



# ENVIRONMENT LEARNING FOR INDOOR MOBILE ROBOTS

PhD Thesis

Juan Andrade Cetto

April 8, 2003

Universitat Politècnica de Catalunya

Department d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial

PhD program:

Enginyeria en Informàtica Industrial / Tecnologies Avançades de la Producció

This thesis was completed at:

Institut de Robòtica i Informàtica Industrial, UPC-CSIC

Thesis advisor: Alberto Sanfeliu

© Juan Andrade Cetto 2003

# Abstract

This thesis focuses on the various aspects of autonomous environment learning for indoor service robots. Particularly, on landmark extraction from sensor data, autonomous map building, and robot localization.

To univocally identify landmarks from sensor data, we study several landmark representations, and the mathematical foundation necessary to extract the features that build them from images and laser range data. The features extracted from just one sensor may not suffice in the invariant characterization of landmarks and objects, pushing for the combination of information from multiple sources. We present a new algorithm that fuses complementary information from two low level vision modules into coherent object models that can be tracked and learned in a mobile robotics context.

Illumination conditions and occlusions are the most prominent artifacts that hinder data association in computer vision. By using photogrammetric and geometric constraints we restrict the search for landmark matches in successive images, and by locking our interest in one or a set of landmarks in the scene, we track those landmarks along successive frames, reducing considerably the data association problem. We concentrate on those tools from the geometry of multiple views that are relevant to the computation of initial landmark location estimates for coarse motion recovery; a desirable characteristic when odometry is not available or is highly unreliable.

Once landmarks are accurately extracted and identified, the second part of the problem is to use these observations for the localization of the robot, as well as the refinement of the landmark location estimates. We consider robot motion and sensor observations as stochastic processes, and treat the problem from an estimation theoretic point of view, dealing with noise by using probabilistic methods.

The main drawback we encounter is that current estimation techniques have been devised for static environments, and that they lack robustness in more realistic situations. To aid in those situations in which landmark observations might not be consistent in time, we propose a new set of temporal landmark quality functions, and show how

by incorporating these functions in the data association tests, the overall estimation-theoretic approach to map building and localization is improved. The basic idea consists on using the history of data association mismatches for the computation of the likelihood of future data association, together with the spatial compatibility tests already available.

Special attention is paid in that the removal of spurious landmarks from the map does not violate the basic convergence properties of the localization and map building algorithms already described in the literature; namely, asymptotic convergence and full correlation.

The thesis also gives an in depth analysis of the fully correlated model to localization and map building from a control systems theory point of view. Considering the fact that the Kalman filter is nothing else but an optimal observer, we analyze the implications of having a state vector that is being revised by fully correlated noise measurements. We end up revealing theoretically and with experiments the strong limitations of using a fully correlated noise driven estimation theoretic approach to map building and localization in relation to the total number of landmarks used.

Partial observability hinders full reconstructibility of the state space, making the final map estimate dependant on the initial observations, and does not guarantee convergence to a positive definite covariance matrix. Partial controllability on the other hand, makes the filter believe after a number of iterations, that it has accurate estimates of the landmark states, with their corresponding Kalman gains converging to zero. That is, after a few steps, innovations are useless. We show how to palliate the effects of full correlation and partial controllability. Furthermore, given that the Kalman filter is an optimal observer for the reconstruction of fully correlated states; it seems pertinent to build an optimal regulator in order to keep the robot as close as possible to a desired motion path when building a map. We show also how the duality between observability and controllability can be exploited in designing such an optimal regulator.

Any map building and localization algorithm for mobile robotics that is to work in real time must be able to relate observations and model matches in an expeditious way. Some of the landmark compatibility tests are computationally expensive, and their application has to be carefully designed. We touch upon the time complexity issues of the various landmark compatibility tests used, and also on the desirable properties of our chosen map data structure. Furthermore, we propose a series of tasks that must be handled when dealing with landmark data association. From model compatibility tests, to search space reduction and hypothesis formation, to the actual association of observations and models.

The work presented in this thesis spans several areas of engineering and computer

science, from new computer vision algorithms, to novel ideas in mobile robot localization and map building. The key contributions are the proposal of a new technique to fuse visual data; the formulation of new algorithms to concurrent localization and map building that take into account temporal landmark quality; new theoretical results on the degree of reconstruction possible when building maps from fully correlated observations; and the necessary techniques to palliate partial observability, partial controllability, and the nonlinear effects when solving the simultaneous localization and map building problem.



# Resum

Aquesta tesi tracta el problema de l'aprenentatge automàtic d'entorns estructurats en robòtica mòbil. Particularment, l'extracció de característiques a partir dels senyals dels sensors, la construcció autònoma de mapes, i l'autolocalització de robots.

S'estudien els fonaments matemàtics necessaris per a l'extracció de característiques a partir d'imatges i registres d'un làser, els quals permeten la identificació unívoca dels elements de l'entorn. Els atributs extrets a partir del senyal d'un sol sensor poden ser insuficients quan es volen caracteritzar els elements de l'entorn de forma invariant; això es pot millorar combinant informació de múltiples fonts. Es presenta un nou algorisme per la fusió d'informació complementària extreta de dos mòduls de visió de baix nivell. Aquesta fusió d'informació produeix descripcions més completes dels objectes de l'entorn, els quals poden ser seguits i apresos dins el context de la robòtica mòbil.

Les variacions en les condicions d'il·luminació i les oclusions fan que l'associació de dades en visió per computador sigui una tasca difícil de completar. Tot i això, l'ús de restriccions geomètriques i fotogramètriques permeten reduir la cerca de correspondències entre imatges successives; i al centrar l'atenció en un reduït nombre de característiques, aquestes poden ser seguides en imatges successives, simplificant així el problema d'associació de dades. Es recalquen les tècniques de la geometria de múltiples vistes que són rellevants pel còmput d'una estimació inicial de la posició dels elements de l'entorn, el que permet la reconstrucció del moviment del robot entre imatges successives; situació desitjable quan no existeix odometria o quan les seves lectures són poc fiables.

Quan els elements de l'entorn s'han extret i identificat, la segona part del problema consisteix en utilitzar aquestes observacions tant per estimar la posició del robot, com per refinar l'estimació dels mateixos elements de l'entorn. El moviment del robot i les lectures dels sensors es consideren com dos processos estocàstics, i el problema es tracta des del punt de vista de la teoria d'estimació, on el soroll inherent als sensors i al moviment del robot es consideren com a seqüències aleatòries.

El principal inconvenient existent en l'ús de tècniques d'estimació pel còmput concurrent de la posició del robot i la construcció d'un mapa, és que fins ara s'ha considerat la seva aplicació únicament en entorns estàtics, i que el seu ús en situacions més realistes ofereix poca robustesa. Es proposa un conjunt de funcions per avaluar la qualitat temporal de les observacions per tal de resoldre les situacions en que les observacions dels elements de l'entorn no siguin consistents en el temps. Es mostra com la utilització d'aquestes proves de qualitat temporal conjuntament amb les proves de compatibilitat espacial milloren els resultats quan es fan servir amb un mètode d'estimació òptima de la construcció concurrent de mapes i l'autolocalització de robots. La idea principal consisteix en emprar un històric dels errors en l'associació de les dades per calcular la possibilitat d'incórrer en nous errors d'associació; i excloure del mapa aquells elements dels quals les observacions no siguin consistents.

Es posa especial atenció en el fet que l'eliminació dels elements inconsistents del mapa no violi les propietats dels algorismes de construcció concurrent de mapes i autolocalització descrits en la literatura; és a dir, convergència asimptòtica i correlació completa.

Aquesta tesi proporciona també un profund anàlisi del model de construcció concurrent de mapes i autolocalització totalment correlat des d'un punt de vista de la teoria de control de sistemes. Partint del fet que el filtre de Kalman no és més que un estimador òptim, s'analitzen les implicacions de tenir un vector d'estats que es revisa a partir de mesures totalment correladas. Es revela de manera teòrica i amb experiments les limitacions d'utilitzar un enfocament per la construcció concurrent de mapes i l'autolocalització a partir de mesures totalment correladas.

El fet de tenir un model parcialment observable inhibeix la reconstrucció total de l'espai d'estats, produint tant mateix una estimació de la posició dels elements de l'entorn que depèn en tot cas de les observacions inicials, i que no garanteix la convergència a una matriu de covariància definida positivament. D'altra banda, el fet de tenir un vector d'estats parcialment controlable fa que, després d'un reduït nombre d'iteracions el filtre cregui tenir una estimació perfecta de l'estat dels elements de l'entorn; amb els corresponents guanys de Kalman convergint a zero. Per tant, després d'un reduït nombre d'iteracions del filtre, les innovacions no s'utilitzen més. Es mostra com reduir els efectes de la correlació total i de la controlabilitat parcial. A més a més, suposant que el filtre de Kalman és un observador òptim per a la reconstrucció dels estats, és pertinent construir un regulador òptim que permeti conduir el robot el més a prop possible a una trajectòria desitjada durant la construcció d'un mapa. Es mostra com la dualitat existent entre l'observabilitat i la controlabilitat es pot fer servir en el disseny d'aquest regulador òptim.

Qualsevol algorisme de construcció concurrent de mapes i autolocalització de robots mòbils que s'ha d'usar en un entorn real ha de ser capaç de relacionar les observacions



i els seus corresponents elements del mapa de manera expedita. Algunes de les proves de compatibilitat de les observacions són costoses des del punt de vista de la seva complexitat computacional, i la seva aplicació s'ha de dissenyar amb especial atenció. Es comenten els costos computacionals de les diferents proves de compatibilitat entre observacions; així com altres característiques desitjables de l'estructura de dades que es fa servir per a la construcció del mapa. A més a més es proposen una sèrie de tasques que han de realitzar-se durant l'associació de dades. Començant per les proves de compatibilitat amb un model bàsic dels elements del mapa, i continuant amb la reducció de l'espai de cerca quan es generen hipòtesis d'associació, així com les proves espacial i temporal d'associació de dades.

El treball que es presenta en aquesta tesi proposa noves tècniques en àrees de l'enginyeria i ciències computacionals, que van des de nous algorismes per la visió per computador, a idees novells de la construcció concurrent de mapes i l'autolocalització de robots mòbils. Les contribucions principals són la proposta d'una nova tècnica per la fusió de dades visuals; la formulació d'un nou algorisme per la construcció concurrent de mapes i l'autolocalització de robots que considera la qualitat temporal dels elements del mapa; nous resultats teòrics en el nivell de reconstrucció possible quan es construeixen mapes a partir d'observacions totalment correladas; i les tècniques necessàries per pal·liar els efectes de l'observabilitat i la controlabilitat parcials, així com els efectes de les no linealitats en la solució del problema de construcció concurrent de mapes i de l'autolocalització.



# Resumen

Esta tesis aborda el problema del aprendizaje automático de entornos estructurados en robótica móvil. Particularmente, la extracción de características a partir de las señales de los sensores, la construcción autónoma de mapas, y la autolocalización de robots.

Se estudian los fundamentos matemáticos necesarios para la extracción de características a partir de imágenes y registros de un láser, las cuales permiten la identificación unívoca de los elementos del entorno. Los atributos extraídos a partir de la señal de un solo sensor pueden ser insuficientes a la hora de caracterizar los elementos del entorno de forma invariante; lo que conlleva a la combinación de información de múltiples fuentes. Se presenta un nuevo algoritmo para la fusión de información complementaria extraída de dos módulos de visión de bajo nivel. Esta fusión de información produce descripciones más completas de los objetos presentes en el entorno, los cuales pueden ser seguidos y aprendidos en el contexto de la robótica móvil.

Las variaciones en las condiciones de iluminación y las oclusiones hacen que la asociación de datos en visión por computador sea una tarea difícil de llevar a cabo. Sin embargo, el uso de restricciones geométricas y fotogramétricas permiten reducir la búsqueda de correspondencias entre imágenes; y al centrar la atención en un reducido número de características, estas pueden ser seguidas en imágenes sucesivas, simplificando así el problema de asociación de datos. Se hace hincapié en las técnicas de la geometría de múltiples vistas relevantes para el cómputo de una estimación inicial de la posición de los elementos del entorno, lo cual permite la reconstrucción del movimiento del robot entre imágenes sucesivas; situación deseable cuando se carece de odometría o cuando sus lecturas son poco fiables.

Una vez que los elementos del entorno han sido extraídos e identificados, la segunda parte del problema consiste en usar estas observaciones tanto para estimar la posición del robot, como para refinar la estimación de los mismos elementos del entorno. El movimiento del robot y las lecturas de los sensores se consideran como dos procesos estocásticos, y el problema se aborda desde el punto de vista de la teoría de estimación, en donde el ruido inherente a los sensores y al movimiento del robot se consideran como secuencias aleatorias.

La principal desventaja existente en el uso de técnicas de estimación para el cómputo concurrente de la posición del robot y la construcción de un mapa, es que hasta ahora se ha considerado su uso en entornos estáticos únicamente, y que su aplicación en situaciones más realistas carece de robustez. Se propone un conjunto de funciones para evaluar la calidad temporal de las observaciones con el fin de solventar aquellas situaciones en que las observaciones de los elementos del entorno no sean consistentes en el tiempo. Se muestra como el uso de estas pruebas de calidad temporal junto con las pruebas de compatibilidad espacial existentes mejora los resultados al usar un método de estimación óptima para la construcción concurrente de mapas y la autolocalización de robots. La idea principal consiste en usar un histórico de los errores en la asociación de datos para el cómputo de la posibilidad de incurrir en nuevos errores de asociación; y eliminar del mapa aquellos elementos cuyas observaciones no sean consistentes.

Se presta especial atención a que la eliminación de elementos inconsistentes del mapa no viole las propiedades de los algoritmos de construcción concurrente de mapas y autolocalización descritos en la literatura; es decir, convergencia asintótica y correlación completa.

Esta tesis proporciona a su vez un análisis en profundidad del modelo de construcción concurrente de mapas y autolocalización totalmente correlado desde un punto de vista de la teoría de control de sistemas. Partiendo del hecho de que el filtro de Kalman no es otra cosa que un estimador óptimo, se analizan las implicaciones de tener un vector de estados que se revisa a partir de mediciones totalmente correladas. Se revela de forma teórica y con experimentos las limitaciones de usar un enfoque para la construcción concurrente de mapas y autolocalización a partir de mediciones totalmente correladas.

El hecho de tener un modelo parcialmente observable inhibe la reconstrucción total del espacio de estados, produciendo a su vez una estimación de la posición de los elementos del entorno que dependerá en todo caso de las observaciones iniciales, y que no garantiza la convergencia a una matriz de covarianza positivamente definida. Por otro lado, el hecho de tener un vector de estados parcialmente controlable, produce después de un reducido número de iteraciones que el filtro crea tener una estimación perfecta del estado de los elementos del entorno; con sus correspondientes ganancias de Kalman convergiendo a cero. Esto es, después de un pequeño número de iteraciones del filtro, las innovaciones no se usan. Se muestra como reducir los efectos de la correlación total y la controlabilidad parcial. Además, dado que el filtro de Kalman es un observador óptimo para la reconstrucción de los estados, es pertinente construir un regulador óptimo que permita conducir al robot lo más cerca posible de una trayectoria deseada durante la construcción de un mapa. Se muestra como la dualidad existente entre la observabilidad y la controlabilidad se puede emplear en el diseño de este regulador óptimo.

Cualquier algoritmo de construcción concurrente de mapas y autolocalización de

robots móviles que deba funcionar en un entorno real deberá ser capaz de relacionar las observaciones y sus correspondientes elementos del mapa de manera expedita. Algunas de las pruebas de compatibilidad de las observaciones son caras desde el punto de vista de su complejidad computacional, y su aplicación debe diseñarse con riguroso cuidado. Se comentan los costes computacionales de las distintas pruebas de compatibilidad entre observaciones; así como otras características deseadas de la estructura de datos elegida para la construcción del mapa. Además, se propone una serie de tareas que debe llevarse a cabo durante la asociación de datos. Partiendo por las pruebas de compatibilidad con un modelo básico de los elementos del mapa, y continuando con la reducción del espacio de búsqueda al generar hipótesis de asociación, así como las pruebas espacial y temporal de asociación de datos.

El trabajo que se presenta en esta tesis propone nuevas técnicas en áreas de la ingeniería y las ciencias computacionales, que van desde nuevos algoritmos de visión por computador, a ideas noveles en la construcción concurrente de mapas y la autocalización de robots móviles. Las contribuciones principales son la propuesta de una nueva técnica para la fusión de datos visuales; la formulación de un nuevo algoritmo para la construcción concurrente de mapas y autocalización de robots que toma en cuenta la calidad temporal de los elementos del mapa; nuevos resultados teóricos en el grado de reconstrucción posible al construir mapas a partir de observaciones totalmente correladas; y las técnicas necesarias para paliar los efectos de la observabilidad y controlabilidad parciales, así como los efectos de las no linealidades en la solución del problema de construcción concurrente de mapas y autocalización.



To Verónica Ana





# Acknowledgments

I would like to thank a number of people for their invaluable help during the completion of this research. First and foremost to my advisor Alberto Sanfeliu, from whom I have learned that in a field such as engineering, theoretically sound ideas are of little value if they are not accompanied with real working implementations. To Avi Kak, for believing in me, and asking me to be part of the Robot Vision Laboratory during my stay at Purdue University.

To Guilherme in the US, and to Carme and Fede in Spain, with whom I have maintained uncountable fruitful discussions during my years as a graduate student. To Chiren, Lynne, Jennifer, Christina, Akio, Yuichi, Jean, and Aleix, for making my stay at RVL a delightful experience; and to Guillem, Francesc, Jaume, Josep María, Adrià, María, and Eduardo, for bearing with me during hectic times at IRI. Also, to Juan Domingo and Juan José, whose external reviews contributed substantially to the revision of an early draft of this manuscript.

This thesis would not have reached completion if it were not thanks to the financial support from various institutions. Primarily, to a graduate studies scholarship from the Mexican Council of Science and Technology and CETYS Universidad. Also, to teaching and research assistantships from the Schools of Liberal Arts and Engineering at Purdue University. And most recently, to an engineering contract from the Spanish Council of Science and Technology under projects TAP-98-0473, and DPI-2001-2223.

I am grateful also to Ana, Vero, Lucio, Carmen, and Miguel for their unconditional emotional support. And most especially, thank you María for being at my side.



# Notation

## Data fusion

$d$	Mahalanobis distance
$\mathbf{d}$	disparity
$\gamma$	initial sample
$\Gamma$	small regions
$\mathbf{n}$	normal vector
$\omega$	initial class
$\Omega$	large regions
$\Omega'$	object regions
$p$	region perimeter
$\mathbf{p}$	a point in a large region
$r$	a point in the disparity image
$R$	image region
$\mathbf{R}$	segmented image regions
$\Sigma$	class covariance
$\mathbf{S}$	smooth surface segments
$\mathbf{V}$	class association votes

## Motion recovery

$\mathbf{a}$	quadratic combination of the elements of $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}'$
$C$	largest consensus set
$\delta$	distance threshold (algebraic or symmetric)
$\mathbf{d}$	disparity vector
$\mathbf{e}, \mathbf{e}'$	right and left epipoles

$\mathbf{f}$	fundamental matrix in vector form
$\mathbf{F}$	fundamental matrix
$\mathbf{h}$	homography in vector form
$\mathbf{H}$	homography
$I_k, I_{k+1}$	consecutive image frames
$\mathbf{K}$	camera intrinsic parameters
$\mathbf{m}$	image point
$N$	number of samples in Ransac
$\tilde{\mathbf{p}}_e$	Euclidean coordinates of a point
$\tilde{\mathbf{p}}_p$	projective coordinates of a point
$\tilde{\mathbf{P}}, \tilde{\mathbf{P}}'$	canonical perspective projection matrices
$\tilde{\mathbf{P}}_e, \tilde{\mathbf{P}}'_e$	perspective projection matrices for Euclidean reconstruction
$S$	random sample set
$\tilde{\mathbf{x}}, \tilde{\mathbf{x}}'$	homogeneous coordinates of image point correspondences
$X, X'$	set of image point correspondences

### Robot localization

$\mathbf{0}$	a vector or a matrix of zeros
$\alpha_d, \alpha_\psi, \alpha_\theta$	dead reconing error factors
$\beta_d, \beta_\psi$	measurement error factors
$\mathbf{f}(\cdot)$	robot motion model
$\mathbf{F}_v$	motion error Jacobian
$\mathbf{F}_x$	motion Jacobian
$\mathbf{h}(\cdot)$	measurement model
$\mathbf{H}_w$	measurement error Jacobian
$\mathbf{H}_x$	measurement Jacobian
$k + 1 k$	<i>a priori</i> estimate, i.e, given observations up until step $k$
$k + 1 k + 1$	<i>a posteriori</i> estimate, i.e, given observations up until step $k + 1$
$\mathbf{K}$	Kalman gain
$m$	total number of filter iterations
$n$	total number of landmarks in the map
$\mathbf{P}_{\cdot  \cdot}$	error covariance estimate

$\mathbf{P}_{ff,\cdot \cdot}$	map error covariance estimate
$\mathbf{P}_{qq,\cdot \cdot}$	landmark quality error covariance estimate
$\mathbf{P}_{rr,\cdot \cdot}$	robot pose error covariance estimate
$\mathbf{R}, \mathbf{t}$	robot pose
$\mathbf{S}$	innovation covariance
$\mathbf{u}_k$	motion command
$u_{q,k}^{(i)}$	landmark identification stamp
$U^n$	set of motion commands
$\mathbf{v}_k$	motion error
$\mathbf{V}_k$	motion error covariance
$\mathbf{w}_k$	measurement error
$\mathbf{W}_k$	measurement error covariance
$\mathbf{x}_f^{(i)}$	landmark with label $i$
$\mathbf{x}_k$	desired robot pose
$x_{q,k}^{(i)}$	landmark quality
$\mathbf{x}_{r,k}$	robot pose
$\mathbf{x}_{r,\cdot \cdot}$	robot pose estimate
$X^n$	desired path
$\tilde{\mathbf{x}}_{r,\cdot \cdot}$	robot pose error estimate
$\mathbf{z}_k$	measurement vector
$\tilde{\mathbf{z}}_{\cdot \cdot}$	measurement error estimate
$\mathbf{z}_k$	hypothetical noise-free measurement
$Z^k$	set of observations up until time step $k$

### Fully correlated maps

$\mathbf{0}_{m \times n}$	a matrix of zeros of size $m$ by $n$
$\mathbf{1}_{m \times n}$	a matrix of ones of size $m$ by $n$
$\alpha$	angle between two subspaces
$I$	regulator performance index
$\mathbf{J}$	Fisher information matrix
$\Lambda(\cdot)$	likelihood function
$\text{Im}\mathbf{Q}$	controllable subspace

$\text{Im}\mathbf{R}^\top$	observable subspace
$\text{Ker}\mathbf{Q}^\top$	uncontrollable subspace
$\text{Ker}\mathbf{R}$	unobservable subspace
$\mathbf{Q}$	controllability matrix
$\tilde{\mathbf{Q}}$	Kalman controllability matrix
$\mathbf{R}$	observability matrix
$\hat{\mathbf{u}}_k$	optimal motion command
$\mathbf{U}$	orthonormal basis
$\mathbf{U}\Sigma\mathbf{V}^\top$	singular value decomposition

### Map building

$A$	set of landmark appearance properties
$D$	set of landmark observations
$F$	set of erased landmarks
$H$	set of landmark match hypotheses
$k$	number of features per landmark
$m$	number of observations
$M$	set of landmark matches
$n$	number of landmarks in the map
$N$	set of new map entries
$T$	total map
$T'$	pruned map

### Landmark extraction

$\alpha_u, \alpha_v$	focal lengths
$\nabla^2$	image Laplacian
$\nabla I$	image gradient
$\nabla_x, \nabla_y$	edge detection operators
$C$	laser scan raw data
$\mathbf{C}$	smoothed square image gradient
$d, \theta$	Hough space features
$\mathbf{d}, \mathbf{D}$	affine motion parameters

$\varepsilon$	energy sum of gray level intensity variances
$E$	inter-pixel image color distances (as graph edges)
$\mathbf{F}$	gradient square
$G$	image graph
$\mathbf{G}$	Gaussian operator (smoothing operator)
$h$	one dimensional Canny operator
$\mathbf{H}$	image Hessian
$\mathbf{I}$	identity matrix
$I, J$	image intensities
$I_x, I_y$	one directional image gradients
$K_B, K_H, K_K$	corner response functions
$\mathbf{m}$	image point
$\tilde{\mathbf{P}}$	perspective projection matrix
$R_i$	image region
$\mathbf{R}, \mathbf{t}$	camera pose
$\Sigma, \Lambda, x_{max}$	Canny criteria
$S$	region pixels
$S'$	border pixels
$\bar{S}$	region complement
$(u_0, v_0)$	principal point
$\mathbf{v}_i$	laser range point
$\overline{\mathbf{v}_i \mathbf{v}_j}$	laser range line
$V$	image pixels (as graph nodes)
$W$	window of interest
$(x, y)$	pixel coordinates
$\tilde{\mathbf{x}}$	homogeneous coordinates of an image point

### Camera calibration

$\alpha_u, \alpha_v$	focal lengths in pixels
$f$	focal length
$I$	image plane
$k_1, k_2$	radial distortion parameters

$O$	origin of camera coordinate frame
$\tilde{\mathbf{P}}$	perspective projection matrix
$\mathbf{q}$	image point coordinates
$\tilde{\mathbf{q}}$	homogeneous image point coordinates
$\mathbf{R}, \mathbf{t}$	camera pose
$(u_0, v_0)$	principal point
$\mathbf{x}$	Euclidean point coordinates
$\tilde{\mathbf{x}}$	homogeneous 3D point coordinates

### **Kalman filter**

$\mathbf{e}_i$	state error estimate
$\mathbf{f}(\cdot)$	nonlinear plant model
$\mathbf{F}$	linear plant model
$\mathbf{h}(\cdot)$	nonlinear measurement model
$\mathbf{H}$	linear measurement model
$\nabla_{\mathbf{x}}, \nabla_{\mathbf{h}}$	model Jacobians
$\mathbf{K}$	Kalman gain
$\mathbf{P}_i$	state error covariance estimate
$\mathbf{u}$	input vector
$\mathbf{v}$	plant noise
$\mathbf{V}$	plant error covariance
$\mathbf{w}$	measurement noise
$\mathbf{W}$	measurement error covariance
$\mathbf{x}$	state vector
$\mathbf{x}_i$	state estimate
$\mathbf{z}$	measurement vector
$\mathbf{z}_i$	measurement estimate
$Z^j$	measurement sequence



# Contents

Abstract . . . . .	iii
Resum . . . . .	vii
Resumen . . . . .	xi
Acknowledgments . . . . .	xvii
Notation . . . . .	xix
List of Tables . . . . .	xxxii
List of Figures . . . . .	xxxiv
<b>1 Introduction</b>	<b>1</b>
<b>2 Landmark Extraction for Mobile Robotics</b>	<b>11</b>
2.1 Low-level processing of 2d intensity images . . . . .	11
2.2 Low-level processing of laser data . . . . .	14
2.3 Landmark characterization . . . . .	14
2.4 Conclusions . . . . .	16
<b>3 Data Fusion for Object and Landmark Characterization</b>	<b>17</b>
3.1 Integration of perceptual grouping and depth using supervised classification . . . . .	18
3.2 Bibliographical notes . . . . .	21

3.3	Conclusions . . . . .	23
<b>4</b>	<b>Combining Information from Sequential Views</b>	<b>25</b>
4.1	Feature tracking . . . . .	26
4.1.1	SSD, SAD . . . . .	26
4.1.2	Illumination changes and occlusions . . . . .	27
4.2	Epipolar geometry . . . . .	27
4.2.1	RANSAC . . . . .	28
4.3	Motion recovery . . . . .	30
4.3.1	Projective reconstruction . . . . .	30
4.3.2	Euclidean reconstruction from knowledge of the scene . . . . .	30
4.3.3	Euclidean reconstruction from the camera intrinsic parameters . . . . .	33
4.3.4	Euclidean reconstruction from a fully calibrated stereo rig . . . . .	35
4.4	Bibliographical notes . . . . .	36
4.5	Conclusions . . . . .	36
<b>5</b>	<b>Robot Localization</b>	<b>37</b>
5.1	Preliminaries . . . . .	39
5.1.1	Full covariance EKF approach to CML . . . . .	39
5.1.2	Map model . . . . .	40
5.1.3	Prediction . . . . .	43
5.1.4	Correction . . . . .	43
5.1.5	Convergence properties of the full covariance EKF approach to CML . . . . .	44
5.1.6	Sequential innovation . . . . .	45
5.1.7	Covariance initialization . . . . .	46

5.1.8	Landmark spatial uncertainty . . . . .	47
5.1.9	Individual compatibility test . . . . .	48
5.1.10	Joint compatibility test . . . . .	49
5.2	Mobile robot platforms . . . . .	53
5.2.1	Monobot . . . . .	53
5.2.2	Planar mobile robot . . . . .	54
5.3	Model assumptions . . . . .	56
5.3.1	Path planning . . . . .	57
5.3.2	Dead reckoning error . . . . .	59
5.3.3	Observation error . . . . .	62
5.4	Temporal landmark validation . . . . .	66
5.4.1	Landmark temporal uncertainty . . . . .	66
5.4.2	Nonlinear model for temporal landmark quality: the exponential decay rule . . . . .	67
5.4.3	Linear model for temporal landmark quality: the data association probability . . . . .	68
5.4.4	Temporal landmark quality test . . . . .	70
5.5	Performance of the CML-LV algorithm . . . . .	75
5.5.1	Sequential innovation . . . . .	75
5.5.2	Partial observations . . . . .	80
5.5.3	Data missassociation . . . . .	82
5.5.4	Landmark validation . . . . .	84
5.6	Divergence . . . . .	92
5.7	Bibliographical notes . . . . .	93
5.8	Conclusions . . . . .	96

<b>6</b>	<b>Fully Correlated Maps</b>	<b>99</b>
6.1	Steady state behavior of EKF-CML . . . . .	101
6.2	Total Fisher information and the Cramer Rao lower bound . . . . .	103
6.3	Partial observability and controllability . . . . .	106
6.4	The monobot . . . . .	108
6.5	The planar robot . . . . .	116
6.6	Complete observability . . . . .	121
6.6.1	A fixed global reference . . . . .	122
6.6.2	An external sensor . . . . .	123
6.6.3	Planar vehicle . . . . .	128
6.7	Controllability . . . . .	128
6.8	Optimal control . . . . .	133
6.9	Divergence . . . . .	139
6.10	Bibliographical notes . . . . .	142
6.11	Conclusions . . . . .	142
<b>7</b>	<b>Map Building</b>	<b>145</b>
7.1	AVL trees . . . . .	145
7.2	Multidimensional search trees . . . . .	146
7.3	Data association . . . . .	148
7.3.1	Model compatibility . . . . .	150
7.3.2	Search space reduction . . . . .	151
7.3.3	Hypotheses formation . . . . .	151
7.3.4	Association . . . . .	152
7.4	Map update . . . . .	153

7.5	Bibliographical notes . . . . .	156
7.6	Conclusions . . . . .	158
<b>8</b>	<b>Experimental Results</b>	<b>159</b>
8.1	Correction of a visual landmark estimate . . . . .	159
8.2	Self localization and map building from visual landmarks . . . . .	162
8.3	Self localization and map building with a laser range scanner . . . . .	167
8.4	Conclusions . . . . .	171
<b>9</b>	<b>Conclusions</b>	<b>173</b>
9.1	Concluding remarks . . . . .	173
9.2	Future research directions . . . . .	175
9.3	Selected publications . . . . .	176
 <b>Appendix A</b>		
	Landmark Extraction . . . . .	179
A.1	Low-level processing of 2d intensity images . . . . .	179
A.1.1	Corners . . . . .	182
A.1.2	Shi and Tomasi features . . . . .	186
A.1.3	Edges . . . . .	187
A.1.4	Image Segmentation . . . . .	192
A.1.5	Depth from stereo . . . . .	194
A.2	Low-level processing of laser data . . . . .	199
A.3	Landmark characterization . . . . .	202
A.4	Bibliographical notes . . . . .	203

**Appendix B**

Camera Calibration . . . . .	205
B.1 Camera modeling . . . . .	206
B.1.1 Perspective projection . . . . .	206
B.1.2 Pattern reference frame . . . . .	207
B.1.3 Pixel size . . . . .	208
B.1.4 Image center . . . . .	209
B.1.5 Non orthogonal image axes . . . . .	209
B.1.6 Lens distortion . . . . .	209
B.2 Camera calibration . . . . .	210
B.2.1 Perspective projection matrix . . . . .	210
B.2.2 Necessary conditions on $\tilde{\mathbf{P}}$ . . . . .	211
B.2.3 Least squares solution . . . . .	212
B.2.4 Scale . . . . .	213
B.2.5 Orthonormality of $\mathbf{R}$ . . . . .	213
B.3 Computing distortion parameters . . . . .	215
B.4 Bibliographical notes . . . . .	215

## Appendix C

The Kalman Filter . . . . .	217
C.1 Recursive state estimation . . . . .	217
C.2 Linear Kalman filter . . . . .	218
C.3 Extended Kalman filter . . . . .	221
C.4 Conditioning . . . . .	222
C.5 Sequential innovation . . . . .	222
C.6 Bibliographical notes . . . . .	224

## Appendix D

Concepts from Linear Algebra . . . . .	225
D.1 Properties of positive semi-definite matrices . . . . .	225
D.2 Linear subspaces . . . . .	225
D.3 Inverse of a block matrix . . . . .	226
D.4 The matrix inversion lemma . . . . .	226
D.5 Matrix inequalities . . . . .	226
D.6 Bibliographical notes . . . . .	227

<b>Bibliography</b>	<b>228</b>
---------------------	------------





# List of Tables

4.1	The random sample consensus algorithm, adapted from [91] to the robust estimation of the fundamental matrix from point correspondences. . . . .	29
5.1	The Pearson $\chi^2$ distribution needed for the computation of the square Mahalanobis distance threshold $d^2 = \chi_{\dim \mathbf{z}^{(p\dots r)}, \alpha}^2$ during the validation of the hypothesis test that measurements $\mathbf{z}^{(p\dots r)}$ correspond to features $\mathbf{x}_f^{(p\dots r)}$ with probability $\alpha = 0.95$ . . . . .	50
5.2	The Pearson $\chi^2$ distribution needed for the computation of the square Mahalanobis distance threshold $d^2 = \chi_{\dim \mathbf{z}^{(p\dots r)}, \alpha}^2$ during the validation of the hypothesis test that measurements $\mathbf{z}^{(p\dots r)}$ correspond to features $\mathbf{x}_f^{(p\dots r)}$ with probability $\alpha = 0.99$ . . . . .	51
7.1	Algorithm for map update. . . . .	154
A.1	Several differentiation masks can be used to approximate the derivatives of an image. . . . .	181
A.2	Epipolar rectification geometric constraints. . . . .	196
A.3	Recursive version of the Douglas and Peucker algorithm to line simplification. . . . .	200
A.4	Algorithm to find the vertex $\mathbf{v}_k$ farthest from the line $\overline{\mathbf{v}_i \mathbf{v}_j}$ , and the squared distance to it. . . . .	200



# List of Figures

1.1	System architecture. . . . .	8
1.2	Marco mobile robot. . . . .	9
3.1	Data fusion steps. . . . .	21
4.1	Three consecutive frames of a sequence acquired by a mobile robot in motion. The rough characteristics of the terrain make dead reckoning computation unreliable. The red dots indicate the set of tracked Shi and Tomasi features. . . . .	27
4.2	Projective reconstruction of structure from sequential views. Point correspondences in between two views of a three-dimensional calibration pattern are computed. Frame (a) shows the projection of all point correspondences on top of an image of a portion of a calibration model (made up of three planes as shown in Figure B.1.2. The camera was then moved, and a second series of snapshots of the calibration pattern were taken. Frame (b) shows the projective reconstruction, estimated only from the computed fundamental matrix. . . . .	31
4.3	Euclidean reconstruction of structure from sequential views. Full metric reconstruction is possible to attain when some ground truth of the scene is available. The recovered camera motion is represented by the plotted coordinate frames. In this example, we consider the knowledge of the real coordinates of only five points of the calibration pattern, two on extremes of first plane, one on the middle of the second plane, and two on opposite extremes of the last plane, all shown in red on frame (b). . . . .	32
4.4	Point correspondences for a calibrated stereo pair for an outdoor mobile robot sequence. . . . .	34
4.5	Recovered robot motion in between frames 3 and 4 of the outdoor image sequence. . . . .	34

5.1	Location estimate with a predictive filter. . . . .	40
5.2	Spatial compatibility tests. Individual and joint compatibility test performance for a map with 3 and 10 landmarks over 100 steps. . . . .	52
5.3	<i>Monobot</i> , a one-dimensional mobile robot. . . . .	54
5.4	Two-dimensional mobile robot motion model. . . . .	55
5.5	Winding path $X^m$ for various values of $m$ . . . . .	58
5.6	Motion noise model. Motion noise regarded as white Gaussian. The ellipse indicates a 95% confidence bound around the a priori motion estimate $\mathbf{x}_{r,k+1 k}$ . . . . .	60
5.7	Dead reckoning error for various values of $\alpha_\psi$ , $\alpha_d$ , $\alpha_\theta$ . . . . .	61
5.8	Measurement noise model. Motion noise regarded as white Gaussian. The ellipse indicates a 95% confidence bound around the landmark estimate $\mathbf{x}_{f,k+1 k}$ . . . . .	64
5.9	Measurement error distribution for various values of $\beta_d$ and $\beta_\psi$ . . . . .	65
5.10	Landmark quality models. . . . .	69
5.11	Landmark quality test for a test run with 10 landmarks and 100 steps. $\alpha = 1$ , $\beta = 1$ , and $a = 0.5$ . The plots show landmarks labeled 2 and 9 being reinitialized near steps 30 and 50 respectively. . . . .	71
5.12	Evolution of the entries in the covariance matrix for a monobot with 3, 2, and 3-then-2 landmarks. . . . .	73
5.13	Asymptotic convergence of the determinant of the map error covariance matrix for a monobot with 3, 2, and 3-then-2 landmarks. . . . .	74
5.14	Full-covariance EKF CML for a path with 100 iterations and 10 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ . . . . .	76
5.15	Difference in localization error between the full covariance EKF version of the algorithm, and the results of applying sequential innovation to the same set of observations. . . . .	77
5.16	Robot localization error estimate. EKF CML with sequential innovation for a path with 100 iterations and 10 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ . . . . .	77

5.17	The landmark noise covariance estimate is monotonically decreasing in the EKF-CML algorithm. . . . .	78
5.18	Landmarks become more and more correlated as iterations take place in the EKF-CML algorithm. The figure shows the increase in the cross correlation coefficients for landmark 1 with respect to each other landmark, in both the $x$ and $y$ directions. . . . .	79
5.19	Full-covariance EKF CML for a path with 100 iterations and 10 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of $2m$ . . . . .	81
5.20	Full-covariance EKF CML for a path with 100 iterations and 10 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of $2m$ . Data missassociation occurs within a radius of $1m$ with a 25% probability. . . . .	83
5.21	Results of using the individual compatibility test. Full-covariance EKF CML for a path with 100 iterations and 10 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of $2m$ . Data missassociation occurs within a radius of $1.0m$ with a 25% probability. . . . .	86
5.22	Results of using the individual compatibility test and the exponential decay rule to test temporal landmark quality, $x_{q,LOW} = 0.6590$ , $x_{q,THLD} = 0.66$ , $x_{q,HIGH} = 0.8659$ . Full-covariance EKF CML for a path with 100 iterations and 10 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of $2m$ . Data missassociation occurs within a radius of $1.0m$ with a 25% probability. . . . .	87
5.23	Results of using the individual compatibility test and the data association probability as temporal landmark quality test, $x_{q,LOW} = 0$ , $x_{q,THLD} = 0.03$ , $x_{q,HIGH} = 1$ . Full-covariance EKF CML for a path with 100 iterations and 10 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of $2m$ . Data missassociation occurs within a radius of $0.5m$ with a 10% probability. . . . .	88

5.24	Robot localization error estimate. Comparison between ICT: Individual spatial compatibility test with limited vision range and data association errors (realistic case of CML); and the improvements proposed, ICT+DAP: EKF-CML-LV with both spatial and temporal landmark quality tests, using the data association probability, and ICT+EDR: EKF-CML-LV with both spatial and temporal landmark quality tests, using the exponential decay rule. The simulation is run over a path with 100 iterations and 10 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of $2m$ . Data missassociation occurs within a radius of $1.0m$ with a 25% probability. . . . .	89
5.25	Robot localization error estimate. Comparison between ICT: Individual spatial compatibility test with limited vision range and data association errors (realistic case of CML); and the improvements proposed, ICT+DAP: EKF-CML-LV with both spatial and temporal landmark quality tests, using the data association probability, and ICT+EDR: EKF-CML-LV with both spatial and temporal landmark quality tests, using the exponential decay rule. The simulation is run over a path with 100 iterations and 50 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of $2m$ . Data missassociation occurs within a radius of $1.0m$ with a 25% probability. . . . .	90
5.26	Robot localization error estimate. Comparison between ICT: Individual spatial compatibility test with limited vision range and data association errors (realistic case of CML); and the improvements proposed, ICT+DAP: EKF-CML-LV with both spatial and temporal landmark quality tests, using the data association probability, and ICT+EDR: EKF-CML-LV with both spatial and temporal landmark quality tests, using the exponential decay rule. The simulation is run over a path with 200 iterations and 50 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of $2m$ . Data missassociation occurs within a radius of $1.0m$ with a 25% probability. . . . .	91
6.1	Final vehicle and landmark localization variances after 500 iterations of CML for a monobot with initial localization variance $\sigma_{r,0 0}^2 = 1$ , and various levels of plant and sensor noise. . . . .	102
6.2	First entry in the total Fisher information matrix for a monobot with variance parameters $\sigma_{r,0 0}^2 = \sigma_v^2 = \sigma_w^2 = 1$ , and various sizes for the measurement vector. . . . .	105
6.3	Controllability and observability of the state space in CML. . . . .	108

6.4	Controllability and observability of a linear one dimensional mobile robot with one landmark. . . . .	109
6.5	$\angle \mathcal{R}(\mathbf{Q})\mathcal{R}(\mathbf{R}^\top)$ . Angle between the observable and controllable subspaces. . . . .	112
6.6	Full-covariance EKF CML for monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . . . . .	114
6.7	Full-covariance EKF CML for monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . . . . .	115
6.8	Reduction of the average monobot localization error $\mathbf{x}_{r,k} - \mathbf{x}_{r,k k}$ with respect to the number of landmarks used. The results correspond to a Montecarlo simulation over 100 CML runs. The dotted lines show the extent of the data for the entire set of runs, and the boxes contain marks at the lower, median and upper quartile. . . . .	117
6.9	Controllability and observability subspaces. . . . .	118
6.10	Full-covariance fully observable CML for a monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . The global reference is observed at the origin, also with $\beta_d = 0.1$ . . . . .	124
6.11	Full-covariance fully observable CML for a monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . The global reference is observed at the origin, also with $\beta_d = 0.1$ . . . . .	125
6.12	Full-covariance fully observable CML for a monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . A fixed external sensor is used for the measurement of the vehicle pose, also with $\beta_d = 0.1$ . . . . .	126
6.13	Full-covariance fully observable CML for a monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . A fixed external sensor is used for the measurement of the vehicle pose, also with $\beta_d = 0.1$ . . . . .	127

6.14	Full-covariance fully observable CML with artificial process noise for a monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . A fixed external sensor is used for the measurement of the vehicle pose, also with $\beta_d = 0.1$ . A pseudo-noise covariance ( $\mathbf{V}_f = 0.05^2\mathbf{I}$ ) is used to prevent the landmark Kalman gains from converging to zero. . . . .	130
6.15	Full-covariance fully observable CML with artificial process noise for a monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . A fixed external sensor is used for the measurement of the vehicle pose, also with $\beta_d = 0.1$ . A pseudo-noise covariance ( $\mathbf{V}_f = 0.05^2\mathbf{I}$ ) is used to prevent the landmark Kalman gains from converging to zero. . . . .	131
6.16	The Cramer Rao lower bound holds only when the eigenvalues of $(\mathbf{P}_{k k} - \mathbf{J}^{-1})$ are positive. . . . .	132
6.17	Full-covariance EKF CML for a path with 1000 steps and 10 landmarks, zoomed view. . . . .	133
6.18	Structure of the optimal state estimator and regulator in CML. . . . .	135
6.19	Optimal state estimation and regulation for a monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . . . . .	136
6.20	Optimal state estimation and regulation for a monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . . . . .	137
6.21	Optimal state estimation and regulation for a monobot in a sinusoidal path from $\mathbf{x}_{r,0 0} = -1m$ to $\mathbf{x}_{100} = -1m$ , with 100 iterations and $\alpha_d = 0.1$ , and $\beta_d = 0.1$ . . . . .	138
6.22	Full-covariance EKF CML for a path with 100 iterations and 10 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ , and full observability. An external sensor measures the robot position at each iteration. . . . .	140
6.23	Full-covariance EKF CML for a path with 100 iterations and 10 landmarks, $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , $\beta_\psi = \beta_d = 0.01$ , full observability with artificial process noise. An external sensor measures the robot position at each iteration, and pseudo noise covariance is added for the landmark location estimates. . . . .	141
7.1	Map representation using AVL trees in higher dimensions. . . . .	147



7.2	A set of nine landmarks. . . . .	148
7.3	Insertion of nine data points in a bidimensional AVL tree. . . . .	149
7.4	Map representation using AVL trees in higher dimensions. . . . .	150
7.5	Wall compatibility tests of a laser range scan. . . . .	151
7.6	Hypothesis search range for walls extracted from a laser range scan. . .	152
7.7	Only the set of landmarks from the map that fall within the robot field of view are compared to the set $D$ of landmarks coming from the landmark extraction module. $M$ is the set of matched landmarks, and during the map update process their uncertainty parameters and existence state are updated. New entries $N$ are added to the map for the elements in $D$ that did not contain a corresponding match in the map; and possibly, a set $F$ of landmarks is removed from the map if their existence state is below a desired threshold. The figure shows the position of the robot at various time instances. The triangles represent the field of view of the robot, and the landmarks are represented by dots. . . . .	155
8.1	Landmark measurements and position estimate. . . . .	160
8.2	Three iterations of concurrent map building and localization over a real scenario with visual landmarks from point correspondences. The image shows how landmark covariance estimates are proportional to the size of the measurement. The further an observation is from the robot, the larger the uncertainty in its location estimate. . . . .	161
8.3	Graphical user interface for the extraction of visual landmarks from stereoscopic images. . . . .	163
8.4	Concurrent Map Building and Localization with 50 visual landmarks. .	164
8.5	Three-dimensional map reconstruction from visual cues. The pictures indicate the four cardinal views of the scenario. . . . .	165
8.6	CML from visual point correspondences. Robust data association is difficult to achieve when using visual data only. A large quantity of image correspondences do not pass the spatial and temporal compatibility tests and end up being removed from the map. . . . .	166

8.7	Mobile robotics graphical user interface developed jointly at IIIA and IRI. The interface helps researches develop and test algorithms on navigation, localization, and robot-human interaction with a variety of mobile robot platforms. The snapshot shows a test run of our map building system. . . . .	168
8.8	The blue dots indicate sensor raw data coming from a laser range finder. The green lines represent walls inferred from consecutive readings. The red lines indicate the estimated robot trajectory. . . . .	169
8.9	Concurrent mobile robot localization and map building. Front and back views of the robot initial position. . . . .	170
8.10	Graphical representation of the map built. . . . .	170
A.1	Corner response functions. . . . .	183
A.2	Feature detection over a typical mobile robotics scene. . . . .	184
A.3	Variance descent approach to corner refinement. . . . .	185
A.4	Shi and Tomasi features over a typical mobile robotics scene. . . . .	187
A.5	Edge detection over a typical mobile robotics scene. . . . .	189
A.6	The Hough transform maps straight lines in the image space into single points in the $(d, \theta)$ space. It can be used to group together unconnected straight line segments produced by an edge operator. . . . .	191
A.7	Color image segmentation . . . . .	193
A.8	Epipolar image rectification makes conjugate epipolar lines collinear and parallel to one of the image axes. This transformation reduces the time complexity of the search for matches in dense stereo from a two dimensional search problem to only one dimension. . . . .	195
A.9	Epipolar image rectification over a typical mobile robotics scene. . . . .	197
A.10	Disparity map computed from the set of rectified images in Figure A.9. Image thanks to F. Moreno. . . . .	198
A.11	Wall extraction by means of line simplification from a typical laser scan. The blue dots in the top view indicate the actual laser readings. The fitted walls are represented by the red and green lines (planes). Image thanks to A. Checa. . . . .	201

B.1	Pinhole camera model. . . . .	207
B.2	Pattern reference frame. . . . .	208



# Chapter 1

## Introduction

Efficient mobile robot navigation is limited mainly by the ability of a robot to perceive and interact with its surroundings in a deliberative way. A desirable characteristic a mobile robot must have are the skills needed to recognize the landmarks and objects that surround it, and to be able to localize itself relative to its workspace. This knowledge is crucial for the successful completion of intelligent navigation tasks. But, for such interaction to take place, a model or description of the environment needs to be specified beforehand.

If a global description or measurement of the elements present in the environment is available, the problem consists on the interpretation and matching of sensor readings to previously stored object models. Moreover, if we know that the recognized objects are fixed and persist in the scene, they can be regarded as landmarks, and can be used as reference points for self-localization. If on the other hand, a global description or measurement of the elements in the environment is not available, at least the descriptors and methods that will be used for the autonomous building of one are required. This is, either the robot has a global map, or it is given the means to learn one.

Consider for example the case of the recently deployed robotic museum tour guides Rhino and Minerva [77, 192]. These robots carry a description of the geometry of the exhibition galleries where they navigate, and of the actual position of the items in display within each gallery. By matching sensory readings to previously stored map representations, they are able to navigate for long periods of time without bumping into the objects in display, and keeping track of their location at all times, even in crowded areas. Another system capable of navigating in cluttered indoor environments is the FINALE system [61, 119, 153]. The robot can navigate and self-localize with respect to a previously stored CAD model of the environment at an average speed of 17m/min by the examination of those portions of the camera image that contain low-level features

in the vicinity of reprojected model features, the extent of the vicinity being determined by the uncertainty in the position of the robot.

In contraposition, consider the case of automated highway systems (AHS) [27, 51, 188, 189, 190]. These navigation systems usually have a model definition of typical road landmarks, such as the centerline or the sidelines, that when compared to sensor data, aid on the automatic detection of lane changes. In the AHS problem domain, an explicit map description of the environment may not be required. Instead, object model descriptions of the possible obstacles are included in the AHS core algorithms in order to reduce the risk for collisions. While local positioning within the lane and with respect to neighboring vehicles is crucial for reactive behaviors such as obstacle avoidance or lane changing, global position estimates are required only if one wants to achieve more intelligent navigation tasks. That is, tasks that include the reach of a particular goal through path planning and execution. Coarse global positioning in AHS is usually accomplished via satellite GPS signals.

The research reported in this thesis focuses on the various aspects of autonomous environment learning for indoor service robots; particularly, on landmark extraction from sensor data, autonomous map building, and robot localization.

Many systems that incorporate human-made models of the environment for indoor mobile robot navigation have been successfully developed, even when only an approximate map is given, or in crowded environments. Some recent examples include the ones by Fox *et al.* [77], Kosaka and Kak [119], Ohya *et al.* [153], or Thrun [192]. However, little success has been attained in the autonomous building of a global, and possibly dynamic, map of the environment for a mobile robot.

Several difficulties arise during self-localization, and consequently, during the autonomous building of an environment map for an indoor mobile robot:

**Lack of global references:** Unlike AHS, it is not common for service robots to include satellite GPS or other global positioning mechanisms. Their current location in the environment has to be interpreted from local sensor readings, without the aid of global references.

**Dead reckoning:** As a robot moves, its position estimate from encoder readings accumulates drift errors, due to wheel slippage or encoder quantization. And, after a small period of time, unless corrected, that estimate is unreliable.

**Sensor uncertainty:** Obstacle and landmark position estimates are restricted by the type of sensors used, and by the finesse of the algorithms used for extraction, location, and identification. For example, in the case of sonar sensors, false reflections are typical, and a large amount of obstacle position readings might be

inaccurate. If on the other hand, computer vision methods are utilized, illumination and pose invariance might be difficult to achieve; and even if these are overcome, the effective and accurate extraction of landmarks might be prohibitive for real-time systems.

**Data association:** The matching of observations with previously learned elements in the environment is a daunting task. More than often, false data association hypothesis are generated in the presence of occlusions. False data association is critical in the sense that it induces large deviations to the robot localization estimate.

**Dynamic environments:** In a restricted number of applications, the environment remains static. However, for the general case, obstacle locations usually change over time, pushing for stochastic map models that continuously update the environment map to reflect these changes.

The problems of landmark extraction, location, and identification can be associated to the well known *object recognition* problem in the computer vision and pattern recognition research communities. We adapt its definition from our own survey [7] to the scope of this study. Given one (or a sequence of) scene image of the environment for a particular robot pose (motion), can we identify and localize those landmarks that are sufficiently visible to the sensory system? To answer this question, we study in **Chapter 2** several landmark representations for various types of sensors. The discussion includes the mathematical foundation necessary to extract points, lines, and regions from images, and lines from laser range data. We conclude with some commentary on the invariant characterization of landmarks for mobile robotics.

This part of the work was carried out during my stay at the Robot Vision Laboratory at Purdue University, and as a consequence of this study, we have contributed the aforementioned survey to the Wiley Encyclopedia of Electrical and Electronics Engineering in its year 2000 edition.

It is often the case, that the information extracted from just one sensor, or only one low level processing technique, does not suffice in the invariant characterization of landmarks and objects; pushing for the combination of information from multiple sources. In **Chapter 3** we introduce a new algorithm that fuses complementary information from two low level vision modules, namely color image segmentation and depth from stereo, into coherent object models that can be tracked and learned in a mobile robotics context. Special attention has been taken so that the time complexity of the algorithm does not exceed that of the individual acquisition modules.

Results from this part of our research were presented at ICPR 2000 in Barcelona [8], and in a collective work on the use of graphs as topological data structures for computer

vision, which appeared in the March 2002 issue of the journal *Pattern Recognition* [172].

We have seen that the search for robust landmark representations with the combination of information from multiple sensors is crucial. However, the comparison of observed landmarks between successive sensory inputs is not an easy task, even with such rich landmark models. Illumination conditions and occlusions are the most prominent artifacts that hinder data association in computer vision. Nevertheless, by using photogrammetric and geometric constraints we can restrict the search for landmark matches in successive images. Furthermore, by locking our interest in one or a set of landmarks in the scene, we could track those landmarks along successive frames, reducing considerably the data association problem. In **Chapter 4** we present a small discussion on these issues, which could give rise to a dissertation topic on its own. In our discussion we concentrate only on those tools from the geometry of multiple views that are relevant to the computation of initial landmark location estimates for motion recovery. A desirable characteristic when odometry is not available or highly unreliable. We present results on projective and Euclidean motion recovery from the tracking of simple landmarks in outdoor environments.

Our study of multiple view geometry is documented in two technical reports. One on stereoscopic vision [144], and one on pattern based camera calibration [5].

Once landmarks are accurately extracted and identified from sensory inputs, the second part of the problem is to use these observations for the localization of the robot, as well as the refinement of the landmark location estimates. We consider robot motion and sensor observations as stochastic processes, and treat the problem from an estimation theoretic point of view. In the robotics and artificial intelligence communities this is known as *simultaneous localization and map building* or as *concurrent map building and localization*, and consists precisely on the incremental refinement of the robot and landmark location estimates from observations and dead reckoning.

The technique requires that the initial location of the robot be known, and works as long as observations can be matched sequentially. Furthermore, once a map is built, we can relocate the robot, turn on their sensors again, and test for multiple localization hypotheses in the quest for global localization.

Both the local and global estimation theoretic approaches to localization deal with noise by using probabilistic methods. The main drawback we encounter is that these techniques have been devised for static environments, and that they lack robustness in more realistic situations; for example, in situations in which illumination conditions change, or in which occlusions are possible. In **Chapter 5** we propose an extension to the estimation theoretic approach to robot localization and map building taking into account the dynamics of the environment. This is achieved by incorporating novel



robust tests for temporal data association.

Spatial compatibility tests are crucial for the solution of data association in localization and map building [150], but they can still be insufficient in situations with moderate scene dynamics. Consider for example the case when a landmark is occluded for a short period of time. A spatial compatibility test would not provide any information on the history of observations of such landmark, and might still be trying to wrongly associate it with a neighboring observed feature. If the localization algorithm succeeds in incorrectly associating the occluded feature, it most probably will induce localization error to the rest of the map, and depending on the algorithm used, this error might propagate recursively, ultimately breaking down the entire localization technique. For this reason, we deem necessary not only spatial, but temporal landmark compatibility tests as well.

To aid in those situations in which landmark observations might not be consistent in time, we propose a new set of temporal landmark quality functions, and show how by incorporating these functions in the algorithm, the overall estimation-theoretic approach to map building and localization is improved. The basic idea consists on using the history of data association mismatches for the computation of the likelihood of future data association, together with the spatial compatibility tests already available.

Special attention is paid in that the removal of landmarks that do not pass the temporal landmark quality tests, does not violate the basic convergence properties of the localization and map building algorithms already described in the literature; namely, asymptotic convergence and full correlation [64, 151].

Once the theoretical foundation for the revised model to map building and localization is set forth, we delve into explicit formulation of motion and measurement models for two mobile robot platforms. First, we study a simple linear one-dimensional mobile robot, the *monobot*. Then, we extend our analysis to the more realistic case of a nonlinear bidimensional planar mobile platform. The latter robot configuration is used to compare the performance of our extended map building algorithm with the original model discussed in the literature; both in the presence of various noise levels, and ultimately, in cases with limited field of view and extreme data missassociation.

Our initial temporal landmark quality studies were first reported in a national meeting on May 2001 [10]. Later on, we came up with a simpler temporal quality model and presented it in Baden-Baden on August 2001 [9]. On the other hand, our study of estimation theory gave birth to a technical report on Kalman filtering [6], and to a pair of contributions on its use in a related application, the tracking of human faces by our mobile robot [145, 146].

The first article covering in detail the theoretic foundation of our new map building algorithm was presented in ICPR 2002 [12]; and a more complete version has been

accepted for presentation at ICRA 2003 [13].

The use of estimation theory gives a clean and concise solution to the simultaneous localization and map building problem. Or that has been the general belief so far. However, some researchers have observed scaling difficulties in the actual implementations of current map building algorithms. As the number of iterations grows to the thousands, Kalman filter based algorithms tend to diverge, in spite of the convergence theorems proved in the literature. The reason for such divergence has been attributed to nonlinearities of the models chosen. We believe that partial observability and partial controllability have also something to do with such behavior.

**Chapter 6** is devoted to an in depth analysis of the fully correlated model to localization and map building from a control systems theory point of view. Considering the fact that the Kalman filter is nothing else but an optimal observer, we analyze the implications of having a state vector that is being revised by fully correlated noise measurements. This is, we end up revealing theoretically and with experiments the strong limitations of using a noise driven estimation theoretic approach to map building and localization in relation to the total number of landmarks used, regardless of the vehicle model chosen.

One revealing conclusion drawn in this chapter is that the number of landmarks plays an important role in the error levels achieved during the localization of a mobile robot purely from noise driven landmark observations. There are directions along the map state vector which cannot be observed, and must be recovered from the correlated observation of motion and landmarks. However, as the number of landmarks grows, these estimates improve; but up to a point in which adding more landmarks to the map model is of no benefit and does no more than introducing noise into the system.

Partial observability hinders full reconstructibility of the state space, in that it makes the final map estimate dependant on the initial observations, and does not guarantee convergence to a positive definite covariance matrix. Partial controllability on the other hand, makes the filter believe after a number of iterations, that it has accurate estimates of the landmark states, with their corresponding Kalman gains converging to zero. That is, after a few steps, innovations are useless. In Chapter 6 we show how to palliate the effects of full correlation and partial controllability in CML.

Furthermore, given that the Kalman filter is an optimal observer for the reconstruction of fully correlated states; it seems pertinent to build an optimal regulator in order to keep the robot as close as possible to a desired motion path when building a map. We show also in Chapter 6 how the duality between observability and controllability can be exploited in designing such an optimal regulator.

Part of our findings about using fully correlated maps will be submitted to IROS

2003. The rest, and their relation with the structure-from-motion problem in computer vision, is waiting for time in our agenda for a journal submission.

Any map building and localization algorithm for mobile robotics that is to work in real time must be able to relate observations and model matches in an expeditious way. Some of the landmark compatibility tests are computationally expensive, and their application has to be carefully designed. In Chapter 5 we touch upon the time complexity issues of the various landmark compatibility tests used, and in **Chapter 7** we concentrate on desirable properties of our chosen map data structure, which in turn allows for expeditious landmark search, insertion and deletion. Furthermore, we propose a series of tasks that must be handled when dealing with landmark data association. From model compatibility tests, to search space reductions and hypothesis formation, to the actual association of observations and models.

The algorithmic details of our map update algorithm appear in an article published in the August 2002 issue of the International Journal on Pattern Recognition and Artificial Intelligence [11]. A schematic overview of the proposed localization and map building architecture is shown in Figure 1.1. Finally, the mobile robot platform Marco, developed as part of this thesis, is shown in Figure 1.2, and was described in great detail in two research reports [173, 174].

Finally, **Chapter 8** contains a collection of images that visually exemplify some of the tools that were developed as part of this thesis.

The work presented in this thesis spans several areas of engineering and computer science, from new computer vision algorithms, to novel ideas in mobile robot localization and map building. The key contributions are the proposal of a new technique to fuse visual data; the formulation of new algorithms to concurrent localization and map building that take into account temporal landmark quality; new theoretical results on the degree of reconstruction possible when building maps from fully correlated observations; and the necessary techniques to palliate partial observability, partial controllability, and the nonlinear effects in CML.

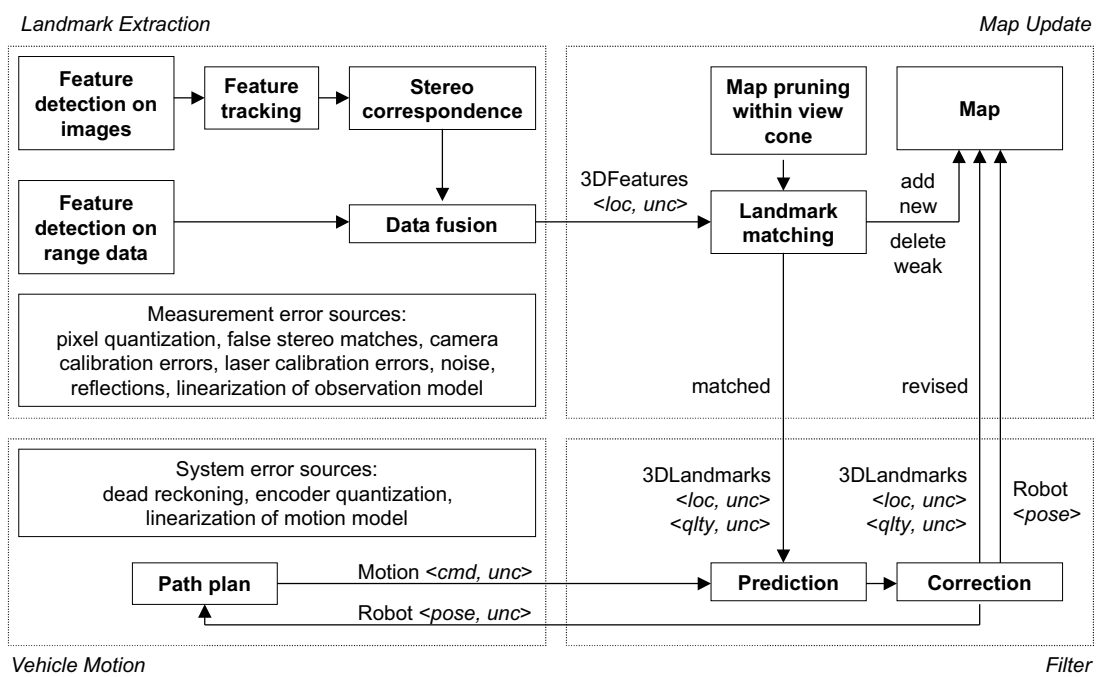


Figure 1.1: System architecture.



Figure 1.2: Marco mobile robot.



## Chapter 2

# Landmark Extraction for Mobile Robotics

In this Chapter, we provide a qualitative discussion of the various tools we have chosen for the extraction of features from vision and laser sensors. These tools constitute the bulk of the input that will be used in our robot localization algorithms discussed in later chapters. More than an extensive review of fundamental computer vision techniques, we only focus on a somewhat detailed explanation of a limited number of algorithms that we have chosen for the extraction of landmarks with the task of mobile robot localization in mind. The advantages and drawbacks of the selected algorithms will be underpinned here; and the details of all of the chosen techniques are presented in Appendix A.

### 2.1 Low-level processing of 2d intensity images

There exist a mirage of methods for edge detection on images, since it is a primary element of computer vision technology; from the very well known Sobel, Prewitt, or Roberts masks that use first-order derivative operations over images, or the *LOG* detector that uses second-order derivatives, to a family of optimally designed methods for edge extraction based on the detection of extremes in the output of the convolution of the image with an impulse response, namely the Canny, Deriche, and Spaceck detectors [7, 35, 36, 74, 137]. Another widely used operator, the Heitger edge detector, uses energy oriented maps, yielding good continuity of features near junctions and precise estimation of gradient orientation [94]. A comparison on the performance of some of these operators is by Heath *et al.* [93]. We have chosen the Canny edge detector as our workhorse for the extraction of edges from images. It is from our own perspective

the most reliable edge extraction algorithm to date. Some of its advantages over other techniques include its optimality by construction, and being isotropic.

All of the above discussed methods produce edges that are not connected, and before attempting higher-level scene interpretation, we need to repair broken edges in the case that noise or other artifacts caused such breakages. Edge repair can sometimes be carried out by expanding and shrinking the edges through morphological operations [24, 165]. Another tool for grouping together the edge elements that form straight lines in an image is based on the concept of the Hough transformation [100] that basically consists on mapping the image space to a space parameterized by the position and orientation of the edges in the image, such that each straight line in the input image maps to a point in the parameterized space. Moreover, an approach to boundary localization called *snakes*, first introduced by Kass *et al.* [115], includes the use of active contours. The classical snakes approach is based on deforming an initial contour curve towards the boundary of the object to be detected, by minimizing an energy function designed so that a local minimum is obtained at the boundary of the object. The use of active contours has been extended, with the use of deformable surfaces, to the processing of three-dimensional data, such as that obtained from stereovision or an array of sonar sensors [42]. The main disadvantage on the use of most of the techniques for edge repair and on the use of active contours in particular is that none of these methods is tailored for its application on real-time systems. The iterative nature of the minimization techniques used on the snakes algorithm make it very time consuming, and unsuitable for our purposes.

Another popular salient feature extracted from images are corners or vertices. Fifteen years ago, solutions to the corner location problem consisted on searching for points with high curvature along edges. These techniques, which relied on the results of preceding edge extraction modules, are no longer popular due to their high computational cost [60, 98, 199]. More recent methods for corner detection are based on the direct computation of the gradients and curvature of an image. One of the first functions used to characterize corner response that did not require a previous edge extraction step is due to Beaudet [28], and consisted basically on the computation of the determinant of the Hessian for each image point. Variations on this approach include that of Kitchen and Rosenfeld [116], Dreschler and Nagel [67], and Harris and Stephen [90]. Another corner detection method based on image surface curvature is by Wang and Brady [208]. Lastly, the extraction of point features from images in the context of tracking from video sequences has gained recent popularity. Shi and Tomasi [181] proposed an optimal salient point feature function derived directly from the model of affine motion.

All of these techniques for corner extraction require the computation of second-order derivatives of the image, which makes them not only very noise sensitive, but also inaccurate for precise localization. To resolve this issue, Deriche and some of his



students have proposed algorithms for vertex localization and refinement, that although initially based on the same principle of using second-order derivatives, they make use of other techniques such as multiresolution, non-maximum suppression, and energy minimization to come up with refined corner localization estimates [32, 59, 60, 84, 142].

When dealing with structured indoor environments, our corner detection module relies on the Beaudet operator with further refinement by energy minimization. We have used, on the other hand, the Shi and Tomasi operator for the extraction of landmarks in outdoor scenes.

The use of more than one camera for the extraction of points of interest from a scene gives several advantages. First, it reduces significantly the consideration of noisy points as landmarks, as it is most likely that the same noise points will not appear in more than one image at the same location due to their stochastic nature. Secondly, when feasible, the three-dimensional coordinates of an image point can be computed by triangulation. However, in order to speed up the matching of stereo correspondences, epipolar rectification of the images is required. The following references provide an in depth analysis of image rectification [80, 131, 159, 160].

There exist multiple stereo matching methods that solve the correspondence problem producing what is known as depth maps. From the commonly used area-based correlation techniques tailored for dense mapping [2, 74, 124], to feature-based methods that compromise density for accuracy during reconstruction [97, 148, 210]. From our point of view, the most impressive demonstration of dense depth map computation comes from the group of Kanade at the Robotics Institute, Carnegie Mellon Univ. Starting with their seminal work [152], Prof. Kanade and his students have now been able to produce dense depth maps in real-time thanks to a video-rate stereo machine developed at the expense of large computational power [113, 149, 204]. There exist many more modest systems that provide acceptable results, including the ones by Aranda [15] or Konolige [117], and our own laboratory version [144].

Once a reliable depth map is obtained, further processing is sometimes required to extract the desired landmarks. A sample of contributions on line labeling and surface segmentation from stereo include the ones by Chen and Lin [46], Huynh and Owens [101], or Venkateswar and Chellapa [205]. Broad reviews on stereo matching techniques can be found on Aranda's PhD thesis [15], or the book by Faugeras [74]. A comparative study on the use of stereovision and laser sensors for mobile robot localization is due to Pérez *et al.* [163].

Geometric reconstruction from stereo matching is only possible provided the cameras have been calibrated a priori. We have implemented a calibration algorithm and it is presented in Appendix B for completeness.

## 2.2 Low-level processing of laser data

We can also use computer vision techniques to process three-dimensional laser points for the extraction of landmarks, in the same manner as with the depth maps computed from stereovision or structured light. The tasks of edge extraction, labeling and segmentation are usually performed via morphological operations. In our survey on object recognition [7], we provide a detailed description on the methods to compute surface normals from sets of points in space via planar patch fitting; on recursive techniques for 3d surface segmentation; and on the characterization of surface segments according to their curvature properties thanks to the *Extended Gaussian Image* [99].

As an example of their applicability, one such system that extracts landmarks purely from range data for grid-based map building relying on morphological processing techniques is presented in an article by Başkent and Barshan [22]. Another example due to Utete *et al.* [202], uses voting techniques for landmark classification from range data.

Due to its simplicity and low computational time complexity, our approach to the detection of straight landmarks from laser data is based primarily in the line simplification algorithm described in [65, 96]. It is discussed in detail in Section A.2; need to say only that our implementation of the algorithm is recursive, and that it has worst case time complexity of  $O(n \log n)$  with respect to the number of points in the laser scan.

## 2.3 Landmark characterization

Once a set of features such as corners, lines or segments has been extracted, we need to store their appearance properties in a feature vector. The kind of feature properties to save depend in the type of feature extracted. In the case of simple geometrical entities such as the lines extracted from range data for example, the end points coordinates would suffice. For the case of image segmented regions, one would like to store other geometric attributes such as its area, perimeter, elongation, moments of inertia, or spatial frequency descriptions such as the discrete cosine transform, Fourier descriptors, wavelets, or eigenimages. Since motion is our main concern, the attributes selected must be as invariant to changes in the feature pose as possible.

We will describe next the specific appearance properties chosen to characterize landmarks for the localization and map building algorithms developed as part of this thesis. For the characterization of image points, the feature vector is simply composed of the image pixel coordinate, as well as the normalized image intensity values for every point in a region of interest around such pixel. The same region that was used during the convolution operations for the extraction of such feature. Typically,  $7 \times 7$ ,  $9 \times 9$  or

$11 \times 11$  window sizes.

Along with the normalized intensity values (three dimensional intensity vectors in the case of color images), we stored also the result of the corner detection operator. Namely, a scalar value with the result of the Beaudet, Harris or Shi-Tomasi corner measure.

Now, if these same points were correlated for the computation of stereo, a few more parameters were added to the feature vector. These are the disparity value obtained, and ultimately the camera centered 3d coordinates of the point.

Lines extracted from monocular and stereo images were characterized by their ending points. So, their representation was seemingly extended from the aforementioned description for points. In the case of lines extracted from laser range data in the other hand, much simpler feature properties were saved. Since depth information is readily available from the laser range scan, the coordinates of points in the scan plane were stored. As an alternative, we also stored the parametric representation of such lines, given by the slope and the  $y$  intercept with respect to the laser coordinate frame. Given the limited properties of a laser scan, these were the only geometric properties they could bear for the description of the environment. Nevertheless, these descriptions constitute sufficient data for the construction of maps and the localization of a mobile vehicle in indoor areas.

Although we have not specifically used image segments for the building of maps, these were used for the extraction and description of object models. The related application of image segmentation along with depth from stereo for object characterization is discussed in detail in Chapter 3. Similarly, Chapter 7 describes the techniques used to infer higher level features such as walls, from the lines extracted from laser range data.

Landmark attributes extracted from sensor data should, under ideal circumstances, be invariant to the robot viewpoint, illumination conditions, and background clutter. If three-dimensional sensors such as stereovision or sonar sensors are utilized, the very nature of the data generated guarantees viewpoint invariance, provided a sufficient number of data points are collected for a given landmark. With such sensors, the shape of a landmark surface will remain the same, as long as the surface remains unobstructed and discounting the effect of highly oblique angles between the surface normal and the line of sight of the sensor. Conversely, lack of invariance poses a bigger problem when using 2d sensors, such as monocular images, and for these cases, it is necessary to use projective invariants and quasi-invariants, such as the angle between two segments and the length ratio of these segments [87].

A description of some invariants for the recognition of planar landmarks from two-dimensional images appears in our survey [7]. Various contributions from our research

group to the study of geometric and projective invariants for three-dimensional object recognition from images are due to Sanfeliu *et al.* [73, 170, 171, 175]. More recent contributions from other authors include the ones by Berthilsson and Heyden [30], Carlsson and Weinshall [38, 39, 40], Gros *et al.* [87], or Rothwell *et al.* [166].

## 2.4 Conclusions

In this Chapter we have presented a qualitative discussion on the various tools and techniques used in the characterization of landmarks from images and laser range data. We deem important to remark at this point that it was not our intention to develop new image or signal processing algorithms, but only to study and implement the ones we have found suitable for the invariant extraction of landmarks in the context of mobile robotics. The details of the algorithms chosen are presented in Appendix A.

## Chapter 3

# Data Fusion for Object and Landmark Characterization

A step forward towards achieving intelligent mobile robotics tasks is the identification of objects. Sometimes, the information extracted from just one sensor does not suffice in the invariant characterization of landmarks and objects; and it may prove useful to combine data from multiple sources in order to be able to produce robust generalization of the objects that lay in the mobile robot workspace. To this aim, we introduce what we believe is a robust algorithm that fuses the type of information extracted from camera images as discussed in Chapter 2, into coherent object models that can be tracked and learned in a mobile robotics context. This new algorithm has two major advantages over other contributions. The time complexity is linear with respect to the number of pixels in the image, and it does not exceed that of the individual acquisition modules.

Different data acquisition methods are tailored at extracting particular characteristics from a scene and by combining their results a more robust description of the environment can be created. We propose here a new method to fuse perceptual groupings extracted from color-based segmentation and depth information from stereo using supervised classification. The merging of data from these two acquisition modules allows for a spatially coherent blend of smooth regions and detail in an image. Depth cues are used to limit the area of interest in the scene and to improve perceptual grouping solving subsegmentation and oversegmentation of the original images. The complexity of the algorithm does not exceed that of the individual acquisition modules. The resulting scene description might then be feed to an object recognition module for scene interpretation.

Our contribution is mostly related to the works by Jain *et al.* [147, 156], but has a critical difference. In Pankanti and Jain's contribution [156], segmentation results of

grey-scale images are used, namely the segmented boundaries, to limit the enforcement of smoothing constraints in the stereo module, and to prevent the propagation of depth values across uniform regions. We believe that the flow of information should also be considered in the opposite direction. This is, depth cues should be considered as an aid to perceptual grouping as well, rather than only using perceptual boundaries to limit the adjustment of depth estimates.

The reasoning behind this assumption is based on the fact that segmentation of intensity images is prone to illumination conditions and surface properties, and the perceptual groupings thus produced by most segmentation algorithm might have considerable error. On the other hand, depth estimation at scene discontinuities is less sensitive to these factors and can be used more robustly to further group segments into higher level perceptual entities that could match an object model, or to discriminate from noisy or undesirable segmented regions.

### 3.1 Integration of perceptual grouping and depth using supervised classification

Consider the set of regions  $\mathbf{R} = \{R_1, R_2 \dots R_N\}$  that indicate good color continuity in the original image as extracted by the color segmentation algorithm described in Section A.1.4. Our implementation of the segmentation algorithm differs from the one presented by Vergès [206] in that regions in  $\mathbf{R}$  are divided in two subsets. If  $|R_j| > t_A$  and  $p_j^2/|R_j| < t_C \Rightarrow R_j \in \Omega$ , otherwise  $R_j \in \Gamma$ .  $\mathbf{R} = \Omega \cup \Gamma$ .  $\Omega$  represents the regions with area greater than  $t_A$  and compactness smaller than  $t_C$ ,  $\Gamma$  represents the detail in the image, and  $p_j$  is the perimeter of  $R_j$ .

To solve for subsegmentation, in each region  $\Omega_j$  the set of points  $r_j = \{\mathbf{p}_i | \mathbf{x}_i \in \Omega_j\}$  is extracted from the depth map, and for each coordinate  $\mathbf{x}_i \in \Omega_j$  from the segmented image a new point  $\mathbf{p}_i$  is added to  $r_j$ , where the corresponding  $d_i$  is obtained from the average disparity from the points in the  $3 \times 3$  window around pixel  $\mathbf{x}_i$  with entry in  $r_j$ . The process is repeated iteratively until all pixels in  $\Omega_j$  have a corresponding entry in  $r_j$ . This is, until all pixels in  $\Omega_j$  have been assigned a depth estimate.

Smooth surface segments  $S_l^j$  are recursively generated by starting at any point  $\mathbf{p}_i \in r_j$  and growing outwards while meeting the following two criteria for the neighboring points  $\mathbf{p}_i$  and  $\mathbf{p}_k$ :

$$\|\mathbf{p}_i - \mathbf{p}_k\| > t_J \quad \text{Jump Edge Criterion} \quad (3.1)$$

$$\frac{\cos^{-1}(\mathbf{n}_i^T \mathbf{n}_k)}{\|\mathbf{p}_i - \mathbf{p}_k\|} > t_U \quad \text{Curvature Criterion} \quad (3.2)$$

The normals  $\mathbf{n}_i$  are computed minimizing the error of fitting a local planar patch in the vicinity of  $\mathbf{p}_i$  [45]. The region  $R_j \in \Omega$  is then replaced in  $\mathbf{R}$  by the segments in  $\mathbf{S}^j$ . The user thresholds  $t_J$  and  $t_U$  are application dependant, and must otherwise be trained by example. A new division of  $\mathbf{R}$  into  $\Omega$  and  $\Gamma$  is necessary.

For the case of oversegmentation we first generate a set of initial classes

$$\omega_l = \{(\mathbf{x}_i, \mathbf{d}_i) | (\mathbf{x}_i, \mathbf{d}_i) \in \Omega_l\} \quad (3.3)$$

and the sets of points to be classified

$$\gamma_m = \{(\mathbf{x}_i, \mathbf{d}_i) | (\mathbf{x}_i, \mathbf{d}_i) \in \Gamma_m\} \quad (3.4)$$

The task at hand is to associate each  $\Gamma_m$  to its parent region  $\Omega_l$  based on their spatial proximity. This classification constitutes the merging of smooth and detail into spatially coherent entities. Consider one detail region  $\Gamma_m$ . We must compute the distance from the points  $(\mathbf{x}_i, \mathbf{d}_i) \in \gamma_m$  to the classes in  $\omega$ . It is clear that the distribution of sample points in the classes in  $\omega$  does not follow any typical probability distribution, but in those areas where the image is rich in detail, the samples resemble a uniform distribution if projected on the image plane; the minimum distance between points being the pixel width and length. This observation suggests the use of a parametric distance measure for classification. The normalized distance from point  $(\mathbf{x}_i, \mathbf{d}_i) \in \gamma_m$  to class  $\omega_l$  is

$$d_{il} = \text{tr} \left( \Sigma_l^{-1} ((\mathbf{x}_i, \mathbf{d}_i) - (\bar{\mathbf{x}}_i, \bar{\mathbf{d}}_i)) ((\mathbf{x}_i, \mathbf{d}_i) - (\bar{\mathbf{x}}_i, \bar{\mathbf{d}}_i))^\top \right) \quad (3.5)$$

where  $(\bar{\mathbf{x}}_i, \bar{\mathbf{d}}_i)$  is the mean vector of class  $\omega_l$ , and  $\Sigma_l$  the covariance matrix. The votes  $V_i = \min_l(d_{il})$  are accumulated for each  $(\mathbf{x}_i, \mathbf{d}_i) \in \gamma_m$ , and the region in  $\Omega$  associated to the class  $\omega_l$  with most votes is considered the parent region for  $\Gamma_m$ .

The result is a new set of regions  $\Omega'$  where each element  $\Omega'_j = \{\Omega_j \cup \Gamma_m \dots \Gamma_n\}$ , represents a region in the scene where color continuity and depth continuity are merged to constitute spatially coherent entities. A set of characteristics can be measured on each of these regions, such as position, normal orientation, curvature, level of detail, area, compactness, etc. The immediate step to follow from these results will be an attempt to learn and recognize these groups of segments as objects.

The time complexity of the depth from stereo module is  $O(dk^2n)$ , where  $k$  is the width of the kernel window,  $d$  is the maximum expected disparity, and  $n$  is the number of pixels in the image. If the edges in  $G$  are sorted in linear time, the segmentation module is bounded by  $O(n)$ .

The time required to compute Equation 3.5 is bounded by  $O(m^2/a)$ , where  $m$  is the number of points in the disparity map associated to regions in  $\Omega$ , and  $a = |\Omega|$ . Given

that the disparity map is dense in the perimeter of  $\Omega_j$  and negligible inside smooth regions

$$m \approx \sum_a p_j k / 2 \quad (3.6)$$

and from the compactness constraint

$$p_j^2 < t_C |R_j| \quad (3.7)$$

The overall cost of the oversegmentation part of the algorithm is bounded by

$$\begin{aligned} O(m^2/a) &\approx O(a\bar{p}_j^2 k^2) \\ &< O(ak^2 t_C |\Omega_j|) \\ &< O(k^2 t_C n) \end{aligned} \quad (3.8)$$

The time complexity of our algorithm is linear with respect to the number of pixels in the image, and is asymptotically comparable to that of the individual data acquisition modules.

Figure 3.1 exemplifies the advantages and drawbacks of our algorithm. The original left color image from the stereo pair is shown in Figure 3.1(a). Both the left and right images were rectified to meet the parallel epipolar constraint. The results of the color segmentation algorithm applied to the rectified left image are shown in Figure 3.1(b). Note that labels assigned to each segment do not resemble the intensity values in the original image, as would be expected from a color based segmentation algorithm. This was done only to ease visual identification of large segments from small or highly compact ones. In this figure, the textured regions represent the small segments that need to be classified as belonging to the nearby objects.

Figure 3.1(c) shows the disparity map obtained from the stereo module. It is most clear from this image how depth information is dense at highly detailed areas in the scene, whereas smooth regions are poorly represented. Although the left-to-right right-to-left constraint could have been relaxed from equality to similarity when creating the depth map, this was not implemented; letting the segmentation module overcome the weaknesses of any typical depth-from-stereo module.

An initial pruning of the segmented image is done based purely on the mean disparity value of each segment. We eliminated from our three-dimensional region of interest those segments that fall too close or too far from the camera by computing their mean disparity. Also those segments with very low points-in-depth-map to segmented-area ratio were discarded, as they do not contain enough disparity information to accurately estimate their depth, and are not suitable for later attempts at object characterization.



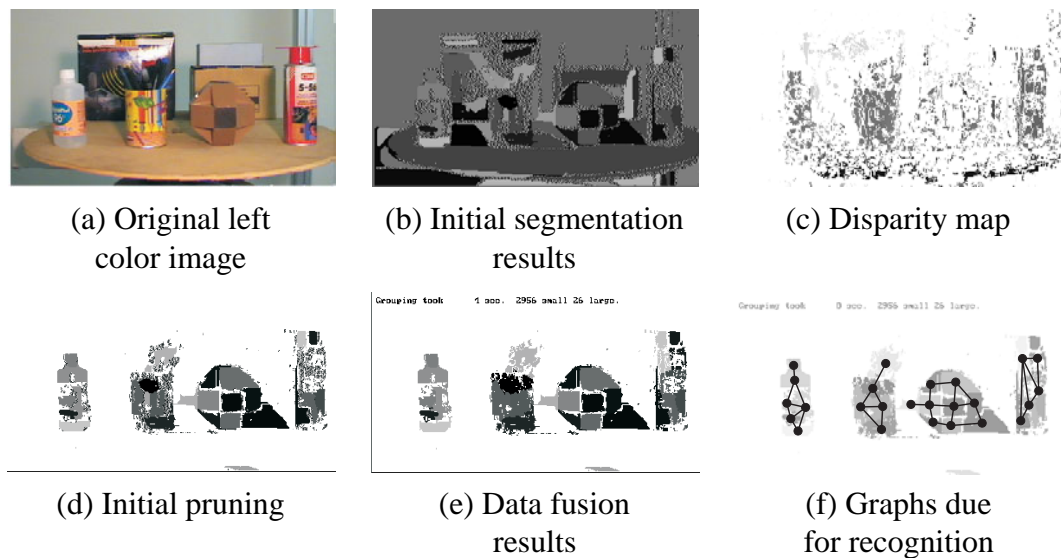


Figure 3.1: Data fusion steps.

Figure 3.1(d) shows how the boxes behind the four objects of interest are virtually eliminated, as well as the table and the vertical bar in the back.

Results from the data fusion algorithm are shown in Figure 3.1(e). The image shows our four objects of interest easily identifiable. These segment groups and their attributes can be used to characterize the objects they represent. Figure 3.1(f) shows a set of graphs representing the hypothesized objects that are due for recognition.

## 3.2 Bibliographical notes

The fusion of 3D information from different acquisition methods allows for more robust scene descriptions. The shortcomings of individual low level processing modules can be overcome in an integrated environment. However, the inherent variability of the acquisition methods and the data formats and noise levels they produce make sensor fusion a challenging task. In the field of computer vision several attempts have been made at coupling data from different sensors. Integration models proposed in the literature vary in the number and type of sensor inputs, in the level at which fusion takes place, and in the rules used for data fusion.

Citing some relevant contributions, in Pankanti and Jain's work [156] an integration

framework that encompasses four vision modules is presented. The merging of data from stereo, shape from shading, perceptual organization, and line labeling is used to estimate accurate depth maps of a scene. Other contributions consider the fusion of 2-D and 3-D data in the form of intensity images and stereo or range data [19, 147], or only 3-D data acquired from stereo and range data [95, 123]. Methods that merge data from stereo and shape from shading include those by Blake *et al.* [31], Bulthoff and Mallot [34], and Cryer *et al.* [52].

Most of the methods that fuse depth information are designed to work at a point or pixel level, whereas those methods that include perceptual grouping in 2-D images are mostly directed towards fusing higher level primitives [156], or more specifically, at labeling primitives such as contours or edges based on their low-level properties, i.e., depth estimates. In the works by Austin and Wallace [19], and Lacroix *et al.* [123], emphasis is made in that an integration architecture should be made at both the pixel and higher perceptual grouping levels.

The methods used for fusing different types of data vary extensively, from ad-hoc implementations that use empirical thresholds to select which sensor contributes to scene formation, to the more elegant techniques of Extended Kalman Filtering to update depth estimates [95, 123], or Bayesian networks to integrate top-down and bottom-up visual processes [177]. Some exploit the fact that one of the data acquisition modules can be viewed as a multiresolution system and have embedded their data fusion techniques in between each level of resolution [156].

The typical data structure used for the representation of objects in the context of object recognition is the attribute graph [7]. In our particular domain, the nodes of such a graph can represent each of the labeled segments in an object, and the edges represent Euclidean distance or any other relationship between these segments. However, some difficulties in the applicability of matching techniques over graphs have reduced their practical use. The search for a specific node or edge in a graph, or for the relationship between two nodes (i.e., shortest path) requires polynomial time. However, the search for a match of a scene graph with a model graph, known as the isomorphism problem, is intractable. This is, its computation has an exponential time complexity. Nevertheless, if we constrain the search for a match with the appropriate heuristic, the time complexity of the problem can be reduced at the expense of a sub-optimal solution [48, 217]. A detailed description on some of these heuristics for graph matching can be found in our survey [7], including discrete relaxation, bipartite matching, or with the use of hash tables.

In our research group, a new type of probabilistic graphs called *function described graphs* has been developed [3, 176, 179, 180]. FDG's augment the capability of first-order probabilistic modeling of structural and attributed information from random graphs by incorporating qualitative knowledge of the second-order probabilities of the graph

elements in the form of binary relations on arcs and nodes. We propose here as an open research topic the possibility of using FDG's for object representation, provided the correct heuristics for fast graph matching are developed for this type of graphs.

### **3.3 Conclusions**

In this Chapter, we have shown how the shortcomings of two individual low-level processing modules can be overcome in an integrated environment in the characterization of objects and landmarks for mobile robotics. The inherent variability of the data formats is tackled by exploiting their individual characteristics. While color-based segmentation methods are robust in smooth regions and tend to fail in areas where detail is prominent, the opposite is true for a depth from stereo module. We have provided the necessary framework to exploit this situation by relabeling those areas where segmentation fails based on depth cues, both in sub- and oversegmentation situations. It is of major relevance that the time complexity of our algorithm is linear with respect to the number of pixels in the image, and does not exceed that of the individual acquisition modules.



## Chapter 4

# Combining Information from Sequential Views

In Chapter 2 we commented on the various types of landmarks that can be extracted from vision and laser range data, one snapshot at a time. We now turn our attention to the tracking of visual landmarks only, but along a set of consecutive images. Firstly, with the purpose of alleviating the association of observations in between views. Secondly, to compute coarse motion estimation when dead reckoning is unreliable or simply unavailable. The material presented in this Chapter represents a tiny portion of the widely known area of structure from motion within the computer vision community. Most of the techniques herein described can be found in traditional multiple view geometry books, such as the one by Hartley and Zisserman [91]. Our contribution resides in using such techniques within the map building and localization context.

In Section 4.1 we revise the area based correlation technique used for establishing image point correspondences. In Section 4.2 we compute the fundamental geometry between two views. The methods described in these two sections are only touched upon given their wide dissemination in modern computer vision literature. They serve as an introductory framework to the technique presented in Section 4.3, in which we study the recovery of camera motion from image correspondences. We use a stratified approach in which an already available technique for the computation of projective reconstruction is presented. Next, we show a simple technique with closed form formulation for the computation of Euclidean reconstruction provided at least five ground truth points in the scene. When no knowledge of the scene is available a priori, we can still recover Euclidean reconstruction up to scale ambiguity. The technique for doing so is presented in Section 4.3.3. However, it is often the case that no metric information of the scene is available. Full metric reconstruction is still achievable if we add metric information directly to our sensor. For example, by using a calibrated stereoscopic rig. We show some exemplary results of the method in Section 4.3.4.

## 4.1 Feature tracking

We consider the tracking of visual features as the search for image correspondences in consecutive frames. Image correspondences can only be evaluated at those points where sufficient detail is present in the scene. That is precisely the reason why we consider corners and Shi and Tomasi features as the visual landmarks to track.

### 4.1.1 SSD, SAD

The matching of image features is usually evaluated by minimizing area based correlations over a window of interest as discussed already in Section A.1.5. However, since what we are estimating is precisely the camera motion for a given image sequence, we do not have access to the entire perspective projection matrices that generated such sequence. Being that the case, it is impossible to perform image rectification beforehand; consequently, one must compute image correspondences taking into account the two dimensions of the image plane.

Given a consecutive pair of image frames  $I_k$  and  $I_{k+1}$ , the match for an image feature at  $\mathbf{m}$  in  $I_k$  would be located at  $\mathbf{m} + \mathbf{d}$  in  $I_{k+1}$ . The disparity vector  $\mathbf{d} = [d_x, d_y]^T$  can be computed either with an  $L_2$  distance, the argument that minimizes the sum of squared differences (SSD) over a window  $W$  of interest

$$\mathbf{d} = \arg \min_{\mathbf{d}} \sum \sum_W (I_k(\mathbf{m}) - I_{k+1}(\mathbf{m} + \mathbf{d}))^2 \quad (4.1)$$

or an  $L_1$  distance, the argument that minimizes the sum of absolute differences (SAD)

$$\mathbf{d} = \arg \min_{\mathbf{d}} \sum \sum_W |I_k(\mathbf{m}) - I_{k+1}(\mathbf{m} + \mathbf{d})| \quad (4.2)$$

Both techniques give similar results. The latter being less computationally expensive. Extensions to these area based correlation methods include the zero mean shift or the normalized cross correlation. For a detailed description of these and other formulas for the computation of image correspondences see our technical report [144]. Figure 4.1 shows three consecutive frames of an indoor image sequence obtained by our mobile robot Marco.

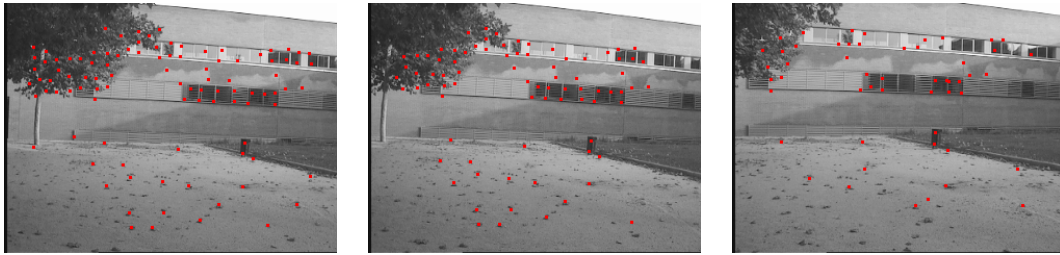


Figure 4.1: Three consecutive frames of a sequence acquired by a mobile robot in motion. The rough characteristics of the terrain make dead reckoning computation unreliable. The red dots indicate the set of tracked Shi and Tomasi features.

### 4.1.2 Illumination changes and occlusions

In computing image correspondences with Equations 4.1-4.2 we are relying strongly on the assumption of small inter-frame motion. However, this simplified motion model is by no means complete, and does not take into account the perspective transformation the patterns of image intensities suffer when a scene is observed from different viewpoints. Furthermore, the model has no way of incorporating the image intensity fluctuations caused by occlusions and illumination variations.

A simple technique to validate image correspondences even in the presence of occlusions, is to maintain a cumulative root mean squared residue of each feature window with respect to a key frame. That is, by maintaining an initial feature appearance model to which compare the tracked feature at all instances. The cumulative residue surpassing a certain threshold may indicate occlusion or sharp illumination fluctuations. The technique is further described in a technical report by Tomasi and Kanade [196].

## 4.2 Epipolar geometry

Given a set of corresponding features  $\mathbf{m}_i \leftrightarrow \mathbf{m}_i + \mathbf{d}_i$  from a sequential pair of images, it is possible to estimate the geometry between the two views given by

$$\tilde{\mathbf{x}}_i'^T \mathbf{F} \tilde{\mathbf{x}}_i = 0 \quad (4.3)$$

in which  $\tilde{\mathbf{x}}_i$  are the homogeneous coordinates of image point  $\mathbf{m}_i$  in  $I_k$ , and  $\tilde{\mathbf{x}}'_i$  are the homogeneous coordinates of its correspondence in  $I_{k+1}$ , the image point  $\mathbf{m}_i + \mathbf{d}_i$ . The matrix  $\mathbf{F}$  is known as the *fundamental* matrix and is a  $3 \times 3$  matrix of rank 2.

The above expression can be written as a vector inner product in the entries of  $\mathbf{F}$   $\mathbf{a}^\top \mathbf{f} = 0$ , where  $\mathbf{a}$  is a 9-dimensional vector made up of quadratic combinations of the elements of  $\tilde{\mathbf{x}}_i$  and  $\tilde{\mathbf{x}}'_i$ , and  $\mathbf{f}$  is a vector made up of the entries of  $\mathbf{F}$ . If more than eight point correspondences are available, we can build an augmented linear system  $\mathbf{A}\mathbf{f} = \mathbf{0}$  by combining each individual inner product into a larger over-determined system of equations. Moreover, since  $\mathbf{F}$  is defined up to scale, we add the additional constraint that  $\|\mathbf{f}\| = 1$ . When more than eight point correspondences are available, a solution, in the least squares sense, to the augmented linear system is obtained by computing the optimum  $\mathbf{f}$  that minimizes the norm  $\|\mathbf{A}\mathbf{f}\|$  subject to the constraint  $\|\mathbf{f}\| = 1$ . This is, the unit normalized singular vector that corresponds to the smallest singular value of  $\mathbf{A}$ . The details of the formulation can be consulted in the book on multiple view geometry by Hartley and Zisserman [91].

One must not forget however that image correspondences, even when tracked efficiently, are the product of an extraction technique with inherent noise sources. When the only source of error is in the accuracy of the measurement of point positions, we can assume a Gaussian noise distribution. However, when point correspondences are mismatched in between frames, such assumption no longer holds, and the mismatched points should be considered as outliers to the Gaussian error distribution.

### 4.2.1 RANSAC

Robust estimation requires the detection of such outliers in the distribution. An increasingly popular technique for the computation of robust estimators is the Random Sampling Consensus (RANSAC) algorithm [76]. The idea of the algorithm is very simple. In the paragraphs that follow we rephrase it in the context of estimating the fundamental matrix between two views.

A minimum set of point correspondences is chosen at random (eight), enough to compute an estimate of  $\mathbf{F}$ . The support for this estimate is measured as a vote count of the number of points in the entire data set that lie within a distance threshold  $\delta$  to our estimate. The distance function being for instance, the algebraic distance

$$\|\mathbf{a}^\top \mathbf{f}\|^2 \tag{4.4}$$

or the symmetric residual error

$$\|\tilde{\mathbf{x}}'_i - \mathbf{F}\tilde{\mathbf{x}}_i\|^2 + \|\tilde{\mathbf{x}}_i - \mathbf{F}^\top \tilde{\mathbf{x}}'_i\|^2 \tag{4.5}$$



---



---

```

Ransac( $X, X'$ )

  repeat  $N$  times
     $S = \text{SelectRandomSampleSet}(X, X')$ 
     $\mathbf{F} = \text{ComputeFundamentalMatrix}(S)$ 
    repeat for entire data set ( $X, X'$ )
      if  $\text{Distance}(\mathbf{x}_i, \mathbf{x}'_i, \mathbf{F}) < \delta$ 
         $\text{AddInliertoConcensusSet}$ 
     $C = \text{SelectLargestConcensusSet}$ 
     $\mathbf{F} = \text{ComputeFundamentalMatrix}(C)$ 
  return  $\mathbf{F}$ 

```

---



---

Table 4.1: The random sample consensus algorithm, adapted from [91] to the robust estimation of the fundamental matrix from point correspondences.

This random selection is repeated a number of times, and the point set with most support (largest number of votes) is deemed the robust fit. The image correspondences within the threshold distance are the inliers, and constitute the consensus set. Thus if a point correspondence is an outlier, the estimated fundamental matrix will not gain much support (low vote count from the rest of the point correspondences). The computation of the final fundamental matrix is then estimated with the entire consensus set.

As for the value of the error threshold  $\delta$ , it can be restricted to say  $6\sigma^2$ , so that there is a 95% probability that a point classified as inlier is so. The standard deviation  $\sigma$  is computed as the average of either Equation 4.4 or 4.5 for the entire data set. This scheme for the selection of the threshold  $\delta$  is discussed in further detail in Section 5.4, for the analogous problem of computing a threshold for the landmark compatibility tests.

The number of samples  $N$  to try is chosen sufficiently high to ensure with a 99% probability that at least one of the random sample sets of points is free from outliers, with an overly pessimistic proportion of outliers of 50%. A value of  $N = 100$  would suffice (see [91]). Table 4.1 summarizes the steps of the algorithm.

## 4.3 Motion recovery

### 4.3.1 Projective reconstruction

With the assumption that  $I_k$  was acquired at the world origin, and that the camera motion incurred before the acquisition of  $I_{k+1}$  is simply  $[\mathbf{R}|\mathbf{t}]$ ; we can compute a canonical set of perspective projection matrices for our pair of images simply with

$$\tilde{\mathbf{P}} = [\mathbf{I}|\mathbf{0}] \quad \tilde{\mathbf{P}}' = [[\mathbf{e}']_{\times}\mathbf{F}|\mathbf{e}'] \quad (4.6)$$

The term  $\mathbf{e}'$  is the left epipole of  $\mathbf{F}$  and is given by the singular vector associated to the null singular value of  $\mathbf{F}^{\top}$ . Moreover,  $[\mathbf{e}']_{\times}$  is a skew symmetric matrix indicating the vector cross product of  $\mathbf{e}'$  (see Equation B.36).

Projective reconstruction of image correspondences  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{x}}'$  can thus be computed with

$$\tilde{\mathbf{y}}_p = \frac{\tilde{\mathbf{x}}(\mathbf{e}'^{\top}\mathbf{e}'\tilde{\mathbf{x}}'^{\top}\tilde{\mathbf{x}}' - (\tilde{\mathbf{x}}'^{\top}\mathbf{e}')^2)}{\tilde{\mathbf{x}}'^{\top}[\mathbf{e}']_{\times}\mathbf{F}\tilde{\mathbf{x}}\mathbf{e}'^{\top}\tilde{\mathbf{x}}'} \quad (4.7)$$

Figure 4.2 shows an example in which we perform projective reconstruction of a calibration pattern from a pair of sequential views of the pattern.

### 4.3.2 Euclidean reconstruction from knowledge of the scene

The next step in the reconstruction process is to obtain the homography  $\mathbf{H}$  that relates projective to Euclidean reconstruction. Given the two view fundamental matrix  $\mathbf{F}$ , the general forms for  $\tilde{\mathbf{P}}$  and  $\tilde{\mathbf{P}}'$  from Equations 4.6, and at least 5 projective  $\tilde{\mathbf{y}}_p$  and Euclidean  $\tilde{\mathbf{y}}_e$  homogeneous point correspondences, we propose the following simple technique for the computation of  $\mathbf{H}$ .

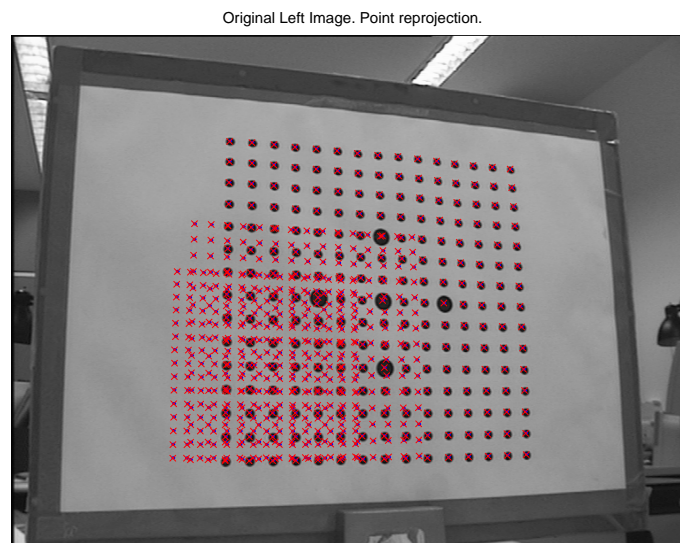
The homogeneous image coordinates of the projective point  $\tilde{\mathbf{y}}_p$  are given by

$$\tilde{\mathbf{x}}_p = \tilde{\mathbf{P}}\tilde{\mathbf{y}}_p \quad \tilde{\mathbf{x}}'_p = \tilde{\mathbf{P}}'\tilde{\mathbf{y}}_p \quad (4.8)$$

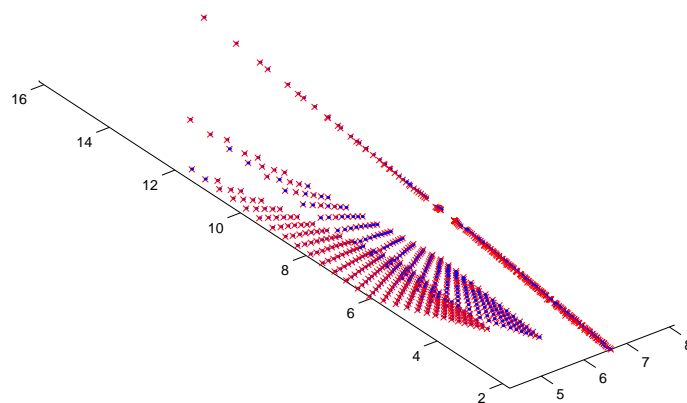
and the same image point but as computed from its Euclidean correspondence obeys the relation

$$\tilde{\mathbf{x}}_e = \tilde{\mathbf{P}}\mathbf{H}\tilde{\mathbf{y}}_e \quad \tilde{\mathbf{x}}'_e = \tilde{\mathbf{P}}'\mathbf{H}\tilde{\mathbf{y}}_e \quad (4.9)$$

Note that the homogeneous scale factor in Equations 4.8 and 4.9 is not equal in general. However, substituting the normalized image coordinates  $\tilde{\mathbf{x}}_1/\tilde{\mathbf{x}}_3, \tilde{\mathbf{x}}_2/\tilde{\mathbf{x}}_3$  from



a) Projection of point correspondences



b) Projective reconstruction

Figure 4.2: Projective reconstruction of structure from sequential views. Point correspondences in between two views of a three-dimensional calibration pattern are computed. Frame (a) shows the projection of all point correspondences on top of an image of a portion of a calibration model (made up of three planes as shown in Figure B.1.2). The camera was then moved, and a second series of snapshots of the calibration pattern were taken. Frame (b) shows the projective reconstruction, estimated only from the computed fundamental matrix.

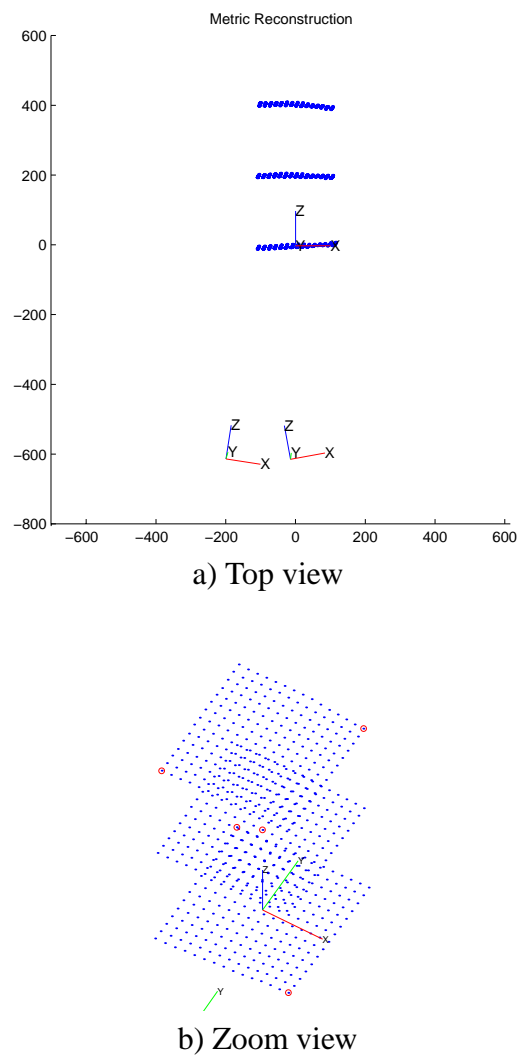


Figure 4.3: Euclidean reconstruction of structure from sequential views. Full metric reconstruction is possible to attain when some ground truth of the scene is available. The recovered camera motion is represented by the plotted coordinate frames. In this example, we consider the knowledge of the real coordinates of only five points of the calibration pattern, two on extremes of first plane, one on the middle of the second plane, and two on opposite extremes of the last plane, all shown in red on frame (b).

the Equation set 4.8 into the Equation set 4.9 we end up with four equations in 16 unknowns, the entries of  $\mathbf{H}$  ( $\mathbf{h}$  in vector form). The resulting system  $\mathbf{A}\mathbf{h} = \mathbf{0}$  has the explicit form

$$\tilde{\mathbf{x}}_{p,1}\tilde{\mathbf{x}}_{e,3} - \tilde{\mathbf{x}}_{e,1}\tilde{\mathbf{x}}_{p,3} = 0 \quad (4.10)$$

$$\tilde{\mathbf{x}}_{p,2}\tilde{\mathbf{x}}_{e,3} - \tilde{\mathbf{x}}_{e,2}\tilde{\mathbf{x}}_{p,3} = 0 \quad (4.11)$$

$$\tilde{\mathbf{x}}'_{p,1}\tilde{\mathbf{x}}'_{e,3} - \tilde{\mathbf{x}}'_{e,1}\tilde{\mathbf{x}}'_{p,3} = 0 \quad (4.12)$$

$$\tilde{\mathbf{x}}'_{p,2}\tilde{\mathbf{x}}'_{e,3} - \tilde{\mathbf{x}}'_{e,2}\tilde{\mathbf{x}}'_{p,3} = 0 \quad (4.13)$$

Piling up as many systems of the form  $\mathbf{A}\mathbf{h} = \mathbf{0}$  as Euclidean and projective point correspondences are available, we end up solving an over determined system of equations, in which again,  $\mathbf{H}$  is made up of the entries of the right singular vector associated to the smallest singular value of  $\mathbf{A}$ .

The last step in the computation of the camera motion is the extraction of the extrinsic parameters from each of the resulting projection matrices  $\tilde{\mathbf{P}}_e = \tilde{\mathbf{P}}\mathbf{H}$  and  $\tilde{\mathbf{P}}'_e = \tilde{\mathbf{P}}'\mathbf{H}$

$$[\mathbf{R}|\mathbf{t}] = [\mathbf{U}^{-1}|\mathbf{B}\tilde{\mathbf{p}}_e] \quad (4.14)$$

with  $\tilde{\mathbf{P}}_e = [\mathbf{P}_e|\tilde{\mathbf{p}}_e]$  and  $\mathbf{UB}$  the QR decomposition of  $\mathbf{P}_e^{-1}$ . The camera motion between image frame  $I_k$  and  $I_{k+1}$  is consequently,  $[\mathbf{R}'|\mathbf{t}'][\mathbf{R}|\mathbf{t}]^{-1}$ .

Figure 4.3 shows the full metric reconstruction of our three-dimensional calibration pattern achieved with this technique. The estimated inter-frame motion of the camera for this example was near  $18cm$  along the  $x$  axis, with a small rotation along the  $y$  axis of about  $18^\circ$ . The entire homogeneous transformation matrix of the incurred camera motion is

$$[\mathbf{R}'|\mathbf{t}'][\mathbf{R}|\mathbf{t}]^{-1} = \begin{bmatrix} 0.9447 & 0.0621 & -0.3220 & 181.9991 \\ -0.0532 & 0.9979 & 0.0363 & -7.0261 \\ 0.3236 & -0.0172 & 0.9460 & 34.0684 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix} \quad (4.15)$$

### 4.3.3 Euclidean reconstruction from the camera intrinsic parameters

Suppose that the intrinsic parameters  $\mathbf{K}$  of the camera that generated the image sequence have been calibrated a priori (equivalent to matrix  $\mathbf{A}$  from Appendix B). Moreover, with the assumption that  $I_k$  was acquired at the world origin, and that the camera

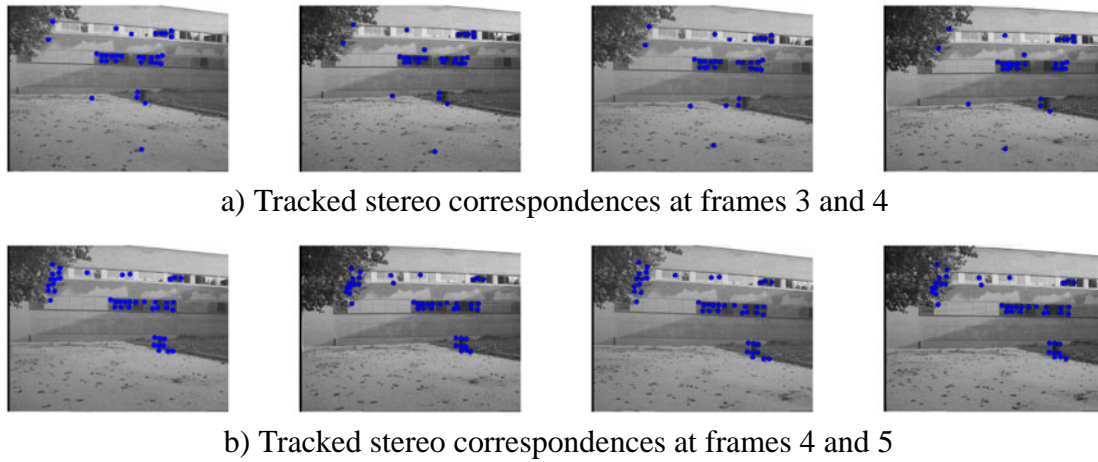


Figure 4.4: Point correspondences for a calibrated stereo pair for an outdoor mobile robot sequence.

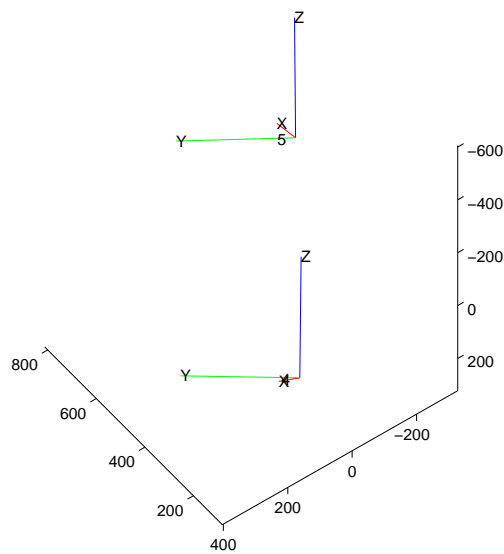


Figure 4.5: Recovered robot motion in between frames 3 and 4 of the outdoor image sequence.

motion incurred before the acquisition of  $I_{k+1}$  is simply  $[\mathbf{R}|\mathbf{t}]$ ; we have that the perspective projection matrices of our pair of images are simply

$$\tilde{\mathbf{P}}_e = \mathbf{K}[\mathbf{I}|\mathbf{0}] \quad \tilde{\mathbf{P}}'_e = \mathbf{K}[\mathbf{R}|\mathbf{t}] \quad (4.16)$$

Knowledge of the fundamental matrix  $\mathbf{F}$  and the intrinsic parameters  $\mathbf{K}$  suffice in estimating the camera motion  $[\mathbf{R}|\mathbf{t}]$  up to a scale, i.e., in obtaining a scaled coarse estimate of dead reckoning from image correspondences. The procedure for the computation of such scaled motion estimate is as follows (see [91]).

The essential matrix  $\mathbf{E} = [\mathbf{t}]_{\times} \mathbf{R}$  is related to the fundamental matrix by the expression

$$\mathbf{E} = \mathbf{K}^{\top} \mathbf{F} \mathbf{K} \quad (4.17)$$

Once the essential matrix is evaluated, we compute its singular value decomposition, which in turn obeys the form  $\mathbf{E} = \mathbf{U} \text{diag}(1, 1, 0) \mathbf{V}^{\top}$ . Furthermore,

$$\mathbf{R} = \mathbf{U} \text{Rot}_z(\pi/2) \mathbf{V}^{\top} \quad \mathbf{t} = \mathbf{U}(0, 0, 1)^{\top} \quad (4.18)$$

However this last result is defined up to scale only, and it is necessary to have some metric information in order to resolve the projective ambiguity in reconstruction. Next we explore the geometry of binocular vision to solve this situation.

#### 4.3.4 Euclidean reconstruction from a fully calibrated stereo rig

It is not always possible to extract metric information from the environment. However, we can still perform full Euclidean reconstruction if we add such metric information to the sensor directly. Such is the case of using fixed binocular vision. Fully calibrating a stereo rig a priori, we estimate the fixed inter-camera displacement  $[\mathbf{R}|\mathbf{t}]$  with the camera calibration technique described in Appendix B. Next, we provide a large enough set of image correspondences in the stereo pair for two consecutive frames, and we build the following set of equations for the computation of the homography  $\mathbf{H}$  describing the entire rig motion

$$\tilde{\mathbf{x}}_1 = \mathbf{K}_1[\mathbf{I}|\mathbf{0}]\tilde{\mathbf{y}} \quad \tilde{\mathbf{x}}'_1 = \mathbf{K}_1[\mathbf{R}|\mathbf{t}]\tilde{\mathbf{y}} \quad (4.19)$$

$$\tilde{\mathbf{x}}_2 = \mathbf{K}_2[\mathbf{I}|\mathbf{0}]\mathbf{H}\tilde{\mathbf{y}} \quad \tilde{\mathbf{x}}'_2 = \mathbf{K}_2[\mathbf{R}|\mathbf{t}]\mathbf{H}\tilde{\mathbf{y}} \quad (4.20)$$

Solving for the Euclidean coordinates of  $\tilde{\mathbf{y}}$  in Equation set 4.19, and plugging into the Equation set 4.20, we build again an overdetermined system of equations of the form

$\mathbf{A}\mathbf{h} = \mathbf{0}$ . The solution for  $\mathbf{h}$ , and consequently  $\mathbf{H}$ , is computed in the same manner as in Section 4.3.2.

Figure 4.4 shows two instances of an outdoor stereoscopic image sequence acquired with our mobile robot. The blue dots in the sequence indicate the tracked image correspondences. The recovered forward vehicle motion is shown In Figure 4.5.

## 4.4 Bibliographical notes

An in depth discussion of the various area based correlation operators used in the computation of images correspondences for stereo and tracking can be found in our technical report [144]. The tracking of Shi and Tomasi features is discussed in [181] and [196]. Robust parameter estimation and stratified reconstruction are covered in great detail in the books by Hartley and Zisserman [91] and Xu and Zhang [219].

## 4.5 Conclusions

In this Chapter we merely touch upon the vast research area of structure from motion. We use tools from the recently developed theory of multiple view geometry for the computation of initial mobile robot motion estimates given a sequential set of image correspondences. Different from current research directions in multiple view geometry, we take advantage of the possibility of fully calibrating our stereo rig in advance. We end up showing in Section 4.3.4 a motion estimation algorithm for a stereoscopic head equipped mobile robot in environments in which odometry is unreliable.



# Chapter 5

## Robot Localization

Perception happens locally, in the egocentric frame of reference of the robot. In order to ensure correspondence between the local representation of the environment built by the landmark extraction processes, and the global representation contained in a map, the robot must estimate its own position with respect to this map.

The use of stochastic models for map building and localization in mobile robotics has gained much popularity in recent years [53, 77, 122]. Of particular interest is the use of predictive filters to estimate the robot position and uncertainty, and to update these estimates from sensor readings while at the same time building an incremental map of the environment [11, 20, 43, 64, 126, 182, 187].

One of the most critical limitations to the application of such estimation-theoretic approaches to map building and localization is the data association problem. Data association refers to the issue of matching observations with previously learned elements from the environment. Some techniques can be used to alleviate the data association problem, such as the tracking of landmarks from one robot position to the next, or by using efficient tests for scene-to-model landmark match hypothesis verification. Obviously there is always a compromise between the possibility of fully invariant landmark characterization and the difficulty to extract such characterizing features from raw sensor data.

As we address issues such as viewpoint invariance and feature extraction from sensor data, it is overwhelming how undesired environment dynamics, occlusions, and sensor noise can still make data association a daunting task. One possibility to overcome the data association problem altogether is with the deployment of uniquely identifiable man-made beacons to aid in localization. Unfortunately, there exist multiple situations where this is not possible, and a map must still be constructed without environment contamination. An alternative approach explored in this work is the use of temporal

and spatial landmark quality measures to validate observations.

We start our discussion by reviewing in Section 5.1 the traditional full covariance extended Kalman filter approach to concurrent map building and localization (EKF-CML in short), based primarily on the works by Smith and Cheeseman [182] and Dissanayake *et al.* [64]. In Section 5.2, explicit formulas for two mobile platforms are presented. First, we show the case of a simple linear one-dimensional mobile robot, the *monobot*. Then, we extend the analysis to the more realistic case of a planar mobile robot. Section 5.3 deals with specific model assumptions of the second more realistic nonlinear mobile robot platform, covering topics such as path planning, and plant and measurement noise models.

Spatial landmark compatibility tests are needed to validate data association hypothesis in terms of the estimated localization error for each landmark. Their use is crucial for the solution of data association in CML [43, 150]. We have realized however, that in situations with moderate scene dynamics, spatial landmark compatibility may not suffice in the search for data association matches. Consider for example the case when a landmark is occluded for a short period of time. A spatial compatibility test would not have any information on the history of observations of such landmark, and might still be trying to wrongly associate it with a neighboring observed feature. If the algorithm succeeds in incorrectly associating the occluded feature, the new observation will not be consistent with the initial measurement, thus producing large error in the estimate for the localization of that landmark, while at the same time underestimating its covariance. Given that the map covariance is fully correlated, starting with the next iteration of the algorithm, that wrong value for the uncertainty would be propagated to the rest of the landmark locations, and that of the robot as well; leading to divergence in the map, and ultimately breaking down the entire estimation approach to CML.

To aid in those situations in which landmark observations might not be consistent in time, we propose a new set of temporal landmark quality models, and show how by incorporating these models, the overall estimation-theoretic approach to CML is improved. With the aid of these models, a new temporal landmark quality test is presented to aid in differentiating between the imprecision in the localization of a landmark, and the uncertainty in its very existence. Thanks to this test we are able to remove weak landmarks from the map. Landmarks that would most likely be a product of false data association or spurious observations, and that if considered, would otherwise induce undesired localization errors. Temporal landmark compatibility is addressed in Section 5.4.

Finally, in Section 5.5, our planar mobile robot configuration is used to evaluate the original full-covariance Extended Kalman Filter algorithm to Concurrent Map Building and Localization as reported by Dissanayake *et al* [64], including the spatial landmark compatibility tests [150], versus our improved algorithm, the EKF-CML-LV, with both

temporal and spatial landmark quality tests, both in the presence of various noise levels, and ultimately, in cases with limited field of view and extreme data missassociation.

## 5.1 Preliminaries

In this Section we review the fundamentals of the stochastic estimation approach to concurrent map building and localization. The material covered summarizes the work of many researchers during the past 15 years, and will constitute a starting point for our view of the mobile robot localization and map building problem.

### 5.1.1 Full covariance EKF approach to CML

Before delving into the mathematical formulation that builds up the full covariance Extended Kalman Filter approach to Concurrent Map Building and Localization, we proceed with a motivational explanation of how such a predictive filter can be used to solve the localization problem.

Imagine, for the sake of simplicity, a planar mobile robot that we believe is located at position  $\mathbf{x}_{r,k|k}$  as shown in Figure 5.1; and that such location is known with some level of uncertainty indicated by the area inside the ellipse  $\mathbf{P}_{rr,k|k}$ . The subscripts  $r$  and  $rr$  indicate the part of the state vector  $\mathbf{x}$  and of the error covariance matrix  $\mathbf{P}$  associated to the robot pose. The ellipse indicates a level curve of equal uncertainty, its principal axes are the eigenvectors of the covariance matrix  $\mathbf{P}_{rr}$ . The term  $k$  is used to indicate the time stamp.

Driving the robot with the motion command  $\mathbf{u}_k$ , we would expect it to end up at location  $\mathbf{x}_{r,k+1|k}$ . Such location estimate, the *a priori* state estimate, possesses a larger level of uncertainty than our previous estimate  $\mathbf{x}_{r,k|k}$ . The reason, is the addition of dead reckoning error and the inaccuracies of our motion model. This increase in uncertainty is exemplified by the larger area inside the ellipse  $\mathbf{P}_{rr,k+1|k}$ .

Once the motion command is complete, the robot sensors acquire information about the environment. More specifically, they measure the distance from the robot to a number of fixed landmarks  $\mathbf{x}_f^{(1)}$ ,  $\mathbf{x}_f^{(2)}$ , and  $\mathbf{x}_f^{(3)}$ . It is from these observations, that the filter corrects its estimate about the location of the robot to  $\mathbf{x}_{r,k+1|k+1}$ , the *a posteriori* estimate; and at the same time, reduces the localization uncertainty  $\mathbf{P}_{rr,k+1|k+1}$ . The true robot location  $\mathbf{x}_{r,k+1}$  is still unknown; however, the idea behind the use of a predictive filter is to minimize the estimation error ( $\mathbf{x}_{r,k+1} - \mathbf{x}_{r,k+1|k+1}$ ). In other words, we want an estimator that keeps the uncertainty ellipse for  $\mathbf{P}_{rr,k+1|k+1}$  as small as possible.

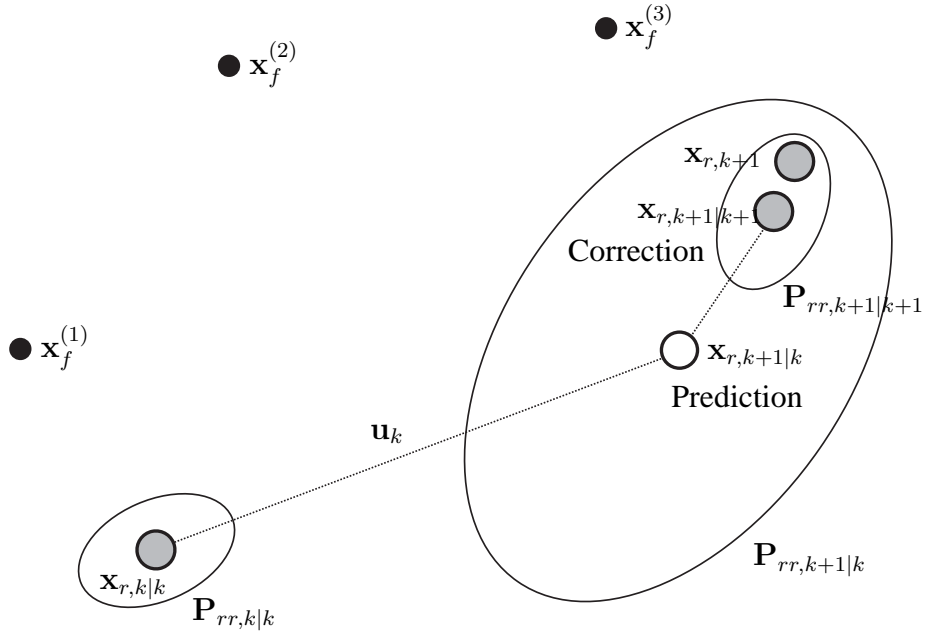


Figure 5.1: Location estimate with a predictive filter.

Fortunately, we can resort to the Kalman filter, a recursive stochastic state estimator for partially observed non-stationary processes that gives an optimal state estimate in the least squares sense. In the typical full-covariance EKF based approach to CML we use such a filter precisely to iteratively estimate the robot and landmark locations, optimally reducing the error between the true robot location and our computed estimate.

### 5.1.2 Map model

Formally speaking, the motion of the robot and the measurement of the map features are governed by the discrete-time state transition model

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k) \quad (5.1)$$

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{w}_k) \quad (5.2)$$

The state vector  $\mathbf{x}_k$  contains the position of the robot  $\mathbf{x}_{r,k}$  at time step  $k$ , and a vector

of stationary map features  $\mathbf{x}_f$ , i.e.,

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_{r,k} \\ \mathbf{x}_f \end{bmatrix} \quad (5.3)$$

The input vector  $\mathbf{u}_k$  is the vehicle control command, and  $\mathbf{v}_k$  is a Gaussian random vector with zero mean and covariance matrix  $\mathbf{V}_k$ , representing unmodeled robot dynamics and system noise. The function  $\mathbf{f}$  is a possibly nonlinear difference equation that models the motion of the robot.

The Gaussian random vector  $\mathbf{w}_k$  represents both, the inaccuracies of the also possibly nonlinear observation model  $\mathbf{h}$ , and the measurement noise with zero mean and covariance matrix  $\mathbf{W}_k$ .

Provided the set of observations  $Z^k = \{\mathbf{z}_1, \dots, \mathbf{z}_k\}$  was available for the computation of the current map estimate  $\mathbf{x}_{k|k}$ , the expression

$$\mathbf{x}_{k+1|k} = \mathbf{f}(\mathbf{x}_{k|k}, \mathbf{u}_k, \mathbf{0}) \quad (5.4)$$

gives an a priori noise-free estimate of the new locations of the robot and map features after the vehicle control command  $\mathbf{u}_k$  is input to the system. Similarly,

$$\mathbf{z}_{k+1|k} = \mathbf{h}(\mathbf{x}_{k+1|k}, \mathbf{0}) \quad (5.5)$$

constitutes a noise-free a priori estimate of sensor measurements.

The EKF approach to CML requires the linearization of both the motion and observation models. Such linearizations are formulated as Taylor series approximations with the higher order terms dropped, i.e.,

$$\mathbf{x}_{k+1} \approx \mathbf{x}_{k+1|k} + \mathbf{F}_x(\mathbf{x}_k - \mathbf{x}_{k|k}) + \mathbf{F}_v \mathbf{v}_k \quad (5.6)$$

$$\mathbf{z}_{k+1} \approx \mathbf{z}_{k+1|k} + \mathbf{H}_x(\mathbf{x}_{k+1} - \mathbf{x}_{k+1|k}) + \mathbf{H}_w \mathbf{w}_{k+1} \quad (5.7)$$

The Jacobian matrices  $\mathbf{F}_x$ ,  $\mathbf{F}_v$ ,  $\mathbf{H}_x$ , and  $\mathbf{H}_w$  contain the partial derivatives of  $\mathbf{f}$  and

$\mathbf{h}$  with respect to  $\mathbf{x}$  and the noises  $\mathbf{v}$  and  $\mathbf{w}$ :

$$\mathbf{F}_{\mathbf{x}} = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{(\mathbf{x}_{k|k}, \mathbf{u}_k, \mathbf{0})} \quad (5.8)$$

$$\mathbf{F}_{\mathbf{v}} = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{v}} \right|_{(\mathbf{x}_{k|k}, \mathbf{u}_k, \mathbf{0})} \quad (5.9)$$

$$\mathbf{H}_{\mathbf{x}} = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{(\mathbf{x}_{k+1|k}, \mathbf{0})} \quad (5.10)$$

$$\mathbf{H}_{\mathbf{w}} = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{w}} \right|_{(\mathbf{x}_{k+1|k}, \mathbf{0})} \quad (5.11)$$

Given that the landmarks are considered stationary, their a priori estimate is simply

$$\mathbf{x}_{f,k+1|k} = \mathbf{x}_{f,k|k} \quad (5.12)$$

Thus, the elements of the non-stationary linear state transition model of the robot and map dynamics in Equations 5.6 and 5.7 take the forms

$$\begin{bmatrix} \mathbf{x}_{r,k+1} \\ \mathbf{x}_f \end{bmatrix} \approx \begin{bmatrix} \mathbf{x}_{r,k+1|k} \\ \mathbf{x}_{f,k|k} \end{bmatrix} + \begin{bmatrix} \mathbf{F}_{\mathbf{x}_r} & \\ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{r,k|k} \\ \tilde{\mathbf{x}}_{f,k|k} \end{bmatrix} + \begin{bmatrix} \mathbf{F}_{\mathbf{v}_r} & \\ & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{r,k} \\ \mathbf{0} \end{bmatrix} \quad (5.13)$$

$$\mathbf{z}_{k+1} \approx \mathbf{z}_{k+1|k} + \begin{bmatrix} \mathbf{H}_{\mathbf{x}_r} & \mathbf{H}_{\mathbf{x}_f} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{r,k+1|k} \\ \tilde{\mathbf{x}}_{f,k+1|k} \end{bmatrix} + \mathbf{H}_{\mathbf{w}} \mathbf{w}_{k+1} \quad (5.14)$$

with the *a posteriori state error*

$$\tilde{\mathbf{x}}_{k|k} = \begin{bmatrix} \tilde{\mathbf{x}}_{r,k|k} \\ \tilde{\mathbf{x}}_{f,k|k} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{r,k} - \mathbf{x}_{r,k|k} \\ \mathbf{x}_f - \mathbf{x}_{f,k|k} \end{bmatrix} \quad (5.15)$$

and the *a priori state error*

$$\tilde{\mathbf{x}}_{k+1|k} = \begin{bmatrix} \tilde{\mathbf{x}}_{r,k+1|k} \\ \tilde{\mathbf{x}}_{f,k+1|k} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{r,k+1} - \mathbf{x}_{r,k+1|k} \\ \mathbf{x}_f - \mathbf{x}_{f,k+1|k} \end{bmatrix} \quad (5.16)$$

In the remaining of this chapter, extensive use of Kalman Filter related notation will be used. A few terms that we deem necessary to keep in mind include: the error covariance matrix  $\mathbf{P}$ , the measurement innovation covariance matrix  $\mathbf{S}$ , and the Kalman gain  $\mathbf{K}$ . If in doubt, please refer to the Notation section at the beginning of this thesis, and ultimately, to Appendix C.

### 5.1.3 Prediction

An a priori prediction of the location of the robot and the state of the map is computed in Equation 5.4 purely from motion commands; consequently increasing the uncertainty of the robot location and that of the map features. In general terms, the a priori estimate to the map state error covariance showing this increase of uncertainty is given by

$$\begin{aligned} \mathbf{P}_{k+1|k} &= E[\tilde{\mathbf{x}}_{k+1|k} \tilde{\mathbf{x}}_{r,k+1|k}^\top] \\ &= \mathbf{F}_x \mathbf{P}_{k|k} \mathbf{F}_x^\top + \mathbf{F}_v \mathbf{V}_k \mathbf{F}_v^\top \end{aligned} \quad (5.17)$$

Writing the map state error covariance matrix in block form, and substituting the corresponding Jacobian matrices, we rewrite Equation 5.17 as

$$\begin{bmatrix} \mathbf{P}_{rr,k+1|k} & \mathbf{P}_{rf,k+1|k} \\ \mathbf{P}_{rf,k+1|k}^\top & \mathbf{P}_{ff,k+1|k} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{x_r} \mathbf{P}_{rr,k|k} \mathbf{F}_{x_r}^\top + \mathbf{F}_{v_r} \mathbf{V}_{r,k} \mathbf{F}_{v_r}^\top & \mathbf{F}_{x_r} \mathbf{P}_{rf,k|k} \\ \mathbf{P}_{rf,k|k}^\top \mathbf{F}_{x_r}^\top & \mathbf{P}_{ff,k|k} \end{bmatrix} \quad (5.18)$$

### 5.1.4 Correction

Assuming that a new set of landmark observations  $\mathbf{z}_{k+1}$  coming from sensor data has been correctly matched to their map counterparts, one can compute the error between the measurements and the estimates with

$$\tilde{\mathbf{z}}_{k+1|k} = \mathbf{z}_{k+1} - \mathbf{z}_{k+1|k} \quad (5.19)$$

This error aids in revising the map and robot locations. The a posteriori state estimate is

$$\mathbf{x}_{k+1|k+1} = \mathbf{x}_{k+1|k} + \mathbf{K}_{k+1} \tilde{\mathbf{z}}_{k+1|k} \quad (5.20)$$

and the Kalman gain is computed with

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1|k} \mathbf{H}_x^\top \mathbf{S}^{-1} \quad (5.21)$$

where  $\mathbf{S}$  is the *measurement innovation matrix*,

$$\mathbf{S} = \mathbf{H}_x \mathbf{P}_{k+1|k} \mathbf{H}_x^\top + \mathbf{H}_w \mathbf{W}_{f,k+1} \mathbf{H}_w^\top \quad (5.22)$$

Finally, the a posteriori estimate of the map state error covariance must also be revised once a measurement has taken place. It is revised with

$$\mathbf{P}_{k+1|k+1} = (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_x) \mathbf{P}_{k+1|k} \quad (5.23)$$

or equivalently, and to guarantee *positive semi-definiteness* of  $\mathbf{P}_{k+1|k+1}$ , with

$$\mathbf{P}_{k+1|k+1} = (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_x) \mathbf{P}_{k+1|k} (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_x)^\top + \mathbf{K}_{k+1} \mathbf{W}_{f,k+1} \mathbf{K}_{k+1}^\top \quad (5.24)$$

The above expression is commonly referred as the Joseph form of the *a posteriori* state error covariance matrix. Its derivation is discussed in detail in Section C.4. The properties of positive semi-definite (psd) matrices are enumerated in Section D.1 for completeness of the discussion.

The contribution to the revision of the robot pose and landmark location estimates is proportional to our degree of trust in the motion and sensor models respectively. If the plant error covariance  $\mathbf{V}$  is large, and the measurement error covariance  $\mathbf{W}$  is small, the EKF-CML algorithm trusts more the observations than dead-reckoning, revising more heavily the robot pose estimate than that of the landmarks. Conversely, when the measurement error covariance is larger than the plant error covariance, the algorithm trusts more on the motion of the robot and ends up revising more heavily the landmark estimates.

### 5.1.5 Convergence properties of the full covariance EKF approach to CML

One important property of the estimation-theoretic approach to CML is that the map is asymptotically convergent. That is, in the original full covariance KF-based CML formulation the map state error covariance submatrix associated with the landmark estimates decreases monotonically as successive observations take place. Formally speaking,

$$\det \mathbf{P}_{ff,k+1|k+1} \leq \det \mathbf{P}_{ff,k|k} \quad (5.25)$$

Another property indicates how *in the limit*, as the number of iterations tends to infinity, the map becomes fully correlated; suggesting that if a landmark location is given, the location of the other landmarks can be deduced with absolute certainty from the map built.

$$\lim_{k \rightarrow \infty} \det \mathbf{P}_{ff,k|k} = 0 \quad (5.26)$$

The third property in the analysis of the covariance in EKF-CML, is that in the limit, the absolute location of the vehicle and map is bounded by the initial vehicle uncertainty. And for a vehicle with no process noise

$$\lim_{k \rightarrow \infty} \mathbf{P}_{rr,k|k} = \mathbf{P}_{rr,0|0} \quad (5.27)$$



These properties were first reported in Newman’s PhD work [64, 151], and will be of use in next Chapter where we talk about partial observability, partial controllability, and filter stability in CML.

### 5.1.6 Sequential innovation

EKF-CML requires that all landmarks in the map be always observed by the vehicle, and be correctly associated to their map counterpart at every iteration. However, it is also possible to revise the state estimate only with partial observations. Independent landmark measurements contribute only to the revision of the map states directly associated to that particular landmark; and, if a sufficient number of independent landmark observations are made, it is still possible to have the system reliably estimate the location of the robot and landmarks iteratively. The technique used to update the state estimate one observation at a time is called *sequential innovation* (see Section C.5). The result of applying sequential innovation to all the landmarks in the map is equivalent to that of using the full covariance extended Kalman filter approach, provided the observations are independent, or that at least they can be *whitened* [213].

The main advantage of using sequential innovation, is that by considering the measurement vector  $\mathbf{z}_{k+1}$  as a set of single measurements  $\mathbf{z}_{k+1}^{(i)}$  that can be treated sequentially, the inversion of the joint measurement innovation covariance matrix  $\mathbf{S}$  is no longer necessary. Instead, a series of smaller individual innovation covariance matrix inverses is computed, reducing considerably the time complexity of the algorithm.

Each landmark measurement  $\mathbf{z}_{k+1}^{(i)}$  provides information for the repair of the state estimate  $\mathbf{x}_{k+1|k}$  in the direction spanned by the  $i$ -th set of columns of  $\mathbf{K}_{k+1}$ , thus, the a posteriori robot and map estimates are given by the sum of the a priori estimate and a linear combination of the independent measurement errors

$$\mathbf{x}_{k+1|k+1} = \mathbf{x}_{k+1|k} + \sum_{i=1}^n \mathbf{K}_{k+1,(i)} \tilde{\mathbf{z}}_{k+1|k}^{(i)} \quad (5.28)$$

Similarly, independent landmark measurement models can be constructed taking only the corresponding  $i$ -th set of rows of the measurement Jacobians  $\mathbf{H}_x$  and  $\mathbf{H}_w$

$$\mathbf{z}_{k+1}^{(i)} \approx \mathbf{z}_{k+1|k}^{(i)} + \mathbf{H}_x^{(i)} \tilde{\mathbf{x}}_{k+1|k}^{(i)} + \mathbf{H}_w^{(i)} \mathbf{w}^{(i)} \quad (5.29)$$

The independent treatment of the observations is only possible if  $\mathbf{W}$  is block diagonal. This is, when the landmark observations taken during the same time interval are

uncorrelated. The  $i$ -th set of columns of  $\mathbf{K}_{k+1}$  is simply computed with

$$\mathbf{K}_{k+1,(i)} = \mathbf{P}_{k+1|k} \mathbf{H}_{\mathbf{x}}^{(i)\top} \mathbf{S}^{(i)-1} \quad (5.30)$$

with the *individual innovation covariance matrices*

$$\mathbf{S}^{(i)} = \mathbf{H}_{\mathbf{x}}^{(i)} \mathbf{P}_{k+1|k} \mathbf{H}_{\mathbf{x}}^{(i)\top} + \mathbf{H}_{\mathbf{w}}^{(i)} \mathbf{W}_{f,k+1}^{(i)} \mathbf{H}_{\mathbf{w}}^{(i)\top} \quad (5.31)$$

The required inverse in Equation 5.30 is of  $\dim \mathbf{z}^{(i)}$ , and is considerably much smaller than the dimensions of the entire measurement vector  $\mathbf{z}$  as required in Equation 5.21. However, if one still wishes to compute the inverse of the whole innovation covariance matrix  $\mathbf{S}$  for more than one landmark at a time, one can resort to an incremental computation of  $\mathbf{S}^{-1}$  as shown in Section 5.1.10.

Another consequence of using sequential innovation is that the model Jacobians can also be re-computed after each observation is incorporated to the filter, thus producing even more accurate state and state covariance estimates.

### 5.1.7 Covariance initialization

Another crucial implementation aspect of the full covariance EKF approach to CML is the initialization of the error covariance matrix  $\mathbf{P}$  as new observations are added to the map. This matrix contains the expected robot and landmark localization error, and will only manifest the asymptotic convergence properties shown in Section 5.1.5 if it is initialized properly.

The function that maps an observation into world coordinates is given by our linearized measurement model, and has the form

$$\begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_f^{(i)} \end{bmatrix} = \mathbf{G} \begin{bmatrix} \mathbf{x}_r \\ \mathbf{z}^{(i)} \end{bmatrix} \quad (5.32)$$

$\mathbf{G}$  is known as the *feature initialization* matrix. Solving for  $\mathbf{G}$  from Equations 5.5 and 5.7, we get

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{H}_{\mathbf{x}_f^{(i)}}^{-1} \mathbf{H}_{\mathbf{x}_r^{(i)}} & \mathbf{H}_{\mathbf{x}_f^{(i)}}^{-1} \end{bmatrix} \quad (5.33)$$

Consequently, the initialization of the corresponding map state error covariance for such landmark is given by

$$\mathbf{P} = \mathbf{G} \begin{bmatrix} \mathbf{P}_{rr} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^{(i)} \end{bmatrix} \mathbf{G}^\top \quad (5.34)$$

Without loss of generality, assume a map with  $n - 1$  landmarks at time step  $k$ . The map state error covariance matrix has the form

$$\mathbf{P}_{k|k} = \begin{bmatrix} \mathbf{P}_{rr,k|k} & \mathbf{P}_{rf^{(1)},k|k} & \cdots & \mathbf{P}_{rf^{(n-1)},k|k} \\ \mathbf{P}_{rf^{(1)},k|k}^\top & \mathbf{P}_{f^{(1)}f^{(1)},k|k} & \cdots & \mathbf{P}_{f^{(1)}f^{(n-1)},k|k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{rf^{(n-1)},k|k}^\top & \mathbf{P}_{f^{(1)}f^{(n-1)},k|k}^\top & \cdots & \mathbf{P}_{f^{(n-1)}f^{(n-1)},k|k} \end{bmatrix} \quad (5.35)$$

Once the robot has observed a sufficiently robust new feature which cannot be associated to any other landmark in the map, it is labeled as the  $n$ -th landmark, and a new row and column must be appended to the map covariance matrix with

$$\mathbf{P}_{rf^{(n)},k|k} = -\mathbf{P}_{rr,k|k}(\mathbf{H}_{\mathbf{x}_f^{(n)}}^{-1}\mathbf{H}_{\mathbf{x}_r^{(n)}})^\top \quad (5.36)$$

$$\mathbf{P}_{f^{(i)}f^{(n)},k|k} = \mathbf{H}_{\mathbf{x}_f^{(i)}}^{-1}\mathbf{H}_{\mathbf{x}_r^{(i)}}\mathbf{P}_{rr,k|k}(\mathbf{H}_{\mathbf{x}_f^{(n)}}^{-1}\mathbf{H}_{\mathbf{x}_r^{(n)}})^\top \quad (5.37)$$

$$\mathbf{P}_{f^{(n)}f^{(n)},k|k} = \mathbf{H}_{\mathbf{x}_f^{(n)}}^{-1}\mathbf{H}_{\mathbf{x}_r^{(n)}}\mathbf{P}_{rr,k|k}(\mathbf{H}_{\mathbf{x}_f^{(n)}}^{-1}\mathbf{H}_{\mathbf{x}_r^{(n)}})^\top + \mathbf{H}_{\mathbf{x}_f^{(n)}}^{-1}\mathbf{W}^{(i)}(\mathbf{H}_{\mathbf{x}_f^{(n)}}^{-1})^\top \quad (5.38)$$

Equations 5.36-5.38 indicate that the initialization of the new feature map error covariance is a function of the actual vehicle position and its accumulated uncertainty.

### 5.1.8 Landmark spatial uncertainty

The estimated uncertainty in the localization of every landmark in the map, as well as that of the robot, is maintained in the state error covariance  $\mathbf{P}$ . Consequently, the uncertainty of its location in observation space is given by the change of basis of  $\mathbf{P}$  plus that of the independent sensor uncertainties  $\mathbf{W}^{(i)}$ . This quantity is called the innovation covariance matrix, and is precisely equivalent to the term already introduced in Equation 5.22 from Section 5.1.4.

For any particular measurement  $\mathbf{z}_{k+1}^{(i)}$ , the squared Mahalanobis distance

$$d_i^2 = \tilde{\mathbf{z}}_{k+1|k}^{(i)\top} \mathbf{S}^{(i)-1} \tilde{\mathbf{z}}_{k+1|k}^{(i)} \quad (5.39)$$

represents a measure of spatial disparity between the observation  $\mathbf{z}_{k+1}^{(i)}$  and the estimated location in robot centered coordinates of the hypothetical landmark match  $\mathbf{x}_f^{(i)}$ .

Two spatial landmark compatibility tests that appear in the literature are the individual compatibility test and the joint compatibility test. The former considers landmark

observations independently and has been widely used not only to validate observations within the framework of the CML problem, but on a wide range of applications. A related example in the computer vision literature, is for example, the evaluation of visual correspondences in the computation of the fundamental matrix of the two-view geometry [91]. The joint compatibility test on the other hand, due to Neira and Tardós [150], considers cross correlated landmark uncertainties when testing match hypotheses, at the expense of higher computational cost.

### 5.1.9 Individual compatibility test

The squared Mahalanobis distance in Equation 5.39 is a weighted squared sum of  $\dim \mathbf{z}^{(i)}$  Gaussian variables and as such, follows a  $\chi^2$  distribution. Moreover, the subset of map features compatible with measurement  $\mathbf{z}_{k+1}^{(i)}$  are the ones that satisfy the  $\chi^2$  compatibility test

$$d_i^2 \leq \chi_{\dim \mathbf{z}^{(i)}, \alpha}^2 \quad (5.40)$$

The number of degrees of freedom in the individual  $\chi^2$  compatibility test is given by the rank of the one landmark innovation covariance matrix  $\mathbf{S}^{(i)}$ . Given that  $\mathbf{S}^{(i)}$  is invertible, it must have full rank. Consequently, the number of degrees of freedom of the test is  $\dim \mathbf{z}^{(i)}$ . So for example, if  $\mathbf{z}^{(i)}$  is an image point, the squared Mahalanobis distance represents the sum of the squared  $u$  and  $v$  measurement errors, weighted by the covariance matrix  $\mathbf{S}^{(i)}$ ; and in this case, the number of degrees of freedom of the individual compatibility test is 2. The term  $\alpha$  indicates the desired confidence level; so for example, a value of  $\alpha = 0.95$  indicates a 95% probability that the observation  $\mathbf{z}^{(i)}$  matches the map feature  $\mathbf{x}_f^{(i)}$ . In other words, a 5% probability that correct data associations will not pass the compatibility test. By increasing the confidence level  $\alpha$ , we are augmenting the rate of observations that pass the compatibility test, even if they correspond to incorrect matches.

One clear drawback of the individual compatibility test is the fact that no correlational information is used. An expansion of  $\mathbf{S}^{(i)}$ , reveals how the square Mahalanobis distance  $d_i^2$  is indeed an individual measure of landmark compatibility, and that no landmark correlation information is taken into account.

$$\begin{aligned} \mathbf{S}^{(i)} = & \mathbf{H}_{\mathbf{x}_r}^{(i)} \mathbf{P}_{r, k+1|k} \mathbf{H}_{\mathbf{x}_r}^{(i)\top} + \mathbf{H}_{\mathbf{x}_f}^{(i)} \mathbf{P}_{rf, k+1|k} \mathbf{H}_{\mathbf{x}_r}^{(i)\top} + \mathbf{H}_{\mathbf{x}_r}^{(i)} \mathbf{P}_{rf, k+1|k} \mathbf{H}_{\mathbf{x}_f}^{(i)\top} \\ & + \mathbf{H}_{\mathbf{x}_f}^{(i)} \mathbf{P}_{ff, k+1|k} \mathbf{H}_{\mathbf{x}_f}^{(i)\top} + \mathbf{H}_{\mathbf{w}}^{(i)} \mathbf{W}_{k+1}^{(i)} \mathbf{H}_{\mathbf{w}}^{(i)\top} \end{aligned} \quad (5.41)$$

To include the landmark correlation information in the test of measurement and map pairings, one must recur to the joint compatibility test.

### 5.1.10 Joint compatibility test

Landmark correlation is taken into account when we test spatial match hypotheses for more than one landmark at a time. The joint measurement innovation covariance for two or more landmarks is

$$\mathbf{S}^{(p\dots r)} = \mathbf{H}_{\mathbf{x}}^{(p\dots r)} \mathbf{P}_{k+1|k} \mathbf{H}_{\mathbf{x}}^{(p\dots r)\top} + \mathbf{H}_{\mathbf{w}}^{(p\dots r)} \mathbf{W}_{k+1} \mathbf{H}_{\mathbf{w}}^{(p\dots r)\top} \quad (5.42)$$

or equivalently,

$$\mathbf{S}^{(p\dots r)} = \begin{bmatrix} \mathbf{S}^{(p\dots q)} & \mathbf{H}_{\mathbf{x}}^{(p\dots q)} \mathbf{P}_{k+1|k} \mathbf{H}_{\mathbf{x}}^{(r)\top} \\ \mathbf{H}_{\mathbf{x}}^{(r)} \mathbf{P}_{k+1|k} \mathbf{H}_{\mathbf{x}}^{(p\dots q)\top} & \mathbf{S}^{(r)} \end{bmatrix} \quad (5.43)$$

The indices  $(p \dots r)$  need not be consecutive. The joint Mahalanobis distance is

$$d_{p\dots r}^2 = \tilde{\mathbf{z}}_{k+1|k}^{(p\dots r)\top} \mathbf{S}^{(p\dots r)-1} \tilde{\mathbf{z}}_{k+1|k}^{(p\dots r)} \quad (5.44)$$

and the joint compatibility test for two or more landmarks is

$$d_{p\dots r}^2 \leq \chi_{\dim \mathbf{z}^{(p\dots r)}, \alpha}^2 \quad (5.45)$$

The joint compatibility test can be computationally expensive when the number of landmarks involved is large. It would require the inverse of a joint measurement covariance matrix of size  $\dim \mathbf{z}^{(p\dots r)} \times \dim \mathbf{z}^{(p\dots r)}$  with complexity  $O((\dim \mathbf{z}^{(p\dots r)})^3)$ . This complexity can be reduced however, as shown next, with an *incremental* computation of  $\mathbf{S}^{(p\dots r)-1}$ .

Without loss of generality, once the inverse up to the  $q$ -th landmark  $\mathbf{S}^{(p\dots q)-1}$  has been computed, from the matrix inverse lemma for block matrices<sup>1</sup>, we have that

$$\mathbf{S}^{(p\dots r)-1} = \begin{bmatrix} \mathbf{S}^{(p\dots q)-1} + \mathbf{S}^{(p\dots q)-1} \mathbf{M} \mathbf{N} \mathbf{M}^\top \mathbf{S}^{(p\dots q)-1} & \mathbf{S}^{(r)-1} \mathbf{M} \mathbf{N} \\ -\mathbf{N} \mathbf{M}^\top \mathbf{S}^{(r)-1} & \mathbf{N} \end{bmatrix} \quad (5.46)$$

with

$$\mathbf{M} = \mathbf{H}_{\mathbf{x}}^{(p\dots q)} \mathbf{P}_{k+1|k} \mathbf{H}_{\mathbf{x}}^{(r)\top} \quad (5.47)$$

$$\mathbf{N} = (\mathbf{S}^{(r)-1} - \mathbf{M}^\top \mathbf{S}^{(r)-1} \mathbf{M})^{-1} \quad (5.48)$$

---

<sup>1</sup>See Appendix D.

Feature model	points on a line	image points, lines on a plane	points, lines, planes in 3d
$\dim \mathbf{z}^{(p)}$	1	2	3
Number of features $n$	$\chi_{n,0.95}^2$	$\chi_{2n,0.95}^2$	$\chi_{3n,0.95}^2$
1	3.84	5.99	7.91
2	5.99	9.49	12.6
3	7.81	12.6	16.9
4	9.49	15.5	21.0
5	11.1	18.3	25.0
10	18.3	31.4	43.8
15	25.0	43.8	61.6
20	31.4	55.8	79.1
30	43.8	79.1	113.1

Table 5.1: The Pearson  $\chi^2$  distribution needed for the computation of the square Mahalanobis distance threshold  $d^2 = \chi_{\dim \mathbf{z}^{(p\dots r)}, \alpha}^2$  during the validation of the hypothesis test that measurements  $\mathbf{z}^{(p\dots r)}$  correspond to features  $\mathbf{x}_f^{(p\dots r)}$  with probability  $\alpha = 0.95$ .

At the end, the computation of the full inverse  $\mathbf{S}^{(p\dots r)-1}$  is simplified with an incremental computation requiring a total of  $r - p$  individual covariance matrices as indicated in Equation 5.48. Values of  $\chi_{\dim \mathbf{z}^{(p\dots r)}, \alpha}^2$  for  $\alpha = 0.95$  and  $\alpha = 0.99$ , are tabulated in Tables 5.1 and 5.2 respectively.

If the regions of uncertainty around each landmark in the map do not overlap, the individual compatibility test (ICT in the sequel) might be sufficient to obtain a robust solution to the data association problem in CML. Such is the case of sparse maps, or when precise sensors are available.

However, in situations with large amounts of clutter, or with very uncertain sensor models, the joint compatibility test (JCT in the sequel) might be more useful. One should not take for granted however, the increase in the computational cost incurred by the JCT. Testing the joint compatibility of tens of landmarks might deem impossible a real time implementation of the algorithm.

Feature model	points on a line	image points, lines on a plane	points, lines, planes in 3d
$\dim \mathbf{z}^{(p)}$	1	2	3
Number of features $n$	$\chi_{n,0.99}^2$	$\chi_{2n,0.99}^2$	$\chi_{3n,0.99}^2$
1	6.63	9.21	11.3
2	9.21	13.3	16.8
3	11.3	16.8	21.7
4	13.3	20.1	26.2
5	15.1	23.2	30.6
10	23.2	37.6	50.9
15	30.6	50.9	70.0
20	37.6	63.7	88.4
30	50.9	88.4	124.1

Table 5.2: The Pearson  $\chi^2$  distribution needed for the computation of the square Mahalanobis distance threshold  $d^2 = \chi_{\dim \mathbf{z}^{(p\dots r)}, \alpha}^2$  during the validation of the hypothesis test that measurements  $\mathbf{z}^{(p\dots r)}$  correspond to features  $\mathbf{x}_f^{(p\dots r)}$  with probability  $\alpha = 0.99$ .

In situations where multiple data association hypotheses are to be verified, the joint compatibility criterion to landmark matching is implemented in branch and bound algorithms that search efficiently the space of compatible landmark association solutions. An exemplary application of one of such heuristics is detailed in the aforementioned contribution by Neira and Tardós [150].

Figure 5.2 exemplifies the use of the ICT and JCT for data association hypothesis validation. The left column of plots in the Figure shows a robot localization sequence with three landmarks. For such a small number of landmark observations, the number of match hypothesis that fail the test is similar in both cases (6.6% for ICT and 8% for JCT). The localization error results of using the ICT and JCT in this case are nearly equivalent.

On the other hand, the right column of plots in the same Figure shows the results of using the ICT and JCT in a localization sequence with 10 landmarks and similar noise parameters. In this case, the number of times that match hypothesis fail the JCT nearly

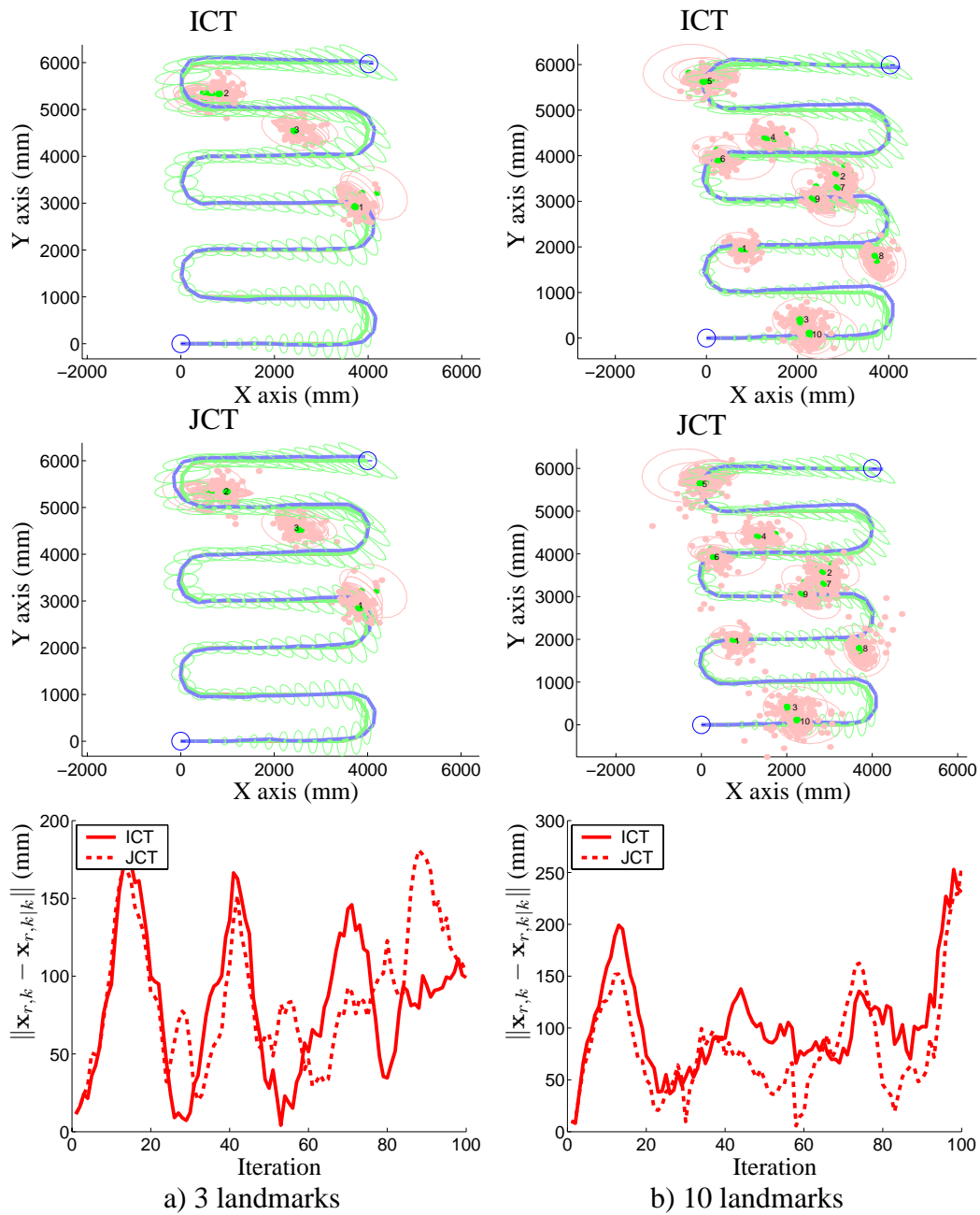


Figure 5.2: Spatial compatibility tests. Individual and joint compatibility test performance for a map with 3 and 10 landmarks over 100 steps.



doubles those of the ICT (30% for JCT vs. 17.7% for ICT); indicating perhaps that the ICT was overconfident when testing for data association. Landmark observations do not pass the joint compatibility test unless data association is accurate for all observations, or until the uncertainty  $\mathbf{P}_{rr}$  has grown to such value as to make the all-landmark Mahalanobis distance  $\tilde{\mathbf{z}}\mathbf{S}^{-1}\tilde{\mathbf{z}}$  sufficiently small. It might happen in the JCT case, that for a large number of the filter iterations (30% in this test run), vehicle location estimates end up being computed from dead reckoning only. It is certainly better however not to revise the robot localization at all, than doing so with incorrect data association. As indicated in the right bottom plot, the overall performance of the algorithm shows some improvement when the JCT is used.

The red dots in the ICT map plots represent all of the observations that pass the test, whereas the red dots in the JCT map plots show all landmark observations, regardless if they passed the test or not. The estimated robot localization error is represented by the green ellipses, which correspond to  $2\sigma$  level curves of the robot pose error covariance estimate. The red ellipses on the other hand, indicate projected  $2\sigma$  bounds for the landmark covariance estimates. The solid red line and the dashed red line in the two bottom plots are used to indicate ICT and JCT localization error, respectively.

## 5.2 Mobile robot platforms

In this section we provide explicit expressions for two mobile robot configurations, a one-dimensional robot moving along a straight line, and a planar wheeled mobile robot on a two-dimensional environment.

### 5.2.1 Monobot

Consider the one-dimensional robot (*monobot*) from Figure 5.3. The robot location is  $x_{r,k}$ , and the motion command is  $u_{r,k}$ . The robot error dynamics is  $v_{r,k}$ , and the vehicle process model is simply

$$x_{r,k+1} = x_{r,k} + u_{r,k} + v_{r,k} \quad (5.49)$$

The map consists of the set  $\mathbf{x}_f$  of static landmark points. The measurement equation for each landmark is

$$z_k^{(i)} = x_f^{(i)} - x_{r,k} - w_k^{(i)} \quad (5.50)$$

with  $w_k^{(i)}$  the landmark observation error.

The study of this simple robot model might seem naive at this point. Its relevance

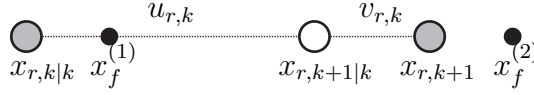


Figure 5.3: *Monobot*, a one-dimensional mobile robot.

will become evident in Chapter 6, where we study in detail the observability and controllability aspects of the stochastic approach to CML. Suffice to say for the time being, that thanks to the linearity of the model, the approximations in Equations 5.6 and 5.7 are exact, and that

$$\mathbf{F}_{x_r} = \mathbf{F}_{v_r} = \mathbf{H}_{x_f^{(i)}} = \mathbf{H}_{w^{(i)}} = 1 \quad (5.51)$$

$$\mathbf{H}_{x_r} = -1 \quad (5.52)$$

## 5.2.2 Planar mobile robot

For the more realistic mobile robot dynamics shown in Figure 5.4, the vehicle state is defined by  $\mathbf{x}_r = [x, y, \theta]^\top$  where  $x$  and  $y$  are the center of the robot with respect to some global coordinate frame, and  $\theta$  is the vehicle orientation. The vehicle control command  $\mathbf{u}_r = [u_x, u_y, u_\theta]^\top$  indicates the desired positional increments to the vehicle pose in robot local coordinates; with vehicle error dynamics  $\mathbf{v}_r = [v_x, v_y, v_\theta]^\top$ . We can observe how for an error free vehicle, the input  $(u_{x,k}, u_{y,k})$  would drive the robot to the position  $(x_{k+1|k}, y_{k+1|k})$  at time step  $k$ . However, due to unmodeled robot dynamics and noise, it ends up in the position  $(x_{k+1}, y_{k+1})$ . This is, the vehicle process model  $\mathbf{f}_r$  is a noise-corrupted discrete-time nonlinear function of the form

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \\ \theta_{k+1} \end{bmatrix} = \begin{bmatrix} x_k + d_u \cos(\theta_k + \psi_u) + d_v \cos(\theta_k + \psi_v) \\ y_k + d_u \sin(\theta_k + \psi_u) + d_v \sin(\theta_k + \psi_v) \\ \theta_k + u_{\theta,k} + v_{\theta,k} \end{bmatrix} \quad (5.53)$$

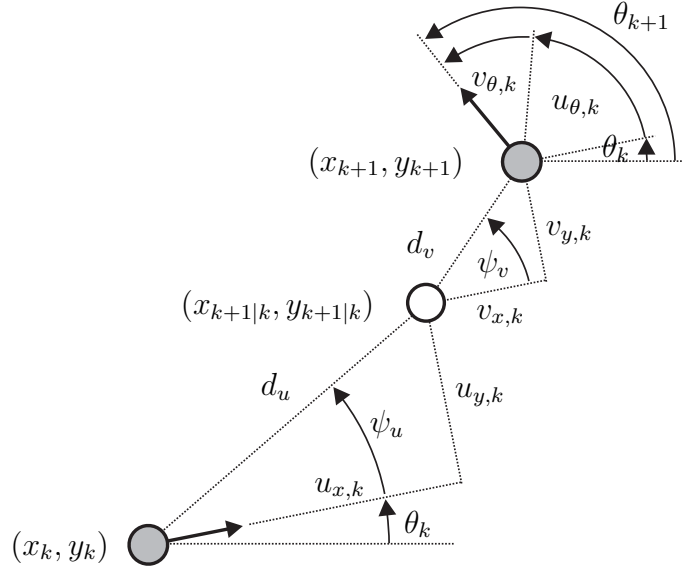


Figure 5.4: Two-dimensional mobile robot motion model.

with

$$d_u = \sqrt{u_{x,k}^2 + u_{y,k}^2} \quad (5.54)$$

$$d_v = \sqrt{v_{x,k}^2 + v_{y,k}^2} \quad (5.55)$$

$$\psi_u = \arctan(u_{y,k}/u_{x,k}) \quad (5.56)$$

and

$$\psi_v = \arctan(v_{y,k}/v_{x,k}) \quad (5.57)$$

Given that the map is considered static, i.e.,

$$\mathbf{f}_f(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k) = \mathbf{x}_f \quad (5.58)$$

and that the landmarks are constituted by two dimensional points  $\mathbf{x}_f^{(i)} = [x_f^{(i)}, y_f^{(i)}]^\top$ ; we are able to formulate the partial derivatives of Equation 5.53 with respect to  $\mathbf{x}$  and  $\mathbf{v}$  for

the state model Jacobians

$$\mathbf{F}_{\mathbf{x}_r} = \begin{bmatrix} 1 & 0 & -d_u \sin(\theta_k + \psi_u) - d_v \sin(\theta_k + \psi_v) \\ 0 & 1 & d_u \cos(\theta_k + \psi_u) + d_v \cos(\theta_k + \psi_v) \\ 0 & 0 & 1 \end{bmatrix} \quad (5.59)$$

$$\mathbf{F}_{\mathbf{v}_r} = \begin{bmatrix} v_{x,k} \cos(\theta_k + \psi_v)/d_v & v_{y,k} \cos(\theta_k + \psi_v)/d_v & 0 \\ v_{x,k} \sin(\theta_k + \psi_v)/d_v & v_{y,k} \sin(\theta_k + \psi_v)/d_v & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5.60)$$

The measurement equation for the  $i$ -th landmark in this configuration is

$$\mathbf{h}^{(i)} = \mathbf{R}^\top (\mathbf{x}_f^{(i)} - \mathbf{t}) + \mathbf{w}^{(i)} \quad (5.61)$$

with  $\mathbf{w}^{(i)}$  the landmark observation error, and

$$\mathbf{R} = \begin{bmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{bmatrix} \quad (5.62)$$

$$\mathbf{t} = [x_k, y_k]^\top \quad (5.63)$$

The  $i$ -th set of rows of the measurement Jacobian in Equation 5.14 for our planar mobile robot are

$$\mathbf{H}_{\mathbf{x}}^{(i)} = \begin{bmatrix} \mathbf{H}_{\mathbf{x}_r}^{(i)} & \mathbf{0}_{2 \times 2(i-1)} & \mathbf{H}_{\mathbf{x}_f}^{(i)} & \mathbf{0}_{2 \times 2(N-i)} \end{bmatrix} \quad (5.64)$$

$$\mathbf{H}_{\mathbf{x}_r}^{(i)} = \begin{bmatrix} -\mathbf{R}^\top & \mathbf{R}^\top (\mathbf{x}_f^{(i)} - \mathbf{t}) \end{bmatrix} \quad (5.65)$$

$$\mathbf{H}_{\mathbf{x}_f}^{(i)} = \mathbf{R}^\top \quad (5.66)$$

$$\mathbf{H}_{\mathbf{w}^{(i)}} = \mathbf{I} \quad (5.67)$$

As a motivational note to the reader, this detailed formulation of the kinematics of the *monobot* and the *planarbot* are to be used in Chapter 6 in which we show how the limitations inherent in the reconstructibility of the map state space are equally present in both vehicle models.

### 5.3 Model assumptions

In this section we argument on the various assumptions made during the simulations and the real world implementations of the EKF-CML-LV algorithm. The discussion includes a description of path planning, as well as details on the plant and measurement

noise models. We discuss the required formulas for our planar mobile robot. The simplification of them to the case of the *monobot* is trivial and shall not be commented further.

### 5.3.1 Path planning

Starting from an initial position  $\mathbf{x}_{r,0}$ , we generate the path plan  $X^m = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ , where each desired robot pose at time step  $k$  is given by the triple  $\mathbf{x}_k = [x_k, y_k, \theta_k]^\top$ . The coordinates  $(x_k, y_k)$  indicate the location of the robot in a fixed reference frame, and  $\theta_k$  gives the desired robot orientation with respect to the aforementioned coordinate frame. The vehicle is to move along this path in  $m$  steps; and if the robot dynamics were error free, the state  $\mathbf{x}_{r,k}$  would coincide exactly with the path element  $\mathbf{x}_k$ . Figure 5.5 shows a simulated winding path  $X^m$  for various values of  $m$ .

Next, we generate a set of motion commands  $U^m = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ , that under ideal circumstances would drive the mobile robot to follow this path. Solving for  $\mathbf{u}_k$  in Equation 5.53, and assuming error-free vehicle dynamics we get

$$\mathbf{u}_k = \begin{bmatrix} d_u \cos \psi_u \\ d_u \sin \psi_u \\ \theta_{k+1} - \theta_k \end{bmatrix} \quad (5.68)$$

with

$$d_u = \sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2} \quad (5.69)$$

$$\psi_u = \arctan\left(\frac{y_{k+1} - y_k}{x_{k+1} - x_k}\right) - \theta_k \quad (5.70)$$

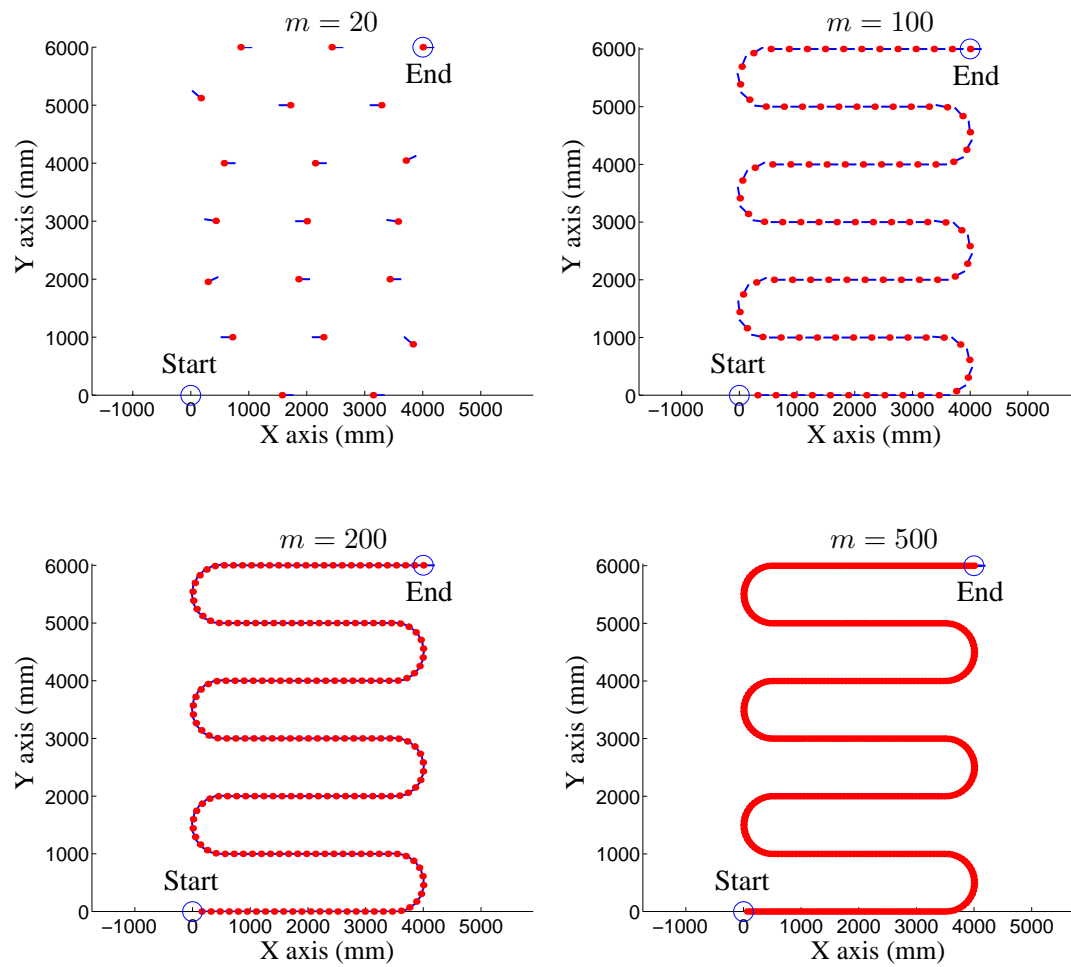


Figure 5.5: Winding path  $X^m$  for various values of  $m$ .

### 5.3.2 Dead reckoning error

Unfortunately, as the robot moves, drift errors due to wheel slippage and encoder quantization will make it diverge from the planned path  $X^m$ . The type of motion that relies solely on encoder readings is called *dead reckoning*, and as indicated in Section 5.1, we model the divergence of a dead reckoning motion from the planned path with zero-mean white Gaussian plant noise.

After extensive experimentation with our mobile robot platforms, we have observed that typical motion commands are divided in three different steps, and that the error is proportional to these steps. One for the direction of motion change, followed by one proportional to the length of the distance traveled, and one proportional to the final heading adjustment. Given the motion model from Figure 5.4, and the directional and angular dead reckoning factors  $\alpha_\psi$ ,  $\alpha_d$  and  $\alpha_\theta$ , the standard deviation in polar coordinates of the vehicle motion error is estimated with

$$\sigma_\psi = \alpha_\psi \psi_u \quad (5.71)$$

$$\sigma_d = \alpha_d d_u \quad (5.72)$$

$$\sigma_\theta = \alpha_\theta u_\theta \quad (5.73)$$

The dead reckoning factors  $\alpha_\psi$ ,  $\alpha_d$ , and  $\alpha_\theta$  serve as indicators of the reliability of using dead reckoning for navigation. Consequently, the robot error dynamics can be simulated by drawing normally distributed noise samples with polar coordinates

$$\psi_v = N(0, \sigma_\psi) \quad (5.74)$$

$$d_v = N(0, \sigma_d) \quad (5.75)$$

$$v_\theta = N(0, \sigma_\theta) \quad (5.76)$$

The noise dynamics in Cartesian coordinates can be computed with

$$[d_v \cos(\theta_k + \psi_v), d_v \sin(\theta_k + \psi_v), v_\theta]^\top \quad (5.77)$$

The nonlinear transformation from polar to Cartesian coordinates destroys the normal properties of the noise distribution. The best we can do is approximate its covariance with

$$\mathbf{R}_u = \begin{bmatrix} \cos(\theta_k + \psi_u) & -\sin(\theta_k + \psi_u) & 0 \\ \sin(\theta_k + \psi_u) & \cos(\theta_k + \psi_u) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5.78)$$

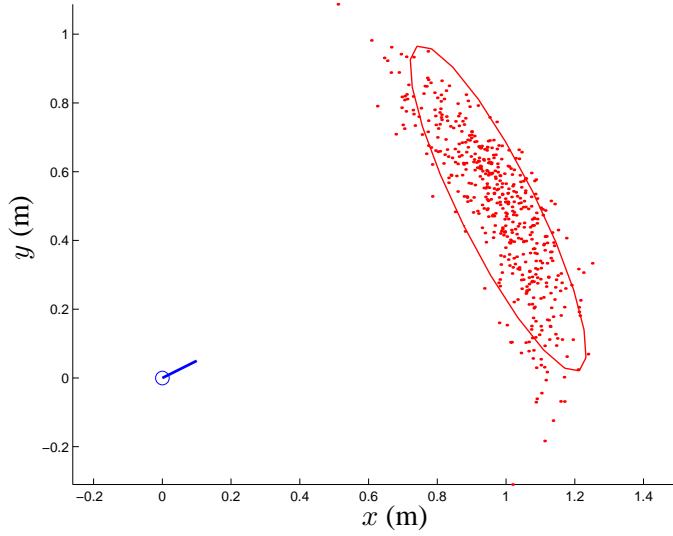


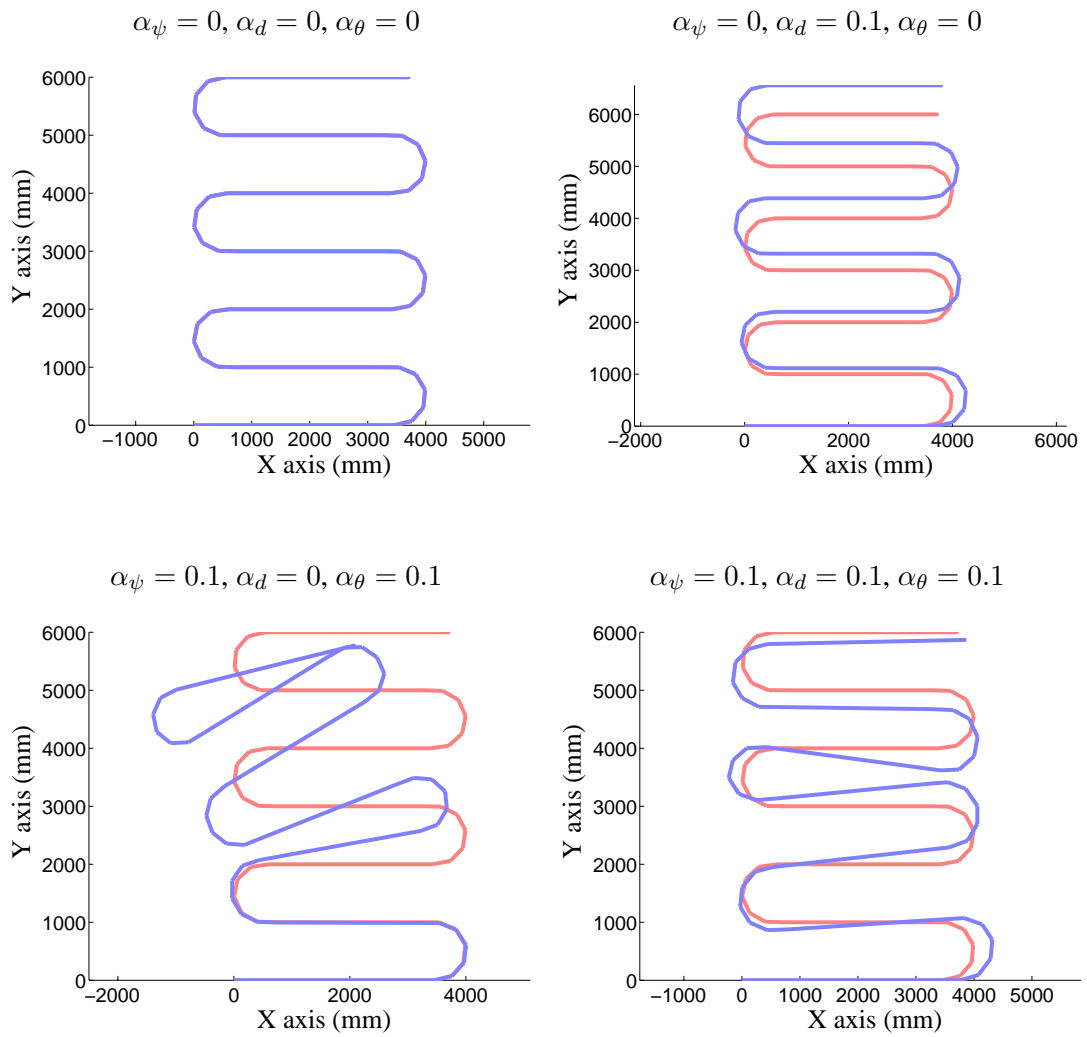
Figure 5.6: Motion noise model. Motion noise regarded as white Gaussian. The ellipse indicates a 95% confidence bound around the a priori motion estimate  $\mathbf{x}_{r,k+1|k}$ .

$$\mathbf{V}_{r,k} = \mathbf{R}_u \begin{bmatrix} \sigma_d^2 & & \\ & (d_u \tan \sigma_\psi)^2 & \\ & & \sigma_\theta^2 \end{bmatrix} \mathbf{R}_u^\top \quad (5.79)$$

Figure 5.6 shows the non Gaussian characteristics of the distribution of 100 noisy motion commands with mean  $\mathbf{u}_k = [1\text{m}, 0.5\text{m}, 0\text{rad}]^\top$ , from an initial pose estimate at  $\mathbf{x}_{r,k|k} = [0\text{m}, 0\text{m}, 0.46\text{rad}]^\top$ , and with  $\alpha_\psi = 0.5$ ,  $\alpha_d = 0.05$ , and  $\alpha_\theta = 0$ . The ellipse represents a  $2\sigma$  level surface of the covariance approximation  $\mathbf{V}_{r,k}$ , computed from Equation 5.79.

Figure 5.7 shows how our simulated planar mobile robot deviates from the desired path  $X^n$  for several values of  $\alpha_\psi$ ,  $\alpha_d$  and  $\alpha_\theta$  over a path plan with 100 steps. The value of the dead reckoning factors depend on the vehicle mechanics and on the characteristics of the terrain surface. The top left plot indicates the effects of having accurate dead reckoning, whereas the rest of the plots show an increase on dead reckoning error due to unreliable encoders or poor soil characteristics. It can be seen in the bottom right plot for example, the deviation in the robot trajectory produced for directional and angular dead reckoning factors that account for as much as 10% of the motion command.



Figure 5.7: Dead reckoning error for various values of  $\alpha_\psi, \alpha_d, \alpha_\theta$ .

### 5.3.3 Observation error

To simulate the observation model dynamics we place  $n$  artificial landmarks in the environment,  $\mathbf{x}_f^{(i)} = [x_f^{(i)}, y_f^{(i)}]^\top$ ,  $i = \{1, \dots, n\}$ . Their location is set randomly obeying a uniform distribution and limited by the robot workspace.

Imagine we had a noise free robot with a noise free sensor. Imagine also that this sensor was capable of observing all landmarks at each time instance. The value of an exact observation for any given landmark taken from this ideal robot could be computed simply by evaluating Equation 5.61 with the hypothetical vehicle states  $\mathbf{x}_k$  from the path plan  $X^n$ , and with zero measurement error:

$$\begin{bmatrix} z_{x,k}^{(i)} \\ z_{y,k}^{(i)} \end{bmatrix} = \begin{bmatrix} \cos(\theta_k) & \sin(\theta_k) \\ -\sin(\theta_k) & \cos(\theta_k) \end{bmatrix} \begin{bmatrix} x_f^{(i)} - x_k \\ y_f^{(i)} - y_k \end{bmatrix} \quad (5.80)$$

Unfortunately, sensors are also prone to error dynamics. And as it turns out, for most sensors used in mobile robot localization (vision, laser, sonar), the measurement error is proportional to the distance and view angle from the robot to the actual landmark location. The further away from the vehicle a landmark is, the larger the error in its observation the sensor will produce. The standard deviation of the measurement error can be computed in exactly the same way as with the motion error model. To distinguish the sensor error factors from the ones for dead reckoning, we use the nomenclature  $\beta_d$  for directional error factor and  $\beta_\psi$  for the angular error factor.

If vision sensors are used for example,  $\beta_d$  would account for the errors in the computation of the disparity in stereo correspondence, which in turn translates to error in the landmark depth estimate. On the other hand,  $\beta_\psi$  would represent camera lens radial distortion effects, and would translate to deviations along a plane perpendicular to the sensor view axis. Such error model is no more than a linear simplification of the perspective projection effect, in which image disparity is inversely proportional to depth and radial distortion involves quadratic terms (see Appendix B). Nevertheless, the approximation is still reasonable, as long as adequate noise levels are used.

The measurement error produced by a sonar range sensor would have a larger value for  $\beta_\psi$  than its laser counterpart; for which for practical purposes angular error can be safely neglected.

Measurement errors can thus be simulated with

$$\sigma_d = \beta_d d_z \quad (5.81)$$

$$\sigma_\psi = \beta_\psi \psi_z \quad (5.82)$$

$$d_w = N(0, \sigma_d) \quad (5.83)$$

$$\psi_w = N(0, \sigma_\psi) \quad (5.84)$$

The simulated measurement in Cartesian coordinates is thus

$$[(d_z + d_w) \cos(\psi_z + \psi_w), (d_z + d_w) \sin(\psi_z + \psi_w)]^T \quad (5.85)$$

Once more, the nonlinear transformation from polar to Cartesian coordinates of the sensor dynamics destroys the normal properties of the noise distribution. The best we can do is approximate its covariance with

$$\mathbf{W}_k = \begin{bmatrix} \cos \psi_z & -\sin \psi_z \\ \sin \psi_z & \cos \psi_z \end{bmatrix} \begin{bmatrix} \sigma_d^2 & \\ & (d_z \tan \sigma_\psi)^2 \end{bmatrix} \begin{bmatrix} \cos \psi_z & \sin \psi_z \\ -\sin \psi_z & \cos \psi_z \end{bmatrix} \quad (5.86)$$

Figure 5.8 shows the distribution of 100 noisy observations for a sensor with pose and orientation  $[0, 0, 0.46]^T$ , and a landmark at  $[1, 0.5, 0]^T$  with  $\beta_d = 0.1$  and  $\beta_\psi = 0.1$ . The ellipse represents a  $2\sigma$  level surface of the covariance approximation  $\mathbf{W}_k$ .

Figure 5.9 shows a simulation of the error produced in the observation for various sets of ten landmarks each, for a sensor with different values of  $\beta_d$  and  $\beta_\psi$ . The top left plot shows an error free observation model, whereas the bottom right plot shows a set of landmark observations that deviate on average 5% the distance between the landmarks and the vehicle in polar coordinates. The large circles are centered on the actual landmark locations, and the dots indicate the observations generated with the aforementioned measurement model. Note how the measurements for the landmarks located in the outer regions of the environment present larger deviations from their true location than those found towards the center of the environment. This artifact is caused by the observation error model chosen; one that is proportional to the distance from the robot to the landmarks.

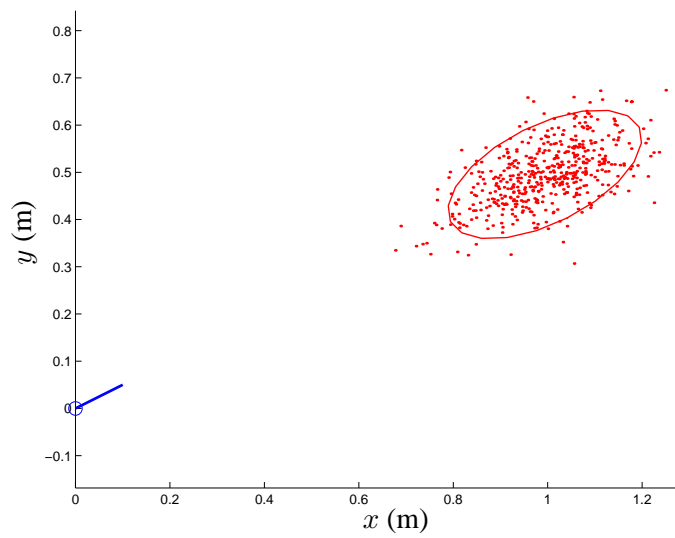


Figure 5.8: Measurement noise model. Motion noise regarded as white Gaussian. The ellipse indicates a 95% confidence bound around the landmark estimate  $\mathbf{x}_{f,k+1|k}$ .

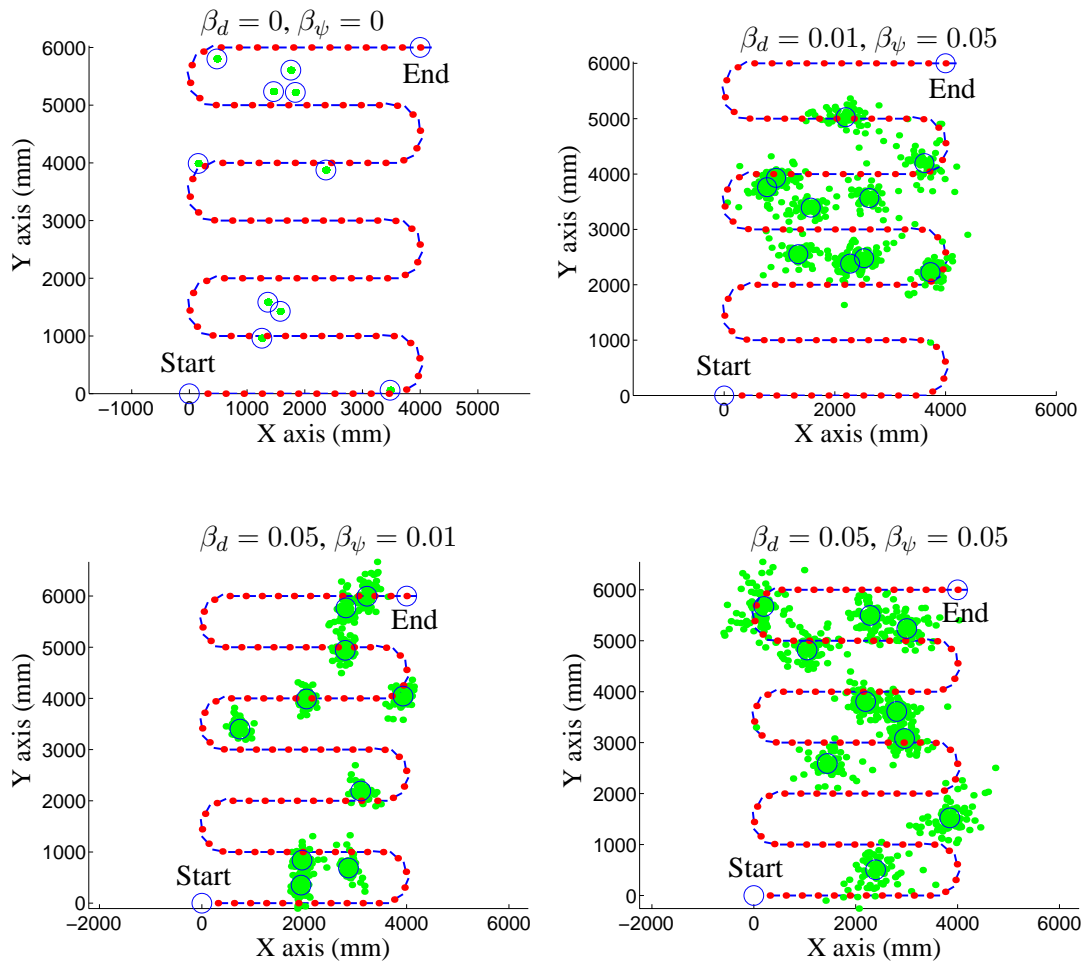


Figure 5.9: Measurement error distribution for various values of  $\beta_d$  and  $\beta_\psi$ .

## 5.4 Temporal landmark validation

Now that we have underpinned the fundamental characteristics the original EKF-CML algorithm, as well as the specifics of the mobile robot platforms used, we turn our attention to the study of appropriate temporal landmark quality functions that help in the alleviation of the data association problem.

We reviewed in Sections 5.1.8-5.1.10 the *spatial* landmark quality tests already present in the literature [150]. We now introduce, in Sections 5.4.1-5.4.4 two new *temporal* landmark quality models, and show the feasibility of using these in a test to validate not only where, but when should we expect future observations for any given landmark.

### 5.4.1 Landmark temporal uncertainty

As indicated in the introductory section to this Chapter, the use of spatial compatibility tests is crucial for the solution of data association in CML, but they can still be insufficient in situations with moderate scene dynamics. Imagine that a spurious landmark observation was sufficiently robust to be added to the map. This landmark could come from a temporary artifact in the scene such as an open door, or an artifact in our sensor such as a shadow or reflection that is persistent only during a small number of iterations.

With the static map model presented so far, once the spurious landmark disappears from the scene, or at least, from the sensor return, we have no means to revise the map and delete it. The two spatial compatibility tests presented would not have any information on the history of observations, and would be trying to associate new observations to that entry in the map. If the algorithm succeeds at incorrectly associating new landmark observations with such previously stored weak feature, i.e., the observation fell inside the uncertainty ellipse described by  $S$  and passed the data association test; it will most likely introduce landmark localization error in the map, while at the same time decreasing the estimated map error covariance. Given that the map covariance is fully correlated, starting with the next iteration of the algorithm, that uncertainty would be propagated to the rest of the landmark locations, and that of the robot as well; ultimately breaking down the entire estimation approach to CML.

For moderately dynamic environments we need some means to erase landmarks from the map once we have realized they are not sufficiently robust, not only during landmark initialization, but during the entire run of the algorithm. To aid in those situations in which the landmark observations might not be consistent in time, we propose a new set of landmark quality functions. The idea is to use these functions as an aid to validate temporal landmark quality.

The first temporal landmark quality function proposed is an exponential decay rule used to learn the persistence of landmark matching. The second function is a linear model used to update the probability of data association from the sequence of landmark matches. These two functions are used to validate landmark observations in time, and ultimately to erase from the map those landmarks that are not sufficiently persistence in the scene. In Section 5.4.4 we analyze the consequences of removing landmarks from the map with respect to the evolution of the fully correlated covariance entries in  $\mathbf{P}$ .

### 5.4.2 Nonlinear model for temporal landmark quality: the exponential decay rule

One possibility in the computation of the temporal landmark quality is to have an exponential decay rule. This way, each landmark in the map will have an associated memory cell that registers how persistent, and how old that landmark is.

Imagine that at the  $k + 1$ -th iteration, the  $i$ -th landmark measurement estimate  $\mathbf{z}_{k+1|k}^{(i)}$  falls inside the current field of view, but none of the entries in the observation vector  $\mathbf{z}_{k+1}$  has similar appearance properties, nor is sufficiently close (in the sense of  $d^2$ ) to pass the spatial landmark compatibility tests. This would be the situation if, for example, the  $i$ -th landmark was learned from a temporary artifact in the scene that was only tracked over sensor data for a short period of time, but is no longer present. With the aid of an exponential decay rule to data association, its quality measure will decay in the absence of observation matches, indicating the map building algorithm that such landmark is no longer present in the scene and should not be considered a relevant feature for robot localization.

We propose a nonlinear update rule for landmark quality of the form

$$x_{q,k+1}^{(i)} = \frac{1}{1 + e^{-\left(\alpha u_{q,k}^{(i)} + \beta x_{q,k}^{(i)}\right)}} \quad (5.87)$$

where  $u_{q,k}^{(i)}$  is the landmark identification stamp

$$u_{q,k}^{(i)} = \begin{cases} 0 & : \text{ failed the spatial data association test} \\ 1 & : \text{ passed the spatial data association test} \end{cases} \quad (5.88)$$

The scalar  $\alpha$  is an input weight used to regulate the contribution of such landmark identification over the previous map configuration, and  $\beta$  is a memory weight used to regulate the contribution of the previous landmark quality state over its new value.

The asymptotic lower and upper bounds of Equation 5.87 can be evaluated by solving the equations

$$x_{q,LOW} = \frac{1}{1 + e^{-(\beta x_{q,LOW})}} \quad (5.89)$$

$$x_{q,HIGH} = \frac{1}{1 + e^{-(1+\beta x_{q,HIGH})}} \quad (5.90)$$

Using a symbolic manipulation math package, we find for example, that for  $\alpha = \beta = 1$ ,  $x_{q,LOW} = 0.6590$ , and  $x_{q,HIGH} = 0.8659$ . Landmark initialization is at the middle of the scale, i.e,  $x_{q,0}^{(i)} = 0.7682$ .

### 5.4.3 Linear model for temporal landmark quality: the data association probability

Another possibility in the computation of the temporal landmark quality is to consider the probability of correct data association of such landmark in the next iteration.

According to the relative frequency definition of probability, if an event (say, the correct association of landmark  $i$ ) occurs  $j$  times in  $k$  trials (observations), and provided  $k$  is sufficiently large, then the probability that the same landmark will be properly matched in the next iteration can be expressed as

$$p_k^{(i)} = \frac{j}{k} \quad (5.91)$$

Now, once a new observation is made, the data association probability will change according to the new landmark association result. This change in probability is represented by the recursive expression

$$p_{k+1}^{(i)} = \frac{p_k^{(i)}k + u_{q,k}^{(i)}}{k+1} \quad (5.92)$$

with  $u_{q,k}^{(i)}$  defined as in Equation 5.88. If we make the notation change  $a = k/(k+1)$ , and  $x_{q,k}^{(i)} = p_k^{(i)}$ , our second model for temporal landmark quality becomes

$$x_{q,k+1}^{(i)} = ax_{q,k+1}^{(i)} + (1-a)u_{q,k}^{(i)} \quad (5.93)$$

For fixed values of  $k$ , the constant  $a$  accounts for a memory weight with a role



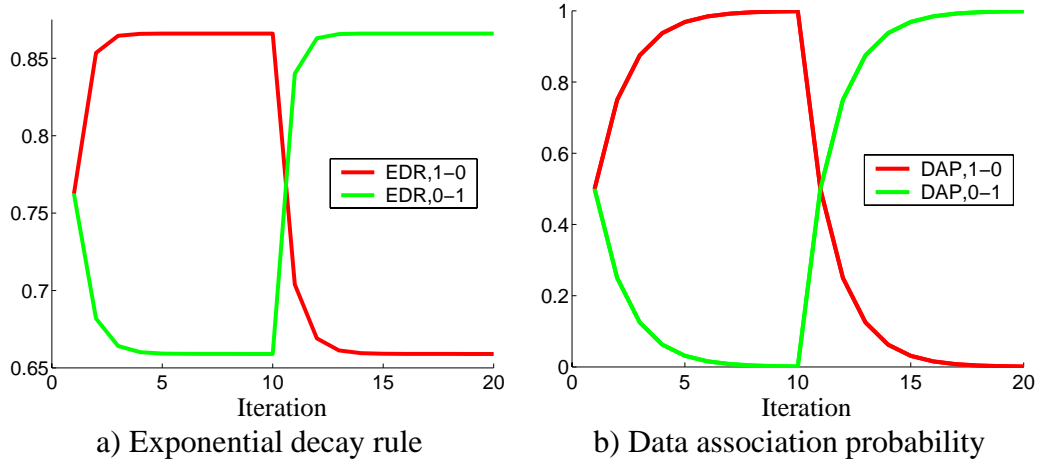


Figure 5.10: Landmark quality models.

similar as those of  $\alpha$  and  $\beta$  from the previously discussed model. It can be fixed to a constant value between 0 and 1, and it indicates the memory length to be used in the computation of the new data association probability. So for example, if a memory window of the last 5 iterations is to be considered, the memory weight becomes  $a = 0.8333$ . In this linear model for temporal landmark quality,  $x_q$  is bounded between 0 and 1, and initialization is made at 0.5.

Both temporal landmark quality measures, the exponential decay rule, and the data association probability were chosen to have similar properties. For instance, to be asymptotically bounded by above and below by

$$x_{q,LOW} \leq x_q \leq x_{q,HIGH} \quad (5.94)$$

Any other function with such monotonicity could be also used as temporal landmark quality function. However, such function must have some way of tuning the memory length of the algorithm. The two functions proposed do so by means of the parameters  $\alpha$ ,  $\beta$ , and  $a$ . The left plot in Figure 5.10 shows the behavior of the exponential decay rule for the parameter values  $\alpha = 1$ , and  $\beta = 1$ . Similarly, the right plot shows the data association probability with parameter  $a = 0.5$ . The labels 0-1 and 1-0 indicate the change in the landmark identification stamp  $\mathbf{u}_q$  from 0 to 1, and from 1 to 0, representing the presence or loss of data association.

### 5.4.4 Temporal landmark quality test

In the same way that the spatial compatibility test is used to validate if observations are consistent with the already learned map entries; the temporal landmark quality test must be used to validate if any map entry is sufficiently robust to be kept in the map. The test verifies if the history of data association has kept the value for the temporal landmark quality above a user defined cut threshold  $x_{q,THLD}$ . All landmarks expected to appear in the current field of view, and for which no occlusion has been predicted, must have their landmark quality measure updated. Furthermore, those landmarks whose temporal quality measure falls below the user defined threshold should be removed from the map. The heuristics needed to handle occlusions, depend on the type of landmarks and sensors used.

The temporal landmark quality test is

$$\mathbf{if} \quad x_q^{(i)} \leq x_{q,THLD} \quad \text{RemoveLandmark}(\mathbf{x}_f^{(i)}) \quad (5.95)$$

In the case of the exponential decay rule with parameters  $\alpha = 1$  and  $\beta = 1$ , for example, the cut threshold  $x_{q,THLD} = 0.66$  is reached once a landmark has not been observed for 5 consecutive iterations, or more if these were not consecutive. Similar effects are obtained when using the data association probability with the parameter value  $a = 0.5$ , and a cut threshold of  $x_{q,THLD} = 0.03$ . Figure 5.11 shows both the exponential decay rule, and the data association probability as landmark quality measures for a test run of 100 steps and 10 landmarks, with slightly modified simulation parameters than the ones used to create Figure 5.2b. In this test run, we have forced 25% of the observations to be misidentified to their closest neighbor. The individual compatibility test catches some but not all of these mismatches, and yields an identification stamp value  $\mathbf{u}_q = 0$  for them. By adding the more restrictive temporal landmark quality tests, those landmarks with a large amount of mismatches end up being removed from the map, and are reinitialized as new landmarks once they become robust again.

We have opted for a simplified heuristic for the removal of a landmark from the map, with the advantage of computational efficiency, but at the expense of suboptimality. Our algorithm simply erases the low quality landmark entries from the state vector and its corresponding row and column in the state error covariance matrix. Once the landmark is robust again, it is considered as a new different landmark, and initialized according to the discussion from Section 5.1.7. Note however that, given the fact that CML is fully correlated, the contribution of a misidentified landmark estimate in revising the error covariance matrix has already propagated to the entire map. The right thing to do, would be to trace back the intermediate results of the algorithm up to the point in

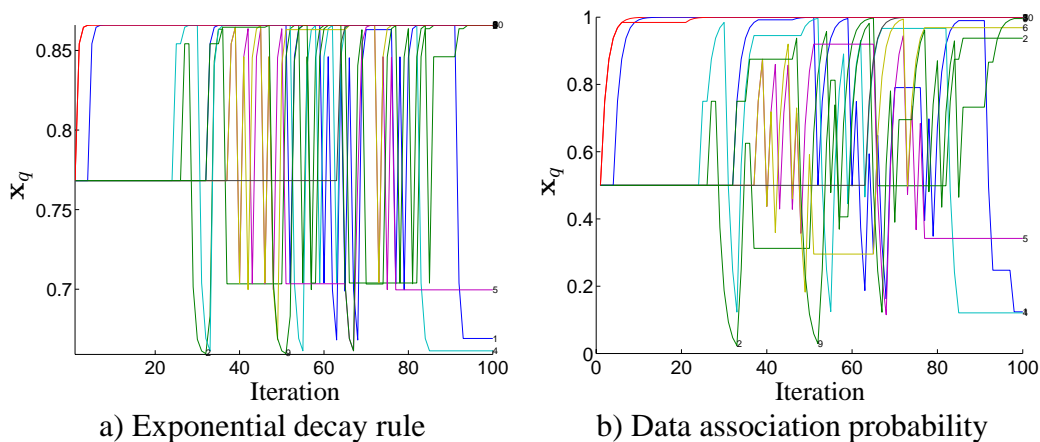


Figure 5.11: Landmark quality test for a test run with 10 landmarks and 100 steps.  $\alpha = 1$ ,  $\beta = 1$ , and  $a = 0.5$ . The plots show landmarks labeled 2 and 9 being reinitialized near steps 30 and 50 respectively.

which the landmark was originally inserted, and to recompute forward once more the map state and map error covariance up to the current iteration, without considering that landmark, as if it had never existed.

Saving the state vector and error covariance matrix for all iterations has space complexity of  $O(kn^2)$ , with  $k$  the number of iterations, and  $n$  the number of landmarks. Furthermore, recomputing the state vector and error covariance matrix from the point at which the spurious landmark was initially inserted, would most likely lead to different values on  $\mathbf{P}$ , and consequently on  $\mathbf{S}$ , producing even different data association results. So, not only the state vector and covariance matrix history must be maintained, but the full measurement data as well, requiring for a full run of the algorithm every time a landmark is found to be spurious. The optimal solution is rather cumbersome, and we have opted for suboptimality, with the aforementioned simplification of just deleting the corresponding entries in  $\mathbf{x}$  and  $\mathbf{P}$ , with the following insight.

Gibbens *et. al.* [83], show how in CML all entries in the covariance matrix  $\mathbf{P}$  depend on the number of landmarks used in the form of the total Fisher information  $I_T$ . This is a measure of the total information per unit time available to the filter. For a monobot with  $n$  landmarks, all with equal measurement covariance  $\mathbf{W}^{(i)} = w$ , the total Fisher information is  $I_T = n/w$ . The more landmarks available, the more information the filter has. That is, the greater the number of landmarks used, the smaller the asymptotic values for the entries in the error covariances.

The removal of a landmark from the state vector in the form discussed is consistent with this observation. We next show, by example, how for the monobot, all the entries in  $\mathbf{P}$  with the removal of a landmark at some point in the algorithm are bounded by below and above by the same entries in  $\mathbf{P}$ , but with and without considering the landmark for the entire run. Let us call  $\mathbf{P}_{\cdot,n}$  the entry in  $\mathbf{P}$  for a map with  $n$  landmarks, and  $\mathbf{P}_{\cdot,n+1,n}$  the entry in  $\mathbf{P}$  for a map that went from  $n + 1$  to  $n$  landmarks via the removal of landmark states. Then, for the entire run of the algorithm

$$\mathbf{P}_{rr,k|k,n+1} \leq \mathbf{P}_{rr,k|k,n+1,n} \leq \mathbf{P}_{rr,k|k,n} \quad (5.96)$$

$$\mathbf{P}_{rf,k|k,n+1}^{(i)} \leq \mathbf{P}_{rf,k|k,n+1,n}^{(i)} \leq \mathbf{P}_{rf,k|k,n}^{(i)} \quad (5.97)$$

$$\mathbf{P}_{ff,k|k,n+1}^{(i,j)} \leq \mathbf{P}_{ff,k|k,n+1,n}^{(i,j)} \leq \mathbf{P}_{ff,k|k,n}^{(i,j)} \quad (5.98)$$

Moreover, by removing a landmark from the map in the form discussed, the asymptotic convergence property from Equation 5.26 is maintained. That is, the revised map is still fully correlated. Figure 5.12 shows the evolution of the entries in the covariance matrix for a monobot with two and three landmarks and noise parameters  $\mathbf{V} = 1$ ,  $\mathbf{W}^{(i)} = 1$ ,  $\mathbf{P}_{rr,0|0} = 1$ . Figure 5.13 shows the evolution of the determinant of the map error covariance for the same monobot configuration, an indication of asymptotic full correlation.

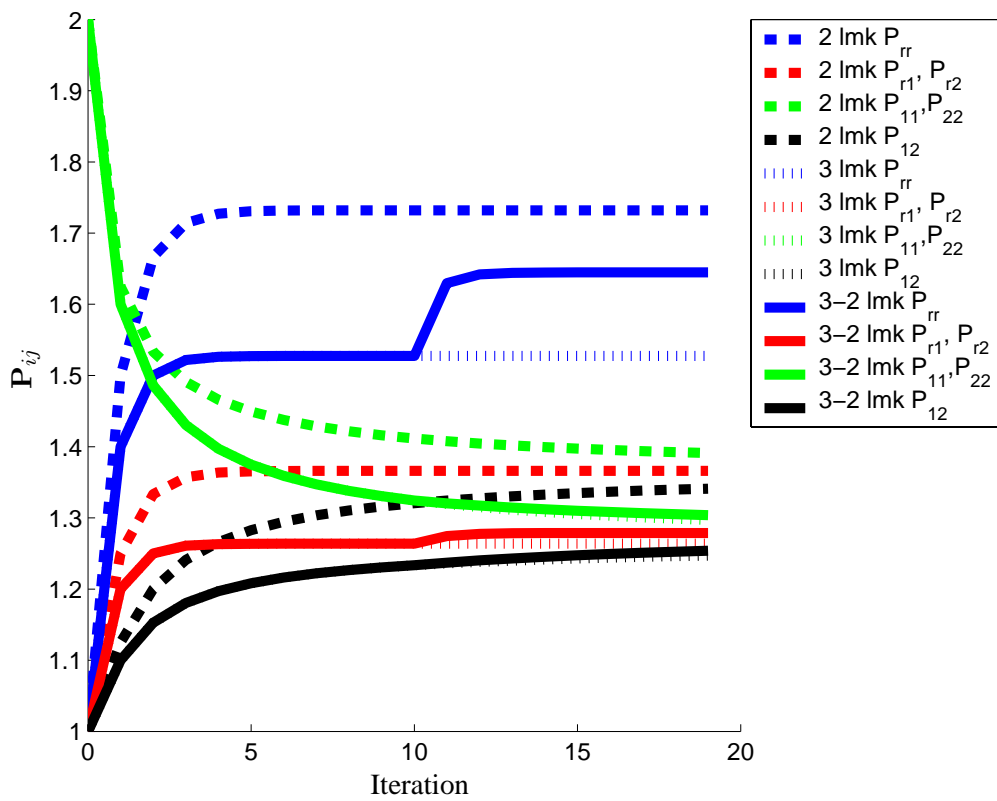


Figure 5.12: Evolution of the entries in the covariance matrix for a monobot with 3, 2, and 3-then-2 landmarks.

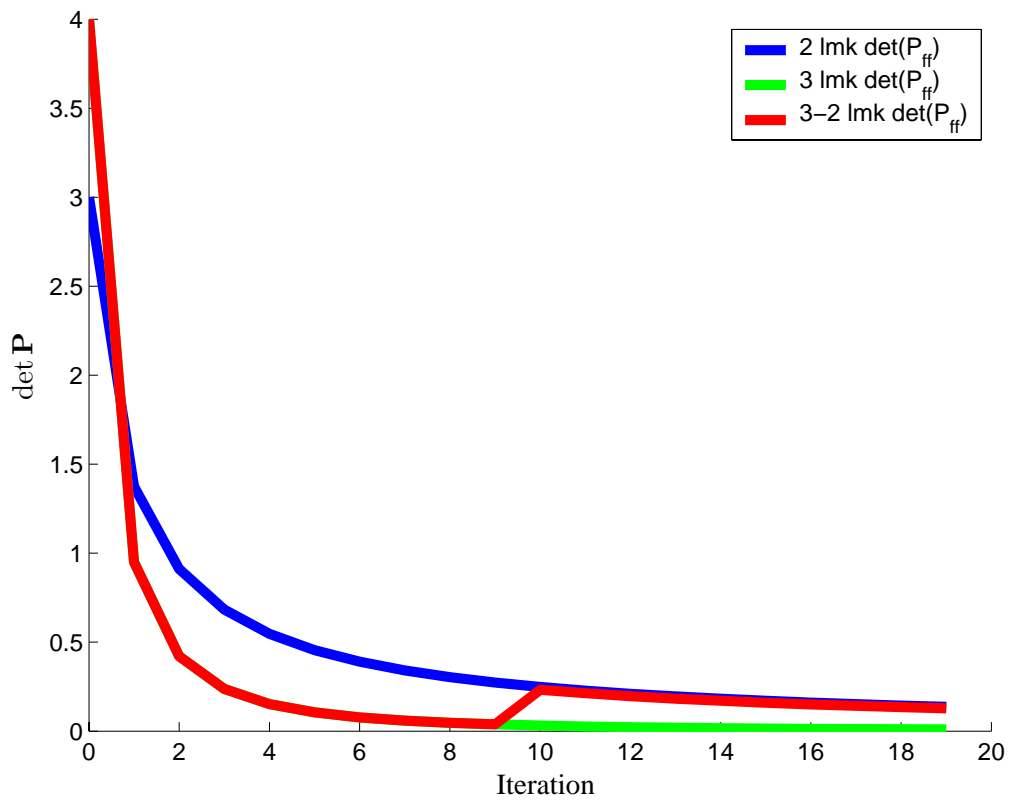


Figure 5.13: Asymptotic convergence of the determinant of the map error covariance matrix for a monobot with 3, 2, and 3-then-2 landmarks.

## 5.5 Performance of the CML-LV algorithm

Having underpinned the characteristics and model assumptions of the full covariance EKF-CML and EKF-CML-LV algorithms (LV stands for landmark validation), we present now results on the improvement in the reconstruction achieved with the modifications presented in this chapter. Namely, we will compare the original algorithm with the modifications induced by sequential innovation and observation range limitations. We will then concentrate on the more realistic limitations produced by erroneous data association, which is considered one of the most critical artifacts that might destroy the viability of the EKF-based method to concurrent localization and map building. This is, when one or more observed landmarks are misidentified, their location estimate might diverge considerably, consequently inducing large errors in the localization of the robot as well. We will show how this problem is alleviated by using both temporal and spatial landmark quality tests to verify the correctness of data association.

### 5.5.1 Sequential innovation

From this point on we will consider as our standard test case, and unless otherwise indicated, a planar robot with 3 dof traversing an environment with 10 landmarks in 100 steps, and with the plant and measurement noise parameters  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , and  $\beta_\psi = \beta_d = 0.01$ . Figure 5.14 shows the results of applying the full-covariance EKF-CML algorithm, and the computationally more efficient version with sequential innovation. In the first two plots in the figure, the blue line indicates the actual robot displacement, and underneath it, is a green version, representing the estimated trajectory. The small green marks along the trajectory are projected ellipses of uncertainty at a distance of  $2\sigma$ . Furthermore, the sparse light red dots indicate landmark measurements, and the green dots indicate their location estimate as computed by the localization algorithm.

As the reader can appreciate in the error plots, both implementations are practically equivalent, the SI implementation being slightly better. The actual difference in the localization of the mobile robot for both implementations is shown in Figure 5.15, with a variation in the localization estimate by  $1.68mm$  on average, for a run of nearly  $40m$ , a deviation of less than  $0.0042\%$ , and for all practical purposes negligible. Given that the two methods are theoretically equivalent, this small divergence is attributed to floating point computations and quantization, and to the better model estimates obtained when re-computing the model Jacobians after each measurement is assimilated by the filter. In order to appreciate the contribution in localization error from each degree of freedom of the vehicle, Figure 5.16 shows in red the components of robot localization error in the EKF-CML-SI run, with  $2\sigma$  bounds signaled as blue dotted lines.

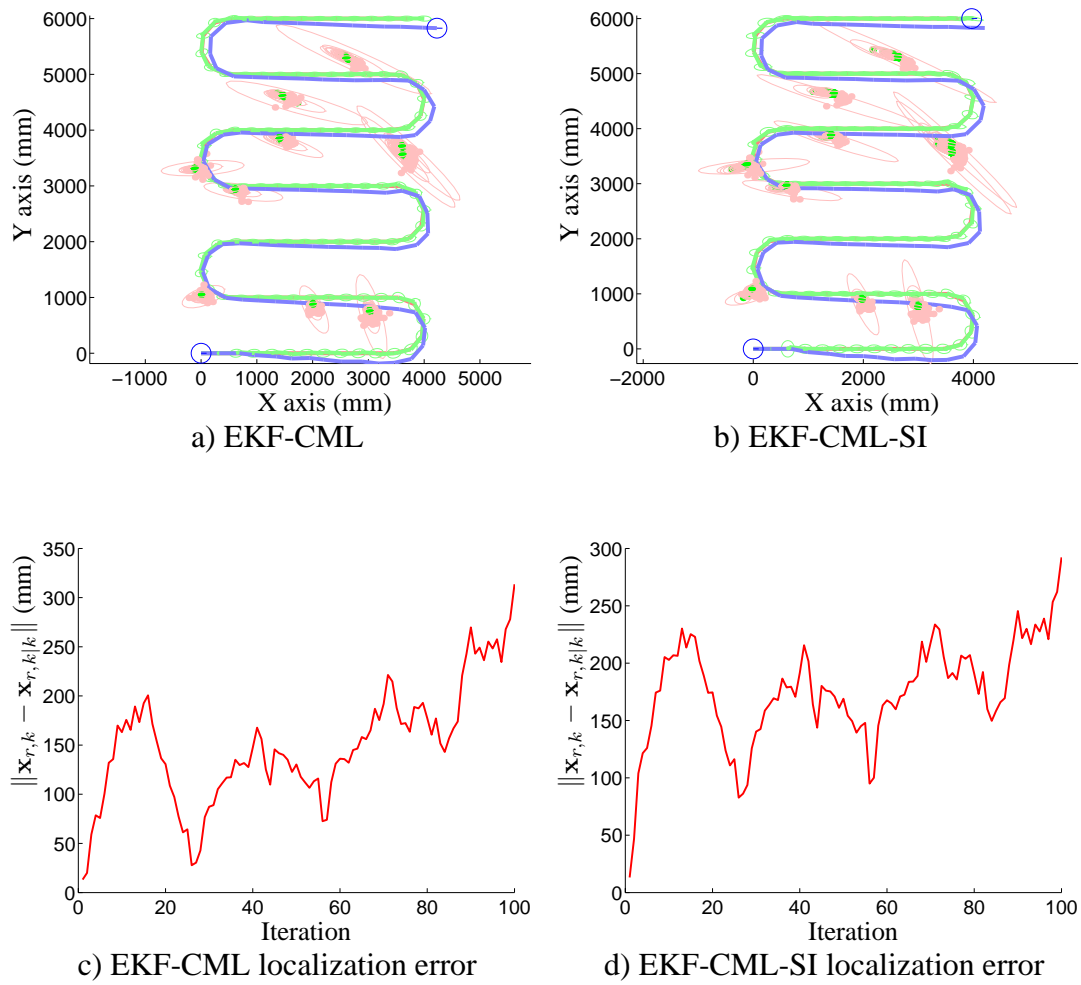


Figure 5.14: Full-covariance EKF CML for a path with 100 iterations and 10 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ .



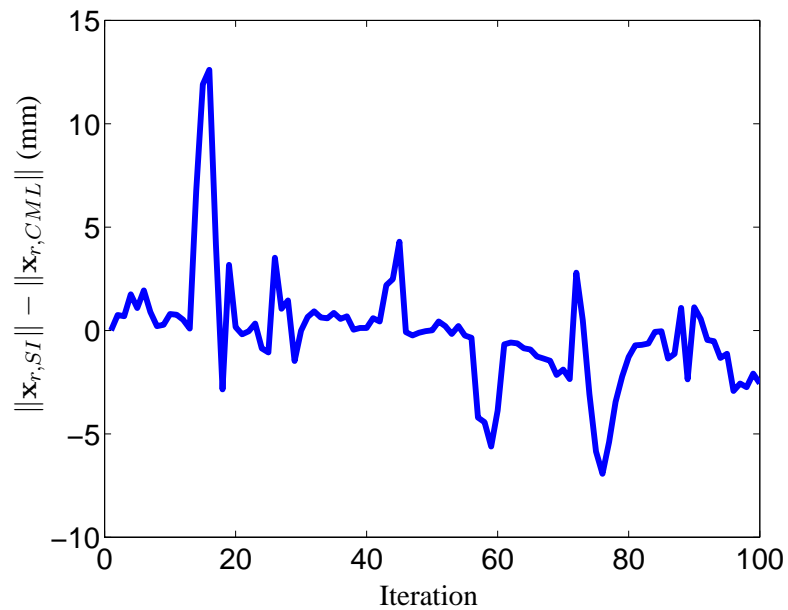


Figure 5.15: Difference in localization error between the full covariance EKF version of the algorithm, and the results of applying sequential innovation to the same set of observations.

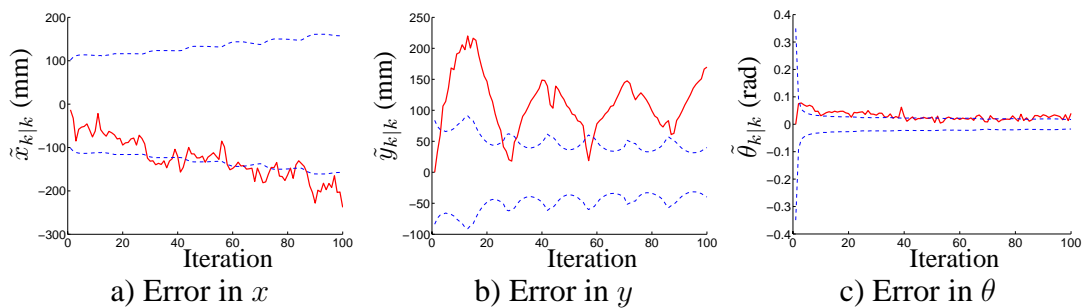


Figure 5.16: Robot localization error estimate. EKF CML with sequential innovation for a path with 100 iterations and 10 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ .

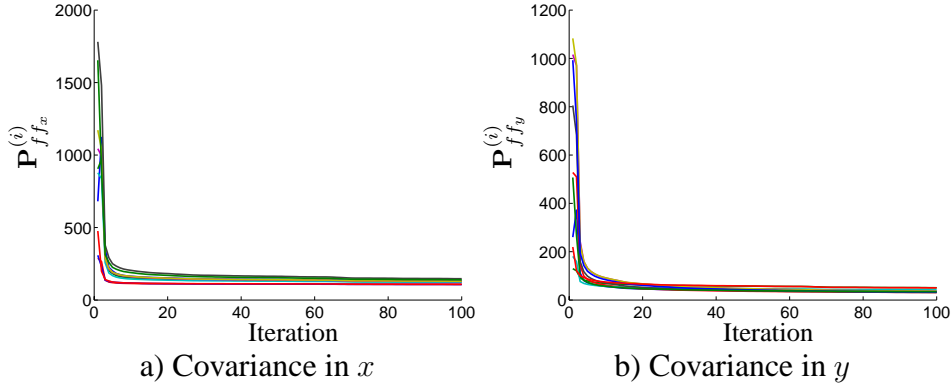


Figure 5.17: The landmark noise covariance estimate is monotonically decreasing in the EKF-CML algorithm.

Figures 5.17 and 5.18 help visualize the behavior of the entries in the error covariance matrix for the EKF-CML-SI case. Figure 5.17 shows for example how the uncertainty in the location of the landmarks reduces monotonically, with each line representing the landmark noise covariance estimate in both the  $x$  and  $y$  axes.

One direct consequence from Equation 5.26 is that the correlation coefficients of the map covariance  $\mathbf{P}_{ff}$  are monotonically increasing. So at each iteration, these correlation coefficients can be evaluated with

$$\rho_{x,k}^{(ij)} = \frac{\mathbf{P}_{ffx,k|k}^{(ij)}}{\sqrt{\mathbf{P}_{ffx,k|k}^{(ii)} \mathbf{P}_{ffx,k|k}^{(jj)}}} \quad (5.99)$$

$$\rho_{y,k}^{(ij)} = \frac{\mathbf{P}_{ffy,k|k}^{(ij)}}{\sqrt{\mathbf{P}_{ffy,k|k}^{(ii)} \mathbf{P}_{ffy,k|k}^{(jj)}}} \quad (5.100)$$

and we show in Figure 5.18 the evolution in time of the correlation coefficients for one such landmark with respect to the rest.

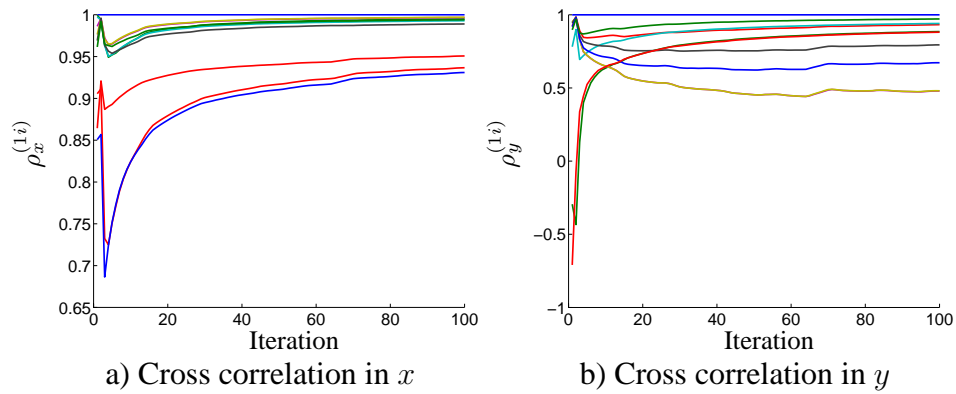


Figure 5.18: Landmarks become more and more correlated as iterations take place in the EKF-CML algorithm. The figure shows the increase in the cross correlation coefficients for landmark 1 with respect to each other landmark, in both the  $x$  and  $y$  directions.

### 5.5.2 Partial observations

Now, imagine a more realistic sensor, one with a limited field of view. Will the fact that not all landmarks can be seen and matched at all times worsen the reconstruction results? The answer is twofold. On the one side, and as will be seen in the next Chapter, when using a reduced number of landmarks, the angle between the observable and controllable subspaces is large, thus limiting the reconstruction of the vehicle and landmark locations. On the other hand, given the measurement noise models proposed, ones that are proportional to the distance to the observed landmark; by neglecting measurements from objects that are far from the robot, we also limit the value of the noise introduced by the measurement model.

These arguments might lead to the formulation of a strategy for the computation of the appropriate number of landmarks to use, optimal in the sense of the reconstruction results (closeness of the observability and controllability subspaces), and at the same time limiting the impact of sensor measurement noise (by eliminating unreliable observations in terms of the distance of the landmarks to the robot).

Let us analyze in greater detail, and with the aid of our standard test case from Section 5.5.1, how a limited field of view undermines the accuracy in localization when using the EKF-CML algorithm. Figure 5.19 shows a run of the algorithm with sequential innovation for a sensor with a limited field of view of  $2m$ . Notice how in between the 5th and 15th iterations approximately, the robot pose estimate depends on observations from 2 landmarks only. The growth in size of the covariance ellipses clearly shows the increase in the uncertainty in localization. By the time new landmark observations enter the field of view, the vehicle has already accumulated a large amount of dead reckoning error. Consequently, all initial landmark position and covariance estimates are biased with such initial localization error.

A similar situation occurs near the end of the simulation, around the 85th iteration. In this case, the map is revised with repeated observations of just one landmark only, and even when the localization estimate remains consistent with such observation, just one 2d landmark does not suffice in localizing a 3dof vehicle. The accumulated deviation is corrected once a previously learned landmark re-enters the field of view.

Consequently, the two most serious artifacts that can hinder accurate localization when the sensors have a limited field of view are, the wrong initialization of landmark estimates due to accumulated error, and the decrease in the dimensionality of the observed references during motion, up to the point in which pose recovery becomes ill-posed.

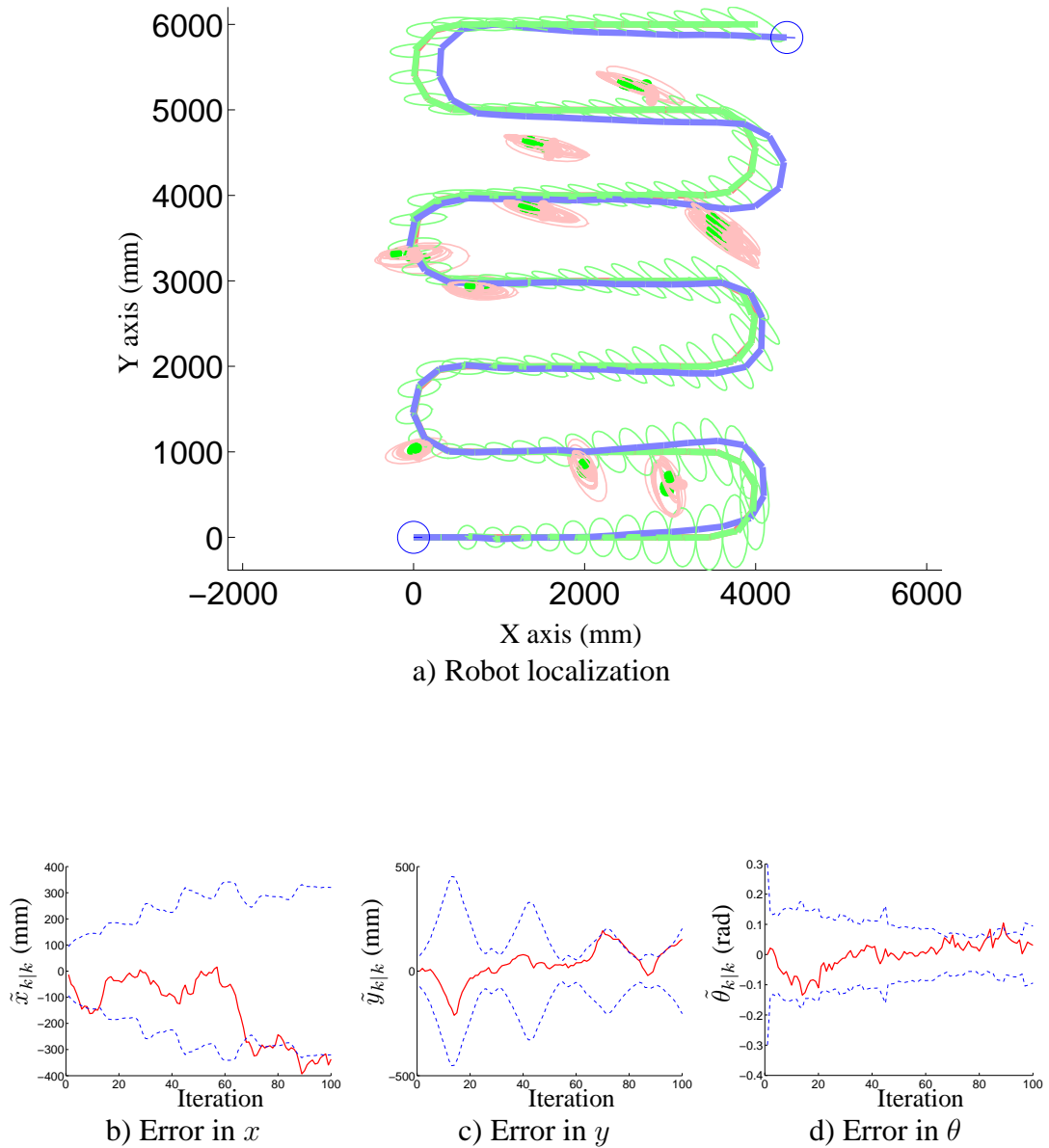


Figure 5.19: Full-covariance EKF CML for a path with 100 iterations and 10 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of  $2m$ .

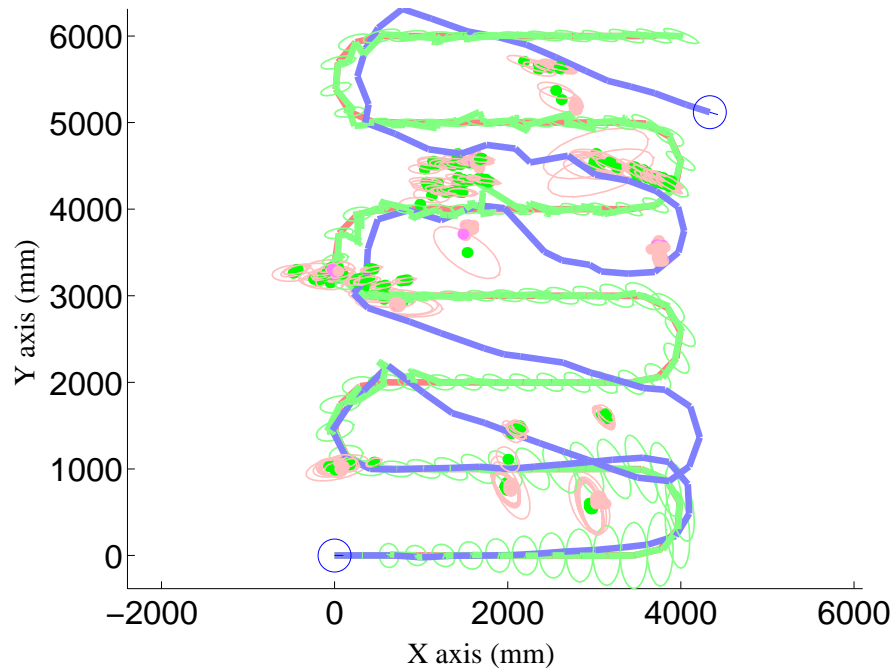
### 5.5.3 Data missassociation

A more drastic situation is the case when the landmark identification module is not error-free. That is the case when we allow the system a percentage of landmark identification mismatches. Imagine that we observe say visual landmarks, such as corners or lines extracted from intensity images, and that our landmark tracking algorithm is not very accurate at matching observations in consecutive frames because of illumination changes or other data association artifacts. To simulate this behavior we rely once more on our standard test case. For the sample run shown in Figure 5.20, landmark matching is performed with a 25 percent probability of missassociation within a  $1m$  radius.

The effects of landmark misidentification are much more pervasive than the lack of observations explained in the previous section. In the sample run shown in Figure 5.20, localization and map building proceeds smoothly until the 30th iteration, when the first mismatch occurs. The algorithm is not able to recover from this failure, and when the previously observed landmarks re-enter the field of view two things happen. On the one hand, the new observations of the already learned landmarks aid in reversing the error trend in localization. On the other hand, the same new observations are used to revise the mere location estimates of those landmarks. Moreover, new landmark observations will be initialized with corrupt robot location estimates.

Given the fact that the map is fully correlated, the effects of landmark mismatch propagate to the localization estimate of all the landmarks in the model.

The estimation theoretic approach to CML, as presented by Smith and Cheeseman, and later formalized by Leonard, Newman, and Durrant-Whyte among others, is very sensitive to data association errors; and as formulated lacks a theoretical foundation to deal with the problem. Efforts have been tailored at correcting the effects of data mismatches, and at finding measures of the spatial compatibility of landmark correspondence. We go one step further, providing a new formulation of temporal landmark quality measures.



a) Robot localization

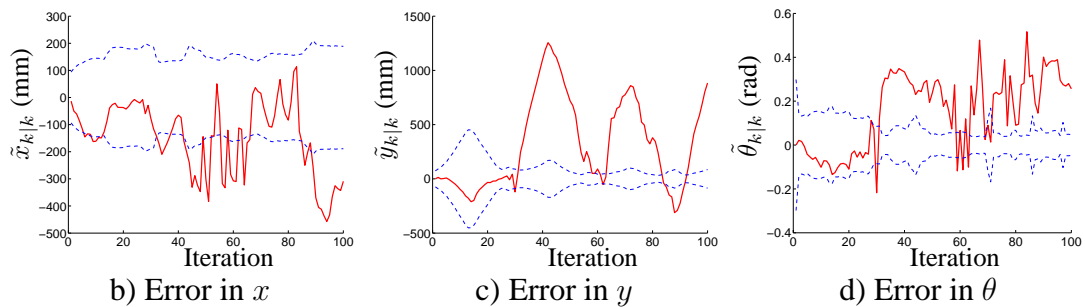
b) Error in  $x$ c) Error in  $y$ d) Error in  $\theta$ 

Figure 5.20: Full-covariance EKF CML for a path with 100 iterations and 10 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of  $2m$ . Data missassociation occurs within a radius of  $1m$  with a 25% probability.

### 5.5.4 Landmark validation

To correct the limitations of the EKF-CML algorithm shown so far, we present now results of the two strategies for the computation of landmark quality herein discussed. First, we will show the results of using the compatibility test to validate landmark observations in terms of their weighted distance to their expected location (in the sense of Mahalanobis  $d^2$ ). Figure 5.21 shows for our base test case, the improvement in the localization of the mobile robot when the spatial compatibility test from Equation 5.40 ( $\chi^2$  goodness of fit test) is performed, with a confidence level of 95%.

The red dots in the map plot represent all of the observations that pass the test. Note how landmarks 2 and 9 were misidentified early after they were initialized, and as a consequence their observations do not pass the spatial compatibility test for the rest of the simulation. The estimated robot localization error is represented by the green ellipses, which correspond to  $2\sigma$  level curves of the robot pose error covariance estimate. The red ellipses on the other hand, indicate projected  $2\sigma$  bounds for the observed landmark covariance estimates.

Also, due to the absence of observations that pass the test during the interval between the 85th and 95th iterations, the vehicle position estimate is revised by dead reckoning only, with the immediate consequence of having the vehicle covariance estimate enlarged at each iteration. Once a landmark passing the test re-enters the robot field of view, the filter kicks in again, with revised vehicle location and error covariance estimates.

Next, we include results of applying the temporal landmark quality test to the above results. Those landmarks whose temporal landmark quality falls below a given threshold are removed from the map. This is, they can no longer be used for localization. Only after observations for those landmarks become robust again, they can be reinitialized. Figure 5.22 shows the improvement of using the exponential decay rule as a function to validate temporal landmark quality, together with individual spatial landmark compatibility tests in CML. Furthermore, Figure 5.23 shows the results of using the data association probability as a function to test the temporal landmark quality.

As with Figure 5.21, the red dots in the plot represent only the observations that pass the test. Note how observations for landmarks 2 and 9 pass now the test, once they are reinitialized and become robust again. The advantage over the previous simulation is that these landmarks will still be used to build the map, providing more information to the filter ( $I_T$ ), and consequently converging to a lower localization error estimate. Figure 5.11 shows the evolution of both the exponential decay rule, and the data association probability as landmark quality measures for this test run.

We plot in Figure 5.24 the norm of the robot localization error in the  $xy$  plane to



show the improvement of using both the temporal landmark quality and spatial landmark compatibility tests, versus using the spatial landmark compatibility test only. The test run is performed on our standard case of the EKF-CML-SI algorithm with a  $2m$  limited sensor range, and 25% data missassociation.

New simulations were run for varying conditions, first by increasing the number of landmarks, and then by also incrementing the number of iterations. Figures 5.25 and 5.26 show the results of comparing the use of spatial landmark validation only versus using temporal and spatial landmark validation.

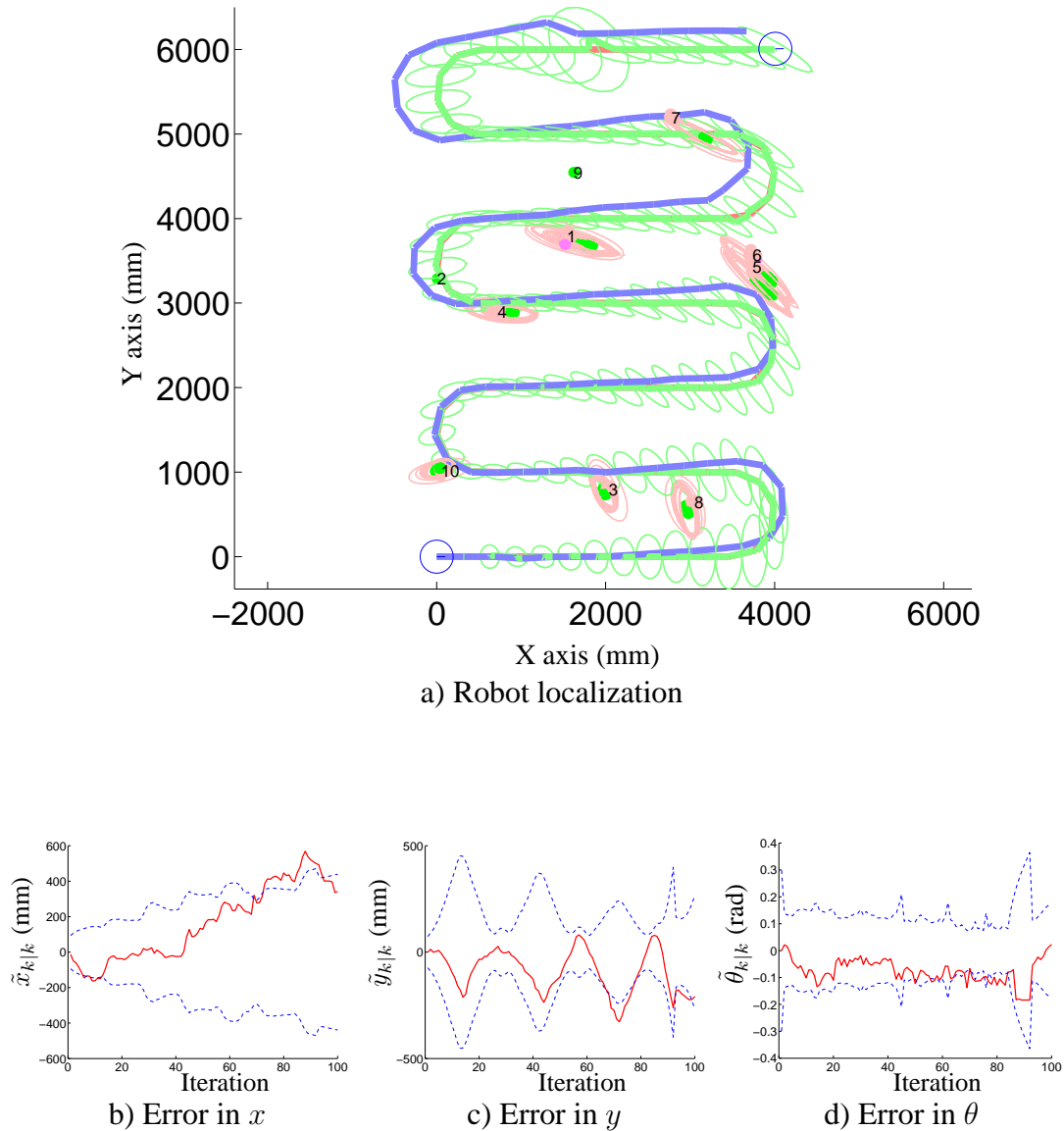


Figure 5.21: Results of using the individual compatibility test. Full-covariance EKF CML for a path with 100 iterations and 10 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of  $2m$ . Data missassociation occurs within a radius of  $1.0m$  with a 25% probability.

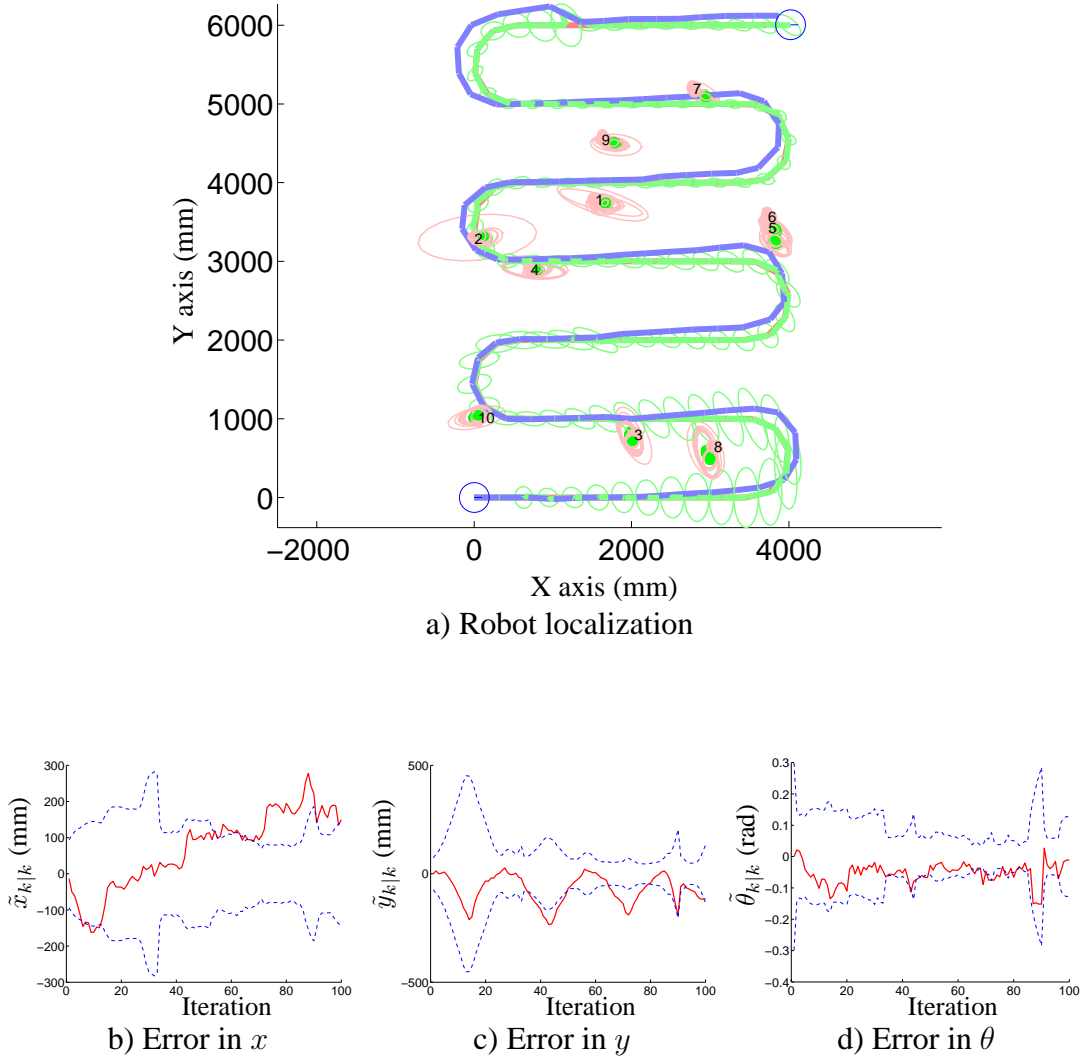


Figure 5.22: Results of using the individual compatibility test and the exponential decay rule to test temporal landmark quality,  $x_{q,LOW} = 0.6590$ ,  $x_{q,THLD} = 0.66$ ,  $x_{q,HIGH} = 0.8659$ . Full-covariance EKF CML for a path with 100 iterations and 10 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of  $2m$ . Data missassociation occurs within a radius of  $1.0m$  with a 25% probability.

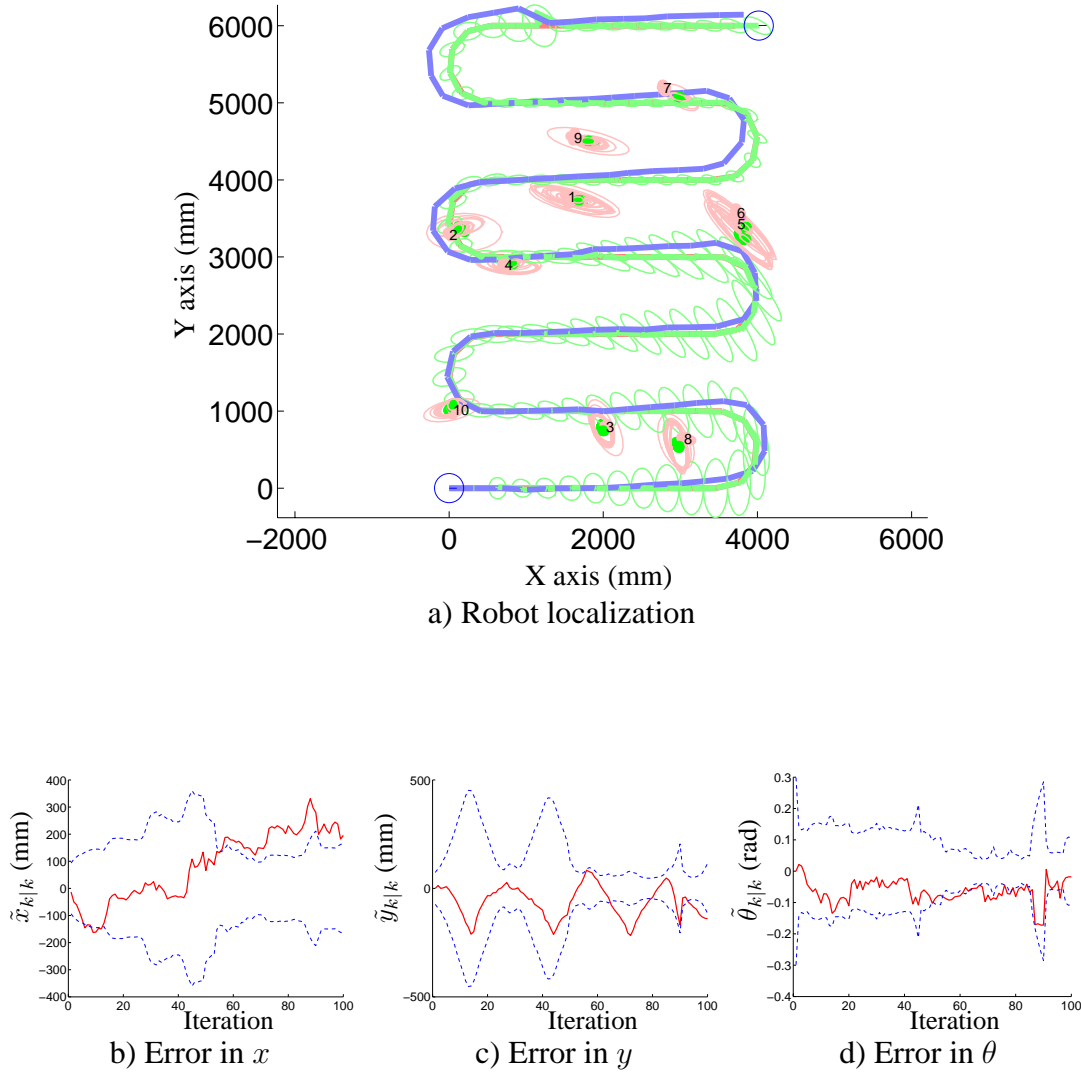


Figure 5.23: Results of using the individual compatibility test and the data association probability as temporal landmark quality test,  $x_{q,LOW} = 0$ ,  $x_{q,THLD} = 0.03$ ,  $x_{q,HIGH} = 1$ . Full-covariance EKF CML for a path with 100 iterations and 10 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of  $2m$ . Data missassociation occurs within a radius of  $0.5m$  with a 10% probability.

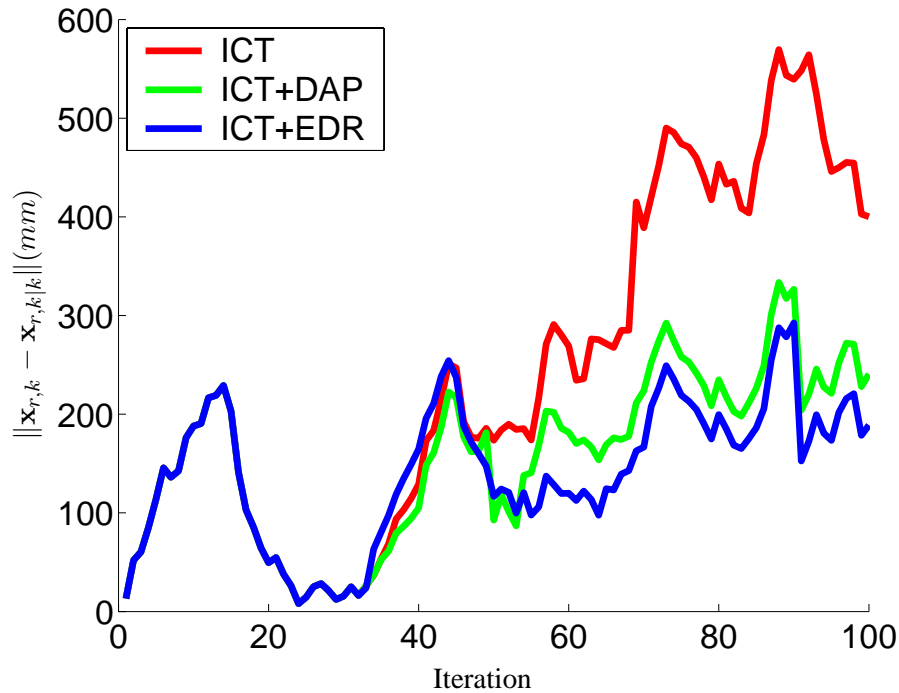


Figure 5.24: Robot localization error estimate. Comparison between ICT: Individual spatial compatibility test with limited vision range and data association errors (realistic case of CML); and the improvements proposed, ICT+DAP: EKF-CML-LV with both spatial and temporal landmark quality tests, using the data association probability, and ICT+EDR: EKF-CML-LV with both spatial and temporal landmark quality tests, using the exponential decay rule. The simulation is run over a path with 100 iterations and 10 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of  $2m$ . Data missassociation occurs within a radius of  $1.0m$  with a 25% probability.

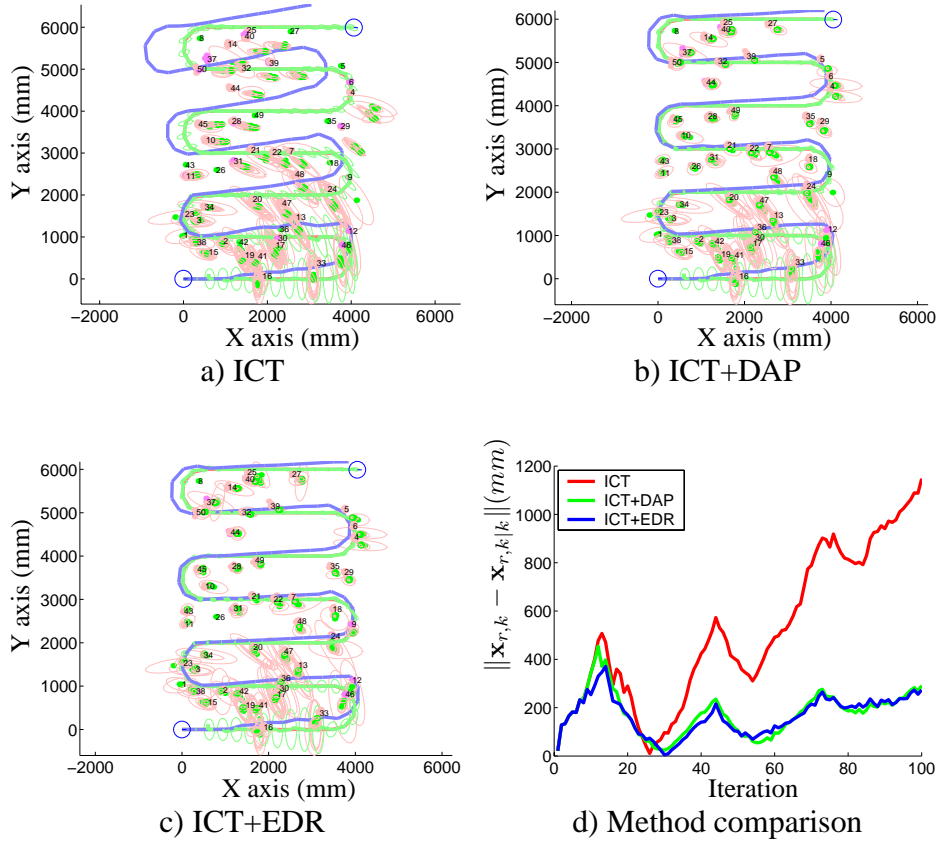


Figure 5.25: Robot localization error estimate. Comparison between ICT: Individual spatial compatibility test with limited vision range and data association errors (realistic case of CML); and the improvements proposed, ICT+DAP: EKF-CML-LV with both spatial and temporal landmark quality tests, using the data association probability, and ICT+EDR: EKF-CML-LV with both spatial and temporal landmark quality tests, using the exponential decay rule. The simulation is run over a path with 100 iterations and 50 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of  $2m$ . Data missassociation occurs within a radius of  $1.0m$  with a 25% probability.

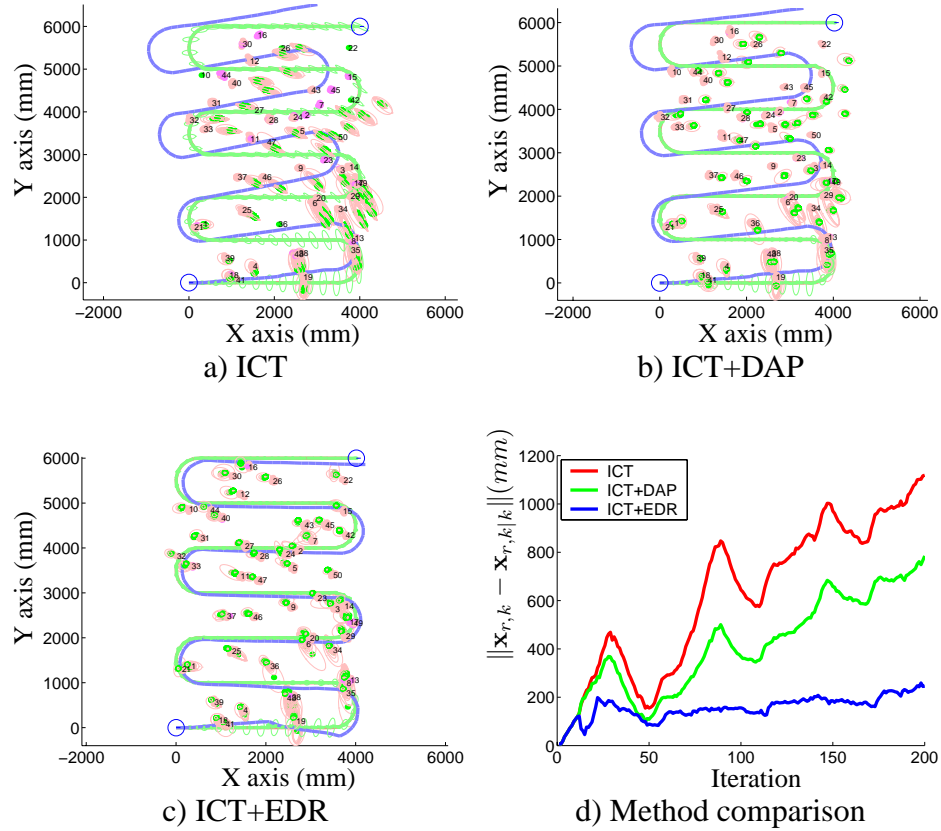


Figure 5.26: Robot localization error estimate. Comparison between ICT: Individual spatial compatibility test with limited vision range and data association errors (realistic case of CML); and the improvements proposed, ICT+DAP: EKF-CML-LV with both spatial and temporal landmark quality tests, using the data association probability, and ICT+EDR: EKF-CML-LV with both spatial and temporal landmark quality tests, using the exponential decay rule. The simulation is run over a path with 200 iterations and 50 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ , and a sensor with a limited radius of observation of  $2m$ . Data missassociation occurs within a radius of  $1.0m$  with a 25% probability.

## 5.6 Divergence

There is unfortunately a tradeoff in using the Extended Kalman Filter in CML. Even with perfect data association, for nonlinear vehicle models, the algorithm diverges in the long run. The divergence has been attributed to the linearization in Equations 5.6 and 5.7. An explanation of this behavior is by Julier and Uhlmann [108], where they show a counter example to the EKF-CML algorithm that did not converge to a fully correlated map in their simulations.

Their vehicle model was similar to our planar mobile robot from Section 5.2.2; a vehicle with three degrees of freedom, two Cartesian coordinates for position, and an orientation angle. Whilst in their model, the variables returned by the sensor were the landmark range and bearing in robot coordinates; in our case, they correspond to Cartesian landmark coordinates in a robot centered frame.

Despite this change of representation from Cartesian to polar coordinates, the observation model also returned two dimensional landmarks. In their simulations they showed how such a vehicle could not reconstruct its location from continuous observations to one such landmark; even, when no motion commands were conveyed to the robot, and when a noise free plant model was assumed.

They suggest that the divergence is due to the linearization of the system needed for the implementation of the EKF, arguing that the no-motion condition for an error-free model is equivalent to having a zero component in the vehicle part of the Kalman gain. Such assumption leads to the equality

$$\mathbf{H}_{\mathbf{x}_r}^{(i)} \mathbf{F}_{\mathbf{x}_r, \mathbf{x}_r} = -\mathbf{H}_{\mathbf{x}_f}^{(i)} \mathbf{x}_f^{(i)} \quad (5.101)$$

Furthermore, they show how Equation 5.101 holds only for the linear case of CML, and breaks down after linearization in the EKF version of the algorithm, and state that “being this a structural condition, no tuning procedure (inflation of plant and measurement noise covariances) would circumvent the problem”. For this reason they push for nonlinear approaches to KF.

To palliate the effects of nonlinearities we analyze in the next Chapter, how the localization error varies with respect to the number of landmarks used. Our study is from the point of view of control systems theory.

We believe that the amount of divergence of the algorithm may also be factored by the partial observability of the state vector, and that by keeping the state error covariance as close to the lower bound  $\mathbf{P}_{rr,0|0}$  as possible, one can reduce the divergence effects. This can only be done if we add more landmarks to the map.



## 5.7 Bibliographical notes

Mobile robot localization has been addressed from different perspectives, divided mainly in the following groups.

Correlation methods that match sensor signals against previously stored maps. An example of a correlation technique that matches vertical lines extracted from a stereo-vision system to a map of such signatures is by Pagès *et al.* [155]. Other systems that match visual data using correlations for self-localization include the ones by Carbonaro and Zingaretti [37], Drocourt *et al.* [68], Hashima *et al.* [92], Mallet and Lacroix [135], Talluri and Aggarwal [186], or Volpe *et al.* [207].

Estimation theoretic approaches that predict and refine the robot position from current and previous sensor readings, past position estimates, and motion commands, as well as uncertainty models of sensors and motion. Such methods are typically robust for local localization, provided that the initial estimate of the robot is sufficiently accurate, and that correct stochastic models for the robot dynamics and the sensors are available. Similar to the talk by Pagès *et al.* [155], but taking into account the stochastic nature of sensor data, an example where vertical lines are extracted from images and matched against a previously stored map of the environment is presented in the work by Ohya *et al.* [153]. In this case, robust self-localization is attained by means of Kalman filtering. Recent contributions in the use of Kalman filtering techniques for self-localization include those of Anousaki and Kyriakopoulos [14], Deans and Hebert [56], Duckett and Nehmzow [70], Kosaka and Kak [119], Lee and Recce [125], and our own contributions deriving from this work [9, 12].

Markov localization techniques are an extension to the above cited method, in which a probabilistic framework is used to maintain a position probability density over the whole set of robot poses. These techniques are better suited to solve the global localization problem, at the expense of stronger assumptions about the nature of the environment than with the Kalman filter. Recent examples that illustrate the use of Markov models for robot localization include the ones by Aycard *et al.* [21], and Fox *et al.* [77]. A thorough description of Markov localization methods can be found in the work by Thrun [192].

Finally, a mirage of other algorithms have been proposed in attempts to solve the robot localization problem. The ones that are less related to the work presented in this thesis vary from fuzzy-based algorithms [58, 169, 168], to the *minmax risk fixed-size confidence set estimate* [136], designed as an optimal set of decision rules in the minmax sense for dealing with the uncertainties in sensor measurements; a topological approach that makes use of a generalized Voronoi graph and graph matching techniques [47]; or a set-theoretic approach, in which the estimate of the position of the robot and

the landmarks is expressed in terms of bounded sets, making use of set membership estimation theory.

The main drawback of most of these approaches is however, that they have not been designed with continuously changing environments in mind. All of these methods treat moving or temporary objects as noisy data or measurement uncertainty. It is only with Kalman filter based methods and Markov localization techniques that by using statistical approaches we can cope with noisy data. Fuzzy based algorithms do the same by estimating qualitative models for sensor and robot uncertainty. We believe that by incorporating time-varying properties to our environment model, we are able to better cope with the dynamics inherent in typical mobile robotics environments.

The study of stochastic models for concurrent map building and localization in mobile robotics has been addressed by several research groups for over fifteen years. Within the KF approach to CML, seminal work by Smith and Cheeseman [182] suggested that as successive landmark observations take place, the correlation between the estimates of the location of such landmarks in a map grows continuously. This observation was ratified recently by Dissanayake *et al.* [64] with a proof showing that the estimated map converges monotonically to a relative map with zero uncertainty. They also showed how the absolute accuracy of the map reaches a lower bound defined only by the initial vehicle uncertainty. With respect to covariance initialization, we would like to express our gratitude to P. Newman for making it clear to us. Thanks to that we were able to reproduce in our simulations the three asymptotic convergence properties of CML described in Section 5.1.5.

In spite of these fundamental convergence properties of the KF approach to CML, there exist some limitations that still hinder full development of CML applications. The three most criticized of these limitations are the time and space complexity of the algorithm, the restriction to unimodal zero mean white Gaussian models of uncertainty, and the data association and landmark quality problems. We go even further in our analysis of the CML problem, and show in the next chapter one more fundamental limitation to the algorithm, that full reconstruction of the map state vector is not possible, regardless of the vehicle model chosen, and that the expected error in state estimation is proportional to the number of landmarks used.

The explicit solution for the monobot CML problem presented by Gibbens *et al.* [83] shed some light on the relation between the total number of landmarks (in the form of the total Fisher information  $I_T$ ), and the asymptotic values for  $\mathbf{P}$ . In the next chapter we treat this topic from a different perspective, with the principles of controllability and observability of linear systems in mind.

On a side issue, special interest has recently been put on the search for methods to reduce the time and space complexity of the CML algorithm, at the expense of a

sub-optimality in the solution; by pruning the map from those landmarks that surpass a certain degree of spatial correlation [63], or by local updating of the map exploiting the advantages of sequential innovation in KF [88]. One requisite for sequential innovation is that observations must be uncorrelated, and they present a suboptimal decorrelation algorithm for that purpose.

Julier *et al.*, also present a suboptimal solution to CML with sequential innovation in KF, reducing the time and space complexity of the algorithm from  $O(n^2)$  to  $O(n)$  [106, 109]. Instead of decorrelating observations, they do away with cross correlations between observations and state estimates, and update the state covariance estimate with a weighting factor that penalizes the lack of information about correlations.

In the search for robust models of uncertainty in CML, a variety of routes have been explored. Durrant-Whyte *et al.* presented an extension to the typical KF-CML algorithm with the use of sums of Gaussians (SOGs), to approximate more general probability distributions for the modeling of unstructured sub-sea terrain [72]. Tardós *et al.* [187] build maps of indoor environments from sonar data, using the Hough transform as a tool to perceptual grouping of sonar returns, and by joining local maps into a global representation.

There is no general consensus on what constitutes a good measure for landmark quality. Intuition suggests measures proportional to the temporal dispersion and inversely proportional to the spatial dispersion of landmarks. We have no knowledge of previous attempts at formulating temporal landmark quality models. There resides precisely one of the most relevant contributions of this thesis. In an attempt to unify landmark quality and data association uncertainty within the structure of the estimation-theoretic approach to CML, we proposed in this Chapter, an augmented map model that incorporates temporal landmark temporal tests to validate the quality of observations.

Dissanayake *et al.* address the issue of landmark robustness as an implementation detail only [64], suggesting a quality measure based on the probability density function of the observations associated to any given landmark, and disregarding the temporal dispersion of such observations. A different, more simple measure of landmark quality is proposed [63], as the trace of the error covariance submatrix  $\mathbf{P}_{f_i}$ .

To aid in data association, Castellanos *et al.* suggested in [43], a  $\chi^2$  test to evaluate scene-to-map landmark match hypotheses based on the squared Mahalanobis distance between map landmarks and observations. Here, the uncertainty in the location of a landmark in the map plays the role of quality assessment. More recently, Neira and Tardós [150] presented a constrained search algorithm for scene to map hypothesis formation, in which the use of the joint compatibility test is argued to supersede the classical nearest neighbor compatibility test typically used for independent matching of sensors and measurements.

Leonard *et al.* [127, 128] have realized that spatial dispersion is not sufficient for landmark quality assessment, and that temporal information is needed as well. They have opted for a delayed decision approach to data association, in which by adding previous robot states to the state vector, they can include in the Kalman filter covariance matrix correlations pertaining estimations from previous vantage points. In this way, they can add non-invariant features such as range-only measurements to aid in data association. A serious disadvantage of this technique is that the map grows linearly in time, increasing the computational burden of the CML algorithm, even when no new features are added to the map. Another serious disadvantage is the divergence of the fundamental motivation of using a Kalman filter to maintain a parametric representation of the history of the uncertainty of robot and landmark localization by means of an estimated full-covariance matrix.

Davison *et al.* have also studied the spatial dispersion of landmark observations as a measure of landmark quality, and suggest that when one has to choose among several observations to update, the one with the largest hyper-ellipsoid volume of dispersion contributes more to the overall reduction of the map error covariance. Similarly, they argue that by diagonalizing the landmark covariance in measurement space  $\mathbf{H}_i\mathbf{P}\mathbf{H}_i^\top$ , they are able to bound the search for scene-to-model landmark matches. These observations led them to active vision heuristics for gaze control when building maps with CML [53, 54]. Another suggestion is the use of robust methods aimed at eliminating observation outliers, such as RANSAC [76], in order to diminish the effect of measurement and data association errors.

General references to linear systems theory include the books by Kailath [110], and DeCarlo [57]. For the study of the Pearson  $\chi^2$  test and other topics of statistics the reader is referred to the books by Peña [55] and Fukunaga [78]. Linear Gaussian models in general are treated extensively in the survey by Roweis and Ghahramani [167], and in the book by Duda *et al.* [71].

## 5.8 Conclusions

This chapter presents a revision of the traditional full-correlation EKF CML algorithm for mobile robot localization and map building. We extend the traditional algorithm by adding temporal landmark quality measures, and a temporal landmark quality test to validate the history of data association. These quality measures permit the maintenance of the map by the elimination of inconsistent observations. The removal of weak landmarks from the state vector and state covariance matrix does not violate the convergence properties of CML. Special attention has been paid in the selection of the

temporal landmark quality models, to guarantee that the uncertainty in the map estimates still reduces monotonically. The proposed solution contributes in simplifying the data association problem in CML.

One drawback that might limit the use of the newly introduced temporal landmark quality test in computer vision based CML is the concept of visibility. Temporal landmark quality can only be revised for those landmarks that are visible to the vehicle. And, depending on the sensor and landmark models used, a visibility condition might be rather difficult to assert. Occlusions of walls inferred from laser data are easy to compute; however to test if a visual landmarks should be present or not in a scene, is rather complicated.



## Chapter 6

# Fully Correlated Maps

Two fundamental aspects in the use of the Kalman Filter for localization are the issues of observability and controllability. Unfortunately, in CML, the state space constructed by appending the robot pose and the landmark locations is fully correlated; a situation that hinders full observability. Moreover, the modeling of map states as static landmarks yields a partially controllable state vector. The identification of these problems, and the steps taken to palliate them, constitute one of the main contributions of this thesis. The bulk of which is covered in this chapter.

In Chapter 5 we dedicated a few paragraphs to the formulation of the EKF-CML equations for a simple linear one-dimensional robot we called, the monobot. The importance of the formulation of a naive case study such as the monobot becomes evident in this Chapter, where we study the behavior of the estimation-theoretic approach to CML from a control systems point of view. We need this simple linear model to show the effects of partial observability and partial controllability, with respect to the number of landmarks used, even when data association is perfect. That is, when all landmarks are correctly identified at each and every iteration.

In Section 6.1 we analyze the steady state behavior of the error state covariance for the monobot CML, and show by example, that the steady state of the filter will always depend on the initial noise parameters. The effect is known as filter instability [194], and is in general an undesirable feature in state estimation.

In Section 6.2 we derive an expression for the total Fisher information in CML, and show how full correlation prohibits the use of the Cramer Rao lower bound for the vehicle and map covariance. The analysis yields a closed form solution for the monobot Fisher information matrix that shows explicitly, the unobservable directions of the map state in CML.

Filter instability, and the singularity of the Fisher information matrix, are both consequences of having partial observability and controllability. Section 6.3 is devoted to the computation of general expressions for the bases of the controllable and observable subspaces in CML. These expressions are later simplified in Sections 6.4 and 6.5 for the monobot and planar vehicle models, respectively. In the end, we show that the angle between these two subspaces is determined only by the total number of landmarks in the map. The result is that as the number of landmarks is incremented, the vehicle pose states get closer to being reconstructible.

The observability condition guarantees a *steady flow* of the information about each state component, and prevents the uncertainty (error state covariance) from becoming unbounded [26]. In Section 6.6 we show how partial observability in CML can be avoided by adding a fixed external sensor to the state model, or equivalently, by setting a fixed landmark in the environment to serve as global localization reference. Full observability yields the existence of a (not necessarily unique) steady state *psd* solution for the error covariance matrix  $\mathbf{P}$ .

The controllability condition allows the process noise to enter into each state component and prevents the covariance matrix from becoming singular [26]. In Section 6.7 we show how having a semi-definite covariance matrix reflects on the filter's belief that it has perfectly accurate estimates of some state components, for which the Kalman gain evaluates to zero, and the innovations are not considered in the revision of the state estimate. To palliate the effects of partial controllability, we propose the addition of artificial positive definite process noise associated to the landmark states.

More than often, map building in mobile robotics is performed in a teleoperation mode, in which a user gives motion commands to the vehicle from a remote station. However, there are times when we want the robot to follow a previously defined trajectory and at the same time build an incremental map of its environment. For this later case, we propose in Section 6.8 the use of a robot motion control law optimal in the sense that it minimizes both the overall state estimation error (robot position plus location of landmarks), and the amount of energy input to the system in the form of motion commands. The simplicity of the approach resides in the existing duality between observability and controllability. That is, the aforementioned control law for robot motion is dual to the mere Kalman filter used for simultaneous map building and localization.

Finally, in Section 6.9 we show how filter divergence in CML, present in the non-linear case [108] can also be palliated to some extent, either by overestimating the plant noise covariance, or with the online computation of adequate models of sensor uncertainty.



## 6.1 Steady state behavior of EKF-CML

We start the discussion with a pictorial representation of the behavior of the EKF-CML algorithm. Consider a monobot with initial vehicle localization variance  $\sigma_{r,0|0}^2 = 1$ , and various levels of motion and sensor noise variances, denoted by  $\sigma_v^2$  and  $\sigma_w^2$ , respectively. For the sake of simplicity, imagine a robot with constant motion uncertainty, regardless of the size of the input command, and a sensor with constant observation uncertainty, regardless of the distance to the landmark being observed. Later, we will take into account more realistic models of uncertainty, and analyze the advantages of having a measurement covariance matrix that can be computed online.

The final vehicle and landmark variances ( $\sigma_{r,k|k}^2, \sigma_{f,k|k}^2, k \rightarrow \infty$ ) depend only on the initial parameters  $\sigma_{r,0|0}^2, \sigma_v^2, \sigma_w^2$ , and on the total number of landmarks  $n$ . The evolution of the error covariance matrix is independent of the state input, and measurements throughout the run of the algorithm. Meaning that, for a monobot with perfect data association and constant motion and sensor uncertainty, the computation of the Kalman gain could even be performed offline. That is, the asymptotic (steady state) behavior of the filter, and its rate of convergence are always the same, regardless of the actual motions and measurements.

Figure 6.1 shows the steady state vehicle and landmark variances of the EKF-CML algorithm applied to a monobot when observations of 1, 2, 3, and 50 landmarks are available. The Figure plots the influence of each of the noise parameters  $\sigma_v^2$  and  $\sigma_w^2$  with respect to the final vehicle and landmark uncertainty.

On the one hand, the final vehicle localization variance depends less on the vehicle plant noise variance  $\sigma_v^2$  than on the measurement noise variance  $\sigma_w^2$ . On the other hand, as the number of landmarks grows, we see a considerable decrease of the steady state value for all the entries in  $\mathbf{P}$ . The landmark-to-vehicle and landmark-to-landmark cross correlation terms ( $\rho_{r,f^{(i)}} \sigma_r \sigma_{f^{(i)}}$  and  $\rho_{f^{(i)}f^{(j)}} \sigma_{f^{(i)}} \sigma_{f^{(j)}}$ ) although not shown on the plots, they all converge to the same values as the landmark variance  $\sigma_{f^{(i)}}^2$ . Meaning that, as  $k$  tends to infinity, the map becomes fully correlated, i.e.,  $\rho_{f^{(i)}f^{(j)}} = 1$ .

The two observations: a) that rate of convergence of  $\mathbf{P}$  is fixed, and b) that the precision of the map (in terms of the asymptotic value of the error covariance) is independent of the plant variance  $\sigma_v^2$ ; were first described by Gibbens *et al.* [83], in which a closed form solution for the computation of Equation 6.1 for the continuous time domain monobot was presented.

Without lose of generality, in the sequel we take the measurement noise Jacobian  $\mathbf{H}_w = \mathbf{I}$ , and drop for clarity, the subscript  $\mathbf{x}$  from the model Jacobians  $\mathbf{F}_x$  and  $\mathbf{H}_x$ .

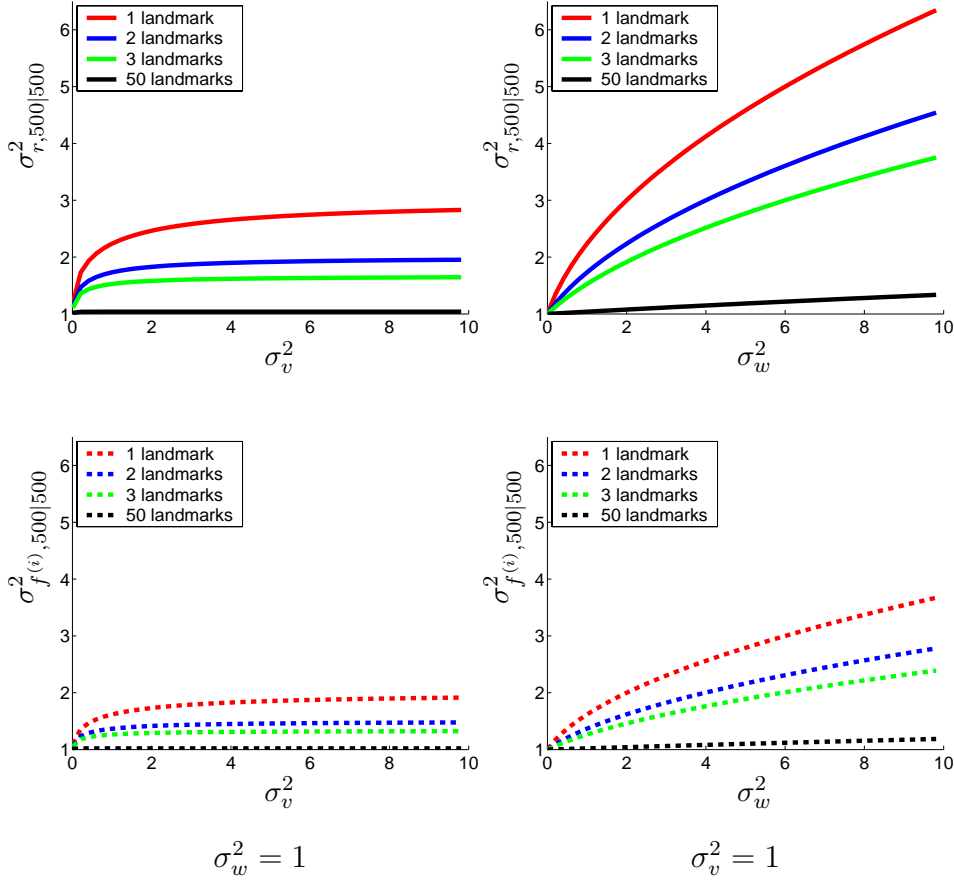


Figure 6.1: Final vehicle and landmark localization variances after 500 iterations of CML for a monobot with initial localization variance  $\sigma_{r,0|0}^2 = 1$ , and various levels of plant and sensor noise.

The steady state covariance matrix is given by the solution of the Riccati equation

$$\mathbf{P} = \mathbf{F}(\mathbf{P} - \mathbf{P}\mathbf{H}^\top(\mathbf{H}\mathbf{P}\mathbf{H}^\top + \mathbf{W})^{-1}\mathbf{H}\mathbf{P})\mathbf{F}^\top + \mathbf{V} \quad (6.1)$$

and for the linear case it is only a function of  $\mathbf{P}_{rr,0|0}$ ,  $\mathbf{V}$ ,  $\mathbf{W}$ , and  $n$ . Note however that, for the nonlinear case, the computation of the Jacobians  $\mathbf{F}$  and  $\mathbf{H}$  will in general also depend on the steady state value of  $\mathbf{x}$ . Moreover, for the linear case, the solution to the Riccati equation is given by the Cramer Rao lower bound [26].

## 6.2 Total Fisher information and the Cramer Rao lower bound

The specific influence of  $\sigma_{r,0|0}^2$ ,  $\sigma_v^2$ ,  $\sigma_w^2$ , and the total number of landmarks  $n$  ( $\mathbf{P}_{rr,0|0}$ ,  $\mathbf{V}$ ,  $\mathbf{W}$ , and  $n$  in the multidimensional case) in the asymptotic value of  $\mathbf{P}$  can be better understood in terms of the total Fisher information available to the filter.

Under the Gaussian assumption for the vehicle and sensor noises, the Kalman filter is the optimal minimum mean square error estimator. And, as pointed out in [26], minimizing the least squares criteria from Equation C.17

$$E[\tilde{\mathbf{x}}_{k+1|k+1}\tilde{\mathbf{x}}_{k+1|k+1}^\top]$$

is equivalent to the maximization of a likelihood function  $\Lambda(\mathbf{x})$  given the set of observations  $Z^k$ ; that is, the maximization of the joint probability density function of the entire history of observations, conditioned on the state  $\mathbf{x}$

$$\Lambda(\mathbf{x}) = p(Z^k) = p(\mathbf{z}_k, Z^{k-1}) = p(\mathbf{z}_k|Z^{k-1})p(Z^{k-1}) = \prod_{i=1}^k p(\mathbf{z}_i|Z^{i-1}) \quad (6.2)$$

Given that the above pdfs are Gaussian, and that  $E[\mathbf{z}_i] = \mathbf{H}\mathbf{x}_{i|i-1}$ , the pdf for each measurement in CML is

$$\begin{aligned} p(\mathbf{z}_i|Z^{i-1}) &= N(\tilde{\mathbf{z}}_{i|i-1}; \mathbf{0}, \mathbf{S}_i) \\ &= (2\pi)^{-\frac{\dim \mathbf{z}}{2}} |\mathbf{S}_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{\mathbf{z}}_{i|i-1}^\top \mathbf{S}_i^{-1} \tilde{\mathbf{z}}_{i|i-1})} \end{aligned} \quad (6.3)$$

That is, the joint pdf of the sequence of measurements  $Z^k$  is equal to the product of the marginal pdfs of the corresponding innovations.

In practice however, it is more convenient to consider the log likelihood function

$$\begin{aligned} \ln \Lambda(\mathbf{x}) &= \sum_{i=1}^k \ln p(\mathbf{z}_i|Z^{i-1}) \\ &= -\frac{1}{2} \sum_{i=1}^k \tilde{\mathbf{z}}_{i|i-1}^\top \mathbf{S}_i^{-1} \tilde{\mathbf{z}}_{i|i-1} + \sum_{i=1}^k \ln |2\pi \mathbf{S}_i| \end{aligned} \quad (6.4)$$

The maximum of  $\ln \Lambda(\mathbf{x})$  is at the value of the state  $\mathbf{x}$  that most likely gave rise to the observed data  $Z^k$ , and is obtained by setting the derivative of Equation 6.4 with

respect to  $\mathbf{x}$  equal to zero, which gives

$$\nabla_{\mathbf{x}} \ln \Lambda(\mathbf{x}) = \sum_{i=1}^k \mathbf{H}^T \mathbf{S}_i^{-1} \tilde{\mathbf{z}}_{i|i-1} \quad (6.5)$$

An intuitive interpretation of the maximum of the log-likelihood in Equation 6.4 is that the best estimate for the state  $\mathbf{x}$ , in the least squares sense, is the one that makes the sum of the entire set of Mahalanobis distances  $\sum_{i=1}^k \tilde{\mathbf{z}}_{i|i-1}^T \mathbf{S}_i^{-1} \tilde{\mathbf{z}}_{i|i-1}$  as small as possible. A measure that is consistent to the spatial compatibility test described in Section 5.1.8.

The Fisher information matrix, a quantification of the maximum existing information in the observations about the state  $\mathbf{x}$  is defined (in [25] and [26]) as the expectation on the dyad of the gradient of Equation 6.4

$$\mathbf{J} = E[(\nabla_{\mathbf{x}} \ln \Lambda(\mathbf{x}))(\nabla_{\mathbf{x}} \ln \Lambda(\mathbf{x}))^T] \quad (6.6)$$

Taking the expectation on the innovation error  $E[\tilde{\mathbf{z}}_{i|i-1} \tilde{\mathbf{z}}_{i|i-1}^T] = \mathbf{S}_i$  in the above formula gives the sum

$$\mathbf{J} = \sum_{i=1}^k \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{W})^{-1} \mathbf{H} \quad (6.7)$$

In the linear case, this expression for the total Fisher information is only a function of  $\mathbf{P}_{rr,0|0}$ ,  $\mathbf{V}$ , and  $\mathbf{W}$ . If on the other hand, the EKF has been used, the Jacobian  $\mathbf{H}$  in Equation 6.6 should be evaluated at the true value of the states  $\mathbf{x}_0, \dots, \mathbf{x}_k$ . Since these are not available, an approximation is obtained at the estimates  $\mathbf{x}_{i|i-1}$ . The pre and post multiplying Jacobian  $\mathbf{H}$  in Equation 6.7 is, in this context, also known as the *sensitivity matrix* [66].

A necessary condition for the estimator (the Kalman filter) to be consistent in the mean square sense is that there must be an increasing amount of information, about the state  $\mathbf{x}$  in the measurements. That is, as  $k \rightarrow \infty$ , the Fisher information has to tend to infinity.

Figure 6.2 shows this for the monobot with constant parameters  $\sigma_{r,0|0}^2 = \sigma_v^2 = \sigma_w^2 = 1$ , and various sizes for the observation vector. Notice how, as the total number of landmarks grows, the total Fisher information also grows, directly relating the number of landmarks to the amount of information available for state estimation in CML.

Solving for the  $k$ -th sum term in  $\mathbf{J}$  for the monobot yields the expression

$$\mathbf{J}_k = \begin{bmatrix} \sum \sum \zeta_{ij} & -\boldsymbol{\zeta} \\ -\boldsymbol{\zeta}^T & \mathbf{S}_k^{-1} \end{bmatrix} \quad (6.8)$$

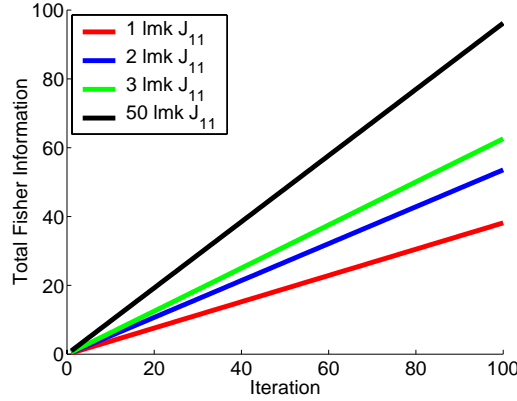


Figure 6.2: First entry in the total Fisher information matrix for a monobot with variance parameters  $\sigma_{r,0|0}^2 = \sigma_v^2 = \sigma_w^2 = 1$ , and various sizes for the measurement vector.

with  $\varsigma_{ij}$  the  $ij$ -th entry in  $\mathbf{S}_k^{-1}$ , and  $\varsigma = [\sum \varsigma_{1i}, \dots, \sum \varsigma_{ni}]$ .

Citing Bar-Shalom *et al.* [26]: “A lower bound on the minimum achievable covariance in state estimation is given by the posterior Cramer Rao lower bound”<sup>1</sup>

$$E[\tilde{\mathbf{x}}_{k+1|k+1} \tilde{\mathbf{x}}_{k+1|k+1}^\top] \geq \mathbf{J}^{-1} \quad (6.9)$$

Unfortunately, it can be easily shown, at least for the monobot case, that the first row (or column) of  $\mathbf{J}$  is equivalent to the sum of the rest of the rows (or columns), producing a singular total Fisher information matrix. In CML the Cramer Rao lower bound cannot be evaluated. CML is unobservable.

In the deterministic case, observability is a binary property. A system is either observable or unobservable. Stochastic observability on the other hand is quantified by the Cramer Rao lower bound. A small condition number on  $\mathbf{J}$  is an indicator of *good observability*; near zero eigenvalues on  $\mathbf{J}$  mean *marginal observability*; and singularity of  $\mathbf{J}$  indicates that part of the state vector is *unobservable*.

Citing once more Bar-Shalom *et al.*: “if the Fisher information matrix is not invertible, then the lower bound from Equation 6.9 will not exist, actually it will have one or more infinite eigenvalues (one in the case of CML), which means total uncertainty in a subspace of the state space, that is, the information is insufficient for the estimation problem at hand.”

<sup>1</sup>See Appendix D for our interpretation of matrix inequalities.

This is a consequence of the form of the Jacobian  $\mathbf{H}$ , i.e, of the full correlation in CML. Zero eigenvalues of  $\mathbf{H}^\top \mathbf{S}^{-1} \mathbf{H}$  are an indicator of partial observability, and the corresponding vectors give the unobservable directions in state space.

So for example, for a one-landmark monobot, the innovation covariance is the scalar  $s = \sigma_r^2 - 2\rho_{rf}\sigma_r\sigma_f + \sigma_f^2 + \sigma_w^2$ , and since  $\mathbf{H} = [-1, 1]$ , the Fisher information matrix in Equation 6.8 evaluates to

$$\mathbf{J} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \sum_{i=1}^k \frac{1}{s_i} \quad (6.10)$$

The unobservable direction of the state space is the eigenvector associated to the null eigenvalue of  $\mathbf{J}$ , we denote it for now  $\mathbf{E}_{\text{KerR}}$  (the name will be clear soon), and evaluates to

$$\mathbf{E}_{\text{KerR}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (6.11)$$

### 6.3 Partial observability and controllability

The solution to the Riccati Equation 6.1 converges to a finite steady state covariance if the pair  $\{\mathbf{F}_x, \mathbf{H}_x\}$  is completely observable. If in addition, the pair  $\{\mathbf{F}_x, \mathbf{F}_v\}$  is completely controllable, then the steady state covariance is a unique positive-definite matrix, independent of the initial covariance  $\mathbf{P}_{0|0}$ .

None of these two conventional results of stochastic estimation theory are valid in CML. We have already seen in the previous Section how due to partial observability the Cramer Rao lower bound cannot be computed.

Partial observability in CML means that the information entering the system in the form of innovations is fully correlated, and that there is no guarantee that the state error covariance is bounded. In other words, given the singularity of the Fisher information matrix  $\mathbf{J}$ , the information extracted from the innovations  $\tilde{\mathbf{z}}$  is insufficient for the entire reconstruction of the state estimate  $\mathbf{x}$ .

To see what part of the state space is compromised by full correlation, we now develop closed form expressions for the bases of the observability and controllability subspaces in CML and relate them to the total number of landmarks used.

The linearized EKF state model can be rewritten in terms of the error states with

$$\tilde{\mathbf{x}}_{k+1|k} = \mathbf{F}_x \tilde{\mathbf{x}}_{k|k} + \mathbf{F}_v \mathbf{v}_k \quad (6.12)$$

$$\tilde{\mathbf{z}}_{k+1|k} = \mathbf{H}_x \tilde{\mathbf{x}}_{k+1|k} + \mathbf{H}_w \mathbf{w}_{k+1} \quad (6.13)$$

Combining the plant and measurement noises into one large input vector, the above error driven system is equivalent to the augmented model

$$\tilde{\mathbf{x}}_{k+1|k} = \mathbf{F}_x \tilde{\mathbf{x}}_{k|k} + \begin{bmatrix} \mathbf{F}_v & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_k \\ \mathbf{w}_{k+1} \end{bmatrix} \quad (6.14)$$

$$\tilde{\mathbf{z}}_{k+1|k} = \mathbf{H}_x \mathbf{F}_x \tilde{\mathbf{x}}_{k|k} + \begin{bmatrix} \mathbf{H}_x \mathbf{F}_v & \mathbf{H}_w \end{bmatrix} \begin{bmatrix} \mathbf{v}_k \\ \mathbf{w}_{k+1} \end{bmatrix} \quad (6.15)$$

and the controllability matrix for such a plant is

$$\mathbf{Q} = \begin{bmatrix} \mathbf{F}_v & \mathbf{0} & | & \mathbf{F}_x \mathbf{F}_v & \mathbf{0} & | & \dots & | & \mathbf{F}_x^{\dim x - 1} \mathbf{F}_v & \mathbf{0} \end{bmatrix} \quad (6.16)$$

Note that Equation 6.14 is already expressed in the so-called Kalman controllable form, i.e.,

$$\begin{bmatrix} \tilde{\mathbf{x}}_{r,k+1|k} \\ \tilde{\mathbf{x}}_{f,k+1|k} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{x_r} & \\ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{r,k|k} \\ \tilde{\mathbf{x}}_{f,k|k} \end{bmatrix} + \begin{bmatrix} \mathbf{F}_{v_r} & \mathbf{0} & | & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{r,k} \\ \mathbf{0} \\ \mathbf{w}_{k+1} \end{bmatrix} \quad (6.17)$$

and the controllability matrix in Equation 6.16 reduces to

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{F}_{v_r} & \mathbf{0} & | & \mathbf{F}_{x_r} \mathbf{F}_{v_r} & \mathbf{0} & | & \dots & | & \mathbf{F}_{x_r}^{\dim x - 1} \mathbf{F}_{v_r} & \mathbf{0} \end{bmatrix} \quad (6.18)$$

Consequently, the dimensionality of the controllable subspace, spanned by the column space<sup>2</sup> of  $\mathbf{Q}$ , ( $\text{Im}\mathbf{Q}$ ), is

$$\text{rank } \mathbf{Q} = \text{rank } \tilde{\mathbf{Q}} = \dim \mathbf{x}_r \quad (6.19)$$

regardless of the number of landmarks in the map. Obviously, the only controllable states are the ones associated with the vehicle motion.

On the other hand, the observability matrix of our error driven system is

$$\mathbf{R} = \begin{bmatrix} \mathbf{H}_x \mathbf{F}_x \\ \mathbf{H}_x \mathbf{F}_x^2 \\ \vdots \\ \mathbf{H}_x \mathbf{F}_x^{\dim x} \end{bmatrix} \quad (6.20)$$

The rank of  $\mathbf{R}$  indicates the dimensionality of the observable subspace, which in

---

<sup>2</sup>See Appendix D for a formal definition of the four fundamental subspaces of linear algebra.

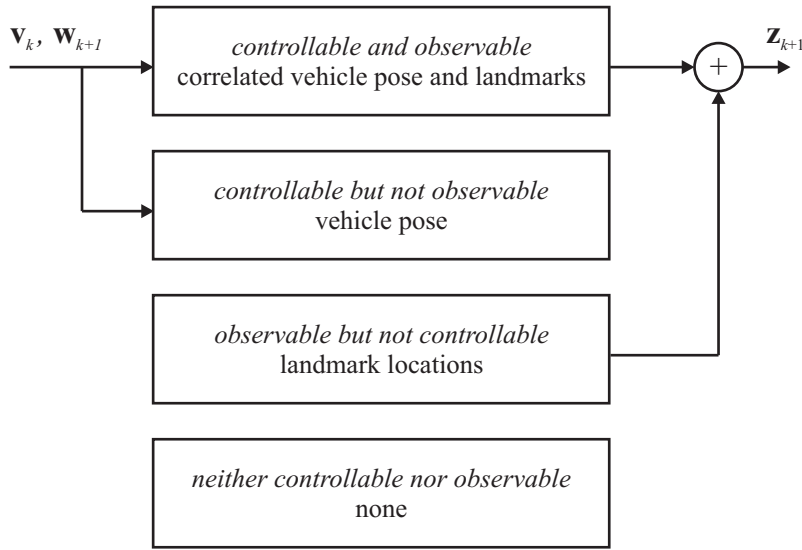


Figure 6.3: Controllability and observability of the state space in CML.

turn, is spanned by the row space of  $\mathbf{R}$ ,  $(\text{Im}\mathbf{R}^\top)$ .

$$\text{rank } \mathbf{R} = \dim \mathbf{x} - \dim \mathbf{x}_{f(i)} \quad (6.21)$$

The decomposition of the state space is portrayed in Figure 6.3. The arrows indicate their roles in the realization of the system.

## 6.4 The monobot

We return our attention now to the simple linear one-dimensional robot from Section 5.2.1. Consider the even more restrictive case in which only one landmark is available. By substituting the resulting expressions for the model Jacobians in Equations 5.51 and 5.52, the controllability and observability matrices reduce to

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (6.22)$$

$$\mathbf{R} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \quad (6.23)$$



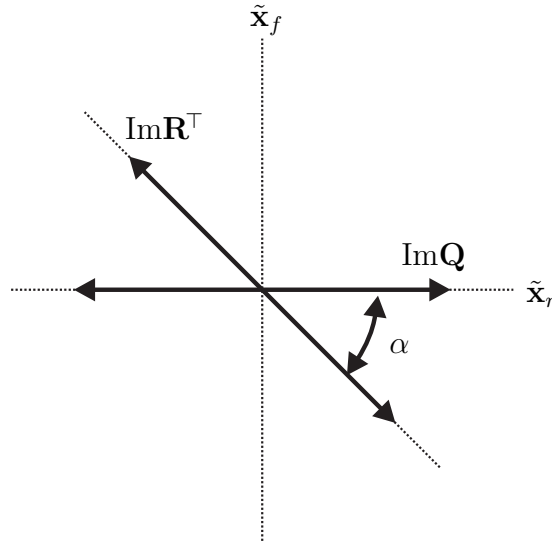


Figure 6.4: Controllability and observability of a linear one dimensional mobile robot with one landmark.

The controllable subspace has a basis of the form  $[q, 0]^T$ , clearly indicating that the only dimension in the state space that can be controlled is the one associated with the motion of the robot.

The observable subspace on the other hand, with basis  $[r, -r]^T$ , shows how the observed robot and landmark locations are fully correlated. This situation is portrayed graphically in Figure 6.4.

The unobservable subspace is the orthogonal complement of  $\text{Im}\mathbf{R}^T$ , and has a basis  $[r, r]^T$ . An expression for it was already derived from the analysis of the total Fisher information matrix and is given in Equation 6.11. The name  $\mathbf{E}_{\text{Ker}\mathbf{R}}$  indicates that it is a basis for the null space of  $\mathbf{R}$ .

The controllable and observable subspaces for the one-landmark one-dimensional CML problem span along a pair of non-orthogonal lines in  $\mathbb{R}^2$ ; with the consequence that the innovations only provide information for a fully correlated reconstruction of the state space.

A measure of the error incurred while trying to reconstruct the state  $\tilde{\mathbf{x}}_r$  from correlated observations is given by the angle between these two subspaces. And, it is evident from Figure 6.4 and Equations 6.22-6.23, that for the one landmark *monobot*, the angle

is

$$\alpha = \angle \text{Im}\mathbf{Q} \text{Im}\mathbf{R}^\top = \pi/4\text{rad}. \quad (6.24)$$

To see the physical interpretation of the angle  $\alpha$  in Equation 6.24, we ask the reader to analyze Figure 6.4 in detail once more. The error between an observed landmark and its prediction must lay along  $\text{Im}\mathbf{R}^\top$ . Such vector quantity, multiplied by its corresponding Kalman gain, is used to revise our estimates of both the robot and landmark locations. However, there is one direction of the state space which is not observed, the one orthogonal to  $\text{Im}\mathbf{R}^\top$  (along  $\text{Ker}\mathbf{R}$ ). The information for the revision of  $\tilde{\mathbf{x}}_r$  and  $\tilde{\mathbf{x}}_f$  along the direction orthogonal to  $\text{Im}\mathbf{R}^\top$  is missing. The angle  $\alpha$  indicates how close noise driven observations are from fully revising the robot part of the state space.

The immediate questions that come to our mind are: what happens if we add more landmarks to the environment? Will the vehicle and landmark location estimates improve or degrade? Will we be able to achieve an uncoupled reconstruction of the entire state space?

Surprisingly enough (and we believe, the entire research community which supports the EKF-based approaches to CML will find these observations crucial), the answer to the above questions is “improve” but “no”.

We will analyze now the two-landmark *monobot* case, and after that, we will introduce and prove a new closed form expression that links the degree of reconstruction in the EKF-CML algorithm to the number of landmarks used in the case of the *monobot*.

The observation Jacobian matrix for the two-landmark *monobot* case becomes

$$\mathbf{H}_x = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \quad (6.25)$$

and the observability matrix of our noise driven linear state system is now

$$\mathbf{R} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \quad (6.26)$$

A possible set of bases for the controllable and observable subspaces are

$$\mathbf{E}_{\text{Im}\mathbf{Q}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad (6.27)$$

$$\mathbf{E}_{\text{Im}\mathbf{R}^\top} = \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} \quad (6.28)$$

and the angle between these two subspaces can be computed as the smallest non null singular value of the product of their orthonormal bases [102, 164, 195].

$$\mathbf{E}_{\text{Im}\mathbf{Q}} = \mathbf{U}_Q \boldsymbol{\Sigma}_Q \mathbf{V}_Q^\top \quad (6.29)$$

$$\mathbf{U}_Q = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad (6.30)$$

$$\mathbf{E}_{\text{Im}\mathbf{R}^\top} = \mathbf{U}_R \boldsymbol{\Sigma}_R \mathbf{V}_R^\top \quad (6.31)$$

$$\mathbf{U}_R = \begin{bmatrix} -0.7071 & 0.4082 \\ 0 & -0.8165 \\ 0.7071 & 0.4082 \end{bmatrix} \quad (6.32)$$

$$\alpha = \angle \text{Im}\mathbf{Q} \text{Im}\mathbf{R}^\top \quad (6.33)$$

$$= \arccos \sigma_{\min}(\mathbf{U}_Q^\top \mathbf{U}_R) \quad (6.34)$$

$$= 163\pi/832\text{rad} \approx 0.6155\text{rad}. \quad (6.35)$$

The decrease in the angle between the controllable and observable subspaces obtained by adding one more landmark to the map, suggests that our measurement noise driven corrections to the map state estimate would reconstruct the vehicle localization estimate closer to the actual value of the vehicle pose.

Following this procedure we computed the value of  $\alpha$  for a three-landmark *monobot* model, further reducing to  $\alpha = \pi/6$ . And, as we add more landmarks to the map, the angle between the observable and controllable subspaces reduces monotonically. Figure 6.5 shows experimentally the decrease in  $\alpha$  as landmarks are added to the map state model. We are ready to introduce the following theorem:

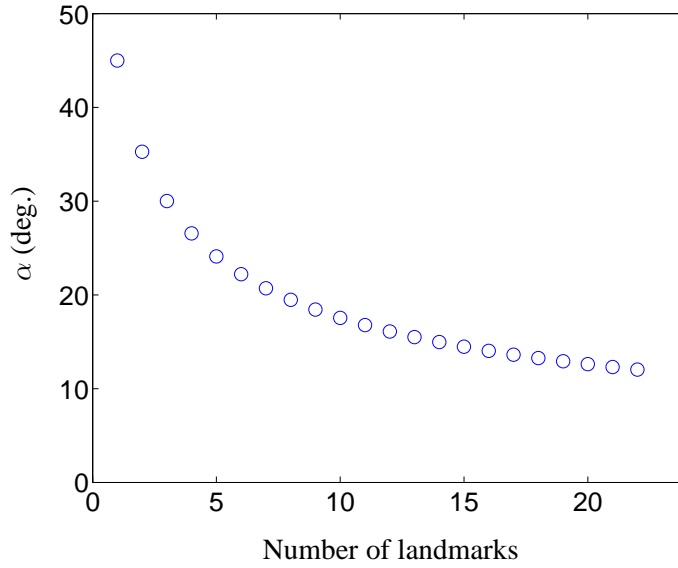


Figure 6.5:  $\angle \mathcal{R}(\mathbf{Q})\mathcal{R}(\mathbf{R}^\top)$ . Angle between the observable and controllable subspaces.

**Theorem 6.4.1.** *The angle between the controllable and observable subspaces in the case of a linear one-dimensional robot in the EKF-CML algorithm depends only on the total number of landmarks used ( $n$ ), and is given by*

$$\alpha = \arccos \sqrt{\frac{n}{n+1}}$$

*Proof.* Carefully following the pattern just described for the computation of  $\alpha$ ; in a generalization for the  $n$ -landmark *monobot* case, a set of possible bases for the controllable and observable subspaces are given by

$$\mathbf{E}_{\text{Im}\mathbf{Q}} = [\mathbf{q}] = \begin{pmatrix} 1 \\ \mathbf{0}_{n \times 1} \end{pmatrix} \quad (6.36)$$

and

$$\mathbf{E}_{\text{Im}\mathbf{R}^\top} = [\mathbf{r}_1 \ \dots \ \mathbf{r}_n] = \begin{pmatrix} \mathbf{1}_{1 \times n} \\ -\mathbf{I} \end{pmatrix} \quad (6.37)$$

respectively.

Moreover, let  $\mathbf{p}$  be the projection of  $\mathbf{E}_{\text{Im}\mathbf{Q}}$  onto  $\mathbf{E}_{\text{Im}\mathbf{R}^\top}$ .  $\mathbf{p}$  is easily computed as the sum of the individual projections of  $\mathbf{q}$  onto each element  $\mathbf{r}_i$  of the basis of the observable

space, i.e.,

$$\mathbf{p} = \sum_{i=1}^n \frac{\mathbf{q}^\top \mathbf{r}_i}{\mathbf{r}_i^\top \mathbf{r}_i} \mathbf{r}_i \quad (6.38)$$

Substituting the basis of the controllable space in Equation 6.38 gives the reduced expression

$$\mathbf{p} = \frac{1}{2} \sum_{i=1}^n \mathbf{r}_i = \frac{1}{2} \begin{bmatrix} n \\ -\mathbf{1}_{n \times 1} \end{bmatrix} \quad (6.39)$$

Finally, the angle between  $\mathbf{p}$  and  $\mathbf{q}$ , and consequently between the two subspaces, is

$$\alpha = \arccos \frac{\mathbf{p}^\top \mathbf{q}}{\|\mathbf{p}\| \|\mathbf{q}\|} = \arccos \sqrt{\frac{n}{n+1}} \quad (6.40)$$

□

As the number of landmarks grows, the observable subspace gets closer to the controllable part of the state space (the vehicle localization states).

$$\lim_{n \rightarrow \infty} \alpha = \lim_{n \rightarrow \infty} \arccos \sqrt{\frac{n}{n+1}} = 0 \quad (6.41)$$

It is unrealistic however, to have an infinite number of landmarks, and a compromise has to be made between the possibility of including as many landmarks as possible, and the amount of information that new observations give. Also one has to bear in mind that as we add more and more landmarks to the map, we will also introduce their associated measurement noise.

It has been argued that the performance of the CML algorithm would be enhanced by concentrating on fewer, better landmark observations [83]. And that is certainly true, little gain (little reduction of  $\alpha$ ) is made from going from 25 to 125 landmarks compared to the move from 1 to 5 or 5 to 25.

In Figures 6.6-6.7 we have plotted the results of using the original fully correlated approach to CML for a monobot that starts at location  $\mathbf{x}_{r,0|0} = -1m$ , and moves along a straight line with a temporal sinusoid trajectory returning to the same point after 100 iterations. Landmarks are located at  $\mathbf{x}_{f(i)} = 1m$ . A plant noise model proportional to the motion command with  $\alpha_d = 0.1$ , and a measurement noise model proportional to the distance from the sensor to the landmark with  $\beta_d = 0.1$  are used.

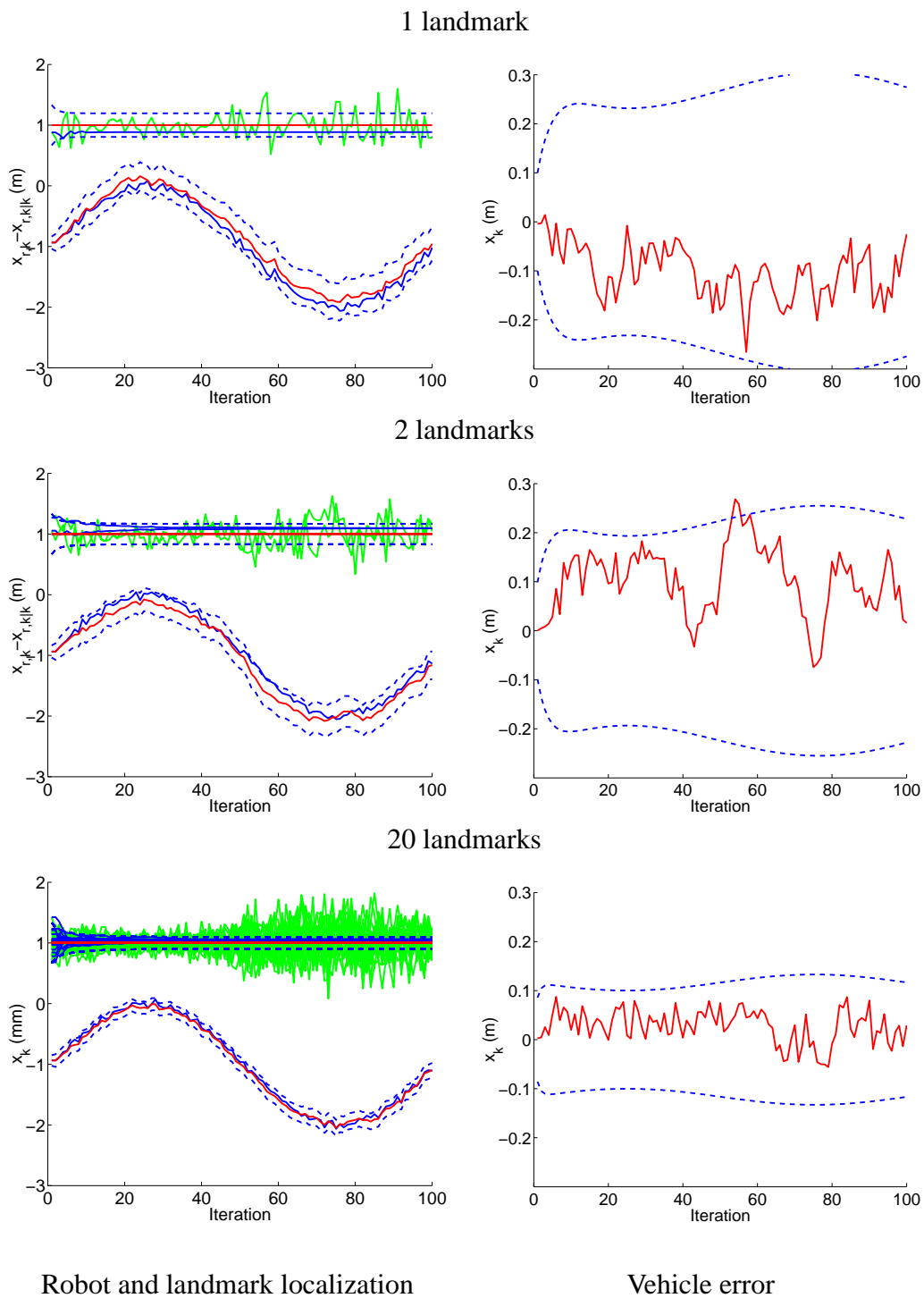


Figure 6.6: Full-covariance EKF CML for monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ .

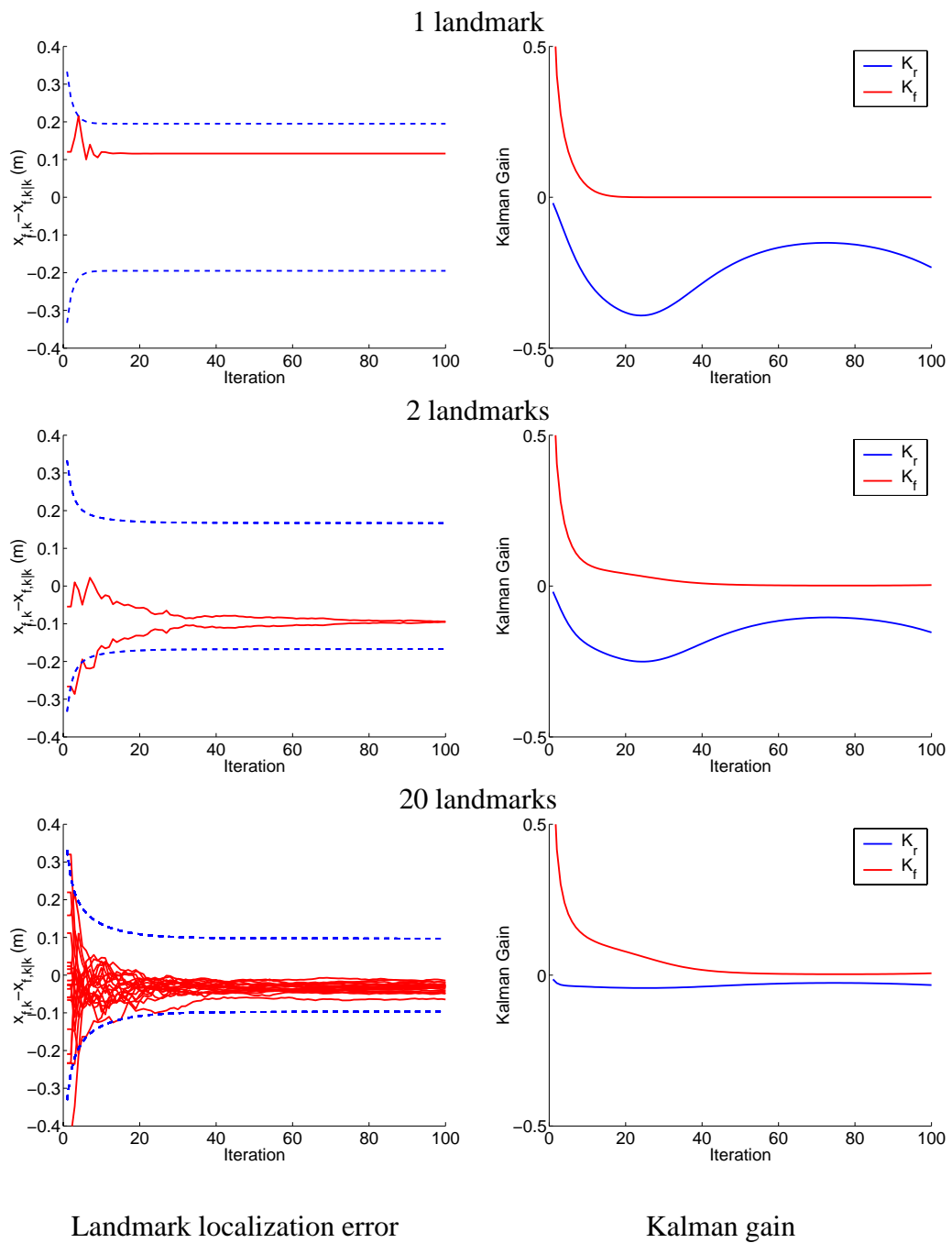


Figure 6.7: Full-covariance EKF CML for monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ .

In the first column in Figure 6.6, the red line indicates the true state trajectory, whereas the blue line shows the estimated one. The green line represents the actual landmark measurements, and the blue dotted line shows  $2\sigma$  bounds on the state estimate.

The effect of partial observability manifests itself in the dependence on the initial conditions. Note how both the vehicle and landmark mean localization errors do not converge to zero. Their steady state value is subject to the error incurred at the first observation. That is, the filter is unstable.

A Montecarlo simulation over 100 CML runs showed however filter unbiasedness, a property of optimal stochastic state estimation (Kalman filter). That is, the average landmark localization error over the entire set of simulations was still zero, thanks to the independence of the initial landmark measurement errors at each test run.

Partial controllability on the other hand, produces a zero Kalman gain for the revision of the landmark estimates. That is, after a few iterations the Kalman filter believes it has a perfectly accurate estimate of the landmark locations, contradictory to the localization error just described. The rate at which the landmark localization Kalman gain approaches zero is dictated by the rate of convergence of the system, i.e., the system's time constant (see the bibliographical notes).

Moreover, the steady state error for the robot and landmark localization is less sensitive to the initial conditions when a large number of landmarks are used. The reason is the same as for the Montecarlo simulation, the observations are independent, and their contribution averages at each iteration in the computation of the localization estimate. The results of the Montecarlo simulation are shown in Figure 6.8 depicting the effect of the increase in the number of landmarks on the average vehicle localization error.

Theorem 6.4.1 about the amount of reconstruction possible in EKF-CML with respect to the number of landmarks applies only to the simple *monobot* case. We will concentrate our attention now on a more realistic case, a planar mobile vehicle.

## 6.5 The planar robot

We will show now how the reconstructibility issues presented for the linear and one-dimensional robot of the previous section, nicely extend when studying more complicated platforms. We investigate now the case of the planar robot presented in Section 5.2.2, a nonlinear wheeled vehicle with three degrees of freedom, and an environment consisting of two-dimensional point landmarks located on the floor.

The dimensionality of the controllable subspace is  $\dim \mathbf{x}_r = 3$ , and for the specific



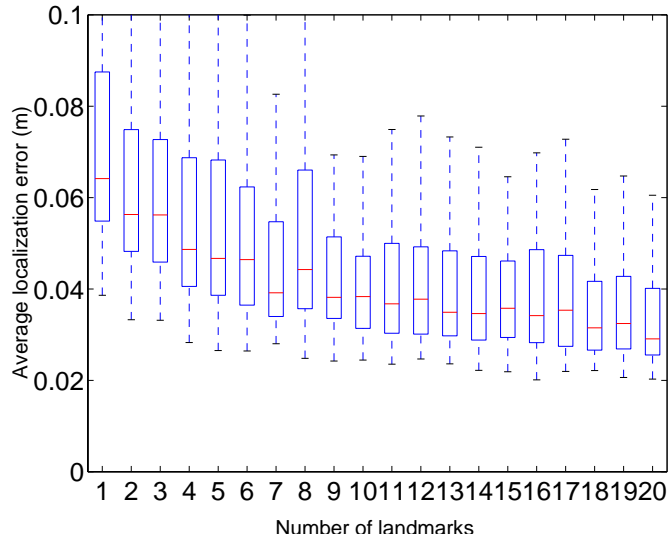


Figure 6.8: Reduction of the average monobot localization error  $\mathbf{x}_{r,k} - \mathbf{x}_{r,k|k}$  with respect to the number of landmarks used. The results correspond to a Montecarlo simulation over 100 CML runs. The dotted lines show the extent of the data for the entire set of runs, and the boxes contain marks at the lower, median and upper quartile.

case in which only one landmark is available, a basis for the controllable subspace is simply

$$\mathbf{E}_{\text{ImQ}} = \begin{pmatrix} \mathbf{I} \\ \mathbf{0}_{2 \times 3} \end{pmatrix}$$

The dimensionality of the observable space is, for this particular configuration, rank  $\mathbf{R} = 3$ . This last result is easily verified with simple symbolic manipulation of the specific expression for the state model from Section 5.2. Furthermore, one possible basis for  $\text{Im}\mathbf{R}^\top$  is

$$\mathbf{E}_{\text{ImR}^\top} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

Our 3dof robot along with a 2d landmark form a map state space in  $\mathbb{R}^5$ , and the null

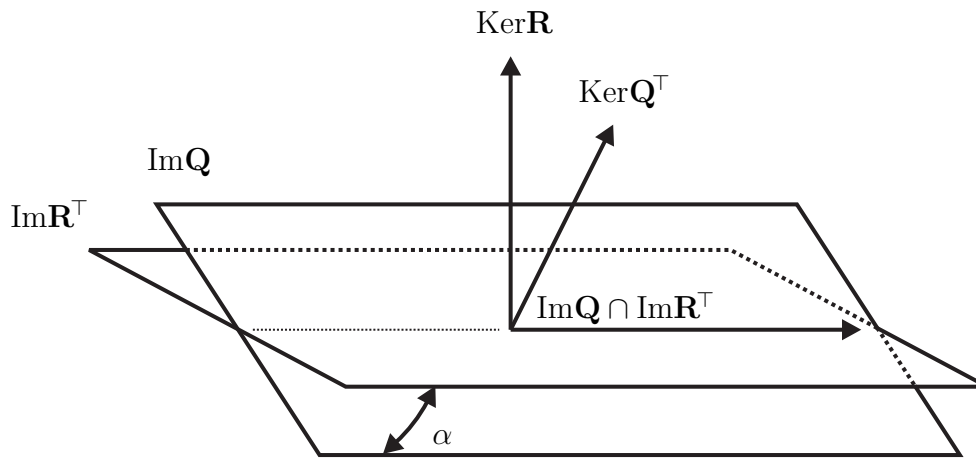


Figure 6.9: Controllability and observability subspaces.

space of  $\mathbf{R}$  (the unobservable subspace) is spanned by

$$\mathbf{E}_{\text{Ker}\mathbf{R}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Recall from the basic properties of the four fundamental subspaces of linear algebra that  $\text{Im}\mathbf{R}^\top + \text{Ker}\mathbf{R} = \mathbb{R}^5$ ,  $\text{Im}\mathbf{R}^\top \cap \text{Ker}\mathbf{R} = \emptyset$ . The only independently observable state is the one associated to the robot orientation  $\theta$ . The other four states, the Cartesian coordinates of the robot and landmark locations depend linearly on each other.

Moreover, even when  $\text{Im}\mathbf{Q}$  and  $\text{Im}\mathbf{R}^\top$  both span  $\mathbb{R}^3$ , we see that the inequality  $\text{Im}\mathbf{Q} \neq \text{Im}\mathbf{R}^\top$  still holds, as in the case of the *monobot*. That is, the observable and controllable subspaces for the one-landmark 3dof-robot CML problem correspond to different three-dimensional subspaces in  $\mathbb{R}^5$ ; and, their intersection represents the only fully controllable and observable state, i.e., the robot orientation. A pictorial representation of this situation is depicted in Figure 6.9, and once more, a measure of the reconstruction error incurred when estimating the vehicle pose from correlated observations is given by the angle between these two subspaces.

Resorting again to the singular value decomposition for the computation of a pair of orthonormal bases for  $\text{Im}\mathbf{Q}$  and  $\text{Im}\mathbf{R}^\top$ , we have that for the one-landmark planar robot

case

$$\mathbf{E}_{\text{Im}\mathbf{Q}} = \mathbf{U}_Q \mathbf{\Sigma}_Q \mathbf{V}_Q^\top \quad (6.42)$$

$$\mathbf{U}_Q = \begin{pmatrix} \mathbf{I} \\ \mathbf{0}_{2 \times 3} \end{pmatrix} \quad (6.43)$$

$$\mathbf{E}_{\text{Im}\mathbf{R}^\top} = \mathbf{U}_R \mathbf{\Sigma}_R \mathbf{V}_R^\top \quad (6.44)$$

$$\mathbf{U}_R = \begin{pmatrix} -0.7071 & 0 & 0 \\ 0 & -0.7071 & 0 \\ 0 & 0 & -1 \\ 0.7071 & 0 & 0 \\ 0 & 0.7071 & 0 \end{pmatrix} \quad (6.45)$$

and, as before,

$$\alpha = \angle \text{Im}\mathbf{Q} \text{Im}\mathbf{R}^\top \quad (6.46)$$

$$= \arccos \sigma_{\min}(\mathbf{U}_Q^\top \mathbf{U}_R) \quad (6.47)$$

$$= \pi/4\text{rad} \quad (6.48)$$

Note the equivalence of  $\alpha$  in Equations 6.24 and 6.48. The immediate inference behind this result is that neither the nonlinearities of the planar mobile robot platform, nor the change in the dimensionality of the model, are related to the amount of reconstruction possible when using a fully correlated CML model.

We should elaborate more on this, because if we can show this result to hold for a map with more landmarks, we would have a powerful argument to defend our hypothesis that the number of landmarks used in CML is directly related to the average localization error in the form of Figures 6.5 and 6.8.

Extending now our study to the two-landmark case; possible orthonormal bases for the controllable and observable subspaces as computed with the aforementioned *svd* are given by

$$\mathbf{U}_Q = \begin{pmatrix} \mathbf{I} \\ \mathbf{0}_{4 \times 3} \end{pmatrix}$$

$$\mathbf{U}_R = \begin{pmatrix} -0.4082 & 0 & 0 & 0 & 0.7071 \\ 0 & 0.4082 & 0 & 0.7071 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ -0.4082 & 0 & 0 & 0 & -0.7071 \\ 0 & 0.4082 & 0 & -0.7071 & 0 \\ 0.8165 & 0 & 0 & 0 & 0 \\ 0 & -0.8165 & 0 & 0 & 0 \end{pmatrix}$$

and once more, the angle between them is

$$\alpha = \arccos \sigma_{\min}(\mathbf{U}_Q^\top \mathbf{U}_R) \quad (6.49)$$

$$\alpha = 163\pi/832. \quad (6.50)$$

Similarly, for a three-landmark model,  $\alpha = \pi/6$ , and as we add more and more landmarks to the environment, the angle between the controllable and observable subspaces reduces monotonically, in exactly the same manner as in the case of the *monobot*. We are now ready to formulate the second theorem in this Chapter

**Theorem 6.5.1.** *The angle between the controllable and observable subspaces in the case of a nonlinear planar robot with 3 degrees of freedom in the EKF-CML algorithm, depends only on the total number of landmarks used ( $n$ ), and is given by*

$$\alpha = \arccos \sqrt{\frac{n}{n+1}}$$

*Proof.* The key to the proof is in Figure 6.9. Notice that thanks to the orthogonality of the four fundamental subspaces, the angle between the observable and controllable subspaces is exactly the same as the angle between their complementary subspaces. This is,

$$\alpha = \angle \text{Ker} \mathbf{Q}^\top \text{Ker} \mathbf{R} \quad (6.51)$$

The controllable subspace has a fixed rank of size  $\dim \mathbf{x}_r = 3$ , regardless of the number of landmarks; and the size of the basis for the observable subspace would depend on  $n$ . Now, the roles are reversed. The dimension of  $\mathbf{E}_{\text{Ker} \mathbf{Q}^\top}$  grows with respect to the number of landmarks, but maintains a very simple form

$$\mathbf{E}_{\text{Ker} \mathbf{Q}^\top} = \begin{bmatrix} \mathbf{0}_{3 \times 2n} \\ \mathbf{I} \end{bmatrix} \quad (6.52)$$

The null complement of the observable subspace on the other hand has a fixed number of columns (just two), and it can be easily shown by inspection that

$$\mathbf{E}_{\text{KerR}} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0}_{1 \times 2} \\ \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix}_{(3+2n) \times 2} \quad (6.53)$$

These are precisely the directions along which our state space is unobservable. Clearly showing that in the EKF-CML model, the Cartesian coordinates of the robot and landmark locations are all fully correlated.

The angle between these two subspaces is again, given by the smallest singular value of an orthonormalized version of the product  $\mathbf{E}_{\text{KerQ}}^\top \mathbf{E}_{\text{KerR}}$ , in which  $\mathbf{E}_{\text{KerQ}}^\top = \mathbf{U}_Q \boldsymbol{\Sigma}_Q \mathbf{V}_Q^\top$ , and  $\mathbf{E}_{\text{KerR}} = \mathbf{U}_R \boldsymbol{\Sigma}_R \mathbf{V}_R^\top$ . This is,

$$\mathbf{U}_Q^\top \mathbf{U}_R = \frac{1}{\sqrt{n+1}} \begin{bmatrix} \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix}_{2 \times 2n} \quad (6.54)$$

$$\alpha = \arccos \sigma_{\min}(\mathbf{U}_Q^\top \mathbf{U}_R) \quad (6.55)$$

$$\alpha = \arccos \sqrt{\frac{n}{n+1}} \quad (6.56)$$

□

Asymptotic stability of the KF means that its solution will gradually become insensitive to its initial conditions. One can see that observability plays a role because, if there are sufficient landmark measurements, the true localization and landmark estimates will be well approximated. Also, controllability will play a role because if the system is not controllable in some modes (as it is), then any number of observations cannot help damp the analysis errors.

## 6.6 Complete observability

In Section 6.2 we characterized the unobservable subspace in CML as the subspace spanned by the null eigenvectors of the total Fisher information matrix. Furthermore, we showed in Sections 6.3-6.5 how the unobservable part of the state space is precisely a linear combination of the landmark and robot pose estimates.

In order to gain full observability we propose to extend the measurement model doing away with the constraint imposed by full correlation. We present two techniques to achieve this. One is to let one landmark serve as a fixed global reference, with its localization uncertainty independent to the vehicle pose.

The second proposed technique is the addition of a fixed external sensor, such as a camera or a GPS, that can measure the vehicle location at all times, independent of the landmark estimates.

Both techniques are based essentially on the same principle. Full observability requires an uncorrelated measurement Jacobian, or equivalently, a full rank Fisher information matrix.

We next present, without loss of generality, the extensions to the monobot CML model in order to obtain full observability.

### 6.6.1 A fixed global reference

The plant model is left untouched, i.e., (from Equations 5.1 and 5.49)

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{u}_k + \mathbf{v}_k \quad (6.57)$$

The measurement model takes now the form

$$\begin{bmatrix} z_k^{(0)} \\ \mathbf{z}_k \end{bmatrix} = \begin{bmatrix} -1 & \mathbf{0}_{1 \times n} \\ -\mathbf{1}_{n \times 1} & \mathbf{I} \end{bmatrix} \mathbf{x} + \begin{bmatrix} w_k^{(0)} \\ \mathbf{w}_k \end{bmatrix} \quad (6.58)$$

One of the observed landmarks is to be taken as a global reference at the world origin. No map state is needed for it. The zero-th superscript in the measurement vector is used for the consistent indexing of landmarks and observations with respect to the original model. It can be easily shown that the observability matrix for this new model is full rank.

The innovation covariance matrix for the augmented system  $\mathbf{S}_{FO,k}$  is of size  $n+1 \times n+1$ , and its inverse can be decomposed in

$$\mathbf{S}_{FO,k}^{-1} = \begin{bmatrix} \varsigma_{FO,00} & \varsigma_{FO,01} & \cdots & \varsigma_{FO,0n} \\ \varsigma_{FO,01} & & & \\ \vdots & & \hat{\mathbf{S}}_k^{-1} & \\ \varsigma_{FO,0n} & & & \end{bmatrix} \quad (6.59)$$

with  $\varsigma_{FO,ij}$  the  $ij$ -th entry in  $\mathbf{S}_{FO,k}^{-1}$ ,  $\boldsymbol{\varsigma}_{FO} = [\sum \varsigma_{FO,1i}, \dots, \sum \varsigma_{FO,1i}]$ , and  $\hat{\mathbf{S}}_k^{-1}$  its sub-matrix associated to the landmarks that are under estimation (excluding the anchor observation).

The  $k$ -th element of the Fisher information matrix sum is

$$\mathbf{J}_{FO,k} = \begin{bmatrix} \sum \sum \varsigma_{FO,ij} & -\boldsymbol{\varsigma}_{FO} \\ -\boldsymbol{\varsigma}_{FO}^\top & \hat{\mathbf{S}}_k^{-1} \end{bmatrix} \quad (6.60)$$

Unlike Equation 6.8, this form of the Fisher information matrix is full rank. Moreover, from the properties of positive definite matrices, if  $\mathbf{J}_{FO,k}$  is positive definite, the entire sum that builds up  $\mathbf{J}_{FO}$  is also positive definite.

Figures 6.10 and 6.11 show the results of applying a full observability to the same monobot model as the one portrayed in Figures 6.6-6.7. Note how the steady state (robot pose and landmark locations) is now unbiased with respect to the initial landmark estimates.

## 6.6.2 An external sensor

Instead of using one of the landmarks as a global reference, one could also use a fixed sensor to measure the position of the robot. For example, by positioning a camera that observes the vehicle at all times. For such cases, the monobot measurement model may take the form

$$\begin{bmatrix} z_k^{(0)} \\ \mathbf{z}_k \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}_{1 \times n} \\ -\mathbf{1}_{n \times 1} & \mathbf{I} \end{bmatrix} \mathbf{x} + \begin{bmatrix} w_k^{(0)} \\ \mathbf{w}_k \end{bmatrix} \quad (6.61)$$

The characteristics on the observability matrix, and the Fisher information matrix, are exactly the same as for the previous case. This new model is once more, fully observable. Figures 6.12 and 6.13 show the results of using an external sensor to measure the vehicle pose. The results are theoretically equivalent to the previous case. The choice of one technique over the other one would depend on the availability of such external sensor, and on its measurement noise covariance characteristics.

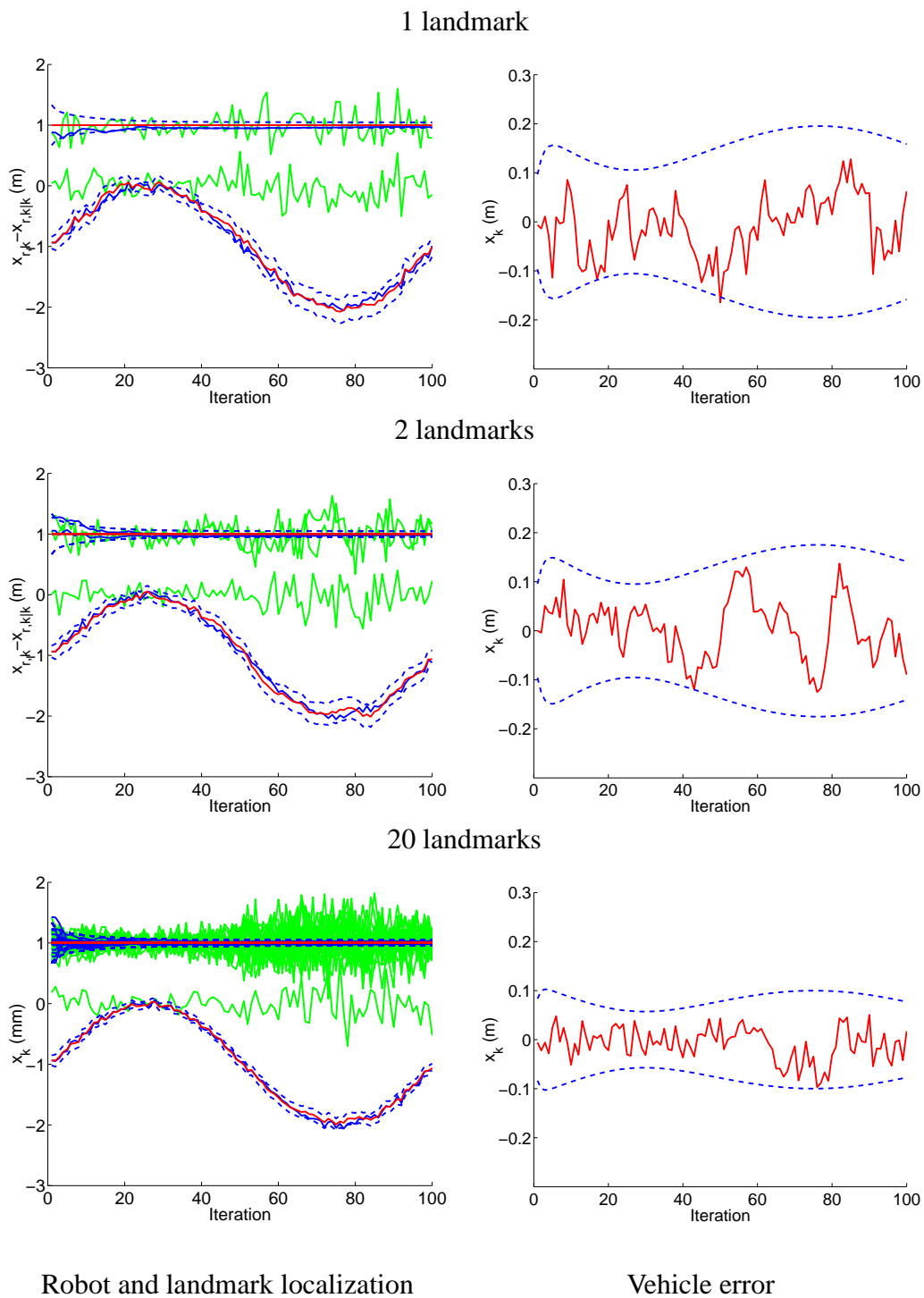


Figure 6.10: Full-covariance fully observable CML for a monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ . The global reference is observed at the origin, also with  $\beta_d = 0.1$ .



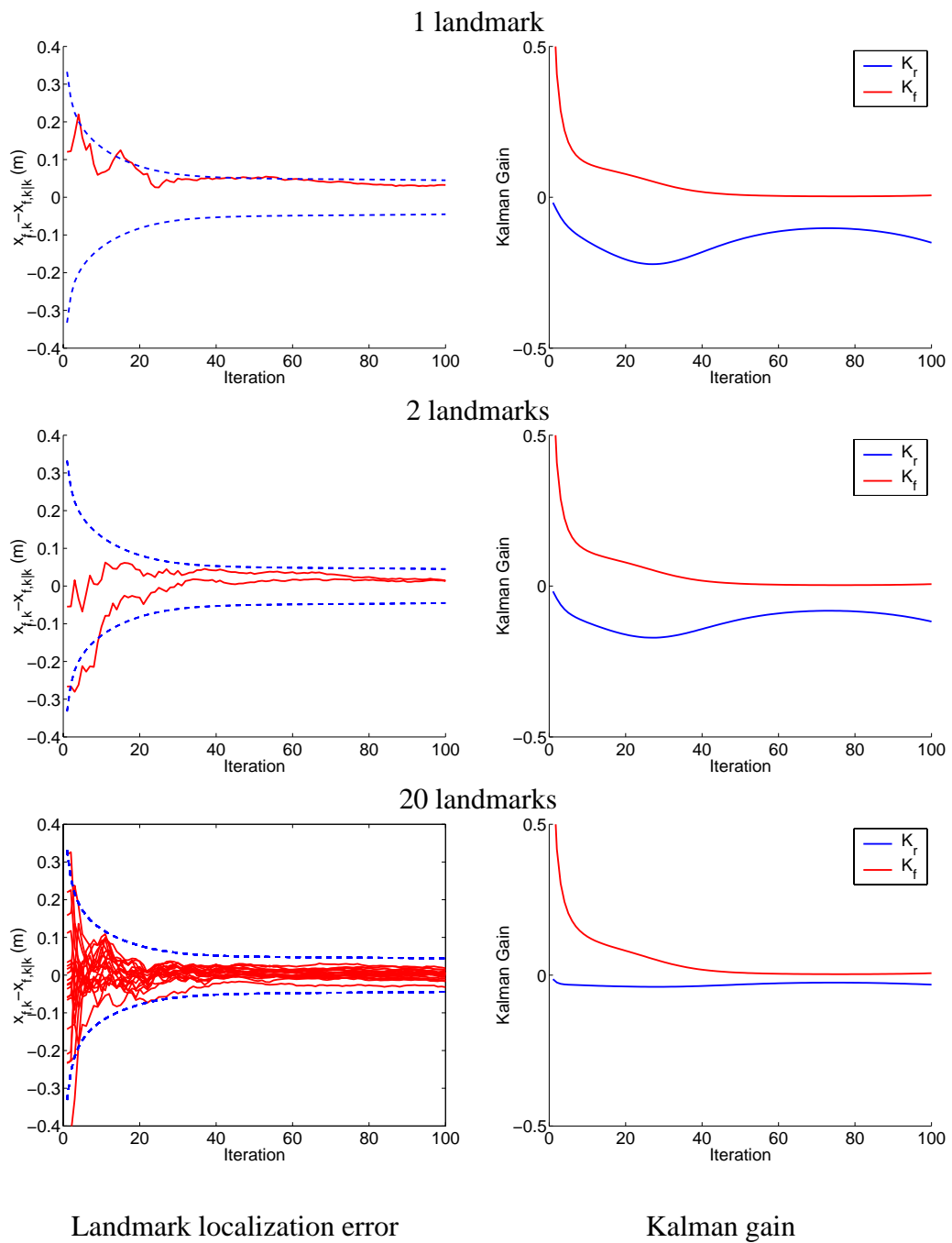


Figure 6.11: Full-covariance fully observable CML for a monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ . The global reference is observed at the origin, also with  $\beta_d = 0.1$ .

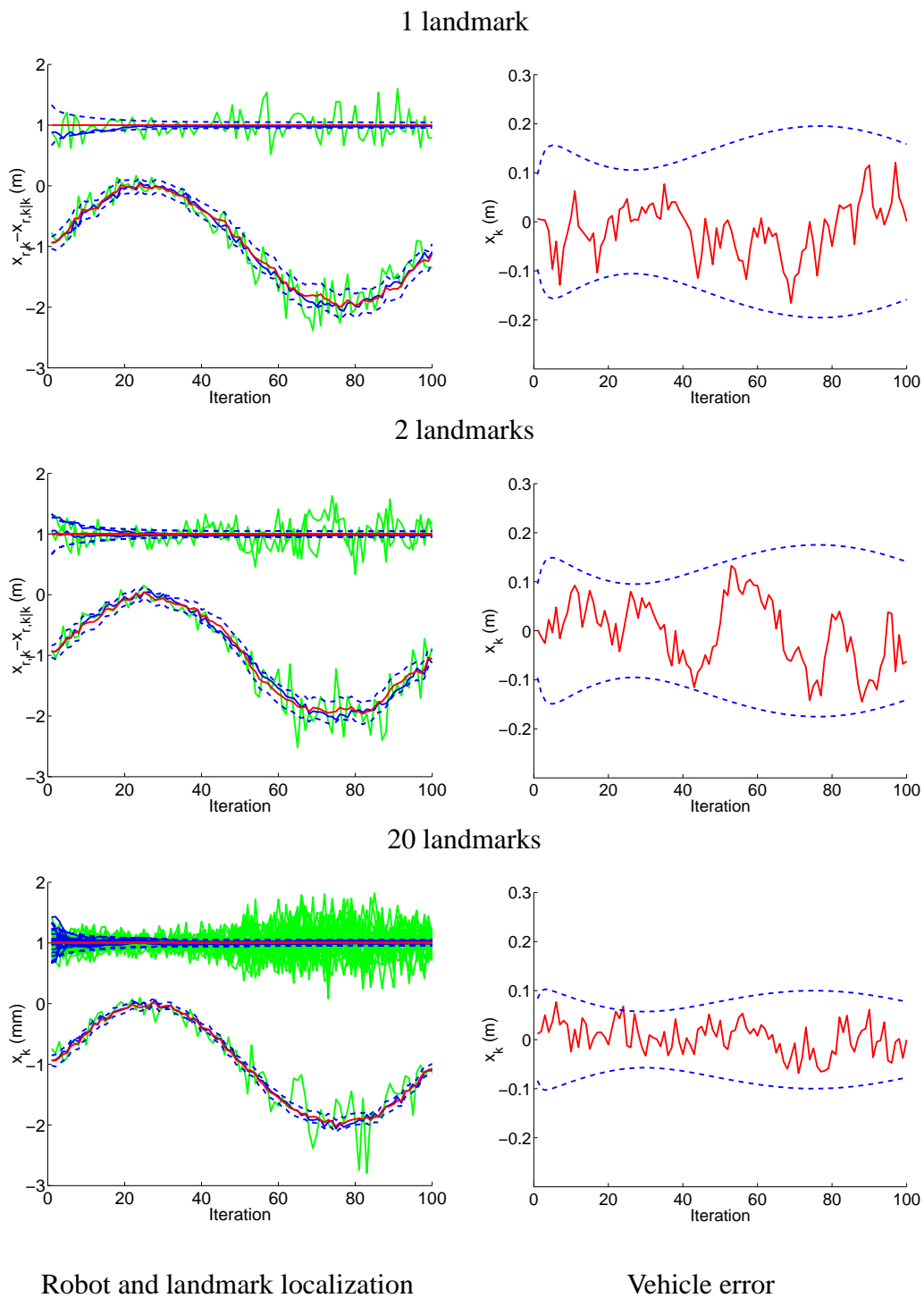


Figure 6.12: Full-covariance fully observable CML for a monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ . A fixed external sensor is used for the measurement of the vehicle pose, also with  $\beta_d = 0.1$ .

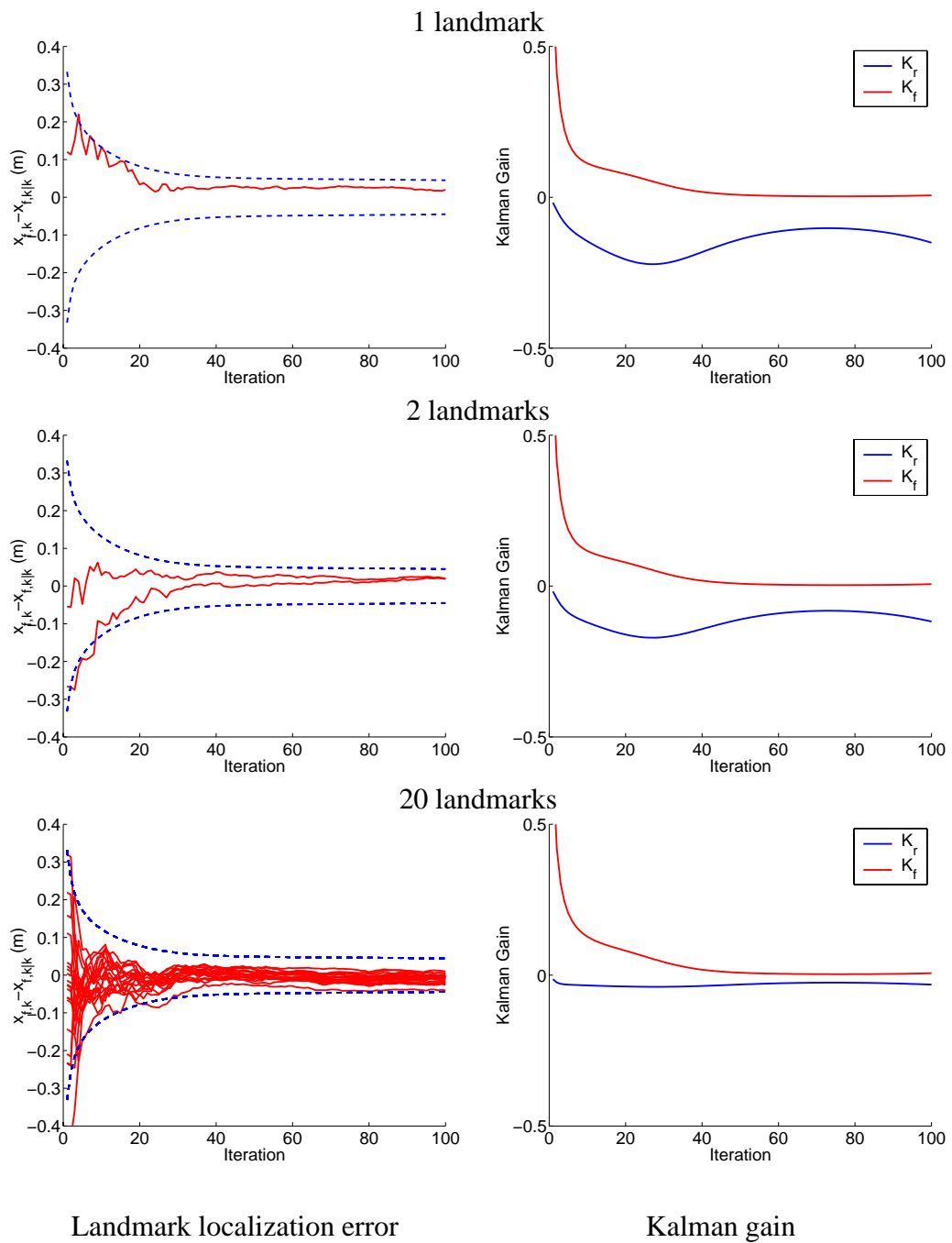


Figure 6.13: Full-covariance fully observable CML for a monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ . A fixed external sensor is used for the measurement of the vehicle pose, also with  $\beta_d = 0.1$ .

### 6.6.3 Planar vehicle

The results from the previous section are easily extensible to more complicated vehicle models. For example, the measurement model of a global reference fixed at the origin, for the nonlinear vehicle from Section 5.2.2 is (from Equation 5.61)

$$\mathbf{h}^{(0)} = -\mathbf{R}^\top \mathbf{t} + \mathbf{w}^{(0)} \quad (6.62)$$

and its corresponding Jacobian is

$$\mathbf{H}_x^{(0)} = [ -\mathbf{R}^\top \quad -\dot{\mathbf{R}}^\top \mathbf{t} \quad \mathbf{0}_{2 \times 2n} ] \quad (6.63)$$

The case of the external sensor is even simpler, the corresponding equations are

$$\mathbf{h}^{(0)} = \mathbf{t} + \mathbf{w}^{(0)} \quad (6.64)$$

$$\mathbf{H}_x^{(0)} = [ \mathbf{I} \quad \mathbf{0}_{2 \times (2n+1)} ] \quad (6.65)$$

In both cases, the symbolic manipulation of Equations 6.63 and 6.65 with a commercial algebra package, produced full rank observability matrices. That is, for the planar mobile robot platform used, only one two-dimensional global reference, or the use of a sensor that can measure the  $xy$  position of the robot, are sufficient to attain full observability in CML.

## 6.7 Controllability

We have seen in Section 6.3 that in the standard CML model, the only controllable states are the ones associated to the vehicle location estimate.

Equation 5.13 assumes the landmarks are fixed elements, for which no process noise is considered. Therefore, their associated noise covariance (its determinant) will asymptotically tend to zero (see Section 5.1.5, Equation 5.26). The filter gain for the landmark states will also tend to zero.

Having a positive definite covariance reflects the belief that there is a perfectly accurate estimate of some states (or linear combinations of them, the ones associated to the null eigenvalues of  $\mathbf{P}$ ). This is, again, because there is no process noise entering these states, and the controllability condition described at the beginning of Section 6.3 pertaining the solution of the Riccati Equation 6.1 is not satisfied.

Since the fully correlated Kalman filter will not in general, yield consistent estimates

for the map states (partial observability and divergence due to nonlinearities produce biased estimates); the situation where the map state covariance (its determinant) tends to zero is undesirable. It will lead in practice to smaller filter-computed variances than the estimated errors.

The situation can be remedied to some extent, by assuming the existence of artificial pseudo-noise associated to the landmark state estimates. This is, we can replace Equation 5.13 with

$$\begin{bmatrix} \mathbf{x}_{r,k+1} \\ \mathbf{x}_f \end{bmatrix} \approx \begin{bmatrix} \mathbf{x}_{r,k+1|k} \\ \mathbf{x}_{f,k|k} \end{bmatrix} + \begin{bmatrix} \mathbf{F}_{\mathbf{x}_r} & \\ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{r,k|k} \\ \tilde{\mathbf{x}}_{f,k|k} \end{bmatrix} + \begin{bmatrix} \mathbf{F}_{\mathbf{v}_r} & \\ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{r,k} \\ \mathbf{v}_{f,k} \end{bmatrix} \quad (6.66)$$

The landmark pseudo-noise  $\mathbf{v}_{f,k}$  is assumed zero mean and white. Any positive definite covariance of this process noise will prevent the filter calculated covariance for the landmark states from converging to zero. The new process noise covariance is computed as the sum of the vehicle model covariance plus that of the pseudo noise.

$$\mathbf{V}_{C,k} = \begin{bmatrix} \mathbf{V}_{r,k} & \\ & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \\ & \mathbf{V}_f \end{bmatrix} \quad (6.67)$$

The positive definite process noise covariance  $\mathbf{V}_{C,k}$  is substituted for  $\mathbf{V}_k$  in Equation 5.17. Figures 6.14 and 6.15 show the results of adding an artificial landmark state covariance to the fully observable model from Figures 6.10 and 6.13

Having a fully observable CML model with artificial process noise makes feasible the computation of the Cramer Rao lower bound. The green dotted lines show the vehicle and landmark associated components of  $\mathbf{J}^{-1}$ . Moreover, Figure 6.16 shows the numerical evaluation of the eigenvalues for the matrix difference  $\mathbf{P}_{k|k} - \mathbf{J}^{-1}$ . When both eigenvalues are positive, the matrix difference is positive definite. That is, we can only trust our second order statistic  $\mathbf{P}_{k|k}$  after a certain number of iterations have elapsed; only after enough information has been input to the filter in the form of innovations.

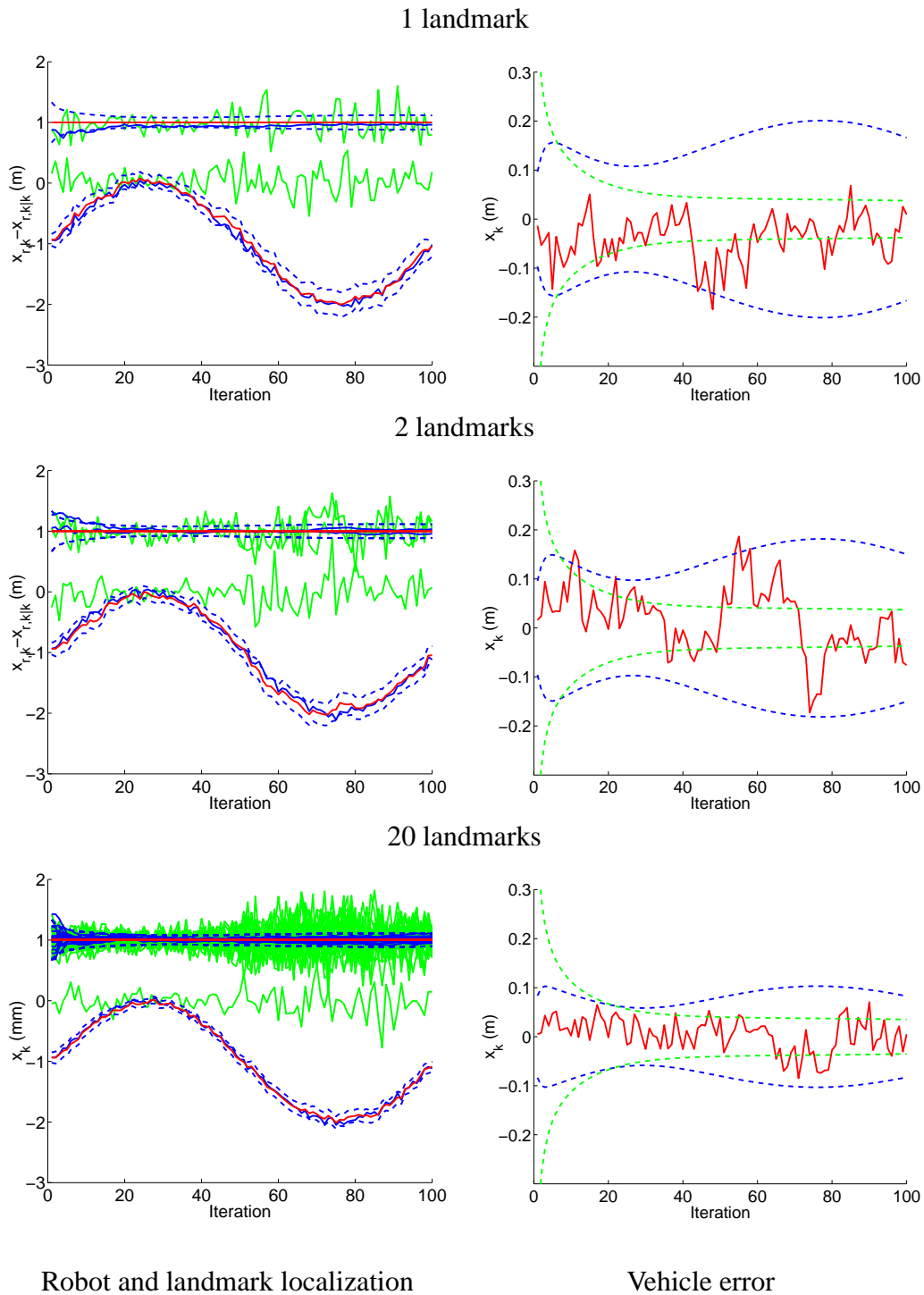


Figure 6.14: Full-covariance fully observable CML with artificial process noise for a monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ . A fixed external sensor is used for the measurement of the vehicle pose, also with  $\beta_d = 0.1$ . A pseudo-noise covariance ( $\mathbf{V}_f = 0.05^2 \mathbf{I}$ ) is used to prevent the landmark Kalman gains from converging to zero.

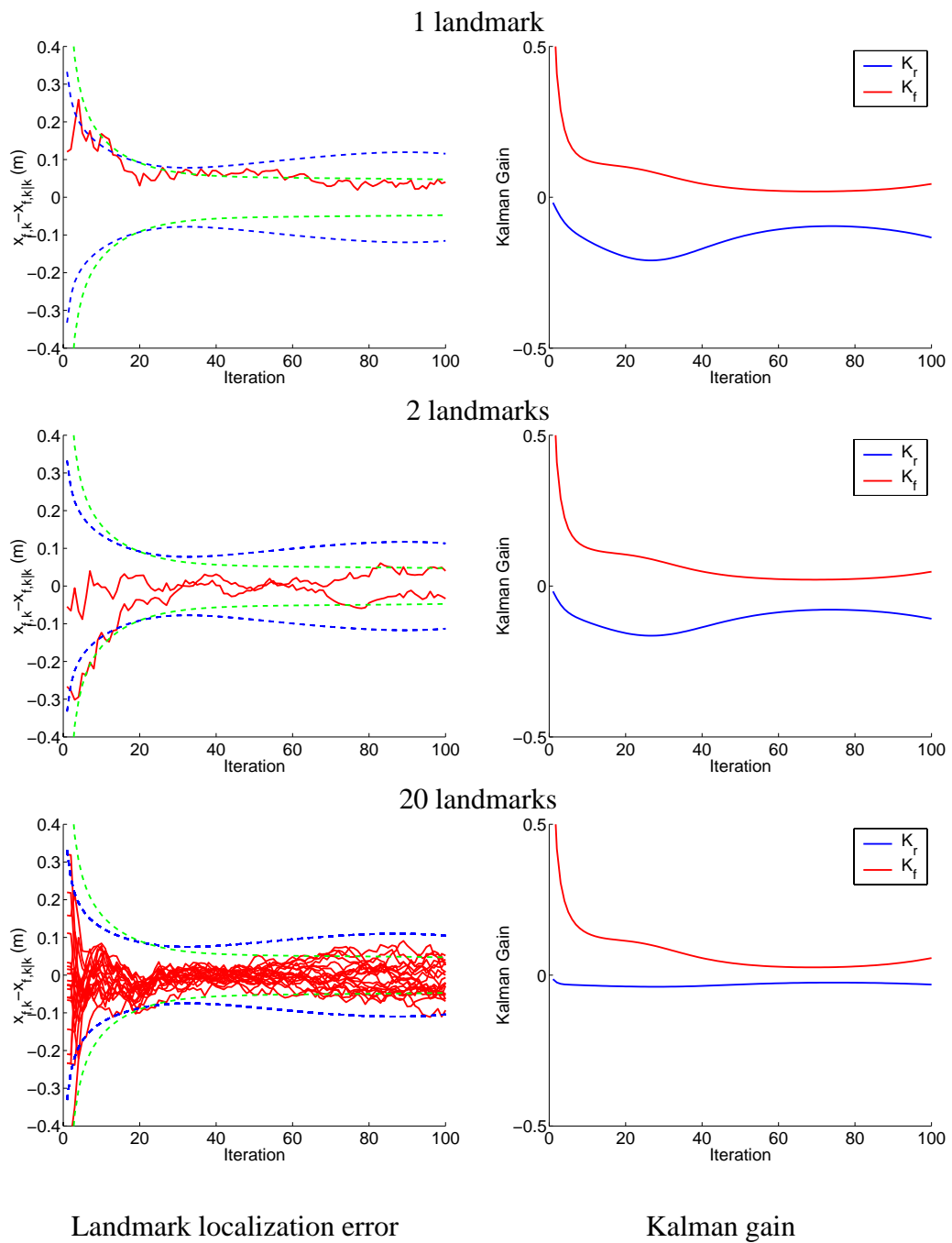


Figure 6.15: Full-covariance fully observable CML with artificial process noise for a monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ . A fixed external sensor is used for the measurement of the vehicle pose, also with  $\beta_d = 0.1$ . A pseudo-noise covariance ( $\mathbf{V}_f = 0.05^2\mathbf{I}$ ) is used to prevent the landmark Kalman gains from converging to zero.

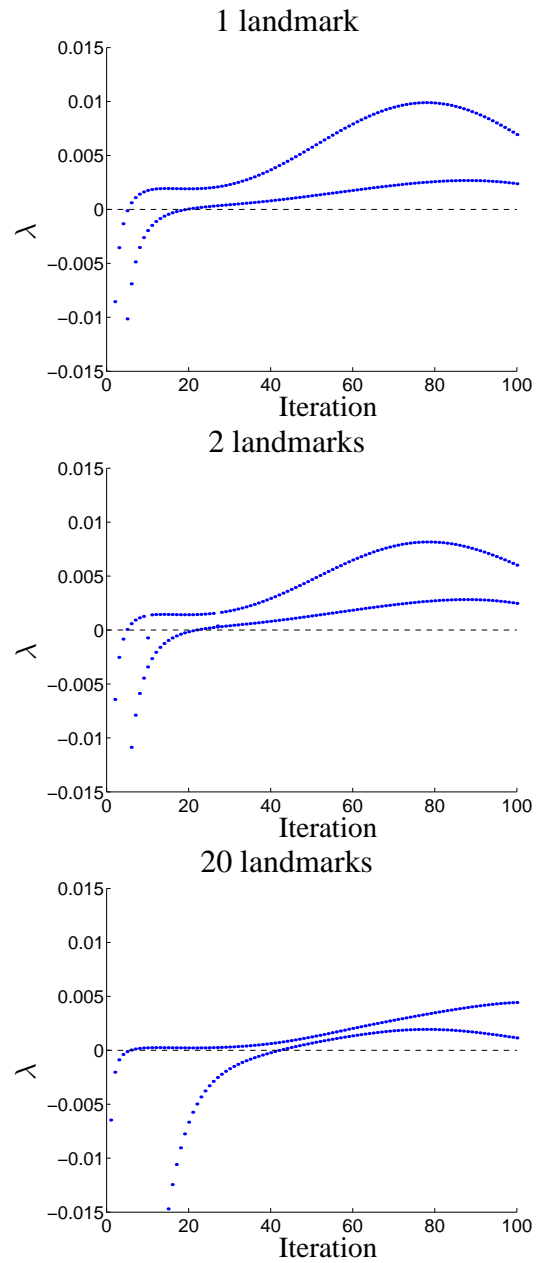


Figure 6.16: The Cramer Rao lower bound holds only when the eigenvalues of  $(\mathbf{P}_{k|k} - \mathbf{J}^{-1})$  are positive.



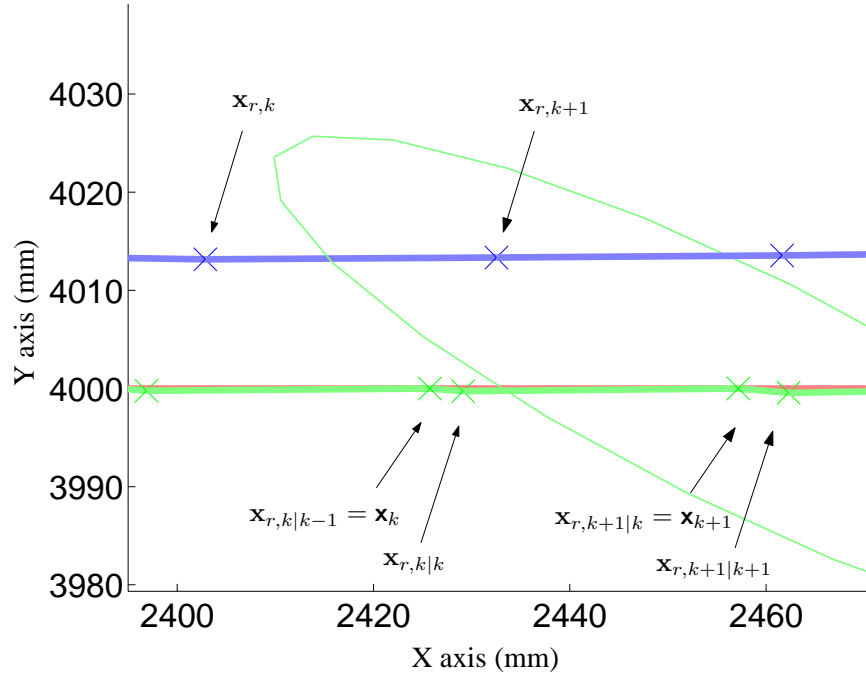


Figure 6.17: Full-covariance EKF CML for a path with 1000 steps and 10 landmarks, zoomed view.

## 6.8 Optimal control

To keep the vehicle follow a desired path, the motion commands  $\mathbf{u}_k$  must be modified at run time to correct the deviation of the state estimates  $\mathbf{x}_{r,k|k}$  from the desired trajectory  $\mathbf{x}_k$ . A simple control law for doing so is given by

$$\mathbf{u}_k = \mathbf{x}_{k+1} - \mathbf{x}_{r,k|k} \quad (6.68)$$

The result is that the a priori robot pose estimate  $\mathbf{x}_{r,k+1|k}$  is always coincident with the path plan. The situation is shown in Figure 6.17, where we have zoomed into a particular spot in a particular test run for the planar robot. The desired path plan is along the  $4000\text{mm}$  level in the  $y$  axis. The green line represents the estimated robot trajectory; whereas the blue line indicates the actual robot trajectory.

It is likely however, to have limited vehicle dynamics; in the sense that it may be unfeasible to exert a given motion command as computed by Equation 6.68. The idea

is to modify the control input in order to obtain realistic motion commands, and at the same time minimize the expected state error.

The stochastic optimization of the linear system in Equations 5.13-5.14 is usually referred as an LQG (Linear Quadratic Gaussian) problem. The LQG can be optimally solved with the estimates obtained from the Kalman filter [4]. That is, the solution to the LQG problem is dual to that of the optimal observer. See Figure 6.18.

The new control input  $\hat{\mathbf{u}}$  should be manipulated to force the system to behave in a desired manner, for example, by minimizing an average quadratic performance index of the form

$$I = E \left[ \sum_{i=0}^{k-1} (\tilde{\mathbf{x}}_i^\top \mathbf{Q}_1 \tilde{\mathbf{x}}_i + \tilde{\mathbf{u}}_i^\top \mathbf{Q}_2 \tilde{\mathbf{u}}_i) \right] \quad (6.69)$$

where  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  are *psd* matrices.

The optimal control that minimizes Equation 6.69 is given by [79]

$$\hat{\mathbf{u}}_k = -\mathbf{L}_k \tilde{\mathbf{x}}_{k|k} \quad (6.70)$$

with

$$\mathbf{L} = (\mathbf{Q}_2 + \mathbf{B}^\top \mathbf{T} \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{T} \mathbf{F} \quad (6.71)$$

and

$$\mathbf{T} = \mathbf{Q}_1 + \mathbf{L}^\top \mathbf{Q}_2 \mathbf{L} + (\mathbf{F} - \mathbf{B} \mathbf{L})^\top \mathbf{T} (\mathbf{F} - \mathbf{B} \mathbf{L}) \quad (6.72)$$

Matrices  $\mathbf{L}$  and  $\mathbf{T}$  in the regulator are dual to the optimal observer matrices  $\mathbf{K}$  and  $\mathbf{S}$ , respectively.

Figures 6.19 - 6.21 show the results of applying optimal control to the fully observable CML model with artificial process noise from Figures 6.14 and 6.15.

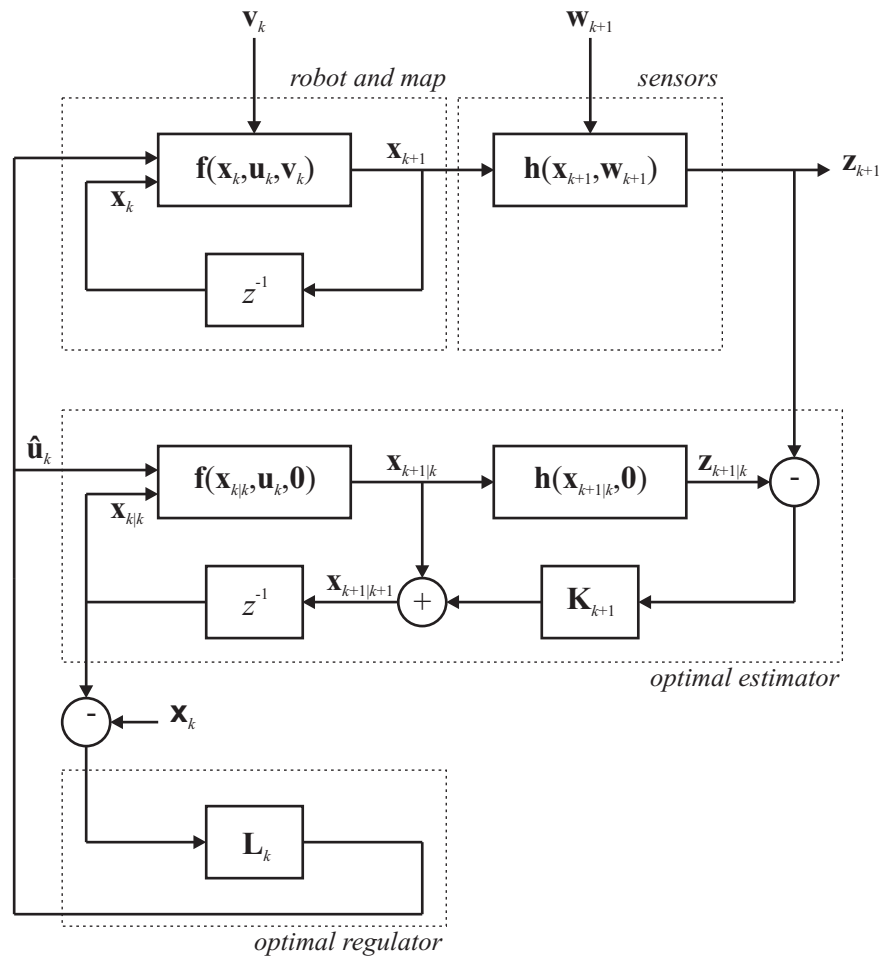


Figure 6.18: Structure of the optimal state estimator and regulator in CML.

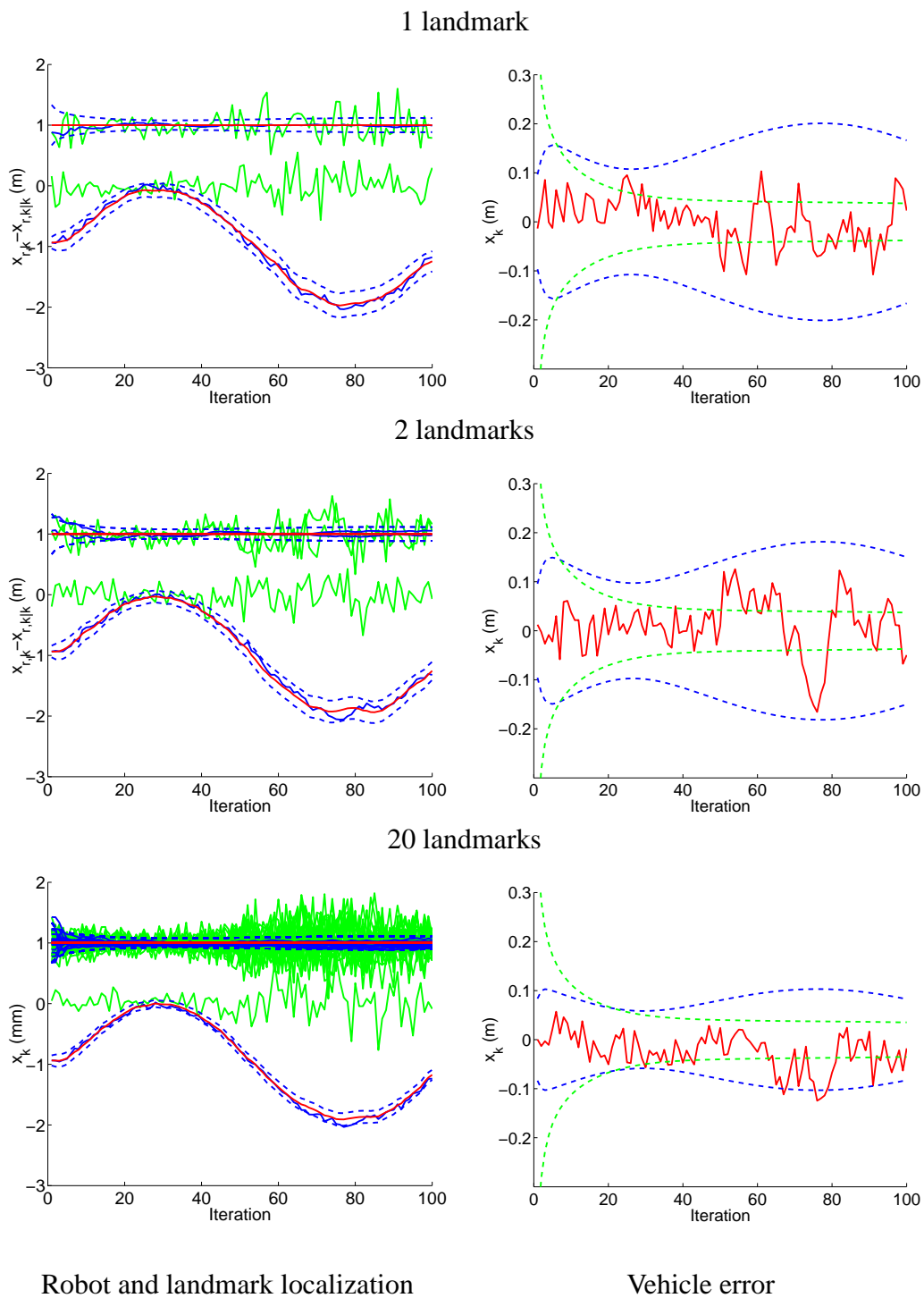


Figure 6.19: Optimal state estimation and regulation for a monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ .

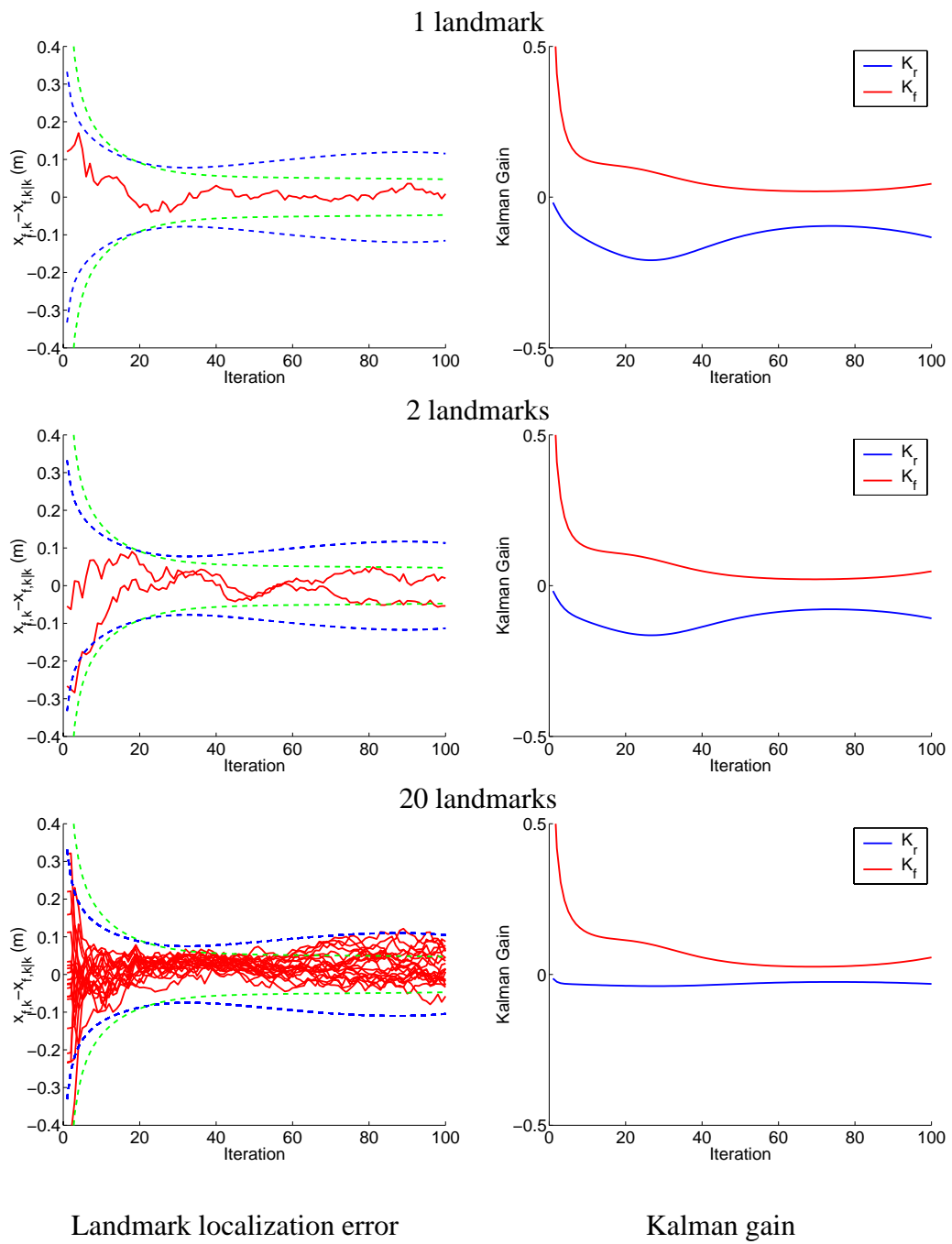


Figure 6.20: Optimal state estimation and regulation for a monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ .

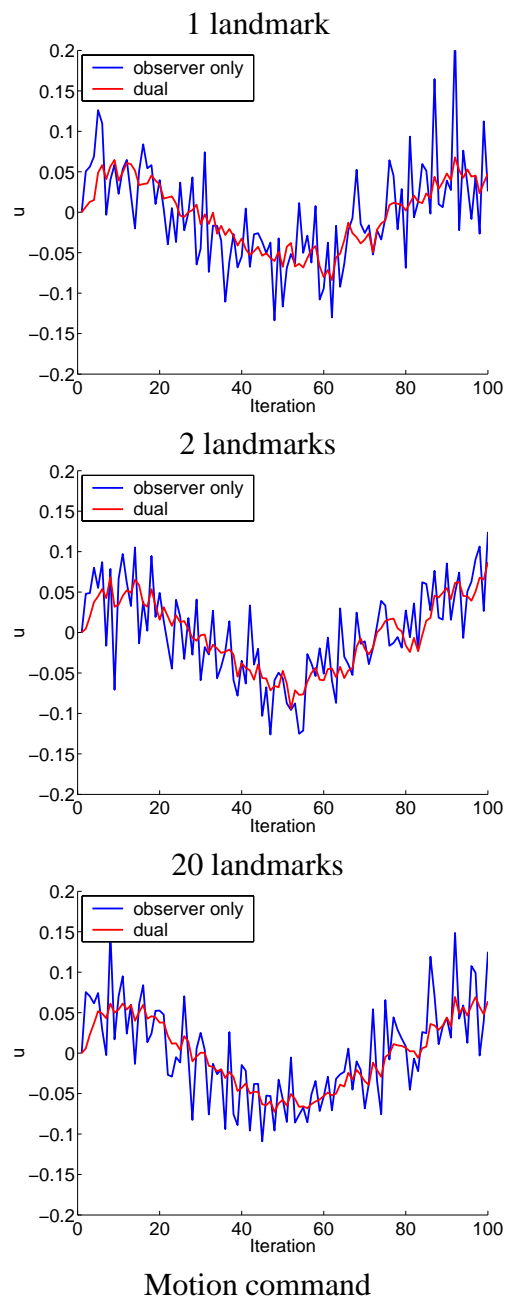


Figure 6.21: Optimal state estimation and regulation for a monobot in a sinusoidal path from  $\mathbf{x}_{r,0|0} = -1m$  to  $\mathbf{x}_{100} = -1m$ , with 100 iterations and  $\alpha_d = 0.1$ , and  $\beta_d = 0.1$ .

## 6.9 Divergence

As mentioned in the previous chapter, Julier and Uhlmann showed concern for the convergence properties of the KF approach to CML. They presented a simulation experiment where CML fails even for a simple configuration [108]. They argue that the divergence of the algorithm is due to the nonlinearities of the model used. Their contribution suggests that robust statistics and nonlinear approaches to KF, such as the Unscented KF [105], should be explored as means to more robust solutions to CML.

Here, we revise the Gaussian assumptions being violated in the linearization process, and propose a couple of measures to deal with them; namely, the over estimation of model errors, and by having more accurate measurement models.

The effects of linearization in the EKF are: a) neglecting the higher order terms in the Taylor series Approximations 5.6-5.7, and b) the evaluation of Jacobians at estimated values of the state rather than the exact values, which are not available.

The over estimation of model errors can be achieved simply by augmenting the process noise covariance matrix  $\mathbf{V}$  with a *psd* matrix  $\mathbf{V}'$

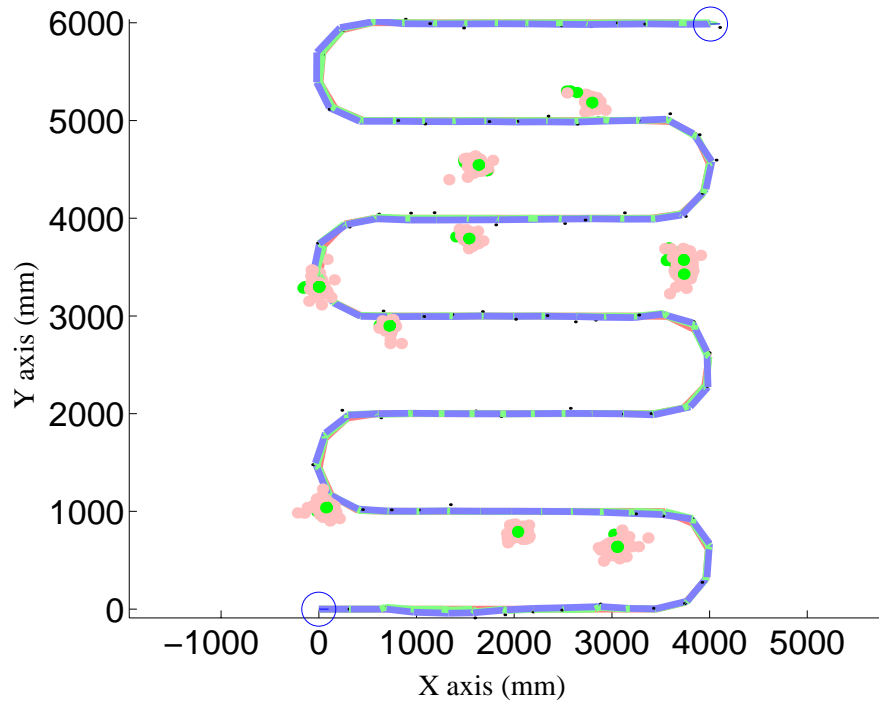
$$\mathbf{V}_D = \mathbf{V} + \mathbf{V}' \quad (6.73)$$

As in Section 6.7, the new process covariance  $\mathbf{V}_D$  is to be used in the evaluation of the a priori error covariance  $\mathbf{P}_{k+1|k}$ .

Having an accurate measurement model would also help in reducing the divergence effects for nonlinear models. The online computation of  $\mathbf{W}_k$  from a series of observations is applicable in this case.

Figure 6.22 shows the results of having full observability by means of an external sensor, and the online computation of the measurement covariance  $\mathbf{W}$ , for a test run of CML with 10 landmarks in 100 iterations using the planar vehicle. The test run parameters are the same as those used in Section 5.5,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ , and  $\beta_\psi = \beta_d = 0.01$ . The small black dots along the trajectory indicate the noise corrupted observations of the robot produced by the external sensor.

Compare the error levels to those of Figure 5.16. The divergence effects are minimized once full observability is guaranteed. Nevertheless, the obtained estimates are still unreliable, unless we can guarantee the filter does not shut off because of full correlation. That is, we need artificial process noise to keep the covariance estimates from converging to zero. We overestimate  $\mathbf{V}$  simply by adding pseudo noise covariance to the landmark estimates, thus making  $\mathbf{V}_D$  positive definite. Figure 6.23 shows the results of having a fully observable CML model with artificial process noise associated to the



a) Robot localization

landmark states for the nonlinear vehicle depicted throughout this thesis.

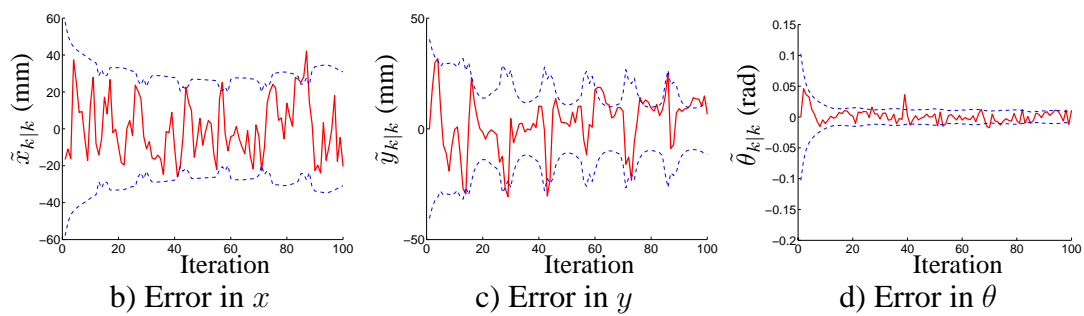
b) Error in  $x$ c) Error in  $y$ d) Error in  $\theta$ 

Figure 6.22: Full-covariance EKF CML for a path with 100 iterations and 10 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ , and full observability. An external sensor measures the robot position at each iteration.



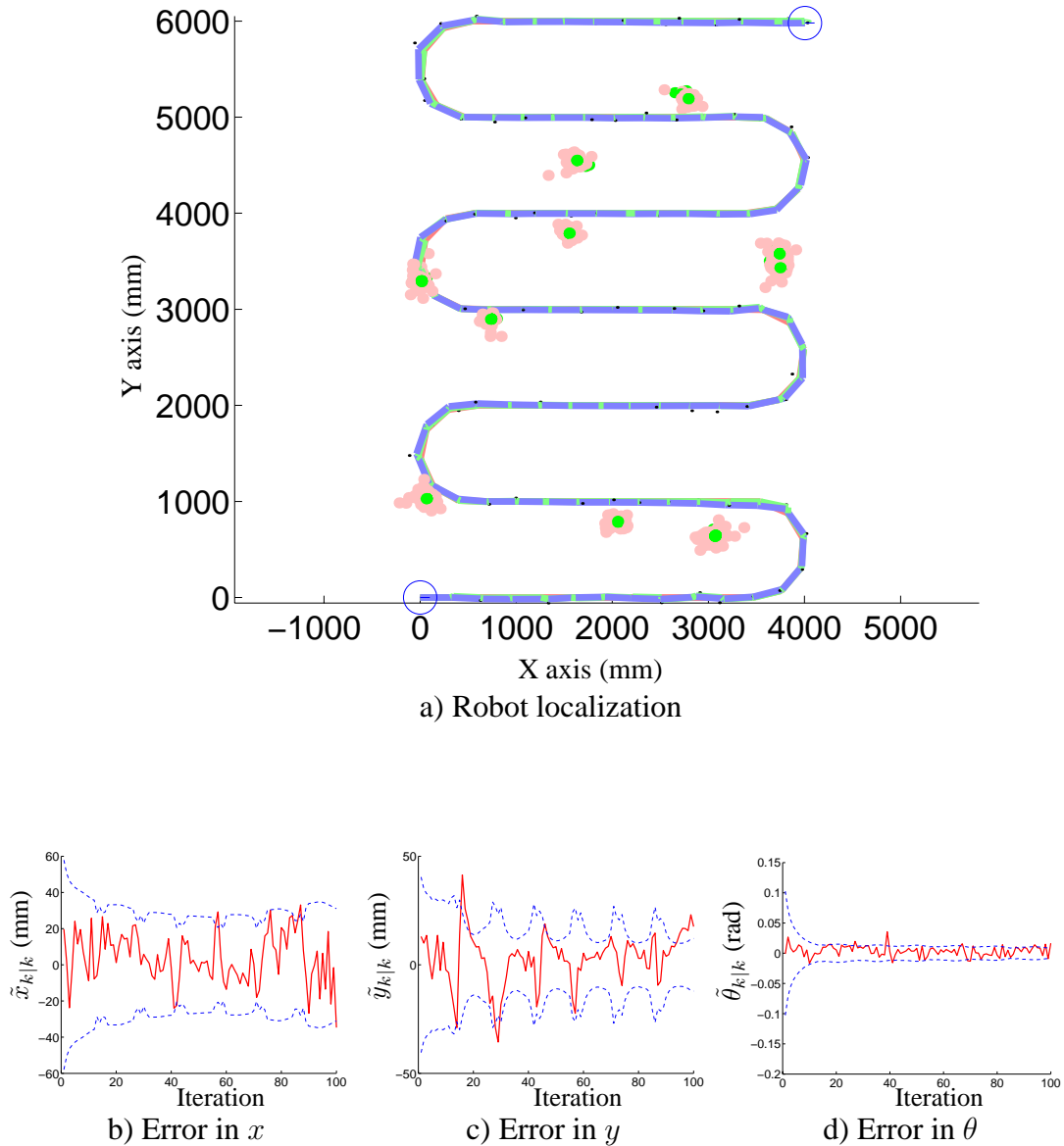


Figure 6.23: Full-covariance EKF CML for a path with 100 iterations and 10 landmarks,  $\alpha_\psi = \alpha_d = \alpha_\theta = 0.1$ ,  $\beta_\psi = \beta_d = 0.01$ , full observability with artificial process noise. An external sensor measures the robot position at each iteration, and pseudo noise covariance is added for the landmark location estimates.

## 6.10 Bibliographical notes

Gibbens *et al.* [83] give a solution to the 1-d Brownian motion CML case. In their solution, the state error covariance is linked to the total number of landmarks in the form of the total Fisher information  $I_T$ .

$$I_T = \sum_1^n \frac{1}{\sigma_w^2} \quad (6.74)$$

The expression indicates the “*informational equivalence of the measurements and the innovations*” [26], and was derived from a simpler likelihood function than the one in Equation 6.2; one that does not contain the fully correlated characteristics of the measurement model.

The issues of controllability and observability in Kalman filter design are covered in the book by Bar-Shalom *et al.* [26], and on the lecture notes by Polavarapu [158]. Southall *et al.* address the issue in the context of tracking for an autonomous agricultural application [183].

General texts on estimation theory include the aforementioned book by Bar-Shalom, and the notes by Todling [194]. The Cramer Rao lower bound and the information matrix are also discussed in an article by Dowski [66].

The duality between the stochastic optimal observer (the Kalman filter) and the stochastic optimal regulator for LQG problems is a standard topic of optimal control theory. Some texts and lecture notes covering the matter include those of Andersen [4], Furuta *et al.* [79], and Wen [214].

## 6.11 Conclusions

The fundamental contributions of this chapter is that we show how the full-correlation of the map model in EKF-CML hinders full observability of the state estimate, and how having a partially controllable state model makes the filter to stop working after a few iterations. Partial observability makes the final map dependant on the initial observations, and does not guarantee convergence to a positive definite covariance matrix. Partial controllability on the other hand, makes the filter believe it has accurate estimates of the landmark states, with their corresponding Kalman gains converging rapidly to zero. That is, after a few steps, innovations are useless for the refinement of landmark locations, and contribute only to the revision of the vehicle pose.

---

Partial observability can easily be solved by anchoring the map to the first landmark observed, or by having an external sensor that sees the vehicle at all times. Partial controllability on the other hand, can be only palliated to some extent by adding artificial process noise to the landmark estimates.

Furthermore, from the formulation of CML as an LQG problem, one can exploit the duality between the Kalman filter and the optimal regulator, with the advantage of obtaining nicer, smoother, motion commands.

Finally, the divergence caused by nonlinearities in CML can be reduced either by assuming larger uncertainty of the vehicle model (augmenting the process noise covariance), and with the online computation of measurement noise covariance matrices.



# Chapter 7

## Map Building

In Chapters 2 and 4 we studied a mirage of techniques for landmark extraction and landmark tracking in mobile robotics. In Chapter 5 we presented a methodology for the estimation and update of the entries in a fully correlated map of the environment with the aid of a Kalman Filter. We studied also the effects dead reckoning robot drift and measurement noise have on such filter. Furthermore, in Chapter 6 we concentrated in the fundamental convergence properties of the algorithm, with respect to the number of landmarks in the map. We now turn our attention to the algorithmic issues for keeping such a fully correlated map of the environment. We will see in detail the requirements of the data structure that must hold the map representation, including methods for the efficient insertion, deletion, and update of landmarks in the map. The link to the techniques used to constrain the search for landmark matches within the estimated field of view, presented in Section 5.4 will become evident. The discussion completes the broad view of the proposed architecture shown in Figure 1.1 in the introductory chapter.

Unlike grid-based techniques, our map must be scale independent, and be able to handle various landmark representations coming from different sensors. Furthermore, the map is designed so that updating can occur in real time in moderately changing environments.

### 7.1 AVL trees

One of the most fundamental data structures that computer science makes available to us for representing the relations between geometric entities in an ordered way are binary trees. The nodes in the tree could be used for example to indicate attribute ranges of learned landmarks; and the arrangement of these nodes within the tree could be such that

searching for a particular landmark or group of landmarks would be done expeditiously. For example, by arranging landmarks in the tree consistent with their global position in the environment.

A binary tree provides the same benefits of a linked list, easy insertion and deletion of nodes with the speed of a binary search. But for a tree to be efficient, it must be balanced, this is, it should have roughly the same number of nodes on every subtree. There exist several types of balanced trees, differentiating on the conditions required to keep it balanced. From AVL and red-black trees that maintain balance by performing rotations on the nodes, to 2-3 trees that maintain balance by allowing more than two branches from each leaf, to the generalization of 2-3 trees to B-trees. For a thorough description of the various types of binary trees and their properties, see for example the book by Cormen *et al.* [50].

We have chosen the Adelson-Velskiĭ-Landis (AVL) definition of balance as the underlying mechanism to maintain a data structure for the construction of our environment maps, because of the low cost of the operations needed to keep the map tree balanced, plus the low cost in the execution of range queries [1, 18].

In an AVL tree, the height of its two subtrees at every node is at most one. This balance condition guarantees worst case times of  $O(\log n)$  for the search, insertion, and deletion of entries. Keeping an AVL tree balanced requires rotations, an operation that switches the parent node on sibling subtrees so as to keep their height equal. Extra work is needed for doing rotations, in comparison to a random binary tree for example. However, because the tree is kept balanced, the average path length for an insertion or a deletion is shorter on average. The overall effect of having the tree balanced outweighs the extra work on keeping it balanced.

## 7.2 Multidimensional search trees

In order to maintain an ordered map of the environment with one dimensional landmarks, such as the ones found by a *monobot* (see Section 5.2.1), an AVL tree would suffice. However, when landmark attributes are of more than one dimension, we must resort to higher dimensional structures.

In an AVL binary tree, entries are ordered in terms of one feature only, say for example, the  $x$  coordinate of a landmark position with respect to a global reference frame. If we want to store elements in a higher dimensional order we have two choices. One is to use a k-d tree, a generalization of binary trees in which each node splits into a pair of subtrees by cycling through the  $k$  keys of the k-dimensional search space; examples of

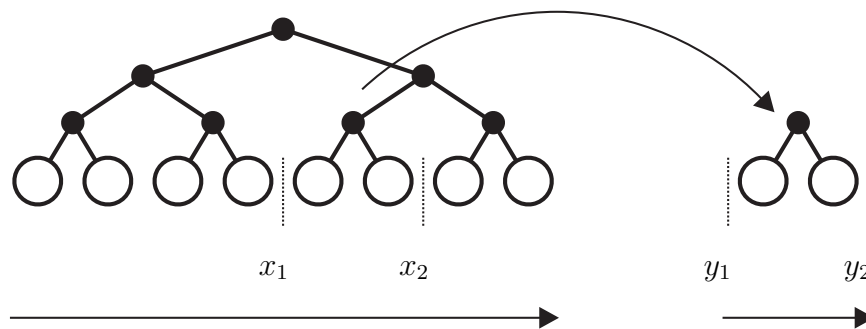


Figure 7.1: Map representation using AVL trees in higher dimensions.

k-d trees are the quadtree for two dimensions, or the octree for three dimensions. However, such a choice is not compatible with the rotation operations needed to maintain the trees balanced.

The second choice is to have auxiliary trees associated to each node of the original AVL tree, one per extra feature. So for example, if our data is 2-dimensional, in the original tree the elements are still stored with respect to one dimension, say the  $x$  coordinate; we call this tree the  $x$ -tree. At every node in the  $x$ -tree an auxiliary tree sorted by  $y$  is appended. See Figure 7.1.

For each landmark in the  $x$ -tree we have  $\log n$   $y$ -trees on the way up the tree to the root. The total space needed for the storage of such a data structure is  $O(n \log^k n)$ , with  $k$  the dimensionality of the data. Insertion and deletion of landmarks in the map take  $O(\log^k n)$ , since for each node in the  $x$ -tree, an instruction takes  $\log n$ , but for each of these we have  $(k - 1) \log n$  auxiliary trees.

Such a data structure is most suitable for the computation of range queries. For a range query we need only to traverse the  $x$ -tree down to the node in which the query is most constrained. Then, we can jump to the  $y$ -tree and continue down to narrow the set of matches. Those entries present in the leaves of the resulting  $x$  and  $y$ -subtrees constitute the query match.

In order to see this, observe the following example. Say we have observed a set of 9 landmarks with coordinates  $(x, y)$  in the following order

$$\mathbf{x}_f = [(1, 0), (2, 1), (2, 0), (1, 1), (0, 0), (1, 2), (0, 1), (0, 2), (2, 2)]^\top$$

The label assignment for these points is indicated in Figure 7.2 (a). Figure 7.3 shows

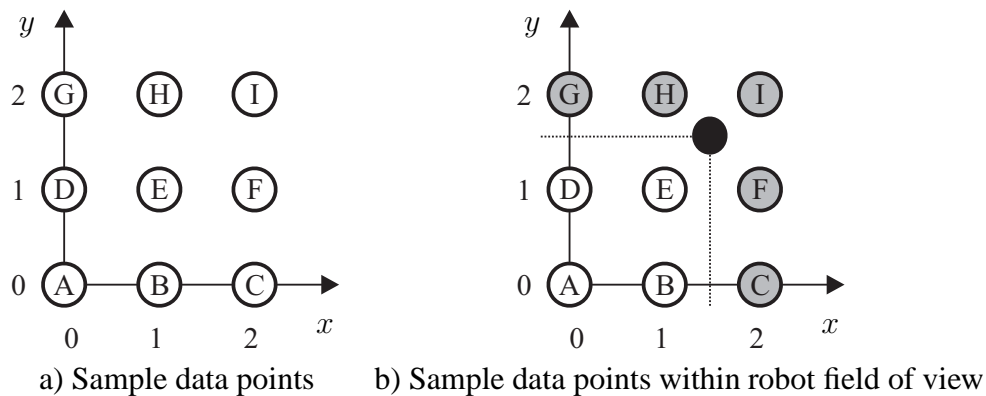


Figure 7.2: A set of nine landmarks.

their insertion into the multidimensional AVL trees.

Now, say for example, that our robot is located at position  $(1.5, 1.5)$  and is heading left-down. From such a vantage point we would expect, for a sensor with a limited field of view, the observation of landmarks with entries  $x < 1.5$  and  $y < 1.5$  only. Figure 7.2 (b) shows the set of landmarks that are within the field of view in white. The rest of the entries in the map are out of the robot field of view, and are shown in grey. Figure 7.4 shows how the query proceeds within the recently built data structure. Once we have located the most constrained node in the  $x$ -tree, we continue narrowing the search space in its corresponding  $y$ -tree. The resulting set of matches is  $\{A, B, D, E\}$ . The search time is  $O(\log^2 n)$ .

### 7.3 Data association

The building of a map of the environment is a continuous process that takes place during robot navigation. The most critical aspect of this process is the association of data from different viewpoints. This is, the matching of a set of observations with the entries already learned and stored in a data structure.

In chapter 2 we commented on the appearance invariant properties visual features must have in order to achieve robust data association, and on the techniques used to match such features in between frames. Furthermore, such association is greatly simplified with the aid of the tracking techniques described in Chapter 4. However visual invariance is not always available. In this section we see the problem of data association



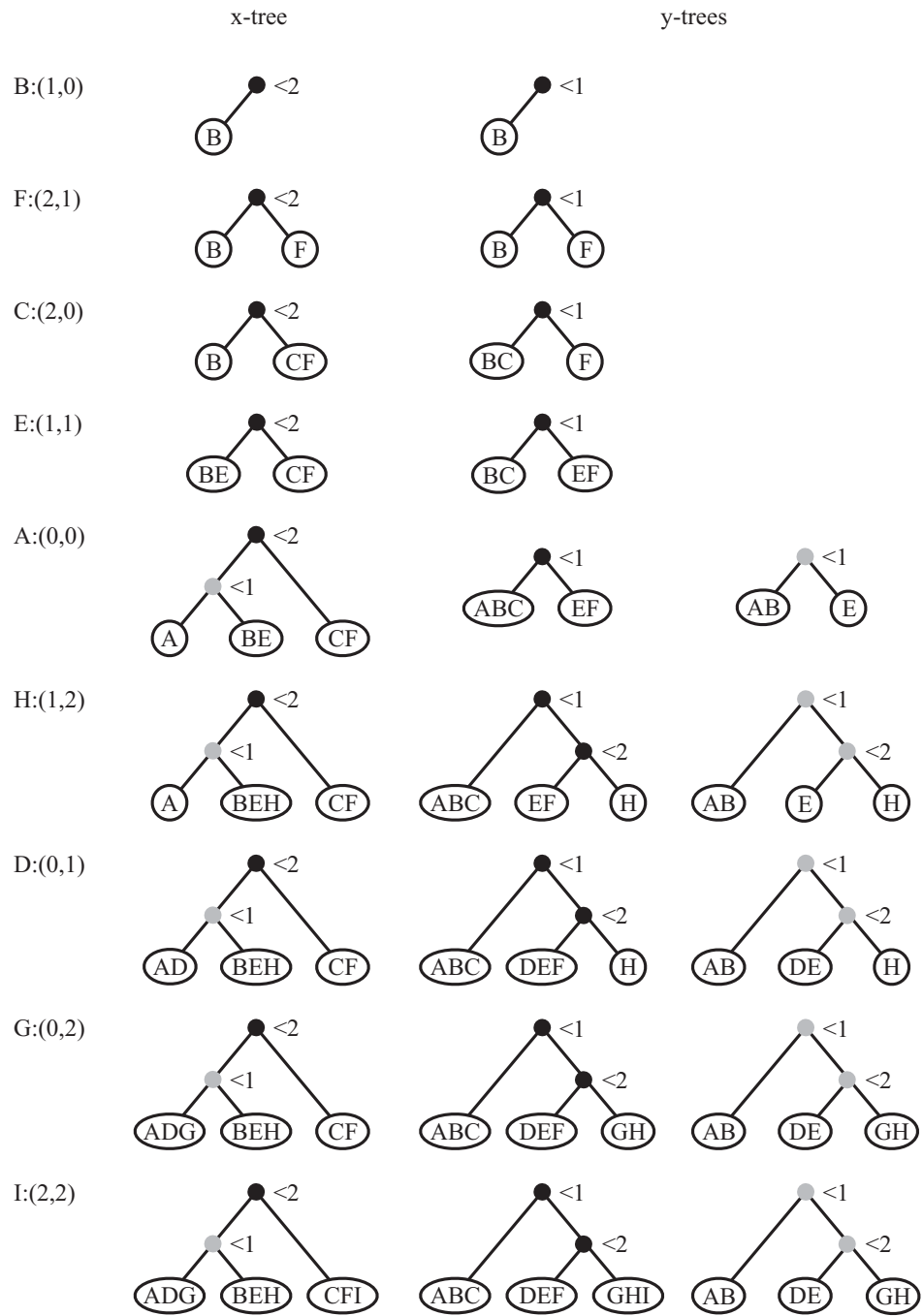


Figure 7.3: Insertion of nine data points in a bidimensional AVL tree.

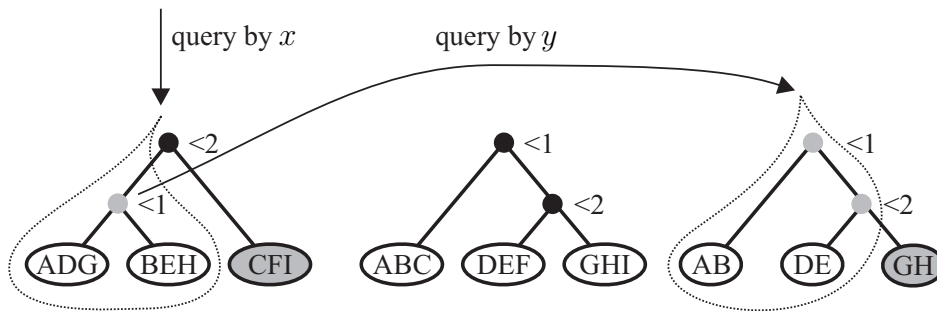


Figure 7.4: Map representation using AVL trees in higher dimensions.

from a different point of view. One that is more model-based than appearance-based. That is, one that relies more on the geometrical properties of the landmarks extracted, rather than on their appearance. Such approach is possible for example with sensors that produce geometrical measurements instead of appearance measurements. That is the case of our laser range scanner. The discussion helps in completing our broad view of data association.

Without loss of generality, we will describe next the various heuristics developed within the scope of this thesis for the efficient matching of straight lines into walls for indoor environments, as scanned with a laser range scanner. We divide the process of data association for laser scanning in four subsequent tasks; from the input of raw data coming from the laser scanner to the association of landmarks. These are model compatibility tests, search space reduction, landmark hypothesis formation, and landmark association. Similar tasks can be devised for other types of landmarks and other types of sensors.

### 7.3.1 Model compatibility

The model compatibility tests are used to validate that the landmark extracted is compliant with a set of properties usually indicated by an expert. In the case of laser range scans of walls as computed with the line simplification algorithm from Section A.2, we have observed that the scan points must not be further apart than  $35cm$ , that there must be at least 5 range points in a straight line, that a wall is usually larger than  $40cm$ , and that observations farther than  $6m$  are unreliable. These tests are explained graphically in Figure 7.5. Once an observation passes these tests, we can proceed to query for compatible matches in the map.

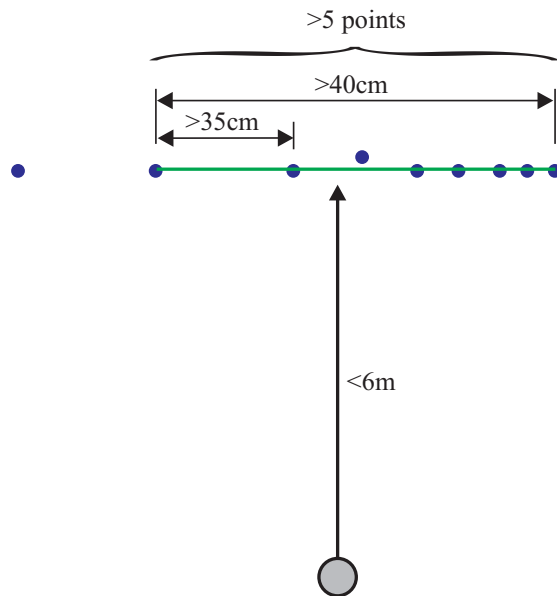


Figure 7.5: Wall compatibility tests of a laser range scan.

### 7.3.2 Search space reduction

When building maps of large areas, the search for landmark match hypotheses might become intractable unless the search space is constrained. A simple and efficient way to reduce the search space is to limit hypothesis tests to be produced locally; that is, within a bounding box around the landmark to be tested.

Thus, the range query into our bidimensional AVL must return all the landmarks in the map with at least one of their endpoints inside a search area. We have observed that for the matching of walls and laser scans, it suffices to bound the search area as much as  $5m$  to each side of the extremes of the observed landmark to be matched. This is exemplified by the large dotted box shown in Figure 7.6.

### 7.3.3 Hypotheses formation

Furthermore, for a wall in the map to be compatible with a landmark observation, it should intersect with at least one of the line segments that form the shaded bounding box shown in Figure 7.6. A typical value for the extension of such bounding box is  $20cm$ . A fast algorithm for the computation of line intersections due to Bentley and

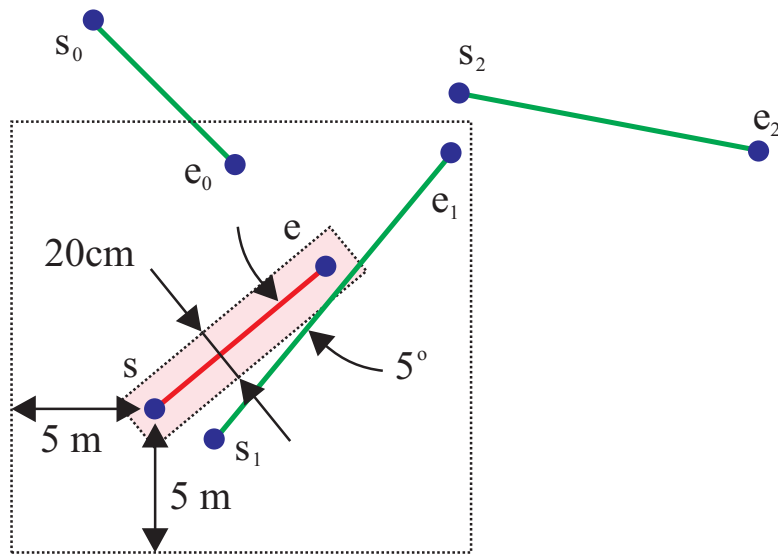


Figure 7.6: Hypothesis search range for walls extracted from a laser range scan.

Ottmann is the *line sweeping* algorithm [29]. It is a standard algorithm in computational geometry, and is typically used in applications such as computer graphics. Suffice to say that the algorithm has time complexity  $O(n \log n + k)$ , where  $n$  is the number of line segments to be tested and  $k$  is the number of intersections reported. The reader is referred to text books such as the one by Preparata and Shamos [162] for a thorough discussion of the technique.

Finally, one more reduction of the search space is needed. In the case of walls as extracted from laser scans, we have observed that the orientation of the matching walls must not diverge by more than 5 degrees. Thus only those landmarks from the reduced search space passing this test are considered for association.

### 7.3.4 Association

Landmark hypothesis verification is performed with the data association tests from Sections 5.1.9 and 5.4.4. Moreover, when contention for more than one landmark match needs to be solved, the more expensive joint compatibility test from Section 5.1.10 proved useful.

## 7.4 Map update

We now generalize the discussion for any type of sensor or landmark models. The result from the landmark extraction module at the  $k$ -th iteration is a set  $D(k)$  of detected landmarks, and its corresponding appearance property vector set  $A(k)$ .

$$D(k) = \{\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(m)}\} \quad (7.1)$$

$$A(k) = \{\zeta_1, \zeta_2, \dots, \zeta_m\} \quad (7.2)$$

To update the map, all landmarks in  $D(k)$  are searched for match hypotheses in the previous map state  $T(k-1)$ . This search is limited to a reduced number of landmarks as indicated in the previous section. Only those map landmarks that fall within the field of view at the robot's current position are considered during the map update process, with  $T'(k)$  the set of map landmarks that fall within the field of view,  $T'(k) \subseteq T(k-1)$ .

If the match hypothesis passes the association tests, and their appearance properties vectors are highly correlated, then we have a scene-to-map landmark match. Once a match is obtained, the distribution parameters for the uncertainty of that landmark position in the map are updated, along with its vector of appearance properties, as well as its landmark quality.

Similarly, we must update the landmark quality of those entries in the search space that did not have a match in  $D$ . We end up removing from the map those entries whose landmark quality falls below a certain threshold. A typical value for the landmark quality threshold is

$$x_q^{(i)} \leq 0.1 \quad (7.3)$$

Table 7.1 shows the steps performed for each iteration of the map building process, and Figure 7.7 illustrates this process. The scene to map landmark matches are indicated by the set  $M(k)$ , and the set of new landmarks that must be added to the map is indicated by the set  $N(k)$ .

$$M(k) = D(k) \cap T'(k) \quad (7.4)$$

$$N(k) = D(k) - M(k) \quad (7.5)$$

The proposed map update scheme has the following advantages over other map learning algorithms

1. The map preserves its topological structure. The prevailing relationships among existing features are their own Euclidean metrics.

---



---

```

MapUpdate( $T, D$ )

  for each  $\mathbf{z}^{(i)}$  in  $D$ 
    if TestModelCompatibility( $\mathbf{z}^{(i)}$ )
       $T' = \text{ReduceSearchSpace}(\mathbf{z}^{(i)}, T)$ 
      for each  $\mathbf{x}_f^{(j)}$  in  $T'$ 
         $H = \text{CreateMatchHypothesis}(\mathbf{z}^{(i)}, \mathbf{x}_f^{(j)})$ 
         $M = \text{max DataAssociationTest}(H)$ 
        if  $M$ 
          UpdateMapMatch( $M$ )
        else
           $N = \text{AddtoMap}(\mathbf{z}^{(i)})$ 

  for each  $\mathbf{x}_f^{(j)}$  in  $T'$ 
    UpdateLandmarkQuality
    if  $x_q^{(j)} < x_{q,THLD}$ 
       $F = \text{RemoveLandmark}(\mathbf{x}_f^{(j)})$ 

```

---



---

Table 7.1: Algorithm for map update.

2. The map is not limited in resolution, as opposed to grid-based maps. This allows the modeling of different size environments without the need to modify its general structure.
3. The dynamic property of the map allows for the robust modeling of changing or noisy environments. It also refrains the map from growing indefinitely, a situation that could affray with system resources (search speed and memory).

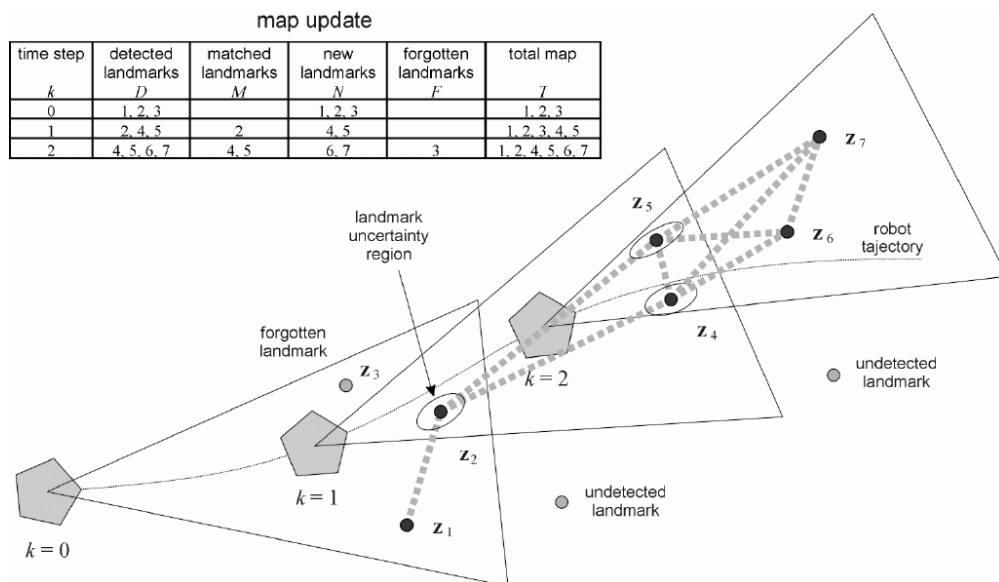


Figure 7.7: Only the set of landmarks from the map that fall within the robot field of view are compared to the set  $D$  of landmarks coming from the landmark extraction module.  $M$  is the set of matched landmarks, and during the map update process their uncertainty parameters and existence state are updated. New entries  $N$  are added to the map for the elements in  $D$  that did not contain a corresponding match in the map; and possibly, a set  $F$  of landmarks is removed from the map if their existence state is below a desired threshold. The figure shows the position of the robot at various time instances. The triangles represent the field of view of the robot, and the landmarks are represented by dots.

## 7.5 Bibliographical notes

### Binary trees

The most fundamental data structure that computer science makes available for the representation of relational information is a binary search tree. Search trees are data structures that support many dynamic-set operations, including search, insertion, deletion, and first order statistics such as maximum and minimum. Basic operations on binary search trees take time proportional to the height of the tree. Thus, if the tree is balanced such operations run in  $O(\log n)$ . For a thorough description of the various types of binary trees and their properties, see the book by Cormen *et al.* [50].

Little work has been done in extending map representations from trees to graphs; the main reason being the inherent intractability of the graph matching problem. To exemplify the use of graphs in mobile robotics, we cite one recent contribution on the representation of the environment with the aid of graphs [69]; and a system that uses fuzzy logic principles for building a grid-based map and graph-searching techniques for trajectory planning [154].

### Map representation techniques

A variety of map representation models have been used in the robotics literature, depending on the task at hand, and on the characteristics of the robot and of the environment. Recent surveys on map generation techniques for mobile robotics include those of Halperin *et al.* [89], Saffiotti [168], and Tsubouchi [201].

The most common type of representation of the environment in indoor mobile robotics is the occupancy-based grid map, where objects extracted from sensor data are mapped onto small cells of a tessellated space fixed to a world reference coordinate system. Recent contributions on grid-based map building include the ones by Anousaki and Kyriakopoulos [14], Lee and Recce [125], or Oriolo *et al.* [154]. Extensions to this method include the storing of sensor signatures on each grid cell to save landmark invariants for later attempts at robot localization [70], or to associate with each cell an occupancy probability distribution [143, 182]. Other recent contributions on grid-based map building include those by Jennings *et al.* [104], and Meikie *et al.* [139]. Moreover, probabilistic approaches that combine map learning and localization over grid maps include those by Thrun *et al.* [77, 191, 193].

Another widely used map representation is based on isolated features, rather than



grid maps. These methods exploit topological properties about the environment, including spatial relations. Probabilistic methods for obstacle parameterization have also been suggested for this type of representation [122], but we consider that the assumptions made on the type of probability density functions proposed for obstacle parameterization may be unrealistic. Two additional stochastic approaches that combine map learning and localization worthy of note are the *SPMap* model [43], based on a general probabilistic representation of uncertain geometric information; and the *BaLL* learning algorithm [191], which uses Bayesian Belief Networks to learn relevant landmarks. A comparative study on the performance of exploration strategies over grid-based maps versus feature-based maps is presented in Lee and Recce's article [125]. More recently, an increased interest in the use of estimation theoretic techniques for the characterization of landmarks has emerged. Exemplars of this are the works by Ayache and Faugeras [20], Dissanayake *et al.* [64, 126], or Smith and Cheeseman [182]. Our thesis, and the related work presented in [11] belong to this group of map representations. However, a differentiating characteristic that separates them from other methods is that the proposed map model takes also into account the temporal landmark uncertainty.

Fuzzy set-based representations have also been proposed to differentiate between the imprecision in the localization of a landmark, and the uncertainty in its very existence [16, 126, 154, 161, 168]. An interesting application is by López-Sánchez [132, 134], in which a troupe of robots merge sensor data for grid-based map building using fuzzy set theory. Some authors have recently proposed the use of goal oriented *cognitive maps* to learn the relationship between successively explored places [17, 82]. However, these methods are usually limited in that changing environments can only be dealt with reactive behaviors.

## Behaviors for scene exploration

A current trend in robotics suggests that high-level tasks such as scene exploration be controlled with the aid of behavior-based systems; that is, to design systems that can switch between different control modes thanks to an arbitrator. Such an arbitrator could be implemented with the use of fuzzy logic [118] or via supervisory control [121]. An arbitrator must be able to decide whether to assign priority to a low-level behavior for obstacle avoidance, or to select a control strategy to carry out its assigned task expeditiously.

A method that has been proposed for the autonomous learning of behaviors by a mobile robot to safely interact with its environment is that of reinforcement learning; see for example the work by Torras [140, 197, 198]. The technique consists on teaching goal-oriented reactive behaviors to a robot by training a neural network that maps the currently perceived situation into the next action. The system rewards actions that

approach the goal while penalizing actions that approach obstacles or deviate from the goal. The system learns at the end, the trajectory that optimizes the total amount of reinforcement obtained along the path to the goal. Other recent goal oriented systems include the one by Cicirelli *et al.* [49], that learns behaviors using reinforcement learning over grid-based maps by fusing vision and sonar data; and the one by Gaussier *et al.* [81] where rather than densely mapped descriptions of the environment, the system learns visual landmarks and performs goal-oriented visual homing. We foresee two major drawbacks on the use of these techniques for environment learning. First, the systems would require extensive training to learn the specifics of one particular setting, and their performance would deteriorate considerably in dynamic environments. Secondly, the very nature of the approach is reactive and does not provide a robust generalization of the environment for more intelligent navigation.

Another tool that might be useful for environment learning borrowed from the artificial intelligence community is that of self-organizing or *Kohonen* maps [86]. Its ability to capture nonlinear statistical relationships in high-dimensional data has proved useful in such diverse fields as robotics, speech recognition, character recognition, and control. In the near future, we plan to study its applicability to our problem domain. Some recent contributions on the use of self-organizing maps in the field of mobile robotics include [130, 178, 185]. Another recent example on the use of an artificial neural network model for robot visual navigation in indoor environments is by Gaussier *et al.* [82].

## 7.6 Conclusions

In this Chapter we touch upon several implementation issues of our map building algorithms. We analyze the various heuristics needed for the validation of the lines extracted by a laser range finder, so they can be characterized as walls. Furthermore, we elaborate on the actual data structure chosen for the creation and update of our map model. We have designed a combined data structure that maintains a multidimensional binary tree balanced.

# Chapter 8

## Experimental Results

In this Chapter we present some experimental results that we hope will broaden the reader's understanding of the implementation details on some of the tools developed as part of our research. We start with a simple experiment that shows the influence of dead reckoning error in the computation of mean landmark estimates with the EKF. Secondly, we present some results of using the EKF-CML algorithm with landmark validation when the vehicle is acquiring visual landmarks with a stereo head. Finally, we conclude the Chapter with an exemplification of using the same algorithm but with landmark observations coming from a laser range finder.

### 8.1 Correction of a visual landmark estimate

The task at hand is to show how the Kalman filter iteratively updates its estimate of the location of a visual landmark. With that purpose in mind, a set of 60 image pairs of an easily identifiable stationary landmark (a circular blob) were taken by the stereo head in our mobile robot MARCO, at a distance of approximately 50cm. Then, the robot was commanded a motion of 10cm backwards, and a rotation of 5° was exerted counterclockwise, both at time step  $k = 61$ ; and a new set of 60 image pairs was acquired. We wanted to show two things. First, how the projection of the measured landmark position in the world reference frame can vary significantly after a motion command due to dead reckoning error. And secondly, how the estimate for the landmark position is refined after each iteration of the filter.

Figure 8.1 shows the detected landmark, as well as its uncertainty region from a

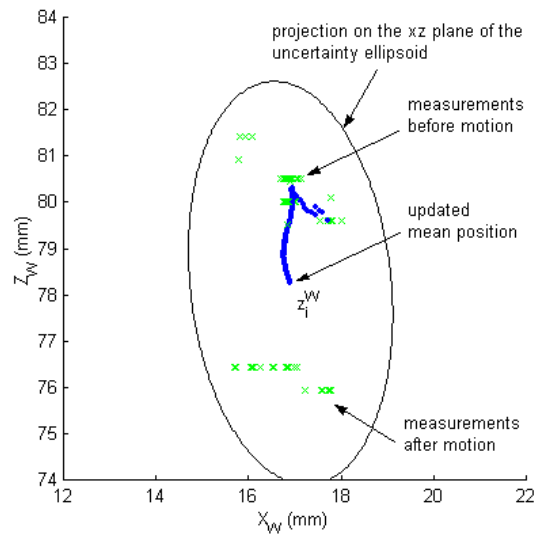


Figure 8.1: Landmark measurements and position estimate.

viewpoint similar to that of the robot stereo head. The green  $\times$  marks indicate the measured landmark positions projected into the world reference frame. The landmark measurements in between motion commands deviate by approximately  $5\text{mm}$  along the vertical direction of the image plane. Several reasons contribute to such variation. Among them, the inaccuracies due to nonlinearities in our motion model, camera and eye-to-base calibration errors, and most importantly, dead reckoning errors. Note also how the landmark position measurements are separated by an apparently constant value along the same  $z$  axis, the vertical image axis, even when the vehicle is still. This amounts to quantization error due to variations by one row in the localization of the landmarks in the images. The blue dotted trajectory indicates the adjustment of the mean landmark position as computed by the Kalman filter; and the ellipse is a  $2\sigma$  projection of its final state error covariance estimate.

A second experiment performed on a more complicated scenario shows how landmark uncertainty is also dependant on the distance at which the sensor is located. Figure 8.2 shows a group of matched salient features from one of a pair of stereo images. The hollow boxes show the features extracted from the image, whereas the filled boxes correspond to features that have been matched properly in the stereo pair. The Figure presents also a top view of the estimated position of the matched landmarks with respect to the robot as well as their position error covariance estimate. The red straight lines indicate the correspondence of scene to map landmarks. The updated robot position is also shown in the plot.

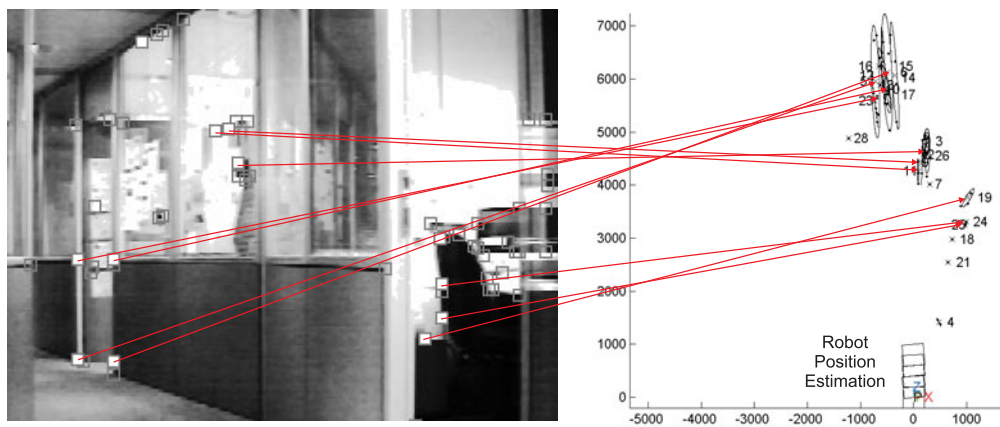


Figure 8.2: Three iterations of concurrent map building and localization over a real scenario with visual landmarks from point correspondences. The image shows how landmark covariance estimates are proportional to the size of the measurement. The further an observation is from the robot, the larger the uncertainty in its location estimate.

## 8.2 Self localization and map building from visual landmarks

Consider the situation in which we allow both the EKF-CML and EKF-CML-LV algorithms iterate over a square motion sequence. The extraction of landmarks is based solely on visual information. The salient feature detector using Beaudet's cornerness measure was used, with further refinement using the variance descent approach as shown in Section A.1.1. These salient features are then pairwise matched in the stereo set by correlation, and by the enforcement of epipolar constraints. The recovered 3D position of a feature with respect to the robot, and an associated vector of appearance properties, constitute a *landmark*, as indicated in Section A.3. The appearance properties, which are used to validate scene to map landmark matches, include the pixel gray-level mean and distribution over a small window around the salient feature, and the energy of the feature computed from the cornerness measure.

The gray level appearance correlation around the pixels of interest weighted by the energy of the Beaudet cornerness measure is used as a similarity metric, and an appearance mapping function is implemented by projecting the window of interest into world coordinates, and recomputing in this new window the appearance properties described before. Figure 8.3 shows the graphical user interface for landmark extraction.

Figure 8.4 shows a run of the full concurrent localization and map building algorithm with 50 landmarks detected. The localization of some of the landmarks might vary considerably due to the characteristics of the sensors. In this case, specular reflections might contribute to false readings. This is exemplified by the coarse localization of some landmarks in the figure with large uncertainty ellipses around them. By pruning the poor quality landmarks using the temporal landmark quality methodology described in this thesis a more accurate map can be constructed, and consequently, better robot localization is achieved. The numbered dots in the figure correspond to the identified landmarks in the scene, and the surrounding ellipses indicate a  $2\sigma$  bounded projection in the  $xy$  plane of their sample covariance  $\mathbf{P}_{ff}^{(i)}$ . The blue thin-line square indicates the desired robot trajectory, and the red thick-line indicates the actual robot trajectory as estimated from landmark correspondences. Figure 8.4 shows a three-dimensional view of the same run with the uncertainty ellipses projected in the room floor, and the landmark height indicated by thin vertical lines.

Figure 8.6 contains images from another experiment, with a sequence showing the extracted landmarks, and the corresponding map of a hallway segment. A large quantity of landmarks do not pass the spatial and temporal quality tests and end up being removed from the map.



Figure 8.3: Graphical user interface for the extraction of visual landmarks from stereoscopic images.

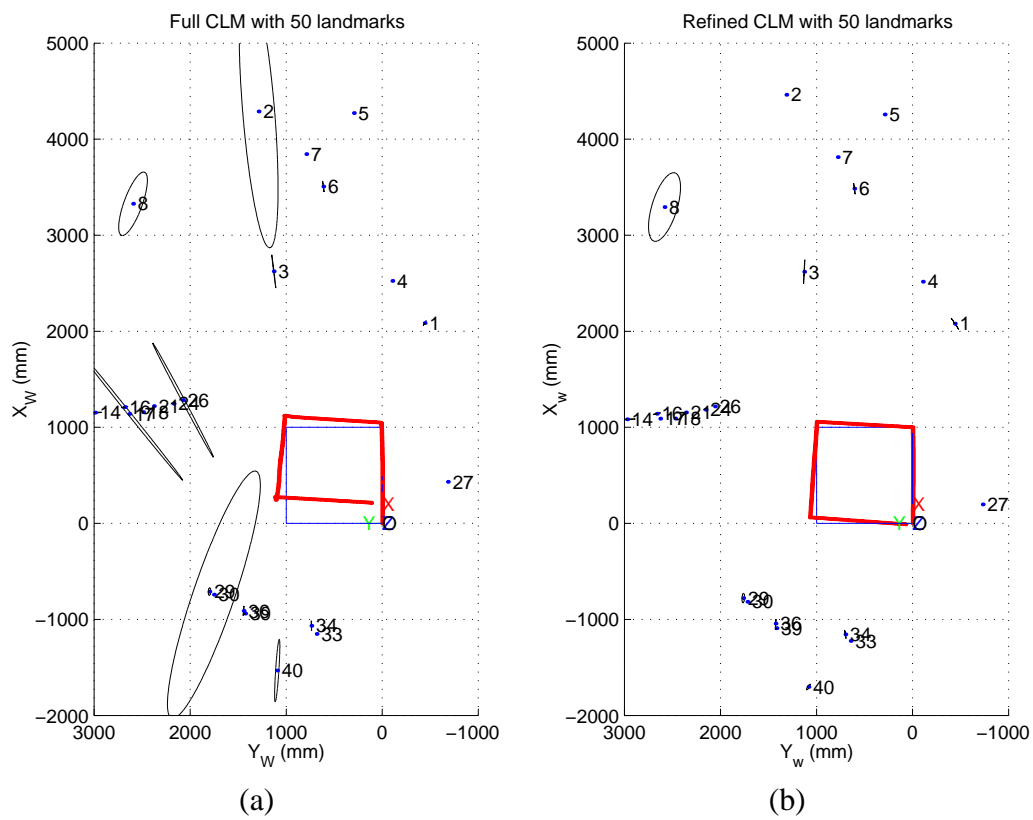


Figure 8.4: Concurrent Map Building and Localization with 50 visual landmarks.



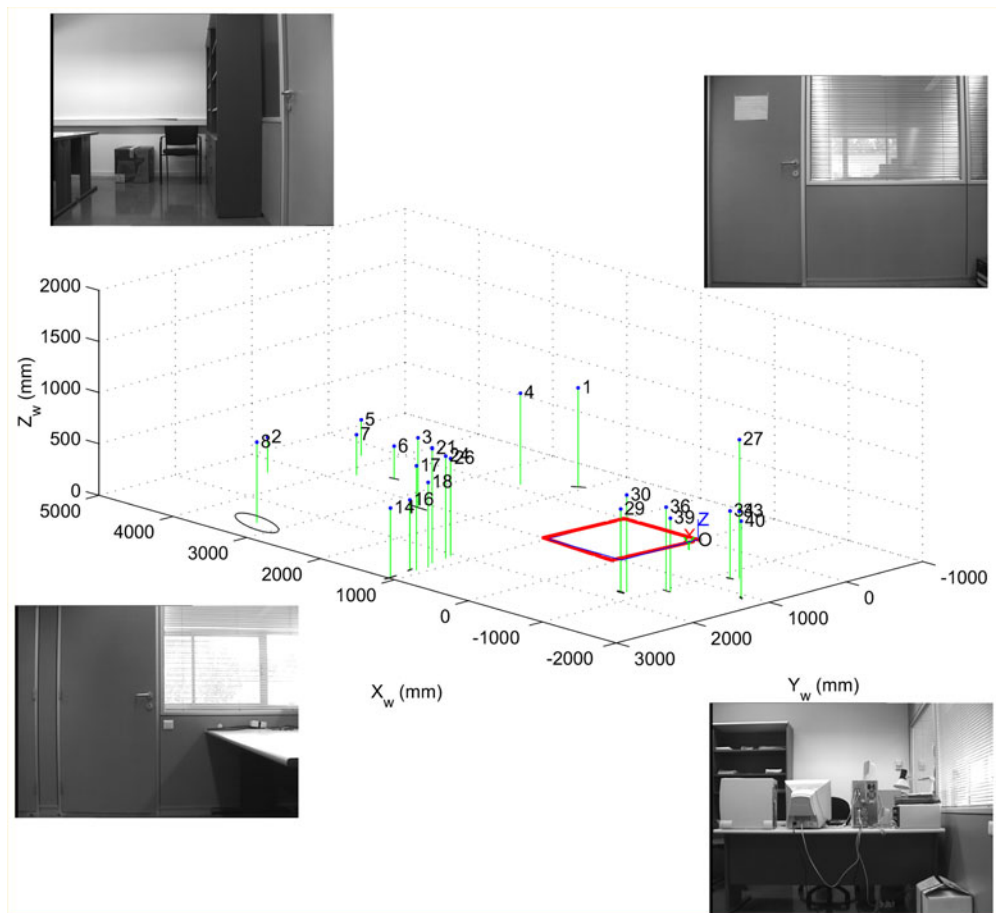
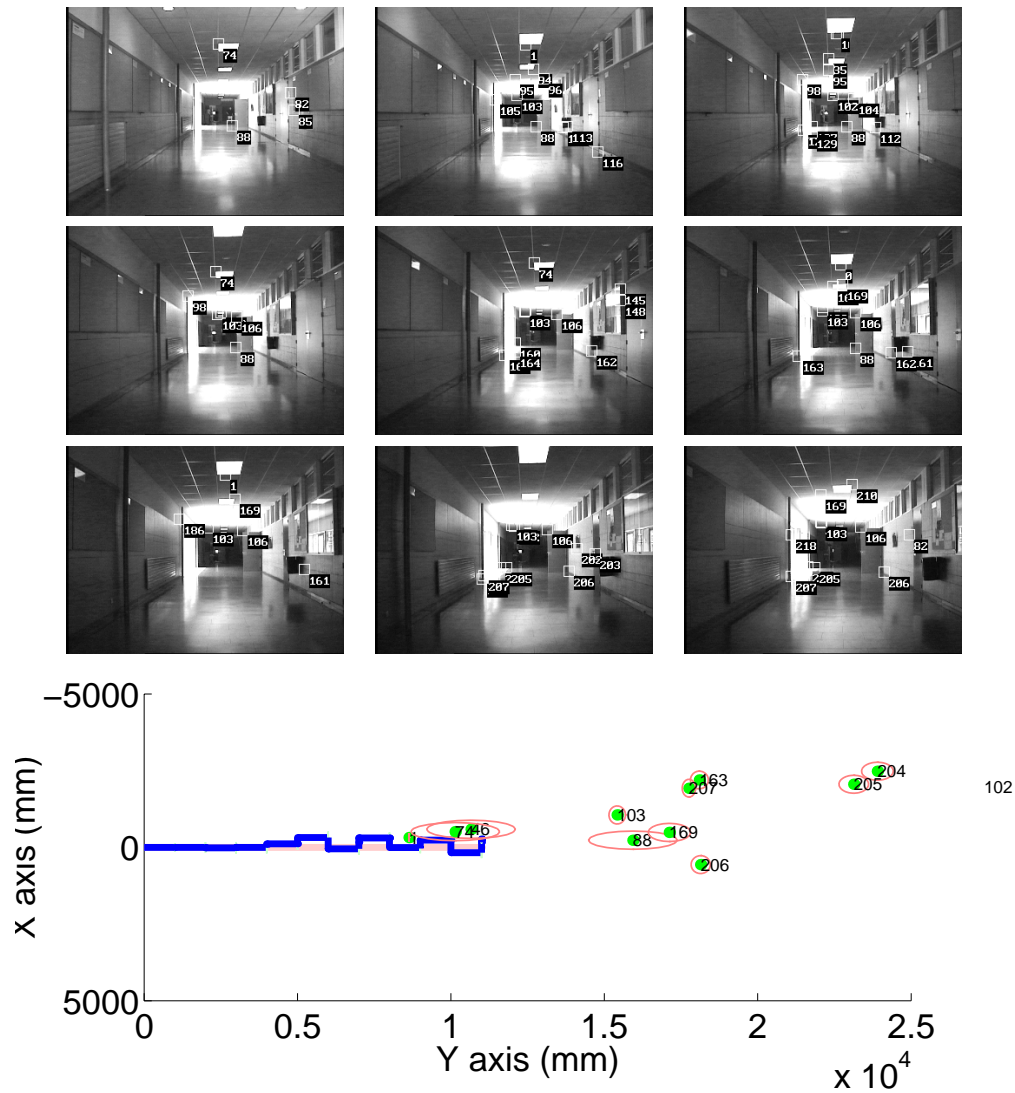


Figure 8.5: Three-dimensional map reconstruction from visual cues. The pictures indicate the four cardinal views of the scenario.



102

Figure 8.6: CML from visual point correspondences. Robust data association is difficult to achieve when using visual data only. A large quantity of image correspondences do not pass the spatial and temporal compatibility tests and end up being removed from the map.

### **8.3 Self localization and map building with a laser range scanner**

During the course of our research we have tested and implemented multiple computer vision algorithms for the extraction and robust characterization of landmarks. Our conclusion is that, changes in illumination conditions and specular reflections, among other artifacts, make solely vision based data association nearly impossible in a realistic mobile robotics scenario. For this reason, we have later invested our attention in the development of landmark extraction algorithms for a more robust type of sensor; robust in the sense that it be illumination independent. The one mostly used in the map building and localization research community is the laser range scanner.

We have developed a complete map building and localization system for our mobile robot Marco equipped with a laser range scanner. Given the interdisciplinary nature in all of our mobile robotics projects at the Institute, many people have contributed in the creation of this tool. Two people deserving special mention are Didac Busquets with whom we have developed the initial versions of our multi-robot multi-purpose graphical user interface, and Albert Checa who has implemented many of the laser raw data heuristics described in the previous Chapter. Figure 8.7 shows a snapshot of the user interface.

Figures 8.9-8.10 show a series of snapshots of the mobile robot Marco during a test run of our concurrent map building and localization algorithms.

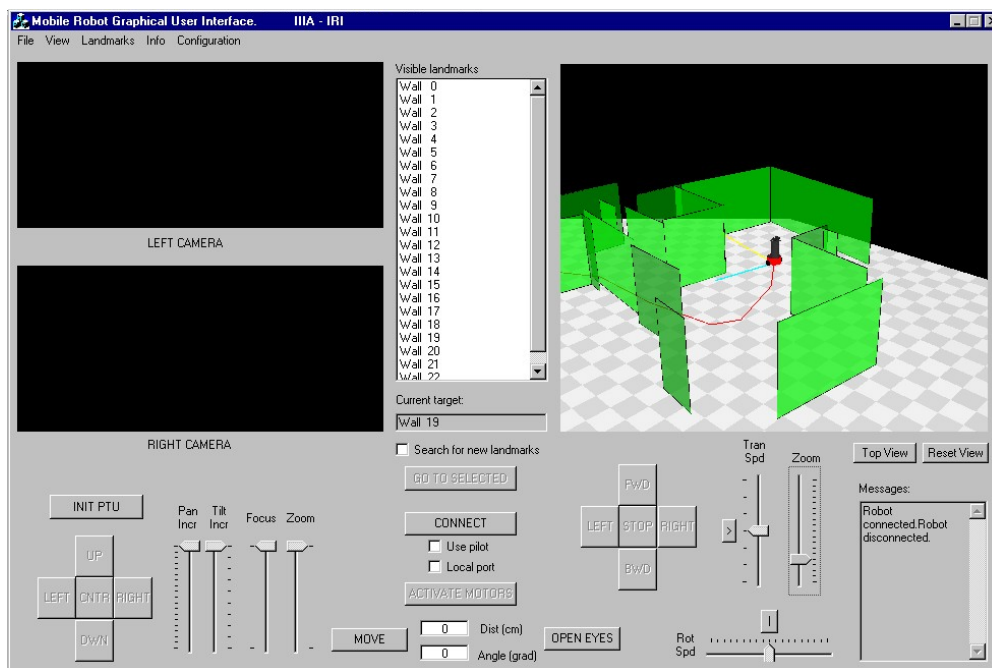


Figure 8.7: Mobile robotics graphical user interface developed jointly at IIIA and IRI. The interface helps researches develop and test algorithms on navigation, localization, and robot-human interaction with a variety of mobile robot platforms. The snapshot shows a test run of our map building system.

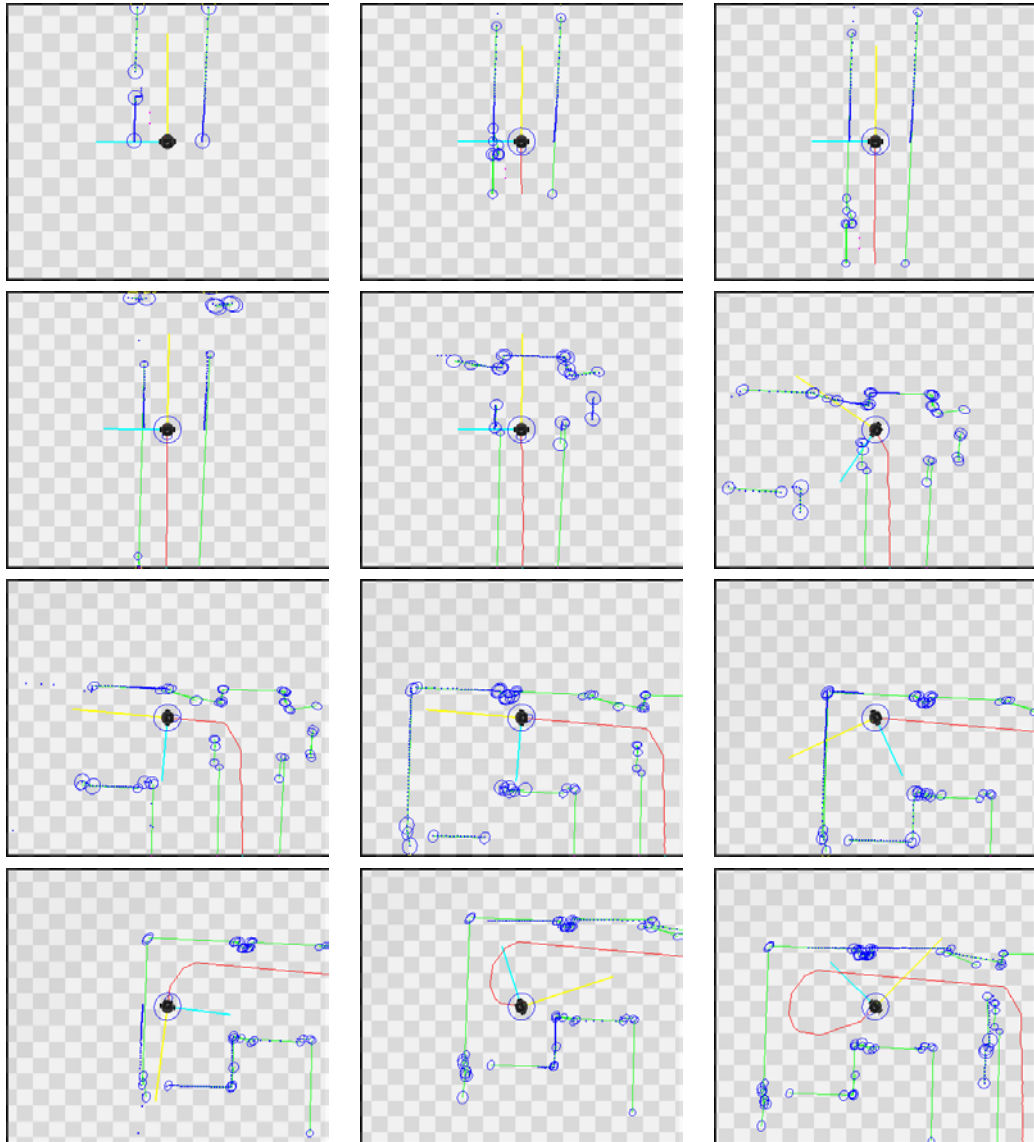


Figure 8.8: The blue dots indicate sensor raw data coming from a laser range finder. The green lines represent walls inferred from consecutive readings. The red lines indicate the estimated robot trajectory.



Figure 8.9: Concurrent mobile robot localization and map building.  
Front and back views of the robot initial position.

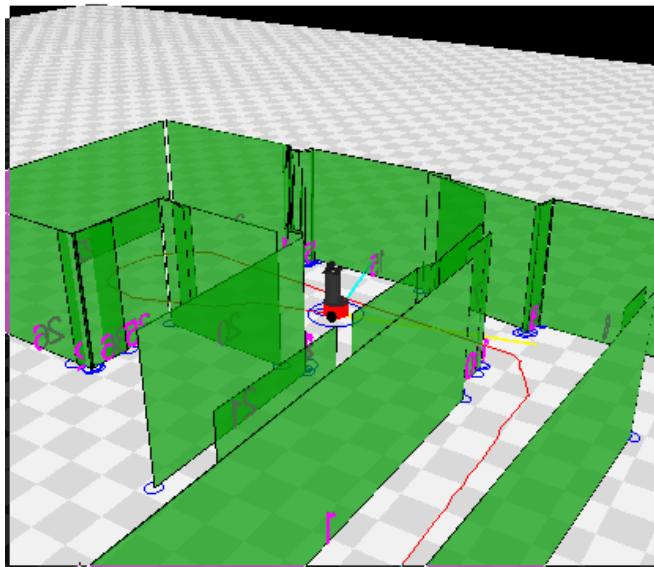


Figure 8.10: Graphical representation of the map built.

## **8.4 Conclusions**

We do not discard altogether the use of visual information in landmark characterization. Instead, we believe that in order to be able to autonomously produce fully rich and usable environment maps by a mobile robot, multiple sensor information must be fused. The algorithm proposed in Chapter 3, and our future research is aimed in that direction.





# Chapter 9

## Conclusions

### 9.1 Concluding remarks

This thesis presents contributions to the robust solution of the problem of autonomous environment learning for indoor service robots. The key contributions can be summarized in three major groups. A new technique to fuse visual data; new algorithms to validate landmarks with respect to their temporal quality; and a set of new theoretical results on the degree of reconstruction possible when building maps from fully correlated observations.

- A new technique to fuse visual data.
  - To univocally identify landmarks from sensor data, we studied several landmark representations, and the mathematical foundation necessary to extract the features that build them from images and laser range data.
  - The features extracted from just one sensor may not suffice in the invariant characterization of landmarks and objects, pushing for the combination of information from multiple sources.
  - We presented a new algorithm that fuses complementary information from two low level vision modules, namely depth from stereo and color-based segmentation, into coherent object models that can be tracked and learned in a mobile robotics context.
- Temporal landmark validation.
  - We consider robot motion and sensor observations as stochastic processes, and treat the problem of concurrent map building and localization from an

estimation theoretic point of view. The main drawback we encountered is that current estimation approaches to CML have been devised for static environments, and that they lack robustness in more realistic situations.

- To validate landmarks in terms of their temporal distribution, we presented two new techniques: the exponential decay rule, and the data association probability.
  - The exponential decay rule is a nonlinear expression that represents a memory cell associated with each landmark, registering its history of observations.
  - The data association probability rule is a more simple, linear expression that models the probability of future correct associations.
  - The simplicity of the linear expression makes the data association probability model a more appealing solution for real time implementations. However, the nonlinear expression yielded consistently better results. Furthermore, both techniques, when used in conjunction with the spatial compatibility techniques present in the literature, outperformed the use of spatial landmark compatibility tests alone.
  - Special care is taken in that the removal of landmarks from the map in terms of their temporal quality does not violate the asymptotic convergence and full correlation properties of CML.
- Building fully correlated maps.
    - This thesis gives also an in depth analysis of the fully correlated approach to CML from a control systems point of view. Considering that the Kalman filter is nothing else but an optimal observer, we end up revealing the strong limitations of using fully correlated noise driven observations to build a map and at the same time estimate the vehicle pose.
    - Full correlation of the map model in CML hinders full observability of the state estimate. Moreover, given that landmarks are fixed, the only controllable states are the ones associated to the vehicle. That is, only portions of the state vector are observable and controllable.
    - Partial observability makes the final map dependant on the initial observations, and does not guarantee convergence to a positive definite covariance matrix.
    - Partial controllability on the other hand, makes the filter believe it has accurate estimates of the landmark states, with their corresponding Kalman gains converging rapidly to zero. That is, after a few steps, innovations are useless.

- Partial observability can easily be solved by anchoring the map to the first landmark observed, or by having an external sensor that sees the vehicle at all times.
- Partial controllability on the other hand, can be palliated only by adding artificial process noise to the landmark estimates.
- As a byproduct, and from the formulation of CML as an LQG problem, one can exploit the duality between the Kalman filter and the optimal regulator, with the advantage of obtaining nicer, smoother, motion commands.
- The divergence caused by nonlinearities in CML can be reduced either by assuming larger uncertainty of the vehicle model by augmenting the process noise covariance, and with the online computation of measurement noise covariance matrices.

## 9.2 Future research directions

As we address issues such as viewpoint invariance and feature extraction from sensor data, it is overwhelming how undesired environment dynamics, occlusions, and sensor noise can still make data association a daunting task; specially when computer vision is used. Another topic requiring further study when using visual cues is the evaluation of landmark visibility with robust handling of occlusions and reflections. In the end, the building of maps purely from visual data is a very challenging problem, and whilst temporal landmark validation aids in palliating its difficulty, it by no means provides a panacea.

We believe that in order to be able to autonomously produce fully rich and usable environment maps by a mobile robot, multiple sensor information must be fused. The algorithm proposed in Chapter 3, and our future research is aimed in that direction.

Revealing the limitations of the fully correlated model to concurrent map building and localization with respect to the partial observability and controllability of the state space has prompted our attention to a similar problem in computer vision, that of computing structure from motion.

The computation of rich and accurate three-dimensional models of a scene from a video sequence is a hot research topic in the computer vision community. The problem resembles that of CML when the structure to be reconstructed and the camera motion are both unknown, regardless of the intrinsic calibration of the camera that generated the sequence.

### 9.3 Selected publications

Next is a list of our published work derived from this thesis, and the relevant citations:

- Book chapters
  - **Object recognition.** A survey on object recognition techniques in computer vision with a complete overview on the techniques for landmark characterization and graph algorithms. This survey comes as a byproduct of our stay at the Robot Vision Laboratory at Purdue University during 1997-1999, and appeared in the Wiley Encyclopedia of Electrical Engineering 2000 [7].
- Peer-reviewed journals
  - **Graph-based representations and techniques for image processing and image analysis.** A commented description on the use of graphs as a topological data structure for computer vision, published in Pattern Recognition, March 2002 [172].
  - **Concurrent map building and localization on indoor dynamic environments.** Algorithmic details of our map update algorithm, published in the International Journal of Pattern Recognition and Artificial Intelligence, May 2002 [11].
- International conferences
  - **Integration of perceptual grouping and depth.** Results on data fusion of color-based segmentation and depth from stereo. Talk presented at the 15th IAPR International Conference on Pattern Recognition, Barcelona, September 2000 [8].
  - **Pattern recognition research at the IRI-CSIC/ESAII group.** Research report presented at the 1st Foro Iberoamericano de Reconocimiento de Formas y Análisis de Imágenes, Barcelona, September 2000 [174].
  - **Localization of human faces fusing color segmentation and depth from stereo.** Use of Kalman filtering for tracking in a related application. Talk presented at the 2001 IEEE International Conference on Emerging Technologies and Factory Automation, Antibes, October 2001. [145].
  - **Learning of dynamic environments by a mobile robot from stereo cues.** These results on the automatic construction of environment maps follow those of [10], and were presented at the 2001 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems in Baden-Baden, August 2001 [9].

- **3D realtime head tracking fusing color histograms and stereovision.** Another contribution on the use of state estimation for tracking in computer vision. Poster presentation at the International Conference on Pattern Recognition, Quebec, August 2002 [146].
  - **Concurrent map building and localization with landmark validation.** First article covering in detail the theoretic foundation of our new temporal landmark quality models. Poster contribution in the International Conference on Pattern Recognition. August 2002 [12].
  - **Concurrent map building and localization with temporal landmark validation.** A more complete version will be presented at the 2003 IEEE International Conference on Robotics and Automation [13].
- National conferences
    - **Topological map learning for a mobile robot in indoor environments.** Preliminary results on the construction of maps for an indoor mobile robot. Talk presented at the 9th Spanish Symposium on Pattern Recognition and Image Analysis, Benicasim, May 2001 [10].
    - **MARCO: A mobile robot with learning capabilities to perceive and interact with its environment.** Research report on the different modules for the mobile robot platform MARCO. Poster contribution at the 9th Spanish Symposium on Pattern Recognition and Image Analysis, Benicasim, May 2001 [173].
  - Technical reports
    - **Desenvolupament d'un sistema d'estereovisió per un robot mòbil.** Research report on depth reconstruction from stereo. Institut de Robòtica i Informàtica Industrial, Universitat Politècnica de Catalunya, March 2001 [144].
    - **Camera calibration.** Research report on pattern based camera calibration. Institut de Robòtica i Informàtica Industrial, Universitat Politècnica de Catalunya, June 2001 [5].
    - **Kalman filter.** Research report on the use of the Kalman filter as an optimal state estimator. Institut de Robòtica i Informàtica Industrial, Universitat Politècnica de Catalunya, March 2002 [6].
    - **Mobile robot laser based map building.** Research report on map building from range data only, (in preparation). Institut de Robòtica i Informàtica Industrial, Universitat Politècnica de Catalunya, [44].



# Appendix A

## Landmark Extraction

This Appendix presents a detailed explanation of the algorithms chosen in Chapter 2 for the extraction of landmarks from images and range data. The discussion starts from the mathematical foundation necessary to extract points, lines, and regions from images, and lines from laser range data; incorporating qualitative examples of how the various algorithms behave over typical mobile robotic scenes.

### A.1 Low-level processing of 2d intensity images

Cameras have become widely popular sensors in mobile robotics. The reason behind is probably twofold. On the one side because of the significant price drop pushed by the consumer electronics market; but from a more philosophical point of view, because of its closeness to human perception. We as humans believe vision to be our most valuable sense, and are thus invariably inclined to incorporate such capabilities in the mobile machines we build.

Features of interest in 2d intensity images such as points or lines are usually indicated by those areas within the image with high discontinuity of the intensity levels. Associated with every pixel at coordinates  $(x, y)$ , there is an intensity value  $I$ . For gray-scale images,  $I$  is a scalar, whereas for color images,  $I$  will usually consist of a functional of the three color components red, green, and blue. The detection of discontinuities in image intensity can be achieved mathematically by computing the derivatives of the image intensity function  $I(x, y)$ , in either their grayscale or color representations.

The gradient of an image at each point is represented by the largest rate of intensity change at that point, and the direction of the gradient is along the direction of steepest

change. Without loss of generality, given the grayscale image intensity  $I(x, y)$ , the image gradient is given by

$$\nabla I(x, y) = \begin{bmatrix} I_x \\ I_y \end{bmatrix} \quad (\text{A.1})$$

The terms  $I_x$  and  $I_y$  indicate the partial derivatives of the intensity function  $I$  with respect to the directions  $x$  and  $y$ . In order to find the derivatives of the image in the  $x$  and  $y$  directions, several operators can be used, the most common being the Sobel operator. Table A.1 shows the image convolution masks that correspond to the Sobel, Prewitt, and Roberts operators. The two-dimensional image intensity function  $I$  has to be convolved with one of these operator sets, each of them possessing different characteristics regarding the sensitivity to the direction of discontinuities. For an image  $I$ , and a two-dimensional operator  $f$  of window size  $W$ , the discrete two-dimensional convolution is given by the double summation

$$I(x, y) * f = \sum_{i \in W} \sum_{j \in W} I(x - i, y - j) f(i, j) \quad (\text{A.2})$$

So for example, the gradient in the  $x$  direction of image  $I$  at point  $(x, y)$  as computed with the Sobel mask is simply

$$I_x(x, y) = I(x, y) * \nabla_x \quad (\text{A.3})$$

Furthermore, it is well known that if needed, the time complexity of the discrete two-dimensional convolution can be reduced from  $\theta(n^2)$  to  $\theta(n \log n)$  by means of the discrete fast Fourier transform. Even when exploited in our filter implementations, the topic is not further discussed here. The interested reader can find a complete overview of the technique in the algorithms book by Cormen *et al.* [50].

The overall gradient magnitude can be computed by the  $L_2$  norm

$$\|\nabla I(x, y)\|_2 = \sqrt{I_x^2 + I_y^2} \quad (\text{A.4})$$

For digital implementations however, the gradient magnitude calculated by using the  $L_\infty$  norm

$$\|\nabla I(x, y)\|_\infty = \max(I_x, I_y) \quad (\text{A.5})$$

gives more uniform results with respect to the direction of the edge [7]. The direction of the gradient at the image point with coordinates  $(x, y)$  is given by

$$\theta = \tan^{-1}(I_y/I_x) \quad (\text{A.6})$$

Quadratic expressions of the first and second order derivatives play also important



---



---

Sobel operators	$\nabla_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$	$\nabla_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$
Prewitt operators	$\nabla_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$	$\nabla_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$
Roberts operators	$\nabla_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$	$\nabla_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

---



---

Table A.1: Several differentiation masks can be used to approximate the derivatives of an image.

roles in the low-level extraction of image features as will be seen later. Suffice to say now that the square of the gradient is given by the rank one matrix

$$\mathbf{F} = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (\text{A.7})$$

that the second order derivative of  $I(x, y)$ , called the image Hessian, is given by

$$\mathbf{H} = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix} \quad (\text{A.8})$$

and that the trace of  $\mathbf{H}$  is called the Laplacian of  $I(x, y)$ , and is indicated by the operator  $\nabla^2$ . Furthermore, a smoothing operation is achieved by convolving the image  $I(x, y)$  with the Gaussian operator

$$\mathbf{G} = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (\text{A.9})$$

### A.1.1 Corners

One of the most simple features that can be extracted from images are corners. We mean by corners those points in an image where there is a steep change in intensity in more than one direction. Such points usually correspond to physical vertices on windows, walls or furniture in the scene, to their shadows, or to any other point with high curvature over any type of surface.

In 1978, Beaudet [28] proposed a rotationally invariant operator for corner extraction from images. The operator was derived from the second order Taylor series of the image intensities  $I(x, y)$ , and corresponds to the determinant of the image Hessian

$$K_B = I_{xx}I_{yy} - I_{xy}^2 \quad (\text{A.10})$$

Later, Kitchen and Rosenfeld [116] formulated a corner response function based on the change of gradient direction along an edge contour multiplied by the local gradient magnitude. Their corner extraction function had the form

$$K_K = \frac{I_{xx}I_y^2 + I_{yy}I_x^2 - 2I_{xy}I_xI_y}{I_x^2 + I_y^2} \quad (\text{A.11})$$

Another widely used operator for the characterization of corners on images is due to Harris [90], and has been formulated from principles of differential geometry

$$K_H = \det \mathbf{C} - k \text{tr} \mathbf{C} \quad (\text{A.12})$$

in which  $\mathbf{C}$  is the matrix that represents a first order Taylor series of the smoothed squared image gradient

$$\mathbf{C} = \mathbf{F} + \sigma^2 \mathbf{H}^2 \quad (\text{A.13})$$

As an exemplar, the top left plot in Figure A.1 shows a three-dimensional representation of an ideal smoothed image corner. The rest of the plots in the Figure indicate the output of the various corner response functions discussed in this section. All of these operators require the computation of second order derivatives on the image, which makes them not only very noise sensitive, but also inaccurate for precise localization. Nevertheless, these functions are usually used to obtain initial corner localization estimates, and to refine these estimates, several heuristics can be used, such as multiresolution, non-maximum suppression, and energy minimization [32, 59, 60, 67, 84, 129, 142]. We have implemented such techniques to refine the corner localization results given by the Beaudet, Kitchen and Harris corner operators. Qualitative results are shown on Figure A.2 for typical grayscale images in a mobile robotics setting.

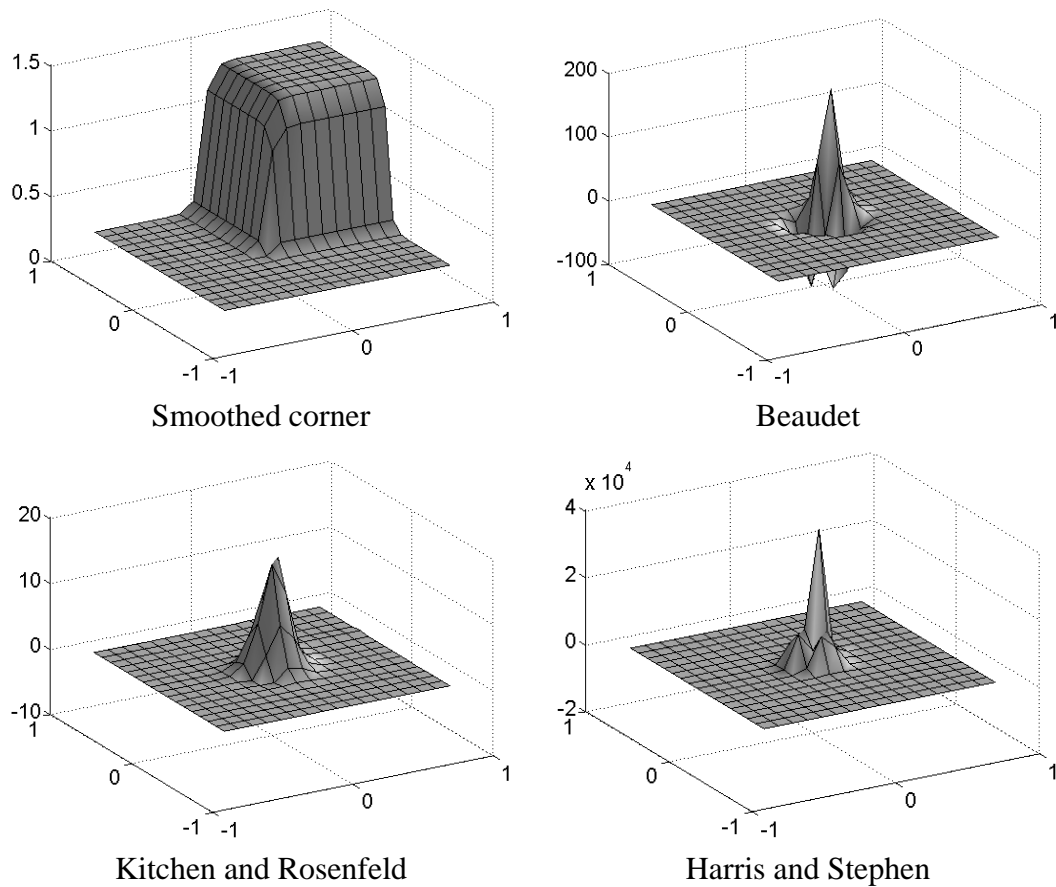


Figure A.1: Corner response functions.

We will now describe briefly our revised version of the variance descent approach originally proposed by Blaszkia and Deriche [32] for the refinement of corner localization on images. Consider the 1-D signal given by the pixels on the perimeter of a window around the hypothesized corner location. We start by computing the two most prominent edges in this 1d signal with multiresolution and non-maximum suppression as suggested by Deriche [59]. These two points are indicated by the white circles in Figure A.3. If the location of these two points is opposite to each other in the perimeter window, the corner is not valid. The detected point probably constitutes an artifact in a line, and the hypothetical corner point is not further considered. Otherwise, we compute the energy sum of gray level intensity variances for the two regions  $R_1$  and  $R_2$



Beaudet Features



Kitchen and Rosenfeld Features



Harris and Stephen Features

Figure A.2: Feature detection over a typical mobile robotics scene.

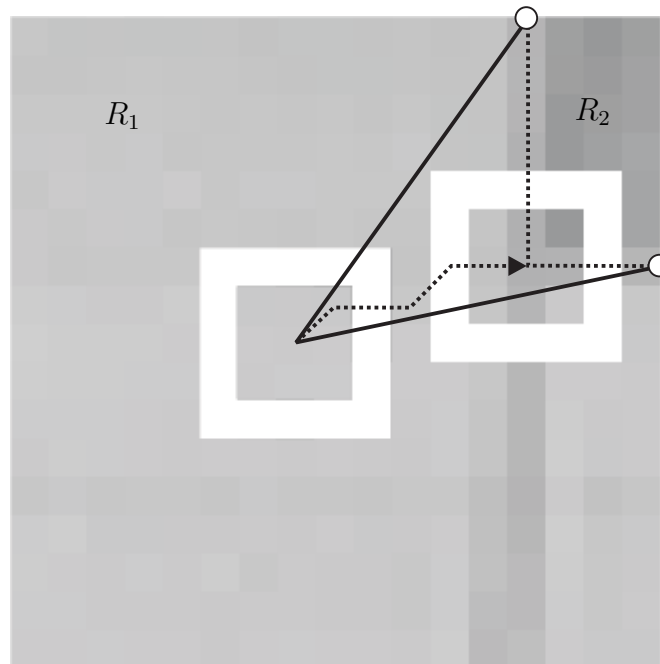


Figure A.3: Variance descent approach to corner refinement.

corresponding to the inside and outside of the corner hypothesis window

$$\varepsilon = \sum_{(x,y) \in R_1} (I(x,y) - \bar{I}_{R_1})^2 + \sum_{(x,y) \in R_2} (I(x,y) - \bar{I}_{R_2})^2 \quad (\text{A.14})$$

Next, we compute the energy sum for the 8 neighboring pixels as new localization hypotheses. The corner location estimate is shifted to the neighbor with the corresponding minimum energy sum. The algorithm is iterated, and converges to the corner location with global energy sum minima. If the two regions that characterize the corner were completely homogeneous, the energy sum would be zero. Otherwise, at the end of the algorithm, the energy sum may still be above a given threshold. In such cases, the refined corner estimate might correspond to a highly textured or noise corrupted area in the image. In those cases, the localized feature is also discarded.

### A.1.2 Shi and Tomasi features

In most cases, the vertex features found with the algorithms described in the last section will correspond to geometrical entities from the environment, such as corners on windows, walls or furniture, or their reflections. Unfortunately, the tracking from one frame to the next of such image features might still be hard to attain. Affine deformations caused by the change in viewpoint, or by the variation of the reflectance conditions contribute to such difficulty. With that in mind, Shi and Tomasi formulated an image feature selection algorithm optimal by construction from the equations of affine motion [181].

Starting from the assumption that a feature in the scene will have similar intensity values on two consecutive images  $I$  and  $J$ , the affine motion  $(\mathbf{D}, \mathbf{d})$  of a window of pixels around such feature  $\mathbf{m} = (x, y)^\top$  from the first image to the second can be represented with

$$I(\mathbf{D}\mathbf{m} + \mathbf{d}) \approx J(\mathbf{m}) \quad (\text{A.15})$$

and with the simplification that for small displacements  $\mathbf{D} = \mathbf{I}$ , a Taylor series approximation of the image intensity change is given by the expression  $I(\mathbf{m}) + \nabla^\top I(\mathbf{m})\mathbf{d}$ . We can then formulate a measure of dissimilarity for a tracked feature around window  $W$  in the two image frames simply as the sum of squared differences

$$\epsilon = \sum \sum_W (I(\mathbf{m}) + \nabla^\top I(\mathbf{m})\mathbf{d} - J(\mathbf{m}))^2 \quad (\text{A.16})$$

Differentiating A.16 with respect to the displacement  $\mathbf{d}$ , and setting the result equal to zero yields the system

$$\sum \sum_W (I(\mathbf{m}) - J(\mathbf{m}))\nabla I(\mathbf{m}) = \sum \sum_W \mathbf{F}(\mathbf{m})\mathbf{d} \quad (\text{A.17})$$

indicating that a feature centered at  $\mathbf{m}$  can be tracked reliably when the above system is well conditioned. We end up choosing as features the points in the image for which  $\sum \sum_W \mathbf{F}(\mathbf{m})$  has both eigenvalues larger than a given threshold. The chosen points will be located near corners, in highly textured regions, or in any other pattern that can be tracked reliably. Figure A.4 indicates the results of applying the Shi and Tomasi operator in our sample robotics setting.



Figure A.4: Shi and Tomasi features over a typical mobile robotics scene.

### A.1.3 Edges

Edges in images are located at pixels with discontinuity of intensity levels in one direction only, and can be easily computed by adding up the result of convolving the image intensity function  $I$  with any of the operator sets in Table A.1. These operators however, are non-isotropic, and they are biased towards detecting edges along the operator direction of steepest change.

On the contrary, another class of edge detectors that are isotropic by design uses second-derivative operations. The best known of these is the Laplacian of the Gaussian (*LoG*) edge detector [137]. In this method, the image intensity function  $I$  is smoothed with a Gaussian function and the edges are detected from the zero crossings of the Laplacian  $\nabla^2$ , the  $*$  symbol indicating the convolution operation.

$$I_{LoG}(\mathbf{m}) = \nabla^2(I(\mathbf{m}) * \mathbf{G}(\mathbf{m})) \quad (\text{A.18})$$

The *LoG* method for edge detection gives one-pixel wide closed contours, with the shortcomings that it can also produce spurious zero crossings caused by points of inflection in the first derivative, and the resulting contours are biased in the vicinity of locations where actual edges form acute angles occur.

The concept of optimality has also been considered in the formulation of more reliable edge detectors. Operators designed as optimum in the detection of extrema in the output of the convolution of the image with an impulse response include the Deriche, Spaceck, and Canny operators [74]; the most widely used being the Canny edge

detector [35, 36].

We will next describe the Canny operator in detail, as it was chosen our workhorse for the extraction of lines from images. It is from our own perspective, the most reliable edge extraction algorithm to date.

In this method, the idea is to maximize the output of the operator  $h(\mathbf{m})$  on an ideal step edge  $e(\mathbf{m})$ , whilst satisfying the robustness to noise, good localization, and uniqueness of response criteria. Without loss of generality, in the 1d case these criteria are given by the following expressions

$$\Sigma = \frac{\int e(-x)h(x) dx}{\sqrt{\int h^2(x) dx}},$$

$$\Lambda = \frac{|h'(0)|}{\sqrt{\int h'^2(x) dx}},$$

$$x_{max} = 2\pi \sqrt{\frac{\int h'^2(x) dx}{\int h''^2(x) dx}}$$

For a detailed derivation of these expressions see Canny's masters thesis and article [35, 36], or the book by Faugeras [74]. Using calculus of variations we can simultaneously find the extrema of  $\Sigma$ ,  $\Lambda$ , and  $x_{max}$  by solving the constrained optimization problem

$$\max \Sigma \Lambda$$

subject to

$$x_{max} = kW$$

which states that the average maximum distance between two local maxima has to be some fraction of the spatial extent of the operator. The solution of the above optimization problem leads to the optimal one-dimensional operator

$$h(x) = e^{-\alpha x}(a_1 \sin \omega x + a_2 \cos \omega x) + e^{\alpha x}(a_3 \sin \omega x + a_4 \cos \omega x) - \frac{\lambda_1}{2}$$

with conditions on  $\alpha$  and  $\omega$ :  $\alpha^2 - \omega^2 = \lambda_1/\lambda_2$  and  $4\alpha^2\omega^2 = (4\lambda_3 - \lambda_2^2)/4\lambda_3^2$ . The optimal operator is found by the method of Lagrange multipliers. The inquisitive reader would find the details in the aforementioned book [74].

Closed expressions can be computed for  $a_1$ ,  $a_2$ ,  $a_3$ , and  $a_4$  as functions of  $\alpha$ ,  $\omega$ ,  $c_3$ , and  $\lambda_1$ , resulting in  $h(x)$  parameterized in terms of  $\alpha$ ,  $\omega$ ,  $c_3$ , and  $\lambda_1$ . The problem of finding the optimal operator has been reduced from an optimization problem in





Sobel Edge Detector



Canny Edge Detector

Figure A.5: Edge detection over a typical mobile robotics scene.

an infinite-dimensional space (the space of admissible functions  $h$ ) to a nonlinear optimization problem with variables  $\alpha$ ,  $\omega$ ,  $c_3$ , and  $\lambda_1$ . These values are obtained using constrained numerical optimization methods. The optimal operator  $h(x)$  computed in this manner resembles the first derivative of a Gaussian  $h(x) = -(x/\sigma^2)e^{(-x^2/2\sigma^2)}$ .

Now, the solution thus found gives a 1d optimal operator for the detection of edges. In order to find 2d edges we still need to combine the results of convolving the original image with two oriented operators along the  $x$  and  $y$  directions, just as with the Sobel and Prewitt masks. The derivation of a 2d optimal isotropic operator from scratch is still an open research problem.

Most edge detection operators produce edges that are not connected. Before any higher-level scene interpretation modules can be brought to bear on an image, it is often

necessary to repair broken edges if such breakages were caused by noise and other artifacts. Edge repair is carried out by expanding and shrinking the detected edges in such a way that any connection made during expansion is not lost during the shrinking operation. In a binary image, two pixels are connected if there is a path of neighboring pixels linking them [24, 165]. They are 4-connected if the path can be followed by traversing along the  $x$  and  $y$  directions only. An 8-connected path is obtained by traversing along the  $x$  and  $y$  directions as well as in diagonal directions. All pixels connected to a given pixel in a set  $S$  of 1's form a connected component of  $S$ . If  $S$  has only one component then  $S$  is simply connected, otherwise it is multiply connected. The border  $S'$  of a set  $S$  is made up of those pixels of  $S$  for which at least one neighbor is in its complement  $\bar{S}$ . The  $i$ -th iteration of an expansion is given by

$$S^{(i)} = S^{(i-1)} \cup \bar{S}'^{(i-1)} \quad (\text{A.19})$$

and the  $i$ -th shrinking iteration is given by

$$\begin{aligned} S^{(i)} &= S^{(i-1)} \sim S'^{(i-1)} \\ &= \overline{\bar{S}^{(i-1)} \cup S'^{(i-1)}} \\ &= S^{(i-1)} \cap \bar{S}'^{(i-1)} \end{aligned} \quad (\text{A.20})$$

Other operations that may be performed on binary images include border following, thinning, and labeling [165]. Border following can be implemented using crack or chain coding schemes. This is, following the border of the elements in  $S$  using 4-connectivity or 8-connectivity. Thinning is similar to shrinking with the exception that the endpoints of the elements in  $S$  should not be deleted from the image. Labeling consists on assigning an entry in a database to every separately connected component of  $S$ .

A powerful and frequently used approach for grouping together the edge elements that form straight lines in an image is based on the concept of Hough transformation [100] that, in its more common implementation, maps a straight line in an image into a single point in  $(d, \theta)$  space,  $d$  and  $\theta$  being the two invariant parameters in the polar coordinate representation of a line. The  $(d, \theta)$  space is also known as the Hough space. A generalization of this approach can also be used for grouping together the detected fragments of smooth curves [23]. For Hough transform based extraction of straight lines, the distance of each edge pixel from the origin is computed by using

$$r = \sqrt{x^2 + y^2} \quad (\text{A.21})$$

and the orientation by

$$\phi = \tan^{-1} \left( \frac{y}{x} \right) \quad (\text{A.22})$$

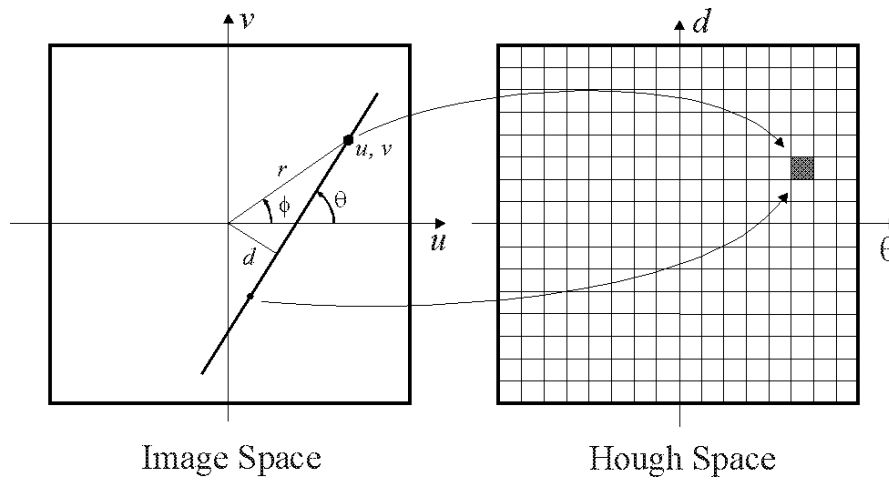


Figure A.6: The Hough transform maps straight lines in the image space into single points in the  $(d, \theta)$  space. It can be used to group together unconnected straight line segments produced by an edge operator.

as indicated in Figure A.6. The edge orientation angle  $\theta$  is obtained from the output of a Sobel or LoG operator, and the perpendicular distance from the image origin to the edge with point  $(x, y)$  and orientation  $\theta$  is

$$d = r \sin(\theta - \phi) \quad (\text{A.23})$$

Once  $d$  and  $\theta$  are computed, the corresponding cell in the Hough space is incremented. After processing the entire image, the lines corresponding to the cells with the highest number of hits are redrawn on top of the original image. Then, a raster scan is performed on the entire image to find the points near this line. This idea can be extended to extract curves from the output of an edge detector. Instead of using the parametric equation of a line, the generalized parametric equation for the desired curve must be used to define the Hough space, i.e., for circles

$$(x - x_0)^2 + (y - y_0)^2 = c^2 \quad (\text{A.24})$$

defines a 3d voting array with perpendicular directions  $x_0$ ,  $y_0$ , and  $c$ . If the output of the edge detector is not a binary image, the update values for the cells on the Hough space may be weighted with the intensity of the pixel being analyzed.

Another approach to boundary localization includes the use of active contours, or

*snakes* [115]. The classical snakes approach is based on deforming an initial contour curve towards the boundary of the object to be detected. The deformation is obtained by minimizing an energy function designed such that a local minimum is obtained at the boundary of the object. This energy function usually involves two terms, one controlling the smoothness and continuity of the contour curve and the other attracting it to the object boundary. The idea of active contours can also be extended to 3d object recognition by using 3d deformable surfaces [42]. In this case, instead of tracking the boundary of an object in a 2d image, the surface representation of the object is computed using 3d information, such as that obtained from a structured-light sensor.

The idea behind edge detection, or any other low-level process, is to prepare the image so that specific image components can be clustered. The clustering of image components into higher level organizations such as contours, each from a single object, is known as grouping or perceptual organization [133]. A grouping process can improve the search for an object in a recognition system by collecting together features that are more likely to come from the object rather than from a random collection of features. Most model-based recognition systems exploit such simple grouping techniques.

#### A.1.4 Image Segmentation

Multiple approaches exist for the segmentation of homogeneous regions from grey-level and color images. In our research, we have chosen a greedy algorithm that uses local variation information [75, 206]. The algorithm is briefly described next. A graph  $G = (V, E)$  is generated for the original image, where the nodes  $v_i \in V$  correspond to each pixel, and each edge  $e_{jk} \in E$  is weighted by a color distance measure among neighboring pixels.

The segmentation algorithm iterates over an increasingly sorted array of the elements in  $E$  to separate  $G$  into a forest of segmented surfaces  $(R_1, R_2, \dots)$  based on a color smoothness criterion, where the external variation must be larger than the internal variation between regions.

$$\text{Ext}(R_1, R_2) \leq \min(\text{Int}(R_1) + \tau(R_1), \text{Int}(R_2) + \tau(R_2)) \quad (\text{A.25})$$

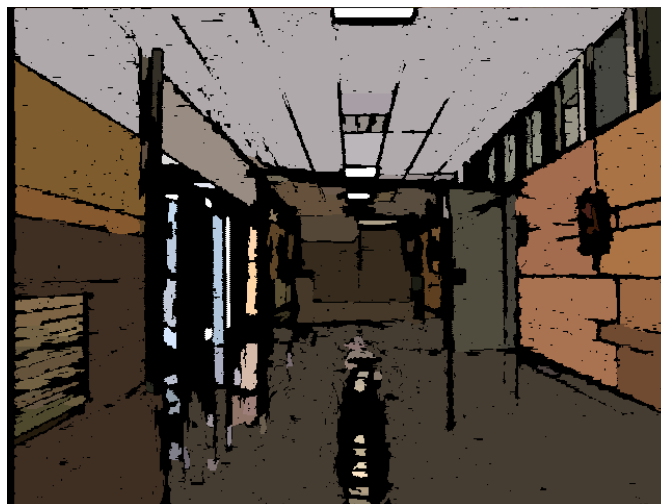
The external variation is defined as the lowest weight connecting two regions

$$\text{Ext}(R_1, R_2) = \min_{v_j \in R_1, v_k \in R_2} e_{jk} \quad (\text{A.26})$$

whereas the internal variation is the maximum weight in any minimum spanning tree



Original Color Image



Segmented Regions

Figure A.7: Color image segmentation

for any given region

$$\text{Int}(R) = \max_{e_{jk} \in \text{MST}(R)} e_{jk} \quad (\text{A.27})$$

The threshold function  $\tau$  controls the degree to which the external variation can actually be larger than the internal variation. For example, by keeping the size of any

given region from growing indefinitely,

$$\tau(R) = k/|R| \quad (\text{A.28})$$

The algorithm maintains the segmentation using a disjoint-set forest with union by rank and path compression [50], keeping the overall time complexity linear with respect to the number of pixels in the image. The details of the algorithm are presented as a contribution of our research group [206]. We show in Figure A.7 a color based segmentation of a typical mobile robotics scene.

### A.1.5 Depth from stereo

It is sometimes possible to extract depth information from images when more than one image is available, either because they are taken with a stereo rig, or because they belong to a temporal series. For stereo vision to work robustly, one must solve the correspondence problem: the matching of pixel pairs in the left and right images that correspond to the same point in 3d space.

Several geometric constraints can be used to alleviate the correspondence problem. One of these constraints is termed the feature constraint, and refers to the fact that what might be an edge in the left image will most likely correspond to an edge in the right image also. Generally speaking, the characteristics of the neighboring pixels for the matching of the left and right image points should be consistent. For Lambertian (completely matte) surfaces, the reflected light is the same in all directions, and as a result, the intensities at two corresponding points in the left and right images would be the same. In practice, few surfaces tend to be Lambertian. By the same token, few surfaces tend to be completely glossy. In reality, for most surfaces the reflected light will vary slowly with the direction of viewing. Another constraint is the epipolar constraint which states that for any point in the left image its possible matches in the right image all lie on the epipolar line, therefore reducing the dimensionality of the search space from two dimensions to one. One epipolar line being the projection on the right image of the line that passes through the center of projection and the pixel point on the left image. Other constraints include uniqueness, continuity, and ordering of points [74].

Full advantage of the epipolar constraint when comparing intensity levels for image correspondence is achieved if the images are rectified such that pairs of conjugate epipolar lines become collinear and parallel to one of the image axes as indicated in Figure A.8. Epipolar image rectification is achieved by computing the linear transformation that projects original images onto a common plane. Given the camera perspective projection matrices  $\mathbf{P}_i = [\mathbf{P}_i | \tilde{\mathbf{p}}_i]$  (see Appendix B) that generated the original images, one must compute a new pair of projection matrices  $\tilde{\mathbf{P}}'_i$  such that image points  $\tilde{\mathbf{x}}_i$  are

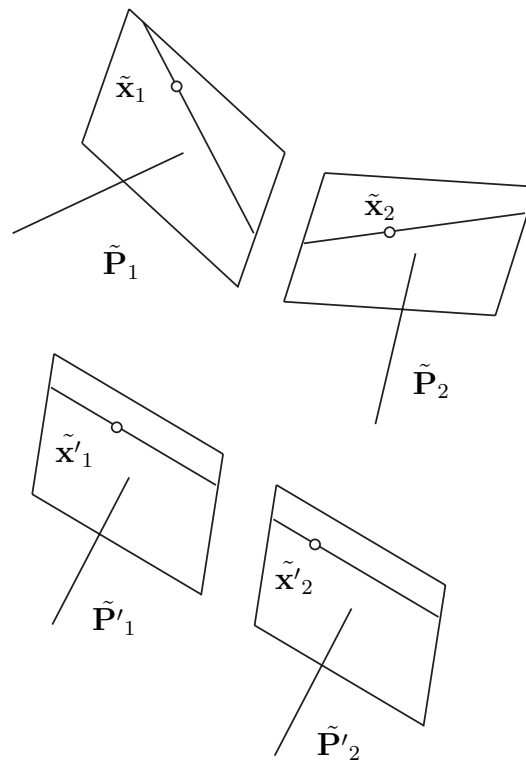


Figure A.8: Epipolar image rectification makes conjugate epipolar lines collinear and parallel to one of the image axes. This transformation reduces the time complexity of the search for matches in dense stereo from a two dimensional search problem to only one dimension.

warped with

$$\tilde{\mathbf{x}}'_i = \mathbf{P}'_i \mathbf{P}_i^{-1} \tilde{\mathbf{x}}_i \quad (\text{A.29})$$

To compute the rectifying projection matrices  $\tilde{\mathbf{P}}'_i$  one must set up a homogeneous system of linear equations with the sufficient constraints to generate the desired rectification. The system of equations explicitly incorporates a set of quadratic expressions for the geometric relations among the elements in  $\tilde{\mathbf{P}}'_i$  that ensure a nontrivial solution. These restrictions are summarized in Table A.2.

The solution to the resulting constrained system of equations can be casted as consecutively solving four generalized eigenvector problems of the form

$$\mathbf{A}\mathbf{x} = \mathbf{0}$$

Scale ambiguity constraint	$\ \mathbf{p}'_{3,i}\  = 1$
Preservation of the optical center	$\mathbf{P}'_i(-\mathbf{R}_i^\top \mathbf{t}_i) + \tilde{\mathbf{p}}'_i = \mathbf{0}$
Sharing of the focal plane and alignment of conjugate epipolar lines	$\mathbf{p}'_{2,1} = \mathbf{p}'_{2,2}, \mathbf{p}'_{3,1} = \mathbf{p}'_{3,2},$ $p'_{24,1} = p'_{24,2}, p'_{34,1} = p'_{34,2}$
Orientation of the new retinal plane	$\mathbf{p}'_{3,1}^\top (\mathbf{p}_{3,1} \times \mathbf{p}_{3,2}) = 0$
Orthogonality of the rectified reference frame	$\mathbf{p}'_{1,i}^\top \mathbf{p}'_{2,i} = 0,$ $\mathbf{p}'_{1,j}^\top \mathbf{p}'_{2,i} = 0$
Fixation of the principal point	$u_0 = \mathbf{p}'_{1,i}^\top \mathbf{p}'_{3,i}, v_0 = \mathbf{p}'_{2,i}^\top \mathbf{p}'_{3,i}$
Preservation of the focal length	$\ \mathbf{p}'_{1,i} \times \mathbf{p}'_{3,i}\ ^2 = \alpha_u^2,$ $\ \mathbf{p}'_{1,j} \times \mathbf{p}'_{3,i}\ ^2 = \alpha_u^2,$ $\ \mathbf{p}'_2 \times \mathbf{p}'_3\ ^2 = \alpha_v^2$

Table A.2: Epipolar rectification geometric constraints.

subject to

$$\|\mathbf{x}\| = k$$

and its solution is given by the eigenvector associated to the smallest eigenvalue of  $\mathbf{A}$ . The details of the method are given in the book by Faugeras [74], and more specifically, in a talk by Fusiello *et al.* [80].

If the original calibration matrices  $\tilde{\mathbf{P}}_1$  and  $\tilde{\mathbf{P}}_2$  that generated the scene are not available, one can still perform epipolar rectification, provided a set of image correspondences, such as the Harris corners, or the Shi-Tomasi features are available. The method requires scene geometry reconstruction and will be analyzed in detail in Section 4.3. Figure A.9 shows a set of 8 virtual landmarks in our exemplary mobile robotics scene, which in turn generate a parallelepiped. Their image correspondences are indicated, as



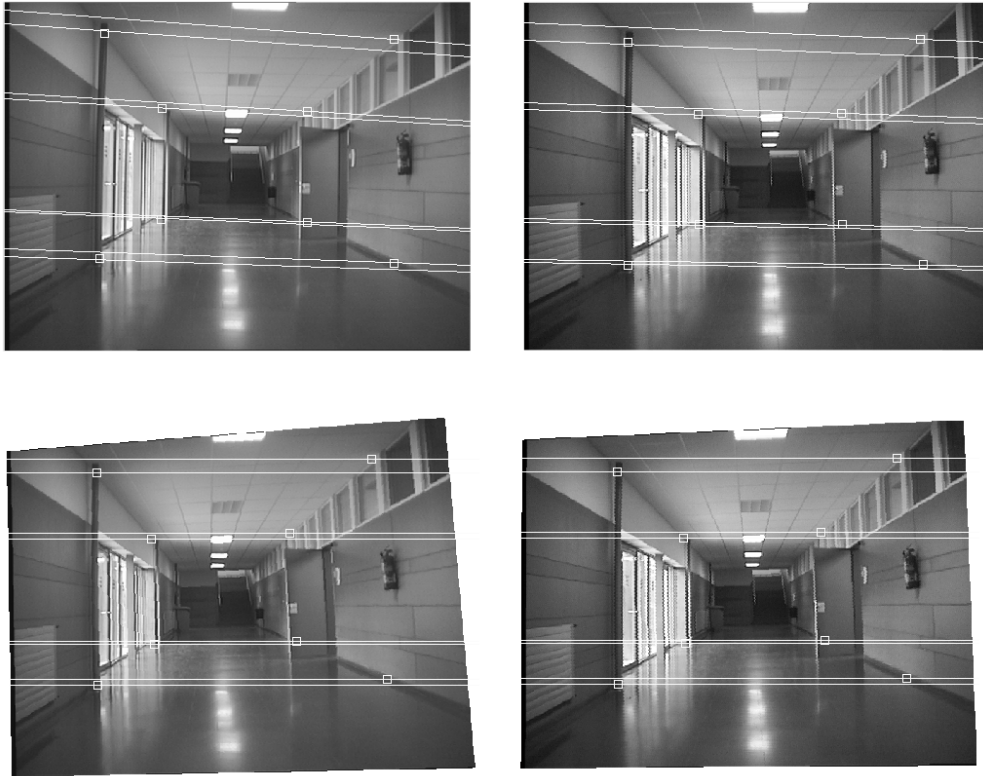


Figure A.9: Epipolar image rectification over a typical mobile robotics scene.

well as their respective epipolar lines. The bottom set of images has been rectified using this latter technique. Note the collinearity of the epipolar lines.

Once epipolar image rectification has been performed, we can proceed to the computation of stereo disparity. Most stereo matching algorithms compute the disparity between corresponding pixels as the distance that minimizes the sum of the squared differences of intensities over a window of interest around the image point  $\mathbf{m}$ , and given epipolar rectification, the dimensionality of the disparity vector reduces to the form  $\mathbf{d} = [d, 0]^T$ .

$$d = \arg \min_d \sum \sum_W (I(\mathbf{m} + \mathbf{d}) - J(\mathbf{m}))^2 \quad (\text{A.30})$$

To eliminate noise and depth estimation errors due to occlusions and reflectance variations, the disparity values resulting from matching points in the left image to points

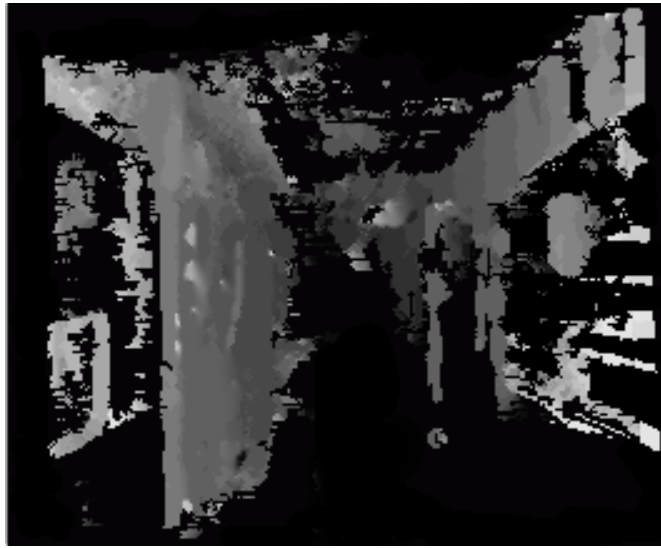


Figure A.10: Disparity map computed from the set of rectified images in Figure A.9. Image thanks to F. Moreno.

in the right image are compared to the disparities obtained when matching in the opposite direction, i.e., from right to left. Only those disparity values that coincide in both directions are used at further processing steps.

The resulting disparity map is a set of points for which a disparity value  $d$  associated to the image point  $\mathbf{m}$ , is inversely proportional to the distance  $z$  at which objects are located from the camera. Given both camera calibration matrices, the computation of the three-dimensional scene point  $\mathbf{x}$  from the image correspondences  $\mathbf{m}$  and  $\mathbf{m} + \mathbf{d}$  is straightforward. Disparity maps obtained with stereo methods based on the area of local regions (correlation, sum of squared differences, or sum of absolute differences) are dense where the scene is rich in detail, and sparse in smooth regions.

A robust approach to stereo vision may use object-level knowledge in the stereo fusion process [120]. This is, the stereo system will use object model knowledge to extract from each of the images higher-level pixel groupings that correspond to discernible features on one or more surfaces (or edges) of the object, and will try to solve the correspondence problem using only these features. Because these features tend to be distinctive, the problem of contention in establishing correspondence is minimized.

The disparity map shown in Figure A.10 was obtained by evaluating Equation A.30

only over the regions of interest on the pair of rectified images from Figure A.9, followed by a set of refining heuristics, including left-to-right and right-to-left correspondence, and region growing with morphological operations. A detailed description of the implementation appears in our technical report [144].

## A.2 Low-level processing of laser data

An efficient algorithm for the extraction of straight line segments from noisy data, originally by Douglas and Peucker [65], was conceived primarily to reduce the number of points required to represent a vector encoded polygonal digitized line. Within the cartography community, this problem is referred as the *line simplification problem*. Douglas and Peucker's approach to line simplification is probably the most cited algorithm when it refers to cartographic generalization. Notwithstanding, the idea of extracting straight line segments from digitized data is applicable not only to cartography and geographic information systems, but to diverse contexts such as freehand drawing simplification or handwrite generalization. We also find the technique nicely applicable to the extraction of straight landmarks from laser and sonar data in mobile robot indoor navigation, as is the case with walls and furniture.

A set of ordered points in a plane,  $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_n$ , forms a polygonal chain of line segments  $\overline{\mathbf{v}_0\mathbf{v}_1}, \overline{\mathbf{v}_1\mathbf{v}_2}, \dots, \overline{\mathbf{v}_{n-1}\mathbf{v}_n}$ . Given a chain  $C$  with  $n$  segments, the line simplification problem asks for the best chain  $C'$ , with fewer segments that represents  $C$  *well*. Furthermore, we assume that the chain  $C$  is simple, that is,  $C$  has no self-intersections.

Representing *well* has many possible meanings. For example, that  $C$  and  $C'$  are close to each other, that the area between  $C$  and  $C'$  be small, or that other measures of curve discrepancy be small. A recursive version of the original algorithm is shown in Table A.3. The key point of the algorithm resides in the technique used to find the appropriate point where to subdivide a line into consecutive line segments. In the original method, subdivision takes place at the farthest point in  $C$  from  $\overline{\mathbf{v}_i\mathbf{v}_j}$ . Table A.4 shows the algorithm to find the farthest point to a line in a stream of planar data points. The notation  $\tilde{\mathbf{v}}$  indicates homogeneous coordinates, i.e., three dimensional points with the last dimension equal to one.

The original algorithm has quadratic worst-case time complexity. However, a revised version of the algorithm by Hershberger and Snoeyink [96] that uses path hulls and the geometric structure of the problem allows for a reduction of the time complexity of the algorithm to  $O(n \log n)$ . In our map building implementation [44] we have opted for this latter version. The reader is referred to the proceedings from the talk by Hershberger and Snoeyink [96] for the details of the algorithm. Figure A.11 shows the results of line simplification from raw depth data with a Leuze Rotoscan RS4 and our

---



---

```

LineSimplification( $C, i, j$ )

  FindSplit( $C, i, j$ )
  if  $d > \epsilon$ 
    LineSimplification( $C, i, f$ )
    LineSimplification( $C, f, j$ )
  else
    return  $\overline{\mathbf{v}_i \mathbf{v}_j}$ 

```

---



---

Table A.3: Recursive version of the Douglas and Peucker algorithm to line simplification.

---



---

```

FindSplit( $C, i, j$ )

   $d = 0$ 
  for  $k = i + 1$  to  $j - 1$ 
     $dk = \frac{|\tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_j \tilde{\mathbf{v}}_k|}{\|\mathbf{v}_i - \mathbf{v}_j\|^2}$ 
    if  $dk > d$ 
       $d = dk$ 
       $f = k$ 
  return  $f, d$ 

```

---



---

Table A.4: Algorithm to find the vertex  $\mathbf{v}_k$  farthest from the line  $\overline{\mathbf{v}_i \mathbf{v}_j}$ , and the squared distance to it.

mobile robot Marco.

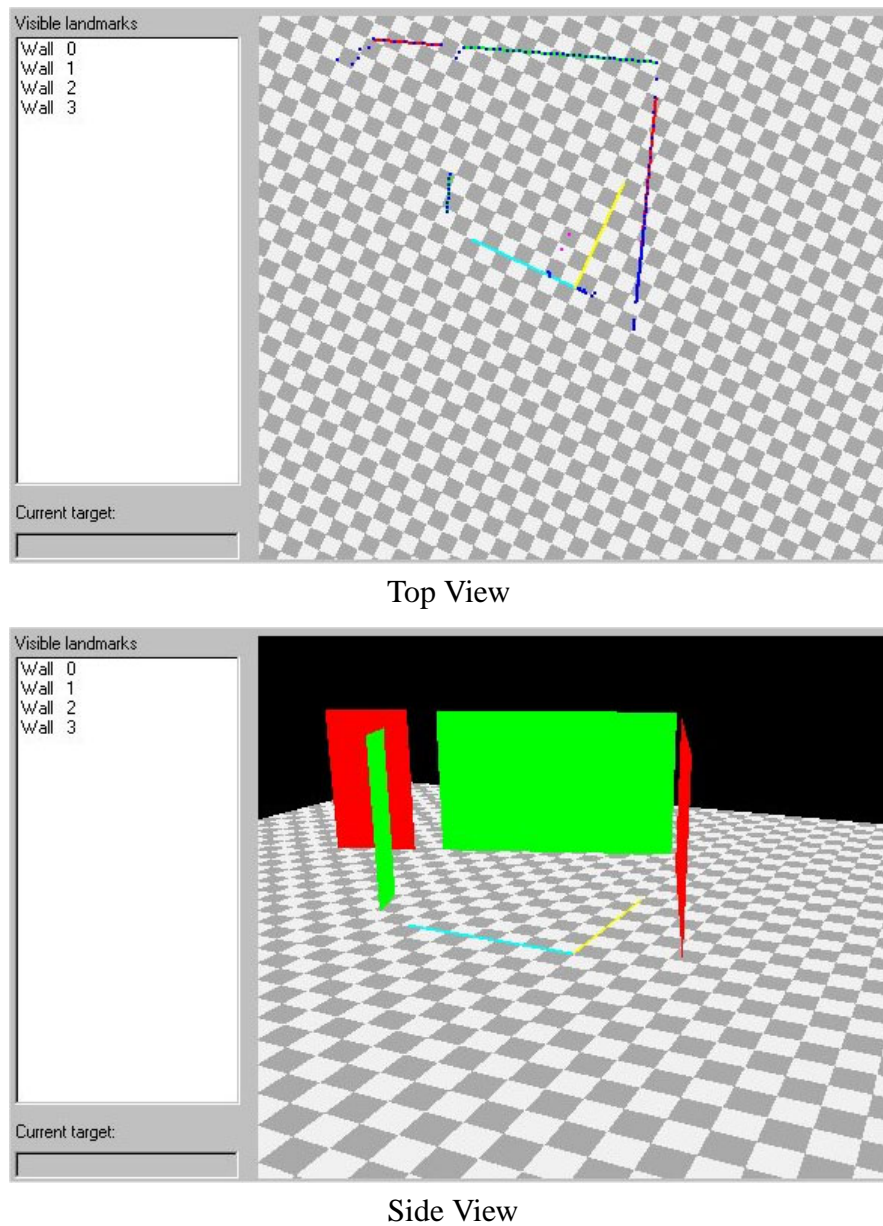


Figure A.11: Wall extraction by means of line simplification from a typical laser scan. The blue dots in the top view indicate the actual laser readings. The fitted walls are represented by the red and green lines (planes). Image thanks to A. Checa.

### A.3 Landmark characterization

Once a set of features such as corners, lines or segments has been extracted, we need to store their appearance properties in a feature vector. The kind of feature properties to save depend in the type of feature extracted. In the case of simple geometrical entities such as the lines extracted from range data for example, the end points coordinates would suffice. For the case of image segmented regions, one would like to store other geometric attributes such as its area, perimeter, elongation, moments of inertia, or spatial frequency descriptions such as the discrete cosine transform, Fourier descriptors, wavelets, or eigenimages. Since motion is our main concern, the attributes selected must be as invariant to changes in the feature pose as possible.

We will describe next the specific appearance properties chosen to characterize landmarks for the localization and map building algorithms developed as part of this thesis. For the characterization of image points, the feature vector is simply composed of the image pixel coordinate, as well as the normalized image intensity values for every point in a region of interest around such pixel. The same region that was used during the convolution operations for the extraction of such feature. Typically,  $7 \times 7$ ,  $9 \times 9$  or  $11 \times 11$  window sizes.

Along with the normalized intensity values (three dimensional intensity vectors in the case of color images), we stored also the result of the corner detection operator. Namely, a scalar value with the result of the Beaudet, Harris or Shi-Tomasi corner measure.

Now, if these same points were correlated for the computation of stereo, a few more parameters were added to the feature vector. These are the disparity value obtained, and ultimately the camera centered 3d coordinates of the point.

Lines extracted from monocular and stereo images were characterized by their ending points. So, their representation was seemingly extended from the aforementioned description for points. In the case of lines extracted from laser range data in the other hand, much simpler feature properties were saved. Since depth information is readily available from the laser range scan, the coordinates of points in the scan plane were stored. As an alternative, we also stored the parametric representation of such lines, given by the slope and the  $y$  intercept with respect to the laser coordinate frame. Given the limited properties of a laser scan, these were the only geometric properties they could bear for the description of the environment. Nevertheless, these descriptions constitute sufficient data for the construction of maps and the localization of a mobile vehicle in indoor areas.

Although we have not specifically used image segments for the building of maps,

these were used for the extraction and description of object models. The related application of image segmentation along with depth from stereo for object characterization is discussed in detail in Chapter 3. Similarly, Chapter 7 describes the techniques used to infer higher level features such as walls, from the lines extracted from laser range data.

## **A.4 Bibliographical notes**

A large portion of the contents in this Appendix has been adapted from our survey in object recognition [7]. For further information on introductory image processing techniques, the reader is referred to the bibliography in that article, and to the following books [24, 41, 85, 103, 137, 165, 209] among others.





# Appendix B

## Camera Calibration

This appendix is a tutorial on pattern-based camera calibration for computer vision. The methods presented here allow for the computation of the intrinsic and extrinsic parameters of a camera.

Camera calibration is a common and vital task for any computer vision application requiring metric reconstruction. The problem consists on that given one or a set of images, we must devise the intrinsic and extrinsic parameters of the camera that rendered such images. We understand for intrinsic parameters, the set of parameters that do not change regardless of the position and orientation of the camera in space, such as the focal length or the lens distortion; and as extrinsic parameters, the homography that defines the pose of the camera with respect to a reference coordinate frame.

There exist many methods for camera calibration depending on the pattern used for calibration; namely, 3-d patterns, 2-d patterns, known camera motion, and unknown camera motion, or the combination of patterns and motion. The degree of reconstruction is tightly dependent on the type of pattern and motions used for calibration, varying from Euclidean to affine to only projective.

In recent years there have been remarkable efforts on the calibration of cameras with lesser to null information about the structure of the scene. This set of methods is commonly referred as self camera calibration, and they rely heavily in the mathematical tools of perspective geometry. The main drawback is that they can provide reconstruction only up to an unknown scale factor. It remains a difficult problem to obtain accurate Euclidean reconstruction of a scene following these techniques. Most if not all of these techniques start from the assumption that the projection of a conic at infinity in the image plane does not change regardless of the camera viewpoint. By exploiting this fact, they separate intrinsic from extrinsic calibration.

Nevertheless, the computation of the intrinsic and extrinsic parameters of a camera with the aid of a 3-d pattern is still of great importance; either to validate these emerging less restrictive calibration methods, or for its use in applications where calibrating with a 3-d pattern is feasible. We will concentrate on the mathematical description and some implementation details of a robust method for camera calibration for Euclidean reconstruction when a 3-dimensional calibration pattern is available. Any other method is out of the scope of this text.

## B.1 Camera modeling

In this section we present the mathematical tools needed to understand the geometry involved in the perception of 3-dimensional points by a CCD camera. We will derive a set of equations that start from a point in  $\mathbb{R}^3$  and end with its corresponding pixel coordinates in an image. The section includes an explanation of the underlying physics related to the image formation process; including perspective projection, rigid transformation, CCD pixel size, non-perpendicular image axes, and image distortion.

### B.1.1 Perspective projection

To date, most computer vision systems that require geometric reconstruction use the pinhole camera model as a projective engine. It is with the aid of projective geometry that we can represent the projection of a point in  $\mathbb{R}^3$  through a focal point and onto an image plane as a linear transformation. This application is a special kind of projection called perspective projection.

Consider a camera-centered coordinate system, where the origin  $O$  represents the camera focal point and the  $z$  axis coincides with the optical axis. The optical axis intersects the image plane  $I$  at a focal distance  $f$  from  $O$  and  $I$  is parallel to the  $xy$  plane, See Fig 1.

Let  $\mathbf{x} \in \mathbb{R}^3$  represent the Euclidean coordinates of any visible point in our camera-centered coordinate system, and let  $\tilde{\mathbf{q}} = [U, V, s]^T \in \mathbb{P}^2$  be the homogeneous coordinates of its perspective projection from  $O$  onto  $I$ . In projective coordinates however,  $\tilde{\mathbf{q}} \sim \lambda \tilde{\mathbf{q}}$ . This is,  $\tilde{\mathbf{q}}$  is only defined up to a scale factor, and any point in the ray that passes through the focal point and  $\mathbf{x}$  is also represented by  $\tilde{\mathbf{q}}$ . For the case when  $s = x_3$  the point lies on  $I$ , and  $\mathbf{q} = [U/s, V/s]^T$  give the coordinates of that image point. Moreover, the homogeneous coordinates of a point with  $s = 0$  represent a ray parallel to the image plane, or an ideal point at infinity in  $I$  for the direction  $[U, V]^T$ .

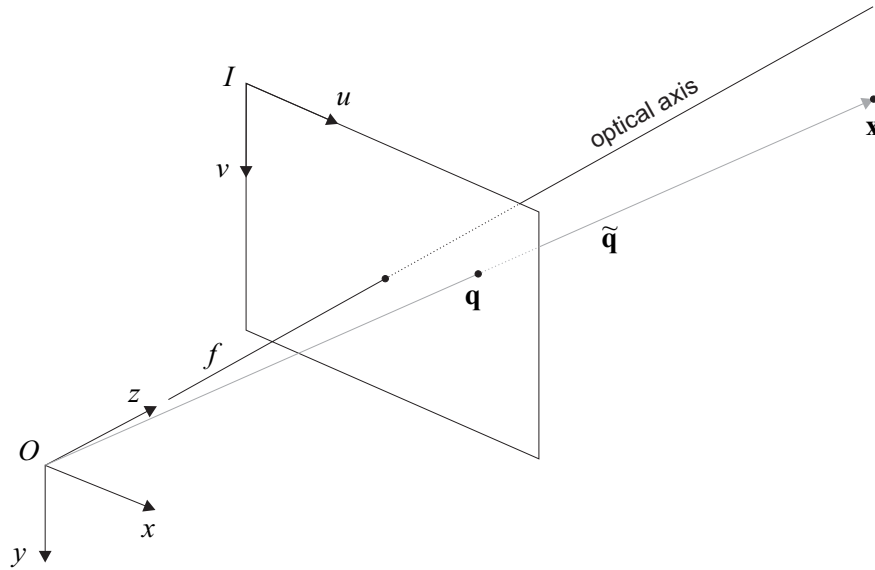


Figure B.1: Pinhole camera model.

To express the application of  $\mathbf{x}$  from  $O$  onto  $I$  as a linear transformation, we must also represent the coordinates of our visible point in homogeneous coordinates, i.e.,  $\tilde{\mathbf{x}} = [x_1, x_2, x_3, 1]^\top$ . Then, the perspective projection of  $\tilde{\mathbf{x}}$  from  $O$  onto  $I$  is given by

$$\tilde{\mathbf{q}} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tilde{\mathbf{x}} \quad (\text{B.1})$$

### B.1.2 Pattern reference frame

Now, consider our point  $\mathbf{x}$  to be an object point from the calibration pattern shown in Figure 2. The coordinates of that same point but with respect to the pattern reference frame are now  $\mathbf{x}_w$ , and the homography that relates pattern-referenced coordinates to camera-centered coordinates is given by

$$\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tilde{\mathbf{x}}_w \quad (\text{B.2})$$

Equation B.2, expressed as a projectivity in  $\mathbb{P}^2$ , is equivalent to  $\mathbf{x} = \mathbf{R}\mathbf{x}_w + \mathbf{t}$ ; the former expression being an Euclidean or rigid body transformation. Rigid body

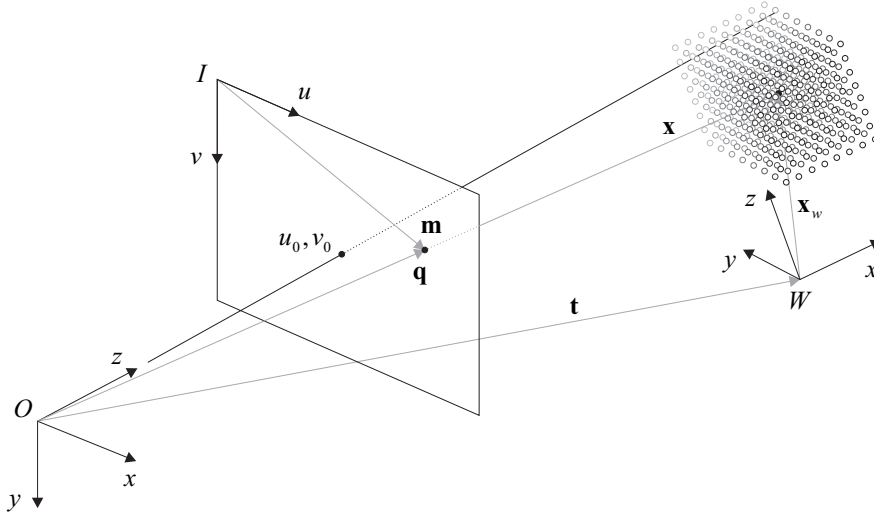


Figure B.2: Pattern reference frame.

transformations are a special case of transformations within the projective group, where lengths and angles are preserved. This is, the point  $x_w$  is rotated by  $\mathbf{R}$  and translated by  $\mathbf{t}$  to obtain  $\mathbf{x}$ . An important property of  $\mathbf{R}$  is that it is orthonormal, i.e.,  $\mathbf{R}^\top \mathbf{R} = \mathbf{I}$ .

### B.1.3 Pixel size

We will denote the image point in pixel coordinates with the symbol  $\mathbf{m}$ . The various subscripts on the following formulae indicate the different artifacts that contribute to the location of that point on the image.

The coordinates of  $\mathbf{q}$  are still given in the same units as our pattern and camera reference frames (say millimeters). To change them so that they represent image pixels, we insert appropriate scaling factors  $\alpha_u = f/k_u$  and  $\alpha_v = f/k_v$ .  $k_u$  and  $k_v$  are the pixel width and height of the CCD in millimeters, respectively.

$$\tilde{\mathbf{m}}_{ps} = \begin{bmatrix} \alpha_u & 0 & 0 & 0 \\ 0 & \alpha_v & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tilde{\mathbf{x}}_w \quad (\text{B.3})$$

### B.1.4 Image center

One last transformation is needed. Often the origin of the image plane is located in the upper left corner of the image, with the positive direction of the  $y$  axis matching the row index of the image coordinates. A translation by  $u_0$  and  $v_0$ , the image center coordinates, is still necessary. On a simplified pinhole model, the column and row image indices indicating the pixel that matches a pattern-referenced object point are thus given by

$$\tilde{\mathbf{m}}_{ic} = \begin{bmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tilde{\mathbf{x}}_w \quad (\text{B.4})$$

### B.1.5 Non orthogonal image axes

If the image axes are not exactly orthogonal, we denote the angle between them by  $\theta$

$$\tilde{\mathbf{m}}_{noia} = \begin{bmatrix} \alpha_u & -\alpha_u/\tan\theta & u_0 & 0 \\ 0 & \alpha_v/\sin\theta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tilde{\mathbf{x}}_w \quad (\text{B.5})$$

### B.1.6 Lens distortion

The simplicity of the pinhole model for camera representation allows for robust camera calibration in most generic situations. However, it is necessary to extend this model if we wish to accommodate for the nonlinear lens distortion artifacts that are present when wide area lenses are used, when zooming, or when the objects of attention are at a distance close to the focal length. The most typical of these artifacts are radial and decentering distortions.

Radial distortion accounts for the displacement of pixel points along radial directions from the center of the lens. Decentering distortion is usually caused by improper lens assembly, and might shift ideal image points in both radial and tangential directions. From these two types of distortion, radial distortion is the most critical, whereas decentering distortion can for all practical purposes be neglected [219].

Distortion correction can be modeled by a power series in the radial distance  $r = \sqrt{(u - u_0)^2 + (v - v_0)^2}$ . The following expression has been derived considering only the first term of such an expansion. Also, the image pixel location  $\tilde{\mathbf{m}}$ , taken from either Equation B.4 or Equation B.5, must be normalized by  $s$  prior to its use in Equation B.6.

The updated pixel location after radial rectification would be

$$\tilde{\mathbf{m}}_{rd} = \tilde{\mathbf{m}} + [k_1, k_2, 0]^\top \tilde{\mathbf{m}}^\top \begin{bmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ -u_0 & -v_0 & u_0^2 + v_0^2 \end{bmatrix} \tilde{\mathbf{m}} \quad (\text{B.6})$$

## B.2 Camera calibration

In this section we will derive an algorithm for the computation of both the intrinsic and extrinsic parameters of a camera. For this algorithm to work, we must provide a set of at least 6 non-coplanar 3-d to image point correspondences. Although in general, the number of correspondences in the set is usually much larger. The method consists on estimating first a least squares approximation for the calibration parameters, and then, a refinement to these parameters is performed based on the fact that the rotation matrix must be orthonormal.

### B.2.1 Perspective projection matrix

We will take as a starting point Equation B.4, and rewrite it as

$$\tilde{\mathbf{m}} = \mathbf{A}\mathbf{H}\tilde{\mathbf{x}}_w = \tilde{\mathbf{P}}\tilde{\mathbf{x}}_w \quad (\text{B.7})$$

where  $\mathbf{A}$  are the intrinsic camera parameters, and  $\mathbf{H}$  is the homography for the position and orientation of the camera-centered coordinate frame with respect to the pattern reference frame. Decomposing the row elements of  $\mathbf{R}$  in

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_1^\top \\ \mathbf{r}_2^\top \\ \mathbf{r}_3^\top \end{bmatrix} \quad (\text{B.8})$$

we rewrite the perspective projection matrix  $\mathbf{P}$  as

$$\tilde{\mathbf{P}} = [\mathbf{P} \quad \tilde{\mathbf{p}}] = \begin{bmatrix} \mathbf{p}_1^\top & p_{14} \\ \mathbf{p}_2^\top & p_{24} \\ \mathbf{p}_3^\top & p_{34} \end{bmatrix} = \begin{bmatrix} \alpha_u \mathbf{r}_1^\top + u_0 \mathbf{r}_3^\top & \alpha_u t_1 + u_0 t_3 \\ \alpha_v \mathbf{r}_2^\top + v_0 \mathbf{r}_3^\top & \alpha_v t_2 + v_0 t_3 \\ \mathbf{r}_3^\top & t_3 \end{bmatrix} \quad (\text{B.9})$$

## B.2.2 Necessary conditions on $\tilde{\mathbf{P}}$

**Theorem B.2.1.** *The following two conditions hold for any given perspective projection matrix  $\tilde{\mathbf{P}}$*

$$\|\mathbf{p}_3\|^2 = 1 \quad (\text{B.10})$$

$$(\mathbf{p}_1 \times \mathbf{p}_3)^\top (\mathbf{p}_2 \times \mathbf{p}_3) = 0 \quad (\text{B.11})$$

*Proof.*  $\|\mathbf{p}_3\|^2 = \mathbf{r}_3^\top \mathbf{r}_3 = 1$ , and  $(\mathbf{p}_1 \times \mathbf{p}_3)^\top (\mathbf{p}_2 \times \mathbf{p}_3) = ((\alpha_u \mathbf{r}_1 + u_0 \mathbf{r}_3) \times \mathbf{r}_3)^\top ((\alpha_v \mathbf{r}_2 + v_0 \mathbf{r}_3) \times \mathbf{r}_3) = -\alpha_u \alpha_v \mathbf{r}_1^\top \mathbf{r}_2 = 0$ .  $\square$

Next, we derive closed form formulae for all the intrinsic and extrinsic camera parameters given a numerically computed  $\tilde{\mathbf{P}}$ , showing that a physically plausible solution exists for the parameters if the two conditions above are satisfied.

$\mathbf{r}_3, t_3$ :

$$\mathbf{r}_3 = \varepsilon \mathbf{p}_3 \quad (\text{B.12})$$

$$t_3 = \varepsilon p_{34} \quad (\text{B.13})$$

$$\varepsilon = \pm 1 \quad (\text{B.14})$$

If the origin of the pattern reference frame is in the positive side of the camera  $z$  axis then  $\varepsilon = 1$ , otherwise  $\varepsilon = -1$ .

$u_0, v_0$ :

$$\begin{aligned} \alpha_u \mathbf{r}_1^\top + u_0 \mathbf{r}_3^\top &= \varepsilon \mathbf{p}_1^\top \\ \alpha_u \mathbf{r}_1^\top \mathbf{r}_3 + u_0 \mathbf{r}_3^\top \mathbf{r}_3 &= \varepsilon \mathbf{p}_1^\top \mathbf{r}_3 \\ u_0 &= \mathbf{p}_1^\top \mathbf{p}_3 \end{aligned} \quad (\text{B.15})$$

$$\begin{aligned} \alpha_v \mathbf{r}_2^\top + v_0 \mathbf{r}_3^\top &= \varepsilon \mathbf{p}_2^\top \\ v_0 &= \mathbf{p}_2^\top \mathbf{p}_3 \end{aligned} \quad (\text{B.16})$$

$\alpha_u, \alpha_v$ :

$$\begin{aligned} \alpha_u \mathbf{r}_1^\top + u_0 \mathbf{r}_3^\top &= \varepsilon \mathbf{p}_1^\top \\ (\alpha_u \mathbf{r}_1 + u_0 \mathbf{r}_3)^\top (\alpha_u \mathbf{r}_1 + u_0 \mathbf{r}_3) &= \mathbf{p}_1^\top \mathbf{p}_1 \\ \alpha_u^2 + u_0^2 &= \mathbf{p}_1^\top \mathbf{p}_1 \\ \alpha_u &= \varepsilon \sqrt{\mathbf{p}_1^\top \mathbf{p}_1 - u_0^2} \end{aligned} \quad (\text{B.17})$$

$$\alpha_v = \varepsilon \sqrt{\mathbf{p}_2^\top \mathbf{p}_2 - v_0^2} \quad (\text{B.18})$$

$\mathbf{r}_1, \mathbf{r}_2$ :

$$\mathbf{r}_1 = (\varepsilon \mathbf{p}_1 - u_0 \mathbf{r}_3) / \alpha_u \quad (\text{B.19})$$

$$\mathbf{r}_2 = (\varepsilon \mathbf{p}_2 - v_0 \mathbf{r}_3) / \alpha_v \quad (\text{B.20})$$

$t_1, t_2$ :

$$t_1 = \frac{\varepsilon p_{14} - u_0 t_3}{\alpha_u} \quad (\text{B.21})$$

$$t_2 = \frac{\varepsilon p_{24} - v_0 t_3}{\alpha_v} \quad (\text{B.22})$$

### B.2.3 Least squares solution

Given a set  $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1 \tilde{\mathbf{x}}_2 \cdots \tilde{\mathbf{x}}_n]$  of points in the pattern-centered coordinate frame, and a set  $\tilde{\mathbf{M}} = [\tilde{\mathbf{m}}_1 \tilde{\mathbf{m}}_2 \cdots \tilde{\mathbf{m}}_n]$  of image correspondences, we can construct the augmented system

$$\tilde{\mathbf{M}} = \tilde{\mathbf{P}} \tilde{\mathbf{X}} \quad (\text{B.23})$$

Substituting the third row of Equation B.23 into its other two rows, we rewrite the system as

$$\begin{bmatrix} -\mathbf{x}_1^\top & \mathbf{0}^\top & u_1 \mathbf{x}_1^\top & -1 & 0 \\ \mathbf{0}^\top & -\mathbf{x}_1^\top & v_1 \mathbf{x}_1^\top & 0 & -1 \\ -\mathbf{x}_2^\top & \mathbf{0}^\top & u_2 \mathbf{x}_2^\top & -1 & 0 \\ \mathbf{0}^\top & -\mathbf{x}_2^\top & v_2 \mathbf{x}_2^\top & 0 & -1 \\ \vdots & & & & \\ -\mathbf{x}_n^\top & \mathbf{0}^\top & u_n \mathbf{x}_n^\top & -1 & 0 \\ \mathbf{0}^\top & -\mathbf{x}_n^\top & v_n \mathbf{x}_n^\top & 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \\ p_{14} \\ p_{24} \end{bmatrix} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_n \end{bmatrix} p_{34}$$

$$\mathbf{A} \mathcal{P} = \mathbf{b} p_{34} \quad (\text{B.24})$$

It can be shown that for a set of at least 6 non-coplanar point matches,  $\mathbf{A}$  has full rank. Nevertheless, the solution for  $\tilde{\mathbf{P}}$  is not unique since, as mentioned before,  $\tilde{\mathbf{P}} \sim \lambda \tilde{\mathbf{P}}$ . Perspective projections can only be computed up to a scale factor. This is,  $\mathbf{A}(\lambda \mathcal{P}) = \mathbf{b}(\lambda p_{34})$ . To overcome this ambiguity we fix  $p_{34} = 1$ , and an initial estimate for  $\mathcal{P}$  is found solving the above overconstrained system

$$\mathcal{P}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \quad (\text{B.25})$$



When computing the least squares estimate no particular attention was paid so that the elements on  $\mathcal{P}^*$  would satisfy the conditions in Equations B.10-B.11. We will now refine our estimate accordingly.

### B.2.4 Scale

First, the elements in  $\mathcal{P}^*$  are scaled so that  $\|\mathbf{p}_3^\top \mathbf{p}_3\| = 1$ . Also, as mentioned earlier, if the origin of the pattern reference frame is in the positive side of the camera  $z$  axis then  $\varepsilon = 1$ , otherwise  $\varepsilon = -1$ .

$$\tilde{\mathbf{P}} = \varepsilon \frac{\begin{bmatrix} \mathcal{P}_1^* & \mathcal{P}_2^* & \mathcal{P}_3^* & \mathcal{P}_{10}^* \\ \mathcal{P}_4^* & \mathcal{P}_5^* & \mathcal{P}_6^* & \mathcal{P}_{11}^* \\ \mathcal{P}_7^* & \mathcal{P}_8^* & \mathcal{P}_9^* & 1 \end{bmatrix}}{\sqrt{\mathcal{P}_7^{*2} + \mathcal{P}_8^{*2} + \mathcal{P}_9^{*2}}} \quad (\text{B.26})$$

and from Equations B.12-B.22, all camera parameters, intrinsic and extrinsic, can be computed. Note that if Equation B.26 has been scaled with  $\varepsilon$  there is no need to use it in Equations B.12-B.22. Our initial estimate for the intrinsic parameters are

$$u_0 = \mathbf{p}_1^\top \mathbf{p}_3 \quad (\text{B.27})$$

$$v_0 = \mathbf{p}_2^\top \mathbf{p}_3 \quad (\text{B.28})$$

$$\alpha_u = \sqrt{\mathbf{p}_1^\top \mathbf{p}_1 - u_0^2} \quad (\text{B.29})$$

$$\alpha_v = \sqrt{\mathbf{p}_2^\top \mathbf{p}_2 - v_0^2} \quad (\text{B.30})$$

### B.2.5 Orthonormality of $\mathbf{R}$

Let

$$\mathbf{D} = \begin{bmatrix} (\mathbf{p}_1 - u_0 \mathbf{p}_3) / \alpha_u \\ (\mathbf{p}_2 - v_0 \mathbf{p}_3) / \alpha_v \\ \mathbf{p}_3 \end{bmatrix} \quad (\text{B.31})$$

be our initial estimate for the rotation part of the extrinsic parameters, computed from Equations B.26-B.30. We will refine  $\mathbf{D}$  so that it satisfies the orthonormality condition. The task at hand is to solve the following constrained minimization problem:

$$\begin{aligned} &\text{minimize} && \|\mathbf{D} - \mathbf{R}\|^2 \\ &\text{subject to} && \mathbf{R}^\top \mathbf{R} = \mathbf{I} \end{aligned}$$

where  $\|\cdot\|$  denotes the Euclidean matrix norm.

Using the quaternion notation our minimization problem is equivalent to solving (see Weng *et al.* [216] for a proof)

$$\begin{aligned} & \text{minimize} && \mathcal{Q}^\top \mathbf{B} \mathcal{Q} \\ & \text{subject to} && \mathcal{Q}^\top \mathcal{Q} = 1 \end{aligned}$$

where  $\mathcal{Q} = [q_0, \mathbf{q}^\top]^\top$  is a unit quaternion such that

$$q_0 = \cos(\theta/2) \tag{B.32}$$

$$\mathbf{q} = \sin(\theta/2) \mathbf{n} \tag{B.33}$$

and  $\mathbf{R}$  can be expressed as a rotation about a unit axis  $\mathbf{n} = [n_1, n_2, n_3]^\top$  by an angle  $\theta$ .

$$\mathbf{B} = \sum \mathbf{B}_i^\top \mathbf{B}_i \tag{B.34}$$

$$\mathbf{B}_i = \begin{bmatrix} 0 & (\mathbf{e}_i - \mathbf{d}_i)^\top \\ (\mathbf{d}_i - \mathbf{e}_i) & [\mathbf{d}_i + \mathbf{e}_i]_\times \end{bmatrix} \tag{B.35}$$

and the notation  $[\cdot]_\times$  is used to represent the vector cross product  $\mathbf{x} \times \mathbf{y}$  as a multiplication of the form  $[\mathbf{x}]_\times \mathbf{y}$ ;  $\mathbf{I} = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$ .

$$[\mathbf{x}]_\times = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix} \tag{B.36}$$

Forming the Lagrangian  $l(\mathcal{Q}) = \mathcal{Q}^\top \mathbf{B} \mathcal{Q} + \lambda(1 - \mathcal{Q}^\top \mathcal{Q})$ , and computing  $\partial l / \partial \mathcal{Q} = \mathbf{B} \mathcal{Q} - \lambda \mathcal{Q}$ , we can see that the solution to our constrained minimization problem is the eigenvector associated with the smallest eigenvalue of  $\mathbf{B}$ .

Finally, the solution for the rotation matrix  $\mathbf{R}$  in terms of the computed quaternion  $\mathcal{Q}$  is given by

$$\mathbf{R} = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \tag{B.37}$$

Once the rotation part of the homography has been computed, we need to recalculate

the rest of the parameters:

$$u_0 = \mathbf{p}_1^\top \mathbf{r}_3 \quad (\text{B.38})$$

$$v_0 = \mathbf{p}_2^\top \mathbf{r}_3 \quad (\text{B.39})$$

$$\alpha_u = \|\mathbf{p}_1^\top \mathbf{p}_1 - u_0^2\| \quad (\text{B.40})$$

$$\alpha_v = \|\mathbf{p}_2^\top \mathbf{p}_2 - v_0^2\| \quad (\text{B.41})$$

$$t = \begin{bmatrix} (p_{14} - u_0 p_{34})/\alpha_u \\ (p_{24} - v_0 p_{34})/\alpha_v \\ p_{34} \end{bmatrix} \quad (\text{B.42})$$

### B.3 Computing distortion parameters

The reconstruction accuracy obtained using the pinhole model is sufficiently robust for most applications. However, if it is still necessary to refine the calibration parameters, one possible solution is to compute estimates for the nonlinear distortion parameters.

The sum of the squared error between the actual image points  $\hat{\mathbf{m}} = [\hat{u}, \hat{v}, 1]^\top$  and their projected estimates  $\mathbf{m} = [u, v, 1]^\top$  can be used as a minimization function for the optimization of the distortion parameters. From Equation B.6

$$\text{minimize} \quad \sum^n \left\| \hat{\mathbf{m}}_i - \mathbf{m}_i - c_i \begin{bmatrix} k_1 \\ k_2 \\ 0 \end{bmatrix} \right\|^2 \quad (\text{B.43})$$

with

$$c_i = \mathbf{m}_i^\top \begin{bmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ -u_0 & -v_0 & u_0^2 + v_0^2 \end{bmatrix} \mathbf{m}_i \quad (\text{B.44})$$

and its minimum at

$$\begin{bmatrix} k_1 \\ k_2 \\ 0 \end{bmatrix}^* = \frac{\sum^n (c_i (\hat{\mathbf{m}}_i - \mathbf{m}_i))}{\sum^n c_i^2} \quad (\text{B.45})$$

### B.4 Bibliographical notes

For an introductory text on geometry the reader is referred to the text by Xambó [218]. A complete overview on projective geometry and its relation to computer vision is presented in a tutorial due to Mohr and Triggs [141]. Comprehensive texts on geometrical

computer vision with different approaches to calibration include those by Faugeras [74], Hartley and Zisserman [91], Jain *et al.* [103], Kanatani [114], or Xu and Zhang [219]. Seminal papers on pattern-based calibration are the ones by Tsai [200] and Weng *et al.* [215, 216].

# Appendix C

## The Kalman Filter

The Kalman Filter developed in the early sixties by R.E. Kalman [111, 112] is a recursive state estimator for partially observed non-stationary stochastic processes. It gives an optimal estimate in the least squares sense of the actual value of a state vector from noisy observations.

### C.1 Recursive state estimation

Consider a discrete-time stochastic process

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k) \quad (\text{C.1})$$

with system input  $\mathbf{u}$  and unmodeled process dynamics plus noise  $\mathbf{v}$ . The task at hand is to find an estimate of the state vector  $\mathbf{x}$ . However,  $\mathbf{x}$  is only accessible from noise distorted sensor measurements

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{w}_k) \quad (\text{C.2})$$

in which as with the process model,  $\mathbf{w}$  represents observation model inaccuracies and sensor noise.

Recursive state estimation consists on iteratively reconstructing the state vector from our knowledge of the process dynamics, the measurement model, and the sensed data.

Let  $\mathbf{x}_{i|j}$  be the estimate of the state  $\mathbf{x}_i$  using the observation information up to and including time  $j$ ,  $\mathbf{Z}^j = \{\mathbf{z}_1, \dots, \mathbf{z}_j\}$ . Given an estimate  $\mathbf{x}_{k|k}$ , and the input to the system

$\mathbf{u}_k$ , the predicted state  $\mathbf{x}_{k+1|k}$  is ideally given by the expectation

$$\mathbf{x}_{k+1|k} = E[\mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k) | \mathbf{Z}^k] \quad (\text{C.3})$$

We call  $\mathbf{x}_{k+1|k}$  the *a priori* estimate of  $\mathbf{x}_{k+1}$ , and compute it from a noise-free version of Equation C.1, the estimate  $\mathbf{x}_{k|k}$ , and the input that hypothetically would drive the process from  $\mathbf{x}_k$  to  $\mathbf{x}_{k+1}$

$$\mathbf{x}_{k+1|k} = \mathbf{f}(\mathbf{x}_{k|k}, \mathbf{u}_k, \mathbf{0}) \quad (\text{C.4})$$

Combining this result with the discrete-time measurement model from Equation C.2, we can also predict a noise-free a priori estimate of the sensor measurements

$$\mathbf{z}_{k+1|k} = \mathbf{h}(\mathbf{x}_{k+1|k}, \mathbf{0}) \quad (\text{C.5})$$

By comparing the actual measurement vector  $\mathbf{z}_{k+1}$  with the predicted data  $\mathbf{z}_{k+1|k}$ , we obtain an observation prediction error which in turn is added in a correction term to the a priori state estimate to produce an *a posteriori* state estimate.

$$\mathbf{x}_{k+1|k+1} = \mathbf{x}_{k+1|k} + \mathbf{K}_{k+1}(\mathbf{z}_{k+1} - \mathbf{z}_{k+1|k}) \quad (\text{C.6})$$

The choice of the gain matrix  $\mathbf{K}$  usually meets some optimality criteria. In the case of the Kalman Filter, the stochastic nature of the process and measurement dynamics is taken into account in the derivation of  $\mathbf{K}$ , producing an optimal linear estimator that minimizes the squared error on the expected value of the state estimate  $\mathbf{x}_{k+1|k+1}$ .

## C.2 Linear Kalman filter

Consider the case when the process and measurement models correspond to a possibly non-stationary<sup>1</sup> discrete-time linear system, and that both the process and sensor noises are zero-mean white<sup>2</sup> and Gaussian with covariance matrices  $\mathbf{V}_k$  and  $\mathbf{W}_k$  respectively, then Equations C.1 and C.1 become

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{u}_k + \mathbf{v}_k \quad (\text{C.7})$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k \quad (\text{C.8})$$

<sup>1</sup>Hence its beauty, compared to its predecessor the Weiner filter that only works for stationary linear systems.

<sup>2</sup>Temporally uncorrelated and with equal power at all frequencies.

$$E[\mathbf{v}_k] = \mathbf{0}, \quad E[\mathbf{v}_k \mathbf{v}_k^\top] = \mathbf{V}_k, \quad \text{and} \quad E[\mathbf{v}_i \mathbf{v}_j^\top] = \mathbf{0}, \quad \forall i \neq j \quad (\text{C.9})$$

$$E[\mathbf{w}_k] = \mathbf{0}, \quad E[\mathbf{w}_k \mathbf{w}_k^\top] = \mathbf{W}_k, \quad \text{and} \quad E[\mathbf{w}_i \mathbf{w}_j^\top] = \mathbf{0}, \quad \forall i \neq j \quad (\text{C.10})$$

The a priori and a posteriori state estimation errors can be written as

$$\mathbf{e}_{k+1|k} = \mathbf{x}_{k+1} - \mathbf{x}_{k+1|k} \quad (\text{C.11})$$

$$\mathbf{e}_{k+1|k+1} = \mathbf{x}_{k+1} - \mathbf{x}_{k+1|k+1} \quad (\text{C.12})$$

and from the linear model in Equation C.7, the noise-free a priori state estimate in Equation C.4 takes the form

$$\mathbf{x}_{k+1|k} = \mathbf{F}_k \mathbf{x}_{k|k} + \mathbf{u}_k \quad (\text{C.13})$$

It follows that the a priori state estimate error is given by

$$\mathbf{e}_{k+1|k} = \mathbf{F}_k \mathbf{e}_{k|k} + \mathbf{v}_k \quad (\text{C.14})$$

Substituting Eq. C.6 and the observation models  $\mathbf{z}_{k+1|k} = \mathbf{H}_{k+1} \mathbf{x}_{k+1|k}$  and  $\mathbf{z}_{k+1} = \mathbf{H}_{k+1} \mathbf{x}_{k+1} + \mathbf{w}_{k+1}$  in Equation C.12, we obtain a recursive expression for the a posteriori state estimation error.

$$\mathbf{e}_{k+1|k+1} = \mathbf{e}_{k+1|k} - \mathbf{K}_{k+1} (\mathbf{H}_{k+1} \mathbf{e}_{k+1|k} + \mathbf{w}_{k+1}) \quad (\text{C.15})$$

The state error covariances are given by the expectations of the square of the state errors.

$$\mathbf{P}_{k+1|k} = E[\mathbf{e}_{k+1|k} \mathbf{e}_{k+1|k}^\top] \quad (\text{C.16})$$

$$\mathbf{P}_{k+1|k+1} = E[\mathbf{e}_{k+1|k+1} \mathbf{e}_{k+1|k+1}^\top] \quad (\text{C.17})$$

Plugging Equation C.14 in Equation C.16 and taking the expectations on  $\mathbf{v}$ , we get the following expression for the a priori state error covariance

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k \mathbf{P}_{k|k} \mathbf{F}_k^\top + \mathbf{V}_k \quad (\text{C.18})$$

For simplicity of notation, in the sequel we rewrite the dependencies  $(k+1|k)$  and  $(k+1|k+1)$  as  $^\ominus$  and  $^\oplus$  respectively, and when no step reference is provided,  $(k+1)$  is assumed. Substituting Equation C.12 in Equation C.17 and taking the expectations on  $\mathbf{w}$  and  $\mathbf{e}^\ominus$ , the a posteriori error covariance takes the form

$$\mathbf{P}^\oplus = \mathbf{P}^\ominus - \mathbf{P}^\ominus \mathbf{H}^\top \mathbf{K}^\top - \mathbf{K} \mathbf{H} \mathbf{P}^\ominus + \mathbf{K} (\mathbf{H} \mathbf{P}^\ominus \mathbf{H}^\top + \mathbf{W}) \mathbf{K}^\top \quad (\text{C.19})$$

The gain matrix  $\mathbf{K}$  is chosen to minimize the a posteriori error covariance. Making the derivative of the trace of  $\mathbf{P}^\oplus$  with respect to  $\mathbf{K}$  equal to  $\mathbf{0}$ , and solving for  $\mathbf{K}$  we get the optimal gain for the computation of Equation C.6, i.e., the Kalman gain.

$$\mathbf{K} = \mathbf{P}^\ominus \mathbf{H}^\top (\mathbf{H} \mathbf{P}^\ominus \mathbf{H}^\top + \mathbf{W})^{-1} \quad (\text{C.20})$$

Plugging Equation C.20 back in Equation C.19 reduces  $\mathbf{P}^\oplus$  to the well known form

$$\mathbf{P}^\oplus = \mathbf{P}^\ominus - \mathbf{K} \mathbf{H} \mathbf{P}^\ominus \quad (\text{C.21})$$

By inspecting the Kalman filter equations the behavior of the filter agrees with our intuition. The Kalman gain is proportional to the uncertainty in the state estimate and inversely proportional to that in the measurements. If sensor readings are very uncertain, and the state estimate is relatively precise, then the Kalman gain has little impact on the update of the state estimate in Equation C.6, and the system relies heavily on the system model. If on the other hand, the uncertainty in the measurement is small and that in the state estimate is large, then  $\mathbf{K}$  is also large, thus trusting more in sensor measurements for the correction of the state estimate.

However, when sensor measurements are uncertain the second term in Equation C.21 is small and the state estimate error covariance sees little reduction. Conversely, accurate sensor measurements contribute considerably in reducing the state estimation error.

Given the initial conditions  $\mathbf{x}_{0|0}$  and  $\mathbf{P}_{0|0}$ , the complete recursion in the Kalman filter is computed iteratively with the following steps:

- Predict the a priori state, error covariance, and observation estimates

$$\begin{aligned} \mathbf{x}^\ominus &= \mathbf{F}_k \mathbf{x}_{k|k} + \mathbf{u}_k \\ \mathbf{P}^\ominus &= \mathbf{F}_k \mathbf{P}_{k|k} \mathbf{F}_k^\top + \mathbf{V}_k \\ \mathbf{z}^\ominus &= \mathbf{H} \mathbf{x}^\ominus \end{aligned}$$

- Compute the Kalman gain and correct the state and state error covariance estimates

$$\begin{aligned} \mathbf{K} &= \mathbf{P}^\ominus \mathbf{H}^\top (\mathbf{H} \mathbf{P}^\ominus \mathbf{H}^\top + \mathbf{W})^{-1} \\ \mathbf{x}^\oplus &= \mathbf{x}^\ominus + \mathbf{K} (\mathbf{z} - \mathbf{z}^\ominus) \\ \mathbf{P}^\oplus &= \mathbf{P}^\ominus - \mathbf{K} \mathbf{H} \mathbf{P}^\ominus \end{aligned}$$



### C.3 Extended Kalman filter

Consider now the case when the process and observation models in Equations C.1 and C.2 are non-linear. The Extended Kalman Filter (EKF) provides a solution by linearizing the process about the current state, and linearizing the measurement model about the predicted observation.

The linearization of  $\mathbf{f}$  about the current estimate  $\mathbf{x}_{k|k}$  can be formulated as a Taylor series with the higher order terms dropped

$$\mathbf{x} \approx \mathbf{x}^\ominus + \nabla \mathbf{f}_{\mathbf{x}}(\mathbf{x}_k - \mathbf{x}_{k|k}) + \nabla \mathbf{f}_{\mathbf{v}} \mathbf{v}_k$$

Similarly, the linearization of the observation model takes the form

$$\mathbf{z} \approx \mathbf{z}^\ominus + \nabla \mathbf{h}_{\mathbf{x}}(\mathbf{x} - \mathbf{x}^\ominus) + \nabla \mathbf{h}_{\mathbf{w}} \mathbf{w}$$

The noise-free estimates  $\mathbf{x}^\ominus$  and  $\mathbf{z}^\ominus$  are given in Equations C.4 and C.5, and the various Jacobian matrices contain the partial derivatives of  $\mathbf{f}$  and  $\mathbf{h}$  with respect to  $\mathbf{x}$  and the noises  $\mathbf{v}$  and  $\mathbf{w}$

$$\begin{aligned} \nabla \mathbf{f}_{\mathbf{x}} &= \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{(\mathbf{x}_{k|k}, \mathbf{u}_k, \mathbf{0})} \\ \nabla \mathbf{f}_{\mathbf{v}} &= \left. \frac{\partial \mathbf{f}}{\partial \mathbf{v}} \right|_{(\mathbf{x}_{k|k}, \mathbf{u}_k, \mathbf{0})} \\ \nabla \mathbf{h}_{\mathbf{x}} &= \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{(\mathbf{x}^\ominus, \mathbf{0})} \\ \nabla \mathbf{h}_{\mathbf{w}} &= \left. \frac{\partial \mathbf{h}}{\partial \mathbf{w}} \right|_{(\mathbf{x}^\ominus, \mathbf{0})} \end{aligned}$$

Following the same discussion as in the previous section but with this new linear model, it is easy to show how the complete recursion for the Extended Kalman Filter involves the following steps:

- Predict the a priori state and observation estimates as well as the a priori state error covariance estimate

$$\begin{aligned} \mathbf{x}^\ominus &= \mathbf{f}(\mathbf{x}_{k|k}, \mathbf{u}_k, \mathbf{0}) \\ \mathbf{P}^\ominus &= \nabla \mathbf{f}_{\mathbf{x}} \mathbf{P}_{k|k} \nabla \mathbf{f}_{\mathbf{x}}^\top + \nabla \mathbf{f}_{\mathbf{v}} \mathbf{V}_k \nabla \mathbf{f}_{\mathbf{v}}^\top \\ \mathbf{z}^\ominus &= \mathbf{h}(\mathbf{x}^\ominus, \mathbf{0}) \end{aligned}$$

- Compute the Kalman gain and correct the state and state error covariance estimates

$$\begin{aligned}\mathbf{K} &= \mathbf{P}^\ominus \nabla \mathbf{h}_x^\top (\nabla \mathbf{h}_x \mathbf{P}^\ominus \nabla \mathbf{h}_x^\top + \nabla \mathbf{h}_w \mathbf{W} \nabla \mathbf{h}_w^\top)^{-1} \\ \mathbf{x}^\oplus &= \mathbf{x}^\ominus + \mathbf{K}(\mathbf{z} - \mathbf{z}^\ominus) \\ \mathbf{P}^\oplus &= \mathbf{P}^\ominus - \mathbf{K} \nabla \mathbf{h}_x \mathbf{P}^\ominus\end{aligned}$$

It is important to note however, that the linearization of the nonlinear process and measurement models in the EKF does not preserve the distributions of the state and measurement random variables as normal. This may lead to difficulties in the implementation and tuning of the EKF, making it only reliable for systems that are almost linear on the time scale interval  $(k, k + 1)$ .

## C.4 Conditioning

It turns out that the recursion in Equation C.21 is ill-conditioned. As the filter converges, the canceling of significant digits on  $\mathbf{P}^\oplus$  may lead to asymmetries or to a non *positive semi definite* (psd) matrix, which cannot be true from the definition in Equation C.17 of the a posteriori error covariance matrix.

An algebraic manipulation that guarantees  $\mathbf{P}^\oplus$  psd is obtained by multiplying Equation C.20 by  $(\mathbf{H}\mathbf{P}^\ominus\mathbf{H}^\top + \mathbf{W})\mathbf{K}^\top$ , rearranging terms

$$\mathbf{K}\mathbf{H}\mathbf{P}^\ominus\mathbf{H}^\top\mathbf{K}^\top - \mathbf{P}^\ominus\mathbf{H}^\top\mathbf{K}^\top + \mathbf{K}\mathbf{W}\mathbf{K}^\top = \mathbf{0} \quad (\text{C.22})$$

and adding Equation C.22 into Equation C.21

$$\mathbf{P}^\oplus = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^\ominus(\mathbf{I} - \mathbf{K}\mathbf{H})^\top + \mathbf{K}\mathbf{W}\mathbf{K}^\top \quad (\text{C.23})$$

The recursivity in Equation C.23 is known as the Joseph form of the a posteriori error covariance matrix, and given its quadratic nature it is obviously psd.

## C.5 Sequential innovation

When combining information from multiple sensors or from multiple data sources, the observation vector can be seen as a collection of  $n$  independent measurements coming

from the same number of independent sources at any particular time instance ( $k + 1$ )

$$\mathbf{z} = \left[ \dots, \mathbf{z}^{(i)\top}, \dots \right]^\top$$

Each of these measurements provides information for the reconstruction of  $\mathbf{x}^\oplus$  in the direction spanned by the  $i$ -th set of columns<sup>3</sup> of  $\mathbf{K}$ , i.e.,  $\mathbf{K}_{(i)} (\mathbf{z}^{(i)} - \mathbf{z}^{\ominus(i)})$ . By the same token, in the linear case the independent measurement models for each source are spanned by the  $i$ -th set of rows of  $\mathbf{H}$

$$\mathbf{z}^{(i)} = \mathbf{H}^{(i)} \mathbf{x} + \mathbf{w}^{(i)}$$

and by the  $i$ -th sets of rows of  $\nabla \mathbf{h}_x$  and  $\nabla \mathbf{h}_w$  for the nonlinear case

$$\mathbf{z}^{(i)} \approx \mathbf{z}^{\ominus(i)} + \nabla \mathbf{h}_x^{(i)} \mathbf{e}^\ominus + \nabla \mathbf{h}_w^{(i)} \mathbf{w}^{(i)}$$

It is possible to process each of these observations independently provided  $\mathbf{W}$  is block diagonal. That is, when the set of measurements taken at the same time interval are uncorrelated. Even when the measurements are correlated, they may always be transformed into uncorrelated data which then may be treated sequentially. The process is called *whitening* (see Fukunaga [78]).

The a posteriori state estimate is given by the combination of its a priori estimate and a linear combination of the independent measurement errors

$$\mathbf{x}^\oplus = \mathbf{x}^\ominus + \sum_i \mathbf{K}_{(i)} (\mathbf{z}^{(i)} - \mathbf{z}^{\ominus(i)})$$

The key advantage of the *sequential innovation* method is that the complexity in the computation of the Kalman gain is reduced considerably. From Equation C.20

$$\mathbf{K}_{(i)} = \mathbf{P}^\ominus \mathbf{H}^{(i)\top} \left( \mathbf{H}^{(i)} \mathbf{P}^\ominus \mathbf{H}^{(i)\top} + \mathbf{W}^{(i)} \right)^{-1} \quad (\text{C.24})$$

The required inverse in Equation C.24 has the dimension of each of the observed variables, and is considerably much smaller than the dimension of the entire measurement vector  $\mathbf{z}$  as required in Equation C.20. When a sensor returns scalar values for each independent measurement, then the inverse in Equation C.24 is replaced by a scalar division.

---

<sup>3</sup>The notations  $(\cdot)^{(i)}$  and  $(\cdot)_{(i)}$  indicate the  $i$ -th set of rows or columns, respectively.

Given the initial conditions  $\mathbf{x}_{0|0}$  and  $\mathbf{P}_{0|0}$ , the complete sequential innovation Kalman filter recursion is computed with the following steps:

- Predict the a priori state, error covariance, and the set of observation estimates

$$\begin{aligned}\mathbf{x}^\ominus &= \mathbf{F}_k \mathbf{x}_{k|k} + \mathbf{u}_k \\ \mathbf{P}^\ominus &= \mathbf{F}_k \mathbf{P}_{k|k} \mathbf{F}_k^\top + \mathbf{V}_k \\ \forall i \quad \mathbf{z}^{\ominus(i)} &= \mathbf{H}^{(i)} \mathbf{x}^\ominus\end{aligned}$$

- For each measurement, iteratively compute the corresponding Kalman gain column and correct the state and state error covariance estimates

$$\begin{aligned}\text{initialize} \quad \mathbf{x}^{\oplus(0)} &= \mathbf{x}^\ominus \\ \mathbf{P}^{\oplus(0)} &= \mathbf{P}^\ominus \\ \forall i \quad \mathbf{K}^{(i)} &= \mathbf{P}^{\oplus(i-1)} \mathbf{H}^{(i)\top} \left( \mathbf{H}^{(i)} \mathbf{P}^{\oplus(i-1)} \mathbf{H}^{(i)\top} + \mathbf{W}^{(i)} \right)^{-1} \\ \mathbf{x}^{\oplus(i)} &= \mathbf{x}^{\oplus(i-1)} + \mathbf{K}^{(i)} \left( \mathbf{z}^{(i)} - \mathbf{z}^{\ominus(i)} \right) \\ \mathbf{P}^{\oplus(i)} &= \mathbf{P}^{\oplus(i-1)} - \mathbf{K}^{(i)} \mathbf{H}^{(i)} \mathbf{P}^{\oplus(i-1)} \\ \text{restore} \quad \mathbf{x}^\oplus &= \mathbf{x}^{\oplus(n)} \\ \mathbf{P}^\oplus &= \mathbf{P}^{\oplus(n)}\end{aligned}$$

## C.6 Bibliographical notes

The reader can find thorough discussions on the Kalman Filter in various articles, books and lecture notes by Brookner [33], Dickmanns [62], Kalman *et al.* [111, 112], Maybeck [138], Newman [151], Papoulis [157], or Welch and Bishop [212], and on its predecessor the Wiener Filter in the books by Carnicer *et al.* [41], or Gonzalez and Woods [85]. One approach to reduce the effect of nonlinearities is to apply iteratively the filter (IEKF) as indicated by Zhang [220]. Another solution is to use the Unscented Kalman Filter (UKF), an extension to the EKF that takes into account the nonlinear transformation of means and covariances [105, 107, 203]. Numerical instability may occur even with the Joseph form of the error covariance matrix. An alternative is the use of the square-root Kalman filter (SKF), in which recursive computations for  $\mathbf{P}^\oplus$  are substituted by equations for a recursion in  $\mathbf{P}^{\oplus 1/2}$  [62]. Sequential innovation in Kalman filtering is discussed in detail by Dickmanns [62], and Welch [211, 213].

# Appendix D

## Concepts from Linear Algebra

### D.1 Properties of positive semi-definite matrices

Each of the following are necessary and sufficient conditions for a real valued symmetric matrix  $\mathbf{A}$  to be positive semi-definite *psd*.

- $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ .
- all the eigenvalues of  $\mathbf{A}$  satisfy  $\lambda_i \geq 0$ .

Moreover, if  $\mathbf{A}$  is *psd*

- $\det(\mathbf{A}) \geq 0$ .
- any principal submatrix of a *psd* matrix is also *psd*.
- $\mathbf{B} \mathbf{A} \mathbf{B}^T$  is also *psd*, for any real valued  $\mathbf{B}$ .
- and if  $\mathbf{B}$  is also *psd*, then  $(\mathbf{A} + \mathbf{B})$  is also *psd*.
- and if  $\mathbf{B}$  is also *psd*, then  $\det \mathbf{A} \leq \det (\mathbf{A} + \mathbf{B})$ .

### D.2 Linear subspaces

The four fundamental subspaces spanned by the columns and rows of a real valued matrix  $\mathbf{A}$  are

- The column space of  $\mathbf{A}$ .  $\text{Im}\mathbf{A} = \{\mathbf{y} \in \mathbb{R}^m : \exists \mathbf{x} \in \mathbb{R}^n, \mathbf{A}\mathbf{x} = \mathbf{y}\}$ .
- The row space of  $\mathbf{A}$ .  $\text{Im}\mathbf{A}^\top = \{\mathbf{y} \in \mathbb{R}^n : \exists \mathbf{x} \in \mathbb{R}^m, \mathbf{A}^\top \mathbf{x} = \mathbf{y}\}$ .
- The null space of  $\mathbf{A}$ .  $\text{Ker}\mathbf{A} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{0}\}$ .
- The left null space of  $\mathbf{A}$ .  $\text{Ker}\mathbf{A}^\top = \{\mathbf{x} \in \mathbb{R}^m : \mathbf{A}^\top \mathbf{x} = \mathbf{0}\}$ .

### D.3 Inverse of a block matrix

The inverse of the nonsingular block matrix

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

has the partitions

$$\mathbf{B}_{11} = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}$$

$$\mathbf{B}_{12} = -\mathbf{B}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$$

$$\mathbf{B}_{21} = -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11}$$

$$\mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$$

### D.4 The matrix inversion lemma

$$\begin{aligned} (\mathbf{A}^{-1} + \mathbf{B}^\top \mathbf{C}^{-1} \mathbf{B})^{-1} &= \mathbf{A} - \mathbf{A} \mathbf{B}^\top (\mathbf{B} \mathbf{A} \mathbf{B}^\top + \mathbf{C})^{-1} \mathbf{B} \mathbf{A} \\ (\mathbf{A} + \mathbf{B} \mathbf{C} \mathbf{B}^\top)^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{B}^\top \mathbf{A}^{-1} \mathbf{B} + \mathbf{C}^{-1})^{-1} \mathbf{B}^\top \mathbf{A}^{-1} \end{aligned}$$

### D.5 Matrix inequalities

The matrix inequality

$$\mathbf{A} \geq \mathbf{B}$$

is to be interpreted as follows:

$$\mathbf{C} = \mathbf{A} - \mathbf{B} \geq \mathbf{0}$$

that is, the difference  $\mathbf{C}$  of the two matrices is positive semi-definite.

## **D.6 Bibliographical notes**

For a comprehensive text on linear algebra the reader is referred to the book by Strang [184].





# Bibliography

- [1] G. M. ADELSON-VELSKIĬ AND E. M. LANDIS, An algorithm for the organization of information. *Soviet Mathematics Doklady*, vol. 3, pp. 1259–1263, 1962.
- [2] M. ADJOUADI AND F. CANDOCIA, A stereo matching paradigm based on the walsh transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 12, pp. 1212–1218, December 1994.
- [3] R. ALQUÉZAR, A. SANFELIU, AND F. SERRATOSA, Synthesis of function-described graphs. In A. Amin, D. Dori, P. Pudil, and H. Freeman, eds., *Proceedings of the International Workshop on Statistical Techniques in Pattern Recognition*, pp. 112–121, Sydney, August 1998.
- [4] D. ANDERSEN, Optimal control. Lecture Notes. 2002. Dept. Control Eng., Inst. Electronic Systems.
- [5] J. ANDRADE-CETTO, Camera calibration. Tech. Rep. IRI-DT-01-02, Institut de Robòtica i Informàtica Industrial, Universitat Politècnica de Catalunya, June 2001.
- [6] J. ANDRADE-CETTO, The Kalman filter. Tech. Rep. IRI-DT-02-01, Institut de Robòtica i Informàtica Industrial, Universitat Politècnica de Catalunya, March 2002.
- [7] J. ANDRADE-CETTO AND A. C. KAK, Object recognition. In J. G. Webster, ed., *Wiley Encyclopedia of Electrical and Electronics Engineering*, supplement 1, pp. 449–470, John Wiley & Sons, New York, 2000.
- [8] J. ANDRADE-CETTO AND A. SANFELIU, Integration of perceptual grouping and depth. In A. Sanfeliu, J. J. Villanueva, M. Vanrell, R. Alquézar, J. O. Eklundh, and Y. Aloimonos, eds., *Proceedings of the 15th IAPR International Conference on Pattern Recognition*, vol. 1, pp. 295–298, IEEE Computer Society, Barcelona, September 2000.

- 
- [9] J. ANDRADE-CETTO AND A. SANFELIU, Learning of dynamic environments by a mobile robot from stereo cues. In *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 305–310, Baden-Baden, August 2001.
- [10] J. ANDRADE-CETTO AND A. SANFELIU, Topological map learning for a mobile robot in indoor environments. In J. S. Sánchez and F. Pla, eds., *Proceedings of the 9th Spanish Symposium on Pattern Recognition and Image Analysis*, vol. 1, pp. 221–226, Universitat Jaume I, Benicasim, May 2001.
- [11] J. ANDRADE-CETTO AND A. SANFELIU, Concurrent map building and localization on indoor dynamic environments. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 16, no. 3, pp. 361–374, May 2002.
- [12] J. ANDRADE-CETTO AND A. SANFELIU, Concurrent map building and localization with landmark validation. In *Proceedings of the 16th IAPR International Conference on Pattern Recognition*, vol. 2, pp. 693–696, IEEE Computer Society, Quebec, August 2002.
- [13] J. ANDRADE-CETTO AND A. SANFELIU, Concurrent map building and localization with temporal landmark validation. Accepted for presentation in 2003 IEEE International Conference on Robotics and Automation.
- [14] G. C. ANOUSAKI AND K. J. KYRIAKOPOULOS, Simultaneous localization and map building for mobile robot navigation. *IEEE Robotics and Automation Magazine*, vol. 6, no. 3, pp. 42–53, September 1999.
- [15] J. ARANDA-LÓPEZ, *Aportació als Mètodes de Seguiment Tridimensional d'Objectes d'Alta Velocitat d'Operació Mitjançant l'Estereovisió*. Ph.D. thesis, Universitat Politècnica de Catalunya, Barcelona, October 1997.
- [16] R. ARAÚJO AND A. T. DE ALMEIDA, Learning sensor-based navigation of a mobile robot in unknown worlds. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 29, no. 2, pp. 164–178, April 1999.
- [17] A. ARLEO, J. DEL R. MILLÁN, AND D. FLOREANO, Efficient learning of variable-resolution cognitive maps for autonomous indoor navigation. *IEEE Transactions on Robotics and Automation*, vol. 15, no. 6, pp. 990–1000, December 1999.
- [18] M. ATALLAH, Algorithm design, analysis, and implementation. Lecture Notes. Spring 1999. Purdue University.
- [19] W. J. AUSTIN AND A. M. WALLACE, Object location by parallel pose clustering. *Computer Vision and Image Understanding*, vol. 72, no. 3, pp. 304–327, December 1998.

- 
- [20] N. AYACHE AND O. FAUGERAS, Maintaining representations of the environment of a mobile robot. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 6, pp. 804–819, December 1989.
- [21] O. AYCARD, J.-F. MARI, AND F. CHARPILLET, Second order hidden markov models for place recognition: New results. In *Proceedings of the IEEE International Conference on Tools with Artificial Intelligence*, pp. 408–415, Taipei, November 1998.
- [22] D. BAŞKENT AND B. BARSHAN, Surface profile determination from multiple sonar data using morphological processing. *International Journal of Robotics Research*, vol. 18, no. 8, pp. 788–808, August 1999.
- [23] D. H. BALLARD, Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, 1981.
- [24] D. H. BALLARD AND C. M. BROWN, *Computer Vision*. Prentice Hall, Englewood Cliffs, 1982.
- [25] Y. BAR-SHALOM AND T. H. FORTMANN, *Tracking and Data Association*, vol. 179 of *Mathematics in Science and Engineering*. Academic Press, Boston, 1988.
- [26] Y. BAR-SHALOM, X. R. LI, AND T. KIRUBARAJAN, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, New York, 2001.
- [27] M. BAYOUTH AND P. KOOPMAN, Functional evolution of an automated highway system for incremental deployment. *Transportation Research Record*, vol. 1651, pp. 80–88, June 1998.
- [28] P. R. BEAUDET, Rotational invariant image operators. In *Proceedings of the 4th IAPR International Conference on Pattern Recognition*, pp. 579–583, Tokyo, 1978.
- [29] J. L. BENTLEY AND T. OTTMANN, Algorithms for reporting and counting geometric intersections. *IEEE Transactions on Computers*, vol. 28, no. 9, pp. 643–647, September 1979.
- [30] R. BERTHILSSON AND A. HEYDEN, Recognition of planar objects using the density of affine shape. *Computer Vision and Image Understanding*, vol. 76, no. 2, pp. 135–145, November 1999.
- [31] A. BLAKE, A. ZISSERMAN, AND G. KNOWELS, Surface description from stereo and shading. In B. Horn and M. Brooks, eds., *Shape from Shading*, Artificial Intelligence Series, chap. 2, pp. 29–52, The MIT Press, Cambridge, August 1989.

- 
- [32] T. BLASZKA AND R. DERICHE, Recovering and characterizing image features using an efficient model based approach. Tech. Rep. TR 2422, Institut National de Recherche en Informatique et en Automatique, November 1994.
- [33] E. BROOKNER, *Tracking and Kalman Filtering made Easy*. John Wiley & Sons, New York, 1998.
- [34] H. H. BULTHOFF AND H. A. MALLOT, Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America A-Optics Image Science and Vision*, vol. 5, pp. 1749–1758, October 1988.
- [35] J. F. CANNY, *Finding Edges and Lines in Images*. Master's thesis, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, June 1983.
- [36] J. F. CANNY, A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, November 1986.
- [37] A. CARBONARO AND P. ZINGARETTI, Landmark matching in a varying environment. In *Proceedings of the Euromicro Workshop on Advanced Mobile Robots*, pp. 147–153, Brescia, 1997.
- [38] S. CARLSSON, The double algebra: An effective tool for computing invariants in computer vision. In *Proceedings of the Joint European-US Workshop on Applied Invariance in Computer Vision*, pp. 145–164, Azores, October 1993.
- [39] S. CARLSSON, Projectively invariant decomposition and recognition of planar shapes. *International Journal of Computer Vision*, vol. 17, no. 2, pp. 193–209, February 1996.
- [40] S. CARLSSON AND D. WEINSHALL, Dual computation of projective shape and camera positions from multiple images. *International Journal of Computer Vision*, vol. 27, no. 3, pp. 227–241, May 1998.
- [41] A. CARNICER, I. JUVELLS, S. VALMITJANA, AND S. BOSCH, *Processament Òptic i Digital d'Imatges*. Edicions de la Universitat de Barcelona, Barcelona, 2001.
- [42] V. CASELLES, R. KIMMEL, G. SAPIRO, AND C. SBERT, Minimal surfaces based object segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp. 394–398, April 1997.
- [43] J. A. CASTELLANOS, J. M. M. MONTIEL, J. NEIRA, AND J. D. TARDÓS, The SPMaP: A probabilistic framework for simultaneous localization and map building. *IEEE Transactions on Robotics and Automation*, vol. 15, no. 5, pp. 948–952, October 1999.

- [44] A. CHECA, J. ANDRADE, AND A. SANFELIU, Construcció de mapes per robots mòbils equipats amb sensors làser de profunditat. Tech. Rep. IRI-DT-03-01, Institut de Robòtica i Informàtica Industrial, Universitat Politècnica de Catalunya, March 2003.
- [45] C. H. CHEN AND A. C. KAK, Robot vision system for recognizing objects in low-order polynomial time. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, no. 6, pp. 1535–1536, November 1989.
- [46] L. H. CHEN AND W. C. LIN, Visual surface segmentation from stereo. *Image and Vision Computing*, vol. 15, pp. 95–106, 1997.
- [47] H. CHOSET AND K. NAGATANI, Topological simultaneous localization and mapping (SLAM): Toward exact localization without explicit localization. *IEEE Transactions on Robotics and Automation*, vol. 17, no. 2, pp. 125–137, April 2001.
- [48] W. J. CHRISTMAS, J. KITTLER, AND M. PETROU, Structural matching in computer vision using probabilistic relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 749–764, August 1995.
- [49] G. CICIRELLI, D. DISTANTE, T. D’ORAZIO, AND G. ATTOLICO, Learning actions from vision-based positioning in goal directed navigation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1715–1720, Victoria, October 1998.
- [50] T. H. CORMEN, C. E. LEISERSON, AND R. L. RIVEST, *Introduction to Algorithms*. MIT Electrical Engineering and Computer Science Series, MIT Press, Cambridge, 1992.
- [51] J. D. CRISMAN AND C. E. THORPE, SCARF: A color vision system that tracks roads and intersections. *IEEE Transactions on Robotics and Automation*, vol. 9, no. 1, pp. 49–58, February 1993.
- [52] J. E. CRYER, P.-S. TSAI, AND M. SHAH, Integration of shape from shading and stereo. *Pattern Recognition*, vol. 28, no. 7, pp. 1033–1043, July 1995.
- [53] A. J. DAVISON AND N. KITA, Sequential localisation and map-building for real-time computer vision and robotics. *Robotics and Autonomous Systems*, vol. 36, pp. 171–183, 2001.
- [54] A. J. DAVISON AND D. W. MURRAY, Simultaneous localisation and map-building using active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 865–880, July 2002.

- [55] D. P. S. DE RIVERA, *Estadística Modelos y Métodos*, vol. 1-2. Alianza Editorial, Madrid, 1997.
- [56] M. DEANS AND M. HEBERT, Experimental comparison of techniques for localization and mapping using a bearing-only sensor. In *Proceedings of the 7th International Symposium on Experimental Robotics*, pp. –, Honolulu, December 2000.
- [57] R. A. DECARLO, *Linear Systems. A State Variable Approach with Numerical Implementation*. Prentice Hall, Englewood Cliffs, 1989.
- [58] K. DEMIRLI AND I. B. TÜRKŞEN, Mobile robot navigation with generalized modus ponens type fuzzy reasoning. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3724–3729, Vancouver, October 1995.
- [59] R. DERICHE, Fast algorithms for low-level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 78–88, January 1990.
- [60] R. DERICHE AND G. GIRAUDON, A computational approach for corner and vertex detection. *International Journal of Computer Vision*, vol. 10, no. 2, pp. 101–124, 1993.
- [61] G. N. DESOUSA AND A. C. KAK, Vision for mobile robot navigation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237–267, February 2002.
- [62] E. D. DICKMANN, Recursive state estimation. Lecture Notes. Spring 1996. California Institute of Technology.
- [63] G. DISSANAYAKE, H. DURRANT-WHYTE, AND T. BAILEY, A computationally efficient solution to the simultaneous localisation and map building (SLAM) problem. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1009–1014, San Francisco, April 2000.
- [64] M. W. M. G. DISSANAYAKE, P. NEWMAN, S. CLARK, H. F. DURRANT-WHYTE, AND M. CSORBA, A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229–241, June 2001.
- [65] D. H. DOUGLAS AND T. K. PEUCKER, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Canadian Cartographer*, vol. 10, no. 2, pp. 112–122, December 1973.

- [66] E. R. DOWSKI, An information theory approach to incoherent information processing. In *Proceedings of the 5th Topical Meeting on Signal Recovery and Synthesis*, OSA Technical Digest, pp. 106–108, March 1995.
- [67] L. DRESCHLER AND H. H. NAGEL, On the selection of critical points and local curvature extrema of region boundaries for interframe matching. In *Proceedings of the 6th IAPR International Conference on Pattern Recognition*, pp. 542–544, 1982.
- [68] C. DROCOURT, L. DELAHOUCHE, C. PEGARD, AND A. CLERENTIN, Mobile robot localization based on omnidirectional stereoscopic vision perception system. In *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1329–1334, Detroit, May 1999.
- [69] T. DUCKETT, S. MARSLAND, AND J. SHAPIRO, Learning globally consistent maps by relaxation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3841–3846, San Francisco, April 2000.
- [70] T. DUCKETT AND U. NEHMZOW, Mobile robot self-localization and measurement of performance in middle-scale environments. *Robotics and Autonomous Systems*, vol. 24, no. 1-2, pp. 57–69, August 1998.
- [71] R. O. DUDA, P. E. HART, AND D. G. STORK, *Pattern Classification*. John Wiley & Sons, Inc., New York, 2nd edn., 2001.
- [72] H. DURRANT-WHYTE, S. MAJUMDER, M. DE BATTISTA, AND S. SCHEDING, A bayesian algorithm for simultaneous localisation and map building. In *Proceedings of the 10th International Symposium on Robotics Research*, Lorne, November 2001.
- [73] W. EMDE AND A. SANFELIU, A robust method of object recognition using projective invariants. In A. Calvo and R. Medina, eds., *Proceedings of the 6th Spanish Symposium on Pattern Recognition and Image Analysis*, pp. 647–654, Cordoba, April 1995.
- [74] O. FAUGERAS, *Three-Dimensional Computer Vision. A Geometric Viewpoint*. The MIT Press, Cambridge, 1993.
- [75] P. F. FELZENSZWALB AND D. P. HUTTENLOCHER, Image segmentation using local variation. In *Proceedings of the 12th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 98–104, Santa Barbara, June 1998.
- [76] M. FISCHLER AND R. BOLLES, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, vol. 24, pp. 381–385, 1981.

- [77] D. FOX, W. BURGARD, AND S. THRUN, Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence Research*, vol. 30, pp. 391–427, November 1999.
- [78] K. FUKUNAGA, *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, 2nd edn., 1990.
- [79] K. FURUTA AND A. SANO, *State Variable Methods in Automatic Control*. John Wiley & Sons, Chichester, 1988.
- [80] A. FUSIELLO, E. TRUCCO, AND A. VERRI, Rectification with unconstrained stereo geometry. In A. F. Clark, ed., *Proceedings of the British Machine Vision Conference*, pp. 400–409, Colchester, September 1997.
- [81] P. GAUSSIER, C. JOULIAN, J. P. BANQUET, S. LEPRÊTRE, AND A. RAVEL, The visual homing problem: An example of robotics/biology cross fertilization. *Robotics and Autonomous Systems*, vol. 30, no. 1, pp. 155–180, January 2000.
- [82] P. GAUSSIER, S. LEPRÊTRE, M. QUOY, A. REVEL, C. JOULIAN, AND J. P. BANQUET, Experiments and models about cognitive map learning for motivated navigation. In J. Demiris and A. Birk, eds., *Interdisciplinary Approaches to Robot Learning*, vol. 28 of *World Scientific Series in Robotics and Intelligent Systems*, chap. 3, pp. 53–94, World Scientific, June 2000.
- [83] P. W. GIBBENS, G. M. W. M. DISSANAYAKE, AND H. F. DURRANT-WHYTE, A closed form solution to the single degree of freedom simultaneous localisation and map building (SLAM) problem. In *Proceedings of the IEEE International Conference on Decision and Control*, pp. 408–415, Sydney, December 2000.
- [84] G. GIRAUDON AND R. DERICHE, On corner and vertex detection. In *Proceedings of the 6th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 650–655, Lahaina, June 1991.
- [85] R. C. GONZALEZ AND R. E. WOODS, *Digital Image Processing*. Addison Wesley, Reading, 1993.
- [86] T. GRAEPEL, M. BURGER, AND K. OBERMAYER, Self-organizing maps: Generalizations and new optimization techniques. *Neurocomputing*, vol. 21, no. 1-3, pp. 173–190, 1998.
- [87] P. GROS, O. BOURNEZ, AND E. BOYER, Using local planar geometric invariants to match and model images of line segments. *Computer Vision and Image Understanding*, vol. 69, no. 2, pp. 135–155, February 1998.



- [88] J. GUIVANT AND M. NEBOT, Navigation in large outdoor unstructured environments. In *Proceedings of the 10th International Symposium on Robotics Research*, Lorne, November 2001.
- [89] D. HALPERIN, L. KAVRAKI, AND J.-C. LATOMBE, Robot algorithms. In M. J. Atallah, ed., *Algorithms and Theory of Computation Handbook*, chap. 21, pp. 1–17, CRC Press LLC, Boca Raton, 1999.
- [90] C. G. HARRIS AND M. STEPHENS, A combined corner edge detector. In *Proceedings of the Alvey Vision Conference*, pp. 189–192, Manchester, August 1988.
- [91] R. HARTLEY AND A. ZISSERMAN, *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, 2000.
- [92] M. HASHIMA, F. HASEGAWA, S. KANDA, T. MARUYAMA, AND T. UCHIYAMA, Localization and obstacle detection for a robot for carrying food trays. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 345–351, Grenoble, 1997.
- [93] M. D. HEATH, S. SARKAR, T. SANOCKI, AND K. W. BOWYER, A robust visual method for assessing the relative performance of edge-detection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 12, pp. 1338–1359, December 1997.
- [94] F. HEITGER, Detection using suppression enhancement. Tech. Rep. TR 163, Communications Technologies Laboratory, Swiss Federal Institute of Technology, 1995.
- [95] Y. HEL-OR AND M. WERMAN, Pose estimation by fusing noisy data of different dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 195–201, February 1995.
- [96] J. HERSHBERGER AND J. SNOEYINK, Speeding up the douglas-peucker line-simplification algorithm. In *Proceedings of the 5th International Symposium on Spatial Data Handling*, pp. 134–143, Charleston, August 1992.
- [97] R. HORAUD AND T. SKORDAS, Stereo correspondence through feature grouping and maximum cliques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 11, pp. 1168–1180, December 1989.
- [98] R. HORAUD AND F. VEILLON, Finding geometric and relational structures in an image. In O. Faugeras, ed., *Proceedings of the 1st European Conference on Computer Vision*, vol. 427 of *Lecture Notes in Computer Science*, pp. 374–384, Springer-Verlag, Antibes, April 1990.

- 
- [99] B. K. P. HORN, Extended gaussian images. *Proceedings of the IEEE*, vol. 72, no. 12, pp. 1671–1686, December 1984.
- [100] P. V. C. HOUGH, Methods and means for recognizing complex patterns. U.S. Patent No. 3,069,654, 1962.
- [101] D. Q. HUYNH AND R. A. OWENS, Line labelling and region segmentation in stereo image pairs. *Image and Vision Computing*, vol. 12, no. 4, pp. 213–225, May 1994.
- [102] I. C. F. IPSEN AND C. D. MEYER, The angle between complementary subspaces. Tech. Rep. NCSU-TR-NA-019501, Mathematics Dept. North Carolina State Univ., Raleigh, January 1995.
- [103] R. C. JAIN, R. KASTURI, AND B. G. SCHUNCK, *Machine Vision*. McGraw Hill, 1995.
- [104] C. JENNINGS, D. R. MURRAY, AND J. J. LITTLE, Cooperative robot localization with vision-based mapping. In *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 4, pp. 2659–2665, Detroit, May 1999.
- [105] S. JULIER, J. UHLMANN, AND H. F. DURRANT-WHYTE, A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, March 2000.
- [106] S. J. JULIER, A sparse weight Kalman filter approach to simultaneous localization and map building. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1251–1256, Maui, November 2001.
- [107] S. J. JULIER AND J. K. UHLMANN, A new extension of the Kalman filter to nonlinear systems. In I. Kadar, ed., *Proceedings of the 11th SPIE International Symposium on Aerospace/Defense Sensing, Simulation, and Controls*, pp. 182–193, International Society for Optical Engineering, Orlando, April 1997.
- [108] S. J. JULIER AND J. K. UHLMANN, A counter example to the theory of simultaneous localization and map building. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 4238–4243, Seoul, May 2001.
- [109] S. J. JULIER AND J. K. UHLMANN, Simultaneous localisation and map building using split covariance intersection. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Maui, November 2001.
- [110] T. KAILATH, *Linear Systems*. Information and System Sciences Series, Prentice Hall, Englewood Cliffs, 1980.

- [111] R. E. KALMAN, A new approach to linear filtering and prediction problems. *Journal of Basic Engineering - Transactions of the ASME*, pp. 35–45, March 1960.
- [112] R. E. KALMAN AND R. S. BUCY, New results in linear filtering and prediction theory. *Journal of Basic Engineering - Transactions of the ASME*, pp. 95–108, March 1961.
- [113] T. KANADE, H. KANO, S. KIMURA, E. KAWAMURA, A. YOSHIDA, AND K. ODA, Development of a video-rate stereo machine. *Journal of the Robotics Society of Japan*, vol. 15, no. 2, pp. 99–105, March 1997.
- [114] K. KANATANI, *Geometric Computation for Machine Vision*. Oxford Engineering Science Series, Clarendon Press, Oxford, 1993.
- [115] M. KASS, A. WITKIN, AND D. TERZOPOULOS, Snakes: Active contour models. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 259–268, London, June 1987.
- [116] L. KITCHEN AND A. ROSENFELD, Gray-level corner detection. *Pattern Recognition Letters*, vol. 1, pp. 95–102, December 1982.
- [117] K. KONOLIGE, Small vision systems: Hardware and implementation. In *Proceedings of the International Symposium on Robotics Research*, pp. 203–212, Hayama, November 1997.
- [118] K. KONOLIGE AND K. MYERS, The saphira architecture for autonomous mobile robots. In D. Kortenkamp, R. P. Bonasso, and R. Murphy, eds., *Artificial Intelligence and Mobile Robots. Case Studies of Successful Robot Systems*, MIT Press, Menlo Park, January 1998.
- [119] A. KOSAKA AND A. C. KAK, Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties. *Computer Vision and Image Understanding*, vol. 56, no. 3, pp. 271–329, November 1992.
- [120] A. KOSAKA AND A. C. KAK, Stereo vision for industrial applications. In S. Y. Nof, ed., *Handbook for Industrial Robotics*, pp. 269–294, John Wiley & Sons, New York, 1999.
- [121] J. KOŠECKÁ, R. BAJCSY, AND M. MINTZ, Control of visually guided behaviors. In C. M. Brown and D. Terzopoulos, eds., *Real-Time Computer Vision*, Cambridge Univ. Press, Cambridge, 1995.
- [122] Y. D. KWON AND J. S. LEE, A stochastic map building method for mobile robot using 2-d laser range finder. *Autonomous Robots*, vol. 7, no. 2, pp. 187–200, 1999.

- [123] S. LACROIX, P. GRANDJEAN, AND M. GHALLAB, Perception planning for a multi-sensory interpretation machine. In *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1818–1824, Nice, May 1992.
- [124] R. A. LANE, N. A. THACKER, AND N. L. SEED, Stretch-correlation as a real-time alternative to feature-based stereo matching algorithms. *Image and Vision Computing*, vol. 12, no. 4, pp. 203–212, May 1994.
- [125] D. LEE AND M. RECCE, Quantitative evaluation of the exploration strategies of a mobile robot. *International Journal of Robotics Research*, vol. 16, no. 4, pp. 413–447, August 1997.
- [126] J. J. LEONARD, H. F. DURRANT-WHYTE, AND I. J. COX, Dynamic map building for an autonomous mobile robot. *International Journal of Robotics Research*, vol. 11, no. 4, pp. 286–292, 1992.
- [127] J. J. LEONARD, P. M. NEWMAN, R. J. RIKOSKI, J. NEIRA, AND J. D. TARDÓS, Towards robust data association and feature modeling for concurrent mapping and localization. In *Proceedings of the 10th International Symposium on Robotics Research*, Lorne, November 2001.
- [128] J. J. LEONARD AND R. RIKOSKI, Incorporation of delayed decision making into stochastic mapping. In *Proceedings of the 7th International Symposium on Experimental Robotics*, Honolulu, December 2000.
- [129] J. J. LITTLE, J. LU, AND D. R. MURRAY, Selecting stable image features for robot localization using stereo. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1072–1077, Victoria, October 1998.
- [130] A. LÖFFLER, J. KLAHOLD, AND U. RÜCKERT, Artificial neural networks for autonomous robot control: Reflective navigation and adaptive sensor calibration. In *Proceedings of the International Conference on Neural Information Processing*, pp. 667–672, Perth, November 1999.
- [131] C. LOOP AND Z. ZHANG, Computing rectifying homographies for stereo vision. In *Proceedings of the 13th IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 125–131, Fort Collins, June 1999.
- [132] M. LÓPEZ-SÁNCHEZ, R. L. DE MANTARAS, AND C. SIERRA, Map generation by means of autonomous robots and possibility propagation techniques. In *Proceedings of the International Conference on Autonomous Agents*, pp. 380–381, Seattle, May 1999.

- 
- [133] D. G. LOWE, *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, Hingham, June 1985.
- [134] M. LÓPEZ-SÁNCHEZ, F. ESTEVA, R. L. DE MÀNTARAS, C. SIERRA, AND J. AMAT, Map generation by cooperative low-cost robots in structured unknown environments. *Autonomous Robots*, vol. 5, pp. 53–61, 1998.
- [135] A. MALLET AND S. LACROIX, Toward real-time 2d localization in outdoor environments. In *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 4, pp. 2827–2832, Leuven, May 1998.
- [136] R. MANDELBAUM, G. KAMBEROVA, AND M. MINTZ, Statistical decision theory for mobile robotics: Theory and application. In *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 17–24, Washington, December 1996.
- [137] D. MARR, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, San Francisco, 1982.
- [138] P. S. MAYBECK, *Stochastic Models, Estimation, and Control*, vol. 1. Academic Press, New York, 1979.
- [139] S. MEIKIE, R. YATES, AND A. HARRIS, Computer vision algorithms for autonomous mobile robot map building and path planning. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, pp. 99–104, Boston, 1997.
- [140] J. MILLÁN AND C. TORRAS, Learning sensor-based navigation. In K. Morik, M. Kaiser, and V. Klingspor, eds., *Making Robots Smarter: Combining Sensing and Action through Robot Learning*, pp. 88–108, Kluwer Academic Publishers, Boston, 1999.
- [141] R. MOHR AND B. TRIGGS, Projective geometry for image analysis. In *Proceedings of the ISPRS Conference*, Vienna, July 1996.
- [142] P. MONTESINOS, V. GOUET, AND R. DERICHE, Differential invariants for color images. In *Proceedings of the 14th IAPR International Conference on Pattern Recognition*, pp. 838–840, Brisbane, August 1998.
- [143] H. P. MORAVEC AND A. ELFES, High resolution maps from wide angle sonar. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 116–121, St. Louis, March 1985.
- [144] F. MORENO, J. ANDRADE-CETTO, AND A. SANFELIU, Desenvolupament d'un sistema d'estereovisió per un robot mòbil. Tech. Rep. IRI-DT-01-01, Institut de

- Robòtica i Informàtica Industrial, Universitat Politècnica de Catalunya, March 2001.
- [145] F. MORENO, J. ANDRADE-CETTO, AND A. SANFELIU, Localization of human faces fusing color segmentation and depth from stereo. In *Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation*, pp. 527–536, Antibes, October 2001.
- [146] F. MORENO, A. TARRIDA, J. ANDRADE-CETTO, AND A. SANFELIU, 3d real-time head tracking fusing color histograms and stereovision. In *Proceedings of the 16th IAPR International Conference on Pattern Recognition*, vol. 1, pp. 368–371, IEEE Computer Society, Quebec, August 2002.
- [147] S. G. NADABAR AND A. K. JAIN, Fusion of range and intensity images on a connection machine (CM-2). *Pattern Recognition*, vol. 28, no. 1, pp. 11–26, January 1995.
- [148] O. NAKAYAMA, A. YAMAGUCHI, Y. SHIRAI, AND M. ASADA, A multistage stereo method giving priority to reliable matching. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1753–1758, Nice, May 1992.
- [149] P. J. NARAYANAN, P. W. RANDEK, AND T. KANADE, Constructing virtual worlds using dense stereo. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3–10, Bombay, January 1998.
- [150] J. NEIRA AND J. D. TARDÓS, Data association in stochastic mapping using the joint compatibility test. *IEEE Transactions on Robotics and Automation*, vol. 17, no. 6, pp. 890–897, December 2001.
- [151] P. M. NEWMAN, *On the Structure and Solution of the Simultaneous Localisation and Map Building Problem*. Ph.D. thesis, The University of Sydney, Sydney, March 1999.
- [152] Y. OHTA AND T. KANADE, Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 139–154, March 1985.
- [153] A. OHYA, A. KOSAKA, AND A. C. KAK, Vision-based navigation by a mobile robot with obstacle avoidance using single-camera vision and ultrasonic sensing. *IEEE Transactions on Robotics and Automation*, vol. 14, no. 6, pp. 969–978, December 1998.
- [154] G. ORIOLO, G. ULIVI, AND M. VENDITTELLI, Real-time map building and navigation for autonomous robots in unknown environments. *IEEE Transactions*

- on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 28, no. 3, pp. 316–332, June 1998.
- [155] J. PAGÈS, J. ARANDA, AND A. CASALS, A 3d vision system to model industrial environments for agv's control. In *Proceedings of the International Symposium on Industrial Robots*, Tokyo, October 1993.
- [156] S. PANKANTI AND A. K. JAIN, Integrating vision modules: Stereo, shading, grouping, and line labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 831–842, September 1995.
- [157] A. PAPOULIS, *Probability, Random Variables, and Stochastic Processes*. McGraw Hill, New York, 3rd edn., 1991.
- [158] S. POLAVARAPU, Data assimilation. Lecture Notes. 2002. Dept. of Physics, Univ. of Toronto.
- [159] M. POLLEFEYS, 3d modeling from images. tutorial. In D. Vernon, ed., *Proceedings of the 6th European Conference on Computer Vision*, vol. 1842, 1843 of *Lecture Notes in Computer Science*, Springer-Verlag, Dublin, June 2000.
- [160] M. POLLEFEYS, R. KOCH, AND L. V. GOOL, A simple and efficient rectification method for general motion. In *Proceedings of the IEEE International Conference on Computer Vision*, vol. 1, pp. 496–501, Corfu, September 1999.
- [161] M. POLONI, G. ULIVI, AND M. VENDITTELLI, Fuzzy logic and autonomous vehicles: Experiments in ultrasonic vision. *Fuzzy Sets and Systems*, vol. 69, pp. 15–27, 1995.
- [162] F. P. PREPARATA AND M. I. SHAMOS, *Computational Geometry*. Texts and Monographs in Computer Science, Springer-Verlag, New York, 1985.
- [163] J. A. PÉREZ, J. A. CASTELLANOS, J. M. M. MONTIEL, AND J. D. TARDÓS, Continuous mobile robot localization: Vision vs. laser. In *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 4, pp. 2917–2923, Detroit, May 1999.
- [164] I. B. RISTESKI AND K. G. TRENČEVSKI, Principal values and principal subspaces of vector spaces with inner product. *Contributions to Algebra and Geometry*, vol. 42, no. 1, pp. 289–300, 2001.
- [165] A. ROSENFELD AND A. C. KAK, *Digital Picture Processing*, vol. 1-2. Academic Press, New York, 1982.
- [166] C. ROTHWELL, O. FAUGERAS, AND G. CSURKA, A comparison of projective reconstruction methods for pairs of views. *Computer Vision and Image Understanding*, vol. 68, no. 1, pp. 37–58, October 1997.

- [167] S. ROWEIS AND Z. GHAHRAMANI, A unifying review of linear gaussian models. *Neural Computation*, vol. 11, pp. 305–345, 1999.
- [168] A. SAFFIOTTI, The uses of fuzzy logic in autonomous robot navigation: a catalogue raisonné. *Soft Computing*, vol. 1, no. 4, pp. 180–197, 1997.
- [169] A. SAFFIOTTI AND L. P. WESLEY, Perception-based self-localization using fuzzy locations. In *Proceedings of the International Workshop on Reasoning with Uncertainty in Robot Navigation*, pp. 368–385, Amsterdam, 1996.
- [170] A. SANFELIU AND M. AÑAÑOS, A cad based vision system for identifying industrial work pieces. In *Proceedings of the IFAC Symposium on Information Control Problems in Manufacturing Technology*, Madrid, September 1989.
- [171] A. SANFELIU, M. AÑAÑOS, AND M. J. DUNJÓ, Integrating driving model and depth for identification of partially occluded 3d models. In *Proceedings of the NATO Advanced Research Workshop on Multisensor Fusion for Computer Vision*, pp. 26–30, Grenoble, June 1989.
- [172] A. SANFELIU, R. ALQUÉZAR, J. ANDRADE, J. CLIMENT, F. SERRATOSA, AND J. VERGÉS, Graph-based representations and techniques for image processing and image analysis. *Pattern Recognition*, vol. 35, pp. 639–650, March 2002.
- [173] A. SANFELIU, J. ANDRADE-CETTO, R. ALQUÉZAR, J. ARANDA, J. CLIMENT, A. GRAU, F. SERRATOSA, AND J. VERGÉS-LLAHÍ, MARCO: A mobile robot with learning capabilities to perceive and interact with its environment. In J. S. Sánchez and F. Pla, eds., *Proceedings of the 9th Spanish Symposium on Pattern Recognition and Image Analysis*, vol. 2, pp. 219–224, Universitat Jaume I, Benicasim, May 2001.
- [174] A. SANFELIU, A. GRAU, J. CLIMENT, R. ALQUÉZAR, F. SERRATOSA, J. ARANDA, J. VERGÉS, AND J. ANDRADE, Pattern recognition research at the IRI-CSIC/ESAI group. In A. Grau, ed., *Proceedings of the 1st Foro Iberoamericano de Reconocimiento de Formas y Análisis de Imágenes*, chap. 29, pp. 345–354, UPC Information, Image and Publication Service, Barcelona, September 2000.
- [175] A. SANFELIU AND A. LLORENS, Projective invariants to identify polyhedric objects. In *Proceedings of the 5th Spanish Symposium on Pattern Recognition and Image Analysis*, pp. 65–77, 1993.
- [176] A. SANFELIU, F. SERRATOSA, AND R. ALQUÉZAR, Clustering of attributed and unsupervised synthesis of function-described graphs. In A. Sanfeliu, J. J. Villanueva, M. Vanrell, R. Alquézar, J. O. Eklundh, and Y. Aloimonos, eds.,



- Proceedings of the 15th IAPR International Conference on Pattern Recognition*, vol. 2, pp. 1026–1029, IEEE Computer Society, Barcelona, September 2000.
- [177] S. SARKAR AND K. L. BOYER, Integration, inference, and management of spatial information using bayesian networks: Perceptual organization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 3, pp. 256–276, March 1993.
- [178] O. SARZEAUD AND Y. STÉPHAN, Data interpolation using kohonen networks. In *Proceedings of the International Conference on Neural Networks*, pp. 197–202, Cuomo, July 2000.
- [179] F. SERRATOSA, *Function-Described Graphs for Pattern Recognition*. Ph.D. thesis, Univ. Politècnica de Catalunya, Barcelona, July 2000.
- [180] F. SERRATOSA AND A. SANFELIU, Function-described graphs applied to 3d object representation. In A. del Bimbo, ed., *Proceedings of the International Conference on Image Analysis and Processing*, vol. 1, pp. 701–708, Florence, September 1997.
- [181] J. SHI AND C. TOMASI, Good features to track. In *Proceedings of the 9th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593–600, Seattle, June 1994.
- [182] R. C. SMITH AND P. CHEESEMAN, On the representation and estimation of spatial uncertainty. *International Journal of Robotics Research*, vol. 5, no. 4, pp. 56–68, 1986.
- [183] B. SOUTHALL, B. F. BUXTON, AND J. A. MARCHANT, Controllability and observability: Tools for kalman filter design. In J. N. Carter and M. S. Nixon, eds., *Proceedings of the British Machine Vision Conference*, pp. 164–173, Southampton, 1998.
- [184] G. STRANG, *Linear Algebra and its Applications*. Saunders College Publishing, Fort Worth, 3rd edn., 1988.
- [185] M. TAKATSUKA AND R. A. JARVIS, Encoding 3d structural information using multiple self-organized feature maps. *Image and Vision Computing*, vol. 19, no. 3, pp. 99–176, February 2001.
- [186] R. TALLURI AND J. K. AGGARWAL, Mobile robot self-location using model-image feature correspondence. *IEEE Transactions on Robotics and Automation*, vol. 12, no. 1, pp. 63–77, February 1996.

- [187] J. D. TARDÓS, J. NEIRA, P. M. NEWMAN, AND J. J. LEONARD, Robust mapping and localization in indoor environments using sonar data. *International Journal of Robotics Research*, vol. 21, no. 4, pp. 311–330, 2002.
- [188] C. E. THORPE, M. H. HEBERT, T. KANADE, AND S. A. SHAFER, Vision and navigation for the Carnegie Mellon Navlab. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 3, pp. 362–373, May 1988.
- [189] C. E. THORPE, M. H. HEBERT, T. KANADE, AND S. A. SHAFER, Toward autonomous driving: The CMU Navlab. part i - perception. *IEEE Expert*, pp. 31–42, August 1991.
- [190] C. E. THORPE, M. H. HEBERT, T. KANADE, AND S. A. SHAFER, Toward autonomous driving: The CMU Navlab. part ii - architecture and systems. *IEEE Expert*, pp. 44–52, August 1991.
- [191] S. THRUN, Bayesian landmark learning for mobile robot localization. *Machine Learning*, vol. 33, no. 1, pp. 41–76, October 1998.
- [192] S. THRUN, Probabilistic algorithms in robotics. *Artificial Intelligence Magazine*, vol. 21, no. 4, pp. 93–109, 2000.
- [193] S. THRUN, D. FOX, W. BURGARD, AND F. DELLAERT, Robust monte carlo localization for mobile robots. *Artificial Intelligence*, vol. 128, pp. 99–141, May 2001.
- [194] R. TODLING, Estimation theory and foundations on atmospheric data assimilation. Tech. Rep. DAO Note 1999-01, Data Assimilation Office. Goddard Space Flight Center, 1999.
- [195] C. TOMASI, Mathematical methods for robotics and vision. Lecture Notes. Spring 2000. Stanford University.
- [196] C. TOMASI AND T. KANADE, Detection and tracking of point features. Tech. Rep. CMU CS 91 132, Carnegie Mellon University, Pittsburgh, April 1991.
- [197] C. TORRAS, Robot adaptivity. *Robotics and Autonomous Systems*, vol. 15, pp. 11–23, 1995.
- [198] C. TORRAS, Robot control. In M. Arbib, ed., *The Handbook of Brain Theory and Neural Networks*, pp. 820–823, The MIT Press, Cambridge, 1995.
- [199] M. TRAJKOVIĆ AND M. HEDLEY, Fast corner detection. *Image and Vision Computing*, vol. 16, pp. 75–87, 1998.

- [200] R. TSAI, A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras. *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, August 1987.
- [201] T. TSUBOUCHI, Nowadays trends in map generation for mobile robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 828–833, Osaka, November 1996.
- [202] S. W. UTETE, B. BARSHAN, AND B. AYRULU, Voting as validation in robot programming. *International Journal of Robotics Research*, vol. 18, no. 4, pp. 401–413, April 1999.
- [203] R. VAN DER MERWE AND E. A. WAN, The square-root unscented Kalman filter for state and parameter-estimation. In *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 3461–3464, IEEE Signal Processing Society, Salt Lake City, May 2001.
- [204] S. VEDULA, S. BAKER, S. SEITZ, AND T. KANADE, Shape and motion carving in 6d. In *Proceedings of the 14th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 592–598, Head Island, June 2000.
- [205] V. VENKATESWAR AND R. CHELLAPA, Hierarchical stereo and motion correspondence using feature groupings. *International Journal of Computer Vision*, vol. 15, no. 3, pp. 245–269, 1995.
- [206] J. VERGÉS-LLAHÍ, J. CLIMENT, AND A. SANFELIU, Color image segmentation solving hard-constraints on graph partitioning greedy algorithm. In A. Sanfeliu, J. J. Villanueva, M. Vanrell, R. Alquézar, J. O. Eklundh, and Y. Aloimonos, eds., *Proceedings of the 15th IAPR International Conference on Pattern Recognition*, vol. 3, pp. 229–232, IEEE Computer Society, Barcelona, September 2000.
- [207] R. VOLPE, T. LITWIN, AND L. MATTHIES, Mobile robot localization by remote viewing of a colored cylinder. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, pp. 257–263, Pittsburgh, 1995.
- [208] H. WANG AND M. BRADY, Real-time corner detection algorithm for motion estimation. *Image and Vision Computing*, vol. 13, no. 9, pp. 695–703, November 1995.
- [209] A. R. WEEKS, *Fundamentals of Electronic Image Processing*. Imaging Science & Engineering, SPIE / IEEE, Bellingham, 1996.
- [210] G. Q. WEI AND G. HIRZINGER, Intensity and feature based stereo matching by disparity parameterization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1035–1040, Bombay, January 1998.

- [211] G. WELCH AND G. BISHOP, SCAAT: Incremental tracking with incomplete information. In T. Whitted, ed., *Computer Graphics. Proceedings of the ACM SIGGRAPH Conference*, pp. 333–344, ACM Press, Los Angeles, August 1997.
- [212] G. WELCH AND G. BISHOP, An introduction to the Kalman filter. Course 8. In *Proceedings of the ACM SIGGRAPH Conference*, ACM Press, Los Angeles, August 2001.
- [213] G. F. WELCH, *SCAAT: Incremental Tracking with Incomplete Information*. Ph.D. thesis, The University of North Carolina at Chapel Hill, Chapel Hill, October 1996.
- [214] J. T. WEN, Optimal control. Lecture Notes. 2002. ECSE, Rensselaer Polytechnic Institute.
- [215] J. WENG, P. COHEN, AND M. HERNIOU, Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 10, pp. 965–980, October 1992.
- [216] J. WENG, T. S. HUANG, AND N. AHUJA, Motion and structure from two perspective views: Algorithms, error analysis, and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 451–476, May 1989.
- [217] R. C. WILSON AND E. R. HANCOCK, Structural matching by discrete relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 634–648, June 1997.
- [218] S. XAMBÓ, *Geometria*. Edicions UPC, Barcelona, 1997.
- [219] G. XU AND Z. ZHANG, *Epipolar Geometry in Stereo, Motion and Object Recognition. A Unified Approach*. Kluwer Academic Publisher, Dordrecht, 1996.
- [220] Z. ZHANG, Parameter estimation techniques: A tutorial with application to conic fitting. *Image and Vision Computing*, vol. 15, no. 1, pp. 59–76, 1997.