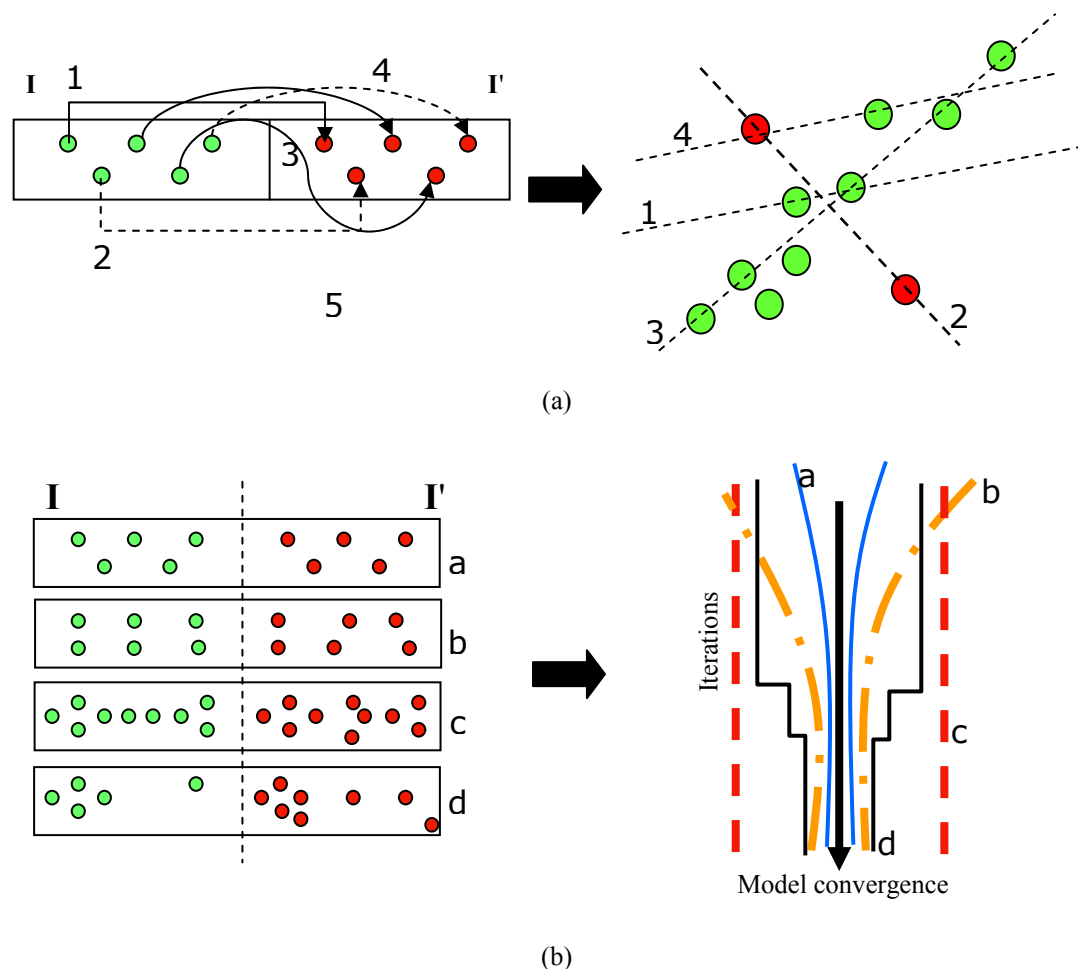


Figure 5.11: Our approach to improve the reliability in the fundamental matrix by generating convergence uncertainty statistics. (a) The correspondence data as it iterates to the parameters of f seeking inlier support. We expect the converged result to be hypothesis-“3”. The uncertainty about f in the iterative convergence produces the time statistics. (b) The convergence of the parameters of f from different feature detectors. The convergence on the parameters of F , provides the space statistics. For reliable estimation of F , our framework helps choose the features that converge faster and with greater confidence adaptively and automatically.

Our proposed algorithm for adaptively leaning optimal methods assumes the implementation of methods in the shape from motion pipeline in Figure 5.7. With that implementation, we present the pseudo code of our procedure towards minimizing the uncertainty about the fundamental matrix below.

- Step 1:** **For** each potential competing pipeline of methods \mathbf{P}_i listed in Figure 5.7 where $i = 1, 2, 3, \dots, k$
- a) Use RANSAC and iterate to a convergence. Collect d -estimated parameters \mathbf{H} of model \mathbf{M} fitted during the iterations of RANSAC.
 - b) Estimate d -variate probability distribution \mathbf{B}_i based on j ($j > 30$) iterations of parameter estimates ($\mathbf{H}_1, \dots, \mathbf{H}_j$) collected.
- End**
- Step 2:** Score Correspondence Outlier Consensus (\mathbf{COC}_i) using the model selection criterion in Section 4.1.
- Step 3:** Compute Model Consensus Score (\mathbf{MCS}_i) by evaluating competing distributions \mathbf{B}_i as a multi-sample clustering problem in Section 4.2.
- Step 4:** Choose the optimal pipeline \mathbf{P}_i with minimum $\mathbf{COC}_i + \mathbf{MCS}_i$.
- Step 5:** Repeat Steps 1-4 every m frames

As mentioned earlier, we are interested in the RANSAC convergence consensus and quantifying the confidence within the convergence process. Particularly, our interest is in identifying the pipeline that is indicative of maximum likelihood of the fundamental matrix parameters with minimum uncertainty, or in simpler words \mathbf{B}_i with minimal variance. This can be mathematically expressed as the minimizer of the criterion (Equation 5.4) that simultaneously considers the likelihood and also penalizes the uncertainty associated with the likelihood of the parameters of model \mathbf{M} .

This model selection criterion in the statistics literature (Bozdogan, 2000) is popularly known as *ICOMP* and derives from the Kullback-Liebler (KL) distance between estimated and unknown underlying probability density. This criterion in theory is inspired from the same source as Kanatani's geometric AIC (Kanatani, 2004) but is able to include the covariance of the model parameters. Without much modification, we are able to apply this criterion in evaluating the confidence in the model fit after the iterative convergence of RANSAC. We note that Equation 5.4 does not involve distributional assumptions and can be applied to even Parzen window estimates of \mathbf{B}_i . The parameters \mathbf{f} and $\hat{\Sigma}_F$ in Equation 5.4 can be computed using one of several methods in (Zhang, 1998), though in our implementation, we use the moments of the distribution \mathbf{B}_i . The correspondence outlier consensus (\mathbf{COC}) is given by:

$$\mathbf{COC} = -2\log(\text{Likelihood of } \mathbf{f}; \mathbf{M}) + 2C_1(F^{-1}(\hat{\Sigma}_F)) \quad (5.4)$$

where F^{-1} is the inverse Fisher information matrix, The C_1 measure and the F^{-1} is computed using Equations 5.5 and 5.6.

$$C_1(F^{-1}(\hat{\Sigma}_F)) = \frac{s}{2} \log \left[\frac{\text{tr}(F^{-1}(\hat{\Sigma}_F))}{s} \right] - \frac{1}{2} \log |F^{-1}(\hat{\Sigma}_F)| \quad (5.5)$$

with s being the rank of F^{-1} , $|\cdot|$ refers to the determinant and tr refers to the trace of the matrix.

$$F^{-1}(\hat{\Sigma}_F) = \begin{bmatrix} \hat{\Sigma}_F & 0 \\ 0 & \frac{2}{j} D_p^+(\hat{\Sigma}_F \otimes \hat{\Sigma}_F) D_p^+ \end{bmatrix} \quad (5.6)$$

with D_p^+ being the Moore-Penrose inverse of vectorized $\hat{\Sigma}_F$, \otimes representing the Kronecker product. The C_1 measure for penalizing uncertainty is obtained by maximizing mutual information in d -dimensions.

We direct the reader to (Bozdogan, 2000) for sampling bias compensating implementation details on the finite sampling form of Equation 5.4. Equation 5.7 is the reduced form of Equation 5.4 for a normal distribution \mathbf{B}_i generated using n_i samples of converging \mathbf{f} stored as y . We would like to note that by choosing the pipeline with minimal **COC** score, we are able to learn the choice of methods, which require lesser number of hypotheses to evaluate at the same time penalize for the uncertainty in the estimated parameters.

$$\text{COC} = \frac{s}{2} \ln \left(\frac{\text{tr}(\hat{\Sigma}_{\mathbf{F}})}{s} \right) - \frac{1}{2} \ln |\hat{\Sigma}_{\mathbf{F}}| + \left(\frac{n_i d}{2} \log(2\pi) + \frac{n_i}{2} \log |\hat{\Sigma}_{\mathbf{F}}| + \frac{1}{2} \text{tr}(\hat{\Sigma}_{\mathbf{F}}^{-1}) \left(\sum_{i=1}^{n_i} (y_i - \mathbf{f})(y_i - \mathbf{f})' \right) \right) \quad (5.7)$$

Though we could stop with minimizing **COC** which is a minor improvement over (Torr, 1998), we note that the choice of the model \mathbf{M} is a significant factor in the fitting process. For example, in an urban environment one feature detector could produce correspondence data that is tracking features far away like the sky and the clouds, while another might be tracking multiple objects such as cars moving in the scene. In the former case, we will end up choosing \mathbf{M} with an affine structure with 5 parameters, and in the latter case up to 18 parameters on 2 fundamental matrices. To avoid this risk in the choice of the model itself, we will quantify the model support from competing pipelines. The model consensus score is obtained by evaluating the data from several pipelines \mathbf{P}_i in a multi-sample clustering framework using Equation 5.8. The idea is to cluster \mathbf{B}_i of each \mathbf{P}_i using information distance between the distributions inspired by the method described in (Bozdogan, 1986). The model consensus score (**MCS**) is as shown below.

$$\text{MCS} = -2 \log(\text{Likelihood of } \mathbf{M}) + 2g \quad (5.8)$$

where g penalizes for the parameter parsimony for \mathbf{M} .

We begin by testing for the first three hypotheses listed below and in the occurrence of Case 2 or 3 alone, evaluate combinatorial subsets in Case 4. The values of g for different cases are listed in Table 5.1.

Table 5.1: Parameter parsimony penalty in the model consensus score.

Cases	g of \mathbf{M}
Case 1: All models \mathbf{M} equivalent (\mathbf{f} and $\hat{\Sigma}_{\mathbf{F}}$ are equal for all competing \mathbf{P}_i) Cluster : $\{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_k\}$	$d + \frac{d(d+1)}{2}$
Case 2: Different \mathbf{M} (Different \mathbf{f} but equal $\hat{\Sigma}_{\mathbf{F}}$ for competing \mathbf{P}_i) Cluster : $\{\mathbf{P}_1\} \{\mathbf{P}_2\} \dots \{\mathbf{P}_k\}$	$kd + \frac{d(d+1)}{2}$
Case 3: All \mathbf{M} 's are different for competing \mathbf{P}_i (Different \mathbf{f} and different $\hat{\Sigma}_{\mathbf{F}}$ for competing \mathbf{P}_i) Cluster : $\{\mathbf{P}_1\} \{\mathbf{P}_2\} \dots \{\mathbf{P}_k\}$	$kd + \frac{kd(d+1)}{2}$
Case 4: Cluster of κ pipelines of k competing \mathbf{P}_i are equivalent (κ \mathbf{f} 's are equivalent) Cluster: $[\{\mathbf{P}_1\} \{\mathbf{P}_2\} \dots \{\mathbf{P}_\kappa\}] \dots [\{\mathbf{P}_{k-1}\} \{\mathbf{P}_k\}]$	$\kappa d + \frac{\kappa d(d+1)}{2}$

The way we interpret these hypotheses is that when all pipelines are essentially leading us to the equivalent fundamental matrix, fusing correspondence data from two feature detectors for example, is not altering the convergence process or the end result. In such cases, the model fitting process can be inferred as stationary and ergodic with high reliability. We evaluate this condition in Case 1, where we treat all samples contributing to \mathbf{B}_i of each \mathbf{P}_i as drawn from a single distribution $\mathbf{B}_{\text{consensus}}$. This case usually occurs when we have interesting features like corners in buildings to track.

However, our method is particularly useful only when Case 2 or Case 3 occurs. Case 2, tells us that the choice of the pipeline is influencing the estimate of \mathbf{F} , and that we might have to be satisfied with the pipeline with minimal **COC** score. Case 3 forces us into the balancing act of which model and whose parameter estimates correspond to minimal uncertainty. We assign the minimal value of Equation 5.8 for the three hypotheses as **MCS** (\mathbf{P}_i). Then, we try to learn the most efficient model by evaluating all combinatorial subsets for this purpose in Case 4, which is actually a nested form of Case 1, 2 and 3. We evaluate all possible clusters (subsets) and assign the minimum value of **MCS** over all clusters only to the κ pipelines with high model support.

This procedure is comparable to modeling the convergence consensus into a mixture model and finding the order of the mixture similar to previous effort (Torr, 1998), the difference being that we can now accommodate for the risk in choosing a particular pipeline. The **COC** score quantifies the risk in the estimation and the **MSC** score quantifies the risk in the model. Our next step is to combine them with the objective of improving the reliability on the process of fundamental matrix estimation.

The total uncertainty score U_F that we will use to make a decision about the choice of algorithms combines two probabilities (uncertainties).

$$U_F(P_i) \propto \Pr(\mathbf{f}; \mathbf{M}, P_i) + \Pr(\mathbf{M}; \mathbf{P}_j; j = 1, 2, 3 \dots k) \quad (5.9)$$

$$U_F(P_i) = \text{COC}(P_i) + \text{MCS}(P_i) \quad (5.10)$$

Both the **COC** and **MCS** scores are constructed using log-likelihood approximations of information complexity and the addition of these scores to decide on optimal P_i is intuitive. In addition, **MCS** and **COC** also provide an extra benefit of informing us, if we actually can improve the uncertainty over the fundamental matrix if we fused several pipelines instead of selecting pipelines. Since the **MCS** computation already considers the reliability aspect of fusion, we are able to note that if **MCS** of selected pipeline P_i is less than its **COC** score, we can improve by fusing the κ pipelines in the maximal cluster. We make this statement following the principles in the study of fusers better than the best sensor in (Rao, 2001). However, in most real world situations that we experiment in the following section, pipeline selection proves to be the promising direction.

In the following paragraphs, we consider navigation applications in indoor and outdoor environments. We present results in two scenarios using mobile robotic platforms. The first scenario is where the switch to better features helps self localization and the second scenario where the quality of features is not guaranteed and we have to learn to use the appropriate hypothesis generator adaptively. Our robotic platforms have direct pose recovery instruments either in the form of a laser range scanner and/or GPS/INS to provide us with a ground truth on the pose. We will demonstrate our proposed framework adapting to monocular depth cues as the scene changes and also guiding the switch of hypothesis generators by learning the quality of the feature correspondences. We compare the results from image-based pose recovery with the hardware-based localization to show the effectiveness of our approach in opportunistically switching features towards better reliability on pose recovery.

We show our test area around Ayres Hall, the mathematics department building at the University of Tennessee that we had experimented earlier in Chapter 2. We deployed our vision-based mobile mapping robot letting the system track nice corners on the building and later encountering vegetation. The reason behind this experiment was to prove that our approach can guide the choice of better feature detectors as the monocular depth cues change in the image.

We have shown the panorama generated using a separate set of images along with the corresponding video frames below the panorama in Figure 5.12 (a). The localization result along with the GPS ground truth in a 7 m path in that area shown in Figure 5.12 (b) compares the deviation in vision-based pose recovery with our adaptive approach to using a single feature detector. We are able to see the drift when the vision-based inference lost linear perspective from the scene. By switching interest points in the scene, we are able to reduce the uncertainty and the drift in localization. In Figure 5.12 (c), we show the Monte-Carlo result of one switch in this dynamic environment to emphasize the uncertainty management. The graphs Figure 5.12 (d) (e) and (f) indicate the reliability on the recovered pose after 100 iterations of convergence on the same frames but using different feature detectors. With these results, we are able to demonstrate that our method chooses feature detectors that guarantee reproducibility with better accuracy in pose recovery. From Figure 5.12 (c), we deduce that, if we did not guide the switch, there is 50% greater chance that we did not estimate the optimal fundamental matrix.

For the indoor case, our robot was equipped with a camera and SICK laser range finder and was intended to traverse a corridor as shown in Figure 5.13. During the course, the robot had to maneuver at different orientations during which our algorithm decided to use 5 point algorithm or 8 point algorithm based on the data. For comparison sake, on offline processing of the image data using each one of the algorithms separately, we realize the significant error we would have accumulated had we not countered for Nister's dilemma in choosing between 5 point and 8 point hypothesis generators (Nister, 2004b). In Figure 5.13, we also tabulate the average error in recovered pose per frame following the definition of error provided in (Nister, 2004b) using the essential matrix.

Our next experiment was to evaluate the robustness of the framework proposed in this paper. We selected 16 video sequences (Figure 5.14), each with 2000 frames containing both man-made structures and natural vegetation. We show a plot of the average uncertainty score on each of these datasets with our adaptive method and a standard method (Harris+ 8 point + RANSAC) in Figure 5.14. From Figures 5.14 (b) and (c) we are able to visualize how the uncertainty management using our framework translates to reliability on the fundamental matrix. Our method is able to do as good as the standard method when the standard method is sufficient, and is able to further improve the reliability in other situations. Thus far, we presented results on real world environments both indoors and outdoors using the same mobile platform and pre-programmed vision module. In the experimental phase we made two interesting observations. The first observation was that, when we replaced our high resolution camera with a noisy cheap one, our method automatically chose a different set of feature detectors and hypothesis generation algorithms adapting to the noise characteristics of the cheap camera.

The second observation exposed a caveat. We realized that the time that we gain by minimizing the number of hypothesis to evaluate using our proposed method depends on how many competing models we evaluate and how often do we evaluate for a scene change. Typically in our experiments, our inference engine feeds the vision system once every 300 frames to operate real-time. A drastically changing scene environment or a fast moving mobile robot might require more frequent updates demanding more computational overhead. In Figure 5.15, we show the overload our framework requires compared to believing one single feature detector and one estimation algorithm. Figure 5.15 (a) is a timing comparison analysis with and without our uncertainty handling scheme. We are able to see that the switch to the optimal features saves time while the statistics generated for guiding the switch is cumbersome depending on how many pipelines we seek for model support. However, the graph in Figure 5.15 (b) is encouraging showing that our proposed method is not going to challenge real-time operation when evaluating 20 pipelines every 50 frames.

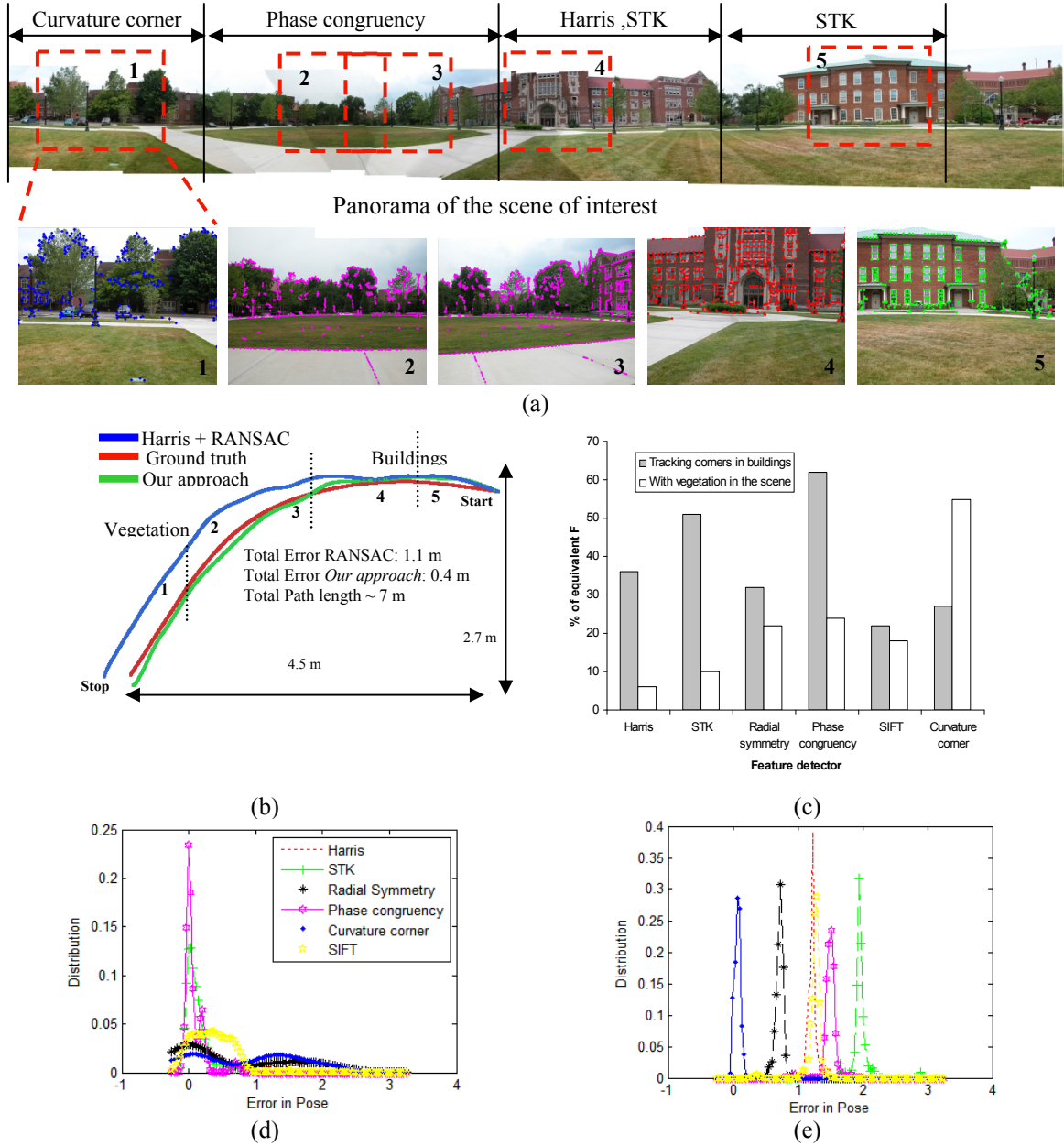


Figure 5.12: Our mobile robotic platform with vision-based pose recovery navigates near the Math building in our University. (a) Our method guiding the switch from one feature detector to another as monocular depth cue changes from buildings to vegetation in this challenging scene of interest. (b) Localization result and comparison on a 7m path. (c) Percentage of equivalent F generated in 100 trials in scenes with vegetation and buildings to show how much improvement we can expect by choosing the optimal feature points. (d) The Monte-Carlo analysis on the recovered pose and the error plotted as a distribution when tracking good corners. (e) Monte-Carlo analysis on the recovered pose when encountering vegetation.

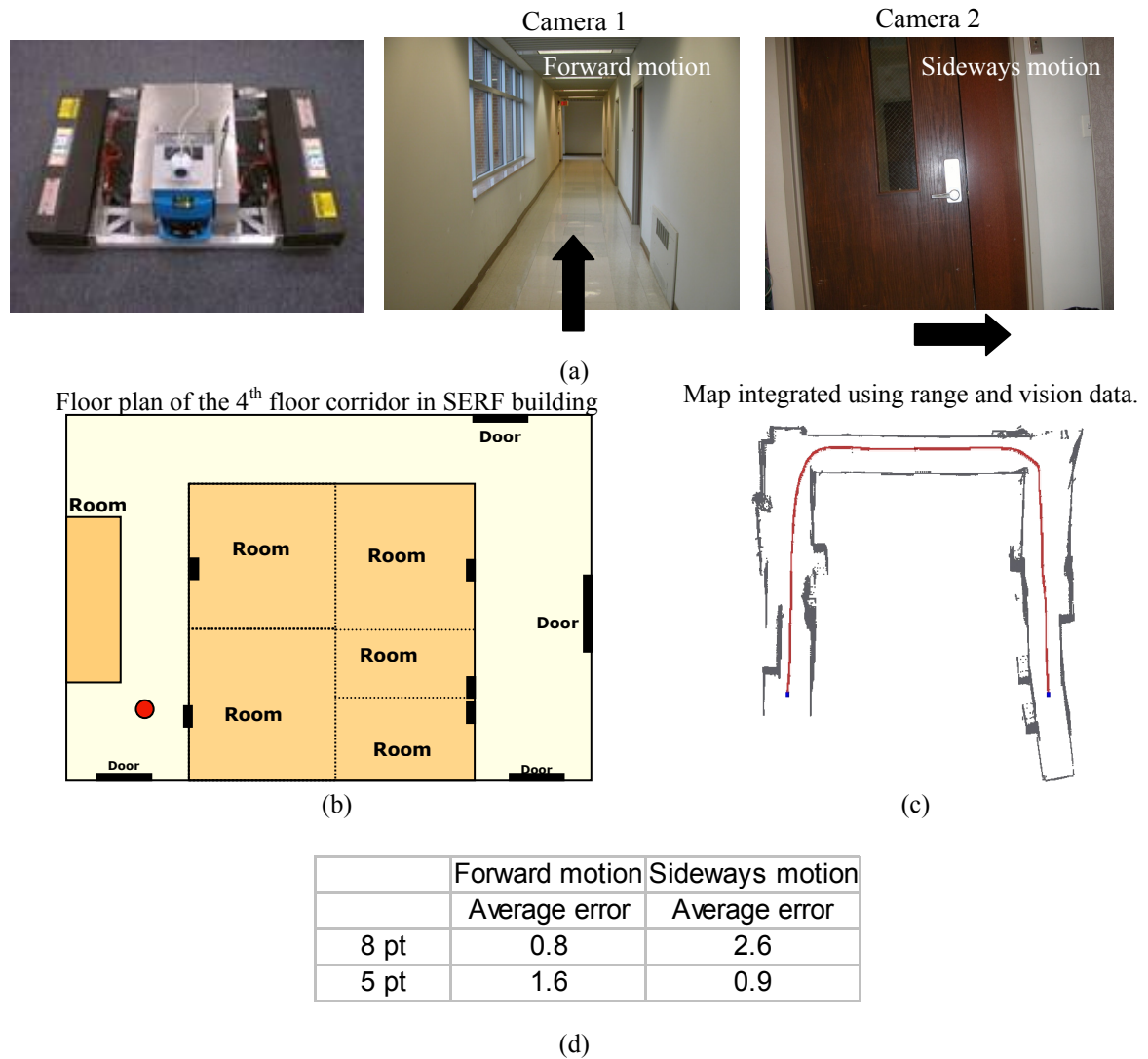
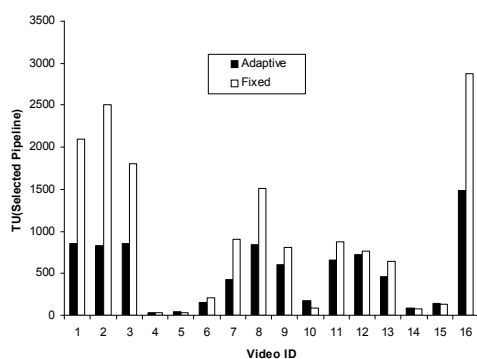


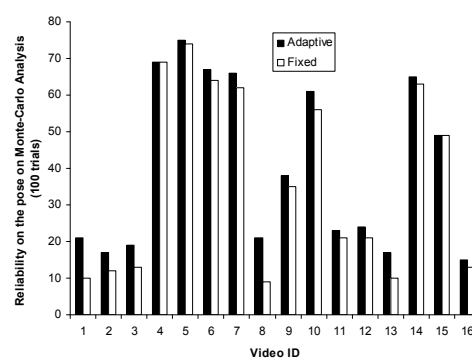
Figure 5.13: Autonomous robot localization with a laser scanner and a vision sensor. (a) The robot trying to autonomously map a corridor at the Science and Engineering building at the University of Tennessee, Knoxville. The localization was performed based on the range and vision measurements. (b) The schematic of the floor plan. (c) The map in black and the localization result (intended path) for mapping. (d) By processing the visual data using our approach we tabulate error results in the frames to show that by switching methods, we can maintain the uncertainty levels across frames during real time localization.



(a)



(b)



(c)

Figure 5.14: Vision-based localization in dynamic environments. (a) Sample frames from several video sequences considered in our experiments. (b) The average uncertainty score after choosing optimal algorithms vs. using Harris corners + RANSAC + 8 pt algorithm. Our approach shows significant improvement over believing on a single pipeline (c) Reliability (Average count on convergence on true pose between frames) on the fundamental matrix after conducting the same experiment 100 times.

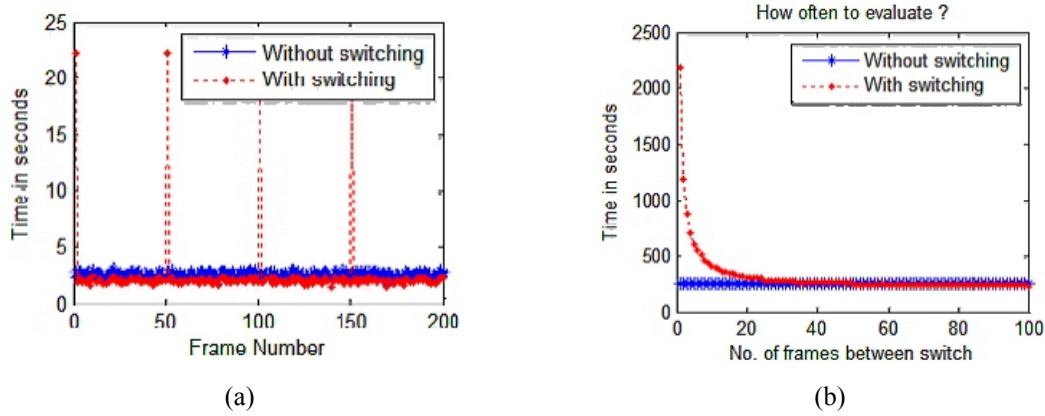


Figure 5.15: Timing analysis for learning competing models. (a) How much of a burden depends on the number of competing pipelines to evaluate. For this particular case, we had 20 competing pipelines, evaluated once every 50 frames. (b) This graph indicates using our framework every 50 frames is not as much as a burden compared to the pose recovery.

We proposed a statistical procedure for uncertainty management when dealing with images from a calibrated camera in pose recovery using the fundamental matrix. We demonstrated results on pose recovery for navigation applications in indoor and outdoor environments. Our method reduces the uncertainty in the convergence of the parameters of the fundamental matrix with the uncertainty of the model itself. The different models generated in the hypothesis-test framework of pose recovery provide the statistics for the confidence in the matches while the models generated by different feature detectors provide the statistics for quantifying the uncertainty in the model. By combining both these uncertainties, we have formulated a generic scheme using model selection theory that will help us choose methods for reliable estimation of the fundamental matrix at the same time acting as a performance measure of pose recovery using image features.

5.4 Summary

We have implemented and demonstrated the intelligence for opportunistic sensing as the ability to evolve a belief policy on sensors within a multi-agent framework to survive/counter concerns of failure in challenging operating conditions. The challenge was to deal with heterogeneous sensing mechanisms functioning at varying degrees of performance and confidence while operating in dynamic conditions. The interaction of the sensors or the sensed object of interest with the environment introduced performance degradation issues were sensing paradigms degenerate to accuracy levels lower than the design specification. By implementing the theory developed in Chapter 4, we have provided a sensor adaptation solution focusing on the localization and mapping problem with mobile robotic platforms. Our objective function for sensor adaptation was formulated as that of ensuring success of a multi-agent system by maximizing information gain (income) considering each sensing unit individually and in synergy, while reducing the risk and cost of information corruption avoiding failure. With our experiments on mobile robotic platforms on both the instrumentation-based and image-based systems in dynamic large scale environments, this Chapter validated the theory further encouraging the potential for application in several critical real world scenarios.

6 Applications: Research Impact

In Chapters 3 and 4, we presented the modular architecture underlying our system design and the methods derived to operate amidst uncertainty in dynamic environments. Rigorous experiments in Chapter 5 further encouraged the real world deployment of our work. This chapter showcases the application of the technology developed thus far in critical safety and security scenarios introduced earlier in Chapter 1. For each of those motivating applications, we brief the state-of-the-art, enlist the shortcomings and emphasize the improvements that our method provides over existing commercial systems. Section 6.1 deals with under-vehicle inspection, Section 6.2 is on road pavement/runway crack inspection and Section 6.3 documents the deployment for large-scale urban and terrain mapping at high accuracy.

6.1 Under-vehicle inspection

In this section, we demonstrate the convergence of 3D sensing technology, modular mobile robotics and computer vision-based automation towards under-vehicle inspection and present the technology ready for deployment at gate-entry terminals, check points and parking lots with military and civilian interests.

6.1.1 State-of-the-art methods

The first idea implemented and marketed for under-vehicle inspection was to use a mirror at the end of a stick as illustrated in Figure 6.1. The mirror-on-a-stick inspection though deters potential attacks, is primitive and exhibits some drawbacks. First, if the security personnel are slow in detecting the bomb, then they are still vulnerable to detonation before the inspection is complete. This weakness is of major concern as the security personnel must be within physical proximity of the vehicle. Second, the physical constraints of the mirror only allow about 40-50% coverage of the vehicle undercarriage. The center line of the vehicle in particular is difficult to reach, and the subsequent viewing angles are oblique. Also, the mirror-on-the-stick system though inexpensive, is only a visual inspection scheme and does not readily lend to archival and automation.

With the mirror-on-the-stick approach proving to be not so efficient with many limitations, the next logical evolution was the “buried sensors” approach. The idea was to embed cameras under ground and acquire images as the target vehicle drives over the sensor suite. The image data is further processed and visualized as high resolution mosaics (Dickson et al., 2002). With such a “buried sensors” approach, the motion of the target vehicle determines the resolution, coverage and completeness of the data required for inspection. Also, the embedded sensor approach consumes more time for operation and maintenance, leading towards the development of mobile robots such as the omni-directional inspection system ODIS (Freiburger et al., 2003) and Spector (Autonomous Solutions Inc., 2005). These low-profile robots mounted with cameras and capable of sneaking under cars and trucks recording images as video frames have proven to be a significant improvement over the mirror-based systems (Smuda et al., 2005). Due to the low clearance of most cars, the field of view using a single camera, as seen in Figure 6.2, becomes too restricted that some components cannot be discerned from the live video stream even by a human inspector. Using several cameras or omni-directional cameras is an option that does not help much in visualizing the data in a meaningful form at the resolution required to identify components.



Figure 6.1: The traditional method for inspecting a vehicle undercarriage is to use a mirror attached to the end of stick. This mirror-on-a-stick approach enables security personnel to search wheel wells and other vehicle cavities. On the right, we show a picture of the scene the inspection personnel views on the mirror.



Figure 6.2: An alternative to a mirror-based inspection is a low-profile mobile robotic platform (called SafeBot) that can navigate under a vehicle while inspection personnel remain at a safe distance. On the right, we see a single frame from the camera on the robot. Due to low ground clearance of the automobile and restricted field of view of the camera on the robot, a single frame only shows a small section of the undercarriage.

Both the robotic approach and the “buried sensors” approach lend to remote inspection and flexibility for archival and automation. But the robotic solution is favored over “buried sensors” approach because the acquired sensor data is only dependent on the robot’s motion (which is measurable or can be controlled by the remote inspector) instead of the target vehicle’s motion. The robotic solution also lends easily to multiple inspections of the same vehicle and enables the inspector to focus on a particular area of interest on a case-by-case basis. Several robots have hence been designed and among several enhancements on the preferred robotic solution, mobile robots with intensity cameras were made independent of illumination and hence operable even during the night by mounting light sources or using night vision cameras or thermal sensors. We show some examples from the data that we collected using such a robotic platform with visual and thermal sensors (Koschan et al., 2004) in Figure 6.3. We observe that the mosaic generated using the images acquired in a single pass provides more coverage compared to the mirror-on-the-stick approach, at the same time, suggesting the need for a better system to maximize visual information from a single pass using as few sensors as possible. Our proposed 3D approach is one such enhancement that also fills most of the other gaps observed with the contemporary systems.

6.1.2 Requirements

The state-of-the-art systems emphasize that the visual cameras alone may not be sufficient for a robust under-vehicle inspection system. In addition to visual information, we see that a thermal sensor can help detect components that are not part of the exhaust system, chemical sensors can help sense explosives in the scene and nuclear sensors can detect radioactivity. The multi-modality sensor data can aid a human inspector and also contribute towards potential automation for the detection of bombs, explosives and contraband. The robotic mapping multi-sensor approach will put the multi-modality measurements in a spatially meaningful perspective reducing the burden of interpretation on the human inspector. We desire implementing such intelligence with the inspection robots reducing the susceptibility to error in any way possible.

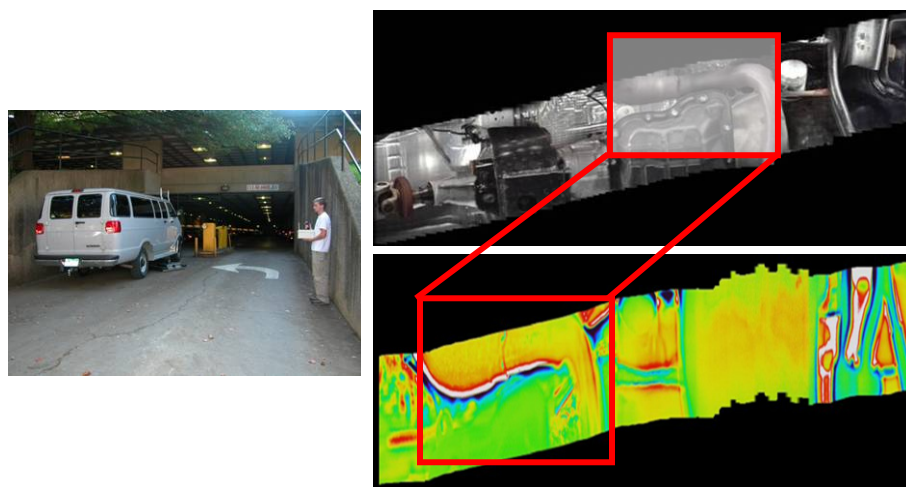


Figure 6.3: The images on the right are high resolution mosaics of visual and thermal data from the robot scanning the carriage. The thermal mosaic at the bottom is color coded, with the hot regions appearing in red and the colder regions in blue. The red boxes show the thermal scene characteristics corresponding to the visual scene. The mosaics were generated using a phase-correlation-based method by (Koschan et al., 2004).

The major contribution with our research is in enhancing existing inspection robots with 3D sensors to acquire geometric information of the under-vehicle scene and hence using the 3D data for automatic detection of anomalous objects in the scene as potential threats. In trying to provide such a robotic system as a reliable solution for under-vehicle inspection, we enlist the following characteristics expected of an under-vehicle inspection system:

- Ability to acquire and communicate data from a safe stand-off distance.
- Maximal coverage with minimal effort reducing the number of inspection iterations.
- Independence from environmental conditions like illumination, temperature, humidity etc.
- Rendering data in a visual and spatially understandable form for a remote inspector to make decisions.
- Flexibility for digital archival and automatic threat detection using simple and efficient algorithms.
- Inexpensive and less cumbersome maintenance and operation.

6.1.3 Results

We deployed two prototypes using two different 3D scanners as different sensor configurations for the SafeBot. We mounted a laser range finder (SICK LMS 200) and a laser-profile scanner (IVP Ranger SC-386) on mobility platforms to acquire the under-vehicle scene data. The mobile robots in both 3D sensor configurations were capable of scanning most automobiles with enough ground clearance for our robotic imaging system to slip under. We have tested our robot under several cars inside parking lots, collecting and visualizing data almost in real time. We present some of those results in Figure 6.4. Using the time-of-flight prototype we are able to model the complete structure information of the undercarriage with a single pass with the geometric accuracy in the order of a few (1-5) centimeters.

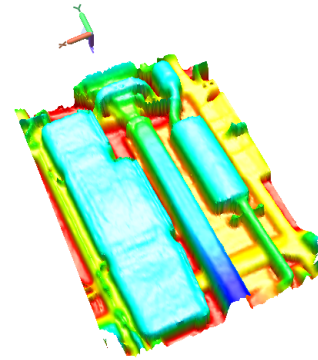
A single pass along the center line of the vehicle takes about 30 seconds to map the entire undercarriage and the observed accuracy varies with the ground clearance of the automobile. We attribute the reason for the below par accuracy to the limitation in the timing electronics of the scanner. The time-of-flight systems also require a minimum stand-off of 25 centimeters at which we get 5-10 millimeters of accuracy which is good enough resolution for the threat detection. With the IVP range profiling sensor, we are able to achieve depth accuracy of 2-5 mm, but with limited field of view, requiring multiple passes under the vehicle and creating an issue of concern with view occlusions.

Having presented the 3D data of the under-vehicle scene, we demonstrate how the 3D data can be used as a visualization bed for multi-sensor data from sensor bricks within the architecture and also localizing potential threat information from nuclear, chemical or biological detectors. The results that we will discuss in this section are based on a simple experiment that we conducted on the Dodge Stratus car. We attached two radioactive Cesium sources under the car, and then used our SafeBot with multiple sensors, to scan the vehicle. We measured the radioactivity in 16 locations around the car seen as blue squares in the Figure 6.5.

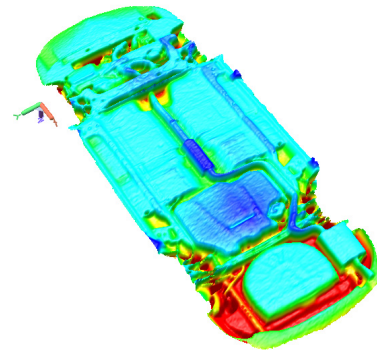
Using a source localization algorithm similar to (Hayes et al., 2002), we were able to determine the number and the location of the radioactive sources in the 3D scene of the undercarriage. Such a capability is of significant assistance to a remote inspector to localize a threat in a physical and spatial sense and act on the potential threat accordingly. Furthermore, we are also able to provide a remote inspector the ability to focus on a particular region and visualize the visual image data and thermal data of the scene on the 3D model from a graphical user interface implemented within this framework. We have shown them separately on the image for clarity reasons though these datasets can be visualized as texture as shown in Figure 6.6. The images from the color cameras are textured on the geometry generated by aligning the profiles from the laser scanners in Figure 6.6.



(a)



(b)



(c)

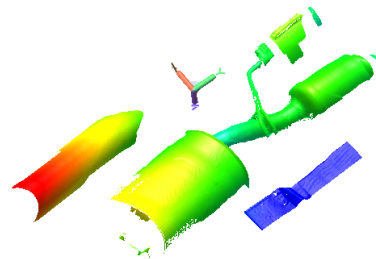


Figure 6.4: 3D under-vehicle inspection using our prototypes. (a) Photograph of the under carriage of the Dodge RAM van and the rendered 3D scene acquired using the time-of-flight system. The 3D models are color coded to emphasize depth. (b) A complete scan of a Dodge Stratus car. (c) Result from another vehicle using the laser profiling sensor configuration.

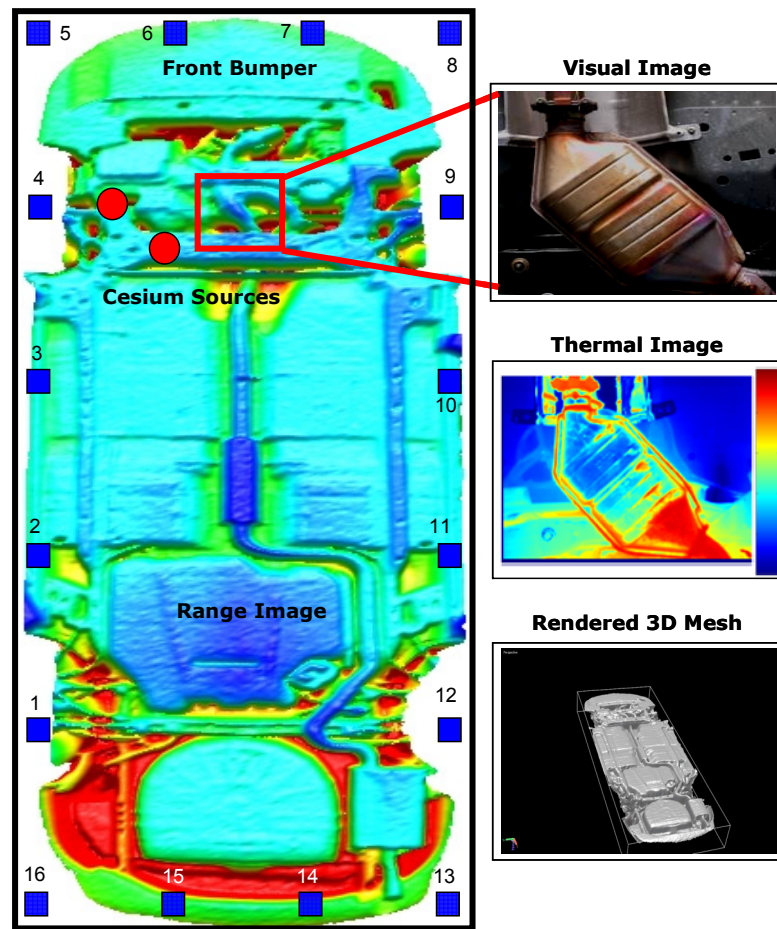


Figure 6.5: Visualization of multi-sensor data on the 3D geometry helps isolate radioactive sources in the scene. We are also able to spatially relate the under-vehicle scene to visual and thermal data. These images show that the data collection using mobile robots and the visualization in a 3D environment brings together the functionality of each of the visual, thermal and range bricks into one single interface for the remote inspector for easy manipulation and interpretation with the multi-sensor data adding to extra scene intelligence.

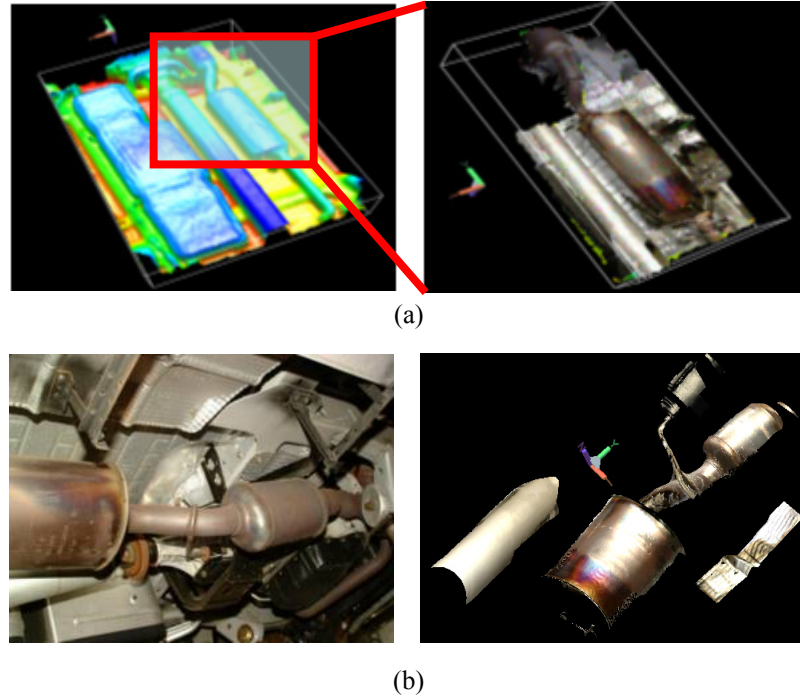


Figure 6.6: Both the 3D scanners lend to easy texture mapping. (a) Texture mapped 3D data from the SICK scanner. Though the entire under carriage is mapped in 3D, the field of view available to the camera on the robot does not span the entire scene in a single pass. Hence, from a single pass, only a part of the scene can be visualized as textured 3D models. (b) Texture mapped on the IVP data. The IVP laser scanning system is a camera-based system that was used to compute range and color maps simultaneously.

We have thus far discussed the construction of the robotic platforms with interchangeable sensors capable of real-time acquisition, processing and integration of multi-sensor data visualized as 3D virtual environments. In the following paragraphs, we will present potential methods of using the 3D information towards scene inference. We will explain a method for scene verification using a popular 3D scene registration algorithm as a significant automation improvement over existing methods.

6.1.4 Improvements over the state-of-the-art

Let us consider the following scenario with John Q. Citizen who works as an analyst at a secure facility for some three-letter agency (TLA) within the U.S. government. John drives to work each day and passes through a security checkpoint to enter the facility. The TLA sticker that John has placed in the lower corner of his windshield validates his access. As he approaches the gate, the TLA security personnel observe the appropriate sticker and wave John into the secure area. This procedure is typical when terrorist threat levels are low, but when threat levels rise, TLA policy requires that security personnel check for the sticker and additionally inspect John's vehicle a little more thoroughly. The assumption is that John is a trusted employee and that his sticker is valid, but the inspectors are looking for bombs or other contraband that may have been hidden on the vehicle without John's knowledge. Essentially, John, and more precisely, his vehicle is a target of opportunity when the vehicle is outside the TLA facility. When John drives to a restaurant for lunch, his vehicle sits in the parking lot where the sticker advertises his association with TLA.

A terrorist might exploit this opportunity by planting a bomb under the vehicle and then waiting for John to return to the TLA checkpoint. At the appropriate moment, the terrorist could remotely detonate the bomb and thereby kill John and the security personnel. The loss of life and the associated destruction would compromise the perimeter security of the entire TLA facility. The 3D scene verification approach applies directly to such a scenario that assumes that we already have a scan of the underside of the automobile, and we are looking for potential modifications made to the undercarriage from the previously archived 3D scans. This approach also extends to the scenario of inspecting unattended vehicles in war zones that may be tagged with improvised explosive devices. We demonstrate the difference shell idea applied to this scenario in Figure 6.7.

A black electronic board that simulates a threat object was attached to the undercarriage of the Dodge van. The vehicle underside was then scanned using our robotic platform. The previously archived 3D dataset was then aligned with the freshly scanned scene with the simulated threat using our implementation of the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992). Once ICP converged on the best possible alignment result, we computed the difference between the two aligned shells.

We show the difference shell in the color coded image Figure 6.7 (d) that highlights the areas of the largest difference indicating the location of the simulated threat object. This approach can be used to detect arbitrary modifications made to the car, like a missing muffler for example. Such an automated approach provides reliable change detection which would have been extremely challenging even for a well-trained human inspector. This ability to quickly, reliably and automatically identify modifications in the undercarriage is a significant additional feature with 3D sensors compared to the traditional camera approach.

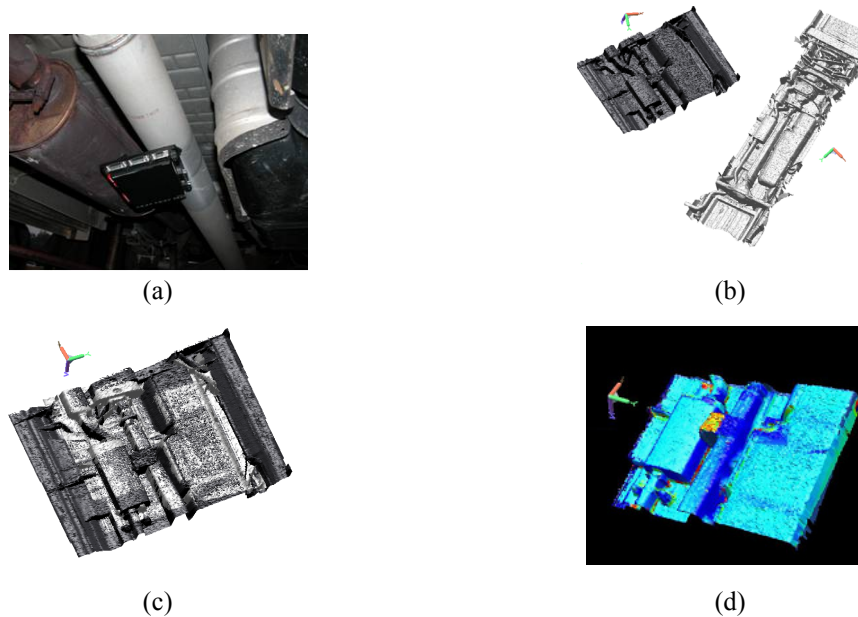


Figure 6.7: Scene verification approach for detecting changes made to the vehicle by comparing with archived scans. (a) An electronic board attached along the center line shaft of the Dodge van. (b) Scan of the under-vehicle scene with and without the electronic board. (c) Registration result of the two shells in (b). (d) The difference shell highlights the anomalous object found. Detecting such a threat automatically in a single pass would have been very difficult using the mirror on the stick approach and the camera-based robots.

We have demonstrated a robotic multi-sensor solution for the under-vehicle inspection scenario to acquire and communicate data from a safe stand-off distance, rendering data in a visual and spatially understandable form for a remote inspector to make decisions. The multiple sensor approach that includes visual cameras, thermal imagers and radioactivity sensors with future potential for biological and chemical detectors in a robust modular architecture, enables visualization of scene characteristics that we call perception intelligence in virtual reality environments. In addition, with the 3D sensing enhancement on our mobile robot, we have shown how the flexibility for archival and automatic threat detection using specific algorithms as technology ready for deployment at places of military and civilian interest. The scene verification approach is also a promising direction for automatic threat detection and localization.

6.2 Airport runway/Pavement crack inspection

The key to successful road surface evaluation lies in identifying different types of distress and linking them to the cause. Recognizing the defects and also understanding the cause based on the appearance helps rate pavement conditions and select cost effective repair measures. To that end, we discuss our approach and present the results from our prototype described in Chapter 3 for mobile 3D data acquisition in this section.

6.2.1 State-of-the-art methods

Related work towards pavement distress, especially on airport runways and army maintained highways dates back to early 1980's. The pavement management system (PMS) idea was proposed by the U.S Army (TM 5-623, 1982) and has since then undergone metamorphosis keeping pace with improving imaging technology. However, transportation departments met with limited real-time success using digital imaging techniques towards automatic crack detection and filling (McGhee, 2004), until the late nineties. Non-visual sensors and several improvements on image-based methods were proposed during this period. We summarize these methods in Figure 6.8 and discuss the advantages and disadvantages of the different types of sensing methodologies.

Analog films have been completely replaced by digital cameras. Among digital systems, video cameras are preferred to line scan methods for the ease of use without special illumination requirements, though line scan methods offer very high resolution data. Such video-based vision systems have two major drawbacks in extension to pavement inspection. They do not provide sufficient depth information and also have ambient illumination requirements. Range sensors that directly give depth measurements have limited field of view while profilometers and acoustic sensors though inexpensive can only provide low resolution and low dynamic range.

In 1987, Mendelsohn listed several of these methods including acoustic sensors and profilometers and suggested that the imaging modality was a promising approach (Mendelsohn, 1987). At that time, the processing and image acquisition speeds challenged the feasibility of a fast and efficient inspection system. Several surveys were conducted to make an assessment of the feasibility of incorporating image acquisition and processing methods for both development and implementation of automated road surface inspection (Howe and Clemena, 1998; Wang, 2000).

The conclusions of the survey encouraged by improving hardware and processing equipment have led to most of the commercial video-based systems available today that basically consist of an array of high speed imaging sensors supported with illumination equipment. The video data from such systems (Meignen, 1997), though promises to be sufficient for distress detection requires spatial information for crack filling after detection and maintenance. A potential solution AMPIS (Chung et al., 2003) was proposed that combined GPS information with video to create GIS-like databases of road surfaces. AMPIS claims improved road network identification, pavement inspection for better maintenance and data management over the base framework of PMS.

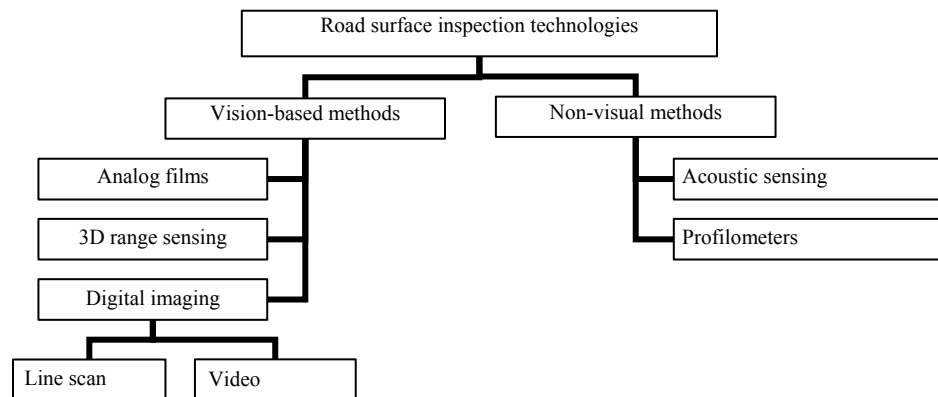


Figure 6.8: Summary of technologies demonstrated for road surface inspection.

Taking a robotic approach, Hass et al. (1992) proposed a system to overcome the shortcomings of the video-based system to make depth measurements by incorporating a laser range sensor. Hass et al. concluded that combining laser range data and video image data can provide overall better accuracy and speed of crack detection although due to the time consuming aspect of laser range sensing in 1992, they demonstrated range imaging for crack verification after the detection using the video-based system. Several 3D approaches have been demonstrated since then. Laurent et al. propose a synchronized laser scanning mechanism to capture high precision 3D range and texture profiles. Bursanescu and Blais (1997) reiterate a 3D optical sensor as the answer to high resolution and high accuracy acquisition and redesign a Biris sensor to meet the specific requirements of the pavement inspection application. They demonstrate six such sensors mounted on a mobile platform acquiring data at normal highway speeds.

6.2.2 Requirements

The main goal of road surface inspection is being able to identify crack patterns, rut depths and the roughness of the cracks. The depth information is of particular significance in airfields because the rating scheme for the runway surface (Walker, 2004) is not just dependent on the length and width of the cracks alone as is the case with pavement distress applications but also on the depth. Crack depths in the order of a few millimeters require high precision distance measurements. Hence, the design requirements for a comprehensive airfield data collection system should address accuracy and precision in three dimensions of measurement, speed of acquisition, time required for post processing, ease of visualization and evaluation.

With current data collection methods conforming the necessity to integrate of several heterogeneous technologies, we further identify the scope for improvements in system design targeting the time of acquisition and processing and list the important characteristics of a real-time deployable system. An ideal road data collection system must operate in real time gathering and post processing speeds. The duration required for data analysis should not overwhelm the time required for acquisition. A single pass data collection should be sufficient for cost-effective distress identification and localization, the critical aspect being the accuracy and robustness of the system and its extendibility to arbitrary terrain. State-of-the-art methods appear to assume relatively planar surfaces in their design. We also note that not only is depth important for detecting and classifying cracks, but also for the detection of object debris and other anomalies like vegetation that one does not expect on an asphalt or concrete runway. Also, a system that can operate without illumination requirements and can accommodate for unexpected GPS failures is desirable for lesser down time for maintenance.

6.2.3 Results

Our 3D imaging system was specifically designed with the requirements listed in Section 6.2.2 in mind. As proof of concept, we tested our system in several pavements with different kinds of cracks and present some of those results. One such area of interest is shown as an inset in Figure 6.9 along with the GPS path on a satellite map (Original image from www.maps.google.com) and the multi-modal integrated dataset in Figure 6.9. The discontinuity in the GPS path shown on the inset image is because we did not get back precisely to the starting point. To draw attention to the resolution at which we have imaged we show some zoomed in images of cracks and rough asphalt surfaces in the same figure.

We have color-coded the depth to emphasize the cracks. The small cracks on the right inset are about 2 cm wide and 1 cm deep while the longitudinal crack in the top-left inset is 3 cm wide and 3 cm deep. We have not shown the entire path (75m) at that high resolution considering the size of the data and memory resources required to render the model. Having emphasized the accuracy that our system is capable of in Figure 6.9, we present the ability to digitize large swaths of data where even large-area alligator cracks can be detected in Figure 6.10. Figure 6.10 (a) zooms in on a small distressed section of a road digitized using our system. The deep alligator cracks and a longitudinal crack are visible in the 3D model. Figure 6.10 (b) demonstrates our ability to detect foreign object debris (red section of the color-coded image) with relative ease compared to commercial video-based systems. With the gray shaded inset, we also note that the dataset in Figure 6.10 (b) shows perceivable geometric details that can differentiate gravel and asphalt surfaces. Figure 6.10 (c) is a texture mapped 3D model of an area inside a parking lot. The video sensor in our system is used to generate the texture and is registered with the 3D range profiles.

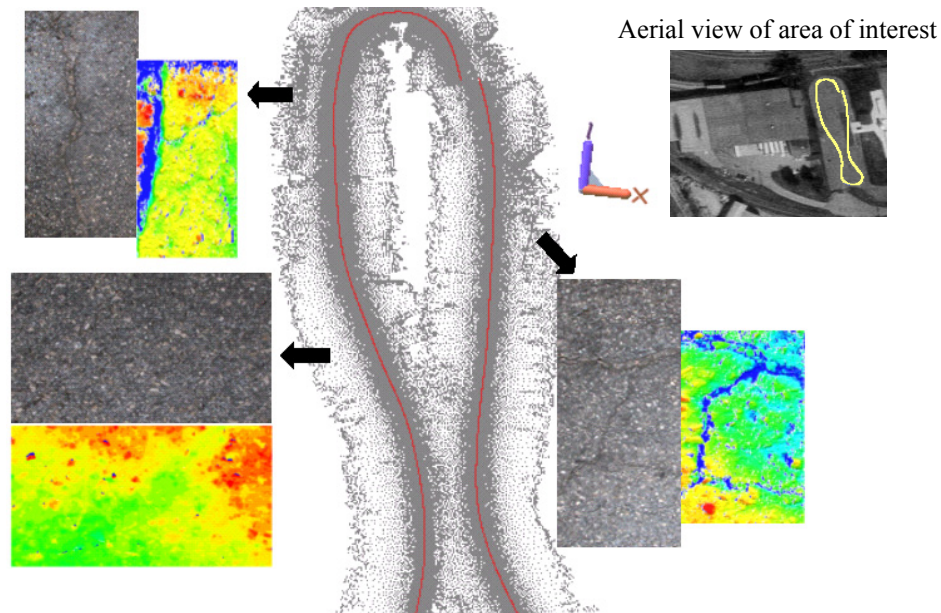
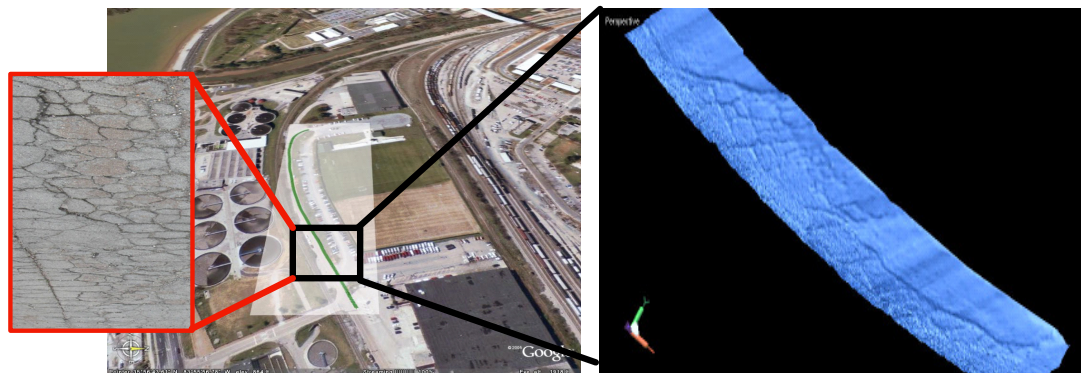
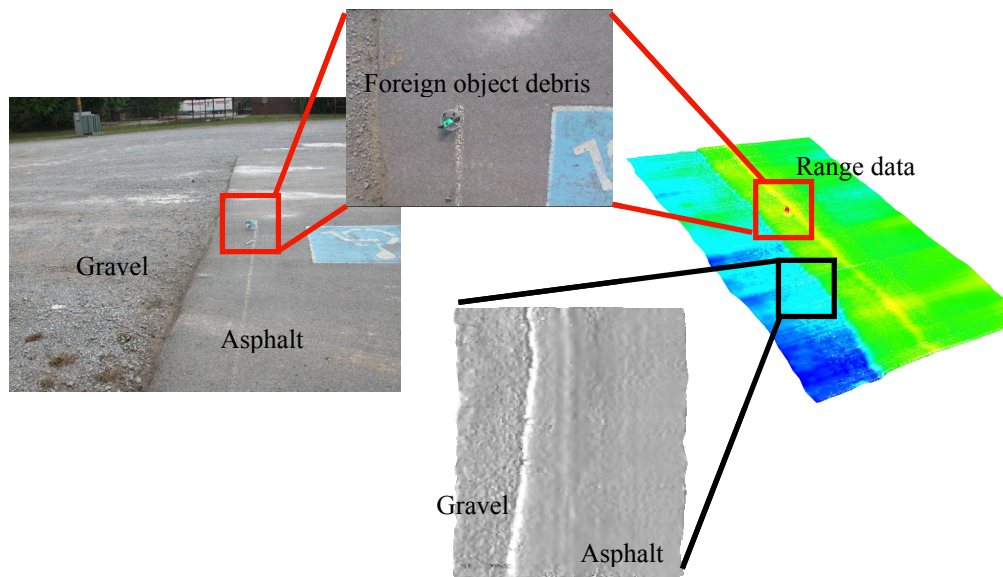


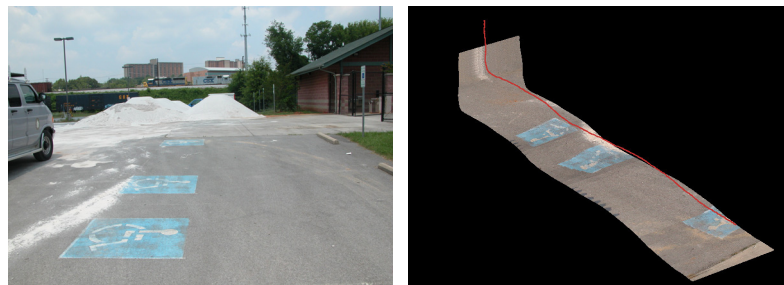
Figure 6.9: Multi-modal integrated 3D data of an area of interest with three small zoomed in sections of areas with different roughness and depth of cracks. The zoomed in sectional views show the color and the color-coded range data side-by-side. Our system is calibrated for high accuracy (order of millimeters) to even sense depth variations caused by the asphalt chips on the surface.



(a)



(b)



(c)

Figure 6.10: Large areas digitized at very high resolution. (a) Alligator cracks detected on the road surface. (b) Detection of foreign object debris based on 3D information. (c) Textured visualization on the 3D geometry still preserves the advantages of a vision-based system. The red contour indicates the path of the imaging system at the time of scanning.

6.2.4 Improvements over the state-of-the-art

With the 3D models that we have integrated crack detection has become easy with simple threshold-based algorithms giving us fast and accurate results. We have overcome the illumination requirements of the contemporary systems and are able to scan driving at 30 miles/hour at 3 mm depth accuracy on the cracks and 6mm inter-profile distance. With data samples from four sensors supplying data at different rates, we have integrated accurate photo-realistic 3D models for surface condition archival and convenient visualization. Our modular design enables the replacement of sensors to map larger areas by sacrificing on the accuracy based on the requirements of the application in hand. Our datasets encapsulate color and geometric information and allow application of existing color-based and geometry-based crack detection/classification algorithms. The fusion of both these modalities guarantees better performance and reliability. Our output of real world terrain as triangle mesh datasets readily feed as input to finite element analysis based vehicle-terrain simulators.

6.3 Large-scale terrain mapping

The motivation for large-scale terrain mapping is towards strategic planning in security situations. Unmanned vehicles have been deployed in several defense and security applications to provide apriori information about unknown unstructured environments with minimal risk to human life (Gage, 1995). These vehicles armored with sensors are capable of avoiding obstacles to navigate in an unknown environment, reporting concerns in different scenarios such as a battlefield (Freiburger et al., 2003), civilian security (Courtright, 1991), disaster management (Murphy, 2004), or in a patrol/surveillance mission (Klarquist, 1999). In such missions, the 3D environment map of the surveyed area of interest is useful feedback for organizing future action and deployment of resources in a much more efficient manner. Hence, we require a modular multi-sensor system and processing package that can be mounted on unmanned vehicles/mobility platforms to generate photo-realistic, geometrically accurate geo-referenced 3D models of the area of interest. Such a system should be able to generate 3D models without making any assumptions about the vehicle trajectory, ambient illumination and should also consider the uncertainties involved in a dynamic unstructured environment. Real-time data collection and processing is also desired.

6.3.1 State-of-the-art methods

In the early attempts towards terrain modeling, large swaths of coarse terrain data were acquired using airborne video systems (Baillard and Maître, 1999). Moving away from air-borne systems to easily accessible ground vehicles, an inexpensive approach of recovering 3D structure of buildings and cityscapes from video (Pollefeys et al., 2000) was demonstrated on cases where the shape could be recovered using stereo principles from successive image frames. Zhao and Shibaski (1997) demonstrated that using range sensors and a line CCD as extra data for registration and integration to create textured 3D models of urban environments was a faster and efficient approach to urban scene modeling compared to the aerial survey that was the state-of-the art at that time.

The MIT City scanning project (Antone and Teller, 2000) that inferred structure using spherical nodules was another effort in that direction. Inspired by Zhao and Shibaski (2001a), Christian Frueh (2001b) came up with the idea for urban mapping using two laser range profilers in an orthogonal arrangement along with digital cameras. He demonstrated the system mounted on a truck and driving at normal highway speeds to collect data that was processed offline. With his orthogonal arrangement, he was able to compute centimeter level accuracy by matching successive laser scans against each other and between the two sensors. The horizontal laser scans were used to approximate a component of the acquisition vehicle's motion. With the vertical scanner providing the façade of the urban structure, he proposed two different

approaches in using information from aerial maps to minimize global localization error using laser scans alone. One of those methods was to use cross correlation and the other a Markov-Monte Carlo technique to acquire 3D models in a matter of few minutes subject to traffic conditions. The two major drawbacks of this approach being the availability of the aerial map and the magnitude of global error that accumulated over just 100m of data.

Zhao and Shibaski (2001a and 2001b) further improved on Konno et al.(2000) who proposed three single-row laser range scanners and six line cameras mounted on a measure vehicle (GeoMaster), with a system equipped with a GPS/INS/Odometer-based navigation system. Their sensor mount outputs three kinds of data sources: laser range points, line images, and navigation data. Either the laser range points or the line images are in the sensor's local coordinate system at the time of measurement. They are synchronized with the navigation data using the sensor's local clock and integrated into 3D models offline. The motivation behind these urban scanning projects described so far are more on digitization than accuracy of digitization with expected errors in the order of a few centimeters. Also these methods did not address the uncertainty in the measurement process and the dynamic environment towards map building.

6.3.2 Requirements

We are focused on digitizing the real world without having to worry about the failure of the GPS or the inconsistencies in vision-based recovery. Hence, an independent modular system with the processing interface dedicated for mapping can expedite the map building process and improve mapping accuracy. Based on the level of detail that we desire in the environment, such a system should be modular and flexible in the system design, making the data collection and processing less cumbersome. Also, the map building process using the unmanned vehicles that are usually operated in stealth mode should be independent of ambient illumination capable of acquiring visual results both during the day and in the night. The hope and promise is that such a system would be faster and more realistic than computer graphics based design.

6.3.3 Results

We tested our system acquiring several miles of data in and around Knoxville, Tennessee and present some of the interesting results using the RIEGL scanner in Figure 6.11. There are three examples shown in that figure. The first one (Figure 6.11 a) is that of a BI-LO shopping center digitized and textured by driving our imaging prototype along the road in the parking lot in front of the shopping center. The second one is the Women's Basket ball hall of fame building near the University of Tennessee campus (Figure 6.11 b). We also show the path of our acquisition platform on a satellite image as an inset in the same figure. These models are accurate up to a few centimeters and extremely dense with each model consisting of a few 100,000 points. Mapping the Hall of fame building was a challenge. The building is along the curve of the road and mapping using image-based techniques was not going to be easy. We also had doubts on the availability of GPS signals as the building was very close to the urban canyon in the downtown area. Our instrumentation and integration method handled the situation resulting in the accurate and photo-realistic model.

We present another model integrated using our system in Figure 6.11 (c). We mapped along a 1 kilometer long path around the mall area on Chapman Highway without any prior knowledge about the area. We have shown zoomed in sections of the Goody's store to indicate the sampling density achievable using our system without having to compromise on the texture quality. Our acquisition took about 10 minutes further emphasizing our capability to quickly produce 3D models of urban environments. For large datasets spanning several miles, we process datasets offline. Though the time taken for processing large datasets can be quite cumbersome, a mile of data usually takes about an hour of processing on off-the-shelf desktop computers.

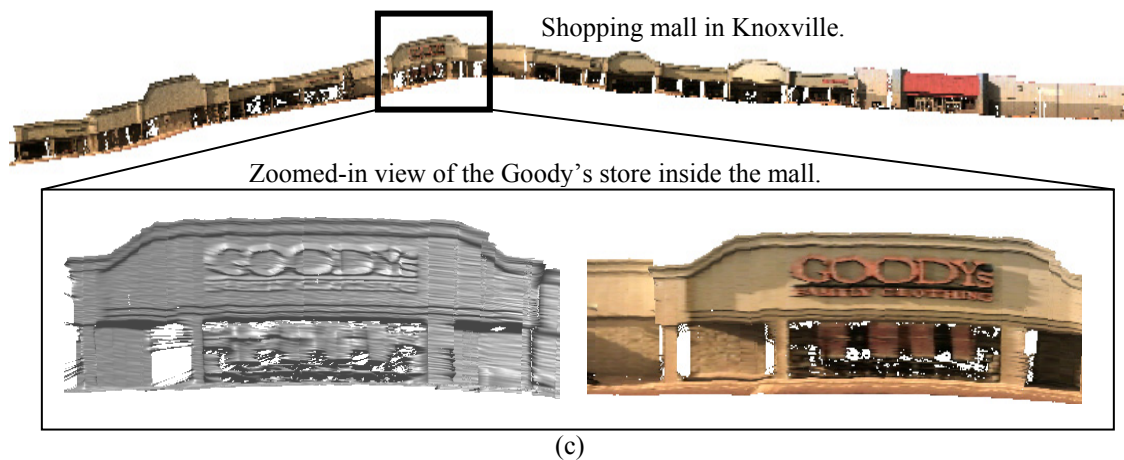
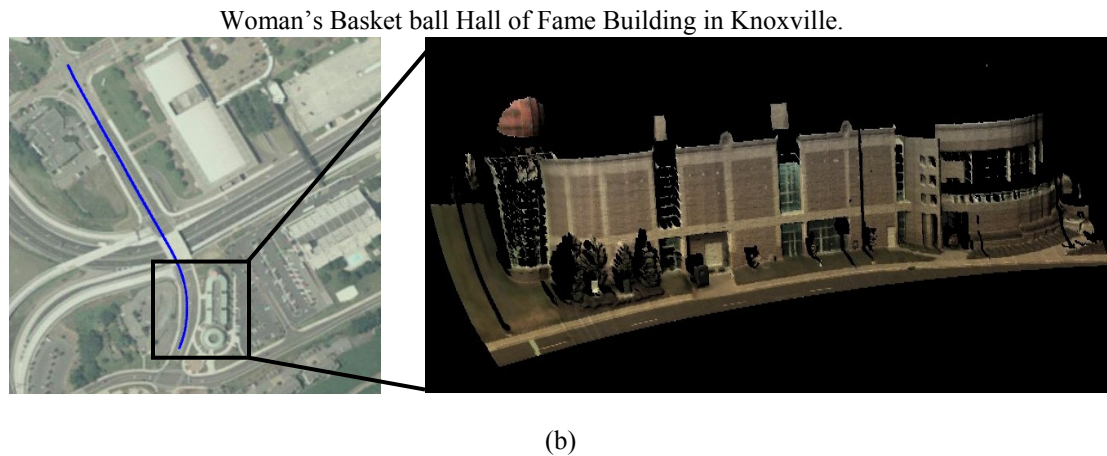
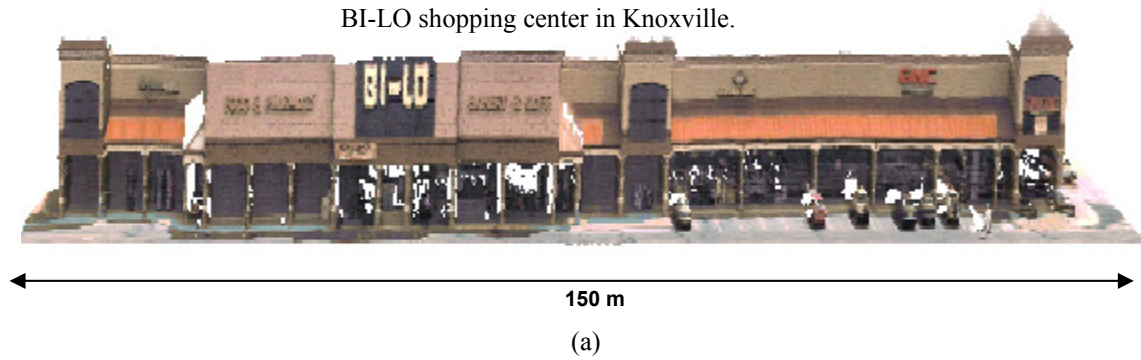


Figure 6.11: Large areas of urban environments digitized at very high geometric resolution with high fidelity texture. (a) BI-LO shopping complex in Knoxville. (b) The women's basket ball Hall of Fame building. (c) The 3D rendering of a shopping mall with the zoomed inset of the Goody's store on Chapman Highway in Knoxville. The sampling density of digitization and the photo-realistic rendering are key enhancements with our systems. Our output models are triangle meshes that are easy to embed in immersive virtual environments.

6.3.4 Improvements over the state-of-the-art

In essence, we have documented mobile mapping prototypes consisting of 4 main components: hardware for 3D geometry and texture acquisition; hardware for positioning and orientation (pose and trajectory) measurement; a mobile platform which moves the sensing package past the environment to be digitized; and software to perform the necessary information fusion to combine the data from different sensing modalities and to process the resulting model to fit the application at hand. While other researchers have developed 3D terrain acquisition systems, these tend to be fixed in regards to the hardware and the fusion methods used. In contrast, our system treats the components independently with the following improvements in the following areas:

- **Accuracy, resolution and photorealism:** Our system promises mm- to cm-level accuracy as required. Also, by using instrumentation for pose estimation and developing methods to handle uncertainty, our system is flexible in a variety of environments; making no assumptions about the environment to be digitized. We do not expect planar structures in the scene and can operate with expected accuracy in most real-world situations. Our contributions over the state-of-the-art is particularly with the accuracy at which we are able to image and simplicity in integration towards efficient processing and realistic visualization.
- **Modularity and robustness** – The modularity inherent in our design allows the system to be as robust to real world environments as the individual components, at the same time being independent of application-specific hardware modifications. That is without too much reconfiguration, our design is capable of easy integration when mounted on an aerial vehicle or a ground vehicle based on the need. The modular design enables us to treat accuracy and resolution as parameters of the system to suit the application in hand.

7 Conclusions

This dissertation work was motivated by the need to build 3D models of dynamic large-scale environments using mobile robotic platforms. We studied different 3D imaging techniques and documented our experience evaluating promising technologies. Our rigorous evaluation aided the design and demonstration of multi-modality multi-sensor system solutions with significant improvements over contemporary commercial systems. Our approach is easy to deploy, more accurate and lends to easier automation and remote visualization. We encountered several challenges in the form of noisy sensors, sensor conflict, performance degradation, potential sensor failure in dynamic environments and had to deal with the statistical validity of measurements from sensors before integrating and digitizing the real world into virtual 3D models. This chapter summarizes the contributions made to the existing literature through our research efforts on automation issues for high accuracy photo-realistic 3D imaging of large-scale dynamic environments and envisions potential directions for future research.

7.1 Dissertation key points

The quintessence of the research effort in this dissertation is finding the answers to the following questions. What are the technologies available today for 3D mapping? How close to reality can we digitize the spatial world around us quickly and cost effectively? What sensors to use for required accuracy and how many? How to build a system that is generic with wide applicability? Is the multi-sensor system the way to go? How to integrate maps using complimentary sensors? How to handle sensor failure and performance degradation in multi-sensory systems? When should a multi-sensor system fuse data and when should it select a believable sensor? How to increase the reliability of 3D inference from image-based mapping? We believe the dissertation provided answers to those questions with efficient prototypes, theoretical derivation and real world experiments to claim the following contributions in system design, data fusion and uncertainty management for 3D imaging using mobile robotic platforms.

- **3D imaging and automation solutions**

The system design contribution of this dissertation is the readily deployable system prototypes built as automation solutions to the applications described in Chapter 1. The modular hardware design architecture that we have implemented enables the trade-off between accuracy and cost for given application specifications. Our expertise in the choice of equipment on mobile robotic platforms combines with the supporting software development for data acquisition, processing and visualization in outperforming commercially available state-of-the-art systems in quickly and automatically generating accurate photo-realistic 3D models in critical safety and security applications.

- **Selfish-Altruist data fusion using information complexity**

The selfish-altruist data fusion scheme was motivated by the need to fuse multi-modal multi-sensor data for localization applications. With particular focus on the state localization of mobile platforms designed for 3D mapping, we realized the inadequacy with current belief propagation methods in two situations (1) when sensors are in conflict (2) when expected uncertainty levels of the sensors are

challenged intermittently while operating in dynamic environments. Our proposed inference framework inspired by model selection theory is a solution that in addition to deciding between sensor fusion and selection identifies reliable and confident sensors online leading to increased belief in the estimated state.

- **Reliability in image-based 3D inference**

Image-based mapping using cameras is by far the most economical approach for vehicle navigation and 3D scene modeling. But current shape from motion algorithms though theoretically mature, have reliability issues. We studied every step in the pose and structure estimation process: the feature extraction, feature matching, motion model estimation, hypothesis generation and iterative refinement and proposed an idea to increase the reliability in pose recovery process by generating ensemble uncertainty samples using different spatial features. The ensemble uncertainty samples were then used to compensate the uncertainty in time-stationary convergence of parameters relating image frames. Our approach proved effective in adaptively switching monocular depth cues and choosing motion models making the shape from motion pipeline less dependent on the scene and relative motion.

7.2 Future directions

The contributions we have claimed in this dissertation appear as refereed technical publications in the form of journal articles and conference proceedings attracting interest both from the academic and industrial community. The positive feedback and the attention we have drawn motivates us further to explore other applications and also improve our current prototypes.

The modular design of our 3D imaging prototypes has separated the hardware and software components in the system allowing the flexibility to function independently or in sync. Though the design was a deliberate and a conscious decision to ease trouble shooting and also enable offline data processing during the development phase, we believe, our integration method is mature and ready for implementation using graphics processing units (GPU). Introducing the GPU as the interface between hardware acquisition and software integration in addition to efficient parallel processing speed will also enable high quality rendering of 3D data. We are encouraged by the results in (Gong and Yang, 2005) to pursue this avenue in future.

The selfish-altruist data fusion theory is inspired by information theoretic concepts with strong foundations in concepts of information complexity. The theory developed is generic for any state-localization problem and does not limit itself to self-localization in 3D mapping. One such immediate extension of our framework could be for tracking people/vehicles in crowded scenes using camera networks. Following the formulation of target localization as a recursive state estimation problem by Isler and Bajcsy (2005) and the in-house expertise in sensor placement and video tracking (Yao, 2008) we see the potential of a target tractability distribution, that includes affine floor plan occupancy, camera coverage, motion uncertainty and the camera load in successfully and consistently tracking multiple people while maintaining and propagating high levels of belief about each target. Our preliminary results in tracking a human in Figure 7.1 are very encouraging. Our method is able to guide the camera handoff on a target adaptively with up to 50% more confidence than a hard-coded switch. We see the potential and hope to pursue the implementation of our inference framework in target and source localization applications involving sensor networks in the future.

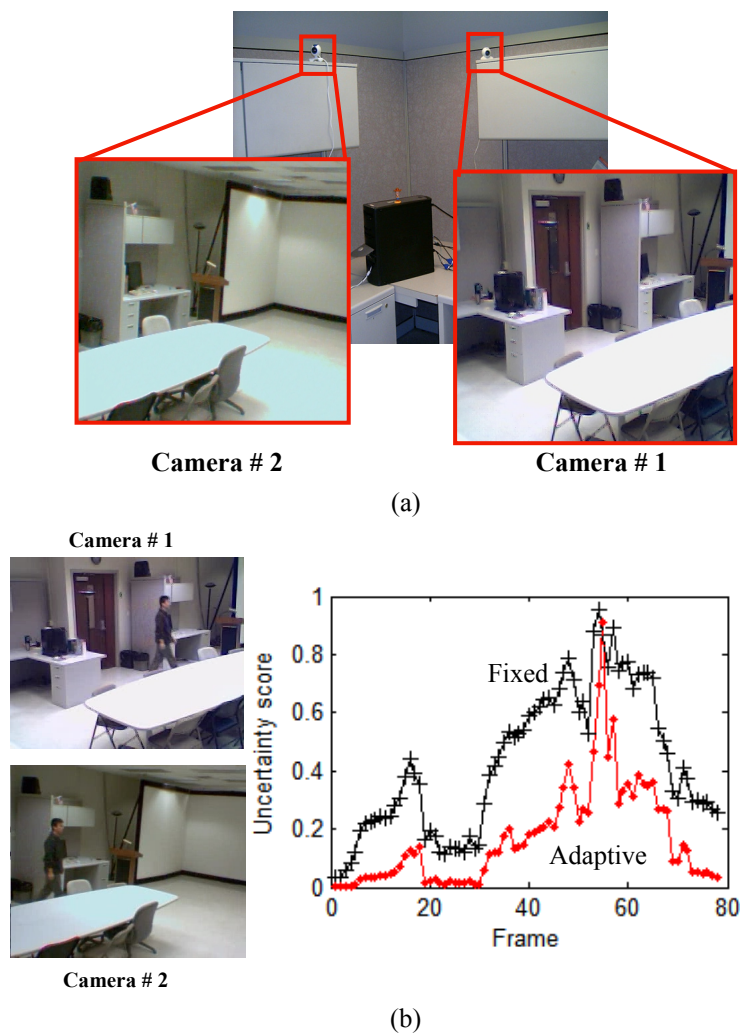


Figure 7.1: Preliminary results implementing our framework to guide camera hand over for target localization. We have plotted the uncertainty from the belief distribution of affine floor plan occupancy (Fleuret et al., 2008) in 80 frames before the complete handoff and show that our method is able to guide the smooth transition maintaining lower uncertainty levels on the target compared to a hard-coded camera switch. (a) The two camera setup and their individual fields of view. (b) The uncertainty score during the belief propagation of a target monitored in two cameras using the hard handoff approach compared with our adaptive framework.

Bibliography

- M.A. Abidi and R.C. Gonzalez, "Data fusion in robotics and machine intelligence," *Academic Press Professional, Inc*, 1992.
- A. Adam, E. Rivlin and I. Shimshoni, "ROR: Rejection of outliers by rotations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 1, pp. 78-84, 2001.
- H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, Vol. 19, No.6, pp. 716- 723, 1974.
- M. E. Antone and S. Teller, "Automatic recovery of relative camera positions in urban scenes," *Computer Vision and Pattern Recognition*, Vol. 2, pp. 282-289, 2000.
- X. Armangué, H. Araujo and J. Salvi, "A review on egomotion by means of differential epipolar geometry applied to the movement of a mobile robot," *Pattern Recognition*, 36, pp. 2927-2944, 2003.
- Autonomous Solutions Inc., "Spector: Under vehicle inspection system," *Product Brochure* 2005.
- R. Axelrod, "The evolution of cooperation," *New York: Basic Books*, ISBN 0-465-02121-2, 1984.
- A. Bacha, C. Bauman, R. Faruque, M. Fleming, C. Terwelp, C. Reinholtz, D. Hong, Al Wicks, T. Alberi, D. Anderson,, S. Cacciola, P. Currier,, A. Dalton, J. Farmer, J. Hurdus, S. Kimmel, P. King, A. Taylor, D. Van Covern, M. Webster, "Odin: Team VictorTango's entry in the DARPA urban challenge" , *Journal of Field Robotics, Special Issue: Special Issue on the 2007 DARPA Urban Challenge*, Vol. 25, No. 8, pp. 467-492, 2008.
- T. Bailey, "Mobile robot localization and mapping in extensive outdoor environments," *PhD thesis*, University of Sydney, Sydney, NSW, Australia, 2002.
- C. Baillard, and H. Maître, "3-D reconstruction of urban scenes from aerial stereo imagery: A focusing strategy," *Computer Vision and Image Understanding*, Vol. 76, No.3, pp. 244-258, 1999.
- P.J. Besl, H.D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14, No. 2, pp. 239-256, 1992.
- F. Blais, "Review of 20 years of range sensor development," *Journal of Electronic Imaging*, Vol. 13(1), pp. 231-240, 2004.
- P. Bogler, "Shafer-Dempster reasoning with applications to multi-sensor target identification systems," *IEEE Transaction on Systems, Man and Cybernetics*, Vol. 17, No. 6, pp. 968-977, 1987.
- G. A. Borges and M.-J. Aldon, "Robustified estimation algorithms for mobile robot localization based on geometrical environment maps," *Robotics and Autonomous Systems*, Vol. 45, No. 3-4, pp. 131-159, 2003.
- J. D. Bossler and R. W. Schmidley, "Airborne integrated mapping system (AIMS): Recent results in applications for large-scale mapping," in *the Proceedings of the Urban and Regional Information Systems Association*, 1997.
- H. Bozdogan, "Multi-sample cluster analysis as an alternative to multiple comparison procedures," *Bulletin of Informatics and Cybernetics*, Vol. 1-2, p. 95-129, 1986.
- H. Bozdogan, "Model selection and Akaike's information criterion," *Psychometrika*, Vol. 53, No.3, pp. 345-370, 1987.
- H. Bozdogan, "Akaike's information criterion and recent developments in information complexity," *Journal of Mathematical Psychology*, Vol. 44, pp. 62-91, 2000.
- E. A. Bretz, "X marks the spot, maybe," *IEEE Spectrum*, 37(04), pp. 26-36, 2000.
- D. M. Buede, "Shafer-Dempster and Bayesian reasoning: A response to "Shafer-Dempster reasoning with applications to multi-sensor target identification systems" ," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 18, No. 6, pp. 1009-1011, 1988.
- W. Burgard, D. Fox, H. Jans, C. Matenar, and S. Thrun, "Sonar-based mapping of large-scale mobile robot environments using EM," in *the Proc. of the International Conference on Machine Learning*, 1999.
- K. P. Burnham and D. R. Anderson, "Model selection and multi-model inference," *Springer-Verlag*, New York, 2002.
- L. Bursanescu and F. Blais, "Automated pavement distress data collection and analysis: A 3D approach. ," in *the Proc. of. Conf. on Recent Advances in 3-D Digital Imaging and Modeling*, pp. 311-317, 1997.

- D. Capriglione, C. Liguori, C. Pianese, A. Pietrosanto, "On-line sensor fault detection, isolation and accommodation in automotive engines," *IEEE Transactions on Instrumentation and Measurement*, Vol. 52, No. 4, pp. 1182-1189, 2003.
- J. A. Castellanos and J. D. Tardos. "Mobile robot localization and map building: a multi-sensor fusion approach," *Kluwer Academic Publishers*, Boston, MA, 2000.
- A. Censi, An accurate closed-form estimate of ICP's covariance, in the *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pp. 3167-3172, 2007.
- J.-P. Chavas, "On sustainability and the economics of survival," *American Journal of Agricultural Economics*, Vol. 75, No. 1, pp. 72-83, 1993.
- H.C. Chung, R. Girardello, T. Soeller, M. Shinozuka, "Automated management for pavement inspection system," in the *Proc. of the SPIE Smart Structures and Materials Symposium*, Vol. 5057, pp. 634-644, 2003.
- D. Cole and P. Newman, "Using laser range data for 3D slam in outdoor environments," in the *Proc. of IEEE Intl. Conference on Robotics and Automation*, pp. 1556-1563, 2006.
- J. Collins, "Good to great: Why some companies make the leap while some others don't," *New York: HarperCollins*, 2001.
- P. Collins and D. Nicholson, "Covariance intersection applied to a Kalman filter problem in the real world," *Proceedings of the SPIE Acquisition, Tracking, and Pointing Conference*, Vol. 3365, 1998.
- M. L. Courtright, "Unmanned vehicles go to war," *Machine Design*, Vol. 63 (25), pp. 60-64, 1991.
- A. Criminisi, I. Reid, A. Zisserman, "Single view metrology," *International Journal of Computer Vision*, Vol. 40, No. 2, pp. 123-148, 2000.
- M. J. E. Crombaghs, R. Brüglemann, and E. J. deMin, "On the adjustment of overlapping strips of laser altimeter height data," *Photogrammetry and Remote Sensing*, 33(B3/1), pp. 230-237, 2000.
- G. Csurka, C. Zeller, Z. Zhang and O. Faugeras, "Characterizing the uncertainty of the fundamental matrix," *Computer Vision and Image Understanding*, Vol. 68(1), pp. 18-36, 1997. Updated version of INRIA Research Report 2560, 1995.
- Y. Cui and S. S. Ge, "Autonomous vehicle positioning with GPS in urban canyon environments," *IEEE Transactions on Robotics and Automation*, Vol. 19, pp. 15-25, 2003.
- CVPR Workshop, "25 years of RANSAC," in the *Proc. of the Intl. Conf. on Computer vision and Pattern recognition*, CDROM.
- DARPA Grand Challenge, *Press Release* <http://www.grandchallenge.org/>, 2005.
- DARPA Grand Challenge Competition (Urban Challenge 2007), Technical evaluation criteria, *Press Release*, 2006.
- A. J. Davison, I. Reid, N. Molton and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Transactions on Pattern analysis and machine intelligence*, Vol. 29(6), pp. 1052-1067, 2007.
- R. Dawkins, "The Selfish Gene," *Oxford University Press*, 1999.
- J. Denzler and C.M. Brown, "Information theoretic sensor data selection for active object recognition and state estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 2, pp. 145-157, 2002.
- P. Dickson, J. Li, Z. Zhu, A. Hanson, E. Riseman, H. Sabrin, H. Schultz and G. Whitten, "Mosaic generation for under-vehicle inspection," *IEEE Workshop on Applications of Computer Vision*, pp. 251-256, 2002.
- A. Doucet, J. F. G. De Freitas and N. Gordon, Editors, "Sequential Monte-Carlo in practice," *Springer Verlag*, New York, 2001.
- B. S. Duran and P. Odell, "Cluster analysis – A survey," *Springer-Verlag*, New York, 1974.
- H. F. Durrant-Whyte, "Consistent integration and propagation of disparate sensor observations," *International Journal of Robotics Research*, Vol. 6, No. 3, pp. 3-24, 1987.
- H. F. Durrant-Whyte, "Sensor models and multi-sensor integration," *International Journal of Robotics Research*, Vol. 7(6), pp. 97-113, 1988.
- F. C. Dyer, N. A. Berry, and A. S. Richard, "Honey bee spatial memory: use of route-based memories after displacement," *Animal Behavior*, Vol. 45, pp. 1028-1030, 1993.

- A. Elfes, "Sonar-based real-world mapping and navigation," *IEEE Journal of Robotics and Automation*, RA-3(3), pp. 249–265, 1987.
- A. Elfes, "Multi-source spatial data fusion using Bayesian reasoning," *Academic Press*, pp. 137–163, 1992.
- M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, 24(6), pp. 381–395, 1981.
- F. Fleuret, J. Berclaz, J. R. Lengagne, P. Fua, "Multi-camera people tracking with a probabilistic occupancy map," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 2, pp. 267–282, 2008.
- D. A. Forsyth and J. Ponce, "Computer vision: A modern approach," *Prentice Hall*, 2003.
- D. Fox, W. Burgard, F. Dellaert, and S. Thrun, "Monte Carlo localization: Efficient position estimation for mobile robots," in *Proc. Natl. Conf. Artificial Intelligence*, pp. 343–349, 1999.
- D. Fox, J. Hightower, L. Liao, D. Schulz and G. Borriello, "Bayesian filtering for location estimation," *IEEE Pervasive Computing*, Vol. 2, pp. 24–33, 2003.
- L.A. Freiburger, W. Smuda, R.E. Karlsen, S. Lakshmanan and B. Ma, "ODIS the under-vehicle inspection robot: Development status update," *Proc. of SPIE Unmanned Ground Vehicle Technology V*, Vol. 5083, pp. 322–335, 2003.
- S. French, "Updating of belief in the light of someone else's opinion," *Journal of Royal Statistical Society Series*, Vol. A 143, pp. 43–48, 1980.
- D. Frere, J. Vandekerckhove, T. Moons, and L. VanGool, "Automatic modeling and 3D reconstruction of urban buildings from aerial imagery," in *the Proc. of the IEEE International Geoscience and Remote Sensing Symposium*, pp. 2593 - 2596, 1998.
- P. Fricker, R. Sandau, and A. S. Walker, "Digital aerial sensors - A new approach," in *the Proc. of the ISPRS Workshop on Sensors and Mapping from Space*, 1999.
- C. Früh and A. Zakhor, "3D model generation for cities using aerial photographs and ground level laser scans," in *the Proc. of Computer Vision and Pattern Recognition*, 2, pp. 31 - 38, 2001a.
- C. Früh and A. Zakhor, "Fast 3D model generation in urban environments," *Proceedings of the International Conference on Multi-sensor Fusion and Integration for Intelligent Systems*, pp. 165–170, 2001b.
- C. Früh and A. Zakhor, "Data processing algorithms for generating textured 3D building façade meshes from laser scans and camera images," *Proceedings of the Intl. Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 834 - 847, 2002.
- D. Gage, "UGV history 101: A brief history of Unmanned Ground Vehicle (UGV) development efforts," *Unmanned Systems Magazine*, Vol. 13, No. 3, pp. 9–16, 1995.
- J.B. Gao and C.J. Harris, "Some remarks on Kalman filters for the multi-sensor fusion," *Information Fusion*, Vol. 3, pp. 191–201, 2002.
- S. B. Gardner and J. K. Uhlmann, "A decentralized data fusion framework for horizontal integration of intelligence data," in *the Proc. of the International Conference on Intelligence Analysis*, 2005.
- T. D. Garvey, J.D. Lowrance and M.A. Fischler, "An inference technique for integrating knowledge from disparate sources". In *Multi-sensor integration and Fusion For intelligent Machines and Systems*, R. C. Luo and M. G. Kay, Eds. *Ablex Publishing Corp., Norwood, NJ*, pp. 309–325, 1995.
- C. Giraud and B. Jouvencel, "Sensor selection: A geometrical approach," in *Intl. Conf. on Intelligent Robots and Systems*, pp. 555–560, 1995.
- M. Gong and Y.-H Yang, "Near real-time reliable stereo matching using programmable graphics hardware," in *the Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Vol.1, pp. 924–931, 2005.
- G. Granosik, M. Hansen and J. Borenstein, "The omni-tread serpentine robot for industrial inspection and surveillance," *International Journal on Industrial Robots, Special Issue on Mobile Robots*, Vol. IR32-2, pp. 139–148, 2005.
- M. S. Grewal, L. R. Weill, and A. P. Andrews, "Global positioning systems, inertial navigation, and integration". *New York: John Wiley & Sons*, 2001.

- B. Grinstead, S. R. Sukumar, D. L. Page, A. F. Koschan, and M. A. Abidi, D. Gorsich, "Mobile scanning system for the fast digitization of existing roadways and structures," *Sensor Review Journal*, Vol. 26, No. 4, pp. 283-289, 2006.
- B. Grinstead, A. Koschan, and M. Abidi, "Fast digitization of large-scale hazardous facilities," in the *Proc. of 10th Int. Conf. on Robotics & Remote Systems for Hazardous Environments*, pp. 521-525, 2004.
- C. Haas, "Investigation of a pavement crack-filling robot," *Report to the Strategic Highway Research Program*, Pittsburgh, PA, 1992.
- J. K. Hackett and M. Shah, "Multi-sensor fusion: a perspective," in the *Proc. of International Conference on Robotics and Automation*, Vol. 3, pp. 1324-1330, 1990.
- D. Hähnel, W. Burgard, and S. Thrun, "Learning compact 3D models of indoor and outdoor environments with a mobile robot," *Robotics and Autonomous Systems*, Vol. 44, pp. 15-27, 2003.
- C. Harris and M. J. Stephens, "A combined corner and edge detector," in the *Proc. of the Alvey Vision Conference*, pp. 147-152, 1988.
- R. I. Hartley, "In defense of the eight-point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19 (6), pp. 580-593, 1997.
- R. I. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," *Cambridge University Press*, 2000.
- A. T. Hayes, A. Martinoli, and R. M. Goodman, "Distributed odor source localization," *IEEE Sensors*, Vol. 2, No. 3, pp. 260-271, 2002.
- X. C. He and N. H. C. Yung, "Curvature scale space corner detector with adaptive threshold and dynamic region of support," in the *Proc. of the Intl. Conference on Pattern Recognition*, Vol. 2, pp. 791-794, 2004.
- D. Hoiem, A. Efros, and M. Hebert, "Automatic photo pop-up," in the *Proc. of ACM SIGGRAPH*, Vol. 24, pp. 577-584, 2005.
- H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," *ACM SIGGRAPH Computer Graphics*, Vol. 26, No. 2, pp. 71-78, 1992.
- B.K.P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of Optical Society of America*, Vol. A4, No.4, pp. 629-642, 1987.
- G.E. Hovland and B.J. McCarragher, "Dynamic sensor selection for robotic systems," in the *Proc. of Intl. Conf. on Robotics and Automation*, Vol. 1, pp. 272-277, 1997.
- R. Howe, and G.G. Clemena, "An assessment of the feasibility of developing and implementing an automated pavement distress survey system incorporating digital image processing," Rep. No. VTRC 98-R1. Virginia Transportation Research Council, 1998.
- T.S Huang, A.N. Netravali, "Motion and structure from feature correspondences: a review," in the *Proc. of the IEEE*, Vol. 82(2), pp. 252-268, 1994.
- Integrated Vision Products, *User documentation: MAPP2500 ranger PCI system*, Version 1.6., Sweden, 2000. (<http://www.sickivp.se/>)
- V. Isler and R. Bajcsy, "The sensor selection problem for bounded uncertainty sensing models," *IEEE Transactions on Automation Science and Engineering*, Vol.3, No. 4, pp. 372- 381, 2005.
- P. Jensfelt and S. Kristensen, "Active global localization for a mobile robot using multiple hypothesis tracking," *IEEE Transactions on Robotics and Automation*, Vol. 17, No. 5, pp. 748-760, 2001.
- A.E. Johnson, C. Yang and L.H. Matthies, "Machine vision for autonomous small body navigation," *Proceedings of the IEEE Aerospace Conference*, Vol. 7, pp. 661-671, 2000.
- S. J. Julier and H. F. Durrant-Whyte, "On the role of process models in autonomous land vehicle navigation systems," *IEEE Transactions on Robotics and Automation*, Vol. 19, pp. 1-14, 2003.
- R.E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME-Journal of Basic Engineering*, Vol. 82, pp. 33-45, 1960.
- K. Kanatani, "Uncertainty modeling and model selection for geometric inference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26(10), pp. 1307-1319, 2004.
- W. N. Klarquist, K.G. Bonner and B. M. Gothard, "Demo III: reconnaissance, surveillance, and target acquisition (RSTA) preliminary design," in the *Proc. of SPIE Mobile Robots XIII and Intelligent Transportation Systems*, Vol. 3525, pp. 232-242, 1999.

- T. Koshizen, "Sensor selection by GMB-REM in real robot position estimation," *Journal of Intelligent and Robotic Systems*, Vol. 27, pp. 275–289, 2000a.
- T. Koshizen, "Improved sensor selection technique by integrating sensor fusion in robot position estimation," *Journal of Intelligent and Robotic Systems*, Vol. 29, No.1, pp. 79-92, 2000b.
- F. Kobayashi, F. Arai, T. Fukuda, "Sensor selection by reliability based on possibility measure," in the *Proc of the IEEE Conf. on Robotics and Automation*, Vol. 4, pp. 2614-2619, 1999.
- T. Konno, "A new approach to mobile mapping for automated reconstruction of urban 3D model," in the *Proc. of the International Workshop on Urban Multi-Media/3D Mapping*, 2000 (CDROM).
- A. Koschan, D. Page, J.-C. Ng, M. Abidi, D. Gorsich, and G. Gerhart, "SAFER under vehicle inspection through video mosaic building," *International Journal of Industrial Robot*, Vol. 31, pp. 435-442, 2004.
- P. Kovesi, "Image features from phase congruency," *Videre: A Journal of Computer Vision Research*, MIT Press, Vol. 1(3), 1999.
- J. R. Krebs, and N. B. Davies (eds.), "Behavioural ecology: An evolutionary approach," 4th ed, Blackwell Science, Oxford. 1997.
- B. Kuipers and Y.-T. Byun, "A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations," *Journal of Robotics and Autonomous Systems*, Vol. 8, pp. 47–63, 1991.
- S. Kullback, and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, Vol. 22, pp. 79-86, 1951.
- T. Kunio, "Feature Point Correspondence of Stereo Images by Monogenic Phase," Communications, Computers and Signal Processing, in the *Proc. of IEEE Pacific Rim Conference*, pp. 272-275, 2007.
- J. J. Leonard and H. F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons," in the *Proc. IEEE Trans. Robotics and Automation*, Vol. 7, No. 3, pp. 376-382, 1991.
- J. J. Leonard, H. F. Durrant-Whyte, I. J. Cox, "Dynamic map building for an autonomous mobile robot," *International Journal of Robotics Research*, Vol. 11, No. 4, pp. 286-298, 1992.
- K. Lerman, "A model of adaptation in collaborative multi-agent systems," *Adaptive Behavior*, Vol. 12(3-4), pp. 187–198, 2004.
- Y. Liu, and S. Thrun, "Results for outdoor-SLAM using sparse extended information filters," in the *Proc. of the IEEE International Conference on Robotics and Automation*, pp.1227–1233, 2003.
- D. G. Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision*, Vol. 60(2), pp. 91-110, 2004.
- G. Loy and A. Zelinsky, "Fast Radial Symmetry for Detecting Points of Interest," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25(8), pp. 959-973, 2003.
- F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Autonomous Robots*, Vol. 4, pp. 333-349, 1997.
- R. Luo and M.-H. Lin, "Robot multi-sensor fusion and integration - Optimum estimation of fused sensor data," in the *Proc. of the IEEE International Conference on Robotics and Automation*, pp. 1076-1084, 1988.
- R.C. Luo, C. -C Yih and K. L. Su, "Multi-sensor fusion and integration: approaches, applications, and future research directions," *IEEE Sensors Journal*, Vol.2, No.2, pp.107-119, 2002.
- Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, "An invitation to 3D vision: From images to geometric models," *Springer Verlag*, 2003.
- S. Mahamud, M. Hebert, Y. Omori, and J. Ponce, "Provably-convergent iterative methods for projective structure from motion," in the *Proc. of the Conference on Computer Vision and Pattern Recognition*, pp. 1018-1025, 2001.
- D. Manandhar and R. Shibasaki, "Geo-referencing of multi-sensor range data for vehicle-borne laser mapping system (VLMS)," in the *Proc. of the 21st Asian conference on Remote Sensing*, pp. 928-931, 2000.
- J. L. Martínez, J. González, J. Morales, A. Mandow, A.J. García-Cerezo, "Mobile robot motion estimation by 2D scan matching with genetic and iterative closest point algorithms," *Journal of Field Robotics*, Vol. 23(1), pp. 21-34, 2006.

- S. Maskell, "A Bayesian Approach to fusing uncertain, imprecise and conflicting information," *Information Fusion Journal*, Vol. 9(2), pp. 259-277, 2008.
- F. Masson, J. Guivant, and E. Nebot, "Robust navigation and mapping architecture for large environments," *Journal of Robotic Systems*, Vol. 20(10), pp. 621-634, 2003.
- K.H. McGhee, "NCHRP Report 334: Automated Pavement Distress Collection Techniques," *Transportation Research Board, National Research Council*. Washington D.C., 2004.
- G.J. McLachlan and T. Krishnan, "The EM algorithm and extensions," *Wiley Series in Probability and Statistics*, New York, 1997.
- D. Meignen, M. Bernadet, H. Briand, "One application of neural networks for defects using video data bases: identification of road distresses," *DEXA Workshop*, pp. 459-464, 1997.
- R. Menzel, K. Geiger, L. Chittka, J. Joerges, J. Kunze and U. Müller, "The knowledge base of bee navigation," *The Journal of Experimental Biology*, Vol. 199, 141-146, 1996.
- D. H. Mendelsohn, "Automated pavement crack detection: An assessment of leading technologies," in *Proc. 2nd North American Conference on Managing Pavements*, Vol. 3. No. 3, pp. 297-314, 1987.
- K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *International Journal of Computer Vision*, Vol. 1(60), pp. 63-86, 2004.
- H. B. Mitchell, "Multi-sensor data fusion," *Springer-Verlag*, 2007.
- D. C. Montgomery, "Design and analysis of experiments", 6th Edition, *Wiley Publishers*, 2005.
- T. Moons, D. Frere, J. Vandekerckhove, and L. V. Gool, "Semi-automatic generation of 3D city models for urban sites from aerial imagery," in the *Proc. of the AGI Conference at GIS 98*, pp. 37-41, 1998.
- A. C. Morris, D. Silver, D. Ferguson, and S. Thayer, "Towards topological exploration of abandoned mines," in the *Proc. of the IEEE International Conference on Robotics and Automation*, pp. 2117-2123, 2005.
- R. R. Murphy, "Activities of the rescue robots at the world trade center from 11-21 September 2001," *IEEE Robotics & Automation Magazine*, Vol. 11, No. 3, 2004.
- N. Naik, "Design, development and characterization of a thermal sensor brick system for modular robotics," *Masters Thesis, University of Tennessee, Knoxville, USA*, 2006.
- Y. Nakamura and Y. Zu, "Geometrical fusion method for multi-sensor robotic systems," in the *Proc. of the IEEE Intl. Conference on Robotics and Automation*, Vol. 2, pp. 668-673, 1989.
- J.-C Ng, "Multi-perspective mosaics and layered representation for scene visualization," *Masters Thesis, Imaging Robotics and Intelligent Systems Lab, University of Tennessee, Knoxville, USA*, 2003.
- D. Nister, O. Naroditsky, J. Bergen, "Visual odometry," in *Proc. of the Intl. Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 652-659, 2004a.
- D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756-770, 2004b.
- J. Nygård, P. Skoglar, M. Ulvklo, and T. Höglström, "Navigation aided image processing in UAV surveillance: Preliminary results and design of an airborne experimental system," *Journal of Robotic Systems*, 21(02), pp. 63-72, 2004.
- G.A. Parker, "Evolutionarily stable strategies. In: Behavioural ecology: An Evolutionary Approach," Second Edition. ed. J. R. Krebs & N. B. Davies, *Blackwell Scientific Publications*, 1984.
- M. Pollefeys, R. Koch, and L. V. Gool, "Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters," *International Journal of Computer Vision*, Vol. 32, pp. 7-25, 1999.
- M. Pollefeys, R. Koch, M. Vergauwen, and L. V. Gool, "Automated reconstruction of 3D scenes from sequences of images," *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 55, No. 4, pp. 251-267, 2000.
- M. Pollefeys, L.V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis and J. Tops, "Video-to-3D," *Proceedings of the Symposium on Photogrammetric Computer Vision*, Vol.A, pp. 252-257, 2002.
- N. S. V. Rao, "On fusers that perform better than best sensor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 904-909, 2001.
- RIEGL Laser measurement systems, *Laser Mirror Scanner (LMS-Z210): Technical documentation and User's instructions*, Edition 00/05, 2000. (<http://www.rieglusa.com>)

- Rissanen, "Modeling by shortest data description," *Automatica*, Vol. 14, pp. 465-471, 1978.
- J. M. Roberts, "Attentive visual tracking and trajectory estimation for dynamic scene segmentation," *PhD Thesis, University of Southampton*, UK, 1994.
- V. Rodehorst and A. Koschan, "Comparison and evaluation of feature point detectors," in *Proc. of the 5th International Symposium Turkish-German Joint Geodetic Days "Geodesy and Geoinformation in the Service of our Daily Life"*, Berlin, Germany, March 2006.
- G. L. Rogova and V. Nimier, "Reliability in information fusion: literature survey," in the *Proc. of the 7th Intl. Conference on Information fusion*, pp. 1158-1165, 2004.
- SICK Laser measurement systems, *LMS 200/211/221/291: Technical Documentation*, 2006 (www.sick.com).
- Y. M. Saiki, E. Takeuchi and T. Tsubouchi, "Vehicle localization in outdoor woodland environments with sensor fault detection," In *Proc. of the IEEE International Conference on Robotics and Automation*, 2008.
- A. Saxena, S. Chung, and A. Ng., "3-D Depth reconstruction from a single still image," *International Journal of Computer Vision*, Vol. 76(1), pp. 53-69, 2008.
- G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, Vol. 6, pp. 461-464, 1978.
- S. Scheduling, E. Nebot, and H. Durrant-Whyte, "The detection of faults in navigation systems: A frequency domain approach," in the *Proc. of the IEEE International Conference on Robotics and Automation*, pp. 2217-2222, 1998.
- C. Schmid, R. Mohr and C. Bauckhage, "Comparing and evaluating interest points," in the *Proc. of the International Conference on Computer Vision*, pp. 230-235, 1998.
- L. Shapiro, A. Zisserman and M. Brady, "Motion from point matches using affine epipolar geometry," in the *Proc. of the European Conference on Computer Vision*, pp. 73-84, 1994.
- D. A. Simon, M. Hebert, and T. Kanade, "Real-time 3-D pose estimation using a high-speed range sensor," in the *Proc. of the International Conference on Robotics and Automation*, Vol. 3, pp. 2235-2241, 1994.
- R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," In *Autonomous Robot Vehicles*, Springer-Verlag, pp. 167-193, 1990.
- B. Smuda, E. Schoenherr, H. Andrusz, and G. Gerhart, "Deploying the ODIS robot in Iraq and Afghanistan," in the *Proceedings of SPIE Unmanned Ground Vehicle Technology VII*, Vol. 5804, pp. 119-129, 2005.
- H. Stewenius, C. Engels and D. Nister, "Recent developments on direct relative orientation," *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4):284-294, 2006.
- C. Strecha, T. Tuytelaars and L. Van Gool, "Dense matching of multiple wide-baseline views," in the *Proc. of the Intl. Conference on Computer Vision*, Vol. 2, pp.1194-1201, 2003.
- P. Sturm and B. Triggs, "A factorization based algorithm for multi-image projective structure and motion," in the *Proc. European Conf. on Computer Vision*, Cambridge, pp. 709-720, 1996.
- S.R. Sukumar, D. L. Page, A. V. Gribok, A. F. Koschan, M.A. Abidi, D. J. Gorsich, and G. R. Gerhart, "Robotic 3D imaging system for under vehicle inspection," *Journal of Electronic Imaging*, Vol. 15, No. 3, 033008, 2006a.
- S. R. Sukumar, S.-J. Yu, D. L. Page, A. F. Koschan, and M. A. Abidi, "Multi-sensor integration for unmanned terrain modeling," in the *Proc. SPIE Unmanned Systems Technology VIII*, Vol. 6230, pp. 65-74, 2006b.
- S.R. Sukumar, D.L. Page, A. Koschan, and M.A. Abidi, "Under vehicle inspection with 3D imaging," in *3D Imaging for Safety and Security*, A. Koschan, M. Pollefeys, and M. Abidi (Eds.), *Springer, Dordrecht*, The Netherlands, pp. 249-278, 2007a.
- S. R. Sukumar, H. Bozdogan, D. L. Page, A. F. Koschan, and M. A. Abidi, "Sensor selection using information complexity for multi-sensor mobile robot localization," in the *Proc. of the IEEE International Conference on Robotics and Automation*, pp. 4158-4163, 2007b.
- S. R. Sukumar, H. Bozdogan, D. L. Page, A. F. Koschan, and M. A. Abidi, "On Handling Uncertainty in the Fundamental Matrix for Scene and Motion Adaptive Pose Recovery," in the *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

- H. Surmann, A. Nüchter, and J. Hertzberg, "An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments," *Robotics and Autonomous Systems*, 45, pp. 181-198, 2003.
- S. Thrun, C. Martin, Y. Liu, D. Hahnel, R. Emery Montemerlo, C. Deepayan, and W. Burgard, "A real-time expectation maximization algorithm for acquiring multi-planar maps of indoor environments with mobile robots," *IEEE Transactions on Robotics and Automation*, Vol. 20(3), pp. 433-442, 2003.
- S. Thrun, Y. Liu, D. Koller, A. Ng, Z. Ghahramani and H. Durrant-Whyte, "Simultaneous localization and mapping with sparse extended information filters," *Intl. Journal of Robotics Research*, 23, pp. 693-716, 2004.
- S. Thrun, W. Burgard and D. Fox, "Probabilistic Robotics," *MIT Press*, 2005.
- S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian, and P. Mahoney, "Stanley, the robot that won the DARPA Grand Challenge," *Journal of Field Robotics*, pp. 661-692, 2006.
- TM 5-623, Department of Army, *Technical Maintenance Management, Technical Manual*, 1982.
- P.H.S. Torr, "Geometric motion segmentation and model selection," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 356(1740), pp. 1321-1340, 1998.
- P.H.S Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, Vol. 78, pp. 138-156, 2000.
- D. S. Touretzky, H.S. Wan and A.D. Redish, "Neural representation of space in rats and robots," in: J.M. Zurada and R.J. Marks (Eds.), *Computational Intelligence: Imitating Life*, *IEEE Press, Piscataway, NJ*, 1994.
- E. Trucco and A. Verri, "Introductory techniques for 3D computer vision", *Englewood Cliffs: Prentice-Hall*, 1998.
- B. Trullier, S. I. Wiener, A. Berthoz, J.-A Meyer, "Biologically based artificial navigation systems: review and prospects," *Progress in Neurobiology*, Vol. 51(5), pp. 483-544, 1997.
- J. Uhlmann, "Covariance consistency methods for fault-tolerant distributed data fusion," *Information Fusion*, Vol. 4, pp. 201-215, 2003.
- C. Urmson, J. Anhalt, D. Bagnell, C. Baker, R. Bittner, M. N. Clark, J. Dolan, D. Duggins, T. Galatali, C. Geyer, M. Gittleman, S. Harbaugh, M. Hebert, T. M. Howard, S. Kolski, A. Kelly, M. Likhachev, M. McNaughton, N. Miller, K. Peterson, B. Pilnick, R. Rajkumar, P. Rybski, B. Salesky, Y.-Woo Seo, S. Singh, J. Snider, A. Stentz, W. R. Whittaker, Z. Wolkowicki, J. Zigar, H. Bae, T. Brown, D. Demitrish, B. Litkouhi, J. Nickolaou, V. Sadekar, W. Zhang, J. Struble, M. Taylor, M. Darms, D. Ferguson, "Autonomous driving in urban environments: Boss and the Urban Challenge," *Journal of Field Robotics*, Vol. 25(8), pp. 423-566, 2008.
- Van Emden, "An analysis of complexity," *Mathematical Centre Tracts*, Vol. 35, 1971.
- D. Walker, "Asphalt airfield pavements, pavement surface evaluation and rating," *University of Wisconsin, Madison, Transportation Information Center*, 2004.
- E. L. Waltz, and D. M. Buede, "Data fusion and decision support for command and control," In *Multisensor integration and Fusion For intelligent Machines and Systems*, R. C. Luo and M. G. Kay, Eds., *Ablex Publishing Corp., Norwood, NJ*, pp. 563-595, 1995.
- K. C.P Wang, "Design and Implementation of Automated Systems for Pavement Surface Distress Survey," *ASCE Journal of Infrastructure Systems*, Vol. 6, No. 1, 24-32, 2000.
- R. Wang and E. Spelke, "Human spatial representation: Insights from animals," *Trends in Cognitive Science*, Vol. 6, No. 9, 2002.
- J.R. Wilson, "A comparison of the sensor brick concept as a modular system architecture to the real time control system as the operational architecture," *Masters Thesis, Imaging Robotics and Intelligent Systems Lab, University of Tennessee, Knoxville, USA*, 2005.

- N. Xiong and P. Svensson, "Multi-sensor management for information fusion: issues and approaches", *Information fusion*, Vol. 3, pp. 163-186, 2002.
- Y. Yao, "Long range automated persistent surveillance," *PhD Thesis, Imaging, Robotics and Intelligence Systems Lab, University of Tennessee, Knoxville, USA*, 2008.
- S.-J. Yu, S.R. Sukumar, A. Koschan, D. L. Page, and M.A. Abidi, "3D reconstruction of road surface using an integrated multi-sensory approach," *Optics and Lasers in Engineering*, Vol. 45, No. 7, pp. 808-818, 2007.
- L. Zadeh, "Fuzzy logic and approximate reasoning," *Synthese*, Vol. 30, pp. 407-428, 1975.
- L. Zhang, "Representation independence, and combination of evidence in the Dempster-Shafer theory," in: R.R. Yager, M. Fedrizzi, F. Kacprzyk (Eds.), *Advances in the Dempster-Shafer Theory of Evidence*, Wiley, New York, pp. 51-69, 1994.
- Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *International Journal of Computer Vision*, Vol. 27(2), pp. 161-195, 1998.
- Y. Zhang, A. Tang, T. Palmer and C. Hazard, "Virtual proving ground - an integrated technology for full vehicle analysis and simulation," *International Journal of Vehicle Design*, Vol. 21 (4/5), pp. 450-470, 1999.
- Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No.11, pp. 1330-1334, 2000.
- W. Zhao, D. Nistér and S. Hsu, "Alignment of continuous video onto 3D point clouds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27 (8), pp. 1305-1318, 2005.
- H. Zhao and R. Shibasaki, "Automated registration of ground-based laser range image for reconstructing urban 3D object," *International Archive on Photogrammetry and Remote Sensing*, Vol. 32 (3-4W2), 1997.
- H. Zhao and R. Shibasaki, "High accurate positioning and mapping in urban area using laser range scanner," in *the Proc. of the IEEE Intelligent Vehicles Symposium*, 2001a.
- H. Zhao and R. Shibasaki, "Reconstructing urban 3D model using vehicle-borne laser range scanners," in *the Proc. of the International Conference on 3D Digital Imaging and Modeling*, pp. 349 - 356, 2001b.
- J. Zhu, Y. Zhu and V. Ramesh, "Error-metrics for camera ego-motion estimation," in *the Proc. of the Intl. Conf. on Computer Vision and Pattern Recognition Workshop*, pp. 67-75, 2005.
- M. Zuliani, C. Kenney and B.S. Manjunath, "A mathematical comparison of point detectors," in *the Proc. IEEE Image and Video Registration Workshop*, 2004.

Vita

Sreenivas Rangan Sukumar is an exploratory researcher with wide interests in science and engineering. He received his Bachelor's degree in Electronics and Communication Engineering from the University of Madras, India in 2002, a Master's and a Doctor of Philosophy degree in Electrical Engineering from the University of Tennessee, Knoxville in 2004 and 2008 respectively. Though he has pursued degrees in electrical engineering with research focus on 3D computer vision and artificial intelligence, he astutely follows the literature on cognitive psychology fascinated by the human mind, reads physics to learn about natural philosophy, investigates creative problem solving for the design of future innovations and studies statistical data analysis for surviving in the present. As an engineer, he motivates himself to be an innovator, believes in bearing the pain of being a pioneer to enjoy the pleasure of being one and enjoys his time exploring ideas and solutions from unseen out-of-the-box perspectives. Academics aside, he is a vegetarian and wishes to lead an organized life of values and ambitions in service to the human community.