

Estimació del moviment de robots mitjançant contorns actius



Guillem Alenyà Ribas

Tesi realitzada al programa de doctorat

Control, Visió i Robòtica (ESAI)

Directora:

Carme Torras

Aquesta tesi es presenta per a l'obtenció del grau de

Doctor

Barcelona, 2007

Universitat Politècnica de Catalunya

Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial

Programa de doctorat:
Control, Visió i Robòtica

Aquesta tesi ha estat realitzada a:

Institut de Robòtica i Informàtica Industrial, CSIC-UPC

Directora de tesi:
Carme Torras

Tribunal de tesi:
Alberto Sanfeliu (Universitat Politècnica de Catalunya) (President)
James L. Crowley (INRIA - Rhone-Alpes)
Rüdiger Dillmann (Universität Karlsruhe)
Norbert Krüger (University of Southern Denmark)
Elisa Martínez Marroquín (Universitat Ramon Llull)

© Guillem Alenyà 2007

Voldria que tots els que llegissin la meva novel·la participessin en la meva emoció. Em fa contenta pensar que entre tants milers de lectors com ha tingut i continua tenint n'hi ha molts que no havien llegit mai res en català i que és llegint-la que han descobert que la nostra era una llengua civilitzada, culta, important.

I would like all that read my novel to participate in my emotion. It makes me happy to think that, among the many thousands of readers that it has had and it continues having, there are many who had never read anything in Catalan and it is by reading my novel that they have discovered that ours is a civilized, cultivated, meaningful language.

Mercè Rodoreda. *La plaça del diamant* (pròleg).

A la Susanna, per la vida que ens dediquem. No m'ho imagino d'altra manera. I per les dues *tesis* que tenim a mitges. La seva inquietud per aprendre és inspiració.

A la Carme. Gràcies per compartir el camí; i donar-me bones eines.
A tots els companys de causa, gràcies per les discussions i per deixar-me pensar en veu alta.

Resum

Aquesta tesi versa sobre l'estimació del moviment d'un robot mòbil a partir dels canvis en les imatges captades per una càmera muntada sobre el robot. El moviment es dedueix amb un algorisme prèviament proposat en el marc de la navegació qualitativa. Per tal d'emprar aquest algorisme en casos reals s'ha fet un estudi de la seva precisió. Per augmentar-ne l'aplicabilitat, s'ha adaptat l'algorisme al cas d'una càmera amb moviments d'orientació i de zoom.

Quan els efectes perspectius no són importants, dues vistes d'una escena captades pel robot es poden relacionar amb una transformació afí (o afinitat), que normalment es calcula a partir de correspondències de punts. En aquesta tesi es vol seguir un enfoc alternatiu, i alhora complementari, fent servir la silueta d'un objecte modelada mitjançant un contorn actiu. El marc es el següent: a mesura que el robot es va movent, la projecció de l'objecte a la imatge va canviant i el contorn actiu es deforma convenientment per adaptar-s'hi; de les deformacions d'aquest contorn, expressades en espai de forma, se'n pot extreure el moviment del robot fins a un factor d'escala. Els contorns actius es caracteritzen per la rapidesa en la seva extracció i la seva robustesa a oclusions parcials. A més, un contorn és fàcil de trobar fins i tot en escenes poc texturades, on sovint és difícil trobar punts característics i la seva correspondència.

La primera part d'aquest treball té l'objectiu de caracteritzar la precisió i la incertesa en l'estimació del moviment. Per avaluar la precisió, primer es duen a terme un parell d'experiències pràctiques, que mostren la potencialitat de l'algorisme en entorns reals i amb diferents robots. Estudiant la geometria epipolar que relaciona dues vistes d'un objecte planar es demostra que la direcció epipolar afí es pot recuperar en el cas que el moviment de la càmera estigui lliure de ciclorotació. Amb una bateria d'experiments, tant en simulació com reals, es fa servir la direcció epipolar per caracteritzar la precisió global de l'afinitat en diferents situacions, com ara, davant de diferents formes dels contorns, condicions de visualització extremes i soroll al sistema.

Pel que fa a la incertesa, gràcies a que la implementació es basa en el filtre de Kalman, per a cada estimació del moviment també es té una estimació de la incertesa associada, però expressada en espai de forma. Per tal propagar la incertesa de l'espai de forma a l'espai de moviment 3D s'han seguit dos camins diferents: un analític i l'altre estadístic. Aquest estudi ha permès determinar quins graus de llibertat es recuperen amb més precisió, i quines correlacions existeixen entre les diferents components. Final-

ment, s'ha desenvolupat un algorisme que permet propagar la incertesa del moviment en temps de vídeo.

Una de les limitacions més importants d'aquesta metodologia és que cal que la projecció de l'objecte estigui dins de la imatge i en condicions de visualització de perspectiva dèbil durant tota la seqüència. En la segona part d'aquest treball, s'estudia el seguiment de contorns actius en el marc de la visió activa per tal de superar aquesta limitació. És una relació natural, atès que el seguiment de contorns actius es pot veure com una tècnica per fixar el focus d'atenció.

En primer lloc, s'han estudiat les propietats de les càmeres amb zoom i s'ha proposat un nou algorisme per determinar la profunditat de la càmera respecte a un objecte qualsevol. L'algorisme inclou un senzill calibratge geomètric que no implica cap coneixement sobre els paràmetres interns de la càmera.

Finalment, per tal d'orientar la càmera adequadament, compensant en la mesura del possible els moviments del robot, s'ha desenvolupat un algorisme per al control dels mecanismes de zoom, capcineig i guinyada, i s'ha adaptat l'algorisme d'estimació del moviment incorporant-hi els girs coneguts del capcineig i la guinyada.

Abstract

This thesis deals with the motion estimation of a mobile robot from changes in the images acquired by a camera mounted on the robot itself. The motion is deduced with an algorithm previously proposed in the framework of qualitative navigation. In order to employ this algorithm in real situations, a study of its accuracy has been performed. Moreover, relationships with the active vision paradigm have been analyzed, leading to an increase in its applicability.

When perspective effects are not significant, two views of a scene are related by an affine transformation (or affinity), that it is usually computed from point correspondences. In this thesis we explore an alternative and at the same time complementary approach, using the contour of an object modeled by means of an active contour. The framework is the following: when the robot moves, the projection of the object in the image changes and the active contour adapts conveniently to it; from the deformation of this contour, expressed in shape space, the robot egomotion can be extracted up to a scale factor. Active contours are characterized by the speed of their extraction and their robustness to partial occlusions. Moreover, a contour is easy to find even in poorly textured scenes, where often it is difficult to find point features and their correspondences.

The goal of the first part of this work is to characterize the accuracy and the uncertainty in the motion estimation. Some practical experiences are carried out to evaluate the accuracy, showing the potentiality of the algorithm in real environments and with different robots. We have studied also the epipolar geometry relating two views of a planar object. We prove that the affine epipolar direction between two images can be recovered from a shape vector when the camera motion is free of cyclorotation. With a battery of simulated as well as real experiments, the epipolar direction allows us to analyze the global accuracy of the affinity in a variety of situations: different contour shapes, extreme visualization conditions and presence of noise.

Regarding uncertainty, since the implementation is based on a Kalman filter, for each motion estimate we have also its covariance matrix expressed in shape space. In order to propagate the uncertainty from shape space to 3D motion space, two different approaches have been followed: an analytical and a statistical one. This study has allowed us to determine which degrees of freedom are recovered with more accuracy, and what correlations exist between the different motion components. Finally, an

algorithm to propagate the motion uncertainty at video rate has been proposed.

One of the most important limitations of this methodology is that the object must project onto the image under weak-perspective visualization conditions all along the sequence. In the second part of this work, active contour tracking is studied within the framework of active vision to overcome this limitation. Both relate naturally, as active contour tracking can be seen as a focus-of-attention strategy.

First, the properties of zooming cameras are studied and a new algorithm is proposed to estimate the depth of the camera with respect to an object. The algorithm includes a simple geometric calibration that does not require any knowledge about the camera internal parameters.

Finally, in order to orientate the camera so as to suitably compensate for robot motion when possible, a new algorithm has been proposed for the control of zoom, pan and tilt mechanisms, and the motion estimation algorithm has been updated conveniently to incorporate the active camera state information.

Índex

1	Introducció	1
1.1	Motivació	1
1.2	Compendi de la tesi	4
1	Introduction	7
1.1	Motivation	7
1.2	Summary of the thesis	10
2	Punt de partida: vector de forma i posa 3D	13
2.1	Introducció	14
2.2	Antecedents	16
2.2.1	Model de càmera	16
2.2.2	L'afinitat que relaciona dues vistes	19
2.2.3	Moviment 3D a partir de l'afinitat	22
2.2.4	Detalls de la implementació	24
2.3	Dues experiències pràctiques	25
2.3.1	Avaluació de l'error en el moviment estimat	25
2.3.2	Enriquiment del seguiment de contorns amb informació inercial	31
2.4	Conclusions	36
3	Recuperació de la direcció epipolar	39
3.1	Introducció	40
3.2	Estudi analític de dues vistes amb perspectiva dèbil	42
3.2.1	El model de càmera	42
3.2.2	Recuperació de la direcció epipolar	43

ÍNDEX

3.2.3	Computant la direcció epipolar a partir dels paràmetres de l'afinitat	44
3.3	Experimentació	45
3.3.1	Consideracions preliminars	45
3.3.2	Efectes de la forma del contorn	47
3.3.3	Relaxació de les condicions de perspectiva dèbil	49
3.3.4	Efectes de la ciclorrotació	53
3.3.5	Sensibilitat al soroll	53
3.3.6	Comparació amb l'algorisme estàndard d'or	54
3.3.7	Experiments amb imatges reals	57
3.3.8	Un comentari: l'interès de reduir l'espai de forma	60
3.3.9	Consideracions d'aplicabilitat	61
3.4	Conclusions	62
4	Estudi analític de la propagació de l'error	65
4.1	Introducció	66
4.2	Propagació als vectors i valors propis	68
4.3	Aproximació de primer ordre	70
4.3.1	Aplicació a l'estimació del moviment 3D	73
4.4	Resultats experimentals	78
4.5	Conclusions	81
5	Estudi de la precisió	83
5.1	Introducció	84
5.2	Monte Carlo	86
5.3	Representació de les rotacions	86
5.4	Avaluant la precisió de les components del moviment	89
5.4.1	Propagació de la incertesa a cada component del moviment	89
5.4.2	Covariància del moviment	96
5.5	Experiments amb seqüències reals	98
5.5.1	La <i>Unscented Transformation</i>	99
5.5.2	Fent servir la UT per estimar la mitjana i la covariància del moviment	101
5.6	Conclusions i treball futur	108

6	La visió activa: profunditat a partir del <i>zoom</i>	111
6.1	Introducció	112
6.2	Encara és vàlid el model de càmera estenopecica?	115
6.3	Cal considerar el efectes de les distorsions en les lents?	118
6.4	Reducció de l'espai de forma	121
6.5	Estimació de la profunditat a partir del moviment induït pel <i>zoom</i> . . .	121
6.5.1	Algorisme de calibratge	122
6.5.2	Com es pot inferir la profunditat	123
6.6	Experiments	125
6.7	Conclusions	131
7	Estimació del moviment en el context de la visió activa	135
7.1	Introducció	136
7.2	Compensant la translació amb el <i>zoom</i>	138
7.2.1	Generant les peticions per al <i>zoom</i>	141
7.2.2	Implicacions per a l'algorisme de recuperació del moviment . . .	143
7.2.3	Resultats experimentals	144
7.3	Moviment d'orientació de la càmera	148
7.3.1	Estratègies de control de l'orientació de la càmera	150
7.3.2	El <i>zoom</i> preserva la resolució dels moviments en el pla frontoparal·lel	153
7.3.3	Implicacions per a l'algorisme de moviment	154
7.3.4	Experiments	158
7.4	Conclusions	164
8	Conclusions i treball futur	169
8.1	Contribucions	170
8.2	Treball futur	172
8	Conclusions and future work	175
8.1	Contributions	176
8.2	Future work	178
A	Rotacions expressades en angles d'Euler	181

ÍNDEX

B Propagació del soroll als valors i vectors propis	183
C Resultats del càlcul de la covariància en el moviment	185
C.1 Càmera en la posició inicial	185
C.2 Moviments simples de la càmera	188
D Mètodes per seleccionar punts sigma	193
Producció científica	195
Bibliografia	198

Capítol 1

Introducció

It is the scientist duty to raise our abilities by increasing our standards for quality and effectiveness, to show what can be done well by pursuing the just possible, and to clarify ruthlessly, independently of the fact that complexity sells better.

És l'obligació del científic augmentar les nostres habilitats incrementant els nivells de qualitat i eficàcia, mostrar què es pot fer bé avançant el just possible, i aclarir sense compassió, independentment del fet que la complexitat es vengui millor.

Edsger W. Dykstra, abril 1986. EDW956-2.

1.1 Motivació

Un robot es pot definir com un aparell mecànic capaç de percebre i manipular l'entorn en què es troba [133]. Cada vegada més, els robots han d'operar en entorns no estructurats i dinàmics, on la capacitat sensorial esdevé crucial per ajudar a solucionar el gran nombre de situacions imprevistes que poden aparèixer. En aquest escenari, la robòtica és cada vegada menys un problema mecànic i esdevé una ciència que té com a objectiu el desenvolupament d'algorismes cada vegada més sofisticats i capaços d'enfrontar-se als diferents problemes que apareixen en aquests entorns. Un gran nombre de factors afegixen soroll i **incertesa** a aquesta tasca, i per tal de dissenyar millors algorismes cal tenir present aquesta incertesa. D'una banda, cal tenir en compte que els sensors tenen la capacitat de percepció limitada, i també que són sensibles a les perturbacions. De l'altra, els actuadors i el sistema de control dels robots no sempre compleixen les consignes de la manera desitjada. Aquestes consignes es calculen basant-se en models

1. Introducció

matemàtics que sovint són simplificacions, ja que només modelen una part de la realitat. I finalment, als robots se'ls demana que actuïn en **temps real**. Això fa que el temps disponible per al còmput estigui limitat i obligui a fer certes aproximacions sacrificant precisió per rapidesa.

Aquest treball versa sobre l'estimació del moviment d'un robot mòbil a partir dels canvis en les imatges captades per una càmera muntada sobre el robot.¹ Fins fa poc, la **visió** no era un dels sensors preferits en aplicacions de robòtica mòbil, sobretot industrial, per darrere de l'odometria, dels sensors d'infraroigs, dels ultrasons, dels làsers i últimament del GPS. Però actualment, gràcies als esforços fets en el camp de la visió artificial, aquesta adquireix més importància i esdevé una eina que cal tenir en compte en robòtica, i en especial per a l'autolocalització i la navegació. Una de les limitacions principals en visió artificial quan es vol recuperar el moviment de la càmera és la inherent pèrdua d'informació en el procés de formació de la imatge: a partir d'informació parcial 2D es vol recuperar el moviment 3D de la càmera. És per això que en general es tendeix a enriquir el sistema de visió amb diverses càmeres, o s'imposen restriccions sobre el moviment o l'escena. És clar que com més informació es té sobre l'entorn, més i millor informació es pot recuperar, però l'enfocament que es vol seguir en aquesta tesi és partir del coneixement del menor nombre possible de paràmetres (per exemple, dels paràmetres de calibratge de la càmera), i donar les claus per aprofitar aquesta informació addicional en el cas que estiguin disponibles.

Quan els efectes perspectius no són importants, dues vistes d'una escena captades pel robot es poden relacionar amb una transformació afí (o afinitat), que normalment es calcula a partir de correspondències de punts. Trobar correctament aquestes correspondències encara és un tema no completament solucionat en visió artificial [90, 126]. En aquesta tesi es vol seguir un enfocament alternatiu, i alhora complementari, fent servir **contorns actius** [17]. Un contorn actiu és una corba que s'inicialitza al voltant de la projecció d'un objecte en la imatge. A mesura que el robot es va movent, la projecció de l'objecte en la imatge va canviant i el contorn actiu es deforma convenientment per adaptar-s'hi. De les deformacions d'aquest contorn, expressades en espai de forma [97], se'n pot extreure el moviment del robot fins a un factor d'escala. Els contorns actius es caracteritzen per la rapidesa en la seva extracció i la seva robustesa en les oclusions

¹Fent un símil biològic, l'objectiu és determinar com s'ha "mogut" analitzant els canvis en el que "veu".

parcials. A més, un contorn és fàcil de trobar fins i tot en escenes poc texturades, on sovint és difícil trobar punts característics i la correspondència que tenen.

El model de càmera emprat és el de **perspectiva dèbil**. Perquè aquest model sigui aplicable cal que es compleixin dues condicions de visualització. Primer, que l'objecte que es mira sigui pla o, si més no, que la profunditat (el que a vegades s'anomena relleu) del contorn de l'objecte sigui petita en comparació amb la distància de la càmera a l'objecte.² Segon, que la projecció del contorn tingui lloc a prop del centre òptic.³ En aquest tipus de visió monocular apareixen típicament diverses ambigüitats. L'algorisme de seguiment que es presenta fa la suposició que el contorn està frontoparal·lel en la primera imatge per discernir l'ambigüitat que hi ha entre profunditat i gir. Com és comú en visió monocular, la translació es recupera fins a un factor d'escala. Aquest factor d'escala depèn tant de la distància inicial entre càmera i objecte com dels paràmetres de calibratge de la càmera.

Hi ha diversos motius per considerar la incertesa en les dades provinents d'un sistema de visió. D'una banda, la resolució limitada de les càmeres i la qualitat del senyal de vídeo provoquen que les dades amb què es treballa siguin imprecises. De l'altra, la qualitat de les imatges també es veu afectada per les condicions d'il·luminació, els reflexos, les distorsions de les lents emprades i, en general, per un nombre elevat de circumstàncies no sempre controlables i/o mesurables [92]. També cal tenir en compte l'error que s'introdueix identificant la posició en la imatge de les característiques que es busquen, en aquest cas el contorn de l'objecte.

Un dels objectius que es persegueix és **caracteritzar la incertesa del moviment** obtingut a partir de les deformacions d'un contorn actiu. Cal conèixer quins factors afecten la recuperació del moviment, la incertesa de cadascuna de les components i trobar les possibles correlacions. Perquè sigui útil en la navegació del robot l'objectiu és trobar un algorisme que permeti avaluar i calcular en temps real la incertesa de cada nou moviment que es computa.

²Utilitzar una càmera de perspectiva dèbil en escenes on hi ha diversos objectes a diferents profunditats generalment no és una bona aproximació. Tanmateix, l'algorisme proposat fixa l'atenció en un sol objecte de l'escena, i per tant només cal que l'objecte en qüestió tingui poc relleu en comparació amb la seva distància a la càmera.

³Habitualment si s'utilitza una càmera amb distància focal gran s'acompleix aquesta restricció.

1. Introducció

Una de les limitacions més importants de l'algorisme de recuperació del moviment basat en el seguiment de contorns actius és que cal que l'objecte estigui visible durant tota la seqüència. Dotant el robot d'un sistema de visió activa es permet que la càmera canviï la seva orientació compensant el moviment del robot. El paradigma de la visió activa, amb tot el que implica respecte a l'atenció visual, combina perfectament amb la proposta de fer servir contorns actius, que de fet són una de les vies per concentrar l'atenció del sistema de visió.

En la mateixa línia, les lents amb *zoom* afegeixen un grau de llibertat extra al sistema de visió, ja que permeten controlar el grau d'obertura de la càmera. Per segons quina aplicació pot ser interessant prendre una imatge més detallada de l'objectiu, o a l'inrevés, una imatge més panoràmica de l'escena. Encara que sembli atractiu afegir un *zoom* a la càmera és un camp no gaire explorat, segurament a causa de la complexitat que afegeix. Primer, perquè quan es fa *zoom* canvia la resolució i, per tant, l'aparença dels objectes, i la correspondència entre imatges és més difícil de solucionar.⁴ Segon, perquè canvien els paràmetres de calibratge de la càmera i, per tant, canvia la correspondència entre les mesures en la imatge i el sistema de coordenades de món. Això és important si es volen fer mesures mètriques. I finalment, perquè canvien els guanyys del sistema de control de l'orientació de la càmera, cosa que en fa més difícil el control.

L'objectiu és fer que l'algorisme de càlcul del moviment sigui més útil encara incorporant una **càmera activa** al robot. El camí seguit passa per proposar un algorisme de control que permeti canviar el *zoom* i orientar la càmera perquè mantingui l'objectiu centrat en la imatge el màxim de temps possible mentre el robot navega lliurement, cosa que amplia considerablement la zona de treball. Cal proposar un nou algorisme que permeti calcular el moviment del robot tenint en compte també el moviment d'orientació de la càmera.

1.2 Compendi de la tesi

Aquesta tesi s'estructura com segueix. Al capítol 2 s'introdueixen els principis sobre els quals es desenvoluparà la resta de la tesi: el seguiment de contorns actius, la pa-

⁴El problema pren importància si es considera que sovint característiques que ressalten basades en punts o cantonades, ho fan només a certes resolucions i no són, per tant, invariants a *zoom*.

rametrització (que anomenarem vector de forma) de la deformació d'aquests contorns en un espai de transformacions del pla de la imatge (que anomenarem espai de forma), i l'algorisme que permet extraure d'aquesta parametrització l'estimació del moviment del robot. Es presenten dos treballs previs realitzats, encaminats a avaluar la viabilitat del seguiment de contorns actius en robòtica mòbil. Primer, en el marc del transport de mercaderies, es presenta una comparació amb un sensor de posicionament làser. La precisió assolida, quan es coneixen els paràmetres de calibratge de la càmera i s'expressa el moviment en coordenades de càmera, és prometedora. Segon, en el marc d'un vehicle de transport de persones, es presenta la col·laboració del sistema de visió amb un sensor inercial. La solució emprada en aquest cas és expressar la informació del sensor inercial en l'espai de deformacions que parametritza el moviment de la càmera.

Un cop analitzada la viabilitat del seguiment de contorns, al capítol 3 es fa un estudi més en profunditat de la geometria de dues vistes i es presenta un nou algorisme que permet calcular la direcció epipolar a partir de dues vistes no calibrades d'un objecte pla. S'observa que la direcció epipolar és un bon indicador de la qualitat del vector de forma calculat i es du a terme una primera aproximació al problema de determinar la precisió de l'algorisme de recuperació del moviment propi fent un bateria d'experiments. En aquests experiments es relaciona la precisió obtinguda en la recuperació de la direcció epipolar amb diferents situacions de relaxació de les condicions de visualització de perspectiva dèbil, la forma del contorn i el soroll d'adquisició. Els experiments reals, amb un braç robòtic Stäubli, permeten validar l'algorisme de recuperació de la direcció epipolar i comparar-lo amb el mètode estàndard basat en correspondències de punts.

Un estudi en profunditat de la precisió de l'algorisme de recuperació del moviment requereix considerar les sis components de la posa. Al capítol 4 es desenvolupen les expressions analítiques per a la propagació de la incertesa suposant que les diferents components són independents. A escala qualitativa els resultats són correctes, però desafortunadament, la incertesa calculada està clarament subestimada ateses les aproximacions que cal fer. Al capítol 5 es mostra un segon enfocament de l'anàlisi de la precisió, aquest cop estadístic, que es veurà que ofereix interessants possibilitats de treball futur. Primer es caracteritza la incertesa amb una simulació de Monte Carlo, que permet observar quines components del moviment es recuperen amb més precisió i les correlacions que hi ha entre les components. Es proposa un algorisme que permet

1. Introducció

fer el càlcul del moviment del robot i de la seva incertesa en temps real, i se'n prova la correctesa amb experiments reals.

Com ja s'ha esmentat, el moviment que es recupera està escalat, a causa d'una sèrie de paràmetres que es consideren desconeguts: calibratge de la càmera i distància inicial. D'aquests, el paràmetre que sembla més difícil d'obtenir és la distància inicial entre la càmera i l'objectiu. Al capítol 6 s'introdueix la càmera amb distància focal variable, i aprofitant el moviment del *zoom* es proposa un nou algorisme, basat en calibratge geomètric, que permet estimar aquesta distància inicial amb un senzill mètode, que no requereix el calibratge dels paràmetres interns de la càmera.

En el conjunt del treball fet s'observa que el repertori de moviments que pot fer el robot és molt limitat: l'objecte ha de trobar-se al davant del robot i els moviments estan restringits gairebé a aproximacions i retrocessos per no perdre'l de vista. Al capítol 7, en primer lloc es proposa un algorisme per controlar el *zoom*. D'aquesta manera es pot mantenir constant la mida de la projecció del contorn en la imatge, i així permetre aproximacions i retrocessos més llargs. Se'n mostra experimentalment la validesa i també es mostra l'algorisme per recuperar el moviment de la càmera. En segon lloc s'explora la possibilitat de canviar l'orientació de la càmera amb un mecanisme de capcineig i guinyada. Això permet al robot fer trajectòries molt més generals, ja que la càmera ara té la capacitat de mantenir l'objecte dins del camp de visió per ella mateixa. D'una banda, es proposa un algorisme que permet tancar els diferents llaços de control del sistema de visió activa, que permeten mantenir l'objecte dins del camp de visió mentre el robot es desplaça. De l'altra, es proposa un nou algorisme que permet recuperar el moviment del robot amb una càmera activa. Els experiments realitzats proven tant la correctesa de l'algorisme de control proposat com la validesa de l'algorisme de càlcul del moviment.

El capítol 8 és el darrer de la tesi i inclou les conclusions, un resum de les aportacions més rellevants portades a terme i una petita discussió sobre el treball futur.

Atès que les aportacions que es fan inclouen diferents àrees de coneixement, la descripció de l'*estat de l'art* no es concentra en un sol capítol, com és habitual, sinó que cada capítol inclou un apartat que exposa la literatura utilitzada sobre cada tema concret.

Chapter 1

Introduction

It is the scientist duty to raise our abilities by increasing our standards for quality and effectiveness, to show what can be done well by pursuing the just possible, and to clarify ruthlessly, independently of the fact that complexity sells better.

Edsger W. Dykstra, april 1986. EDW956-2.

1.1 Motivation

A robot can be defined as a mechanical device with the ability of perceiving and manipulating the environment [133]. More and more, robots have to work in unstructured and dynamic environments, where the sensory capacity becomes crucial to help solving the great number of unforeseen situations that can turn up. In this scenario, robotics is no longer just a problem of mechanics and becomes a science, with the goal of developing more sophisticated algorithms to confront directly the different problems that appear in these environments. A great number of factors add noise and **uncertainty** to this task, and in order to design better algorithms it is necessary to consider this uncertainty. On the one hand, it is necessary to take into account that sensors have limited perception, and also that they are sensitive to external disturbances. On the other hand, the actuators and the robot control system not always fulfil the commands in the expected way. These commands are calculated based on mathematical models that often are simplifications. And, finally, robots are expected to act in **real-time**. This implies that time for the calculation is limited, which forces to make certain approximations sacrificing accuracy for speed.

1. Introduction

This work deals with the estimation of the motion of a mobile robot by observing the changes in the images acquired by a camera mounted on the own robot¹. Until little ago, **vision** was not one of the favourite sensors in applications of mobile robotics, especially in industrial robotics, behind odometry, infrared sensors, ultrasound, laser, and lately GPS. But at present, thanks to the efforts made in the field of artificial vision, it has acquired more importance and has become a tool to take into account in robotics, and specifically in self-localization and navigation problems. One of the main limitations of artificial vision, when the goal is to compute the motion of the camera, is the inherent loss of information in the process of image formation: from 2D partial information we want to retrieve the complete 3D camera motion. For this reason, the general trend is to endow the vision system with several cameras, or to impose restrictions on the motion or the scene. Of course, as more information about the environment is available, more and better motion information can be recovered. However, the path followed in this thesis is to start from the knowledge of the least possible number of parameters (for example, of the calibration parameters of the camera), and to give the clues to be able to use this additional information in the case that it becomes available.

When perspective effects are not significant, two views of a scene can be related by an affine transformation (or affinity), which is usually calculated from point correspondences. Finding correctly these correspondences is a subject still not completely solved in artificial vision [90, 126]. In this thesis we want to evaluate an alternative, and at the same time complementary, technique that uses **active contours** [17]. An active contour is a curve that is initialized around the projection of an object in the image. When the robot moves freely, the projection of the object in the image changes and the active contour deforms itself conveniently to adapt to it. From the deformations of this contour, expressed in shape space form [97], the robot motion can be extracted up to a scale factor. Active contours are very easy to track, and are also robust to partial occlusions. Moreover, a contour is easy to find even in poorly textured scenes, where characteristic points and their correspondences are difficult to find.

The camera model used is the **weak-perspective** one. Two visualization conditions should be fulfilled to be able to use this model. First, the target object should be planar or, at least, the depth (sometimes called relief) of the object contour should be small in

¹Making a biological comparison, the goal is to determine how it has “moved” analyzing the changes in what it “sees”.

comparison with the distance from the camera to the object². Second, the projection of the contour should lie near the optical center³. In this type of monocular vision typically several ambiguities appear. The tracking algorithm presented makes the assumption that the object is frontoparallel in the first image to discern the existing ambiguity between depth and turning. As usual in monocular vision, translations are recovered up to a scale factor. This scale factor depends on the initial distance between the camera and the object as well as on the camera calibration parameters.

There are several reasons for considering the uncertainty in the data coming from a vision system. On the one hand, the limited resolution of the cameras and the quality of the video signal lead to lack of precision. On the other hand, the quality of the images is also affected by the lighting conditions, the reflections, the lens distortions, and in general by a high number of circumstances not always controllable and/or measurable [92]. Finally, the errors produced when obtaining the position of the landmarks in the image, in our case the position of the object contour, should be taken into account.

One of the thesis objectives is to **characterize the uncertainty** of the motion obtained from the deformations of an active contour. It is necessary to determine which factors influence motion recovery, to estimate the uncertainty of each motion component, and to find the possible correlations between them. To be useful for robot navigation, the algorithm developed should allow to evaluate and to calculate the uncertainty of each new computed motion in real time.

One of the most severe limitations of the motion recovery algorithm based on the tracking of active contours is that the object should keep visible during all the motion sequence. Providing the robot with an active vision system dissociates robot and camera motion, allowing the camera to change its orientation, thus offsetting the robot motion. The active vision paradigm, with all its implications with respect to visual attention, combines perfectly with the proposal of using active contours, which in fact is one of the ways to concentrate the attention of the vision system.

²Using weak-perspective camera models in scenes where there are several objects at different depths is generally not a good approach. However, the proposed algorithm fixes the attention on an object of the scene alone, and therefore it is only necessary that this object has little depth relief in comparison with its distance to the camera.

³This restriction is usually satisfied using a camera with a large focal distance.

1. Introduction

Along the same line, zoom lenses add an extra degree of freedom to the vision system, allowing to control the field of view of the camera. For some applications it can be interesting to take more detailed images of the target, or conversely, a more panoramic image of the scene. Even if it seems attractive to add a zoom to the camera, it is a field hardly explored, probably because of the added complexity. First, because zoom changes the resolution of the image and, therefore, the appearance of the objects, making the correspondence between images more difficult to solve⁴. Second, because the camera calibration parameters change, and accordingly, the correspondence between measures in the image and world coordinates also changes. This is important if metric measures are expected. And, finally, because zoom changes the gains of the orientation control system, which becomes more difficult to control.

A final goal is to increase the applicability of the motion recovery algorithm by incorporating an **active camera** to the robot. The path followed is to propose a control algorithm to change the zoom and to turn the camera to keep the target projection centered on the image if possible while the robot moves freely, broadening considerably the robot workspace. It becomes necessary to propose a new algorithm to estimate robot motion taking into account also the change in orientation of the camera.

1.2 Summary of the thesis

This thesis is structured as follows. Chapter 2 introduces the principles used to develop the rest of the thesis: the tracking of active contours, the parametrization (called shape vector) of the deformation of this contour in a space of transformations of the image plane (called shape space), and the algorithm to extract the motion estimation from this parametrization. Two introductory works are also presented, aimed at estimating the feasibility of the tracking of active contours in mobile robotics. First, within the framework of goods transportation, a comparison with a positioning laser sensor is presented. The accuracy attained, when the parameters of calibration of the camera are known and the motion is expressed in camera coordinates, is promising. Second, within the framework of people transportation, the collaboration of the vision system with an inertial sensor is presented. The solution adopted in this case is to express

⁴The problem gains importance if one considers that often salient characteristics based on points or corners are salient only for certain resolutions and they are not, therefore, zoom invariant.

the inertial information in the space of deformations that parameterizes the camera motion.

Once the feasibility of egomotion recovery from contour tracking has been established, in Chapter 3 two-view geometry is revisited, and a new algorithm to calculate the epipolar direction from two uncalibrated views of a planar object is presented. It is also noticed that the accuracy of the epipolar direction is a good measure of the quality of the shape vector. A first approach to the problem of determining the accuracy of the motion recovery algorithm is conducted by making a series of experiments. In these experiments, the accuracy obtained in the recovery of the epipolar direction is related to different situations of relaxation of the weak-perspective visualization conditions, the form of the contour and the acquisition noise. The real image experiments, using a Stäubli robotic arm, allow us to validate the algorithm of recovery of the epipolar direction and to compare it with the standard method based on point correspondences.

An in-depth study of the accuracy of the algorithm requires to consider the complete set of 6 pose components. Chapter 4 presents the analytical expressions derived for the propagation of the uncertainty, assuming independence between the different components. Results are correct at a qualitative level, but unfortunately the calculated uncertainty is clearly underestimated due to the approximations which it is necessary to make. In Chapter 5 a statistical approach to the analysis of the accuracy is presented, which will be seen to offer interesting future work possibilities. First, the uncertainty is characterized through a Monte Carlo simulation, that allows to observe which motion components are recovered with more or less accuracy, as well as the correlations between them. An algorithm to calculate robot motion and its uncertainty in real time is also proposed, and its correctness is assessed through real-world experiments.

As already mentioned, the recovered motion, in fact the translation components, is scaled due to the fact that some parameters are considered unknown: calibration parameters of the camera and the initial distance. Of these, the parameter that seems more difficult to obtain is the initial distance between the camera and the target. The camera with variable focal distance is introduced in Chapter 6. Using zoom motions, a new algorithm, based on geometric calibration, is proposed to estimate this initial distance with a simple method, which does not require the calibration of the camera internal parameters.

1. Introduction

In the work carried until now, the repertoire of motions that the robot can make while keeping the object into the field of view is very limited: the object has to be in front of the robot and motion is restricted almost to approaches and recedings. Chapter 7 presents, in the first place, an algorithm to control the zoom and recover the motion. The objective is to keep the size of the projection of the contour constant in the image, and allow longer approach and receding motions. Its validity is shown experimentally. In the second place, the possibility of changing the orientation of the camera with a pan-and-tilt mechanism is explored. This allows the robot to make more general trajectories, since the camera now has itself the capacity to maintain the object in the field of view. On the one hand, an algorithm to close the different control loops of the active vision system is presented, which allows to maintain the object in the field of view while the robot travels. On the other hand, a new algorithm is proposed to recover the robot motion by using an active camera. The experiments carried out show the correctness of the proposed control algorithm as well as the good performance of the motion recovery algorithm.

Chapter 8 is the last one and includes the conclusions, a summary of the most relevant contributions of this thesis, and a small discussion about future work.

Since the contributions made touch upon different areas of knowledge, the state of the art is not concentrated in one chapter, as usual, but each chapter includes a state-of-the-art section.

Capítol 2

Punt de partida: vector de forma i posa 3D

“... et, pour la première fois, je voisai les sons que j’avais en tête: «Maman!» Extase de la mère. Et comme il ne fallait vexer personne, je me hâtaï d’ajouter: «Papa!». Attendrissement du père. [...] Je pensai qu’ils n’étaient pas difficiles. [...] À croire qu’ils avaient un doute sur leur propre identité: n’étaient-ils donc pas sûrs de s’appeler Papa et Maman? Ils semblaient avoir eu tant besoin que je le leur confirme.”

Amélie Nothomb, *Metaphysique des tubes*.

“...i per primera vegada vaig articular els sons que duia al cap: «mama!» Èxtasi de la mare. I com que no calia humiliar a ningú, vaig afanyar-me a afegir de seguida: «papa!». Entendrimient del pare. [...] Vaig pensar que no eren gent difícil. [...] Qualsevol diria que tenien un dubte sobre la seva identitat: que no estaven segurs de dir-se papa i mama? Semblava ben bé que els feia molta falta que jo els ho confirmés.”

Amélie Nothomb, *La metafísica dels tubs*.

Extracte

Per deduir el moviment d’una càmera a partir de les imatges captades, una alternativa interessant a buscar les correspondències de punts és fer el seguiment de contorns actius. Les deformacions d’un contorn actiu en una seqüència d’imatges

2. Punt de partida: vector de forma i posa 3D

provocada per una càmera en moviment es pot codificar en un espai de transformacions afins, anomenat espai de forma, amb un vector de forma. D'aquest vector de forma se'n pot extreure el moviment 3D de la càmera. En primer lloc, s'estudia si és factible usar aquesta aproximació en situacions reals i s'explora la col·laboració del sistema de visió amb un algorisme de posicionament làser, comparant-ne la precisió. Els resultats són prometedors tot i que és necessari conèixer més a fons quins factors determinen la precisió. En segon lloc, s'enriqueix l'estratègia de cerca del contorn actiu amb la informació d'un sensor inercial i es mostra que són complementaris.

2.1 Introducció

Les fonts tradicionals d'informació per als algorismes de visió artificial han estat punts, línies, cantonades, característiques locals més refinades com ara SIFT [90], que sembla que actualment representen l'estat de l'art en la definició de característiques, o una combinació de totes elles per tal d'identificar objectes més complexos a l'escena [121]. Alternativament, el que es vol explorar en aquest treball és la possibilitat d'utilitzar també contorns actius com a trets informatius, especialment en aplicacions que no exigeixen que la precisió sigui molt alta, com pot ser el cas d'algunes aplicacions de navegació de robots.

Durant les dues dècades passades, s'han proposat diversos algorismes per resoldre el problema del posicionament de robots. Es poden classificar en dos grups [19]: posicionament absolut i posicionament relatiu. Els mètodes de posicionament absolut calculen la posició del robot i l'orientació en l'espai de treball detectant alguna fita coneguda en l'entorn. Es poden distingir dos subgrups més, depenent de si utilitzen fites naturals [16, 127] o artificials [70, 119]. Els algorismes basats en fites naturals exploten trets característics ja presents en l'ambient. Pel contrari, les fites artificials es posen en llocs coneguts en l'entorn de treball amb l'únic propòsit de facilitar la navegació del robot. Normalment això és car, tant en termes del cost de preparar l'entorn i mesurar la posició de les fites com en termes del sensor, ja que habitualment s'utilitzen sensors d'alta resolució.

Els mètodes de posicionament relatiu, d'altra banda, computen la posició del robot i l'orientació a partir d'una configuració inicial i, consegüentment, sovint es classifiquen com a mètodes d'estimació del moviment. Encara es pot fer una segona distinció,

segons se segueixi un enfocament incremental o no incremental. Entre els primers, hi ha els mètodes basats en odometria i els basats en la integració d'informació inercial. El principal defecte que pateixen són els errors que van acumulant, que fan que per a desplaçaments grans l'error en el posicionament que produeixen sigui massa elevat per poder ser útils.

Es presenta un mètode d'estimació del moviment que depèn de fites naturals. No és incremental i, per això, no pateix del desavantatge de l'error acumulatiu. Utilitza les imatges proporcionades per una sola càmera. És ben sabut en visió monocular que si no es disposa d'informació complementària les translacions només es poden recuperar fins a un factor d'escala. L'algorisme assumeix que les condicions de visualització es poden modelar amb un model de càmera de perspectiva dèbil. Les condicions de visualització en aquest model són, primer, que els punts de l'objectiu són prop del raig de projecció (es pot complir fàcilment amb una càmera amb un camp de visió petit), i segon, que la variació de profunditat de l'objectiu (el relleu) és petit comparat amb la seva distància a la càmera. Aquest model de càmeres ha estat àmpliament utilitzat abans [21, 81, 125] i fins i tot es prefereix quan els efectes perspectius no són clars [136].

Els contorns actius són una eina usual per a la segmentació en l'anàlisi d'imatges mèdiques. L'habilitat per seguir ràpidament contorns actius parametritzats com a B-Splines va ser desenvolupada per Blake i Isard [17] en el marc de l'aprenentatge de la dinàmica i dels contorns deformables. Considerant objectes no deformables, Martínez [97] va mostrar que els contorns podrien ser adequats per recuperar qualitativament el moviment propi d'una càmera, aplicat a un robot amb potes [98]. En aquests treballs, la inicialització de la B-Spline la fa manualment un operador.

L'algorisme emprat per recuperar el moviment té els quatre passos següents. Primer, el contorn actiu s'ha d'inicialitzar. Una manera comuna de representar contorns actius és utilitzant B-Splines [18], on la corba queda definida bàsicament amb una llista de punts de control. La inicialització del contorn \mathbf{Q}_0 la fa manualment un operador, o bé es fa amb tècniques d'inicialització automàtica, com la proposada per Cham [26], i que un cop provada dona bons resultats. Ja que s'assumeixen condicions de visualització de perspectiva dèbil, l'algorisme de seguiment només ha de permetre deformacions afins del contorn inicial i, per això, el procés d'inicialització és important, ja que determina la família de deformacions afins \mathbf{W} a les quals el contorn podrà adaptar-se.

2. Punt de partida: vector de forma i posa 3D

Segon, un algorisme [17, 97] basat en el filtre de Kalman mesura la nova posició del contorn en la imatge \mathbf{Q} i computa el vector de forma \mathbf{S} que millor explica les mesures obtingudes. La mida del vector de forma és constant, independentment del nombre de punts de control que s'utilitzen per parametritzar la B-Spline, i de la complexitat que tingui.

Tercer, a partir del vector de forma \mathbf{S} es computa el moviment 3D de la càmera. S'ha demostrat [97] que és possible obtenir la informació de rotació i translació utilitzant tècniques de descomposició de SVD per a diferents models de càmeres afins, situant el sistema de referència en diferents llocs i amb diferents graus de coneixement del calibratge de la càmera, fins i tot amb càmeres amb *zoom* [99].

A la secció 2.2 es presenten els conceptes bàsics que més endavant es faran servir. Això inclou la introducció del model de perspectiva dèbil, una introducció a les deformacions afins que codifiquen les deformacions dels contorns actius en espai de forma i l'algorisme per obtenir el moviment 3D a partir de la codificació en l'espai de forma, inclòs el cas d'una càmera amb *zoom* on la distància focal és variable. Seguidament es presenten dos treballs realitzats. El primer vol obtenir una primera aproximació a l'error que es comet, per a la qual cosa compara el sistema de visió amb un sistema de posicionament làser (secció 2.3.1) en el marc d'un experiment real amb un robot paletitzador. El segon (secció 2.3.2) presenta la col·laboració amb un sensor inercial. Per tal d'expressar la informació inercial en espai de forma cal derivar les transformacions inverses que porten d'espai 3D a espai de posa. Un experiment amb un robot de transport de persones il·lustra els resultats obtinguts.

2.2 Antecedents

2.2.1 Model de càmera

Un model de càmera de perspectiva dèbil (o ortogràfic escalat [8]) ψ representa la projecció d'un punt de l'escena al pla de la imatge de la manera següent. Primer projecta el punt de l'escena ortogràficament al pla de profunditat mitjana \mathcal{R}^C (pla paral·lel al pla de la imatge \mathcal{R} que conté el centroid C de l'objecte de l'escena), i seguidament projecta perspectivament des d'aquest pla frontoparal·lel \mathcal{R}^C en la imatge \mathcal{R} (figura 2.1). Així, la segona projecció perspectiva (de centre òptic propi P) simplement introdueix

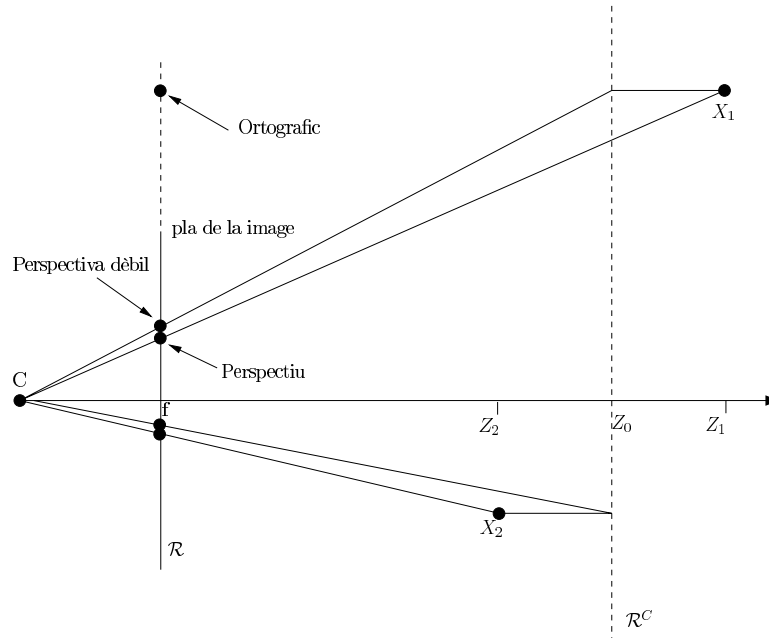


Figura 2.1: Projecció afí comparada amb la projecció de perspectiva completa i el model de càmera ortogràfica. Les projeccions afins (i en particular una projecció de perspectiva dèbil) donen una millor aproximació a la projecció de perspectiva que l'altre model simplificat, la càmera ortogràfica. En el model de càmera afí, es fa una projecció paral·lela al pla de profunditat mitjana abans de fer una projecció de perspectiva global (escalat). Per exemple, en el model paraperspectiu, aquesta projecció és en la direcció paral·lela a la direcció de projecció perspectiva del centre de l'objectiu. Tanmateix, en el model de perspectiva dèbil, la projecció paral·lela al pla de profunditat mitjana es fa en direcció perpendicular al pla de la imatge. Sota la condició que l'objectiu es vegi amb un camp de visió petit, aquestes dues direccions fixes de projecció paral·lela donen una bona aproximació per a tots els punts de l'objectiu, comparat amb el model de càmera ortogràfica.

un factor d'escala. Mòdul aquest factor d'escala, la càmera de perspectiva dèbil és una càmera ortogràfica: tots els raigs de projecció són paral·lels a la direcció ortogonal al pla de la imatge \mathcal{R} .

Es pren ψ com a sistema de referència global associat a la càmera de perspectiva dèbil, i s'escull de la manera següent: es pren un sistema de referència qualsevol al pla de la imatge \mathcal{R} i s'hi suma un vector ortogonal a \mathcal{R} per obtenir un sistema de referència afí global. Sota aquest sistema de referència, les coordenades x i y de la imatge coincideixen amb les coordenades globals X i Y , respectivament, i la coordenada Z coincideix amb la direcció de visualització, de manera que la projecció de perspectiva dèbil està donada

2. Punt de partida: vector de forma i posa 3D

per

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z_0} \begin{bmatrix} X \\ Y \end{bmatrix}, \quad (2.1)$$

on f és la distància focal i Z_0 és la distància mitjana de l'objectiu a la càmera (és a dir, el pla \mathcal{R}^C té l'equació $Z = Z_0$). Quan la variació de profunditat de l'objectiu és petita, comparada amb Z_0 , i el punt principal és a prop del centroide de la projecció de l'objectiu, llavors el model de càmera de perspectiva dèbil dóna una aproximació de la projecció perspectiva [67].

Suposem que l'objectiu està immòbil a l'escena i que la càmera es trasllada segons \mathbf{T} i gira segons \mathbf{R} al voltant de l'objectiu, i possiblement fa un *zoom*,¹ cosa que dóna una segona càmera de perspectiva dèbil ψ' . El nou sistema de coordenades afí associat amb la segona càmera està definit per les files de \mathbf{R} i el nou origen a $-\mathbf{R}^\top \mathbf{T}$, de manera que ψ' respon a l'expressió

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{f'}{Z'_0} \begin{bmatrix} X' \\ Y' \end{bmatrix}, \quad (2.2)$$

on $[X, Y, Z]^\top = \mathbf{R}[X', Y', Z']^\top + \mathbf{T}$, f' és la nova distància focal, i Z'_0 és la distància mitjana entre l'objectiu i la segona càmera.

Les dues vistes de l'escena coplanària es poden explicar per l'afinitat donada per [99]

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \mathbf{M} \begin{bmatrix} x \\ y \end{bmatrix} + \mathbf{t}, \quad (2.3)$$

on

$$\mathbf{M} = \frac{f'}{f} \frac{Z_0}{Z_0 + T_z} \begin{bmatrix} R_{11} & R_{21} \\ R_{21} & R_{22} \end{bmatrix}, \quad (2.4)$$

$$\mathbf{t} = \frac{f'}{Z_0 + T_z} \begin{bmatrix} T_x \\ T_y \end{bmatrix} + \begin{bmatrix} u - u' \\ v - v' \end{bmatrix}, \quad (2.5)$$

on R_{ij} són elements de la matriu de rotació 3D \mathbf{R} , T_i són els elements del vector de translació 3D \mathbf{T} , Z_0 és la distància inicial de la càmera a l'objecte, f és la distància focal de la càmera a la imatge inicial, f' és la distància focal en la segona imatge, (u_0, v_0) és el punt principal en la imatge inicial i (u', v') és la seva posició en la segona imatge. Cal fer notar que quan es canvia la distància focal la posició del punt principal també canvia [83], i això queda clarament expressat a l'equació. Una de les tècniques més

¹L'efecte més evident d'una càmera que fa *zoom* és que la distància focal i el punt principal canvi en [84].

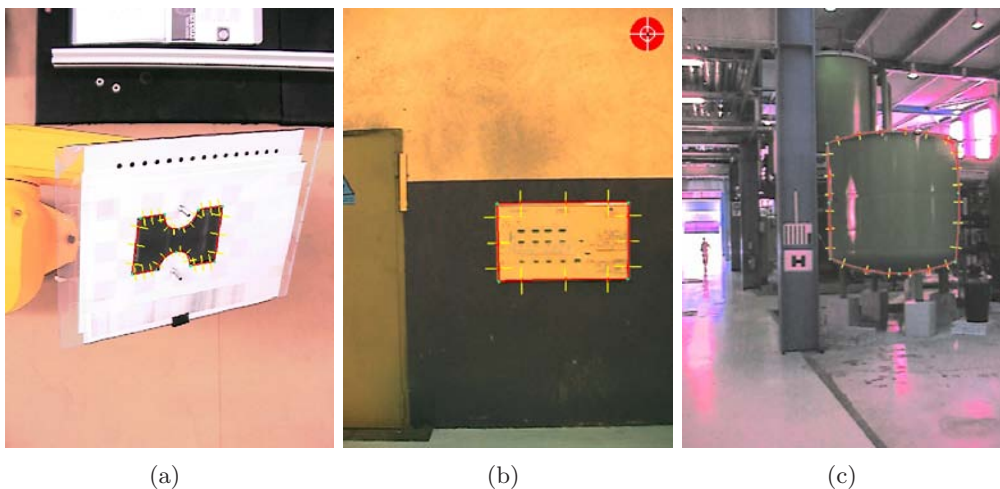


Figura 2.2: Imatges reals amb un contorn actiu inicialitzat.

comunes quan els canvis en la posició del punt principal són petits² és la de considerar-ne la mitjana [87] calculada amb posicions diferents del *zoom*.

És fàcil veure que quan la distància focal no canvia, per exemple quan es disposa d'una càmera amb lents fixes com les utilitzades en les seccions 2.3.1 i 2.3.2, les equacions (2.4) i (2.5) deriven a les presentades a [17, 97].

2.2.2 L'afinitat que relaciona dues vistes

Un contorn actiu és una corba que es defineix al pla de la imatge i que s'ajusta a la projecció del contorn d'un objecte que es fa servir de característica rellevant a la imatge (figura 2.2). Quan la càmera es mou, la projecció del contorn en la imatge canvia, i el contorn actiu es deforma per ajustar-se a la nova projecció. Aquesta deformació es pot codificar amb un vector de forma, tal com s'explicarà més endavant. El contorn es representa com una Spline,³ tal com és fa comunament en gràfics per computador [55], $\mathbf{d}(s) = (d_x(s), d_y(s))^\top$, on ambdues $d_x(s)$ i $d_y(s)$ són corbes B-Spline. Es poden escriure en funció dels seus punts de control \mathbf{Q} ,

$$\mathbf{d}(s) = \begin{bmatrix} d_x(s) \\ d_y(s) \end{bmatrix} = \begin{bmatrix} \mathbf{B}(s)\mathbf{Q}^x \\ \mathbf{B}(s)\mathbf{Q}^y \end{bmatrix} = \begin{bmatrix} \mathbf{B}(s) & \mathbf{0}^\top \\ \mathbf{0}^\top & \mathbf{B}(s) \end{bmatrix} \begin{bmatrix} \mathbf{Q}^x \\ \mathbf{Q}^y \end{bmatrix} = \mathbf{U}(s)\mathbf{Q}, \quad (2.6)$$

²Això depèn de la fabricació de les lents.

³Funció polinòmica a trossos que interpola una sèrie de nodes, de manera que cada tros és un polinomi de grau n i en cada node els dos polinomis concurrents tenen totes les seves derivades iguals fins a l'ordre $n-1$.

2. Punt de partida: vector de forma i posa 3D

on \mathbf{Q}^x i \mathbf{Q}^y són els vectors columna de les components x i y dels punts de control, $\mathbf{0}$ és un vector columna de zeros, i $\mathbf{U}(s) = \mathbf{I} \otimes \mathbf{B}(s)$ és el producte de Kronecker entre la matriu identitat \mathbf{I} i el vector fila $\mathbf{B}(s)$ de la funció de base de la B-Spline [17, 55].

Quan el contorn s'observa des de dos punts de vista diferents, l'equació (2.3) es pot generalitzar per descriure'n la relació com

$$\mathbf{d}'(s) = \mathbf{M}\mathbf{d}(s) + \mathbf{t}, \quad (2.7)$$

on $\mathbf{d}'(s)$ i $\mathbf{d}(s)$ són el contorn en les dues vistes. Combinant l'expressió (2.6) amb l'equació (2.7), s'obté

$$\mathbf{d}'(s) = \mathbf{M}\mathbf{U}(s)\mathbf{Q} + \mathbf{t}. \quad (2.8)$$

Ara, definint $\mathbf{1}$ com un vector columna d'uns, es pot observar que $\mathbf{B}(s)\mathbf{1} = 1$ per la propietat de l'envolvent convexa de les corbes B-Spline, i altre cop usant l'equació (2.6), la diferència entre $\mathbf{d}'(s)$ i $\mathbf{d}(s)$ es pot escriure com

$$\begin{aligned} \mathbf{d}'(s) - \mathbf{d}(s) &= t_x \mathbf{U}(s) \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} + t_y \mathbf{U}(s) \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} + (M_{11} - 1) \mathbf{U}(s) \begin{bmatrix} \mathbf{Q}^x \\ \mathbf{0} \end{bmatrix} \\ &+ M_{12} \mathbf{U}(s) \begin{bmatrix} \mathbf{Q}^y \\ \mathbf{0} \end{bmatrix} + M_{21} \mathbf{U}(s) \begin{bmatrix} \mathbf{0} \\ \mathbf{Q}^x \end{bmatrix} + (M_{22} - 1) \mathbf{U}(s) \begin{bmatrix} \mathbf{0} \\ \mathbf{Q}^y \end{bmatrix}. \end{aligned} \quad (2.9)$$

D'altra banda, la diferència entre $\mathbf{d}'(s)$ i $\mathbf{d}(s)$ és, a partir de l'equació (2.6),

$$\mathbf{d}'(s) - \mathbf{d}(s) = \mathbf{U}(s)(\mathbf{Q}' - \mathbf{Q}). \quad (2.10)$$

Per tant, comparant aquest resultat amb l'expressió (2.9), es pot concloure que la diferència en termes de punts de control $\mathbf{Q}' - \mathbf{Q}$ es pot escriure com la combinació lineal de sis vectors. Fent servir notació matricial, es pot escriure

$$\mathbf{Q}' - \mathbf{Q} = \mathbf{W}\mathbf{S}, \quad (2.11)$$

on

$$\mathbf{W} = \left(\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \begin{bmatrix} \mathbf{Q}^x \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{Q}^y \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{Q}^x \end{bmatrix}, \begin{bmatrix} \mathbf{Q}^y \\ \mathbf{0} \end{bmatrix} \right) \quad (2.12)$$

és la *matriu de forma* i \mathbf{S} és un vector amb els sis paràmetres de la combinació lineal, el *vector de forma*

$$\mathbf{S} = (t_x, t_y, M_{11} - 1, M_{22} - 1, M_{21}, M_{12})^\top, \quad (2.13)$$

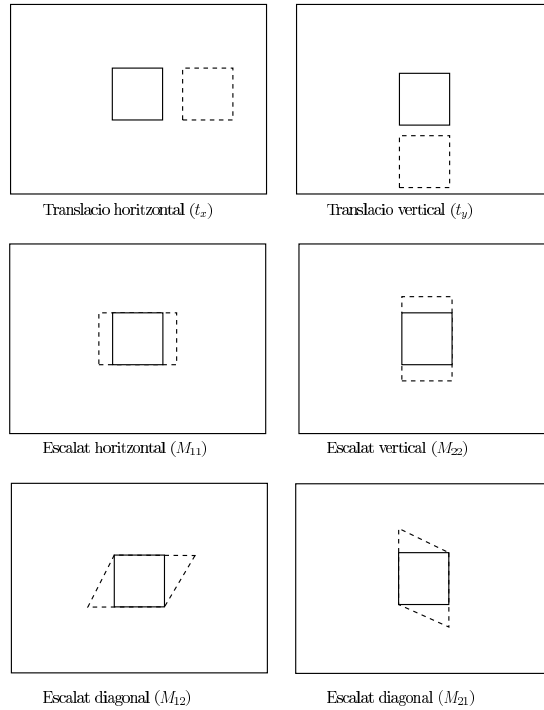


Figura 2.3: Representació gràfica del significat de cada grau de llibertat del vector de forma induït pel moviment de la càmera.

que codifica la relació entre dues vistes afins diferents del contorn planar.

La figura 2.3 (reproduïda de [97]) mostra gràficament l'efecte de cada paràmetre: dues translacions i quatre escalats.

Un cop s'ha obtingut la representació compacta en termes de punts de control, es fa servir un filtre de Kalman per fer el seguiment del contorn durant tota la seqüència d'imatges [17] i s'obté un nou vector de forma per a cada nova imatge.

A tall d'exemple, s'ha dut a terme un petit experiment per tal de veure quina forma tenen els diferents vectors de forma que codifiquen diferents moviments de la càmera. A la figura 2.4 es pot veure el vector de forma corresponent a un desplaçament del robot cap a un objectiu mentre es pertorba lleugerament la direcció. A la figura 2.4(a) es pot observar com la pertorbació de la direcció que correspon a translacions T_x i T_y del robot es codifica en les dues primeres components del vector de forma. La translació T_z , que correspon al moviment d'aproximació que fa el robot cap a l'objectiu, queda codificada en la part vectorial del vector de forma (figura 2.4(b)) i no correspon a una

2. Punt de partida: vector de forma i posa 3D

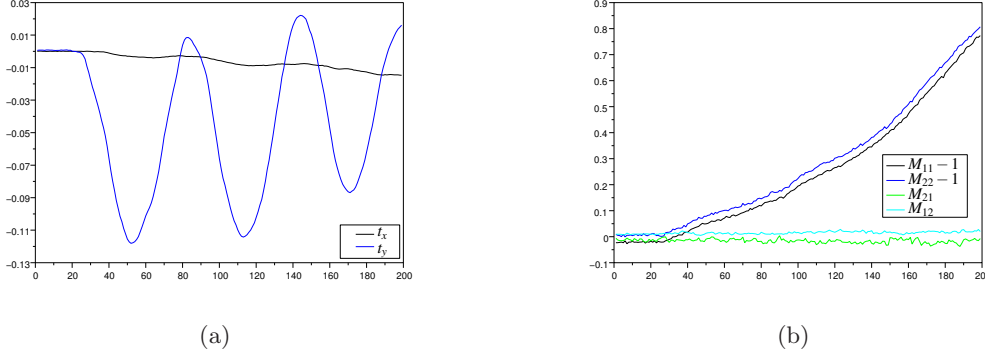


Figura 2.4: Representació dels vectors de forma extrets en una seqüència real d'un moviment d'un robot.

sola component, sinó a les components M_{11} i M_{22} a la vegada. Observant de nou la figura 2.3 es pot veure que aquestes components corresponen a l'escalat horitzontal i vertical. Sembla lògic que un moviment d'aproximació, que fa créixer la mida de la projecció de l'objecte en la imatge, correspongui a increments iguals en les components que codifiquen l'escalat. Aquests vectors de forma corresponen a un dels experiments que es mostraran més endavant (secció 2.3.1).

2.2.3 Moviment 3D a partir de l'afinitat

Seguidament es presenta com es pot obtenir la rotació i la translació 3D codificades per l'afinitat definida per $\mathbf{M} = [M_{i,j}]$ i $\mathbf{t} = (t_x, t_y)$ [99]. Si es representa la matriu de rotació amb angles d'Euler ZXXZ,

$$\mathbf{R} = \mathbf{R}_z(\phi)\mathbf{R}_x(\theta)\mathbf{R}_z(\psi), \quad (2.14)$$

l'equació (2.4) es pot reescriure com

$$\begin{aligned} \mathbf{M} &= \frac{f'}{f} \frac{Z_0}{Z_0 + T_z} \mathbf{R}_z|_2(\phi)\mathbf{R}_x|_2(\theta)\mathbf{R}_z|_2(\psi) = \\ &= \frac{f'}{f} \frac{Z_0}{Z_0 + T_z} \mathbf{R}_z|_2(\phi) \begin{bmatrix} 1 & 0 \\ 0 & \cos\theta \end{bmatrix} \mathbf{R}_z|_2(\psi) \end{aligned}$$

on $\mathbf{R}|_2$ denota la submatriu 2×2 de \mathbf{R} . Llavors,

$$\mathbf{M}\mathbf{M}^T = \mathbf{R}_z|_2(\phi) \begin{bmatrix} L & 0 \\ 0 & L\cos^2\theta \end{bmatrix} \mathbf{R}_z|_2^{-1}(\phi) \quad (2.15)$$

on

$$L = \left(\frac{f'}{f} \frac{Z_0}{Z_0 + T_z} \right)^2.$$

Aquesta darrera equació mostra que θ es pot calcular a partir dels vectors propis de la matriu \mathbf{MM}^T ,⁴ que s'anomenaran (λ_1, λ_2) :

$$\cos\theta = \sqrt{\frac{\lambda_2}{\lambda_1}}, \quad (2.16)$$

on λ_1 és el valor propi més gran. L'angle ϕ es pot deduir dels vectors propis de \mathbf{MM}^T ; el vector propi \mathbf{v}_1 que correspon al valor propi més gran correspon a la primera columna de $\mathbf{R}_z|_2(\phi)$:

$$\mathbf{v}_1 = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}. \quad (2.17)$$

Aïllant $\mathbf{R}_z|_2(\psi)$ de l'equació (2.15),

$$\mathbf{R}_z|_2(\psi) = \frac{f'}{f} \left(1 + \frac{T_z}{Z_0} \right) \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\cos\theta} \end{bmatrix} \mathbf{R}_z|_2(-\phi)\mathbf{M}, \quad (2.18)$$

i observant, a l'equació (2.15), que

$$\frac{f}{f'} \left(1 + \frac{T_z}{Z_0} \right) = \frac{1}{\sqrt{\lambda_1}},$$

es pot trobar, primer, $\sin\psi$ i, després, ψ .

Un cop trobats els angles θ, ϕ i ψ es pot reconstruir la matriu de rotació \mathbf{R} aplicant l'equació (2.14).

La translació escalada⁵ en la direcció Z es pot calcular com

$$\frac{T_z}{Z_0} = \frac{f'}{f} \frac{1}{\sqrt{\lambda_1}} - 1. \quad (2.19)$$

La resta de components de la translació 3D es poden calcular a partir de \mathbf{t} fent servir l'equació (2.5):

$$\frac{T_x}{Z_0} = \frac{t_x - (u' - u)}{f\sqrt{\lambda_1}}, \quad (2.20)$$

$$\frac{T_y}{Z_0} = \frac{t_y - (v' - v)}{f\sqrt{\lambda_1}}. \quad (2.21)$$

⁴Es pot aconseguir una derivació semblant si s'expressa la rotació amb angles d'Euler ZYZ, però desafortunadament no es pot seguir el mateix guió amb altres expressions de la rotació, com per exemple XYZ.

⁵Com és habitual en un sistema monocular, si no es disposa d'informació addicional la translació només es pot recuperar fins a un factor d'escala. Més endavant, al capítol 6, es mostra un algorisme per mesurar la distància inicial de la càmera a l'objecte i eliminar aquest factor d'escala.

2. Punt de partida: vector de forma i posa 3D

Les equacions presentades permeten calcular una estimació del moviment 3D de la càmera a partir de l'afinitat que el parametriza, fins i tot quan la càmera varia els seus paràmetres interns. Aquestes mateixes expressions es poden derivar quan el sistema de coordenades està centrat en la càmera [97]. En aquest cas, les expressions per trobar les rotacions no varien i les translacions es poden calcular amb:

$$\begin{aligned}\frac{T_x}{Z_0} &= \frac{t_x - (u' - u)}{f\sqrt{\lambda_1}} - R_{13}, \\ \frac{T_y}{Z_0} &= \frac{t_y - (v' - v)}{f\sqrt{\lambda_1}} - R_{23}, \\ \frac{T_z}{Z_0} &= \frac{f'}{f} \frac{1}{\sqrt{\lambda_1}} - R_{33}.\end{aligned}$$

Es pot veure fàcilment que, quan la càmera no mou el *zoom*, i per tant la seva distància focal no varia, llavors $f' = f$. En aquest cas, el punt principal tampoc varia, i si es considera $u' = u$ i $v' = v$ s'arriba a les expressions ja conegudes [97] per calcular el moviment per a una càmera passiva.

2.2.4 Detalls de la implementació

L'objectiu de l'algorisme de seguiment és mantenir la referència d'un contorn objectiu al llarg d'una seqüència d'imatges i estimar el vector de forma que codifica la deformació del contorn. La implementació de l'algorisme s'ha fet amb un filtre de Kalman. A causa de la seva representació com a B-Spline, el contorn es divideix naturalment en seccions, cadascuna entre dos nodes consecutius. Per al seguiment, es defineixen alguns punts d'interès de manera equidistant al llarg de cada secció de contorn. Passant a través de cada punt i normal al contorn, es defineix un segment de recta. La recerca del contorn es fa només per als píxels sota aquests segments normals, i el resultat és el pas de mesura de l'algorisme del filtre de Kalman. Això provoca que el sistema sigui molt ràpid, ja que el processament d'imatge és local, cosa que evita l'ús d'algorismes costosos de segmentació. Una vegada que els elements de la frontera estan situats al llarg de tots els segments de recerca, el filtre Kalman calcula el vector de forma que millor explica la deformació que s'ha mesurat, que és sempre una deformació afí del contorn inicial.

En la implementació que s'ha fet, la llargada dels segments de cerca està determinada per la covariància estimada en la iteració anterior del filtre Kalman. Això es fa

projectant la matriu de covariàncies a la recta normal al contorn en el punt donat. Si el seguiment troba bones transformacions afins que expliquen canvis en la imatge, les disminucions de la covariància fan que els segments de cerca encongeixin. D'una banda, això és una bona estratègia de mesura, ja que quan es té confiança que el seguiment es fa correctament les característiques del contorn es busquen més localment i el soroll en la imatge afecta menys al sistema. Però, d'altra banda, aquesta solució no és la millor per seguir canvis grans en la projecció de la imatge, com per exemple el que es té després d'un moviment ràpid del robot.

Un cop estimat el vector de forma que millor aproxima la deformació del contorn actiu, la posa és calcula amb l'algorisme 2.1. A efectes pràctics, aquest algorisme serveix tant per al cas calibrat com per al cas no calibrat. Quan no es disposa de la distància focal es poden utilitzar les posicions del controlador del *zoom* (com es veurà al capítol 7) per obtenir una aproximació, i si no es coneix la distància inicial entre l'objecte i la càmera cal tenir en compte que els valors de les translacions estan escalats per Z_0 .

2.3 Dues experiències pràctiques

2.3.1 Avaluació de l'error en el moviment estimat

Seguidament es presenta un experiment per determinar la viabilitat de l'algorisme de seguiment de contorns en un robot real i un entorn real. Es vol avaluar, a grans trets, si sembla possible que l'algorisme proposat tingui utilitat per a la navegació de robots. Per tal de calcular el moviment precís del robot es fa servir un goniòmetre làser i es compara amb l'estimació del moviment produïda per l'algorisme presentat anteriorment (algorisme 2.1). Per obtenir informació mètrica cal conèixer els paràmetres de calibratge f i estimar la distància inicial Z_0 entre càmera i objectiu.

Reducció de la matriu de forma

Quan el moviment que cal parametritzar està restringit possiblement no calen els sis graus de llibertat de la matriu de forma definida a l'equació (2.12).

En l'experiment que es presenta a continuació es fa servir un robot planar amb tres graus de llibertat. En aquest cas, l'espai de moviment es pot parametritzar amb dues

2. Punt de partida: vector de forma i posa 3D

Entrada : \mathbf{S} , Z_0 , f i opcionalment f', u, u', v, v'

Sortida : posa 3D = $\{T_x, T_y, T_z, \phi, \theta, \psi\}$

- 1 **si** no hi ha paràmetres opcionals **aleshores**
 es considera una càmera sense *zoom*, de manera que s'inicialitza
 $u = u' = v = v' = 0 \quad f' = f$

fi

- 2 Reompondre \mathbf{M} a partir de la part vectorial de \mathbf{S} com

$$\mathbf{M} = \begin{bmatrix} S_3 + 1 & S_6 \\ S_5 & S_4 + 1 \end{bmatrix}$$

- 3 Trobar els vectors propis λ_1 i λ_2 i el valor propi \mathbf{v}_1 a partir de la descomposició en valors singulars de $\mathbf{M}\mathbf{M}^T$
- 4 La rotació θ es troba a partir dels valors propis:

$$\theta = \text{acos}\left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right)$$

- 5 La rotació ϕ es calcula a partir de les components del vector propi \mathbf{v}_1 :

$$\phi = \text{atan2}(v_{12}, v_{11})$$

- 6 Per calcular ψ primer cal reompondre les matrius de rotació

$$\mathbf{R}_{\text{psi}} = \frac{1}{\lambda_1} \mathbf{R}_{\mathbf{z}} |_2(-\phi) * \begin{bmatrix} 1 & 0 \\ 0 & 1/\cos(\theta) \end{bmatrix} * \mathbf{M}$$

i ψ es calcula amb:

$$\psi = \text{atan2}(\text{Rpsi}_{21}, \text{Rpsi}_{11})$$

- 7 Les translacions es calculen aplicant

$$T_x = \frac{S_1 - (u - u')}{f\sqrt{\lambda_1}} Z_0$$

$$T_y = \frac{S_2 - (v - v')}{f\sqrt{\lambda_1}} Z_0$$

$$T_z = \frac{Z_0 f'}{f\sqrt{\lambda_1}} - Z_0$$

Algorisme 2.1 : Algorisme per computar la posa a partir del vector de forma.

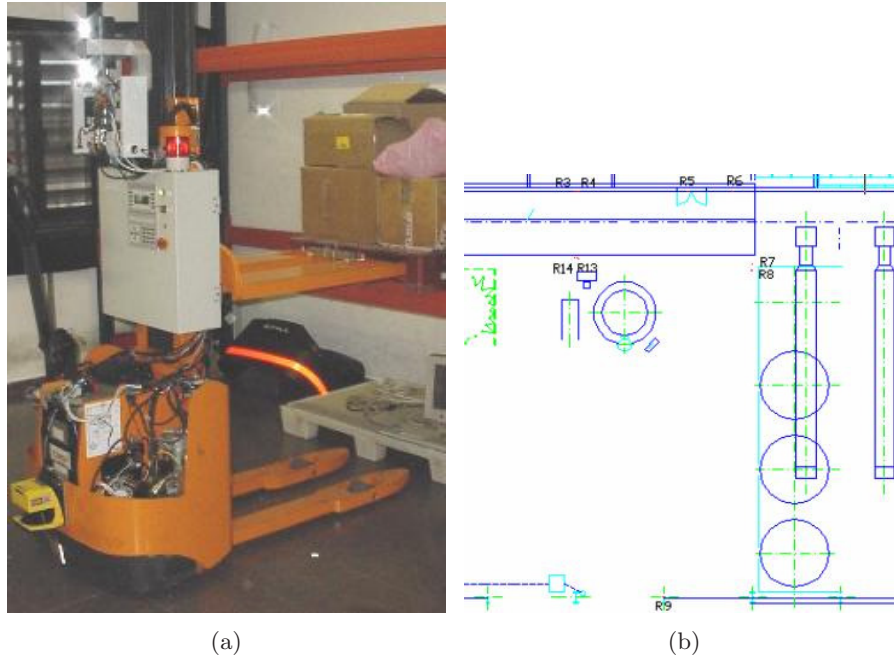


Figura 2.5: Portapalets robotitzat Still EGV-10 emprat en l'experiment real en un magatzem. Un enregistrator recollia informació d'odometria, posicionament làser i imatges. (b) Plànol detallat de la zona del magatzem on es va dur a terme l'experiment.

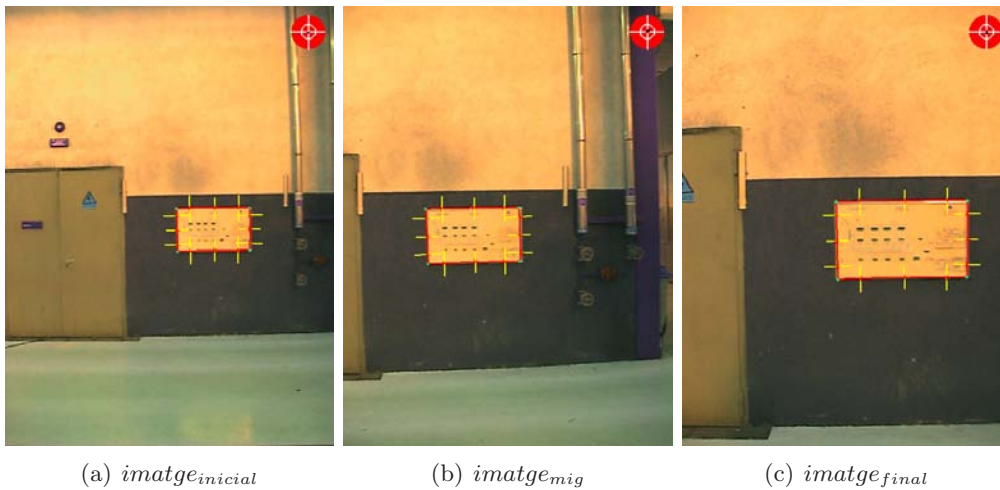


Figura 2.6: Imatges de l'experiment durant una translació llarga del robot. El contorn actiu es fixa en un panell d'informació que s'utilitza d'objectiu per estimar el moviment de la càmera.

translacions (T_x , T_z) i una rotació (R_y). Òbviament, la resta de moviments no són possibles. Si s'apliquen aquestes restriccions a les equacions (2.4 i 2.5) de la deformació

2. Punt de partida: vector de forma i posa 3D

afí d'un contorn es pot deduir un nou espai de forma que correspon a una matriu de forma de tres graus de llibertat

$$\mathbf{W} = \left(\left[\begin{array}{c} \mathbf{1} \\ \mathbf{0} \end{array} \right], \left[\begin{array}{c} \mathbf{Q}_0^x \\ \mathbf{0} \end{array} \right], \left[\begin{array}{c} \mathbf{0} \\ \mathbf{Q}_0^y \end{array} \right] \right), \quad (2.22)$$

i un vector de forma

$$\mathbf{S} = (t_x, M_{11} - 1, M_{22} - 1)^T. \quad (2.23)$$

Perquè això es pugui aplicar en situacions reals, cal que l'objecte que es fa servir com a objectiu estigui centrat en la imatge i que la direcció d'avançament del robot coincideixi amb l'eix òptic de projecció de la càmera. Òbviament, si no es compleixen aquestes condicions una translació de la càmera fa que la projecció de l'objectiu es desplaci verticalment al pla de la imatge. Per aquesta raó, cal ampliar la família de deformacions afins per incloure els desplaçaments verticals aplicant aquestes noves restriccions a les equacions (2.4 i 2.5) amb la qual cosa s'obté la matriu de forma

$$\mathbf{W} = \left(\left[\begin{array}{c} \mathbf{1} \\ \mathbf{0} \end{array} \right], \left[\begin{array}{c} \mathbf{0} \\ \mathbf{1} \end{array} \right], \left[\begin{array}{c} \mathbf{Q}_0^x \\ \mathbf{0} \end{array} \right], \left[\begin{array}{c} \mathbf{0} \\ \mathbf{Q}_0^y \end{array} \right] \right), \quad (2.24)$$

i el vector de forma

$$\mathbf{S} = (t_x, t_y, M_{11} - 1, M_{22} - 1)^T. \quad (2.25)$$

Experiment

El robot mòbil utilitzat en aquest experiment és un portapalets Still EGV-10 modificat (vegeu la figura 2.5(a)). De sèrie, aquest és un vehicle guiat manualment amb ajut en la tracció. Per robotitzar-lo, es va afegir un motor a l'eix de direcció amb tota l'electrònica necessària per al control dels dos eixos.

L'experiència pràctica es va fer en un magatzem de la companyia cervesera DAMM al Prat de Llobregat, a Barcelona. Durant l'experiència, el robot es va guiar manualment. Un programa enregistrator de dades guardava els senyals simultanis següents: la posició obtinguda utilitzant la triangulació dinàmica d'un goniòmetre làser, els reflexos captats i els senyals d'odometria proporcionats pels codificadors. En la mateixa freqüència, s'enviava un senyal de sincronisme a la càmera i es captava una imatge.

Les dades recopilades van permetre aplicar diferents processos per tal de comparar una nova tècnica de posicionament làser amb l'algorisme proposat basat en visió [2].

Encara que aquest experiment estava dissenyat en dos passos —recollida de dades i anàlisi de dades— les implementacions dels dos algorismes permeten l'execució a la màxima freqüència permesa per cada sensor, és a dir, 20 fps per al subsistema càmera i 8 Hz per al subsistema làser.

En l'experiment que es presenta el conjunt de dades consta de dues-centes imatges. S'inicialitza manualment un contorn actiu en un tauler d'informació que apareix en la primera imatge de la seqüència escollida (figura 2.6). Tal com s'ha presentat anteriorment, l'algorisme de seguiment troba la deformació afí del contorn inicial que millor encaixa amb la posició actual de la projecció del contorn i produeix un vector de forma (2.25). Els vectors de forma corresponents a aquest experiment s'han mostrat anteriorment a la figura 2.4. Fent servir les equacions (2.19-2.21) i les equacions (2.14-2.18) es pot estimar el moviment 3D. Si es pot conèixer la distància inicial a l'objectiu i es tenen els valors del calibratge de la càmera, es pot obtenir una reconstrucció mètrica del moviment. En aquest experiment, el valor de la profunditat inicial es va calcular amb el sensor làser, aprofitant que l'objectiu (el tauler d'informació) era sobre la mateixa paret que algunes de les marques catadiòptriques. Es va mesurar un valor inicial de 7.7 m. El moviment que es va realitzar és una translació d'aproximadament 3.5 m al llarg de la direcció d'avanç del robot, pertorbada per petits canvis de direcció. Les translacions obtingudes T_x , T_y i T_z es poden veure a la figura 2.7(a). Es pot observar que, tot i que el vector de forma inclou t_y , la translació T_y que es dedueix queda correctament al voltant de zero.⁶ Es pot reconstruir la trajectòria del robot (figura 2.7(b)) posant els valors computats T_x i T_z en correspondència al pla de moviment.

Mentre que les mesures obtingudes amb el làser són globals, el sistema de visió proporciona un posicionament relatiu a la posició inicial [98]. Per comparar les dues estimacions es transformen les mesures del làser en incrementals i s'estima la transformació que posa en correspondència els dos sistemes de referència. La comparació de les dues estimacions es mostra a la figura 2.8.

A la figura 2.8(a) es pot veure l'error comès pel sistema de visió considerant que la translació real és la mesurada pel sistema làser. El que es pot observar és que l'estimació de la translació T_z és més sorollosa que les de les translacions T_x i T_y (figura 2.8(a)).

⁶Naturalment, el seu valor depèn del grau d'alineació que hi ha entre la càmera i el pla de moviment del robot.

2. Punt de partida: vector de forma i posa 3D

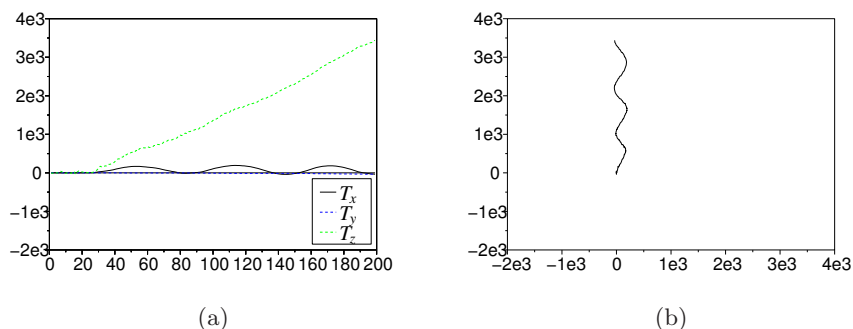


Figura 2.7: (a) Evolució de les translacions T_x , T_y i T_z (en mil·límetres). (b) Trajectòria computada (en mil·límetres) en el pla XZ .

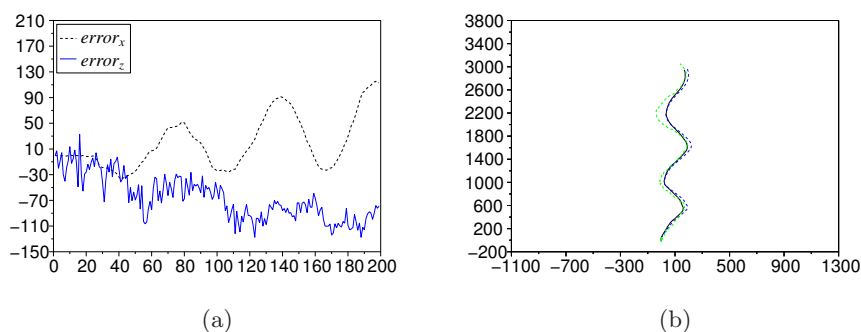


Figura 2.8: Comparació entre els resultats obtinguts amb l'algorisme proposat d'estimació del moviment i amb el de posicionament làser. (a) Diferència en mil·límetres entre l'estimació de translació proporcionada pel làser i l'algorisme proposat a cada imatge. (b) Trajectòries en mil·límetres en el pla XZ . La línia negra correspon a la trajectòria estimada pel làser; la blava, a la trajectòria del làser rectificada amb la informació de la distància entre els sensors, i la verda, a la trajectòria computada amb l'algorisme proposat.

També s'aprecia que l'error en la component X és més gran a mesura que el valor d'aquesta translació augmenta. Cal observar que en aquest cas es viola una de les condicions del model de perspectiva dèbil, ja que els punts del contorn estan allunyats del centre òptic de projecció.

Finalment, per comparar gràficament els dos mètodes, a la figura 2.8(b) es mostren tres trajectòries sobreposades: la mesurada amb el sistema de visió, la mesurada amb el goniòmetre làser i aquesta darrera rectificada amb la informació del posicionament del làser respecte de la càmera.

Aquest experiment mostra que l'estimació del moviment proporcionada per l'algorisme proposat té una precisió raonable, suficient per a la navegació de robots. Per poder comparar les dues estimacions de moviment ha estat necessari proporcionar a l'algorisme de visió la distància inicial a l'objectiu Z_0 i els paràmetres del calibratge de la càmera f . Òbviament, si no es disposa d'aquestes dades les translacions només es poden obtenir fins a un factor d'escala. Amb moviments escalats encara és possible obtenir informació útil per a la navegació de robots, per exemple, el temps per al contacte⁷ [98]. Els paràmetres interns de càmeres es poden calcular amb un procés previ de calibratge de la càmera o en línia amb mètodes d'autocalibratge. Més endavant, al capítol 6, es mostra un algorisme que s'ha desenvolupat per obtenir la distància inicial aprofitant una càmera amb *zoom*.

2.3.2 Enriquiment del seguiment de contorns amb informació inercial

Els sensors inercials i la visió són dos tipus de sensors que es complementen naturalment. Una velocitat baixa del robot és una situació favorable per a l'algorisme de seguiment de contorns, ja que els canvis que es produeixen entre dues imatges consecutives es poden mesurar adequadament. Però és un cas desfavorable per a un sensor inercial, ja que és menys precís per a velocitats baixes i sovint apareixen problemes de deriva en la posició a causa de la integració que cal fer de la velocitat. En canvi, una velocitat alta del robot és un cas favorable per a un sensor inercial, que la pot mesurar molt bé, però un cas desfavorable per a l'algorisme de seguiment. Utilitzant el bon condicionament dels sensors inercials precisament per detectar velocitats relativament altes, es mostra com es pot millorar l'algorisme de seguiment introduint la dinàmica mesurada pel sensor inercial.

S'ha vist a la secció 2.2.2 que l'algorisme de seguiment troba el vector de forma que millor explica la deformació del contorn al pla de la imatge. Per integrar la informació del sensor inercial dins de l'algorisme de seguiment cal expressar la informació del sensor inercial en espai de matriu de forma. Aquestes expressions corresponen a les transformacions inverses de les expressions presentades a la secció 2.2.3⁸.

⁷*Time To Contact*, TTC.

⁸La transformació inversa ens serà útil més endavant per estendre el desenvolupament del capítol 5 en el marc de l'UKF.

2. Punt de partida: vector de forma i posa 3D

Escalant la covariància d'acord amb la informació inercial

La implementació de l'algorisme de seguiment utilitza la covariància del filtre de Kalman per escalar les línies normals al contorn, que s'utilitzen per portar a terme el pas de mesura de l'algorisme (secció 2.2.4). Quan hi ha un moviment bruscat l'algorisme de seguiment no és capaç de mantenir l'objectiu i, per tant, no es pot calcular el moviment. Es vol aprofitar la informació del sensor inercial, que mesura molt bé aquests moviments bruscos, per canviar la longitud de les línies de cerca convenientment. L'objectiu és allargar a zona de cerca quan es detecta un moviment gran.

Considereu un sensor inercial de tres graus de llibertat que proporciona coordenades (x, y, θ) . Per simplicitat, el sensor es col·loca sota la càmera amb els eixos de coordenades alineats amb els de la càmera.⁹D'aquesta manera, els eixos X i Y del sensor inercial corresponen respectivament als eixos Z i X de la càmera, i les rotacions tenen lloc sobre el mateix eix. El moviment mesurat es pot expressar llavors com a translació

$$\mathbf{T} = \begin{pmatrix} v_x \\ 0 \\ v_z \end{pmatrix}, \quad (2.26)$$

i la rotació

$$\mathbf{R} = \begin{bmatrix} \cos v_\theta & 0 & -\sin v_\theta \\ 0 & 1 & 0 \\ \sin v_\theta & 0 & \cos v_\theta \end{bmatrix}. \quad (2.27)$$

Combinant les equacions (2.4, 2.5) amb la versió amb el sistema de referència a la càmera de les equacions (2.26, 2.27), les dades mesurades es poden expressar en espai de forma com

$$M_{11} = \frac{Z_0}{Z_0 R_{33} + T_z} R_{11} = \frac{Z_0}{Z_0 \cos v_\theta + v_z} \cos v_\theta \quad (2.28)$$

$$M_{21} = \frac{Z_0}{Z_0 R_{33} + T_z} R_{21} = 0$$

$$M_{12} = \frac{Z_0}{Z_0 R_{33} + T_z} R_{12} = 0$$

⁹Altrament, el que caldria és un procés de calibratge entre els dos sensors per tal de conèixer exactament la matriu homogènia de transformació entre els sistemes de coordenades del sensor inercial i de la càmera.

$$M_{22} = \frac{Z_0}{Z_0 R_{33} + T_z} R_{22} = \frac{Z_0}{Z_0 \cos v_\theta + v_z} \quad (2.29)$$

$$t_1 = \frac{1}{Z_0 R_{33} + T_z} (Z_0 R_{13} + T_x) = \frac{1}{Z_0 \cos v_\theta + v_z} (-Z_0 \sin v_\theta + v_x) \quad (2.30)$$

$$t_2 = \frac{1}{Z_0 R_{33} + T_z} (Z_0 R_{23} + T_y) = 0$$

Per a cada segment de cerca normal al contorn el factor d'escala es computa com

$$\mathbf{E} = \sqrt{\mathbf{N}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T) \mathbf{N}} \quad (2.31)$$

on \mathbf{N} són les coordenades de les línies normals al contorn, \mathbf{H} és el vector de les mesures i \mathbf{P} és la matriu 6×6 de covariància.

Cal fer notar que, atès que la covariància canvia a cada iteració, l'escala de les normals de cerca s'ha de recalculer en cada iteració. Es pot observar que el valor que aquest algorisme fixa per a cada normal de cerca és diferent depenent de l'orientació de la normal, tenint en compte la informació direccional de la covariància del filtre Kalman.

Ja que l'objectiu és escalar la covariància, els denominadors en les equacions (2.28 - 2.30) es poden eliminar. Aquestes equacions es poden reescriure en forma de vector de forma com

$$\mathbf{S} = \begin{pmatrix} t_1 & 0 & M_{11} - 1 & M_{22} - 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} -Z_0 \sin v_\theta + v_x & 0 & -v_z & Z_0 (1 - \cos v_\theta) - v_z & 0 & 0 \end{pmatrix}.$$

Per a velocitats de rotació petites $\sin v_\theta$ es pot aproximar per v_θ i per tant

$$\mathbf{S} = \begin{pmatrix} -Z_0 v_\theta + v_x & 0 & -v_z & Z_0 v_\theta^2 / 2 - v_z & 0 & 0 \end{pmatrix}. \quad (2.32)$$

El sensor inercial representa les dades en la direcció X dins del rang $[v_{xmin} \dots v_{xmax}]$. Per simplificar la notació es pot considerar que el rang és simètric $|v_{xmin}| = |v_{xmax}|$. Les mesures del sensor es poden reescalar per produir mesures en el rang $[\bar{v}_{xmin} \dots \bar{v}_{xmax}]$. Un valor v_x mesurat amb el sensor inercial es pot reescalar fent servir

$$\bar{v}_x = |v_x| \frac{\bar{v}_{xmax} - \bar{v}_{xmin}}{v_{xmax}} + \bar{v}_{xmin} \quad (2.33)$$

2. Punt de partida: vector de forma i posa 3D

Seguint el mateix raonament, les components de la informació inercial en forma de vector de forma també es poden reescalar. Per al primer component es té

$$t_{1max} = Z_0 v_{\theta max} + v_{x max}, \quad (2.34)$$

i l'expressió

$$\bar{t}_1 = |t_1| \frac{\bar{t}_{1max} - \bar{t}_{1min}}{t_{1max}} + \bar{t}_{1min} = |t_1| f_{t_1} + \bar{t}_{1min} \quad (2.35)$$

La informació inercial es pot afegir a l'algorisme de seguiment reescalant la matriu de covariància amb una matriu que representa la informació inercial de la manera següent

$$\mathbf{E} = \sqrt{\mathbf{N}^T (\mathbf{HVPV}^T \mathbf{H}^T) \mathbf{N}} \quad (2.36)$$

on \mathbf{V} és la matriu escalada de les mesures del sensor inercial definida com

$$\mathbf{V} = \begin{bmatrix} t_1 & & & & & & & 0 \\ & t_2 & & & & & & \\ \cdot & & M_{11} - 1 & & & & & \cdot \\ \cdot & & & M_{22} - 1 & & & & \cdot \\ & & & & M_{12} & & & \\ 0 & & & & & & M_{21} & \end{bmatrix} \begin{bmatrix} f_{t_1} \\ f_{t_2} \\ f_{M_{11}} \\ f_{M_{22}} \\ f_{M_{12}} \\ f_{M_{21}} \end{bmatrix} + \begin{bmatrix} \bar{t}_{1min} \\ \bar{t}_{2min} \\ \bar{M}_{11min} \\ \bar{M}_{22min} \\ \bar{M}_{12min} \\ \bar{M}_{21min} \end{bmatrix}. \quad (2.37)$$

En la propera secció d'experiments, els valors màxims i mínims s'han fixat a 1 i 2 respectivament.

Experiments combinant visió i informació inercial

En aquests experiments s'utilitza un robot mòbil RobuCab de Robosoft. Com es pot veure a la figura 2.9, és un vehicle mòbil relativament gran amb capacitat per a quatre persones. Es pot utilitzar en dos modes: en mode de conducció cotxe (roten les rodes de només un eix) i en mode bidireccional (les quatre rodes giren).

Per simplificar el sistema de control, es fa servir el mode de conducció de cotxe, però es poden obtenir millors resultats, en termes d'angle de gir menor, amb el mode de conducció bidireccional, ja que s'incrementa l'angle màxim de rotació que el vehicle és capaç de seguir. Per a aquest experiment es munta un sistema de visió monocular DragonFly de Point Grey, i les imatges es tracten amb el sistema de seguiment descrit. Per mesurar les rotacions sobre l'eix Y es fa servir un sensor inercial Gyrostar, de Murata. Per mesurar les acceleracions lineals als eixos X i Z s'utilitza un doble



Figura 2.9: Plataforma mòbil Robucab usada en l'experiment.

acceleròmetre ADXL d'Analog Devices. Tots aquests sensors estan connectats a una targeta dissenyada especialment per a aquest experiment que compta amb un processador AVR utilitzat per fer conversions A/D, descodificació de PWM i integració dels senyals d'acceleració. Té també un termòmetre per a la correcció tèrmica de les dades. Aquest sensor 'intel·ligent' que s'ha fabricat proporciona l'acceleració i també en fa la integració per obtenir la velocitat mitjana i la posició. El biaix, típic en aquesta classe de càlculs d'integració, es restaura periòdicament amb la informació obtinguda per la fusió dels altres sensors. Aquesta targeta comparteix memòria amb una targeta basada en un processador PowerPC MPC555, que està connectada a través d'un bus CAN al PC de processament de control i visió. Tot el sistema s'executa en un nucli de Linux en temps real sobre un ordinador industrial Pentium 233 MHz. S'ha utilitzat una nova aproximació a la programació distribuïda [111] per programar el control del robot i per a la intercomunicació del processament de control i de visió, aprofitant el sistema operatiu de temps real.

Podria semblar que ja que el robot és planar es pot fer servir la forma reduïda de l'espai de forma (2.24). Però en aquest cas no és possible, ja que es tracta principalment d'un robot d'exterior i per tant no es pot assumir que es mou en un pla. A més, disposa de suspensions que fan que, depenent de la càrrega, la seva distribució i el seu moviment, hi hagi moviment en gairebé els sis graus de llibertat. Per tant, cal fer servir la matriu que codifica tot l'espai de forma (2.12).

En aquest experiment el robot s'ha programat perquè pugui conduir autònomament

2. Punt de partida: vector de forma i posa 3D

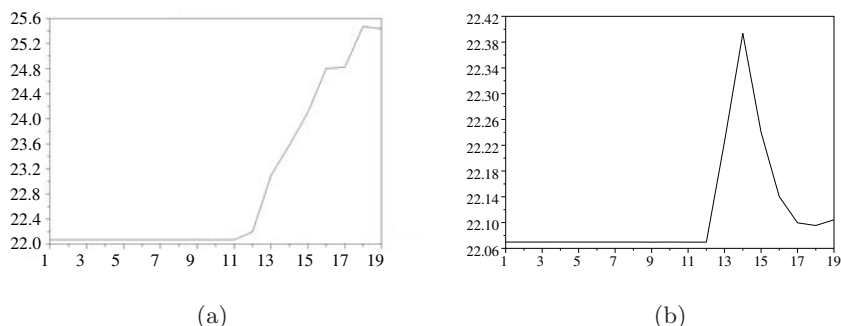


Figura 2.10: Traces de la seqüència de matrius de covariància resultat de seguir un contorn (a) sense informació inercial i (b) utilitzant la informació inercial. En el segon cas el contorn no es perd i el seguiment pot continuar.

seguint un camí filoguiat. D'aquesta manera, la trajectòria es pot repetir fàcilment i es poden fer tots els experiments necessaris en condicions molt similars. El camí seguit en l'experiment que es mostra consta d'un segment de recta, una corba i un altre segment de recta.

Primer s'utilitza l'algorisme sense informació inercial. En el primer segment recte el contorn se segueix bé, però quan comença el gir i la projecció del contorn es mou més ràpidament al pla de la imatge l'algorisme de seguiment perd l'objectiu. Es fa servir la traça de la matriu de covariància com a mesura. Com es pot observar en els resultats de la figura 2.10(a), en aquest experiment la traça de la matriu de covariància augmenta contínuament.

Segon, s'utilitza l'algorisme que inclou la informació inercial en el seguiment. En aquest experiment, el seguiment no perd l'objectiu i acaba la seqüència donant els valors de posa correctament. Com es pot veure en la representació de la traça de la covariància en la figura 2.10(b), la covariància augmenta al començament del gir però disminueix de pressa, cosa que mostra que l'algorisme de seguiment ha mantingut l'objectiu malgrat la seva translació ràpida d'un costat a l'altre de la imatge.

2.4 Conclusions

Aquest capítol s'ha iniciat presentant els resultats d'uns treballs previs [17, 97, 99] en què el model de càmera de perspectiva dèbil s'ha emprat per definir un espai de forma

afí amb què es parametriza el moviment realitzat per una càmera, i posteriorment s'ha mostrat com es pot estimar el moviment 3D a partir d'aquest espai de forma. Fins al moment, aquest algorisme mai no havia estat comparat amb altres algorismes de posicionament, i l'error i la precisió no havien estat mesurats abans.

La primera estimació de l'error que es comet en la recuperació del moviment s'ha obtingut amb un experiment amb un robot transpaletitzador que s'ha fet en un entorn real, on també s'ha explorat la possibilitat de combinar la flexibilitat del sistema de visió amb la precisió d'un sistema de posicionament làser. S'ha pogut comprovar numèricament que l'error més gran es comet en la translació T_z i que el seu valor no és mai superior al 3%, i s'ha vist que l'algorisme de visió presentat pot ser una bona alternativa a la navegació làser en fragments de la navegació on no calgui un grau de precisió molt elevat.

Finalment, s'ha proposat un nou algorisme que aprofita la informació d'un sensor inercial per millorar el seguiment. Ha calgut definir les transformacions inverses d'espai 3D a espai de forma per poder expressar la informació inercial a espai de forma, i s'ha proposat un algorisme per escalar convenientment la zona de cerca. S'ha validat l'algorisme amb un experiment real amb un robot RobuCar de transport de persones.

En els treballs realitzats s'ha observat el potencial de l'algorisme d'estimació del moviment mitjançant contorns actius. S'ha constatat que cal fer un estudi més en profunditat de la precisió que es pot obtenir, i de quins factors influeixen en aquesta precisió: distància inicial, efectes perspectius, soroll en l'adquisició...

Capítol 3

Recuperació de la direcció epipolar

«Això, en teoria, també serà veritat; però a la pràctica és fals». Amb aquest sofisma s'admeten les raons i tanmateix es neguen les conseqüències.

A. Schopenhauer, *L'art de tenir sempre la raó*, (Estratagema 33), Empúries, 2005.

«"That's all very well in theory, but it won't do in practice.» In this sophism you admit the premisses but deny the conclusion.

A. Schopenhauer, *The Art of Controversy*, (stratagem 33), (Translated by T. Bailey Saunders, at Gutenberg project)

Extracte

Per aprofundir en el coneixement de l'algorisme presentat en el capítol anterior el primer pas és estudiar la geometria que relaciona dues vistes d'un objecte planar en el context de condicions de visualització afins. Utilitzant resultats de geometria projectiva, es demostra que la direcció epipolar afí es pot recobrar a condició que el moviment de la càmera sigui lliure de ciclorrotació (moviment que correspon a la rotació sobre l'eix de projecció). Aquest resultat es demostra analíticament, i se sotmet a experimentació. L'experiment consta d'un robot Staübli que sosté un objecte planar davant d'una càmera, i s'utilitza per obtenir seqüències d'imatges amb moviments ben coneguts, que serveixen per avaluar el rendiment del mètode i trobar les seves limitacions en la pràctica. L'algorisme es beneficia dels avantatges

3. Recuperació de la direcció epipolar

d'utilitzar un contorn actiu per fer el seguiment, que permet fer una aplicació robusta i un processament simple (resolent una equació de segon ordre senzilla), i és vàlid fins i tot per a escenes pobrament texturades. La direcció epipolar ha resultat ser un bon indicador de la qualitat de l'afinitat que es calcula. Els experiments realitzats han servit per avaluar quines situacions són més o menys favorables.

3.1 Introducció

En la darrera dècada s'ha dedicat un treball extens a estimar la geometria epipolar que relaciona dues vistes de la mateixa escena. Els mètodes proposats solucionen el problema de manera diferent depenent del següent: a) el model de càmeres que assumeixen, b) els tipus d'escenes a què s'apliquen, i c) com es mesura el moviment visual en el pla de la imatge.

El model de càmera relaciona el moviment mesurat en el pla de la imatge amb el moviment 3D de la càmera. Habitualment s'han utilitzat diferents models de càmera depenent de les condicions de visualització per emular el procés de creació de la imatge [15, 101]. S'ha demostrat que el model de perspectiva completa -en versió calibrada (càmera perspectiva) i no calibrada (càmera projectiva)- és massa general quan els efectes de perspectiva disminueixen i no són apreciables. En condicions de visualització de perspectiva dèbil (camp de visió petit o variació de la profunditat en l'escena petita comparada amb la seva distància mitjana a la càmera), els models de càmera simplificats, com l'ortogràfic, l'ortogràfic escalat, o la seva generalització per al cas no calibrat, el model de càmera afí, proporcionen una aproximació avantatjosa a la càmera de perspectiva completa, ja que eviten computar paràmetres que les condicions de visualització fan que estiguin mal condicionats.

L'estimació de la geometria epipolar en el cas afí ha rebut molta atenció en la literatura [81, 125]. Tanmateix, la majoria dels mètodes coneguts assumeixen el següent:

1. l'escena conté **informació de profunditat**, de manera que els algorismes fallen quan la configuració de l'escena s'aproxima a una estructura planar. Per exemple, els experiments amb l'algorisme estàndard d'or¹ (secció 3.3) mostren que per

¹Per definició, un estàndard d'or és el mètode, procediment o mesura que s'accepta àmpliament i que s'usa com la millor comparació per avaluar mètodes, procediments o mesures nous.

obtenir resultats acceptables el relleu (diferència de profunditats) de l'objecte ha de ser del mateix ordre de magnitud que la llargada/amplada; i

2. l'escena està suficientment **texturada** per permetre l'estimació visual del moviment a partir de correspondències de punts.

Aquí s'explora un *enfocament alternatiu* aplicable quan aquestes dues suposicions no són aplicables.

Els intents previs per superar la suposició 1), i per tant calcular la geometria epipolar entre dues imatges a partir d'un pla dominant, han utilitzat una càmera de perspectiva completa. És ben sabut que dues vistes d'un pla es relacionen per una colineació segons el model de perspectiva completa [62]. Uns quants autors han utilitzat aquest fet per proposar algorismes per al calibratge de càmeres [132], autocalibratge [44, 93] i extracció d'estructura i moviment des de vistes no calibrades de punts sobre plans [14, 29] o corbes planars [79]. Tanmateix, quan disminueixen els efectes de perspectiva, la relació entre dues vistes d'una estructura planar es converteix en una afinat, que invalida els mètodes basats en colineacions. Que en tinguem coneixement, no hi ha treballs previs que eludeixin la suposició 1) en condicions de visualització afí.

Pel que fa a la suposició 2), el que es proposa és calcular visualment el moviment entre dues vistes utilitzant contorns actius, i així es poden abordar escenes molt poc texturades. L'afinat que relaciona dues vistes, que aquí es proposa obtenir seguint un contorn planar, és teòricament equivalent a la que resulta de tres correspondències de punts. Tanmateix, com ja s'ha esmentat, l'algorisme que es presenta eludeix la suposició 2) i es beneficia d'altres avantatges dels contorns actius, com la seva robustesa a oclusions i soroll, així com la simplicitat del procediment de seguiment, al mateix temps que evita la correspondència de punts.

Primer s'explora quina informació de la geometria epipolar afí es pot inferir a partir de la deformació afí de la projecció d'un contorn rígid i planar en dues vistes de perspectiva dèbil, fent servir l'anàlisi estratificada del moviment per a condicions de visualització afí introduïda per Koenderink i Van Doorn [81] i revisitada per Shapiro et al. [125]. Això posa la base per obtenir els paràmetres de moviment en una segona fase. Es mostra que, amb un moviment 3D lliure de ciclorrotació, la direcció epipolar es pot recuperar de la relació de les dues vistes afins del contorn. Es realitzen una

3. Recuperació de la direcció epipolar

sèrie d'experiments per provar la sensibilitat del mètode davant de diferents condicions. Aquests experiments posen de manifest que la direcció epipolar, a part de l'interès que té ja que relaciona dues vistes, serveix per tenir una mesura de la qualitat de l'afinitat que es computa. D'aquesta manera, amb només un nombre es pot avaluar la qualitat de l'afinitat calculada (per exemple, davant de situacions adverses fora de condicions de perspectiva dèbil, amb soroll...), cosa que fins ara era molt difícil d'acomplir, ja que s'havia de manejar tota la posa 6D, introduïda al capítol 2.

El capítol s'organitza de la manera següent. La secció 3.2 conté l'estudi analític de dues vistes de perspectiva dèbil i proporciona la base per a la recuperació de la direcció epipolar. La secció 3.3 es dedica a l'experimentació, utilitzant seqüències d'imatges tant sintètiques com reals. Es mostren múltiples experiments, que inclouen la relaxació de les condicions de visualització de perspectiva dèbil, i es mostra que la direcció epipolar serveix també per mesurar la qualitat de l'afinitat que codifica la deformació del contorn. Finalment, a la secció 3.4 es resumeix la contribució feta.

3.2 Estudi analític de dues vistes amb perspectiva dèbil

3.2.1 El model de càmera

El model de càmera que es fa servir és el de perspectiva dèbil, introduït a la secció 2.2.1. Seguint el mateix tipus de raonament, si es considera l'equació $aX + bY + c = Z$ d'un pla global \mathcal{S} , llavors les dues vistes de l'escena coplanària es poden explicar per l'afinitat donada per (2.3)

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \mathbf{M} \begin{bmatrix} x \\ y \end{bmatrix} + \mathbf{t},$$

amb

$$\mathbf{M} = s \frac{f'}{f} \begin{bmatrix} R_{11} + aR_{13} & R_{12} + bR_{13} \\ R_{21} + aR_{23} & R_{22} + bR_{23} \end{bmatrix}, \quad (3.1)$$

$$\mathbf{t} = -\frac{f'}{Z'} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \end{bmatrix} \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} + c \begin{bmatrix} R_{13} \\ R_{23} \end{bmatrix}, \quad (3.2)$$

on $s = Z_0/Z'$ és el factor d'escala que explica la variació de profunditat ($s > 1$ si la segona càmera s'acosta a l'objecte, i $s < 1$ si s'allunya), i $R_{i,j}$ són els elements de la matriu de rotació \mathbf{R} .

3.2 Estudi analític de dues vistes amb perspectiva dèbil

Una direcció $\mathbf{v} = [x, y]^\top$ de la primera imatge \mathcal{R} correspon, per l'afinitat citada, amb la direcció $\mathbf{M}\mathbf{v}$ de la segona imatge \mathcal{R}' . Gràcies al fet que els sistemes de referència afins escollits en les dues càmeres coincideixen pel desplaçament, podem superposar les dues imatges i té sentit considerar les direccions invariants per \mathbf{M} .

3.2.2 Recuperació de la direcció epipolar

Considerem un sistema de coordenades ortonormal associat a la primera imatge (per exemple, coordenades de píxel normalitzades, quan es coneixen la relació d'aspecte i el biaix de la càmera). La matriu de rotació sobre l'eix unitari $[\cos \alpha, \sin \alpha, 0]^\top$ i l'angle ρ té la forma

$$\mathbf{R} = \begin{bmatrix} (1 - \cos \rho) \cos^2 \alpha + \cos \rho & \cos \alpha \sin \alpha (1 - \cos \rho) & \sin \alpha \sin \rho \\ \cos \alpha \sin \alpha (1 - \cos \rho) & (1 - \cos \rho) \sin^2 \alpha + \cos \rho & -\cos \alpha \sin \rho \\ -\sin \alpha \sin \rho & \cos \alpha \sin \rho & \cos \rho \end{bmatrix}. \quad (3.3)$$

Per això, la matriu \mathbf{M} és

$$\mathbf{M} = s \frac{f'}{f} \begin{bmatrix} (1 - \cos \rho) \cos^2 \alpha & \cos \alpha \sin \alpha (1 - \cos \rho) \\ + \cos \rho + a \sin \alpha \sin \rho & + b \sin \alpha \sin \rho \\ \cos \alpha \sin \alpha (1 - \cos \rho) & (1 - \cos \rho) \sin^2 \alpha \\ -a \cos \alpha \sin \rho & + \cos \rho - b \cos \alpha \sin \rho \end{bmatrix}, \quad (3.4)$$

on $\mathbf{a} = [\cos \alpha, \sin \alpha]^\top$ és la direcció de l'eix de rotació i el vector ortogonal $\mathbf{e} = [-\sin \alpha, \cos \alpha]^\top = \mathbf{a}^\perp$ és la direcció epipolar. Un càlcul directe mostra que

$$\mathbf{M}\mathbf{e} = s \frac{f'}{f} (\cos \rho + \sin \rho (a \sin \alpha - b \cos \alpha)) \mathbf{e}, \quad (3.5)$$

la qual cosa dóna una prova analítica del resultat següent:

Teorema 3.1. Si el moviment rígid entre dues càmeres de perspectiva dèbil se suposa que és lliure de ciclorrotació, llavors la direcció epipolar \mathbf{e} pot ser recobrada com un dels dos vectors propis de la part vectorial \mathbf{M} de l'afinitat que explica dues vistes d'una escena planar.

Com a conseqüència, la direcció $\mathbf{a} = \mathbf{e}^\perp$ de l'eix de rotació també es pot recuperar.

La figura 3.1 il·lustra el resultat citat. Es mostren dues vistes \mathcal{R} i \mathcal{R}' d'un objecte planar en forma de "H", relacionades per una rotació sobre un eix paral·lel al pla de la imatge (per exemple, lliure de ciclorrotació). Per poder-ho il·lustrar més fàcilment,

3. Recuperació de la direcció epipolar

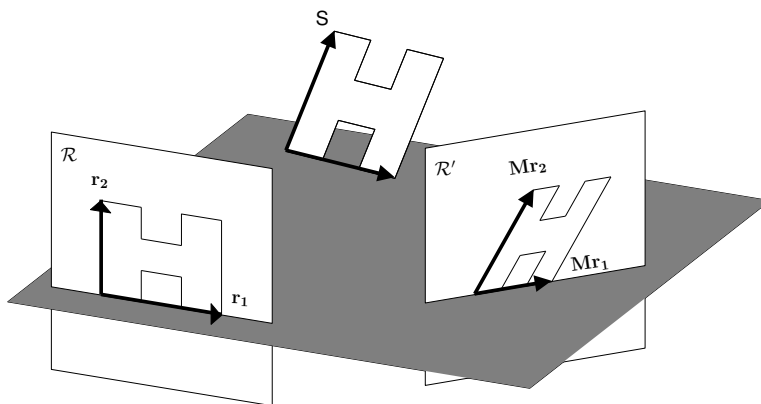


Figura 3.1: Il·lustració gràfica del teorema 3.1. Vegeu el text per als detalls.

es tria la base $\{\mathbf{r}_1, \mathbf{r}_2\}$ alineada amb els eixos principals de la “H”, i l’eix de rotació es tria paral·lel a \mathbf{r}_2 . Així, el pla gris de la figura 3.1 escombrat per \mathbf{r}_1 és invariant a la rotació. Cal notar que, llavors, la direcció epipolar és la de \mathbf{r}_1 a \mathcal{R} i la de \mathbf{Mr}_1 a \mathcal{R}' , i la seva perpendicular dins de cada imatge és la direcció de l’eix de rotació.

3.2.3 Computant la direcció epipolar a partir dels paràmetres de l’afinitat

Es fixa un sistema de coordenades qualsevol en la imatge (per exemple coordenades de píxel, ja que l’ortonormalitat no s’exigeix) i se suposa que l’afinitat que explica les dues vistes té l’expressió

$$\mathbf{x}' = \mathbf{M}\mathbf{x} + \mathbf{t} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \quad (3.6)$$

En la virtut del teorema 3.1, la direcció epipolar és un dels vectors propis de \mathbf{M} . Un vector propi $[1, w]^\top$ de \mathbf{M} satisfà l’equació

$$M_{12}w^2 + (M_{11} - M_{22})w - M_{21} = 0. \quad (3.7)$$

Si el moviment està en les hipòtesis del teorema 3.1, llavors (3.7) ha de tenir dues solucions reals, w_1 i w_2 , i la direcció epipolar és $\mathbf{e} = [1, w_i]^\top$, per a algun $i \in \{1, 2\}$ (o $[0, 1]^\top$, en cas que $M_{12} = 0$).

3.3 Experimentació

S'han fet dos conjunts d'experiments per avaluar la precisió del mètode proposat. El primer conjunt utilitza seqüències d'imatges sintètiques generades simulant el moviment de la càmera i computant les projeccions segons un model de càmera de perspectiva completa. Utilitzant aquest conjunt, s'avalua la sensibilitat de l'algorisme proposat: primer, respecte a les característiques de la forma de contorn; segon, relaxant cadascuna de les diferents condicions del model de perspectiva dèbil; tercer, introduint ciclorrotació al moviment, i en quart lloc, afegint quantitats variables de soroll a les projeccions del contorn en el pla de la imatge.

La manera més comú de calcular la geometria epipolar afí és utilitzant l'algorisme estàndard d'or [62]. Aquesta tècnica exigeix trobar les correspondències en les imatges de com a mínim quatre punts no coplanars. Utilitzant el conjunt de proves sintètic, es mostren els efectes de configuracions quasi coplanars sobre aquest mètode, i es comparen els resultats amb els de l'algorisme proposat, que computa la direcció epipolar afí a partir del seguiment d'un contorn planar.

En el segon conjunt d'experiments s'utilitzen imatges reals preses per un braç robòtic que es mou al llarg d'un camí calibrat, amb la qual cosa es mostra el rendiment de l'algorisme proposat en condicions realistes. També es mostren les conseqüències de reduir la dimensió del vector de forma de 6 a 5 quan es parametritzen diferents deformacions del contorn.

3.3.1 Consideracions preliminars

Representació de contorn

Com s'ha esmentat en el capítol 2.2.2, el contorn $\mathbf{d}(s)$ en (2.6) es representa utilitzant B-Splines. Breument, la seqüència de punts de control, juntament amb la multiplicitat dels nòduls que uneixen els diferents segments de corba polinòmica, es fan servir per representar el contorn [17]. La multiplicitat del nòdul determina la suavitat de la corba en aquell punt. En la nostra representació amb B-Splines quadràtics, la multiplicitat doble s'utilitza per modelar les cantonades. Com es pot veure en les figures 3.2(a) i 3.2(b), només doblant la multiplicitat d'un nòdul la forma del contorn actiu canvia, fins i tot si la seqüència de punts de control és la mateixa. Així doncs, tant el conjunt

3. Recuperació de la direcció epipolar

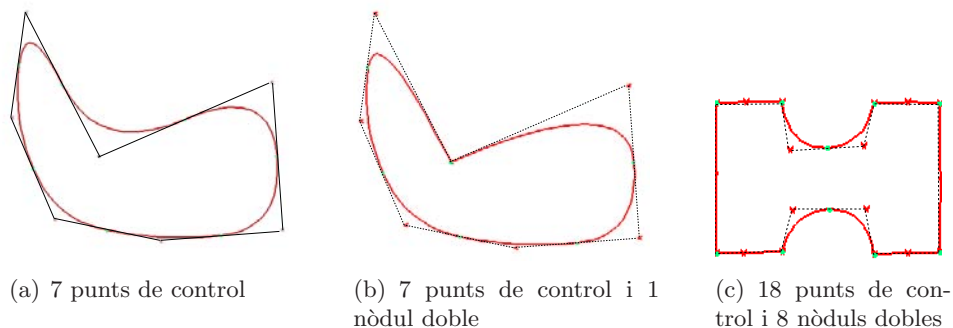


Figura 3.2: Contorns representats amb B-Splines: a (a) i a (b) es mostra que diferents contorns es poden modelar amb el mateix polígon de control, excepte per les multiplicitats dels nòduls, mentre que (c) mostra el modelat d'un contorn en forma de "H" utilitzat en els experiments.

de punts de control com les multiplicitats dels nòduls s'han d'utilitzar per obtenir el vector de forma \mathbf{S} de l'equació (2.11).

Això es té en compte quan s'utilitza el seguiment de contorns en els experiments que utilitzen imatges reals. Tanmateix, en els experiments que utilitzen imatges sintètiques, es computa la pseudoinversa de la matriu de forma amb només els punts de control, sense tenir en compte la multiplicitat dels nòduls. Aquesta simplificació és vàlida perquè el contorn en forma de "H" utilitzat en els experiments (figura 3.2(c)) conté molts nòduls dobles i el polígon de control s'aproxima molt a la projecció del contorn real. El mateix és cert per a la forma quadrada que també s'utilitza en els experiments sintètics.

El model de càmera emprat per generar imatges sintètiques

Quan les imatges sintètiques es generen utilitzant un model de càmeres afí (és a dir, assumint condicions de visualització de perspectiva dèbil), la direcció epipolar es recupera exactament amb l'algorisme proposat, i en demostra així la validesa. Tanmateix, ens agradaria avaluar els resultats del mètode en condicions més generals. Amb aquest objectiu, es generen un conjunt d'imatges sintètiques de prova utilitzant un model de càmera de perspectiva completa (representat anteriorment a la figura 2.1). En aquest cas, naturalment, els efectes perspectius afecten a la recuperació de la direcció epipolar de les maneres que s'analitzen en les seccions següents.

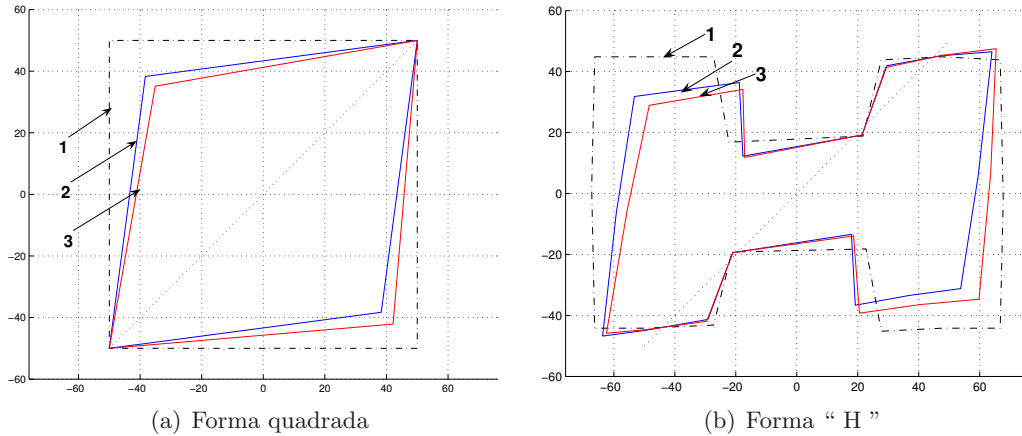


Figura 3.3: Diferència en la projecció de dos polígons de control per als dos models de càmeres dels quals es parla, després de rotar la càmera 40° segons un eix a 45° situat sobre el centre de l'objecte. 1) Contorn original, 2) projecció de perspectiva dèbil i 3) projecció de perspectiva completa.

A la figura 3.3 es pot veure una il·lustració de les diferències entre els polígons de control d'un contorn projectats amb els dos models de càmeres (perspectiva dèbil i perspectiva completa) per a dos objectes diferents: un quadrat simple (figura 3.3(a)) i una forma en "H" (figura 3.3(b)). A la figura, (1) representa el patró original, (2) correspon a la projecció de perspectiva dèbil i (3) mostra la projecció amb perspectiva completa. Aquestes projeccions s'obtenen després d'una rotació de les càmeres de 40° sobre un eix centrat en el centroid de l'objecte i que té una inclinació de 45° , com el representat en cada figura per una línia discontinua. L'objecte està situat a una distància de 500 mm des del pla d'imatge de les càmeres, i la distància focal es fixa a 5 mm. Cal notar que s'escull aquesta configuració intencionadament, lluny de les suposicions de perspectiva dèbil, per poder observar els efectes de perspectiva en la projecció dels dos models.

3.3.2 Efectes de la forma del contorn

Per determinar la influència de la forma del contorn sobre la recuperació de la direcció epipolar s'utilitzen els dos contorns diferents de la figura 3.3. El moviment que se simula per tal de poder calcular la direcció epipolar és una rotació de la càmera de 40° sobre un eix l'orientació del qual varia de 0° a 360° . Es fixa Z_0 a 1500 mm.

3. Recuperació de la direcció epipolar

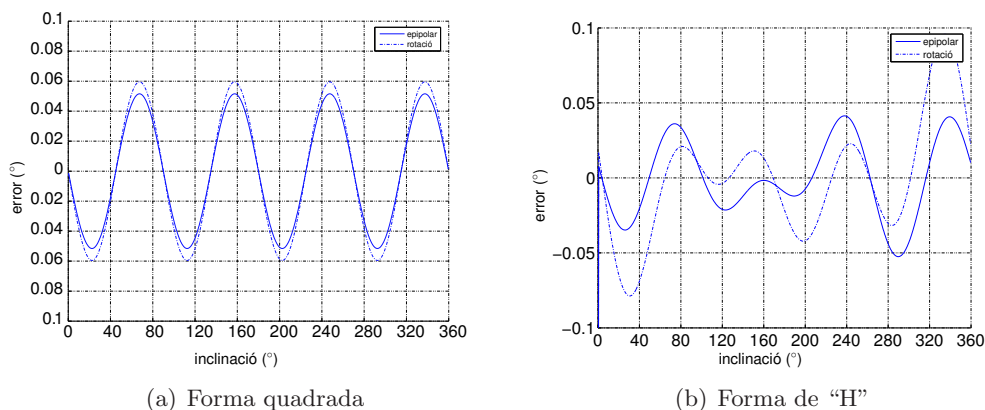


Figura 3.4: Error en la recuperació de la direcció epipolar després d’una rotació de la càmera de 40° sobre un eix a través del centre de l’objecte que varia de 0° a 360° en orientació. La línia contínua mostra l’error en la recuperació de la direcció d’epipolar, mentre que la línia discontinua indica l’error en la recuperació de l’orientació de l’eix de rotació.

La figura 3.4(b) mostra l’evolució de l’error que es comet en la recuperació de la direcció epipolar. Es pot observar que l’error roman molt baix per a totes les configuracions: menys de 0.1° . Per a la forma quadrada, l’error mostra una periodicitat de 90° , i és zero per a les rotacions específiques de $0^\circ, 45^\circ, \dots, 360^\circ$. En properes seccions es veurà que aquest efecte es conserva fins i tot després que es relaxin les condicions de perspectiva dèbil. Tanmateix, per a la forma més complexa en “H”, l’error de recuperació d’epipolar no es distribueix igualment en els quatre quadrants, i les rotacions per a les quals l’error són zero no tenen la mateixa periodicitat pel que fa al model quadrat simple.

A partir d’aquest experiment i d’altres de similars, concloem que la forma del contorn que se segueix té una petita influència sobre la distribució de l’error per a moviments de càmeres diferents. Això es produeix presumiblement a causa de les diferències en les simetries de la forma i en el nombre de punts de control que conformen el polígon de control.

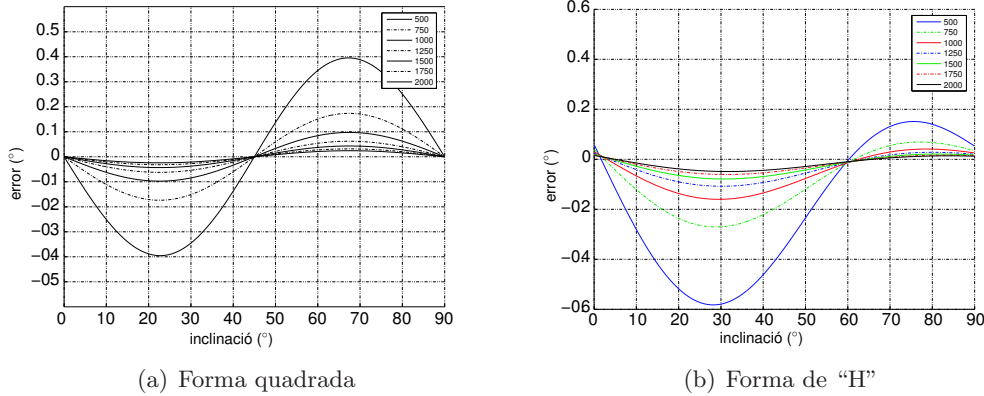


Figura 3.5: Error en la recuperació de la direcció epipolar quan la distància des de la càmera fins a l'objecte varia de 500 a 2000 mm. Es mostren els resultats per a una rotació de la càmera de 40° sobre un eix fixat sobre l'objecte i amb una inclinació que va de 0° a 90° .

3.3.3 Relaxació de les condicions de perspectiva dèbil

Translació al llarg del centre òptic: distància a l'objectiu

S'analitza com afecta la distància des de la càmera fins a l'objecte Z_0 en el càlcul de la direcció epipolar. Disminuint la distància els efectes de perspectiva es fan evidents, i consegüentment, *a priori* hauria d'augmentar l'error en la recuperació de direcció epipolar. Per a aquest experiment es consideren les distàncies 500, 750, 1000, 1250, 1500, 1750 i 2000 mm. La més petita, 500 mm, correspon a una situació extrema per al model de perspectiva dèbil, ja que les distorsions en el polígon de control projectat són importants (vegeu la figura 3.3(b)). Per a valors de profunditat més grans, les condicions afins se satisfan millor, i computen, per tant, un error menor, com es mostra a la figura 3.5. Cal fixar-se en el fet que fins i tot en les condicions més adverses l'error de recuperació queda per sota de 0.6° .

Rotació de la càmera: relleu

El model de perspectiva dèbil suposa que la diferència de profunditat entre els punts de l'objecte, el relleu, és petita comparada amb la distància Z_0 que hi ha des de la càmera fins al centroid de l'objecte.

3. Recuperació de la direcció epipolar

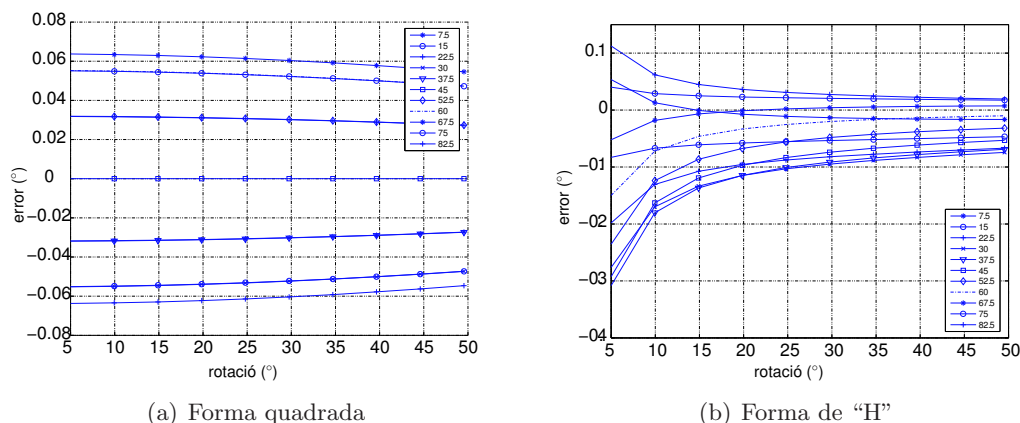


Figura 3.6: Error en la recuperació de la direcció epipolar per a inclinacions de l'eix de rotació diferents quan les rotacions a través d'aquest eix van de 5° a 50° .

Començant des d'una vista frontoparal·lela del contorn, la diferència en profunditat dels punts es relaciona directament amb la quantitat de rotació que la càmera realitza sobre un eix situat sobre el contorn. Així, es fa un experiment per analitzar aquest efecte fixant un eix de rotació en l'objecte a diverses inclinacions ($7.5^\circ, 15^\circ, \dots, 82.5^\circ$) i llavors girant la càmera sobre cadascun d'aquests eixos de 5° a 50° , amb la mateixa distància focal i valors de Z_0 emprats en l'experiment anterior.

La figura 3.6 mostra els resultats obtinguts. Congruent amb la figura 3.4 per a la forma quadrada, l'error en l'estimació de la direcció epipolar és zero per a qualsevol posicionament de la càmera resultant d'una rotació sobre un eix que té una inclinació de 45° . Per a la forma de "H", l'error zero ocorre a valors d'inclinació diferents per a quantitats diferents de rotació. Es pot observar que, per a les dues formes, les rotacions més grans indueixen uns errors de recuperació més petits. Una explicació geomètrica d'aquest efecte és que, començant des d'una posició de l'objecte frontoparal·lela, com més gran és la quantitat de rotació més propera és la configuració dels raigs de projecció a una projecció paral·lela (figura 3.7). Per això, es pot deduir que els efectes de perspectiva que no es poden modelar produeixen errors més grans per a quantitats de rotació petites.

En resum, en el cas ideal l'error en la recuperació de la direcció epipolar a causa de la diferència de profunditat entre els punts és insignificant.

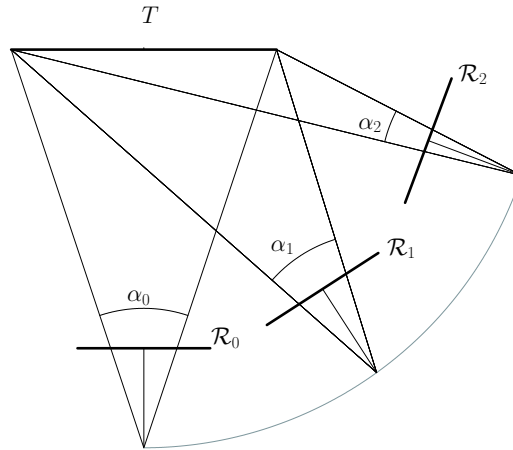


Figura 3.7: Començant amb un objecte frontoparal·lel T , es pren una primera vista \mathcal{R}_0 i llavors dues altres vistes \mathcal{R}_1 i \mathcal{R}_2 ; \mathcal{R}_2 després d'una quantitat més gran de rotació que \mathcal{R}_1 . El màxim angle α_i entre els raigs de projecció disminueix (i, així, la projecció es torna més propera a una projecció paral·lela) a mesura que la quantitat de rotació augmenta.

Translació lateral: projecció del contorn no centrada

Usant el model de càmeres de perspectiva dèbil se suposa que les distàncies dels punts de control del contorn al raig principal són petites, una altra vegada en relació amb la distància Z_0 des de la càmera fins a l'objecte. Aquesta condició es pot satisfer amb un camp de visió reduït i mantenint la projecció de l'objecte projectat centrada en la imatge (suposant que el centre d'imatge estigui a prop del punt principal). Per avaluar els efectes de relaxar la condició que l'objecte estigui centrat, se simulen una sèrie de translacions de la càmera paral·leles al pla de la imatge que s'estenen des de -175 mm fins a 175 mm, de manera que es cobreix la totalitat de l'àrea de la imatge simulada. Es mostren els resultats per a diferents orientacions cada 30° ($0^\circ, 30^\circ, \dots, 150^\circ$) i els casos especials a 45° i 135° .

Comparat amb les altres condicions prèviament avaluades, les translacions de la càmera que ocasionen desplaçaments grans de l'objecte en la imatge porten a errors més grans en el càlcul de la direcció epipolar. Com es pot veure a la figura 3.8, comparant 3.8(a) i 3.8(b), els efectes de forma de contorn són insignificants respecte de l'error provocat per no tenir centrada la projecció de l'objecte en la imatge. A més, els errors de recuperació de la direcció epipolar són més significatius per a translacions al llarg

3. Recuperació de la direcció epipolar

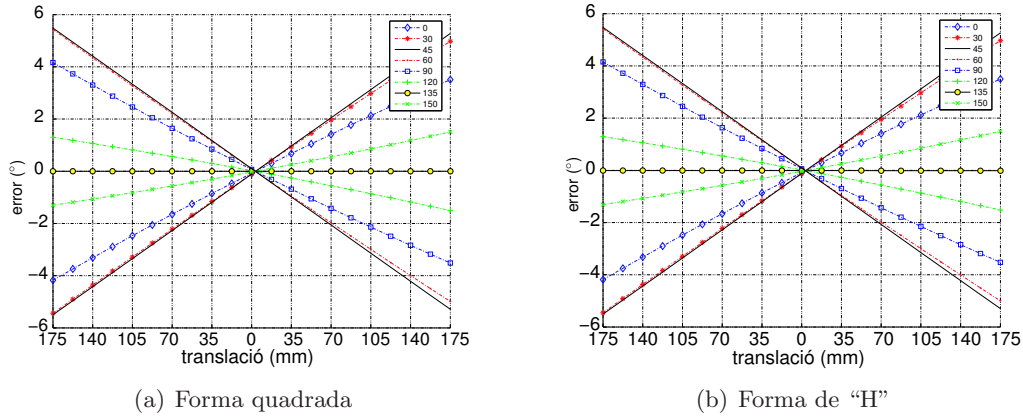


Figura 3.8: Error en la recuperació de la direcció epipolar per a una rotació fixa de 40° sobre un eix inclinat 45° sobre el pla de l'objecte, i amb una distància inicial de 5 m, quan després de la rotació es fan translacions laterals de la càmera al llarg de les direccions llistades a la llegenda.

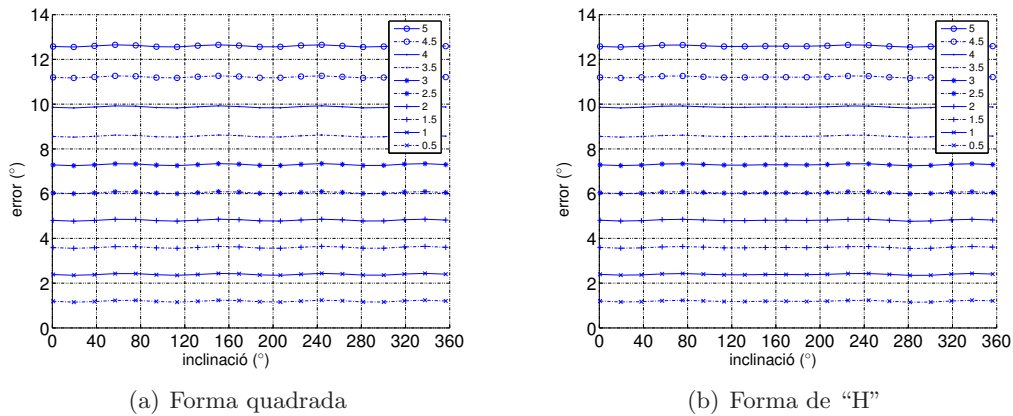


Figura 3.9: Efectes d'introduir ciclorrotació al moviment de la càmera. Cada línia és l'error de recuperació després d'una rotació de la càmera sobre un eix definit sobre l'objecte, i l'orientació de qui varia de 0° a 360° en el pla, i de 0.5° a 5° fora del pla (augmentant la quantitat de ciclorrotació).

de la direcció orthogonal a la direcció epipolar (45°) i insignificants al llarg de la direcció epipolar.

3.3.4 Efectes de la ciclorrotació

La restricció que s'imposa en l'algorisme presentat suposa que no hi ha ciclorrotació en el moviment de la càmera. Com es pot observar fàcilment a la figura 3.1, si es realitza un moviment de la càmera que conté ciclorrotació no es conserva cap direcció en la imatge. El que es pretén és quantificar l'error en la recuperació de la direcció epipolar mentre la quantitat de ciclorrotació present en el moviment augmenta.

La figura 3.9 mostra els errors resultants de repetir el mateix experiment de la secció 3.3.2 però introduint graus diferents de ciclorrotació. Per a cada experiment, la rotació és de 40° sobre un eix a través del centre de l'objecte, l'orientació del qual varia de 0° a 360° , i on la component de ciclorrotació s'estén de 0.5° a 5° . Com era d'esperar, l'error en la direcció epipolar augmenta severament respecte a la quantitat de ciclorrotació, i per tant, cal complir la restricció imposada a l'algorisme.

3.3.5 Sensibilitat al soroll

L'afinitat \mathbf{M} que explica dues vistes s'aproxima a la identitat per a moviments petits, de manera que el càlcul de la direcció epipolar basada en \mathbf{M} hauria de ser molt sensible davant les pertorbacions en les projeccions dels punts de control. Per avaluar aquesta sensibilitat, es duen a terme una sèrie d'experiments considerant rotacions que s'estenen de 5° a 50° sobre un eix orientat a -45° al pla de l'objecte, per al qual l'error de recuperació s'acosta a zero en condicions lliures de soroll (figura 3.6). S'afegeix soroll gaussià amb desviació estàndard igual a múltiples de 0.25 píxels a ambdues components x i y de les projeccions en la imatge dels punts de control, i es fa una simulació de Monte Carlo amb 10000 mostres. Els resultats es poden veure en la figura 3.10. Per a cada forma, hi ha deu gràfiques que corresponen als angles de rotació de 5° a 50° i, dins de cada gràfica, es representen, per a cada nivell de soroll de 0.25 a 1.00 píxels, la direcció epipolar mitjana amb una línia horitzontal, i la seva desviació estàndard com un segment vertical.

Cal notar que, d'acord amb l'efecte il·lustrat en la figura 3.7, les rotacions més grans porten a errors de recuperació més petits. A més, com era d'esperar, les quantitats més altes de soroll produeixen una recuperació més pobre, encara que es pot observar que per a rotacions més grans que 20° la recuperació és bastant bona fins i tot en presència

3. Recuperació de la direcció epipolar

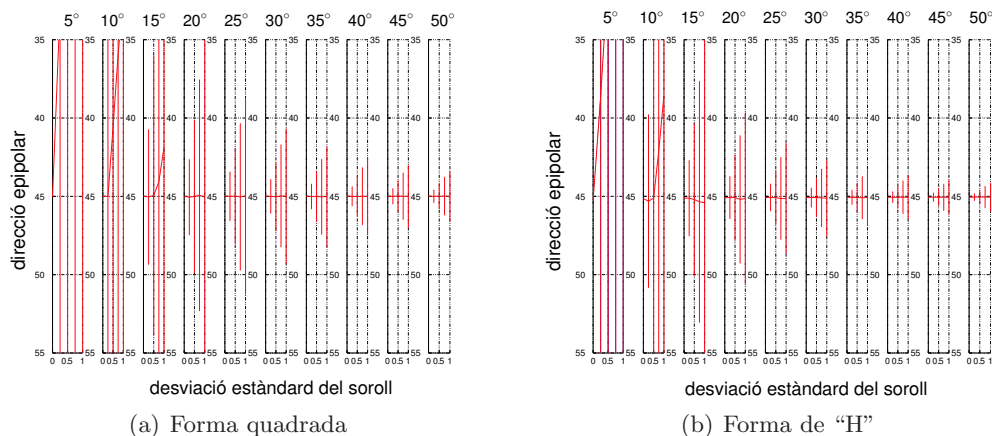


Figura 3.10: Mitjana aritmètica (línia orientada horitzontalment) i desviació estàndard (línia vertical) de les direccions epipolars computades en presència de soroll gaussià amb desviacions estàndards de 0.25 a 1.00 píxels. Els deu gràfics mostrats per a cada forma corresponen a rotacions que s’estenen de 5° a 50° sobre un eix orientat a -45° .

d’un soroll considerable. Al contrari, per a rotacions més petites que 10° , l’algorisme proposat no és fiable en condicions sorolloses. Observeu que els resultats per a la forma de “H” són molt millors, especialment per a petites rotacions, que els obtinguts amb el quadrat, a causa del fet que amb un nombre més gran de punts de control l’estimació de l’afinitat és molt més robusta al soroll.

3.3.6 Comparació amb l’algorisme estàndard d’or

Fins a aquest punt, s’han analitzat els efectes de diversos factors (referent a la forma, el model de càmera emprat...) que influeixen en la precisió de l’algorisme proposat. Ara, i encara en una escena simulada, es vol fer una comparació amb l’algorisme estàndard d’or (EdO) per computar la geometria epipolar afí [62]. Aquest mètode, al contrari que el nostre algorisme, necessita que les correspondències de punts que utilitza no provenguin de punts coplanars per tal de computar amb la màxima versemblança possible l’estimació de la matriu fonamental afí. En teoria, només amb quatre punts no coplanars és suficient per computar la geometria epipolar afí amb l’algorisme EdO. En realitat, el seu rendiment depèn fortament de la quantitat d’informació no coplanària proporcionada, tant pel que fa a la gamma de profunditats com al nombre de punts

utilitzats. L'objectiu d'aquesta secció és establir experimentalment la quantitat d'informació de profunditat exigida per l'EdO per proporcionar resultats de recuperació de direcció epipolar equivalents a l'algorisme proposat.

Primer es dissenya un experiment en el qual s'afegeix una sèrie de dos a dotze punts suplementaris al contorn en forma de "H", amb la qual cosa es fa variar la seva distància respecte al pla de contorn. Els paràmetres de la càmera es fixen a 500 mm de profunditat entre càmera i objecte, i a una distància focal de 767 píxels. Com en l'experiment anterior, el moviment de la càmera s'aconsegueix mitjançant una rotació de 40° sobre un eix posat a una orientació de 45° sobre el pla de l'objecte. Els resultats es mostren a la figura 3.11. Es pot veure que, quan la profunditat d'aquests punts augmenta, l'error en el càlcul de la direcció epipolar disminueix. A més, s'ha observat que el nombre i la localització xy d'aquests punts tenen poc efecte en el resultat del càlcul de la direcció epipolar. El diagrama conté la representació dels errors que resulten en el càlcul de la direcció epipolar afí amb l'algorisme EdO per a diferents nombres de punts fora del pla, i un llindar de tall on es mostra l'error en la recuperació de la direcció epipolar utilitzant l'algorisme proposat en les mateixes condicions experimentals (evidentment, els punts addicionals fora del pla del contorn no s'utilitzen en l'algorisme proposat). Com es mostra en el diagrama, per a les condicions experimentals donades, els resultats de l'algorisme proposat són comparables als resultats de l'algorisme EdO quan els punts extrems es posen aproximadament a una distància igual a la mida de l'objecte (120 mm en aquest cas).

Cal observar la importància de la paral·laxi en el càlcul de la matriu fonamental afí en l'algorisme EdO. A mesura que els punts de l'objecte s'acosten a una configuració coplanar, el vector de paral·laxi, que determina la direcció epipolar, es redueix monotònicament en llargada. Conseqüentment, la precisió de la direcció de la recta es redueix de la mateixa manera, i la covariància d'una matriu fonamental afí aproximada augmenta. Aquesta situació no ocorre en l'algorisme proposat, ja que s'ha ideat precisament per computar la direcció epipolar afí des de dues vistes d'un objecte pla.

El segon experiment avalua la sensibilitat al soroll de les dues tècniques per computar la direcció epipolar afí, i consisteix a repetir el mateix experiment diverses vegades afegint soroll gaussià a les projeccions de polígon de control. L'experiment es realitza amb l'algorisme proposat utilitzant la forma de "H" planar, i per a l'algorisme d'EdO

3. Recuperació de la direcció epipolar

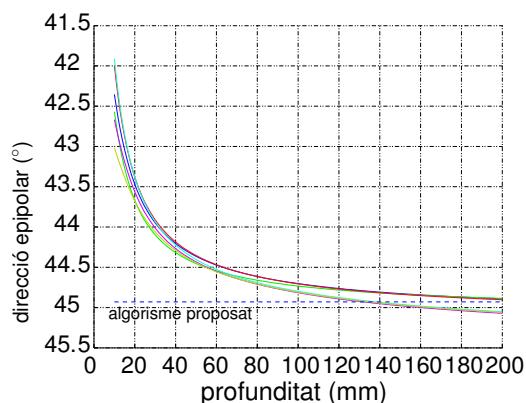


Figura 3.11: Direcció epipolar computada amb l'algorisme EdO en el cas que es posen 2, 4... 12 punts fora del pla (una corba per a cada nombre) situats a profunditats creixents damunt el contorn en forma de "H". Observeu que el factor més important no és el nombre de punts, sinó la profunditat a què es posen.

cas	soroll (píxels)					
		0.00	0.25	0.50	0.75	1.00
contorn planar "H" amb l'algorisme proposat	μ	-44.97	-44.95	-44.96	-45.04	-45.07
	σ	0	0.193	0.492	0.552	0.876
contorn planar "H" i 2 punts 120mm fora del pla (EdO)	μ	-43.76	-43.77	-43.78	-43.74	-43.85
	σ	0	0.143	0.245	0.436	0.486
2 capes del contorn "H" a 0mm i 120mm (EdO)	μ	-44.91	-44.91	-44.89	-44.92	-44.91
	σ	0	0.068	0.140	0.210	0.324
Valors de profunditat aleatoris per als punts de polígon de control (EdO)	μ	-44.81	-45.75	-45.03	-44.51	-44.84
	σ	0	0.294	0.486	0.712	0.625

Taula 3.1: Recuperació de direcció epipolar utilitzant l'algorisme proposat i l'algorisme EdO per a nivells creixents de soroll gaussià aplicats al punts en la imatge. La rotació real de la càmera és de 40° sobre un eix a 45° centrat en l'objecte.

utilitzant: a) el contorn "H" planar i dos punts fora del pla a una distància de 120 mm, b) 24 punts dividits en dos estrats, formant dues formes de "H" a distàncies diferents, i c) a profunditat aleatòria dels 12 punts que formen la "H". La taula 3.1 mostra una comparació dels resultats obtinguts per a una desviació estàndard creixent de 0 a 1 píxel en el soroll expressant en coordenades d'imatge.

Es pot observar que els resultats més coherents s'obtenen amb l'algorisme proposat

i amb l'experiment EdO que utilitza dos estrats de punts de control, mentre que l'EdO amb només dos punts fora del pla produeix una estimació esbiaixada. Tanmateix, no s'observen grans diferències en la precisió en condicions sorolloses.

3.3.7 Experiments amb imatges reals

Presentem ara els resultats obtinguts emprant seqüències d'imatges en una escena controlada. L'objectiu d'aquest treball no és a seguir l'objecte, sinó a computar l'afinitat i utilitzar-la per calcular la direcció epipolar que explica les dues vistes. És per això que es facilita la fase de seguiment fent avançar l'objecte simple de la figura 3.12 (és clar que en situacions reals no és usual trobar objectes tan ben definits!). L'objecte es munta en l'element terminal d'un braç manipulador, de manera que se sap exactament el moviment que es realitza, i el treball es focalitza en avaluar la precisió de la direcció recobrada en situacions diferents. Es pot veure fàcilment que un moviment de l'objecte amb el braç equival a un moviment de la càmera.

L'escenari de l'experiment consta d'un manipulador RX60 Stäubli que sosté el dibuix d'un objecte al seu efector final. Aquest objecte és una figura planar artificial en forma de "H" que combina cantonades i línies rectes i corbes, que es poden seguir fàcilment amb el nostre algorisme de seguiment de contorns actius. Les imatges s'adquireixen utilitzant una càmera *firewire* Sony DWL500 amb una distància focal calibrada de 767 píxels. La zona de treball del braç està molt limitada (comparada amb un robot mòbil, per exemple), i això restringeix el repertori d'experiments en què podem actuar, però convé pagar aquest preu a canvi de la precisió que s'obté en el moviment. La distància inicial des de la càmera fins a l'objecte s'ha establert en 500 mm. Això correspon al cas extrem que es descriu a la secció 3.3.3, figura 3.5 i, per això, s'està provant l'algorisme proposat en condicions relaxades de perspectiva dèbil, com les estudiades en simulació a la secció 3.3.3. Les imatges adquirides tenen evidents efectes de perspectiva, com es pot observar a les figures 3.12 i 3.13.

El primer experiment inclou el moviment equivalent de la càmera provocat per una rotació de 40° sobre un eix en l'objecte en diversos angles d'inclinació mostrejats a intervals de 15° . D'aquesta manera, es poden comparar amb els resultats de la figura 3.5(b) en cas que la distància és igual a 500mm . Es fan cinc experiments, tots començant des de la posició frontoparal·lela mostrada en la figura 3.12(a). El contorn

3. Recuperació de la direcció epipolar

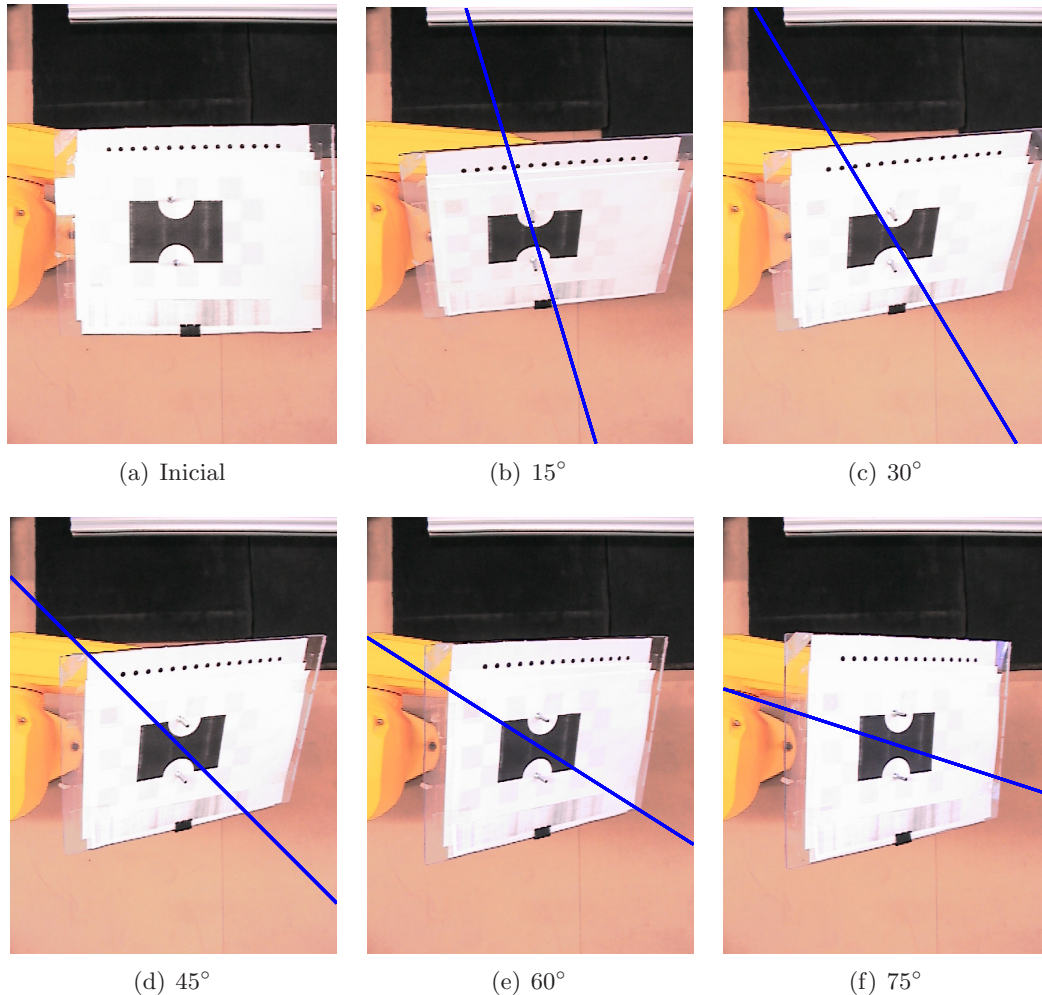


Figura 3.12: El primer experiment amb imatges reals correspon al càlcul de la direcció epipolar entre parells de vistes inclosa la imatge inicial i cadascuna de les altres cinc. Corresponen a rotacions de 40° de la càmera sobre un eix en l'objecte amb inclinacions mostrejades a intervals de 15° . La direcció epipolar computada es mostra gràficament com la recta que passa a través del centre de l'objecte.

se segueix fins a cadascuna de les vistes finals mostrades en les posicions restants de la figura 3.12(a) que generen una seqüència de direccions epipolars. La direcció epipolar computada al final del moviment per l'algorisme proposat es mostra gràficament com una línia recta a través del centre de l'objecte.

La taula 3.2 presenta els valors numèrics obtinguts en el càlcul de la direcció epipolar. La desviació estàndard es computa adquirint 300 imatges en la posició final,

direcció epipolar	-15	-30	-45	-60	-75
$\bar{\theta}$	-16.63	-31.01	-45.00	-57.63	-72.04
σ	0.14	0.09	0.14	0.19	0.13

Taula 3.2: Mitjana aritmètica i desviació estàndard en graus de la direcció epipolar computada per l'algorisme proposat en imatges reals.

calculant els corresponents vectors de forma i computant cadascuna de les direccions epipolars. Cal notar que totes les desviacions estàndards són molt similars i que els valors mitjans es desvien més del valor real a mesura que l'angle se separa de 45° d'inclinació. Això s'hauria d'interpretar, considerant també la figura 3.5, com l'error introduït a causa dels efectes de perspectiva que no poden ser modelats per la càmera de perspectiva dèbil. De fet, en condicions veritables de perspectiva dèbil, els errors són molt més baixos, com es veu amb l'encongiment de les corbes d'error a la figura 3.5, quan la distància Z_0 de la càmera a l'objectiu augmenta.

També s'analitzen tres seqüències addicionals relaxant altres condicions de perspectiva dèbil, de manera similar al cas simulat. La primera seqüència, etiquetada “no centrada”, comença en la posició inicial frontoparal·lela (figura 3.13(a)) i acaba en una posició descentrada, després d'una translació de 100 mm al llarg de l'eix x del sistema de coordenades del robot, i d'una rotació de 40° sobre un eix a 45° d'inclinació (figura 3.13(b)). De manera coherent amb els resultats simulats (vegeu la secció 3.3.3), aquesta translació lateral de la càmera és, de bon tros, la violació de les condicions de perspectiva dèbil que té l'efecte més important sobre el càlcul de la direcció epipolar. Vegeu els valors numèrics en la Taula 3.3, primera fila.

El segon experiment, etiquetat “no frontoparal·lel A”, correspon a la mateixa rotació descrita en l'experiment anterior, però la posició inicial no és frontoparal·lela. La seqüència comença amb l'objecte rotat 20° , com es mostra a la figura 3.13(c) i, després d'una altra rotació de 20° , acaba a 40° (figura 3.13(d)). Es pot veure que el resultat és una mica pitjor que el de l'experiment anterior, però amb una desviació estàndard similar.

Finalment, l'últim experiment, etiquetat “no frontoparal·lel B”, correspon a una seqüència amb el contorn inicialment no frontoparal·lel, rotat 40° (figura 3.13(d)) i

3. Recuperació de la direcció epipolar

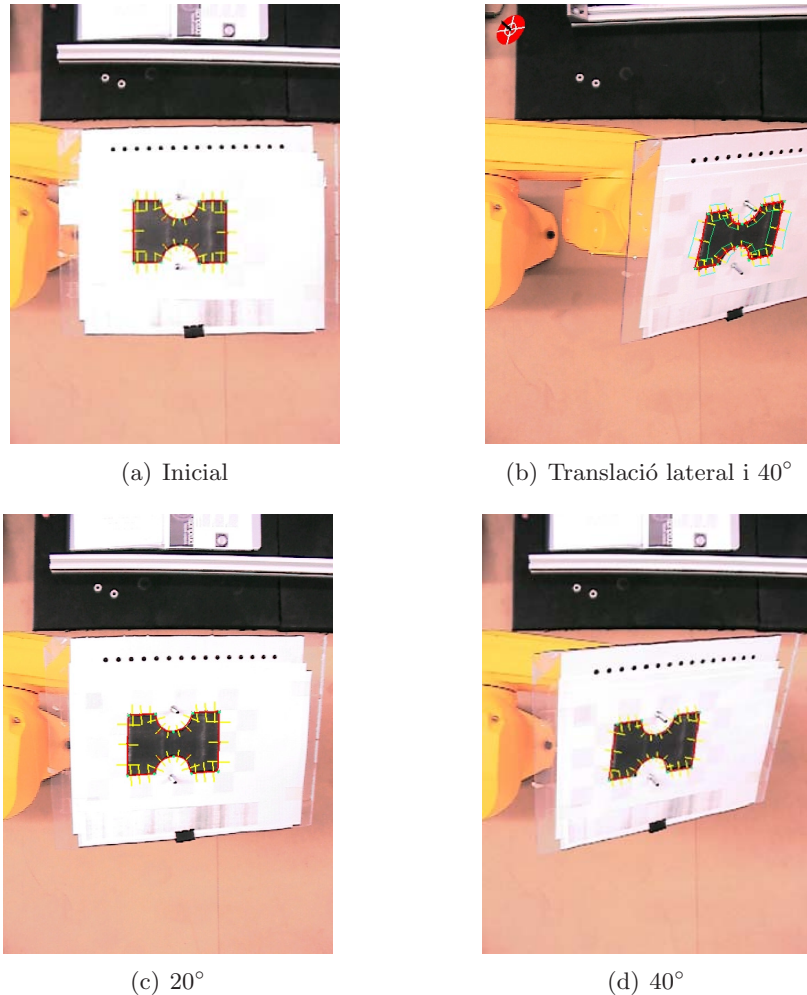


Figura 3.13: Experiment amb imatges reals que relaxen diferents condicions de perspectiva dèbil. La primera seqüència, que tracta un objecte descentrat, comença a (a) i acaba a (b). La seqüència següent comença en una posició no frontoparal·lela a (c) i acaba a (d). L'última seqüència, que prova també una inicialització no frontoparal·lela, comença a (d) i finalitza a (a).

una posició final frontoparal·lela (figura 3.13(a)). El resultat és molt similar al de l'experiment anterior.

3.3.8 Un comentari: l'interès de reduir l'espai de forma

Utilitzant un objecte frontoparal·lel i centrat, la matriu \mathbf{M} a (2.7) resulta simètrica. Podem explotar aquest fet per reduir de 6 a 5 la dimensió del vector de forma (2.13)

<i>Frames</i>	θ	σ
No centrat	-34.65	0.13
No frontoparal·lel A	-43.89	0.09
No frontoparal·lel B	43.96	0.10

Taula 3.3: Mitjana aritmètica i desviació estàndard de la direcció epipolar computada sobre imatges reals quan es relaxen algunes condicions de la perspectiva dèbil.

utilitzat per parametritzar les deformacions del contorn

$$\mathbf{W} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} Q^x \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ Q^y \end{bmatrix}, \begin{bmatrix} Q^y \\ Q^x \end{bmatrix} \right). \quad (3.8)$$

Reproduint els experiments de la secció 3.3.2 amb aquesta matriu de forma reduïda s'obtenen solucions exactament perpendiculars per a la direcció epipolar i per a l'eix de rotació. Això és així perquè la part vectorial de la deformació de perspectiva dèbil es força a ser simètrica, fins i tot amb els efectes de perspectiva presents. Com a resultat, els errors computats són els mateixos per a les dues direccions. Les figures 3.14(a) i 3.14(b) mostren els valors tant de la direcció epipolar com de l'eix de rotació, però les dues corbes estan sobreposades una sobre de l'altra. Observeu que els angles en els quals l'error és 0° es conserven. Això pot no suposar un gran benefici quan les condicions de visualització de perspectiva dèbil es compleixen, però quan els efectes de perspectiva són forts, en reduir l'espai de forma disminueixen en gran manera les distorsions que aquests efectes provoquen en la recuperació de la direcció epipolar.

3.3.9 Consideracions d'aplicabilitat

En tota aquesta secció experimental s'ha fet un esforç per relaxar gradualment les condicions de visualització fora del model de perspectiva dèbil amb l'objectiu d'avaluar la validesa de l'algorisme proposat fins i tot en condicions adverses. A la secció 3.3.3 s'ha vist que l'error introduït per a translacions petites al llarg de l'eix òptic és insignificant, i a la secció 3.3.3 s'ha constatat que la no-planaritat de l'objecte no s'exigeix realment en condicions normals, ja que és petita en comparació amb la distancia entre càmera i objecte. Contràriament, a la secció 3.3.3 s'ha mostrat que les translacions laterals que envien les projeccions del contorn fora del centre de la imatge introdueixen errors

3. Recuperació de la direcció epipolar

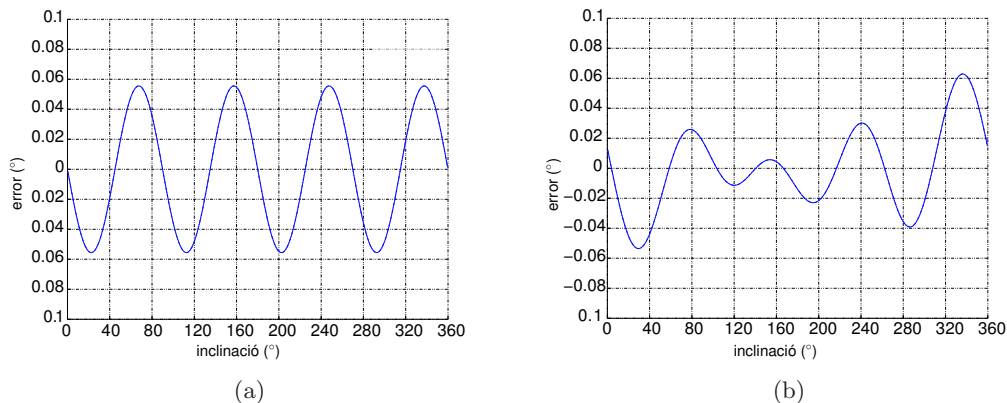


Figura 3.14: Les mateixes figures que a la figura 3.4, però utilitzant un vector de forma 5D en comptes d'un de 6D.

crítics. En una aplicació realista, es necessita assegurar que l'objecte es projecti sempre al centre de la imatge. Per fer això, s'hauria d'utilitzar un sistema de visió activa que s'ocupés de mantenir el contorn centrat en la imatge, independentment del moviment de la càmera (als capítols 6 i 7 es desenvolupa aquesta idea).

A més, a la secció 3.3.5 s'ha vist que la direcció epipolar no es pot computar correctament per a rotacions petites en presència de soroll. En la implementació amb les seqüències d'imatges reals i el braç robòtic, l'estimació acurada de la direcció epipolar es pot obtenir a partir de rotacions de 15° o, equivalentment, una vegada la càmera es trasllada lateralment aproximadament una cinquena part de la seva distància a l'objecte mentre es manté la projecció de l'objecte centrada en la imatge.

Finalment, si podem suposar que l'objecte és frontoparal·lel en alguna imatge de la seqüència de moviment, es pot considerar l'ús de la matriu reduïda de forma definida a la secció 3.3.8.

3.4 Conclusions

El resultat principal és que, en condicions de perspectiva dèbil i assumint que el moviment de la càmera està lliure de ciclorrotació, la direcció epipolar es pot recuperar a partir de l'afinitat que explica dues vistes d'una escena planar.

Per estudiar el rendiment del mètode proposat s'han dissenyat dos conjunts d'experiments. En els dos, l'afinitat que explica les dues vistes s'obté seguint un contorn planar des d'una vista fins a l'altra, i llavors es calculen els paràmetres de l'afinitat a partir de la deformació observada del contorn, codificada com un vector de forma 6D (o 5D). Primer, s'han utilitzat imatges sintètiques per avaluar els resultats en un ambient controlat de soroll, relaxant una per una les restriccions del model de visualització, per llavors comparar la precisió del nostre algorisme amb l'algorisme de referència, que necessita correspondències de punts no coplanars.

El resultat de la comparació ha estat molt encoratjador, ja que amb menys informació d'escena (només amb un pla) i amb un processament molt més simple (resolent una senzilla equació de segon grau), es pot obtenir la direcció epipolar amb precisió similar. Controlant la informació de profunditat de l'escena s'ha establert la frontera en què els dos algorismes comparats tenen un comportament semblant, que és a partir d'on l'algorisme proposat té un millor rendiment que l'EdO. És important recordar, tanmateix, que el mètode proposat és menys general, en el sentit que exigeix un moviment de la càmera lliure de ciclorrotació.

El segon conjunt d'experiments ha utilitzat moviments ben coneguts per adquirir les seqüències d'imatges utilitzades per validar l'algorisme proposat en condicions realistes. Atès que l'objectiu és mostrar la recuperació de la direcció epipolar, i no demostrar la robustesa del seguiment de contorns, s'han emprat seqüències d'imatges simples i amb moviments ben coneguts per tenir una bona base per a la comparació amb les direccions epipolars reals. En aquests experiments, les condicions de visualització han estat expressament forçades, amb efectes perspectius evidents, per tal de comprovar la validesa del mètode fins i tot en aquestes condicions extremes.

Els experiments realitzats han mostrat que la direcció epipolar, més enllà del seu interès natural de relacionar dues vistes, és una bona mesura per avaluar la qualitat de l'afinitat computada amb els contorns actius. Fins al moment, l'única manera d'obtenir resultats intuïtius de la qualitat del procés de modelatge del moviment segons una matriu de forma era computant la posa 6D sencera i, tot i que també es pot obtenir una matriu de covariància associada a la posa, és un procés ferregós. Tot i que un estudi en profunditat de la precisió requereix utilitzar la posa sencera (vegeu els capítols 4 i 5), la direcció epipolar és un bon indicador de la qualitat de l'afinitat que s'obté, i permet

3. Recuperació de la direcció epipolar

fer una primera aproximació al problema de determinar la precisió de l'algorisme, ja que es pot relacionar l'error comès amb diferents situacions de relaxació de les condicions de perspectiva dèbil, d'efectes perspectius, de soroll en la imatge...

Capítol 4

Estudi analític de la propagació de l'error

«What you have told us is rubbish. The world is really a flat plate supported on the back of a giant turtle.» The scientist gave a superior smile before replying, «What is the turtle standing on?» «You're very clever, young man, very clever,» said the old lady. «But it's turtles all the way down!»

«El que ens heu dit són ximpleries. En realitat, el món és una safata plana sostinguda sobre la closca d'una tortuga gegant». El científic esbossà un somriure de superioritat abans de replicar: «I sobre què s'aguanta la tortuga?». «Ah, sou molt llest, jove, molt llest, -digué la dama-, però hi ha tortugues fins a baix de tot!».

Stephen W. Hawking, *A brief history of time.*

Extracte

Al capítol anterior s'ha mostrat que la direcció epipolar es pot utilitzar com a indicador global de qualitat del vector de forma que codifica la deformació d'un contorn corresponent a un moviment del robot. Un estudi en profunditat de la precisió requereix utilitzar la posa sencera computada a partir del vector de forma. Del procés d'obtenció del vector de forma, amb un filtre de Kalman, s'obté també una covariància associada, que està expressada en espai de forma. En aquest capítol es desenvolupen les expressions analítiques que permeten propagar aquesta covariància de l'espai de forma a l'espai de posa.

4.1 Introducció

Al capítol 2 s'ha vist com es pot obtenir el moviment d'una càmera a partir de la parametrització \mathbf{S} de la deformació d'un contorn actiu en l'espai de forma. També s'ha observat la necessitat de trobar una manera de calcular la incertesa de cada estimació del moviment. Resulta interessant tant per saber quina confiança es té en l'estimació del moviment, com per poder implementar estratègies de fusió i col·laboració amb altres sensors. Amb el filtre de Kalman es té una estimació de la covariància Σ_S , però tal com s'ha vist aquesta covariància està expressada en espai de forma. Cal, doncs, trobar un algorisme per propagar aquesta covariància de l'espai de forma a l'espai 3D.

La propagació de la incertesa es pot abordar bàsicament des de dos punts de vista: l'analític i l'estadístic. Un mètode analític té l'avantatge que troba una expressió tancada, amb els beneficis que això comporta. El problema és que arribar a la solució a vegades comporta fer aproximacions que introdueixen un biaix en el resultat, però s'obté un algorisme generalment ràpid. En canvi, els mètodes estadístics són més senzills d'implementar, però generalment són més costosos de computar. Un estudi estadístic de la propagació de la incertesa es duu a terme al capítol 5.

El mètode analític considera que entre les dades d'entrada i les de sortida hi ha una funció f . En el cas general, aquesta funció no té perquè ser lineal, però ha de ser possible estimar-ne el jacobià. Aquesta relació es pot aproximar per un desenvolupament en sèrie de Taylor fins a primer ordre. Aquesta aproximació s'ha fet servir, per exemple, per estimar paràmetres [31, 37, 148], per millorar els resultats en sistemes de cerca guiada [139], i recentment per localitzar robots [60, 123]. Aquest mètode té l'avantatge de ser ràpid de computar, però cal un esforç previ de derivació més gran en comparació amb els mètodes estadístics. Té el desavantatge que només s'aproxima l'error, de manera que sempre hi ha un error residual que no es pot eliminar.

L'algorisme que cal analitzar, que computa la posa a partir de l'afinitat que parametriza la deformació afí d'un contorn, fa servir els valors i vectors propis de la matriu \mathbf{M} de l'afinitat. En la implementació de l'algorisme, els valors propis necessaris per solucionar l'equació 2.15 es troben amb el mètode de la descomposició en valors singulars (SVD). Per estimar la propagació de la incertesa en aquest cas, una possibilitat és estimar les derivades parcials de les tres matrius resultat de SVD respecte de cada

element de la matriu original. Es pot computar el jacobià de la SVD [100, 109], fins i tot quan les matrius estan en casos degenerats. A [109] s'aplica la derivació trobada per propagar l'error en tres problemes comuns de la visió artificial: autocalibració, estimació de la geometria epipolar i estimació 3D de moviment rígid. A vegades, però, és més útil trobar directament l'expressió de la propagació de les incerteses en els valors i vectors propis. Hi ha estudis clàssics basats en la sensibilitat de vectors i valors singulars a pertorbacions en la matriu original [131, 146]. En aquesta línia, Weng [144, 145], basant-se en aproximacions de primer ordre, estableix la relació que hi ha entre la incertesa de la matriu original i la incertesa en els seus vectors i valors propis, i aplica aquesta derivació a l'estudi de l'error en un algorisme d'estimació del moviment rígid d'una càmera. En l'algorisme que cal analitzar la posa es calcula a partir dels valors i vectors propis combinats algebraicament amb elements de l'afinitat original. Per tant, només amb aquests mètodes no és suficient per conèixer la incertesa de la posa.

Un altre camí possible és emprar àlgebres de Lie [116]. Drummond i Cipolla [48] les utilitzen per estimar el jacobià de tota la transformació no lineal que passa de l'espai de deformacions afins en la imatge a l'espai de moviments 3D, i ho apliquen al servocontrol visual. Desafortunadament, aquesta aproximació és vàlida només en determinades condicions i per a moviments infinitesimals. El jacobià obtingut en les condicions més generals del nostre algorisme resulta singular i, atès que no és invertible, no es pot aplicar.

Es proposa un algorisme que divideix la propagació de la incertesa en dues etapes. En primer lloc, es propaga a través de la SVD per tal de trobar les expressions de la incertesa associades als valors i vectors propis (secció 4.2). En segon lloc, es pren l'expressió que serveix per trobar cadascun dels paràmetres del moviment 3D i es fa servir l'aproximació de primer ordre per trobar l'expressió final de la propagació de la incertesa per a cadascun dels elements de la posa (secció 4.3). L'algorisme es formalitza a la secció 4.4 i es mostren els resultats d'uns experiments fets amb simulació. Finalment s'exposen les conclusions (secció 4.5).

4.2 Propagació als vectors i valors propis

Coneguda la incertesa d'una matriu, el teorema de Weng *et. al.* [144, 145] descrit a continuació mostra com es pot aproximar fins a primer ordre la incertesa dels valors i vectors propis de la matriu. Més endavant, Earnshaw i Blostein [50] van estendre la derivació del teorema per aproximar la incertesa fins a segon ordre.

El resultat útil del teorema (vegeu l'apèndix B) és el següent: sigui $\mathbf{A} = [a_{ij}]$ una matriu diagonalitzable i simètrica $n \times n$, i Σ_A una matriu d'incertesa, considerant una aproximació de primer ordre es pot expressar la incertesa al vector propi més petit com

$$\boldsymbol{\delta}_{v_1} = \mathbf{H}\boldsymbol{\Delta}_1\mathbf{H}^T\Sigma_A\mathbf{v}_1 \quad (4.1)$$

on $\mathbf{H} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ és la matriu de vectors propis i $\boldsymbol{\Delta}_1 = \text{diag}(0, (\lambda_1 - \lambda_2)^{-1}, \dots, (\lambda_1 - \lambda_n)^{-1})$ és la matriu associada al valor propi més petit. La incertesa d'aquest valor propi s'expressa com

$$\delta_{\lambda_1} \cong \mathbf{v}_1^T \Sigma_A \mathbf{v}_1 \quad (4.2)$$

La matriu d'incertesa Σ_A es pot reescriure en forma vectorial $\boldsymbol{\delta}_A$ apilant-ne totes les columnes. D'aquesta manera, es pot reescriure (4.1) com

$$\boldsymbol{\delta}_{v_1} \cong \mathbf{H}\boldsymbol{\Delta}_1\mathbf{H}^T\Sigma_A\mathbf{v}_1 = \mathbf{H}\boldsymbol{\Delta}_1\mathbf{H}^T[v_1 I_n v_2 I_n \dots v_n I_n] \boldsymbol{\delta}_A \triangleq \mathbf{G}_{v_1} \boldsymbol{\delta}_A \quad (4.3)$$

on v_i són les components del vector propi, \mathbf{I}_n és la identitat de dimensió n , i $\boldsymbol{\delta}_A$ és la versió vectorial de la matriu d'incertesa Σ_A de la matriu \mathbf{A} .

De la mateixa manera es pot reescriure (4.2) com

$$\delta_{\lambda_1} \cong \mathbf{v}_1^T \Sigma_A \mathbf{v}_1 = \mathbf{v}_1^T [v_1 I_n v_2 I_n \dots v_n I_n] \boldsymbol{\delta}_A \triangleq \mathbf{P}_{\lambda} \boldsymbol{\delta}_A \quad (4.4)$$

La incertesa del vector propi \mathbf{v}_2 associat al segon valor propi λ_2 és

$$\boldsymbol{\delta}_{v_2} \cong \mathbf{H}\boldsymbol{\Delta}_2\mathbf{H}^T\Sigma_A\mathbf{v}_2$$

on

$$\boldsymbol{\Delta}_2 = \text{diag}((\lambda_2 - \lambda_1)^{-1}, 0, (\lambda_2 - \lambda_3)^{-1}, \dots, (\lambda_2 - \lambda_n)^{-1})$$

De \mathbf{M} a \mathbf{MM}^T

El teorema esmentat explica com es propaga la incertesa d'una matriu \mathbf{M} als seus valors i vectors propis. Però la matriu que es fa servir en els càlculs de la recuperació del moviment 3D és \mathbf{MM}^T i la incertesa que es coneix, $\boldsymbol{\Sigma}_M$, és la de la matriu \mathbf{M} . Per tant, cal veure com es relacionen $\boldsymbol{\Sigma}_M$ i $\boldsymbol{\Sigma}_{\mathbf{MM}^T}$.

Es fa un canvi de variable que permet expressar de manera més compacta el resultat. Si es defineix $\mathbf{N} \triangleq \mathbf{M}^T$, llavors $\mathbf{A} = \mathbf{MM}^T = \mathbf{N}^T\mathbf{N}$. El soroll que pertorba aquesta matriu es pot aproximar linealment i s'expressa com

$$\boldsymbol{\Sigma}_A = \boldsymbol{\Sigma}_{\mathbf{N}^T\mathbf{N}} \cong \mathbf{N}^T\boldsymbol{\Sigma}_N + \boldsymbol{\Sigma}_N^T\mathbf{N}$$

En forma matricial això és:

$$\begin{pmatrix} \delta_{11}^A & \delta_{12}^A \\ \delta_{21}^A & \delta_{22}^A \end{pmatrix} = \begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix}^T \begin{pmatrix} \delta_{11}^N & \delta_{12}^N \\ \delta_{21}^N & \delta_{22}^N \end{pmatrix} + \begin{pmatrix} \delta_{11}^N & \delta_{12}^N \\ \delta_{21}^N & \delta_{22}^N \end{pmatrix}^T \begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix}.$$

Posant les matrius en forma vectorial i reorganjant els termes es pot escriure

$$\begin{pmatrix} \delta_{11}^A \\ \delta_{12}^A \\ \delta_{21}^A \\ \delta_{22}^A \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} N_{11} & 0 \\ 0 & N_{11} \end{pmatrix} + \begin{pmatrix} N_{11} & 0 \\ N_{12} & 0 \end{pmatrix} & \begin{pmatrix} N_{21} & 0 \\ 0 & N_{21} \end{pmatrix} + \begin{pmatrix} N_{21} & 0 \\ N_{22} & 0 \end{pmatrix} \end{pmatrix} \begin{pmatrix} \delta_{11}^N \\ \delta_{12}^N \\ \delta_{21}^N \\ \delta_{22}^N \end{pmatrix}$$

i tenint en compte que $\mathbf{N} \triangleq \mathbf{M}^T$, resulta

$$\begin{pmatrix} \delta_{11}^A \\ \delta_{12}^A \\ \delta_{21}^A \\ \delta_{22}^A \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} M_{11} & 0 \\ 0 & M_{11} \end{pmatrix} + \begin{pmatrix} M_{11} & 0 \\ M_{21} & 0 \end{pmatrix} & \begin{pmatrix} M_{12} & 0 \\ 0 & M_{12} \end{pmatrix} + \begin{pmatrix} M_{12} & 0 \\ M_{22} & 0 \end{pmatrix} \end{pmatrix} \begin{pmatrix} \delta_{11}^M \\ \delta_{12}^M \\ \delta_{21}^M \\ \delta_{22}^M \end{pmatrix}.$$

De forma compacta, cada element del vector de soroll és

$$\boldsymbol{\delta}_A = \mathbf{G}_A\boldsymbol{\delta}_M, \tag{4.5}$$

on \mathbf{G}_A pot reescriure's com:

$$\mathbf{G}_A = [\mathbf{F}_{ij}] + [\mathbf{G}_{ij}],$$

essent $[\mathbf{F}_{ij}]$ i $[\mathbf{G}_{ij}]$ matrius 2×2 , on cada element \mathbf{F}_{ij} i \mathbf{G}_{ij} és a la vegada una submatriu 2×2 . En concret, $\mathbf{F}_{ij} = \mathbf{M}_{ij}\mathbf{I}_2$ i \mathbf{G}_{ij} és una matriu en què en la columna i -èsima hi ha el vector \mathbf{M}_j i en la resta de columnes, zeros. Aquestes equacions són lleugerament diferents de les trobades per Weng [145] atès el canvi de variable realitzat.

4. Estudi analític de la propagació de l'error

Ara, observant (4.5) es pot trobar l'expressió final de la incertesa als vectors propis de $\mathbf{A} = \mathbf{M}\mathbf{M}^T$ respecte de la incertesa coneguda de la matriu \mathbf{M} , si es reescriu (4.3):

$$\boldsymbol{\delta}_{v_1} = \mathbf{G}_{v_1} \boldsymbol{\delta}_{MM^T} = \mathbf{G}_{v_1} \mathbf{G}_{MM^T} \boldsymbol{\delta}_M \triangleq \mathbf{D}_{v_1} \boldsymbol{\delta}_M. \quad (4.6)$$

De la mateixa manera, (4.4) es reescriu com

$$\boldsymbol{\delta}_{\lambda_1} = \mathbf{P}_{\lambda_1} \boldsymbol{\delta}_{MM^T} = \mathbf{P}_{\lambda_1} \mathbf{G}_{MM^T} \boldsymbol{\delta}_M \triangleq \mathbf{Q}_{\lambda_1} \boldsymbol{\delta}_M. \quad (4.7)$$

4.3 Aproximació de primer ordre

Un cop s'ha trobat l'expressió de la incertesa dels valors i vectors propis, el que cal és propagar-la pel que resta de l'algorisme d'estimació del moviment, que són funcions derivables (2.16-2.18 i 2.19-2.21), per a les quals és possible calcular el jacobià. El mètode més comú per trobar una aproximació de la matriu de covariància $\boldsymbol{\Sigma}_y$ del resultat d'aplicar la funció $\mathbf{y} = f(\mathbf{x})$ és fer servir [32, 51]:

$$\boldsymbol{\Sigma}_y = \frac{\partial f}{\partial \mathbf{x}^0} \boldsymbol{\Sigma}_x \frac{\partial f}{\partial \mathbf{x}^0}{}^T. \quad (4.8)$$

Per poder aplicar aquest mètode cal que la funció $f(\mathbf{x})$ sigui explícita i diferenciable, i cal conèixer la matriu de covariància $\boldsymbol{\Sigma}_x$ de les dades. Clarke [32] cita el teorema en què es basa aquesta afirmació i en mostra alguns exemples, i la seva demostració es pot trobar a [24, p. 68].

Aquest mètode no es pot aplicar a l'algorisme proposat per computar la posa a partir de les deformacions afins, ja que no hi ha una funció explícita, el que es té és un procediment o algorisme. A més, no és diferenciable atès que el procediment inclou el fet d'escollir el més petit dels valors propis obtinguts.

Com s'ha vist, per calcular la posa es fan servir els valors i vectors propis de la SVD de la matriu \mathbf{M} . A la secció anterior s'ha mostrat com es pot propagar la incertesa d'una matriu als valors i vectors propis de la seva descomposició en valors singulars. S'observa que cada element de la posa es calcula com a combinació d'aquests valors i vectors propis, i d'elements del vector de forma original. En aquest cas, no es té la matriu de covariància de tots aquests paràmetres, i per tant, tampoc no és factible fer servir directament (4.8).

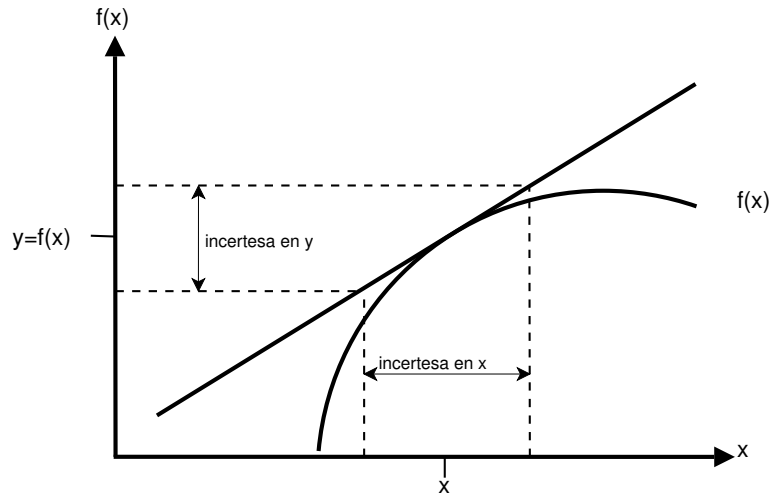


Figura 4.1: Interpretació gràfica de la incertesa

Una de les tècniques d'aproximació més utilitzades es basa en l'expansió en sèrie de Taylor d'una funció. A vegades, aplicada a la propagació de la incertesa s'anomena *mètode dels moments*, ja que es propaga i analitza la incertesa fent servir la mitjana (primer moment), la desviació típica (segon moment) i a vegades moments d'ordre més gran (coeficient d'asimetria, coeficient de curtosi...) de les distribucions de probabilitat. Una de les formes més utilitzades s'anomena *primer-ordre segon-moment* (FOSM¹), ja que s'estima fins al segon ordre de la distribució amb una aproximació de primer ordre de la funció. Es poden fer aproximacions d'ordre més gran, però no és gaire freqüent, ja que els càlculs són bastant més complexos i cal conèixer prèviament els moments d'ordre elevat de les dades d'entrada, cosa que no sempre és factible.

El mètode dels moments es pot aplicar per trobar la incertesa de cadascun dels paràmetres de la posa 3D un cop coneguda la incertesa dels valors i vectors propis de la SVD de la matriu \mathbf{M} .

4. Estudi analític de la propagació de l'error

Una sola variable

Si una variable x es mesura amb una incertesa δ_x , i es fa servir per calcular una funció $y = f(x)$, la incertesa δ_y és

$$\delta_y \approx \left| \frac{\partial f}{\partial x} \right| \delta_x. \quad (4.9)$$

Aquesta equació es pot explicar geomètricament d'una manera intuïtiva observant la figura 4.1. Si es coneix la funció $y = f(x)$ se'n pot fer la representació gràfica. Es pot aproximar aquesta funció amb una recta tangent al punt x i de pendent la derivada parcial.² Sabent la mesura d'entrada i el seu interval de confiança es pot trobar el valor i l'interval del resultat. Com es pot observar, la incertesa depèn tant de la incertesa del valor d'entrada com del pendent de la funció (o de l'aproximació per la línia tangent).

Diverses variables

En el cas que f depengui de més d'una variable, la incertesa δ_y es troba combinant les aportacions que fa cadascuna de les diferents variables

$$\delta_y \approx \sqrt{\sum_{j=1}^m \left(\frac{\partial f}{\partial x_j} \delta_{x_j} \right)^2 + 2 \sum_{j=1}^m \sum_{k=j+1}^m r_{x_j x_k} \left(\frac{\partial f}{\partial x_j} \delta_{x_j} \right) \left(\frac{\partial f}{\partial x_k} \delta_{x_k} \right)}, \quad (4.10)$$

on $r_{x_j x_k}$ és el coeficient de correlació entre les dues variables.

De manera semblant al cas d'una variable, es pot interpretar geomètricament com l'aproximació de f pel pla tangent m -dimensional avaluat a les mitjanes de totes les m variables de que depèn, i que té pendent de valor les derivades parcials $\partial f / \partial x_i$.

Diverses variables independents

Si les variables són independents, els termes de la covariància són nuls i l'equació anterior es pot reescriure com

$$\delta_y \approx \sqrt{\sum_{j=1}^m \left(\frac{\partial f}{\partial x_j} \delta_{x_j} \right)^2} \quad (4.11)$$

¹first-order second-moment.

²Tenint en compte que el pendent de la gràfica pot ser tant positiu com negatiu, cal emprar el valor absolut de la derivada parcial.

4.3.1 Aplicació a l'estimació del moviment 3D

En primer terme es considera que les variables del vector de forma són independents, i així es poden aplicar unes equacions de propagació més simples. En el cas que els resultats siguin prometedors, llavors es pot ampliar el raonament tenint en compte les correlacions entre les variables.

Translació en Z

El valor de la translació en l'eix Z es troba amb (2.19)

$$\frac{T_z}{Z_0} = \frac{1}{\sqrt{\lambda_1}} - 1$$

Primer cal trobar-ne la derivada

$$\frac{\partial_{T_z/Z_0}}{\partial \lambda_1} = \frac{-1/2}{\lambda_1 \sqrt{\lambda_1}}$$

i llavors, substituint aquesta derivada a (4.9), s'obté

$$\delta_{T_z/Z_0} = \left| \frac{\partial_{T_z/Z_0}}{\partial \lambda_1} \right| \delta \lambda_1 = \frac{1/2}{\lambda_1 \sqrt{\lambda_1}} \delta \lambda_1 \quad (4.12)$$

Translació en X

L'expressió per calcular la translació en l'eix X és (2.20)

$$\frac{T_x}{Z_0} = \frac{t_x}{f \sqrt{\lambda_1}}$$

Les derivades parcials són

$$\begin{aligned} \frac{\partial_{T_x/Z_0}}{\partial t_x} &= \frac{1}{f \sqrt{\lambda_1}} \\ \frac{\partial_{T_x/Z_0}}{\partial \lambda_1} &= \frac{-t_x}{2f \lambda_1 \sqrt{\lambda_1}} \end{aligned}$$

i substituint a (4.9) s'obté

$$\begin{aligned} \delta_{T_x/Z_0} &= \sqrt{\left(\frac{1}{f \sqrt{\lambda_1}} \delta t_x \right)^2 + \left(\frac{t_x}{2f \lambda_1 \sqrt{\lambda_1}} \delta \lambda_1 \right)^2} \\ &= \frac{1}{f \sqrt{\lambda_1}} \sqrt{(\delta t_x)^2 + \left(\frac{-t_x \delta \lambda_1}{2 \lambda_1} \right)^2} \end{aligned} \quad (4.13)$$

4. Estudi analític de la propagació de l'error

Translació en Y

De manera anàloga al punt anterior, es troba l'expressió de la incertesa a la translació en l'eix Y com

$$\delta_{T_y/Z_0} = \frac{1}{f\sqrt{\lambda_1}} \sqrt{(\delta_{t_y})^2 + \left(\frac{-t_y\delta_{\lambda_1}}{2\lambda_1}\right)^2} \quad (4.14)$$

Rotació θ

La rotació θ es computa amb (2.16)

$$\cos\theta = \sqrt{\frac{\lambda_2}{\lambda_1}},$$

Les seves derivades parcials són

$$\begin{aligned} \frac{\partial\theta}{\partial\lambda_2} &= \frac{-1}{2\sqrt{\frac{\lambda_2}{\lambda_1}}\lambda_1\sqrt{1-\frac{\lambda_2}{\lambda_1}}} \\ \frac{\partial\theta}{\partial\lambda_1} &= \frac{\lambda_2}{2\sqrt{\frac{\lambda_2}{\lambda_1}}\lambda_1^2\sqrt{1-\frac{\lambda_2}{\lambda_1}}} \end{aligned}$$

i llavors, substituint a (4.9) s'obté

$$\begin{aligned} \delta_\theta &= \sqrt{\left(\frac{1}{2\sqrt{\frac{\lambda_2}{\lambda_1}}\lambda_1\sqrt{1-\frac{\lambda_2}{\lambda_1}}}\delta_{\lambda_2}\right)^2 + \left(\frac{\lambda_2}{2\sqrt{\frac{\lambda_2}{\lambda_1}}\lambda_1^2\sqrt{1-\frac{\lambda_2}{\lambda_1}}}\delta_{\lambda_1}\right)^2} \\ &= \frac{1}{2\lambda_1\sqrt{\frac{\lambda_2}{\lambda_1}}\sqrt{1-\frac{\lambda_2}{\lambda_1}}} \sqrt{\delta_{\lambda_2}^2 + \left(\frac{\lambda_2\delta_{\lambda_1}}{\lambda_1}\right)^2} \end{aligned} \quad (4.15)$$

Rotació ϕ

El valor de ϕ s'obté directament de les components del vector propi $\mathbf{v}_1 = (v_x, v_y)^T$ corresponent al valor propi més gran (2.17). Si es pren $\phi = \arccos(v_x)$, l'expressió que cal utilitzar és

$$\delta_\phi = \frac{1}{\sqrt{1-v_x^2}}\delta_{v_x} \quad (4.16)$$

4.3 Aproximació de primer ordre

Amb aquesta expressió no es pot recuperar el signe del resultat. Per tal de tenir en compte el signe es pot fer servir la funció $\phi = \arctan\left(\frac{v_y}{v_x}\right)$. Les derivades parcials són

$$\begin{aligned}\frac{\partial\phi}{\partial v_x} &= \frac{-v_y}{v_x^2 + v_y^2} \\ \frac{\partial\phi}{\partial v_y} &= \frac{v_x}{v_x^2 + v_y^2}\end{aligned}$$

La incertesa es troba com

$$\delta\phi = \frac{1}{v_x^2 + v_y^2} \sqrt{(v_y^2 \delta v_x^2 + v_x^2 \delta v_y^2)} \quad (4.17)$$

Experimentalment es pot comprovar que els valors obtinguts amb les dues equacions (4.16) i (4.17) són molt semblants.

Rotació ψ

Per tal de trobar ψ es fa servir l'element R_{11} de la matriu definida a (2.18). Desenvolupant, es pot trobar l'expressió següent:

$$R_{11}(\psi) = \left(1 + \frac{T_z}{Z_0}\right) (\cos(\phi)M_{11} + \sin(\phi)M_{21})$$

on sabem que

$$1 + \frac{T_z}{Z_0} = \frac{1}{\sqrt{\lambda_1}}$$

Atès que \mathbf{R} és una matriu de rotació, es pot calcular el valor de la rotació amb $\psi = \arccos(R_{11})$. Per trobar la propagació de la incertesa es calculen les derivades parcials respecte de λ_1 , ϕ , M_{11} i M_{21} :

$$\begin{aligned}\frac{\partial\psi}{\partial\lambda_1} &= \frac{1/2(\cos(\phi)M_{11} + \sin(\phi)^2 M_{21}^2)}{\lambda_1^{3/2} \sqrt{\left(-\frac{(-\lambda_1 + \cos(\phi)^2 M_{11}^2 + 2 \cos(\phi)M_{11} \sin(\phi)M_{21} + M_{21}^2 - M_{21}^2 \cos(\phi)^2)}{\lambda_1}\right)}} \\ \frac{\partial\psi}{\partial\phi} &= \frac{(\sin(\phi)M_{11} - \cos(\phi)M_{21})}{\sqrt{\lambda_1} \sqrt{\left(-\frac{(-\lambda_1 + \cos(\phi)^2 M_{11}^2 + 2 \cos(\phi)M_{11} \sin(\phi)M_{21} + M_{21}^2 - M_{21}^2 \cos(\phi)^2)}{\lambda_1}\right)}} \\ \frac{\partial\psi}{\partial M_{11}} &= \frac{-\cos(\phi)}{\sqrt{\lambda_1} \sqrt{\left(-\frac{(-\lambda_1 + \cos(\phi)^2 M_{11}^2 + 2 \cos(\phi)M_{11} \sin(\phi)M_{21} + M_{21}^2 - M_{21}^2 \cos(\phi)^2)}{\lambda_1}\right)}}\end{aligned}$$

4. Estudi analític de la propagació de l'error

$$\frac{\partial_{\psi}}{\partial M_{21}} = \frac{-\sin(\phi)}{\sqrt{\lambda_1} \sqrt{\left(-\frac{(-\lambda_1 + \cos(\phi))^2 M_{11}^2 + 2 \cos(\phi) M_{11} \sin(\phi) M_{21} + M_{21}^2 - M_{21}^2 \cos(\phi)^2}{\lambda_1} \right)}}$$

Fent servir (4.11) es troba l'expressió de la propagació de la incertesa a ψ :

$$\delta_{\psi} = \sqrt{\left(\frac{\partial_{\psi}}{\partial \lambda_1} \delta_{\lambda_1} \right)^2 + \left(\frac{\partial_{\psi}}{\partial \phi} \delta_{\phi} \right)^2 + \left(\frac{\partial_{\psi}}{\partial M_{11}} \delta_{M_{11}} \right)^2 + \left(\frac{\partial_{\psi}}{\partial M_{21}} \delta_{M_{21}} \right)^2} \quad (4.18)$$

De manera semblant al cas anterior, per tal de tenir en compte la informació de signe, es fa ús de la funció *atan2*. La funció és

$$\psi = \arctan\left(\frac{-R_{21}}{R_{11}}\right) = -\text{atan}\left(\frac{\cos(\phi)M_{21} - \sin(\phi)M_{11}}{\cos(\theta)(\cos(\phi)M_{11} + \sin(\phi)M_{21})}\right)$$

Primer cal trobar les derivades parcials respecte de ϕ, θ, M_{11} i M_{21} :

$$\frac{\partial_{\psi}}{\partial \phi} = \frac{(M_{21}^2 + M_{11}^2) \cos(\theta)}{\text{den}}$$

on *den* és

$$\begin{aligned} \text{den} &= \cos(\theta)^2 \cos(\phi)^2 M_{11}^2 + 2 \cos(\theta)^2 \cos(\phi) M_{11} \sin(\phi) M_{21} \\ &+ \cos(\theta)^2 M_{21}^2 - \cos(\theta)^2 M_{21}^2 \cos(\phi)^2 + \cos(\phi)^2 M_{21}^2 \\ &- 2 \cos(\phi) M_{11} \sin(\phi) M_{21} + M_{11}^2 - \cos(\phi)^2 M_{11}^2 \end{aligned}$$

$$\frac{\partial_{\psi}}{\partial \theta} = \frac{\sin(\theta)(-2 \cos(\phi)^2 M_{21} M_{11} - \cos(\phi) M_{21}^2 \sin(\phi) + \sin(\phi) M_{11}^2 \cos(\phi) + M_{11} M_{21})}{\text{den}}$$

$$\frac{\partial_{\psi}}{\partial M_{11}} = \frac{M_{21} \cos(\theta)}{\text{den}}$$

$$\frac{\partial_{\psi}}{\partial M_{21}} = \frac{-M_{11} \cos(\theta)}{\text{den}}$$

Fent servir (4.11) es troba l'expressió de la propagació de la incertesa a ψ :

$$\delta_{\psi} = \sqrt{\left(\frac{\partial_{\psi}}{\partial \phi} \delta_{\phi} \right)^2 + \left(\frac{\partial_{\psi}}{\partial \theta} \delta_{\theta} \right)^2 + \left(\frac{\partial_{\psi}}{\partial M_{11}} \delta_{M_{11}} \right)^2 + \left(\frac{\partial_{\psi}}{\partial M_{21}} \delta_{M_{21}} \right)^2} \quad (4.19)$$

- Entrada** : contorn actiu 3D \mathbf{X} , model de càmera, moviment 3D \mathbf{H} , desviació típica del soroll σ
- 1 . **Sortida** : $\delta_{T_x/Z_0}, \delta_{T_y/Z_0}, \delta_{T_z/Z_0}, \delta_\phi, \delta_\theta, \delta_\psi$
 - 2 . El patró \mathbf{Q}_0 es troba projectant el contorn inicial 3D X al pla de la imatge \mathcal{R}
 - 3 Determinar la matriu de forma \mathbf{W} i la seva pseudoinversa \mathbf{W}^{-1}
 - 4 Pertorbar \mathbf{X} segons la transformació \mathbf{H} : $\mathbf{X}_H = \mathbf{H}\mathbf{X}$ i projectar el nou contorn al pla de la imatge \mathcal{R} per trobar el nou conjunt de punts de control \mathbf{Q}
 - 5 **per** $i=1$ **fins** n **fer**
 - 6 Generar un vector de soroll gaussià δ_i amb mitjana nul·la i desviació típica σ , i pertorbar cada punt del patró $\mathbf{Q}_i = \mathbf{Q} + \delta_i$
 - 7 Trobar els vectors de forma corresponents usant $\mathbf{S}_i = \mathbf{W}^{-1}(\mathbf{Q}_i - \mathbf{Q}_0)$
 - 8 **fi**
 - 9 Determinar la matriu de covariància Σ_S a partir dels n vectors de forma
 - 10 Fixar la incertesa δ_M de la matriu de deformació afí a partir de l'arrel quadrada dels elements de la diagonal de Σ_S
 - 11 Calcular els valors intermedis $\mathbf{v}_1, \lambda_1, \lambda_2$ a partir de $SVD(\mathbf{M}\mathbf{M}^T)$ i les seves incerteses associades δ_{v_1} i $\delta_{\lambda_1}, \delta_{\lambda_2}$ utilitzant (4.6) i (4.7)
 - 12 Aplicar les equacions 4.12, 4.13, 4.14, 4.15, 4.19 i 4.17.

Algorisme 4.1 : Algorisme de propagació analítica de la incertesa a l'espai 3D amb inicialització guiada de la incertesa a l'espai de deformació afí.

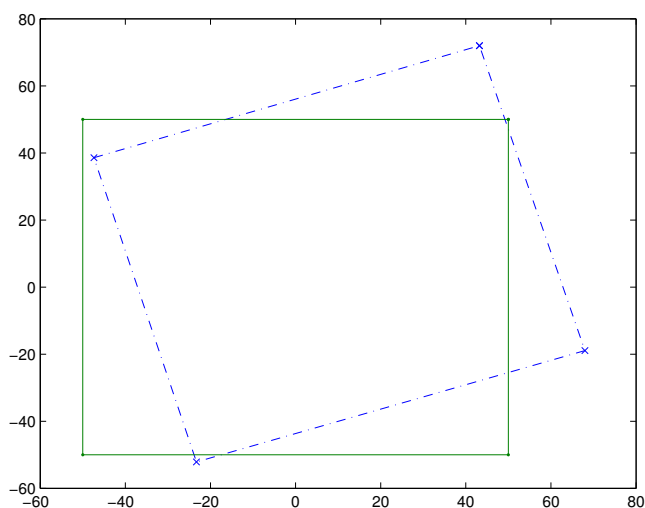


Figura 4.2: Projecció del contorn original i del contorn deformat (línia discontinua) per a l'experiment detallat al text.

4.4 Resultats experimentals

La validació de les equacions de propagació analítica es porta a terme generant soroll en l'espai de deformacions afins en la imatge i propagant-lo en l'espai de posa. El procediment que cal seguir es detalla a l'algorisme 4.1.

Es defineixen els punts de control d'un contorn; per a aquesta prova, un simple quadrat (figura 4.2) que es col·loca a una distància inicial $Z_0 = 5000$ mm. Es projecten els punts de control sobre la imatge fent servir un model perspectiu de càmera ($f = 50$ mm) i amb els punts resultants es defineix la matriu de forma \mathbf{W} . Al contorn original s'hi aplica una transformació en espai 3D i es projecta a la càmera. La transformació la defineix una translació de 100 mm en tots els eixos $[100 \ 100 \ 100]$, i una rotació de la càmera en un eix centrat a l'objecte parametritzada amb el quaternió de valor 30° i vector $[100 \ 100 \ 100]$. La matriu homogènia resultant és

$$\begin{bmatrix} 0.910 & -0.244 & 0.333 & 100 \\ 0.333 & 0.910 & -0.244 & 100 \\ -0.244 & 0.333 & 0.910 & 100 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

La diferència entre els punts projectats del contorn original \mathbf{Q}_0 i del contorn deformat \mathbf{Q}_i serveixen per calcular el vector de forma següent:

$$\mathbf{S} = \mathbf{W}^{-1}(\mathbf{Q} - \mathbf{Q}_0) = \begin{bmatrix} 0.0981 \\ 0.0979 \\ -0.1023 \\ -0.1135 \\ 0.3315 \\ -0.2456 \end{bmatrix} \quad (4.20)$$

Per inicialitzar la incertesa Σ_S a l'espai de posa es generen diversos vectors de forma \mathbf{S}_i afegint soroll gaussià i blanc amb desviació típica $\sigma = 0.5$ píxels als punts de control \mathbf{Q}_i un cop l'objecte ha estat projectat a l'imatge. Amb el conjunt de vectors

	mitjana
λ_1	0.96224
λ_2	0.79940
$v_1(x)$	0.45630
$v_1(y)$	0.57655
δ_{λ_1}	0.01899
δ_{λ_2}	0.00034
$\delta_{v_1(x)}$	0.02786
$\delta_{v_1(y)}$	0.03517

Taula 4.1: Taula de resultats de l'experiment de propagació. Valors propis, el primer vector propi i el resultat de la propagació de l'error per a cadascun.

de forma es troba la matriu de covariància següent:

$$\Sigma_S = 10^{-2} \begin{bmatrix} \mathbf{0.062648} & 0.000943 & -0.004347 & -0.018289 & 0.005601 & -0.023623 \\ 0.000943 & \mathbf{0.063509} & -0.009374 & 0.004545 & -0.002565 & -0.012103 \\ -0.004347 & -0.009374 & \mathbf{25.00790} & -0.194582 & -0.125238 & -0.122556 \\ -0.018289 & 0.004545 & -0.194582 & \mathbf{25.18097} & 0.248811 & -0.057509 \\ 0.005601 & -0.002565 & -0.125238 & 0.248811 & \mathbf{24.87217} & 0.145663 \\ -0.023623 & -0.012103 & -0.122556 & -0.057509 & 0.145663 & \mathbf{24.89749} \end{bmatrix} \quad (4.21)$$

A la diagonal de la matriu (4.21) hi ha els valors de la variància de cada element del vector de forma. Es fan servir aquests valors com a inicialització de la incertesa, que es caracteritza amb la desviació típica. Tal com s'ha dissenyat l'algorisme, primer cal trobar la propagació de la incertesa als valors i vectors propis de la descomposició de \mathbf{MM}^T aplicant els resultats del teorema presentat a la secció 4.2. Aquests resultats intermedis es mostren a la taula 4.1. S'observa que la desviació típica calculada analíticament corresponent al valor propi λ_1 és molt més elevada que la calculada per a λ_2 . Finalment, un cop estimats tots els valors i incerteses necessàries, es pot calcular la incertesa de cadascun dels elements de la posa fent servir les equacions (4.12, 4.13, 4.14, 4.15, 4.19 i 4.17).

A la taula 4.2 es mostren els resultats finals de l'experiment. Es pot veure en primer lloc quin és el moviment real aplicat a l'objecte (primera fila) i quina és la posa que es recupera sense aplicar soroll (segona fila). Pel que fa als valors, les rotacions es recuperen amb força precisió, però en els valors recuperats per a la translació s'aprecia

4. Estudi analític de la propagació de l'error

	T_x/Z_0	T_y/Z_0	T_z/Z_0	$R_z(\phi)$	$R_x(\theta)$	$R_z(\psi)$
moviment real	100	100	100	53.7939	24.4000	-36.2060
posa sense soroll	96.298	96.050	98.045	50.276	24.249	-32.349
posa mitjana computada	96.315	96.068	97.291	50.274	-24.270	-32.347
σ posa computada	0.54581	0.54621	25.93528	2.35903	0.96261	2.35784
σ propagació analítica	0.00020	0.00020	0.01005	0.04229	0.02191	0.04470

Taula 4.2: Taula de resultats de l'experiment de propagació

un cert error (alguns mil·límetres per a la distància inicial entre l'objecte i la càmera de 5 metres). Com es veurà més endavant, no es pot recuperar completament la translació perquè la projecció deixa d'estar centrada en la imatge i, per tant, les condicions de perspectiva dèbil es deixen de complir. Com era d'esperar, la mitjana de les poses calculades amb una simulació de Monte Carlo (tercera fila), en que es generen 50000 mostres amb un soroll de $\sigma = 0.5$ píxels, és molt semblant a l'obtinguda sense soroll.

Pel que fa la desviació típica computada, com es pot observar a la segona part de la taula 4.2, els valors obtinguts per simulació (quarta fila) i els calculats (darrera fila) són quantitativament molt diferents, ja que els valors calculats analíticament són molt menors als obtinguts per simulació. No obstant això, es pot observar que mantenen la mateixa pauta qualitativa. Pel que fa a les translacions, els valors són coherents entre si, amb les desviacions de les translacions T_x i T_y molt semblants i menors a la desviació calculada per T_z , que és aproximadament 50 vegades més gran, tant en el cas simulat com en l'analític. Respecte de les rotacions, la simulació de Monte Carlo mostra que les desviacions de $R_z(\phi)$ i $R_z(\psi)$ són semblants i més grans que la de $R_x(\theta)$, que és la mateixa pauta que segueixen els valors calculats analíticament. Desafortunadament, com en el cas de les translacions, aquests valors calculats estan subestimats.

Amb la simulació de Monte Carlo realitzada es pot calcular la matriu de covariància a partir de les múltiples poses obtingudes. Això permet veure el valor real de les covariàncies entre les components de la posa i avaluar si és necessari desenvolupar-ne les expressions analítiques. La matriu de covariància és

$$\Sigma_{Rt} = 10^{2*} \begin{bmatrix} 0.002979 & 0.002383 & -\mathbf{0.126427} & 0.000045 & -0.003155 & -0.000033 \\ 0.002383 & 0.002983 & -\mathbf{0.126314} & 0.000144 & -0.003143 & -0.000139 \\ -\mathbf{0.126427} & -\mathbf{0.126314} & 6.726390 & -0.005424 & \mathbf{0.167982} & 0.005321 \\ 0.000045 & 0.000144 & -0.005424 & 0.055650 & -0.000141 & -\mathbf{0.055391} \\ -0.003155 & -0.003143 & \mathbf{0.167982} & -0.000141 & 0.009266 & 0.000143 \\ -0.000033 & -0.000139 & 0.005321 & -\mathbf{0.055391} & 0.000143 & 0.055594 \end{bmatrix}.$$

Com es pot veure, hi ha covariàncies entre elements de la posa que no són negligibles. Primer, el valor de la covariància entre les rotacions $R_z(\phi)$ i $R_z(\psi)$ és tan gran com el de la mateixa variància de $R_z(\phi)$ (elements (6, 4) i (4, 4) de la matriu). Segon, el valor de les covariàncies entre les translacions T_x o T_y i T_z (elements (3, 1) i (3, 2)) és més gran que el valor de les seves variàncies (elements (1, 1) i (2, 2)). I tercer, la covariància entre la translació T_x i la rotació $R_x(\theta)$ és més gran que la variància de la mateixa rotació (elements (5, 3) i (5, 5)). La conclusió és que caldria considerar el càlcul de les covariàncies entre les diferents components de la posa.

4.5 Conclusions

En aquesta secció s'han derivat unes expressions analítiques per a la propagació de la incertesa suposant la independència entre les components del vector de forma. La propagació s'ha dividit en dues parts: el càlcul de la incertesa als vectors i valors propis de la SVD de \mathbf{MM}^T , i la propagació per a les operacions algebraiques posteriors. El fet d'aproximar fins a primer ordre, tant la propagació a través de SVD com a través de les expressions algebraiques, fa que els valors obtinguts estiguin clarament subestimats, si es comparen amb els valors obtinguts amb una simulació de Monte Carlo. Tanmateix, qualitativament els valors són coherents. Tant els resultats de la simulació (que es consideren els valors reals), com els resultats de la propagació analítica, coincideixen en el fet que les components amb més incertesa són la translació T_z i les rotacions $R_z(\phi)$ i $R_z(\psi)$.

La covariància esperada de la posa ha mostrat que hi ha diversos valors elevats de covariància entre components de la posa. Per tal de fer l'estudi complet caldria considerar, d'una banda, la dependència entre les variables del vector de forma (que s'ha vist que no és molt alta), i de l'altra, el càlcul de les expressions de les covariàncies entre les diferents components de la posa. Tant una consideració com l'altra són difícils

4. Estudi analític de la propagació de l'error

de plantejar, i augmenten considerablement la complexitat dels càlculs. Com es veurà al proper capítol, l'aproximació estadística a la propagació de la incertesa ofereix alguns avantatges, i enceta algunes línies interessants de treball futur.

Capítol 5

Estudi de la precisió

Most people, including mathematicians, are amateur thinkers in the sense that they have not been taught how to think efectively. They have not been told to throw the crutch away and, as a result, have never learned how to run.

Moltes persones, incloent-hi els matemàtics, són pensadors afeccionats en el sentit que no se'ls ha ensenyat com han de pensar amb eficàcia. No se'ls ha dit que llencin la crosseja i, com a resultat, mai no han après a córrer.

Edsger W. Dykstra, December 1978. EDW696-1.

Extracte

En aquest capítol se segueix un enfocament estadístic per analitzar la precisió de l'algorisme d'estimació del moviment. Les simulacions de Monte Carlo mostren que les translacions paral·leles al pla de la imatge i la rotació sobre l'eix òptic es recuperen millor que les translacions al llarg d'aquest eix, que a canvi són més acurades que les rotacions fora del pla. Pel que fa a les covariàncies, només els tres graus de llibertat menys precisos semblen estar relacionats. Per obtenir el moviment i la seva covariància associada en un sistema robòtic real es recorre a un altre mètode estadístic, la *Unscented Transformation* (UT), que amb l'espai de forma complet requereix tractar només tretze mostres per cada vista. Se'n valida l'ús mitjançant simulacions de Monte Carlo en experiments simulats, i s'utilitza per calcular la incertesa en experiments reals mostrats en els capítols anteriors:

5. Estudi de la precisió

moviments curts amb un braç robòtic Staübli en una escena de laboratori controlada i translacions llargues amb un vehicle industrial en un magatzem. En l'últim cas, les incerteses obtingudes estan al voltant d'un 3%, que és una estimació del moviment prou precisa en operacions de transferència.

5.1 Introducció

La importància que s'ha donat al soroll en visió artificial ha augmentat progressivament al llarg dels anys. Els primers algorismes de geometria visual se centraven en el nombre mínim de punts que calia per aplicar un algorisme determinat. Més endavant, es va incorporar la redundància en les dades per poder tractar amb imatges reals, en què el soroll és una component important. Avui dia, la propagació de l'error i les tècniques d'estimació de la incertesa s'estan aplicant com a pas necessari per intentar reduir la incertesa d'una manera activa.

Els algorismes per recuperar la geometria epipolar i el moviment propi també han seguit aquesta tendència general. Per exemple, encara que se sap que vuit correspondències de punts són suficients per obtenir la matriu fonamental [63, 89], l'algorisme de referència EdO utilitza més correspondències, redundants, que porten a una estimació més robusta a la pràctica [62, 88]. Un primer pas per tractar explícitament els errors és detectar els valors anòmals.¹ Els dos algorismes més populars són el *Least Mean Squares* (LMedS), de qui Zhang [148] dóna una descripció detallada, i el *Random Sample Consensus* (RANSAC), proposat per Fischler i Bolles [54]. Torr i Murray [137] en proporcionen un estudi comparatiu.

El següent pas és modelar el soroll en les dades d'entrada per analitzar com es propaga en el resultat. Una bona introducció amb exemples aplicats a casos reals es pot trobar a [32] i a [36]. Si no es té informació *a priori*, sovint se suposa que la incertesa de les dades d'entrada obeeix una distribució gaussiana [62]. La interdependència entre les variables, quan està quantificada, normalment es representa per mitjà d'una matriu de covariàncies [28, 103]. Alguns estudis [22, 80] demostren que utilitzar la covariància per caracteritzar la incertesa, per exemple en la localització de correspondències de punts, ajuda a dissenyar algorismes millors que si no es té en compte. Fins i tot, la

¹En anglès *outliers*.

incertesa en el mateix procés d'estimació de la matriu de covariàncies també afecta la qualitat del resultat.

Una vegada que es modela el soroll a les dades d'entrada, la propagació de la incertesa generalment s'estudia o bé analíticament o bé estadísticament. Tal com s'ha vist al capítol anterior, els estudis analítics sovint requereixen que s'obtingui el jacobià de la relació entre les dades d'entrada i el resultat que, per a funcions no lineals, sovint exigeix recórrer a aproximacions lineals.

Des del punt de vista estadístic, una de les eines més potents i simples és la simulació de Monte Carlo [46]. Es basa a mostrejar densament l'espai d'entrada i executar l'algorisme per a cada mostra. Un dels problemes és que cal determinar el nombre de mostres que s'utilitzaran. Sovint es pren un nombre molt alt per assegurar que es mostreja tot l'espai d'entrada i que el resultat no estarà esbiaixat i, per tant, s'utilitza quan es necessiten pocs resultats o quan el temps de computació no és un problema. Per tal de reduir el temps de còmput, Julier i Uhlmann han proposat la *Unscented Transformation* (UT), que intenta trobar determinísticament el conjunt de mostres més petit que capturi la distribució estadística de les dades. Aquesta representació és particularment interessant, ja que obre la possibilitat d'utilitzar més endavant un filtre de Kalman no lineal amb la UT (*Unscented Kalman Filter* [74], UKF), o un filtre de partícules [57] (PF), també amb la seva versió UT (*Unscented Particle Filter* [140], UPPF). La conveniència de la UT per a un problema particular es pot provar amb simulació de Monte Carlo i, si és vàlida, aquesta transformació pot comportar estalvis de temps considerables.

L'objectiu d'aquest capítol és analitzar estadísticament la precisió de l'algorisme de recuperació del moviment de la càmera mostrat a la secció 2.2. Primer, a la secció 5.3 es mostra que cal canviar la representació de les rotacions. L'anàlisi de la precisió es porta a terme a la secció 5.4, on s'analitza la precisió utilitzant simulació de Monte Carlo. A la secció 5.4.2 s'analitzen les correlacions que han aparegut entre les components de translació i rotació.

Un segon propòsit d'aquest capítol és obtenir un algorisme per estimar la matriu de covariàncies del sis graus de llibertat del moviment aplicable en temps real per utilitzar-lo en aplicacions de robòtica. Per això, a la secció 5.5.1 es presenta la *Unscented Transformation*, que permet, escollint determinísticament un nombre molt limitat de

5. Estudi de la precisió

punts, propagar la covariància a l'espai de moviment. Per demostrar que l'algorisme és utilitzable es porten a terme una sèrie d'experiments (secció 5.5.2). Finalment, a la secció 5.6 es presenten algunes conclusions i perspectives futures.

5.2 Monte Carlo

Tot i que ja s'ha fet servir anteriorment, es presenta breument aquest mètode estadístic. L'estimació per mostreig aleatori és un mètode general i molt fàcil d'implementar, però normalment és molt costós des de punt de vista computacional. A més, només proporciona la solució per al vector de dades d'entrada que es fa servir. Si les dades canvien, cal refer tota la simulació. El mètode és bastant simple [46]: s'assumeix que es té una funció, un vector de dades d'entrada, i que la incertesa Σ es pot modelar a partir d'una distribució de probabilitat coneguda. El vector d'entrada es pertorba repetidament amb el soroll modelat Σ per tal d'obtenir n vectors de dades diferents. Per a cadascun d'aquests vectors es calcula el valor corresponent aplicant la funció $y = f(x, q, \dots)$. Finalment, es pot examinar la distribució dels valors obtinguts de y i es poden calcular els estadístics (generalment la mitjana i la covariància).

Aquest mètode permet trobar la incertesa encara que la funció f sigui complexa o fins i tot desconeguda (per exemple, un procés de minimització [32]), però requereix molt d'esforç de computació, ja que actua per força bruta. A més, no és fàcil descriure l'aportació de cada variable a la incertesa global. Usualment, tal com s'ha emprat al capítol anterior, la simulació de Monte Carlo es fa servir per avaluar la precisió dels mètodes analítics.

5.3 Representació de les rotacions

Tal com es pot veure a l'equació 2.14, fins ara les rotacions s'han codificat mitjançant angles d'Euler $\mathbf{R} = \mathbf{R}_z(\phi)\mathbf{R}_x(\theta)\mathbf{R}_z(\psi)$. És ben conegut que aquesta representació té el problema de la *gimbal lock*: quan dos eixos estan alineats apareix un problema d'indeterminació. En aquesta representació això passa quan la segona rotació $\mathbf{R}_x(\theta)$ està a prop de la rotació nul·la. El resultat és que petites variacions de la posa de la càmera al voltant de la posició inicial no corresponen a valors continus de les components de

5.3 Representació de les rotacions

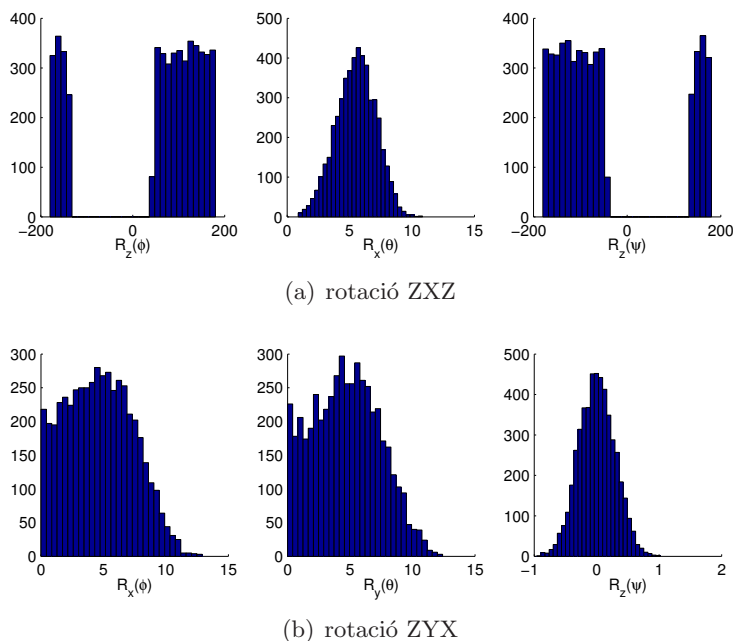


Figura 5.1: Histogrames de les components de 5000 rotacions calculades afegint soroll $\sigma = 0.5$ píxels als punts de control de la projecció d'un contorn. A les rotacions ZXZ, petites variacions de la posa corresponen a valors discontinus de les components $R_z(\phi)$ i $R_z(\psi)$. En canvi, les mateixes variacions corresponen a valors continus en la representació ZYX.

la rotació (vegeu $\mathbf{R}_z(\phi)$ i $\mathbf{R}_z(\psi)$ a la figura 5.1(a)). Aquest problema apareix freqüentment, tant a l'inici de la seqüència d'imatges com també quan el robot es mou cap a l'objecte fixat.

Es proposa convertir la representació en una codificació guinyada-capcineig-balanceig.² Aquesta representació s'utilitza freqüentment en el camp de navegació, i a vegades també s'anomena rumb-horitzó-inclinació [122].³ Es fa servir la forma

$$\mathbf{R} = \mathbf{R}_z(\psi)\mathbf{R}_y(\theta)\mathbf{R}_x(\phi) = \begin{bmatrix} c\psi c\theta & s\psi c\phi + c\psi s\theta s\phi & s\psi s\phi - c\psi s\theta c\phi \\ -s\psi c\theta & c\psi c\phi - s\psi s\theta s\phi & c\psi s\phi + s\psi s\theta c\phi \\ s\theta & -c\theta s\phi & c\theta c\phi \end{bmatrix}, \quad (5.1)$$

²en anglès *roll, pitch, yaw*

³en anglès *heading, attitude, bank*

5. Estudi de la precisió

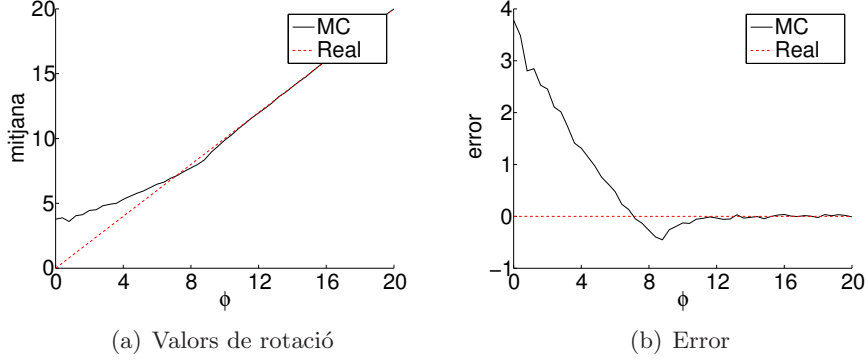


Figura 5.2: Error sistemàtic en la component R_x . La línia contínua representa els valors obtinguts amb Monte Carlo i la línia discontinua són els valors reals. El mateix es pot aplicar a la component R_y .

on $s\psi$ i $c\psi$ representen el sinus i el cosinus respectivament de ψ . La solució inversa és

$$\phi = \text{atan2}(R_{32}, R_{33}) \quad (5.2)$$

$$\theta = \text{atan2}(-R_{31}, \sqrt{R_{32}^2 + R_{33}^2}) \quad (5.3)$$

$$\psi = \text{atan2}(R_{21}, R_{11}). \quad (5.4)$$

Generalment, per tal de representar tot l'espai de rotacions possibles, cadascun dels elements de la rotació es restringeix per pertànyer al rang $[0..2\pi]rad$ per ψ i ϕ , i a $[0..\pi]rad$ per θ .

De fet, el seguiment d'un objecte planar per una càmera que rota segons l'eix X o Y més enllà de $\pi/2$ rad no té sentit, ja que en aquesta posició tots els punts de control de l'objecte són colineals i la forma de l'objecte es perd. A més, a causa de l'ambigüitat *Necker reversal* [62], no és possible determinar el signe de les rotacions entorn d'aquests eixos. Conseqüentment, i sense pèrdua de generalitat, es poden restringir els valors de les rotacions $R_y(\theta)$ i $R_x(\phi)$ per tal que pertanyin al rang $[0..\frac{\pi}{2}]rad$ i deixar $R_z(\psi)$ al rang de valors $[0..2\pi]rad$. Amb aquesta representació s'ha aconseguit desplaçar la *gimbal lock* fins a $\cos(\theta) = 0$, que queda fora del rang de valors possibles.

Amb l'eliminació de signe esmentada s'introdueix un biaix per a l'estimació de rotacions $\mathbf{R}_x(\phi)$ i $\mathbf{R}_y(\theta)$ petites. En presència de soroll i quan la rotació realitzada per la càmera és petita, les rotacions negatives es computen positives. D'aquesta manera el càlcul d'una posa mitjana, com la que es presenta en aquest capítol, està esbiaixat. La

5.4 Avaluant la precisió de les components del moviment

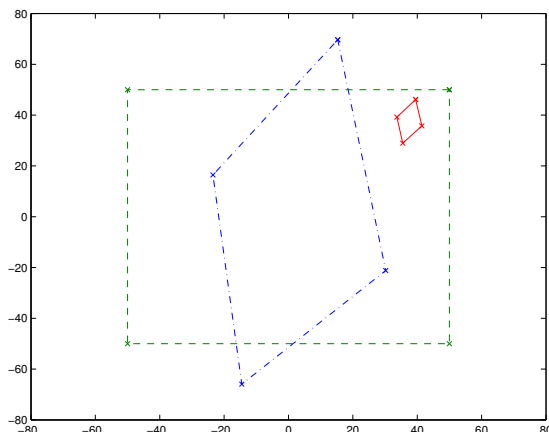


Figura 5.3: Projecció original del contorn (línia discontinua), projecció de contorn després del moviment combinat detallat a la secció 5.4.1 (línia contínua), i projecció de contorn després del moviment combinat excepte la translació (línia de punts i guions), per apreciar millor la deformació extrema a la que s'arriba amb les rotacions en els experiments. Es poden observar alguns efectes de perspectiva evidents.

figura 5.2(a) mostra els resultats d'un experiment en què la càmera fa una rotació de 0 a 20° sobre l'eix X d'un sistema de coordenades situat en l'objecte. Clarament, els valors $\mathbf{R}_{\mathbf{x}}(\phi)$ computats per la simulació de Monte Carlo són més propers als veritables a mesura que l'angle de rotació augmenta. La figura 5.2(b) resumeix els errors que en resulten. La comparació dels dos valors permet avaluar la quantitat d'error sistemàtic introduït per la representació de rotació.

En resum, l'espai de rotació proposat és significativament menor que l'anterior, però hem mostrat que és prou gran per representar totes les situacions possibles. També s'aconsegueix evitar la *gimbal lock* desplaçant-la fora del rang. Com es pot veure a la figura 5.1, les variacions petites en la posa resulten en variacions petites en les components de rotació. Conseqüentment, el moviment es pot calcular coherentment com una mitjana i la covariància.

5.4 Avaluant la precisió de les components del moviment

5.4.1 Propagació de la incertesa a cada component del moviment

Els experiments sintètics estan dissenyats de la manera següent. Es defineix un objecte planar en l'espai 3D i es projecta al pla de la imatge utilitzant un model de càmera

5. Estudi de la precisió

de perspectiva completa.⁴ Es trien un conjunt de punts de control que ressegueixen la projecció de l'objecte, de manera que es defineix la parametrització B-Spline del contorn (figura 5.3). Si es forcen les condicions de visualització possiblement apareixen efectes de perspectiva en els punts projectats (com en una situació real), però l'afinitat emprada no els pot modelar (només pot aproximar el conjunt de punts el millor possible). Si apareixen efectes perspectius, aquests efectes es modelen amb les deformacions afins que es permeten (figura 2.3) i per tant s'introdueix algun error en l'estimació del moviment que cal avaluar. Per a aquests experiments la càmera es posa a 5000 mm de l'objecte i la distància focal es fixa a 50 mm.

S'han dut a terme diversos experiments amb moviments diferents. Una vegada que la càmera s'ha mogut, s'afegeix soroll gaussià amb mitjana zero i $\sigma = 0.5$ als punts de control per simular el soroll d'adquisició de la càmera. Una simulació de Monte Carlo serveix per trobar la mitjana i la covariància del moviment en cada experiment, utilitzant com a transformació l'algorisme presentat a la secció 2.2. En cada simulació de Monte Carlo s'utilitzen 5000 mostres.

Efecte del soroll en la recuperació d'un moviment simple

L'objectiu és determinar experimentalment el grau d'exactitud que s'assoleix (error mitjà i incertesa) amb l'algorisme de recuperació de posa per a cada component del moviment de la càmera, és a dir, les translacions T_x, T_y i T_z , i les rotacions R_x, R_y i R_z . Els primers dos experiments impliquen una translació lateral de la càmera paral·lela als eixos X o Y . Amb la configuració escollida de la càmera, la translació lateral de fins a 300 mm porta la projecció de l'objecte des del centre de la imatge fins al límit de la imatge. Els errors en les estimacions del moviment es presenten a les figures 5.4(a) i 5.4(c), i com era d'esperar són iguals per a les dues translacions. Això és perquè es considera que els píxels són quadrats i que la relació d'aspecte és unitària. Observeu que mentre la càmera s'allunya de la posició inicial l'error en la translació augmenta, així com la incertesa corresponent. L'explicació és que les condicions del model de perspectiva dèbil estan menys satisfetes quan l'objecte no està centrat. Tanmateix, el màxim error en la mitjana aritmètica és aproximadament 0.2%, i la pitjor desviació

⁴Per modelar millor el procés de formació de la imatge en la simulació es fa servir un model més complet, el de perspectiva completa, que el model que suposa l'algorisme de recuperació del moviment, el de perspectiva dèbil, per tal d'obtenir imatges sintètiques més reals.

5.4 Avaluant la precisió de les components del moviment

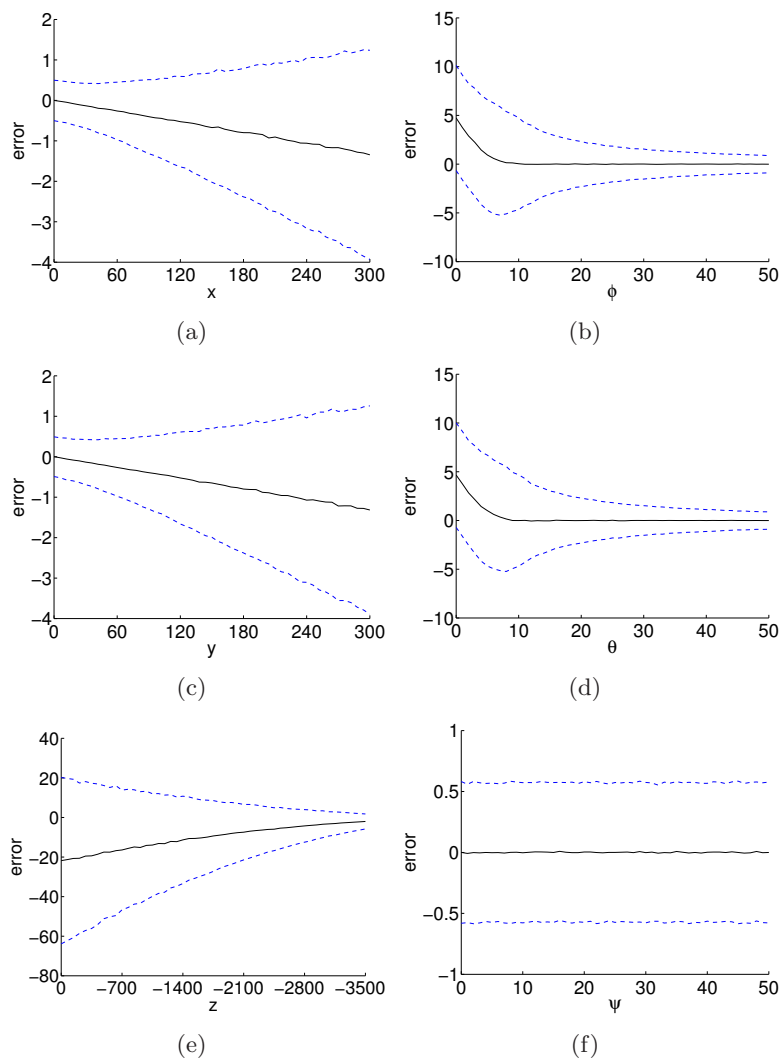


Figura 5.4: Mitjana de l'error (línies contínues) i desviació típica 2σ (línies discontinúes) per a moviments purs al llarg i al voltant dels tres eixos de coordenades. Els errors i desviacions en les translacions T_x i T_y són equivalents, petits mentre la projecció està centrada i augmentant a mesura que s'allunya del centre, i pitjors en la translació T_z (encara que milloren si la càmera s'aproxima a l'objecte). Els errors a R_x i R_y per a rotacions petites són grans, mentre que per a rotacions grans els errors són menys significatius. L'error en la rotació R_z és insignificant.

estàndard és 0.6%, i per tant es pot considerar que les translacions laterals es recuperen força correctament. Tal com s'ha mostrat al capítol 3, el signe de l'error depèn de la forma de l'objecte i de l'orientació de l'eix de rotació.

El tercer experiment implica una translació al llarg de l'eix òptic Z . Des de la

5. Estudi de la precisió

distància inicial $Z_0 = 5000$ mm la càmera es trasllada fins a $Z = 1500$ mm, que és una translació de -3500 mm. Amb aquesta translació es poden apreciar clarament els efectes sobre l'estimació del moviment d'aproximar-se a l'objecte. Els errors i els valors de la incertesa es mostren a la figura 5.4(e). A mesura que la càmera s'acosta a l'objecte, la mitjana de l'error i la seva desviació típica disminueixen. Aquest és un efecte del procés de projecció.⁵ Com era d'esperar, la precisió en les estimacions d'aquesta translació són pitjors per a les dels eixos X i Y .

Els propers dos experiments inclouen rotacions de la càmera al voltant dels eixos X i Y d'un sistema de coordenades fixat en l'objecte. Les figures 5.4(b) i 5.4(d) mostren els resultats. Com era d'esperar, els resultats obtinguts són similars per a aquests dos experiments. S'utilitza la representació alternativa de la rotació presentada a la secció 5.3, i per tant els valors R_x i R_y estan restringits. Tal com s'ha presentat, no es pot recuperar el signe d'aquestes rotacions, que es calculen al mateix costat de la rotació nul·la, i per tant s'introdueix un biaix. A la pràctica aquest biaix no suposa cap perjudici respecte de l'antiga codificació de les rotacions, ja que, com es mostrarà en els experiments amb imatges reals, el soroll present en la fase d'adquisició i seguiment emmascara aquestes rotacions petites, i l'algorisme és incapaç de distingir rotacions de menys de 10° .

El darrer experiment d'aquesta secció implica rotacions de la càmera al voltant de l'eix Z . Els errors computats (figura 5.4(f)) mostren que aquesta component es pot recuperar acuradament, atès que els errors en la mitjana aritmètica són insignificants i la desviació estàndard corresponent es manté també prop de zero.

Efecte del soroll en la recuperació d'un moviment compost

En el proper experiment se segueix una trajectòria combinant tots els moviments al llarg i al voltant dels tres eixos de coordenades. Respecte de l'últim experiment, el moviment de translació en l'eix Z s'ha invertit per anar de 5000 mm a 8500 mm. Això és perquè l'aproximació de la càmera fins a 1500 mm realitzat en l'experiment anterior, combinat amb la translació lateral, faria que el contorn de l'objecte es projectés fora de la imatge.

⁵La resolució en mil·límetres corresponent a un píxel depèn de la distància de l'objecte a la càmera. Quan l'objecte és proper a la càmera, es poden captar fàcilment petites variacions. En canvi, quan la

5.4 Avaluant la precisió de les components del moviment

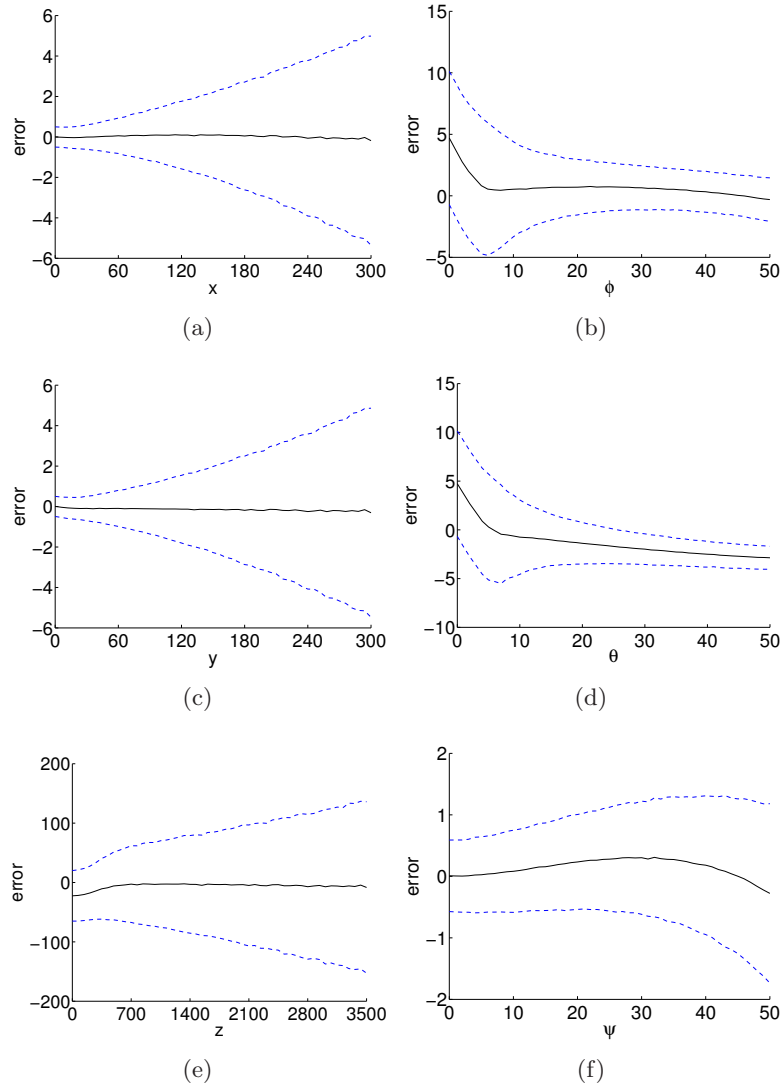


Figura 5.5: Error mitjà (línies contínues) i desviació típica 2σ (línies discontinúes) per a una càmera amb un moviment combinat al llarg i al voltant dels eixos de coordenades.

Òbviament, la translació lateral no es pot mesurar de la mateixa manera quan la càmera s'acosta (figura 5.4(a)) que quan s'allunya. Al final del moviment de la càmera, la projecció del contorn està gairebé centrada, tot i haver fet la translació lateral, ja que alhora s'ha fet un moviment d'allunyament. Com es pot observar a les figures 5.5(a) i 5.5(c), el resultat és que l'error en la recuperació de translació lateral es manté al voltant

distància és gran calen moviments més grans perquè la càmera pugui detectar-los.

5. Estudi de la precisió

de 0. De manera congruent amb el moviment d'allunyar-se, la incertesa augmenta a mesura que la càmera és més lluny de l'objecte. Comparant la figura 5.5(a) amb la figura 5.4(a) es pot observar que la incertesa creix en els dos casos, però, com ja s'ha explicat, per raons diferents.

L'error en la recuperació de la translació en profunditat es mostra a la figura 5.5(e). És gairebé nul al llarg de tot el moviment, excepte al començament de la seqüència, quan la càmera gairebé no s'ha mogut. Més endavant es demostra que això és degut al biaix introduït per la representació de rotació, juntament amb una correlació entre els moviments. Tan aviat com les rotacions R_x i R_y es recuperen correctament, la translació també es recupera. Com era d'esperar, a mesura que la càmera s'allunya la incertesa augmenta. A la figura 5.4(e) sempre hi ha un error en la trajectòria, perquè les rotacions són sempre nul·les. En canvi, en aquest experiment (figura 5.5(e)), hi ha un biaix només al començament.

Els resultats per a les rotacions R_x i R_y són molt similars als de l'experiment anterior. La incertesa al final de la seqüència és una mica més gran a causa de l'augment en la distància entre càmera i objecte. La mateixa raó és aplicable a la incertesa computada per a la rotació R_z (figura 5.5(f)), que també augmenta. D'altra banda, a causa de la rotació extrema al final de la seqüència ($R_x = 50^\circ$ i $R_y = 50^\circ$), apareix un error negligible en l'estimació de la rotació R_z .

Sensibilitat a la quantitat de soroll en un moviment combinat

Ens agradaria mesurar la incertesa del moviment recuperat com una funció de la quantitat de soroll que s'afegeix als punts de control del contorn projectat. L'organització de l'experiment és la mateixa que la presentada en la darrera secció. El moviment de la càmera es defineix amb una translació de 100 mm al llarg de cadascun dels eixos del sistema de coordenades, i amb una rotació de 30° al voltant de l'eix $(1, 1, 1)$ en un sistema de referència centrat en l'objecte. Als punts de control del contorn se'ls afegeix soroll gaussià amb mitjana zero i una desviació típica que va des de $\sigma = 0.1$ fins a $\sigma = 1.0$ en passos de 0.1, cosa que produeix un conjunt de vectors de forma. Per a cada vector de forma es pot computar el moviment corresponent i trobar la covariància en l'espai de moviment 3D. Els resultats es resumeixen a la taula 5.1, on es mostren les desviacions típiques calculades a partir de la matriu de covariància.

5.4 Avaluant la precisió de les components del moviment

σ	T_x	T_y	T_z	R_x	R_y	R_z
0.1	0.1105	0.1100	5.2330	0.1982	0.1939	0.0602
0.2	0.2195	0.2185	10.3813	0.3939	0.3850	0.1193
0.3	0.3292	0.3289	15.6153	0.5925	0.5794	0.1791
0.4	0.4390	0.4377	20.7633	0.7910	0.7710	0.2383
0.5	0.5464	0.5465	25.8702	0.9855	0.9616	0.2968
0.6	0.6589	0.6576	31.1632	1.1824	1.1513	0.3612
0.7	0.7681	0.7663	36.3466	1.3787	1.3463	0.4193
0.8	0.8800	0.8786	41.6336	1.5787	1.5415	0.4810
0.9	0.9944	0.9927	47.0746	1.7858	1.7412	0.5449
1.0	1.0991	1.0979	52.0709	1.9856	1.9338	0.6007

Taula 5.1: Desviacions típiques de les sis components de moviment per a nivells creixents de soroll afegit als punts de control del contorn.

De manera coherent amb la intuïció, a mesura que augmenta el soroll que s'afegeix als punts de control també augmenta la incertesa a les components del moviment. Una característica important que es pot observar és que la incertesa augmenta en la mateixa proporció en totes les components del moviment. Cal recordar que un soroll de $\sigma = 1.0$ implica una pertorbació en els punts de control projectats de ± 2 píxels, que és un soroll considerable. En aquesta situació, les incerteses en les components T_x , T_y i R_z , —per exemple, els moviments dins del pla frontoparal·lel— són molt petites. Les incerteses en les components R_x i R_y són més grans. Clarament, la component amb més incertesa és T_z . Cal recordar que, com s'ha mostrat prèviament, el rendiment de l'algorisme en aquesta component depèn de la distància inicial, que en aquest experiment era de 5000 mm.

Precisió relativa de cadascuna de les components del moviment

En resum, els resultats obtinguts són congruents amb la intuïció si s'observa el model de projecció. Les translacions laterals de la càmera T_x i T_y produeixen canvis més grans en la projecció del contorn al pla de la imatge que no pas la translació al llarg de l'eix òptic T_z , i per tant, també s'obté una millor recuperació del moviment i una incertesa menor en les dues primeres translacions comparades amb la darrera. Les rotacions R_z al voltant l'eix òptic provoquen canvis grans en la imatge, i consegüentment, es recuperen millor que les altres dues rotacions pures, R_x i R_y . Les variàncies obtingudes

5. Estudi de la precisió

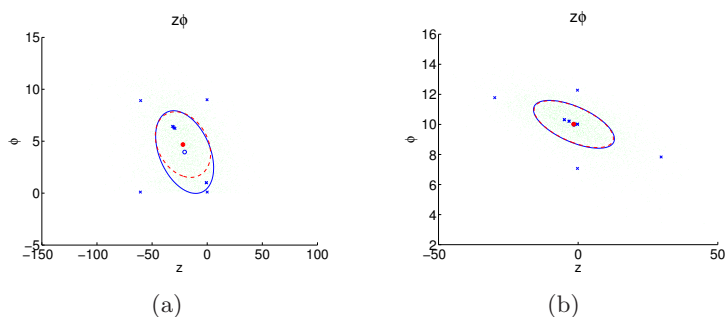


Figura 5.6: Representació gràfica de les submatrius de covariància 2×2 per mitjà de l'el·lipse d'error. Els punts petits són els resultats projectats (moviments recuperats) de transformar vectors de forma d'entrada amb una simulació de Monte Carlo. El punt més gran representa la mitjana i l'el·lipse discontinua la incertesa. Sobreposats, es mostren els resultats computats amb UT, que s'explica a la secció 5.5.1, on les creus representen els punts de sigma transformats i l'el·lipse contínua la incertesa.

difereixen per als diversos moviments. Els errors i les variàncies més grans ocorren quan la projecció de contorn no està centrada en la imatge, atès que es violen les restriccions del model de perspectiva dèbil. Si la distància a l'objecte és petita, s'assoleix més precisió gràcies al fet que augmenta la resolució, però en canvi apareixen els efectes de perspectiva. Les rotacions petites fora del pla es calculen malament, però a mesura que la rotació augmenta l'error i la variància disminueixen. Finalment, les rotacions al pla de la imatge es recuperen correctament amb una variància petita.

5.4.2 Covariància del moviment

Per tal d'obtenir estimacions de la covariància del moviment s'han realitzat diversos experiments, en què s'inclouen moviments nuls de la càmera i, anàlogament a la secció anterior, moviments al llarg i al voltant dels tres eixos del sistema de coordenades. Els resultats de tots els experiments són gairebé coincidents en determinar que no hi ha correlacions entre les components del moviment. Hi ha només dues excepcions, que es presenten seguidament. La relació exhaustiva de tots els resultats es pot trobar a l'apèndix C.

La figura 5.6(a) correspon a un experiment sense cap moviment de la càmera, en què el soroll s'ha afegit als punts de control del contorn. Com es pot observar, les components R_x i T_z del moviment estan correlacionades. Això es pot explicar de la manera següent. Quan la rotació se sobrevalora una mica, la projecció de l'objecte

5.4 Avaluant la precisió de les components del moviment

hauria de ser més petita del que realment és. Per compensar aquest engrandiment, l'algorisme sobreestima la translació T_z , de manera que es computa que la càmera és més a prop del que realment és. Anteriorment, ja s'ha mostrat que la representació de les rotacions introdueix un biaix per a rotacions petites, i per tant, per a rotacions petites aquesta correlació és molt evident a causa de la sobreestimació induïda. A mesura que el valor de la rotació augmenta, la correlació també existeix, però l'efecte que té és menys visible. Això es pot observar a la figura 5.6(b), que correspon a un experiment en què la càmera s'ha girat 10° segons un sistema de coordenades centrat en l'objecte. Aquí es veu que, gràcies al fet que la rotació és més gran, la translació es recobra correctament. Les mateixes observacions es poden aplicar a les components de moviment R_y i T_x .

Aquesta correlació explica la subvaloració de la translació T_z presentada prèviament a la figura 5.4(e), que apareix perquè, prop de la rotació nul·la, R_x i R_y estan sobrevalorades.

La figura 5.7(a) mostra la segona font d'error detectada amb l'estudi de la propagació de la covariància. Quan es fa un moviment de rotació de la càmera al voltant de l'eix Y , apareix una lleugera translació al llarg de l'eix Y . Això es pot explicar analitzant el procés de projecció assumit en el model de càmera de perspectiva dèbil. Quan la càmera gira, alguns dels punts 3D de l'objecte queden més propers a la càmera que d'altres. La figura 5.7(b) il·lustra aquest fet. Per simplicitat de la figura, és més fàcil representar la rotació de l'objecte que la rotació de la càmera, però és fàcil comprovar que les dues situacions són equivalents. La projecció dels punts més llunyans és una mica més propera a l'eix de projecció que no pas la dels punts més propers. Conseqüentment, es computa una petita translació. Anàlogament, si la rotació és al voltant de l'eix X , la translació es computa llavors al llarg de l'eix X . La màxima translació ocorre a $R_x = 45^\circ$, però es pot observar que aquests errors es mantenen sempre molt petits (en els experiments amb seqüències reals es podrà comprovar que aquests errors són mil·limètrics). Això no és una correlació entre variables, sinó un efecte de les diferències entre el model de perspectiva dèbil assumit i el model de perspectiva completa que realment s'ha emprat per trobar les projeccions del contorn en la imatge.⁶

⁶Si la distància inicial és més gran (menor diferència entre la profunditat de l'objecte i la distància a la càmera, com demana el model de càmera afí), aquest efecte és més petit, ja que el model de càmera afí s'acosta més al de perspectiva completa i les translacions laterals que es computen són cada vegada

5. Estudi de la precisió

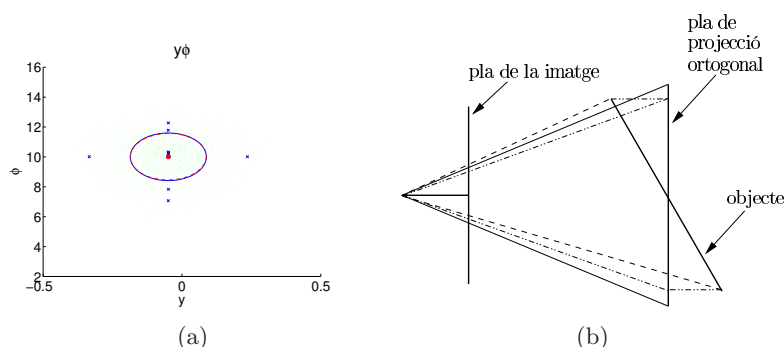


Figura 5.7: (a) Representació gràfica de la submatriu 2×2 de covariància per mitjà de l'el·lipse d'error que relaciona la rotació ϕ al voltant de l'eix X i la translació al llarg de l'eix Y . En realitat, el moviment es compon de només una rotació, però també se'n dedueix una petita translació. (b) Efecte de rotar un objecte planar (que és equivalent a rotar una càmera al voltant d'un objecte). El model de perspectiva dèbil suposa que les projeccions en la imatge haurien de ser on es creuen els raigs puntejats amb el pla de la imatge, però realment són on es projecten les línies ratllades. Consegüentment, es computa una translació lateral quan en realitat només s'ha fet una rotació. Cal notar que el valor de la translació depèn de la distància inicial.

Aquestes són les úniques dues correlacions presents entre les components de moviment recuperades amb l'algorisme proposat. Com ja s'ha esmentat anteriorment, el conjunt complet de resultats experimentals, inclosos tots els moviments al llarg i al voltant dels eixos coordinats i totes les representacions gràfiques de les submatrius 2×2 de covariància, es poden trobar a l'apèndix C.

5.5 Experiments amb seqüències reals

Les simulacions de Monte Carlo utilitzades anteriorment són una eina simple i potent. Tanmateix, sovint només es poden aplicar quan es necessiten pocs resultats o quan el temps de càlcul no és una restricció. Ens agradaria trobar un algorisme per propagar la incertesa per a cada estimació del moviment de la càmera que s'obté. En l'actual implementació de l'algorisme proposat es treballa amb vint imatges per segon.

L'alternativa que s'ha adoptat, l'anomenada *Unscented Transformation*, utilitza un nombre petit de mostres que representen la distribució estadística de les dades, cosa que redueix considerablement el temps de computació exigut per a la propagació d'incertesa.

més negligibles, com es podrà observar a la secció 5.5.2 d'experiments reals.

5.5.1 La *Unscented Transformation*

La *Unscented Transformation* (UT) va ser proposada per Julier i Uhlmann [73, 74]. És un mètode per propagar els estadístics a través d'una funció general no lineal. La UT no es restringeix a assumir que la distribució d'incertesa de les dades és gaussiana. Es defineix de la manera següent: primer, es trien determinísticament un conjunt de punts, anomenats punts sigma (*sigma points*), que s'espera que capturin els estadístics de la distribució de les dades. Segon, els punts es propaguen fent servir la funció no lineal. Finalment, es poden calcular els estadístics amb els punts propagats i així estimar els estadístics del resultat de la funció no lineal.

Normalment, aquesta tècnica s'utilitza en un paradigma del Filtre Kalman per fer l'estructura de predicció i actualització recursiva. Aquest és l'anomenat *Unscented Kalman Filter* (UKF) [73]. Cal notar que no és necessari calcular les derivades parcials de la transformació. A més, la complexitat d'aquest algorisme és la mateixa que la de l'EKF. Julier i Uhlmann [74] van demostrar els beneficis de l'UKF en el context de l'estimació de l'estat per a control no lineal, i Wan i Van Der Merwe [143] van mostrar-ne l'aplicació en problemes d'estimació de paràmetres. També van desenvolupar una formulació on es propaga l'arrel quadrada de la matriu de covariàncies en lloc de la matriu mateixa [141]. Amb aquesta aproximació, els algorismes tenen millors propietats numèriques (principalment en el marc d'aplicació de l'UKF), i la seva complexitat es redueix per a problemes d'estimació de paràmetres. També van desenvolupar una extensió del concepte dels filtres de punts sigma per treballar amb filtres de partícules i sumes de gaussianes [140, 142]. Lefebvre *et al.* [85] van proposar una interpretació alternativa de la UT com una regressió lineal estadística, que és útil per justificar els beneficis que s'obtenen amb UKF.

S'han proposat diferents algorismes per a la tria determinística dels punts sigma. Es pot trobar un resum dels diferents mètodes amb els avantatges i inconvenients que tenen a l'apèndix D. Aquí es fa servir l'algorisme que originalment es va desenvolupar: la solució simètrica [73]. Aquest algorisme requereix de $2N_x + 1$ punts sigma, on N_x és la dimensió del vector de dades d'entrada. Un punt sigma se situa en la mitjana i els altres se situen simètricament al voltant, dos en cada dimensió dels eixos principals de la covariància. Es pot representar com un punt sigma al centre de cada cara d'un

5. Estudi de la precisió

hipercub imaginari. A vegades s'anomena UT de segon ordre, ja que la transformació garanteix que es preserven la mitjana i la covariància (els primers dos moments).

La variable aleatòria N_x -dimensional \mathbf{x} amb mitjana $\bar{\mathbf{x}}$ i matriu de covariància $\Sigma_{\mathbf{x}}$ es pot aproximar amb el conjunt de punts:

$$\begin{aligned} \mathbf{x}^0 &= \bar{\mathbf{x}} \\ \mathbf{x}^i &= \bar{\mathbf{x}} + \left(\sqrt{\frac{N_x}{1-w^0} \Sigma_{\mathbf{x}}} \right)_i \text{ per } i = 1, \dots, N_x \\ \mathbf{x}^{i+N_x} &= \bar{\mathbf{x}} - \left(\sqrt{\frac{N_x}{1-w^0} \Sigma_{\mathbf{x}}} \right)_i \text{ per } i = 1, \dots, N_x \end{aligned} \quad (5.5)$$

amb els pesos

$$\begin{aligned} w^0 & \\ w^i &= \frac{1-w^0}{2N_x} \text{ per } i = 1, \dots, N_x \\ w^{i+N_x} &= \frac{1-w^0}{2N_x} \text{ per } i = 1, \dots, N_x \end{aligned} \quad (5.6)$$

on $(\sqrt{N_x \Sigma_{\mathbf{x}}})_i$ és la fila o columna i -èsima⁷ de l'arrel de la matriu $N_x \Sigma_{\mathbf{x}}$, i w^i és el pes associat amb el i -èsim punts sigma. Els pesos han de complir la condició $\sum w^i = 1$.

Per convenció, el primer punt sigma \mathbf{x}^0 correspon al punt situat en la mitjana. El pes w^0 assignat a aquest punt controla d'alguna manera la situació dels altres punts. Si el pes és positiu, la resta de punts tendeixen a allunyar-se de l'origen i preserven, per tant, el valor de la covariància. Al contrari, si el pes és negatiu, els punts es mouen de manera que s'apropen a l'origen [72]. Aquest mecanisme ajusta els moments d'ordre elevat de la distribució.

Amb aquest esquema, la mitjana i la covariància de la variable \mathbf{y} es poden calcular a partir dels punts sigma transformats, $\mathbf{y}^i = f(\mathbf{x}^i)$, segons

$$\begin{aligned} \bar{\mathbf{y}} &= \sum_{i=0}^{2n} w^i \mathbf{y}^i \\ \Sigma_{\mathbf{y}} &= \sum_{i=0}^{2n} w^i \{ \mathbf{y}^i - \bar{\mathbf{y}} \} \{ \mathbf{y}^i - \bar{\mathbf{y}} \}^T. \end{aligned} \quad (5.7)$$

⁷Dependent de com està formada la matriu Σ , cal fer servir les files o les columnes. Si $\Sigma = \mathbf{A}\mathbf{A}^T$, llavors els punts sigma es formen amb les files d' \mathbf{A} . En canvi, si $\Sigma = \mathbf{A}^T\mathbf{A}$, els punts sigma es formen amb les columnes d' \mathbf{A} .

5.5.2 Fent servir la UT per estimar la mitjana i la covariància del moviment

En l'algorisme d'estimació del moviment propi proposat, l'espai d'entrada és l'espai de forma 6-dimensional (2.12), que es transforma a través de les equacions (2.16 - 2.21) en les tres components de translació i les tres de rotació que componen el moviment de la càmera. Per propagar les covariàncies amb la UT calen $2d + 1 = 13$ punts sigma segons el mètode de selecció simètric, essent $d = 6$ la dimensió de l'espai entrada.

El procediment per calcular les covariàncies utilitzant la UT és el següent. Primer, s'inicialitza manualment sobre la imatge un contorn actiu en què s'especifiquen alguns punts de control. Això defineix una matriu de forma \mathbf{W} segons (2.12). L'algorisme de propagació de la covariància, especificat a l'algorisme 5.1, procedeix de la manera següent. A cada iteració s'adquireix una nova imatge. Amb un filtre de Kalman es calcula la deformació afí del contorn actual respecte al contorn inicial, codificat com un vector de forma, així com una covariància associada. Basant-se en aquests resultats, l'algorisme UT selecciona 13 punts sigma en l'espai de forma, als quals s'aplica la transformació no lineal per trobar el moviment 3D i la covariància que els correspon.

Entrada : Inversa \mathbf{W}^{-1} de la matriu de forma, punts de control inicials \mathbf{Q}_0
Sortida : Posa \mathbf{RT} i covariància Σ_{RT}

- 1 Adquirir una nova imatge
- 2 Amb el filtre de Kalman, predir on es troba el contorn, mesurar on es troben els punts de control \mathbf{Q} del contorn projectat, i estimar el vector de forma corresponent usant $\mathbf{S} = \mathbf{W}^{-1}(\mathbf{Q} - \mathbf{Q}_0)$
- 3 Trobar els *tretze* punts sigma \mathbf{x}^i en espai de forma (5.5) i els seus pesos w^i (5.6) a partir del vector de forma \mathbf{S} i la covariància Σ_S estimats pel filtre de Kalman
- 4 **per** $i=1$ **fins** 13 **fer**
- 5 Calcular la posa corresponent al punt sigma \mathbf{x}^i
- 6 **fi**
- 7 Calcular la posa \mathbf{RT} i la seva covariància Σ_{RT} aplicant (5.7)

Algorisme 5.1 : Algorisme d'estimació de la posa i la seva covariància amb la UT.

Per tal de validar la derivació d'aquest algorisme, s'han realitzat els mateixos experiments sintètics que els duts a terme amb la simulació de Monte Carlo a la secció 5.4.2 i l'apèndix C, i s'ha calculat la corresponent covariància amb el darrer algorisme utilitzant la UT. Els resultats s'inclouen a les figures 5.6, 5.7, i C.1 a C.6. En tots els

5. Estudi de la precisió

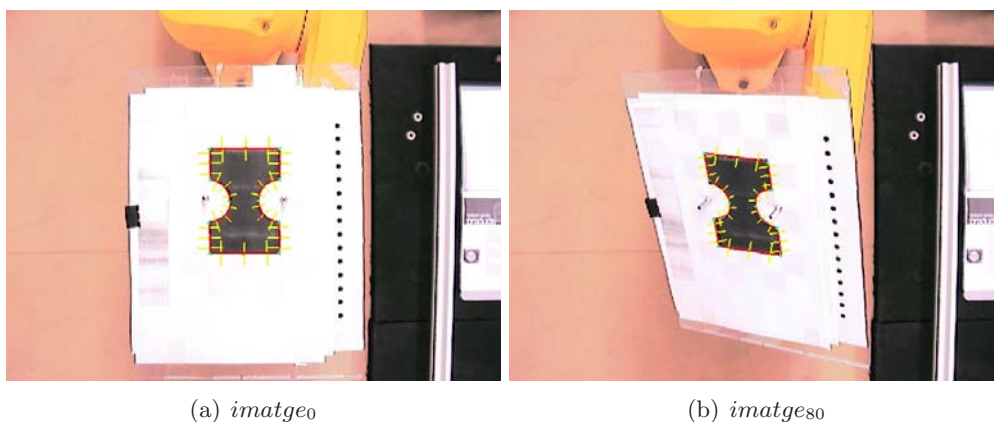


Figura 5.8: Imatge inicial i rotació màxima per a un dels experiments amb el braç robòtic.

experiments la covariància estimada amb la UT és molt similar a l'obtinguda amb la simulació de Monte Carlo i, per això, es pot concloure que és correcte utilitzar la UT per calcular la covariància del moviment.

S'ha realitzat dos conjunts d'experiments amb robots reals. En el primer, un objecte sostingut per un braç robòtic Staübli es fa girar davant d'una càmera immòbil. En el segon, una càmera es munta sobre un vehicle portapalets Still EGV-10, que segueix una trajectòria per apropar-se a un objecte mentre oscil·la lleugerament, cosa que ocasiona una translació llarga.

Experiments de rotació usant un braç robòtic

En el primer experiment es calcula la incertesa del moviment en un entorn d'experimentació que s'ha presentat anteriorment a la secció 3.3: un braç robòtic Staübli que sosté un objecte artificial. En aquest experiment l'objecte es fa girar 40° (i s'atura cada zero graus) al voltant d'un eix definit sobre l'objecte a 45° , i després s'hi aplica una segona rotació per retornar-lo a la posició inicial. La figura 5.8 mostra l'objecte en la posició inicial i en el punt de màxima rotació. Observeu que aquest moviment és equivalent a rotar la càmera -45° sobre el mateix eix. A la figura 5.9 es pot observar l'evolució de les *sis* components del moviment al llarg de tota la seqüència d'imatges, on es mostra tant l'estimació amb la UT com la transformació del vector de forma (anomenada "directe").

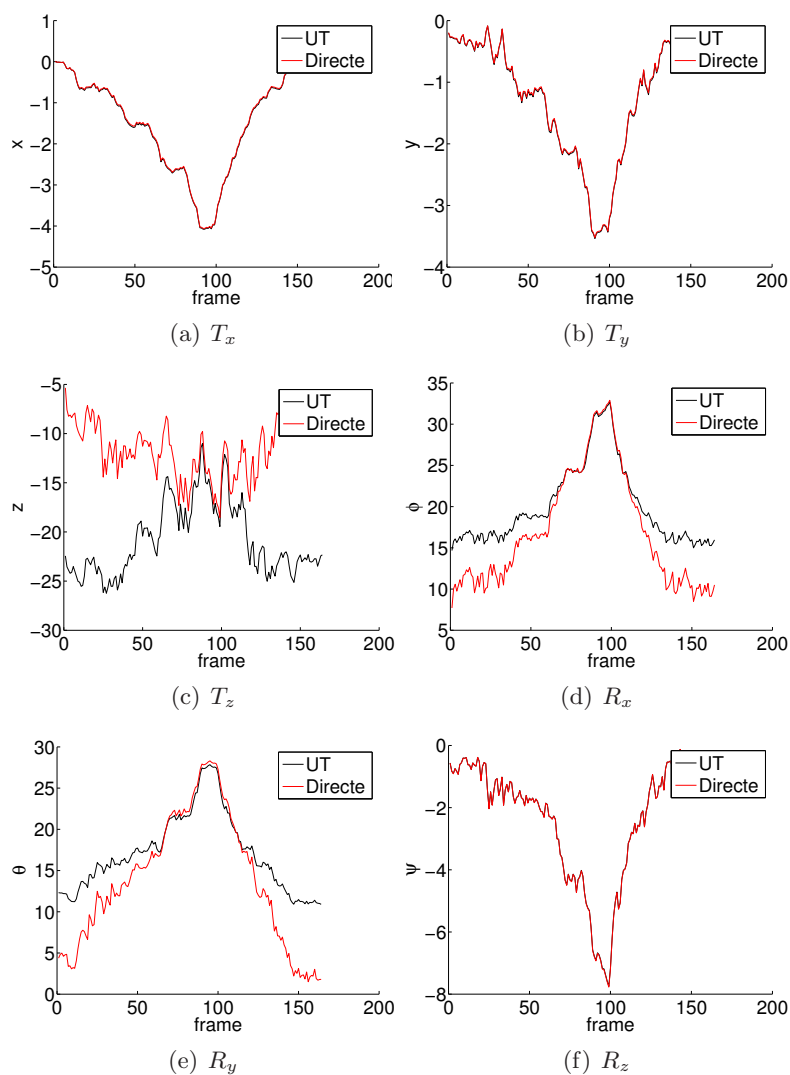


Figura 5.9: Components del moviment que es recupera en l'experiment de la rotació amb imatges reals. El moviment consta d'una rotació de 40° al voltant d'un eix inclinat 40° frontoparal·lel a la càmera i centrat en l'objecte, seguit d'una segona rotació de -45° al voltant del mateix eix. En vermell hi ha els resultats obtinguts amb l'algorisme original, i en negre els valors del moviment amb l'algorisme que utilitza la UT.

De manera congruent amb els resultats sintètics (vegeu la figura 5.7(b), on es computa una lleugera translació quan es fa només una rotació), es computen petites translacions T_x i T_y (figures 5.9(a) i 5.9(b)) tot i que en realitat no es fan. Gràcies al calibratge realitzat, aquests resultats es poden expressar en mil·límetres i es pot concloure que els errors de translació computats (de com a màxim 4 mm) són insignificants. Com

5. Estudi de la precisió

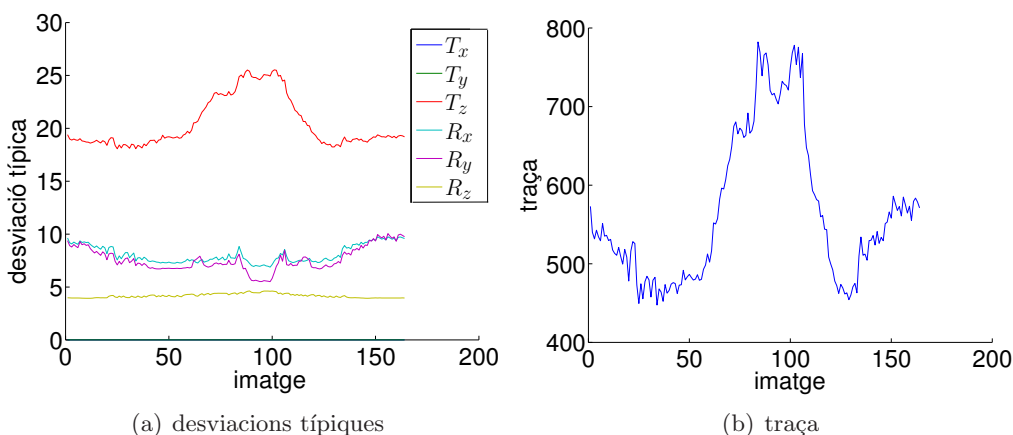


Figura 5.10: (a) Desviacions típiques computades amb la UT per a l'experiment de rotació, i (b) Traça de la matriu de covariàncies que en resulta.

era d'esperar, les translacions T_z es recuperen amb més error. El procés de calibratge determina una distància inicial des de la càmera fins a l'objecte de $Z_0 = 500$ mm, així que la precisió en la recuperació d'aquesta translació està entre l'1% i el 3%, que concorda amb els resultats obtinguts en la simulació.

Les rotacions R_x i R_y no es recuperen correctament per sota de 15° a causa del soroll en l'adquisició i el procés de seguiment del contorn. El biaix a causa de la representació de la rotació també contribueix a afavorir aquest error inicial. Entre les imatges 50 i 100 es poden observar clarament les pauses a cada 10° , i com coincideixen les rotacions computades amb la UT i directament a partir del vector de forma. Per a la rotació R_z coincideixen al llarg de la seqüència sencera.

A la figura 5.10(a) es mostren les desviacions típiques aproximades per a la seqüència de moviment sencera. Les desviacions per a T_x i T_y són gairebé nul·les. La desviació més important s'obté per a la component T_z . Es pot observar que la desviació augmenta al mig de la seqüència, on la rotació és més gran. Això és degut a l'efecte de perspectiva explicat anteriorment a la figura 5.5(f). Es pot observar que les desviacions per a les components R_x i R_y disminueixen lleugerament quan augmenten els valors de la rotació, i tornen a la seva posició inicial quan l'objecte retorna a la posició inicial, on s'hauria de recuperar la rotació nul·la. Com era d'esperar, a causa de la correlació entre les rotacions R_x i R_y , i la translació T_z , la incertesa en T_z també disminueix lleugerament.

La figura 5.10(b) mostra la traça de les matrius de covariàncies. La traça es pot

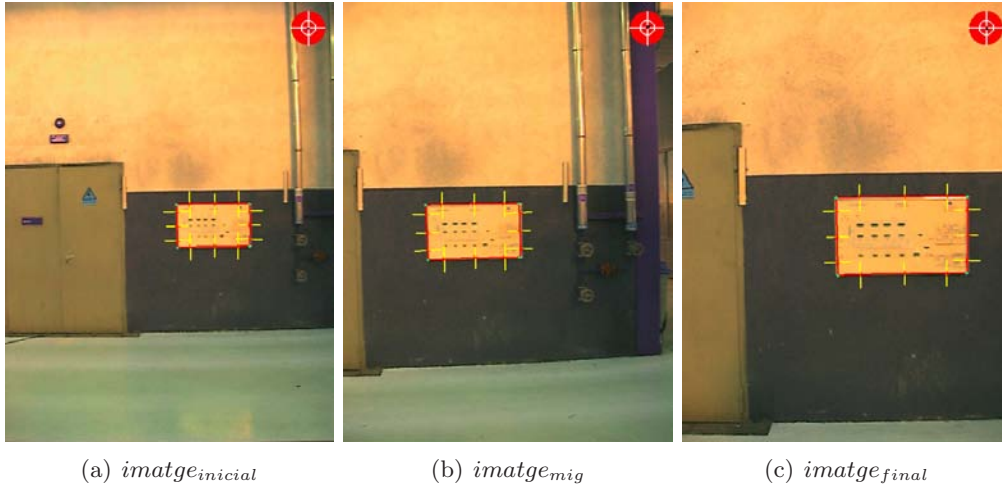


Figura 5.11: Experiment real en què el robot efectua una translació llarga mentre oscil·la lleugerament.

considerar una estimació aproximada de la mida de la covariància [128], i serveix aquí per il·lustrar el comportament global de la incertesa. En la covariància calculada amb l'algorisme proposat, la traça està fortament influïda per la incertesa a la component T_z . La incertesa global disminueix a la primera part de la seqüència, un cop es calculen millor les rotacions, però al mig de la seqüència la incertesa global augmenta a causa de la incertesa de T_z .

Translació llarga amb un vehicle industrial

El segon experiment utilitza les dades que es van recollir en una experiència realitzada en el magatzem d'una fàbrica en un ambient real, presentada ja al capítol 2.3.1. El vehicle portapalets robotitzat que es va emprar disposava d'un làser de posicionament i s'han utilitzat els resultats del posicionament làser per comparar-los amb els del moviment estimat per l'algorisme de visió. Per obtenir resultats mètrics va ser necessari calibrar la càmera i calcular la distància inicial a l'objecte amb el làser. Es va seleccionar com a objecte un tauler d'informació, i es va mesurar que la distància inicial entre la càmera i el tauler era de 7700 mm. L'experiment consta d'una aproximació de 3500 mm amb una lleugera oscil·lació lateral. Tres de les imatges de la seqüència es poden veure a la figura 5.11 (reproduïda de la figura 2.6): a la posició inicial, a meitat de la seqüència i a la posició final. L'anàlisi dels resultats del moviment ja s'ha presentat a

5. Estudi de la precisió

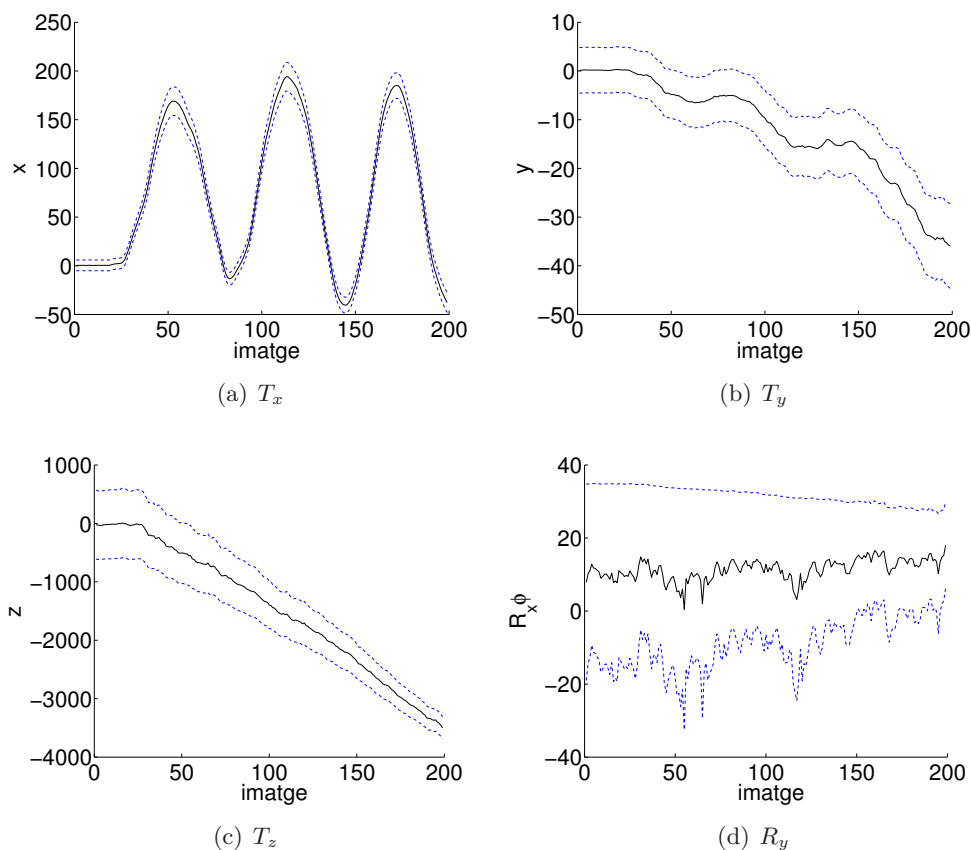


Figura 5.12: Components de moviment recuperades per l'experiment amb un vehicle portapalets. La línia contínua representa el valor del moviment (en mil·límetres per a les translacions i graus per a la rotació) i les línies de punts són les fites 2σ per a la seqüència sencera.

la secció 2.3.1, on també s'ha utilitzat un espai de forma reduït de 4 graus de llibertat. Aquí es presenta la covariància per a cada estimació del moviment.

La figura 5.12 mostra els resultats obtinguts. Es pot observar que la incertesa a la component T_x augmenta quan la distància entre el contorn i el centre de la imatge també augmenta. Això s'ha explicat a la secció 5.4.1 i és degut a la no-satisfacció de les suposicions del model de càmera aff. També s'obté una translació en la direcció T_y . Com ja s'ha explicat anteriorment (secció 2.3.1), això és causat per desalineacions entre els sistemes de referència de la càmera i del robot. Com passa per a la component T_x , la incertesa de T_y augmenta quan augmenta la distància del contorn al centre de la imatge, però en aquest cas els valors són més petits i aquest efecte no es veu fàcilment.

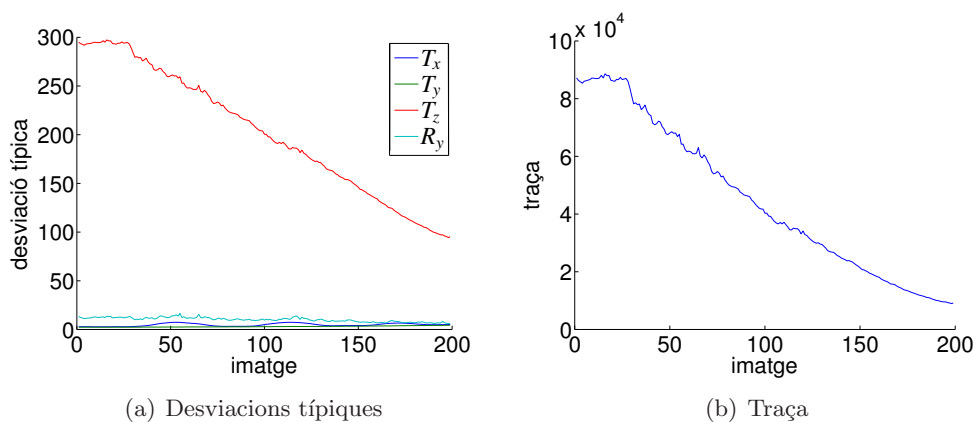


Figura 5.13: (a) Desviacions típiques computades amb la UT, i (b) traça de la matriu de covariància.

Com era d'esperar, la incertesa en la translació T_z disminueix a mesura que el robot s'acosta a l'objecte. Desafortunadament, en aquest experiment les rotacions eren molt petites i l'algorisme no ha estat capaç de recuperar-les. Conseqüentment, la incertesa aproximada per a aquesta component és molt gran (vegeu la figura 5.12(d)).

Es representen les desviacions típiques de les components del moviment i la traça de la matriu de covariàncies a la figura 5.13. Com en el cas anterior, la traça està dominada per la incertesa de T_z . Es pot observar com varia la desviació estàndard de la component amb distància. Comparat amb l'experiment previ, en què la distància inicial era de 500 mm i la desviació típica estava entre 20 i 25, la precisió en l'estimació de T_z i la seva desviació típica són similars, ja que els valors de desviació estan entre 300 i 100 per a distàncies des de 7700 mm fins a 3500 mm.

En resum, en aquesta secció experimental s'ha mostrat que la incertesa en la recuperació del moviment propi es pot calcular en temps real. Dissenyant un algorisme que explota aquesta capacitat, s'ha analitzat la precisió de l'enfocament proposat en la pràctica, i s'ha arribat a la conclusió que s'obté al voltant del 3% d'error per a trajectòries llargues. D'aquesta manera, sembla que l'algorisme d'estimació del moviment propi proposat pot ser utilitzat en vehicles portapalets per a operacions de transferència en què no cal una precisió elevada, aprofitant que no cal preparar l'entorn amb marques artificials, en combinació amb el posicionament làser, més precís en operacions de càrrega i descàrrega [2].

5.6 Conclusions i treball futur

Aquest capítol s'ha dedicat a l'anàlisi dels errors i la seva propagació en l'algorisme d'estimació de la posa. Primer de tot, s'ha vist que la representació de les rotacions en angles d'Euler ZXZ en aquest cas no resulta adequada perquè no és contínua. S'ha vist que és possible fer servir una representació ZYX , ja que es pot restringir el rang de valors, i les rotacions en aquest rang són contínues. La indeterminació de signe en les rotacions al voltant dels eixos X i Y introdueix un biaix quan les rotacions es calculen com una mitjana. En els experiments reals hem vist que, amb l'actual sistema d'adquisició d'imatges, i sobretot, amb l'actual algorisme de seguiment, aquestes rotacions petites de fet no es poden recuperar degut al soroll.

S'ha analitzat la precisió de l'algorisme d'estimació del moviment mitjançant una simulació de Monte Carlo, i els resultats obtinguts són congruents amb la intuïció. Les translacions laterals T_x i T_y produeixen canvis més grans en la projecció del contorn en la imatge i, per tant, es poden recuperar amb més precisió que la translació T_z al llarg de l'eix òptic. Les rotacions R_z al voltant de l'eix de projecció també provoquen canvis grans en la imatge, i per tant es recuperen millor que les altres dues rotacions pures, R_x i R_y . Les variàncies obtingudes per als moviments 3D difereixen molt segons la component del moviment. Els errors més grans i les majors variàncies ocorren quan la projecció del contorn no està centrada en la imatge, ja que es violen les suposicions de perspectiva dèbil. Si la distància a l'objecte és petita, s'assoleix més precisió però apareixen els efectes de perspectiva. Desafortunadament, les rotacions fora del pla no es poden recuperar correctament i conseqüentment es computa una incertesa molt alta, però a mesura que la rotació augmenta, l'error i la variància disminueixen. Les rotacions al pla de la imatge es recobren correctament, amb variància petita. S'ha vist que les úniques correlacions existents són entre les rotacions R_x i R_y i la translació T_z . Es pot observar que, quan no hi ha rotació i els valors de les rotacions R_x i R_y estan sobreestimats, el valor de T_z està subestimat. En canvi, quan R_x i R_y són estimades amb més precisió, T_z també ho és. S'ha observat que una rotació R_x provoca una lleugera translació en T_y , però com s'ha exposat, això no és producte d'una correlació sinó de l'assumpció del model de perspectiva dèbil. En els experiments reals s'ha demostrat que aquest efecte és negligible.

En els experiments reals s'ha emprat la *Unscented Transformation* per computar la incertesa del moviment de diferents robots. S'ha aconseguit una implementació dels algorismes de seguiment, d'estimació del moviment propi i d'estimació de la covariància que funcionen en temps real. El primer experiment, dut a terme amb un braç robòtic Staübli, ha servit per avaluar el moviment i la incertesa per a petites translacions i, principalment, per a grans rotacions. El segon conjunt d'experiments reals, fets en un magatzem d'una fàbrica amb un vehicle portapalets, s'han utilitzat per validar l'algorisme d'estimació del moviment en translacions de llarg abast. Desafortunadament, els resultats obtinguts per a les components de rotació de la posa no són tan bons com s'esperava, ja que no es poden recuperar petites rotacions atès el soroll introduït en el procés d'adquisició de les imatges i en l'algorisme de seguiment.

La continuació del treball està clarament orientada per les conclusions a les quals s'ha arribat aquí. D'una banda, els experiments simulats suggereixen que la projecció de l'objecte s'hauria de mantenir centrada en la imatge —complint millor les suposicions del model de càmera de perspectiva dèbil— per assolir més precisió. D'altra banda, els experiments reals mostren que el ventall d'aplicabilitat de l'algorisme proposat és limitada perquè la projecció del contorn s'ha de mantenir sempre dins de les imatges de la seqüència. Una solució és canviar el contorn que s'empra com a objecte quan desapareix de la imatge i inicialitzar-ne un de nou. Una altra solució, que explorarem en els propers capítols, és mantenir la projecció dins la imatge amb l'ús d'una càmera amb moviment panoràmic (capcineig i guinyada). Això permetrà moviments més generals del robot i a la vegada permetrà obtenir una incertesa menor per a cada moviment.

També ens hem adonat que la mida de la projecció de l'objecte en la imatge s'hauria de quedar dins d'uns marges raonables perquè l'algorisme sigui capaç de fer el seguiment i deduir informació vàlida. La sèrie de translacions d'aproximació en els experiments al magatzem eren com a màxim de 5 metres. Això també és una limitació. Al capítol vinent també s'explora l'ús d'una càmera que fa *zoom* per mantenir la mida de la projecció de l'objecte constant en la imatge. Això presenta alguns desafiaments, com ara que el *zoom* complica el control dels mecanismes de capcineig i guinyada, ja que depenent de la distància inicial (que s'assumeix desconeguda), els guanys que cal aplicar a l'algorisme de control són diferents.

Capítol 6

La visió activa: profunditat a partir del *zoom*

-Tot l'art, vaig continuar, consisteix doncs a buscar la manera més fàcil i eficaç que l'ànima pugui realitzar la conversió que ha de fer. No es tracta de donar-li la facultat de veure, doncs ja la té. Però el seu òrgan no està dirigit en la bona direcció, no mira cap a on hauria s'adreçar-se: això és el que s'ha de corregir.

Plató, *República VII*, 518d. (R. Verneaux, *Textos dels grans filòsofs. Edat antiga*, Herder, Barcelona, 1982).

-There might be an art, an art of the speediest and most effective shifting or conversion of the soul, not an art of producing vision in it, but on the assumption that it possesses vision but does not rightly direct it and does not look where it should, an art of bringing this about.

Plato, *Republic VII*, 518d. (James Adam, *The Republic of Plato*, Cambridge University Press, 1902).

Extracte

S'ha observat que el repertori de moviments que pot realitzar el robot sense que la càmera perdi de vista l'objecte és limitat. En els propers dos capítols s'introdueix el concepte de la visió activa, que ha de servir per fixar l'atenció de la càmera en l'objecte mentre el robot té més llibertat de moviments gràcies a la capacitat d'orientació que s'afegeix a la càmera. Prèviament, però, en aquest capítol, es

6. La visió activa: profunditat a partir del *zoom*

fa èmfasi en la capacitat menys biològica d'un sistema de visió activa: el *zoom*. Fins ara, s'ha mostrat que el moviment propi d'un robot es pot estimar, excepte un factor d'escala, a partir de les deformacions d'un contorn actiu en el pla de la imatge. Una de les components que cal conèixer per eliminar el factor d'escala és la distància inicial entre la càmera i l'objecte que s'utilitza d'objectiu. Si no es disposa de coneixement *a priori*, la solució usual és obtenir "d'alguna manera" aquesta distància. Es proposa un nou algorisme per obtenir la distància inicial aprofitant el grau de llibertat addicional que representa el *zoom*. Inspirat en la calibració geomètrica, l'algorisme que es proposa no necessita els paràmetres de calibratge de la càmera, ni tampoc cal cap suposició sobre l'estabilitat de l'eix òptic quan es canvia la distància focal. Els experiments realitzats sobre un robot mòbil mostren la correctesa de l'enfocament.

6.1 Introducció

Segons Marr [94], la visió és el procés que crea, donat un conjunt d'imatges, una completa i acurada representació de l'escena i les seves propietats. Marr va formular aquesta definició seguint els corrents majoritaris de la intel·ligència artificial dels anys vuitanta, que principalment pretenien trobar metodologies i representacions generalistes. No va tenir en compte, però, que els sistemes de visió existents, i en particular l'humà, són actius.

D'una banda, la visió passiva considera que l'observador no pot triar com vol veure l'escena i que s'ha de limitar, per tant, al que se li ofereix. De l'altra, els sistemes de visió activa tenen mecanismes que permeten controlar activament algun dels diversos paràmetres d'una càmera (o d'un sistema de diverses càmeres), com ara: l'orientació, tant horitzontal com vertical, el focus, el *zoom* i l'obertura. Dit d'una altra manera, la visió activa permet centrar l'atenció, percebre selectivament l'espai o canviar la resolució. Per tant, la visió activa utilitza la capacitat de canviar els paràmetres d'adquisició per obtenir imatges millors. En aquest sentit, cal entendre que una imatge és millor quan permet obtenir informació que facilita la resolució d'una tasca concreta. Per tant, sembla lògic pensar que el disseny d'un sistema de visió activa està íntimament relacionat amb la tasca que cal accomplir. Aquesta visió del problema, lligada íntimament a la tasca, juntament amb el fet que es va veure que no calia computar totes les coses cada vegada sinó computar només el que calia en cada moment, van portar a desenvolupar els primers sistemes de visió activa, coneguts com *purposive vision* [8] i *animate*

vision[12].

La tasca que motiva aquest capítol és trobar la distància inicial entre la càmera i l'objecte que es fa servir d'objectiu. Aquesta mesura representa un dels factors d'escalat que afecta l'algorisme de recuperació del moviment propi. En visió monocular, és ben sabut que no és possible discernir *a priori* el moviment d'un objecte proper i petit del d'un objecte llunyà i gran. Per obtenir la reconstrucció mètrica del moviment cal disposar d'informació addicional. En l'algorisme presentat, el factor d'escala depèn tant de la distància focal de la càmera com de la distància inicial des de la càmera fins a l'objecte. La distància focal es pot obtenir fàcilment amb un algorisme de calibratge o amb mètodes d'autocalibratge, disponibles fins i tot per a càmeres que fan *zoom* [87]. En canvi, la distància inicial des de la càmera fins a l'objecte és més difícil d'obtenir. A [2] es va utilitzar un làser i altres autors han proposat, per exemple, utilitzar el sensor de profunditat d'una càmera autofocus [53], combinar càmeres amb escàners làser [78], fer servir càmeres estereo, calcular els tensors trifocals [135], obtenir la profunditat a partir d'imatges desenfocades [120] i a partir del *zoom* [84, 91].

De totes les capacitats de què disposa un sistema de visió activa la que és útil per resoldre la tasca proposada és el *zoom*. En aquest capítol es proposa un algorisme que aprofita el moviment induït en la imatge per una càmera estàtica que fa *zoom* per calcular la distància inicial entre la càmera i un objecte.

El *zoom* ja ha estat utilitzat anteriorment per deduir la profunditat. En *depth from zooming*, tant la càmera com l'escena han de ser estàtiques, i per tant es pot fer la consideració que la deformació en la imatge prové exclusivament del canvi en el *zoom*. Amb aquesta idea, Ma i Olsen [91] van proposar un mètode per recuperar la informació de profunditat a partir de la variació en la distància focal i del flux òptic. Es van adonar que l'equació que descriu el desplaçament obtingut fent un *zoom* és similar a la que descriu la translació d'una càmera al llarg de l'eix òptic. Van assumir un model de càmera de lent prima, que avui dia se sap que no és el model més adequat per modelar la formació de la imatge en una càmera amb *zoom* [135]. En la seva formulació matemàtica, suposaven que la translació aparent de l'objecte és deguda exclusivament a la variació de distància focal, entesa des del model de càmera estenopec.¹ Lavest *et.al.* [84] van mostrar que això no és correcte. En el seu treball

¹*Pinhole.*

6. La visió activa: profunditat a partir del *zoom*

utilitzen el model de càmeres de lent gruixuda, que és més acurat, ja que imita el procés de canvi de distància focal. És interessant la correspondència que estableixen entre un model de lent gruixuda i una configuració equivalent amb el model estenopecic. Per obtenir bones dades de reconstrucció, els cal un procés de calibratge molt acurat, que inclou paràmetres intrínsecs (amb distorsió radial) i extrínsecs. Estaven forçats a utilitzar una càmera amb lents de molta qualitat, el que s'anomena una càmera axial, ja que suposaven que l'eix òptic era estable durant tota la seqüència en què es feia *zoom*.

Rodin i Ayache [115] van introduir un mètode de calibratge que no exigeix una càmera axial física. Se serveixen d'un mètode de rectificació geomètric per posar en correspondència totes les imatges d'una seqüència, però el mètode que proposen no té en compte les distorsions, i a més la base de triangulació que s'utilitza és molt petita (només 50 mm).

Més tard, Lavest *et.al.* [83] van proposar un mètode de reconstrucció implícit que utilitza un procediment de calibratge geomètric basat en dos patrons. Originalment, el mètode va ser desenvolupat per Martins *et.al.* [95] per resoldre el problema de retro-projecció, i Gremban *et.al.* [61] en van estendre l'ús per incloure també una solució al problema de la projecció. La idea és trobar, sense cap model de càmeres explícit, el raig en l'espai que defineix la línia de mira d'un píxel donat. Per calibrar, Lavest *et.al.* van utilitzar una taula micromètrica per traslladar el patró de calibratge, ja que el mètode de reconstrucció que proposen exigeix un procés de calibratge amb un posicionament del patró d'alta precisió. L'algorisme considera que es pot triangular un nou punt en la imatge,² que cal trobar amb molta precisió, amb les dades de calibratge per trobar la localització 3D del punt. Aquest mètode té els avantatges de tenir en compte totes les distorsions, el desplaçament del centre òptic produït en fer *zoom*, i de no exigir l'estimació dels paràmetres intrínsecs de la càmera. Una crítica que sovint rep [11, 106] és que no té en compte els efectes de desenfocament que apareixen en algunes situacions quan es fa *zoom*. Precisament, el desenfocament s'ha fet servir també per estimar la profunditat, tant activament actuant sobre el focus de la càmera, com aprofitant el mecanisme d'autofocus de la càmera per deduir la profunditat [59]. El problema d'aquest tipus d'enfocament és que per evitar el problema d'haver de calibrar amb totes

²Localitzat manualment a [83] i per mitjà d'un algorisme iteratiu a [43].

6.2 Encara és vàlid el model de càmera estenopecica?

les distàncies focals possibles només permet un nombre molt petit de nivells de *zoom*.

El capítol segueix de la manera següent. A la secció 6.2 es presenten diversos models de càmeres que s’han fet servir per modelar el *zoom*. A la secció 6.3 es mostren els efectes sobre la imatge de fer *zoom* i es fa una petita discussió sobre les distorsions que apareixen. L’espai de deformacions afins reduït que es fa servir per modelar el moviment induït per *zoom* es presenta a la secció 6.4. A la secció 6.5 es presenta l’algorisme proposat de calibratge i el mètode per inferir la profunditat. La secció 6.6 es dedica a presentar alguns experiments realitzats amb imatges reals preses des d’un robot mòbil. Finalment, a la secció 6.7 es poden trobar les conclusions i algunes idees sobre l’aplicabilitat de l’algorisme en altres treballs actuals que requereixen el coneixement de la profunditat inicial.

6.2 Encara és vàlid el model de càmera estenopecica?

Fins al moment s’ha considerat que el model de lents estenopecic era suficientment adequat per modelar el funcionament de la càmera. Però, quan s’introdueix el *zoom* al sistema de visió, es fan variar els paràmetres interns de la càmera i cal revisar si el model de lents estenopecic encara és adequat per modelar la nova càmera. Una òptica d’una càmera generalment és un mecanisme que inclou diverses lents amb propietats diferents. En el camp de la visió artificial, generalment es consideren quatre models de lents diferents: estenopecic, de lent fina, de lent gruixuda i geomètric [56].

El model estenopecic és el més simple i, clarament, també el més utilitzat [51, 62, 92]. En aquest model s’assumeix que tots els rajos passen per un sol punt, el punt focal. Pel que fa a la càmera, es pot modelar com una projecció perspectiva o bé afí, que en les seves versions calibrades i no calibrades donen lloc a diferents matrius de projecció. En general, és inadequat per modelar una càmera amb *zoom*, tot i que s’han proposat diversos algorismes per calibrar una càmera amb *zoom* amb aquest model [42]. És ben conegut que es pot utilitzar un model de lent estenopecic per modelar una càmera amb *zoom* si l’algorisme considera que la distància entre l’objecte i el pla de projecció no és un paràmetre fix [83, 84].

El segon model és el model de lent fina, que assumeix, com el seu nom indica, una lent infinitament fina. Aquesta lent es modela a partir d’un pla. Els rajos que incideixen

6. La visió activa: profunditat a partir del *zoom*

sobre la lent amb un angle determinat surten amb l'angle canviat. Amb aquest model es tenen en compte efectes com la profunditat de camp, que no es poden modelar amb el model estenopecic, però s'ha demostrat que no és adequat per modelar càmeres amb *zoom* [84].

El tercer model s'anomena de lent gruixuda. En aquest model es representa la lent de la càmera amb dos plans, els anomenats plans principals [69]. Els rajos de llum incideixen en el primer pla, viatgen paral·lels a l'eix òptic fins al segon pla i surten amb un angle igual al d'incidència. Tots els rajos provinents d'un mateix punt incideixen en llocs diferents i amb angles diferents sobre el primer pla, i després de sortir per punts diferents del segon pla i amb angles diferents entre ells es troben en un sol punt. Com es pot veure, depenent d'on es col·loca la retina s'obté una imatge més o menys enfocada del punt. És molt interessant la correspondència que es pot establir entre configuracions de lents amb el model de lent gruixuda i el seu equivalent amb el model estenopecic [84].

El darrer model és el model geomètric. És el més complet de tots i també el més difícil d'obtenir, ja que intenta modelar tot el conjunt de lents i mecanismes que formen l'òptica de la càmera. Quan es fa *zoom* es produeixen diversos moviments relatius entre les diferents lents que componen l'òptica, i per tant, per poder utilitzar aquest model, cal conèixer molt acuradament com està dissenyada la lent. Actualment hi ha diversos paquets de programari que ajuden tant a dissenyar òptiques com a modelar-les, però obtenir informació acurada per generar un model útil és difícil. És el model menys utilitzat en visió artificial.

Un cop presentats els diferents models cal veure més detalladament com es comporta el model utilitzat fins ara, l'estenopecic, quan es té una càmera amb *zoom*. És ben conegut que quan es fa *zoom* canvia tant la distància focal com el punt principal (o focus d'expansió). El canvi de punt principal segons la distància focal es pot mesurar fàcilment, i depèn del procés de fabricació i de la qualitat de la lent.

A la figura 6.1(a) es pot veure el model estenopecic d'una càmera amb dues distàncies focals diferents si es considera que en canviar la distància focal de f_1 a f_2 només canvia la posició del pla principal. Naturalment es tracta d'una figura errònia, però sovint és la idea que es té d'un canvi de distància focal. Aquest error prové de la convenció de posar el pla principal davant del centre de projecció. Si es fa a l'inrevés i el pla de

6.2 Encara és vàlid el model de càmera estenopeica?

projecció queda fix, el que es mou en realitat és la lent que es representa pel punt de projecció, i el model estenopeic corresponent és el que es pot veure a la figura 6.1(b).

Pel que fa al control dels mecanismes d'orientació i *zoom* de la càmera, els algorismes que es proposaran més endavant (7) utilitzen resultats parcials de l'algorisme de seguiment per generar les noves posicions. Es tracta d'algorismes en llaç tancat, i atès que només es consideren els canvis en la projecció per al control, es pot continuar fent servir el model estenopeic.

En canvi, l'algorisme de recuperació del moviment propi que s'ha proposat al capítol 2 només té en compte el canvi de distància focal entre dues posicions de *zoom* diferents, però no preveu que el centre òptic també varia en una càmera amb *zoom*. En conseqüència, cal saber que si no es canvia de model de càmera s'introdueix un factor d'escalat addicional a la informació del moviment propi. Cal diferenciar clarament entre la component de rotació i la de translació de la posa. Les rotacions no queden afectades en utilitzar aquest model [87], i només les translacions, i sobretot T_z , queden afectades pel desplaçament no modelat del centre òptic.

Ara bé, aquest escalat té molt poca influència en les condicions de visualització del model de perspectiva dèbil. Els experiments realitzats per Tordoff [135] amb dues càmeres amb *zoom* diferents mostren que la mida relativa d'un objecte en la imatge a mesura que es canvia la distància focal és gairebé idèntica a l'esperada en un model estenopeic per a distàncies focals curtes. Per a distàncies focals llargues, la diferència augmenta a mesura que augmenta la distància focal. Però si la distància de la càmera a l'objecte és gran, el model estenopeic s'ajusta millor als valors obtinguts, de manera que l'error és molt menor. Els errors observats són prou petits per permetre, per exemple, calibrar una càmera amb *zoom* que només gira [64], on es considera que no hi ha translació.

Si calgués modelar més adequadament la càmera amb *zoom*, caldria utilitzar el model de lent gruixuda, que és millor que el de lent prima. Per implementar el model geomètric caldria tenir informació molt precisa de la construcció de la lent, i el model resultant seria molt costós i vàlid únicament per a aquella òptica. En el model de lent gruixuda la dificultat rau a determinar la posició correcta dels plans conjugats que representen la lent. Aquesta distància hauria de poder-se mesurar en un procés de calibratge [135]. Un cop trobats es pot calcular el model corresponent de càmera

6. La visió activa: profunditat a partir del *zoom*

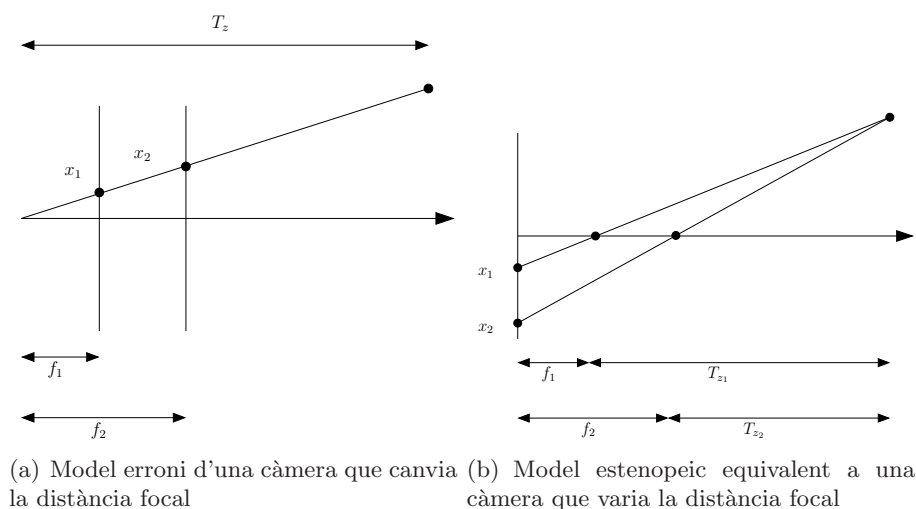


Figura 6.1: Model estenopeic erroni i model correcte d'una càmera en què varia la distància focal. Com es pot veure, no només canvia la distància focal, sinó que també canvia el centre de projecció. En realitat, també canvia el punt principal (que a vegades s'anomena focus d'expansió), però aquest efecte no es mostra a la figura.

estenopeica i, per tant, no caldria definir un nou algorisme de recuperació del moviment propi, ja que es podrien fer servir les mateixes equacions desenvolupades per al model estenopeic.

Malgrat tot, atès que fins ara no s'ha disposat de cap informació sobre els paràmetres de calibratge de la càmera i que tampoc es coneix la distància inicial ni cap característica de l'objecte, les components de translació de la posa computada han estat sempre valors relatius (com és típic en visió monocular), escalats per un factor desconegut. No sembla, per tant, un gran inconvenient introduir un factor d'escalat addicional, i encara més si se sap que aquest factor té molt poca influència.

6.3 Cal considerar el efectes de les distorsions en les lents?

Un altre dels efectes òptics que poden aparèixer en modificar la distància focal de la càmera és la distorsió radial.³ Generalment, modelar la distorsió radial en lents que fan *zoom* és bastant complicat. La relació entre la distància focal i la distorsió generalment

³No tant pel fet de modificar la distància focal com pel fet que ara el robot pot aproximar-se molt més a l'objecte sense que la projecció d'aquest surti de la imatge.

6.3 Cal considerar el efectes de les distorsions en les lents?

no és simple. Tanmateix, hi ha alguns intents de modelar aquesta distorsió per a casos particulars [33].

En els sistemes de càmera-lent comuns la distorsió de la projecció de l'objecte en el pla de la imatge es produeix quan la distància es molt curta. Un tractament més general dels efectes de distorsió, aberració, i les seves causes i conseqüències es pot trobar a [20]. La distorsió generalment s'expressa com el desplaçament d'un punt de la imatge respecte de la distància radial al centre de distorsió [20], i per això moltes vegades s'anomena distorsió radial. Generalment, el centre de distorsió i el centre de projecció són molt propers.

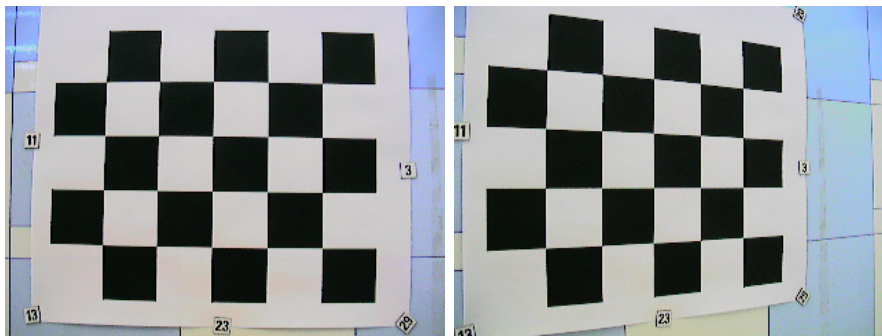
Per avaluar els efectes de la distorsió s'han pres diverses imatges d'un patró de calibratge en diferents configuracions (figura 6.2). Quan la càmera està frontoparal·lela a l'objecte es poden apreciar millor les distorsions radials.⁴ Com es pot veure a la figura 6.2(a), on la distància entre càmera i objecte era de tant sols 30 cm, les línies horitzontals que passen prop del centre de la imatge són rectes, mentre que les més allunyades són corbades. El mateix es pot dir de les línies verticals. La figura 6.2(c) es una imatge presa amb la càmera més allunyada. Per tal que la projecció sigui molt semblant el *zoom* de la càmera s'ha variat, de manera que s'ha augmentat la distància focal. Com es pot observar, amb la càmera allunyada (de fet es va col·locar a tant sols 180 cm) els efectes de distorsió són molt menors, gairebé inapreciables.

Quan la càmera no està en posició frontoparal·lela respecte de l'objecte, els efectes de perspectiva són molt més importants que no pas els efectes de distorsió. Això es pot observar comparant les figures 6.2(b) i 6.2(d). Quan la càmera està realment molt a prop (6.2(b)) el que s'observa clarament és l'efecte de fuga de la perspectiva. Si la càmera està més lluny (en aquest cas, de fet, a només 180 cm) el efectes de perspectiva són molt menors.

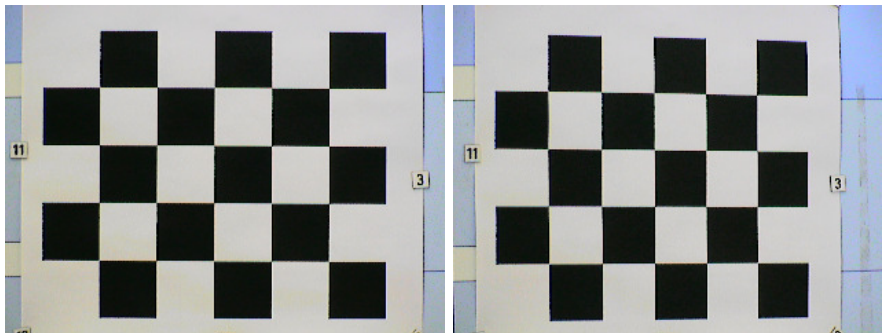
Com s'ha vist, els efectes de la distorsió radial són menyspreables, sobretot si es tenen en compte les condicions de visualització de perspectiva dèbil. Tanmateix, és possible analitzar-ne els efectes potencials sobre l'extracció del moviment. Primer de tot, cal tenir en compte que el contorn de l'objecte s'ajusta en cada nova imatge segons

⁴En una escena general es pot observar fàcilment la distorsió radial si es disposa d'un objectiu d'ull de peix (de distància focal molt curta). Per això cal una lent *especial*. En una càmera amb una lent *normal* per observar l'efecte de la distorsió cal acostar molt la càmera a l'objecte. Com es pot veure, es tracta de reduir la proporció entre distància focal i distància a l'objecte.

6. La visió activa: profunditat a partir del *zoom*



(a) dist=30 cm, zoom=64, frontoparal·lel (b) dist=30 cm, zoom=64, no frontoparal·lel



(c) dist=180 cm, zoom=1120, frontoparal·lel (d) dist=180 cm, zoom=1120, no frontoparal·lel

Figura 6.2: Imatges en què canvia la distància i el *zoom* i que permeten observar els efectes perspectius i de distorsió radial. Si la distància és curta (a), es poden observar els efectes de la distorsió radial, mentre que si l'objecte està allunyat (c), aquests efectes són menys evidents. Si la càmera no està frontoparal·lela a l'objecte (b) per a distàncies curtes, s'observa principalment l'efecte de perspectiva, mentre que si l'objecte està una mica més allunyat (d), aquest efecte és molt més subtil.

el model definit en la primera imatge. Això fa que el vector de forma que parametriza la deformació del contorn sigui fins a cert punt immune a petits efectes, tant de distorsió de la imatge com, fins i tot, de perspectiva.

6.4 Reducció de l'espai de forma

Coneixent que el moviment que es provoca en la imatge induït només per canvis en el *zoom* és restringit, es pot reduir l'espai de forma presentat a la secció 2.2.2. Per construir el nou espai de deformació cal tenir en compte dos aspectes. Primer, l'efecte de fer *zoom* en un factor ρ és el de traslladar un punt a la imatge x al llarg de la línia que va des del punt principal de la càmera v_0 al punt $x' = \rho u + (1 - \rho)v_0$. A efectes pràctics, es pot explicar multiplicant la matriu de calibració de la càmera que correspon a la primera imatge pel factor ρ . En l'espai de forma reduït que es vol construir, es pot introduir com un grau de llibertat amb el vector

$$\begin{bmatrix} \mathbf{Q}^x \\ \mathbf{Q}^y \end{bmatrix} \quad (6.1)$$

que codifica directament l'escala de l'afinitat.

Segon, tal com ja s'ha anunciat anteriorment (capítol 6.2), l'eix òptic en un càmera que fa *zoom* no és constant [84], i això té l'efecte de canviar la posició del punt principal cada vegada que es modifica la distància focal. Per poder modelar aquests moviments cal introduir translacions laterals en l'espai de deformació que es dissenya.⁵ La matriu de forma resultant és

$$\mathbf{W}_{\text{zoom}} = \left(\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \begin{bmatrix} \mathbf{Q}^x \\ \mathbf{Q}^y \end{bmatrix} \right), \quad (6.2)$$

i el vector de forma és

$$\mathbf{S} = [t_x, t_y, \rho]. \quad (6.3)$$

6.5 Estimació de la profunditat a partir del moviment induït pel *zoom*

Tal com es mostrarà, l'algorisme que es presenta comparteix els avantatges principals de l'algorisme de calibratge de dos patrons [83, 95]: no cal estimar els paràmetres intrínsecs

⁵Es pot derivar matemàticament de manera semblant a com es fa a [97].

6. La visió activa: profunditat a partir del *zoom*

de la càmera, i no cal fer cap suposició sobre l'estabilitat de l'eix òptic entre les diferents distàncies focals. Justament, considerar que el punt principal, i per tant l'eix òptic, no és estable és el que ens permet obtenir l'estimació de profunditat. Cal observar que, en el model estenopec equivalent a una càmera que fa *zoom*, no es conserva ni la direcció de projecció ni tampoc el centre de projecció [87]. Per dur a terme el desenvolupament només se suposa que l'eix òptic varia sempre de la mateixa manera entre dues distàncies focals determinades [147]. També se suposa que la relació entre dues vistes de la mateixa escena captada per una càmera estàtica que fa *zoom* es pot aproximar acuradament per una homotècia planar (un canvi d'escala i una translació). Tal com s'ha explicat anteriorment, el factor d'escala (és a dir, la raó de l'homotècia) representa el canvi de distància focal, mentre que la translació explica el desplaçament del punt principal a causa de la no-estabilitat de l'eix òptic.

A més, l'algorisme proposat supera una de les dificultats essencials dels algorismes existents: funciona correctament en condicions de visualització afí. A més, des d'un punt de vista computacional, és un algorisme de calibratge eficient: evita els sovint costosos càlculs de minimització,⁶ ja que les dades d'entrada són les raons de les tres homotècies planars. Amb l'algorisme que es presenta, l'estimació d'aquestes raons depèn de la restricció de l'espai de forma afí, que parametrilitza la deformació de l'objecte projectada en la imatge (vegeu la secció 6.4), que tant es pot calcular amb les dades provinents d'un mètode de seguiment d'un contorn actiu [17] com amb les correspondències de punts provinents, per exemple, d'un mètode basat en transferència afí [136].

6.5.1 Algorisme de calibratge

L'algorisme de calibratge geomètric que es proposa es compon dels passos següents. Primer se situa un objecte planar a certa distància z_1 dins del camp de visió d'una càmera amb un *zoom* A . Llavors, la càmera canvia el *zoom* a B . El moviment induït pel *zoom* es pot parametrilitzar amb un vector de forma (6.3) en què el tercer paràmetre del vector és la raó ρ_1 de l'homotècia h_1 que explica aquestes dues vistes (del *zoom* A al *zoom* B). Aquest procés es repeteix a certa distància z_2 de la càmera: un objecte

⁶Es pot considerar que el pas previ per trobar el vector de forma és una minimització, però computacionalment el cost és molt reduït ja que només cal trobar la pseudoinversa d'una matriu (nombre de punts de control \times mida de l'espai de forma).

6.5 Estimació de la profunditat a partir del moviment induït pel *zoom*

planar (pot ser diferent del precedent) és captat per la càmera que fa un *zoom*, del *zoom A* al *zoom B*, i es calcula un segon vector de forma, que representa l'homotècia h_2 que explica les vistes inicials i finals, i s'obté la raó ρ_2 .

Si la càmera veu un objecte planar nou (a una distància desconeguda z), es pot prendre una imatge en *zoom A* i fer *zoom* fins al *zoom B*. Llavors, es pot computar un nou vector de forma (i la seva raó ρ) que explica la vista inicial i final. El que es demostrarà més endavant és que la raó entre les profunditats $\frac{z_2 - z_1}{z - z_1}$ es pot obtenir a partir de la raó entre les homotècies anteriors, aplicant $\frac{\rho(\rho_2 - \rho_1)}{\rho_2(\rho - \rho_1)}$. D'aquesta manera s'obté una estimació directa de la profunditat desconeguda z , sense necessitat de conèixer cap paràmetre addicional de la càmera. A més, s'evita l'ús tediós d'instruments mètrics, com ara una taula micromètrica, en el procés de calibratge, ja que l'orientació relativa entre els plans que contenen els dos objectes de calibratge no és rellevant. Tampoc no hi ha cap necessitat d'utilitzar patrons de calibració, i és per això que els dos objectes utilitzats en el calibratge poden ser objectes familiars en l'escena (com una porta, una finestra, un tauler d'anuncis...). Es proposa superar el problema de computar acuradament la raó de l'homotècia que explica les vistes inicial i final d'una càmera que fa un *zoom* reduint la dimensió del vector de forma, que codifica la relació afí entre les dues vistes (vegeu la secció 6.4).

6.5.2 Com es pot inferir la profunditat

Ara es mostra, tal com ja s'ha anunciat anteriorment, com s'utilitza la no-estabilitat de l'eix òptic entre les diferents distàncies focals per inferir l'estimació de la profunditat.

Suposeu que la direcció de l'eix òptic quan la distància focal és A difereix lleugerament de la direcció de l'eix òptic quan la distància focal és B . Llavors, hi ha un raig òptic l a la imatge amb *zoom A* que passa per un punt de l'imatge x la direcció del qual és la mateixa que la direcció de l'eix òptic a_B a la imatge amb *zoom B* (vegeu la figura 6.3).

Aquest raig l és proper a l'eix òptic en la imatge amb *zoom A*, i talla els plans de calibratge en els punts X_1 i X_2 , i el pla on es troba l'objecte en el punt X . Per tant, la raó simple d'aquests punts $(X_1, X_2, X) = \frac{d(X_1, X_2)}{d(X_1, X)}$ (on $d(Y_1, Y_2)$ és la distància entre dos punts Y_1 i Y_2) és una estimació de la raó de les profunditats $\frac{z_2 - z_1}{z - z_1}$.

6. La visió activa: profunditat a partir del zoom

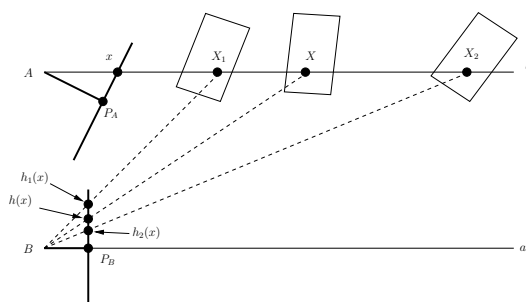


Figura 6.3: Una càmera que fa un zoom estàtic veu la mateixa escena amb el zoom A i el zoom B . La variació de l'eix òptic entre les dues distàncies focals s'ha augmentat per exhibir els trets rellevants (vegeu 6.5.1) que serveixen per inferir la profunditat en l'algorisme de la secció 6.5.2.

Els punts de l'escena X_1 , X_2 i X es projecten en la imatge amb zoom B als punts de la imatge $h_1(x)$, $h_2(x)$ i $h(x)$, respectivament (vegeu la figura 6.3). Recordem que l'objectiu és determinar la raó simple dels punts de l'escena (X_1, X_2, X) des dels punts de la imatge $h_1(x)$, $h_2(x)$ i $h(x)$. Això es pot fer aplicant el resultat següent de geometria projectiva:

Teorema 6.1. *Donat el punt de fuga P d'una línia de l'escena r , amb tres punts de referència Y_1, Y_2, Y_3 , llavors la raó simple (Y_1, Y_2, Y_3) es pot computar des dels seus punts transformats $\bar{Y}_1, \bar{Y}_2, \bar{Y}_3$ de la manera següent: s'escull un punt O a la imatge (que no estigui sobre la línia \bar{r}) i una línia a la imatge s (que no passi per O) paral·lela a la línia que uneix O i P ; per $i = 1, 2, 3$, determinar el punt Y'_i que pertany a s i a la recta que uneix O i \bar{Y}_i ; llavors $(Y_1, Y_2, Y_3) = (Y'_1, Y'_2, Y'_3)$ (vegeu la Fig. 6.4).*

El cas que ens ocupa es pot definir fàcilment posant en correspondència $r = l$ i $Y_1 = X_1, Y_2 = X_2, Y_3 = X$. El punt de fuga de l (el punt a la imatge que correspon al punt de l'infinit de l) és el punt principal $P_B = P$ a la imatge amb zoom B . La suposició que l'eix òptic variï sempre de la mateixa manera entre el zoom A i el zoom B és equivalent a $P_B = h_1(P_A) = h_2(P_A) = h(P_A)$, on P_A és el punt principal en la imatge amb zoom A . En conseqüència, si es defineix un sistema de referència en la imatge centrat en $P = P_B$, amb el primer vector en la direcció de \bar{r} i amb llargada unitària $d(x, P_A)$, llavors $h_1(x)$, $h_2(x)$ i $h(x)$ tenen coordenades $(\rho_1, 0)$, $(\rho_2, 0)$ i $(\rho, 0)$, respectivament. Triant, per exemple, $O = (0, -1)$ i la recta $x = 1$, i aplicant el teorema 6.1, s'obté el resultat desitjat:

$$(X_1, X_2, X) = \frac{\rho(\rho_2 - \rho_1)}{\rho_2(\rho - \rho_1)}. \quad (6.4)$$

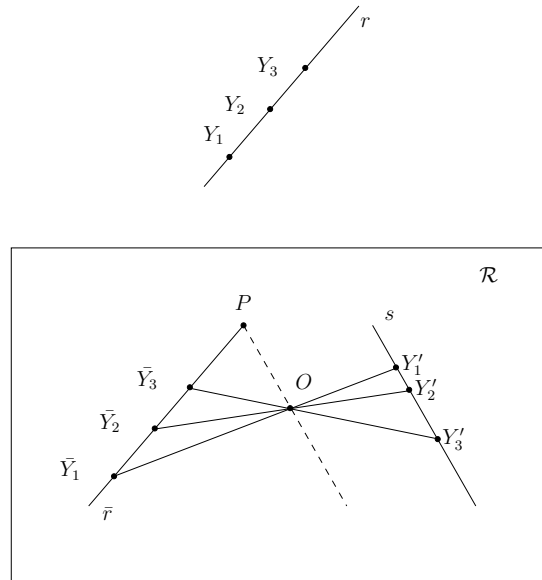


Figura 6.4: Línia de l'escena r amb tres punts de referència Y_1, Y_2, Y_3 projectats a la imatge \mathcal{R} a \bar{r} i $\bar{Y}_1, \bar{Y}_2, \bar{Y}_3$ respectivament. P és el punt de fuga de r . Alguns punts i línies auxiliars es dibuixen damunt d' \mathcal{R} per obtenir la igualtat de raons simples $(Y_1, Y_2, Y_3) = (Y'_1, Y'_2, Y'_3)$, tal com se suposa al teorema 6.1

6.6 Experiments

La correctesa de l'algorisme proposat s'ha provat en diversos experiments amb imatges reals adquirides amb una càmera digital amb *zoom* Sony DFW-VL500. En el manual de la càmera s'especifica que el *zoom* es pot controlar segons unes posicions predefinides que van de 40 a 1432, corresponent a distàncies focals que van des de 5.5 fins a 64 mm. Es munta la càmera sobre una plataforma mòbil Pioneer (vegeu la figura 6.5). Les translacions realitzades amb el robot es calculen aproximadament amb marques al terra. Tot i que el focus de la càmera es manté constant, no s'han observat problemes de desenfocament en cap de les combinacions de *zoom*-distància que s'ha utilitzat. La idea és, tal com és usual en altres mètodes de calibratge, fer el calibratge geomètric amb un objecte natural abans de començar la tasca, i fer servir aquest calibratge durant el funcionament en temps real per trobar la distància d'algun objecte nou. Els passos per computar aquesta profunditat desconeguda es detallen a l'algorisme 6.1.

El primer conjunt d'experiments es du a terme utilitzant un senzill requadre. L'objectiu és validar l'algorisme d'estimació de la profunditat i no l'algorisme de seguiment,

6. La visió activa: profunditat a partir del *zoom*



Figura 6.5: Robot mòbil Pioneer 3AT emprat en els experiments, on es veu el posicionament de la càmera.

1 Calibratge

3 per $i=1$ fins 2 fer

5 Col·locar la càmera a la distància coneguda z_i de l'objecte de calibració

7 Computar el vector de forma \mathbf{S}_i produït per la deformació existent entre la imatge captada a *zoom A* i la captada a *zoom B*

8 fi

9 Mesura

11 Col·locar la càmera a una distància desconeguda de l'objecte

13 Computar el vector de forma \mathbf{S} produït per la deformació existent entre la imatge captada a *zoom A* i la captada a *zoom B*

15 Amb $\mathbf{S}_1, \mathbf{S}_2$ i \mathbf{S} trobar la distància desconeguda z a partir de (6.4)

$$z = z_1 + \frac{\rho_2(\rho - \rho_1)}{\rho(\rho_2 - \rho_1)} (z_2 - z_1)$$

Algorisme 6.1 : Passos de l'algorisme per estimar la profunditat de la càmera respecte d'un objecte. Observeu que només cal fer una calibració per poder estimar la profunditat de diversos objectes.

i per això es tria l'objecte de manera que faciliti al màxim el seguiment. Es trien dues posicions de *zoom* determinades: 500 i 1160. Per obtenir el primer vector de forma dels dos necessaris en el procés de calibratge geomètric, es col·loca la càmera a 120 cm de l'objecte i s'inicialitza manualment un contorn actiu al voltant de la projecció del rectangle en la imatge. Com es pot observar a la figura 6.6, en aquest cas el contorn actiu està format per quatre punts de control (un a cada costat del rectangle) i quatre punts dobles (a cadascuna de les quatre cantonades) formant segments rectes. La seqüència que porta des de la imatge inicial (figura 6.6(a)) fins a la imatge final (figura 6.6(b)) es compon de dues-centes imatges, per a les quals s'ha computat cadascun dels vectors de forma corresponents. D'aquesta manera es completa el tercer pas de l'algorisme 6.1. Seguidament es col·loca el robot de manera que la càmera estigui a 200 cm de l'objecte, i es repeteix el procediment inicialitzant un nou contorn actiu (figures 6.6(c) i 6.6(d)), que s'assembla a l'anterior (ja que la forma de l'objecte no ha canviat), però que no és exactament el mateix, i de fet no té perquè ser el mateix. Un cop obtinguts els dos vectors de forma acaba el procés de calibratge.

Per fer la mesura, en aquest experiment la camera se situa aproximadament a 160 cm de l'objecte. Es repeteix el procediment per calcular la nova raó i s'inicialitza un nou contorn actiu. Es porta la càmera des del *zoom* 500 fins al *zoom* 1160, i es calcula el nou vector de forma. La profunditat de l'objecte es pot computar a partir de les dues raons que s'han determinat durant el procés de calibratge i de la raó computada en el procés de mesura. En aquest experiment el resultat va ser de 157.9 cm, és a dir, un error de tan sols 2.1 cm, que representa un error de l'1.3%.

En el segon conjunt d'experiments es fan servir com a fites *naturals* els calaixos d'una taula i un tamboret, d'on s'extreu la informació de calibratge i de test. En aquest conjunt d'experiments, el robot pren un parell d'imatges amb el *zoom* en posicions 40 i 708,⁷ a distàncies 240, 280, 320, 360 i 400 cm respecte dels calaixos de la taula. Des de la figura 6.7(a) fins a la 6.7(d), els parells d'imatges corresponen a les distàncies 240 i 360 cm. Per a la distància 280 s'utilitza també un tamboret de fusta (vegeu les figures 6.7(e) i 6.7(f)) per validar que el mètode proposat és només dependent de la càmera que fa un *zoom*, i no de l'objecte de calibratge, que pot ser qualsevol.

⁷Observeu que només es coneix la posició del *zoom* segons el controlador intern, però no es té cap coneixement de la distància focal que representa. Només se suposa, tal com s'ha dit anteriorment, que la mateixa posició de *zoom* correspon físicament a la mateixa distància focal.

6. La visió activa: profunditat a partir del *zoom*

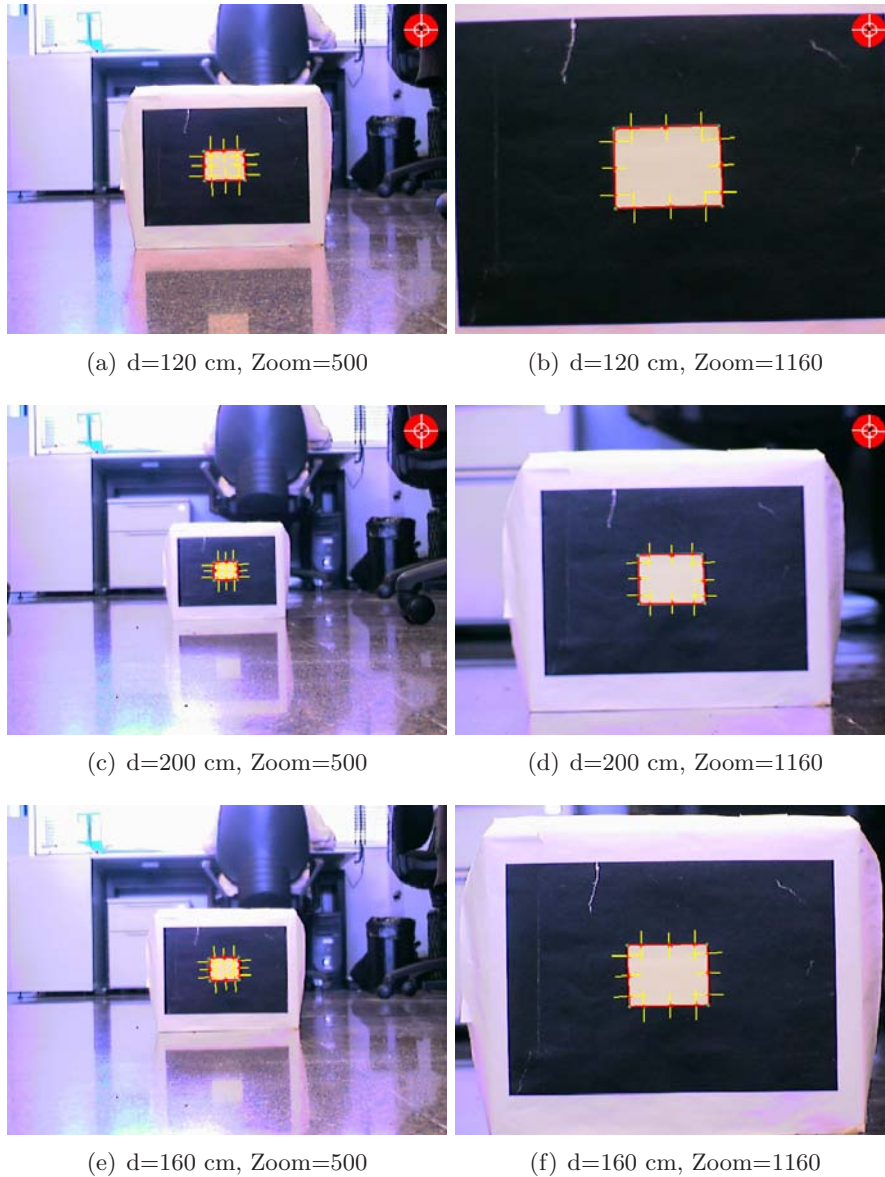


Figura 6.6: Primer conjunt d'experiments utilitzant contorns actius. Per a cada posició de la càmera calen dues imatges per estimar el factor d'escala. (a) (b) Primer parell de calibratge. (c)(d) Segon parell de calibratge. (e)(f) Parell de mesura. En cada parell d'imatges s'ha tornat a inicialitzar el contorn actiu per mostrar que l'algorisme no depèn del contorn actiu utilitzat.

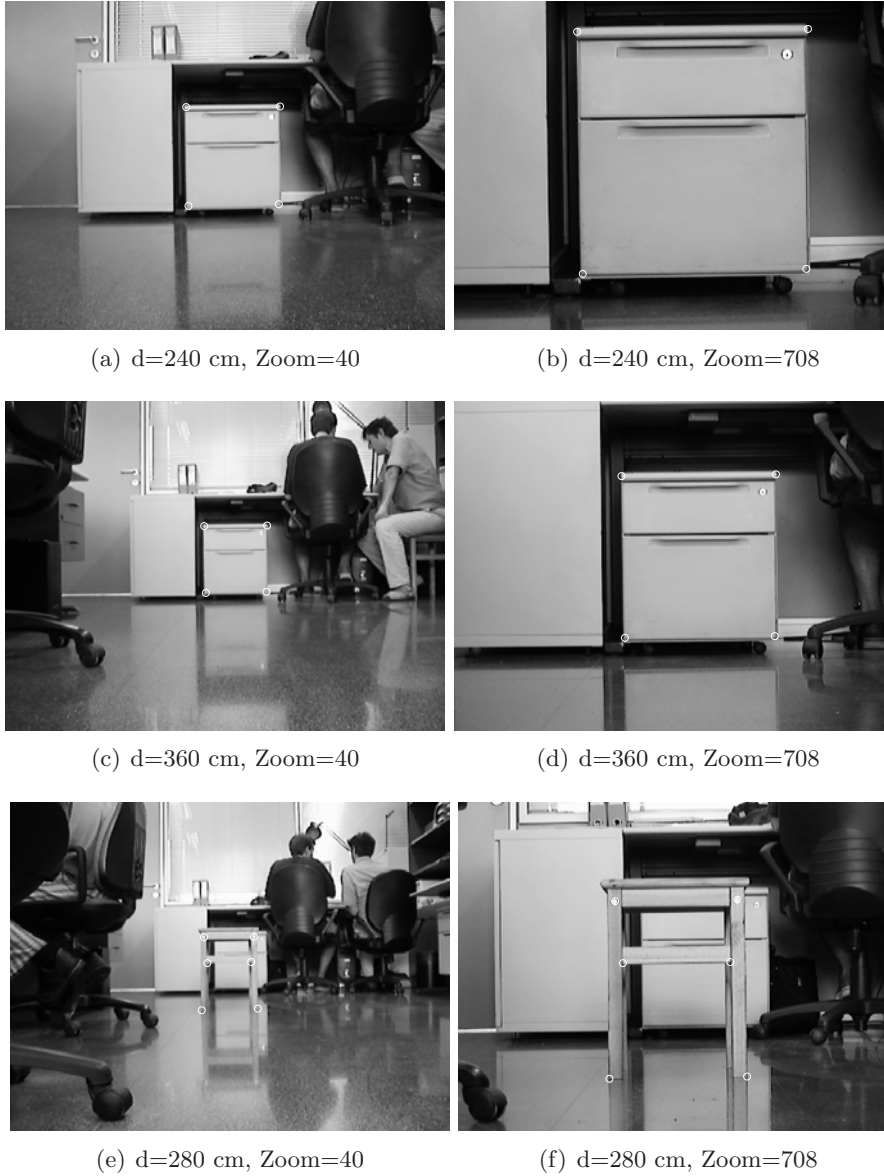


Figura 6.7: Per a cada posició de la càmera calen dues imatges per calcular el factor d'escala. Les imatges mostrades corresponen a l'experiment etiquetat 4 a la taula 6.1. (a) (b) Primer parell de calibratge. (c)(d) Segon parell de calibratge. (e)(f) Parell de mesura. Observeu que l'objecte de calibratge i l'utilitzat per fer la mesura no són iguals, i també que s'ha utilitzat un nombre diferent de mesures per calcular el vector de forma, *quatre* per a les imatges dels calaixos i *sis* per al tamboret.

6. La visió activa: profunditat a partir del *zoom*

Taula 6.1: Resultats del segon conjunt d'experiments on s'utilitzen distàncies de calibratge diferents i objectes diferents.

Exp. ID	Cal1	Cal2	Estimat	Mesurat
1	240	360	277.6	280
2			321.4	320
3			401.7	400
4			269.8	280
5	240	320	288.2	280
6			357.8	360
7			281.6	280
8	320	400	367.7	360

En aquest experiment es vol mostrar que la codificació del moviment en un espai de forma és aplicable també quan s'utilitzen correspondències de punts. Es pot establir una relació entre característiques puntuals extrems d'un objecte i un contorn si es considera la llista de punts com el conjunt de punts de control d'una B-Spline. Com a conseqüència, el mètode presentat, que obté un parametrització del moviment, es pot aplicar també en mètodes de correspondències de punts.

Per calibrar s'extreuen manualment quatre punts per a cada imatge dels calaixos, que serveixen per construir el vector de forma corresponent. Per a les imatges del tamboret, en canvi, s'extreuen sis punts, per demostrar que no importa a partir de quants punts s'extreu la parametrització en vector de forma. Es pot observar que, ja que el mètode per obtenir el vector de forma multiplicant per la pseudoinversa es pot veure com una minimització [17], si es tenen més punts es pot obtenir més precisió.

Alguns dels resultats es resumeixen a la taula 6.1. Les columnes etiquetades Cal1 i Cal2 indiquen les dues distàncies conegudes amb què s'ha realitzat el calibratge geomètric, i les altres dues columnes mostren la distància estimada per l'algorisme presentat i la mesurada manualment. Per als experiments etiquetats 1 i 2, la càmera està situada respectivament a 280 i 320 cm del calaix. Aquestes profunditats estan entre les dues distàncies de calibratge (240 i 360 cm), i la profunditat que es computa amb l'algorisme és prou correcte en cada experiment. En l'experiment etiquetat 3, la càmera se situa més lluny que la segona distància de calibratge (fora del rang de calibratge), i la profunditat també es recobra amb un error petit. Amb aquests paràmetres de cali-

bratge es realitza un quart experiment (numerat 4) utilitzant els sis punts extrems de les imatges dels tamborets. En aquest cas, la profunditat també es recobra raonablement, encara que pitjor que en els casos previs.

En els experiments 5, 6 i 7 el rang de la zona de calibratge s'escurça i s'utilitzen les distàncies de calibratge 240 i 320 cm. Quan la distància està dins de la zona de calibratge, com en l'experiment 5, l'error és del mateix ordre que en l'experiment previ. En canvi, quan la càmera està situada més lluny que la segona distància de calibratge (experiment 6), la profunditat es recobra correctament, però amb més error, comparable amb el de l'experiment 3. Aquest és un efecte típic en calibratge geomètric: la profunditat es recobra correctament dins del rang definit entre la primera i la segona distància de calibratge, ja que l'algorisme està interpolant; fora d'aquesta zona, la profunditat també es pot inferir, extrapolant la informació, però l'error augmenta a mesura que la distància augmenta. En l'experiment 7 s'utilitzen els punts del tamboret.

Finalment, amb l'experiment 8, es prova l'efecte de moure les dues distàncies de calibratge més lluny. El calibratge es duu a terme amb imatges preses a 320 i 400 cm. Es fa una prova posant la càmera al mig, i s'obté una profunditat correcta.

6.7 Conclusions

S'ha presentat un mètode simple per determinar la distància d'una càmera respecte a una fita. La deformació en la imatge provocada quan la càmera fa *zoom* es codifica amb un vector de forma de tres graus de llibertat en un espai de forma presentat, on el tercer element correspon a l'escala de l'homotècia associada. Aquest simple valor d'escala s'enregistra a cada pas del calibratge. Quan es computa una nova deformació a partir del *zoom* sobre un objecte, que pot ser diferent, se'n pot comparar l'escala amb les escales de calibratge i, coneixent la profunditat del calibratge, es mostra com es pot deduir la profunditat de l'objecte actual amb una operació simple.

Amb els experiments amb imatges reals s'ha demostrat la validesa de l'algorisme. La distància entre posicions de calibratge determina una zona calibrada on l'algorisme és més precís. Fora d'aquesta zona, l'algorisme també infereix la profunditat, però és menys precís a mesura que la distància augmenta. S'ha mostrat que es pot trobar el

6. La visió activa: profunditat a partir del *zoom*

vector de forma amb el seguiment de contorns actius, i també amb algorismes de correspondències de punts. En aquest darrer cas, calen un mínim de tres correspondències de punts per construir l'afinitat, però més correspondències proporcionen una millor estimació del vector de forma. Aquí s'han presentat experiments que utilitzen quatre i sis correspondències de punts entre imatges que fan *zoom*. S'ha demostrat que es pot calcular el vector de forma requerit a partir d'objectes diferents i utilitzant un nombre diferent de correspondències de punts.

Realitzant els experiments pràctics s'ha observat que en fer *zoom* a vegades la projecció de l'objecte surt fora de la imatge. Per a propòsits pràctics, és convenient calibrar amb algunes posicions de *zoom* diferents per ser capaç de trobar un rang de posicions de *zoom* que contingui l'objecte en les dues imatges i per a la qual es té informació de calibratge.

El nostre objectiu ha estat principalment eliminar de l'algorisme de moviment propi la incertesa d'escala, comú en tots els sistemes de visió monocular. Però aquest mètode es pot utilitzar també per a uns altres propòsits, per exemple, la inicialització dels controladors del sistema d'orientació del nostre sistema de visió activa (que es mostra al capítol 7). Els experiments amb el control del dispositiu mostren que la precisió obtinguda és prou acurada per inicialitzar els controladors en una bona zona de resposta.

A [2] es calculava la distància inicial amb un làser, i a [1] amb un patró de calibratge. Alguns altres algorismes es podrien beneficiar de l'estimació de la distància inicial d'una fita donada. Ens permetem enumerar-ne alguns. Davison [39] calcula la profunditat d'una fita en visió monocular utilitzant refinaments successius amb un filtre de partícules. Per adquirir l'escala de l'escena en la primera imatge s'utilitza un objecte conegut. L'algorisme presentat es podria utilitzar per canviar l'objecte conegut per qualsevol objecte de l'escena. Recentment, Caballero *et. al.* [25] han presentat un odòmetre visual monocular per a vehicles aeris. Proposen mesurar la distància entre la càmera i els diversos objectes utilitzats en els experiments amb un sonar o un sensor làser, però finalment en els experiments que presenten ho fan manualment. Sola [130] proposa resoldre el problema de la inicialització de la profunditat amb una aproximació del filtre de suma gaussià, i Jensfelt *et. al.* [71] proposen excloure directament del

procés de SLAM⁸ els trets per als quals la profunditat no ha estat determinada. Tots aquests mètodes no poden extreure informació significativa quan hi ha poca disparitat entre característiques, per exemple en moviments d'aproximació de robots que veuen objectes distants.

Òbviament, per a mapes tradicionals basats en punts no és pràctic realitzar el reposicionament del *zoom* cada vegada que cal introduir un nou punt al mapa. Tanmateix, l'algorisme presentat és útil per a les situacions en què cal conèixer una profunditat de referència, com les esmentades anteriorment.

⁸SLAM (*Simultaneous Localization and Mapping*) és el procediment pel qual es construeix un mapa i s'hi localitza el robot, tot al mateix temps.

Capítol 7

Estimació del moviment en el context de la visió activa

Com si fos versemblant que la Natura, tan sol·lícita a vetllar pels mosquits i àdhuc per les herbes i per les floretes, s'hagués endormiscat només en el cas de l'home, obligant-lo a refiar-se de les ciències, que Theuth, divinitat hostil al gènere humà, inventà per emmenar-lo a la total perdició!

Erasme de Rotterdam, *Elogi de la follia* XXXII. (Jaume Molina, MOLU 10, Edicions 62, 1982.)

As if this had the least face of truth, that Nature that was so solicitously watchful in the production of gnats, herbs, and flowers should have so slept when she made man, that he should have need to be helped by sciences, which that old devil Theuth, the evil genius of mankind, first invented for his destruction.

Desiderius Erasmus, *The Praise of Folly*. (Translated by John Wilson, 1668, at Guttenberg project.)

Extracte

Una de les limitacions més importants de l'algorisme de recuperació del moviment propi és que cal que el contorn sigui completament visible, en condicions de perspectiva dèbil, durant tota la seqüència. Aprofitant els avantatges de la visió activa, en aquest capítol es posen les bases per superar aquesta limitació, d'una banda proveint la càmera de *zoom*, que permetrà compensar la translació del robot al

7. Estimació del moviment en el context de la visió activa

llarg de l'eix òptic, i de l'altra, controlant la direcció de visualització amb un dispositiu que permet orientar la càmera. Els algorismes de control que es proposen, tant per al mecanisme de *zoom* com per al de capcineig i guinyada, minimitzen un senyal d'error obtingut directament de mesures en la imatge, sense que calgui més informació 3D que l'estimació de la distància inicial (que es pot obtenir amb l'algorisme presentat al capítol anterior). Un altre avantatge és que el senyal d'error s'obté com un resultat parcial de l'algorisme d'estimació del moviment, i per tant no afegeix cap càrrega de còmput. Els resultats experimentals validen aquesta extensió de l'algorisme d'estimació del moviment que possibilita la visió activa.

7.1 Introducció

L'algorisme de recuperació del moviment propi presentat requereix que la projecció de l'objecte estigui sempre en la imatge. Amb una càmera passiva això és sovint massa restrictiu, ja que els moviments que pot realitzar el robot es limiten gairebé només a aproximacions i recessions de l'objectiu. Una de les solucions més prometedores que s'ha considerat és proporcionar capacitat d'orientació a la càmera i dissenyar un algorisme de control per conservar la projecció de l'objectiu centrada en la imatge (o com a mínim dins de la imatge) en la seqüència sencera. Un dels problemes principals que presenta l'algorisme de control és que caldrà aplicar guanys diferents depenent de la distància des de la càmera fins a l'objectiu.

En el camp de la visió activa, seguint la inspiració biològica, s'han fet diversos treballs per emular en plataformes artificials dos dels reflexos oculars presents en els organismes vius: el moviment sacàdic i el seguiment suau [114]. El moviment sacàdic implica un moviment ràpid que canvia sobtadament el punt de fixació. El seguiment suau és un moviment dels "ulls" que permet fer el seguiment continu d'un objecte mentre es mou, de manera que sempre estigui centrat en la imatge. Els apartats següents es desenvolupen centrant l'interès en aquest darrer moviment.

Una càmera activa té un grau de llibertat addicional, comparat amb l'ull biològic: el *zoom*.¹ Tot i que sembla una capacitat molt interessant en un sistema de visió activa els treballs fets fins ara s'ocupen o bé del *zoom* o bé del moviment d'orientació, però poques vegades del moviment conjunt. Una de les causes principals és que les característiques que ressalten en una imatge obtinguda amb un *zoom* determinat no tenen perquè ser

¹L'acomodació, el moviment de l'ull per enfocar objectes a distàncies diferents, es pot identificar amb el mecanisme d'enfocament de les lents, però el moviment del *zoom* no té referent biològic.

salients en una segona imatge obtinguda amb un valor de *zoom* diferent. En general, aquestes característiques no són invariants a canvis de *zoom*. Una altra de les causes és que el canvi de *zoom* canvia els paràmetres de calibratge de la càmera i complica els algorismes de calibratge i autocalibratge. Finalment, el control d'un mecanisme de capcineig i guinyada quan hi intervé el *zoom* també és més complicat. És per això que molts dels treballs amb càmera activa només tenen en compte els moviments d'orientació de la càmera i no el *zoom* [40, 105, 114], o bé els algorismes proposats són estratificats, separant clarament el moviment del *zoom* i el de l'orientació [129].

Des del punt de vista del control sovint s'ha tractat diferent el moviment sacàdic i el seguiment suau. L'estratègia ha estat, en general, generar dos algorismes de control diferents i definir un tercer algorisme per canviar entre aquestes dues estratègies de control. Generalment no es considera una altra de les característiques biològiques: la resolució de l'ull biològic no és uniforme, mentre que la del sensors comercials sí que ho és (el que s'anomena visió foveal).² Encara que s'ha implementat, implícitament i explícitament [82, 104], a efectes pràctics la visió foveal és convenient només en alguns casos [114].

El capítol s'estructura de la manera següent. A la secció 7.2 s'introdueix l'ús del *zoom* per compensar la translació del robot, i preservar així la mida de la projecció de l'objecte en la imatge. La funció d'error que permet implementar l'algorisme de control es presenta a la secció 7.2.1. A la secció 7.2.2 es discuteixen diferents implementacions de l'algorisme de control i quins canvis cal introduir a l'algorisme d'estimació del moviment. Els resultats experimentals es presenten a la secció 7.2.3. Els canvis d'orientació de la càmera es tracten a la secció 7.3. Primer s'exposen els diferents mètodes i es compara la funció d'error que es proposa amb la més utilitzada normalment (secció 7.3.1). A la secció 7.3.2 es mostra com, quan es combina el moviment d'orientació amb el control del *zoom*, la resolució en el pla de l'objecte no canvia. L'algorisme per controlar el capcineig i la guinyada es presenta a la secció 7.3.3, i els experiments, realitzats amb un robot mòbil, es presenten a la secció 7.3.4. Finalment, les conclusions i treballs futurs s'exposen a la secció 7.4.

²Cal esmentar també els sistemes purament electrònics, on es pren una regió d'interès de la imatge, s'expandeix i se serveix com si fos la imatge presa per la càmera, de manera que dóna la sensació (o il·lusió) que es disposa d'un sistema mecànic d'orientació [136].

7.2 Compensant la translació amb el *zoom*

Si s'imagina la situació d'un robot que ha fixat com a objectiu un objecte que és davant seu, és fàcil veure que el moviment de translació que pot fer sense perdre de vista l'objecte és limitat. La projecció de l'objecte a la imatge ha de tenir una mida adequada, ni gaire petita ni gaire gran, per tal que l'algorisme de seguiment pugui funcionar. Això implica que la distància màxima entre el robot i l'objectiu depèn dels paràmetres de l'òptica de la càmera i de la mida de l'objectiu.

L'automatització del control del *zoom* és una opció molt prometedora per a sistemes de visió en aplicacions generals i en aplicacions robòtiques en particular. Desafortunadament, el control del *zoom* d'una càmera no ha rebut l'atenció que es podria esperar, tenint en compte com enriqueix les competències d'un sistema de visió. La possibilitat de canviar la mida de la projecció a la imatge d'un objectiu no només permet analitzar objectes a una resolució més alta [68, 124], sinó que també permet millorar el seguiment [64] i, en conseqüència, l'estimació del moviment 3D i els resultats de reconstrucció [118]. L'interès més gran per al cas que es presenta és que modificar activament el paràmetre de *zoom* permet moviments del robot molt més grans del que seria possible amb una càmera de distància focal fixa, ja que es pot controlar la mida de la projecció de l'objectiu a la imatge perquè sigui favorable per a l'algorisme de seguiment. El més natural per aconseguir aquest objectiu és trobar un algorisme que permeti compensar els moviments de translació del robot amb canvis de *zoom*. Per tant, mesures de l'estil de l'àrea ocupada per la projecció de l'objecte en la imatge no són adequades. Es pot observar a la figura 7.1 com, utilitzant el model estenopec i les condicions de visualització de perspectiva dèbil, els canvis en la distància focal provoquen els efectes desitjats: la imatge d'un objecte es pot conservar si es troba la distància focal f' que neutralitza l'efecte de la variació de distància $Z_0 - Z'$ entre la càmera i l'objecte.

S'han proposat estratègies diferents per al control del *zoom*. Fayman. *et. al.* [52, 53] consideren un objectiu planar i un robot que es pot traslladar només al llarg de l'eix òptic. Per obtenir una projecció en la imatge de dimensió constant proposen una tècnica, que anomenen *zoom tracking*, dirigida a conservar la proporció f/Z entre la distància focal i la distància a l'objectiu. Fan servir un model de càmera de lent gruixuda i suposen coneguts de manera precisa els paràmetres de calibratge de la càmera.

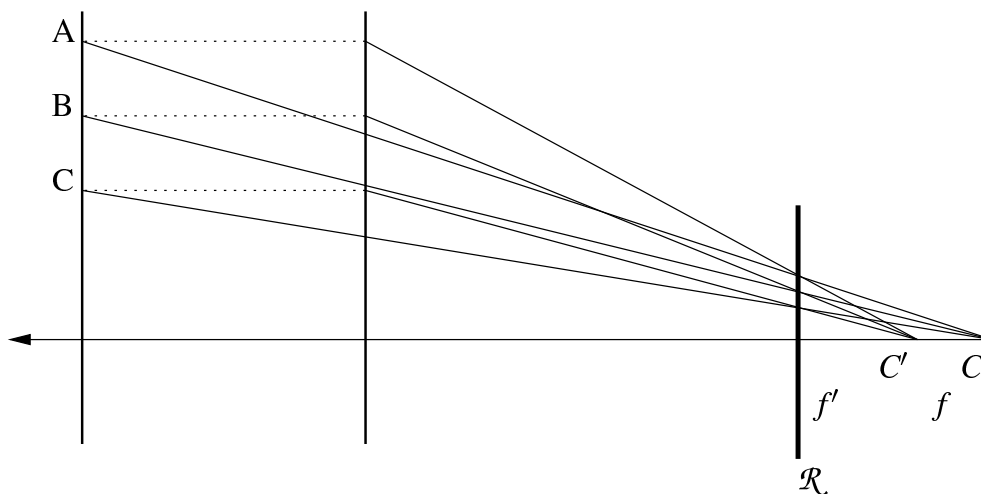


Figura 7.1: Quan la distància entre una càmera i un objecte frontoparal·lel al pla de la imatge canvia de Z a Z' , es pot trobar una nova distància focal f' que fa canviar el centre de projecció de C a C' de manera que la projecció de l'objecte al pla de la imatge \mathcal{R} sigui la mateixa.

Els autors sostenen que, en un sistema de visió activa, les translacions perpendiculars a l'eix òptic es poden estabilitzar amb tècniques de registre d'imatges [23], però desafortunadament no es proporciona cap solució pel que fa a les rotacions. Les peticions de *zoom* es calculen amb els components divergents o convergents del flux òptic (és a dir, només amb mesures en la imatge), o també utilitzant la profunditat proporcionada per un sensor d'abast (és a dir, depenent de la informació de l'escena).

Tordoff i Murray [135, 136] també encaren el problema de fixar la mida de l'objectiu en la imatge. Consideren el moviment dels robots en general pel que fa a la càmera, tant el model perspectiu com el model afí. Amb el primer model, el model perspectiu, només tracten el cas de càmeres que fan rotacions pures (com per exemple les càmeres de seguretat o vigilància). Això es deu al fet que l'algorisme que proposen necessita fer un autocalibratge continu i aquest autocalibratge es fa utilitzant el mètode d'Agapito *et. al.* [42], que només funciona en el cas de càmeres que roten. L'algorisme per generar demandes al *zoom* que proposen també conserva la proporció f/Z . Els autors identifiquen problemes de l'algorisme quan l'objectiu és planar, llunyà, o en situacions en què els efectes de perspectiva no són presents o són discrets (comuns en les càmeres de seguretat o vigilància).

Amb el segon model, el model de càmera afí, no cal el procediment d'autocalibratge.

7. Estimació del moviment en el context de la visió activa

Per trobar l'objecte en l'escena es fa servir el mètode de la transferència afí [134], amb l'avantatge que no cal veure tots els punts al llarg de la seqüència d'imatges³ [66, 96]. Amb aquest enfocament afí, la necessitat de calcular la profunditat d'un punt únic es canvia per la recuperació de l'escala de base afí. En aquest cas, quan s'utilitza un model de càmera afí en general i de perspectiva dèbil en particular, f/\bar{Z} és la relació que s'ha de conservar, on \bar{Z} és la profunditat mitjana de tots els punts que s'utilitzen de l'objecte. Tordoff [135] demostra que conservar aquesta relació és un avantatge en condicions sorolloses, ja que la relació depèn d'una propietat de la projecció global en comptes d'una propietat dependent de punts individuals. Per calcular l'escala proporciona quatre algorismes diferents: dos de basats en la imatge i dos de basats en l'escena, i un altre algorisme per triar entre els anteriors segons les condicions de visualització i així ser capaç d'afrontar objectius planars. Separa els algorismes de control de l'atenció i de demanda de *zoom*, tot i que el darrer utilitza informació que proveeix el primer. Breument, el procediment que proposa és el següent. Primer cal un preprocessat de la imatge per tal de detectar punts singulars i separar el fons de l'escena de l'objecte. Llavors, els punts seleccionats s'utilitzen per computar la transferència afí, i finalment, s'escull un dels algorismes proposats per recobrar l'escala afí. Un dels problemes més grans és el de decidir quins són els punts que convé utilitzar durant el càlcul de la transformació afí. Recentment, Micheloni i Foresti [102] han provat de donar-hi una solució proposant un procediment d'agrupació dels punts.

Denzler *et. al.* [45] proposen un mètode basat en l'entropia per trobar el paràmetre òptim de *zoom* per tal de reduir la incertesa de posicionament (dins del marc del filtre de Kalman) d'un objecte 3D que se segueix. A efectes pràctics, guiar la distància focal amb la incertesa té l'efecte sobre la imatge d'allargar la distància focal (*acostar-se*) quan la càmera està estàtica, i reduir-la (*allunyar-se*) quan la càmera està en moviment. Aquesta aproximació necessita que l'objectiu sigui visible simultàniament per dues càmeres. Els autors resolen aquest problema no actualitzant l'estimació de l'estat del filtre quan l'objectiu es perd amb una de les dues càmeres. Desafortunadament, el seu algorisme no sembla que tingui avantatges notables en l'estimació de posició quan el moviment és lent, ja que els errors de calibratge i el soroll d'adquisició tenen un impacte molt important. Tanmateix, com manifesten, l'ús del *zoom* permet augmentar la

³Hayman [64] ha demostrat l'eficàcia de l'algorisme de transferència afí per fer seguiment quan s'utilitzen sistemes, tant monoculars com estèreo, amb *zoom*.

mida de la projecció de l'objectiu en la imatge augmentant generalment la robustesa de l'algorisme de seguiment. Aquesta és justament la motivació del treball que es presenta a continuació.

Un dels problemes amb valors grans de distància focal és que el camp de visió és petit, cosa que sovint ocasiona imatges desenfocades. Una opció és fer servir el mecanisme d'autofocus present en algunes càmeres, però no és gaire utilitzat perquè fa variar la distància focal i complica el calibratge.⁴ El desenfocament per a algorismes basats en correspondències de punts sovint pot ser crític, ja que provoca un nombre menor de correspondències positives i un nombre més alt de correspondències errònies. Tanmateix, amb l'algorisme de seguiment utilitzat, basat en contorns actius, tot i haver-hi desenfocament encara es poden obtenir bones estimacions del moviment [99].

7.2.1 Generant les peticions per al *zoom*

Entre les estratègies proposades fins ara es poden distingir dos grups segons la informació que utilitzen: basades en l'escena 3D o basades en la imatge 2D. Generalment, en el cas d'un objectiu planar, els mètodes 3D solen fallar perquè el moviment planar afí és ambigu.⁵ En l'algorisme presentat s'ha imposat la restricció de tenir l'objectiu planar frontoparal·lel a la càmera en la primera imatge de la seqüència, amb la qual cosa s'elimina aquesta ambigüitat. Al capítol 2 s'ha mostrat que la translació que es produeix al llarg de l'eix òptic es pot recuperar a partir del vector de forma afí (2.13). Aquest valor es pot utilitzar com a mètode per generar les demandes per al *zoom*.

Els mètodes basats en la imatge 2D se serveixen de càlculs i mesures en la imatge. L'efecte de fer *zoom* per un factor f'/f és la translació del punt de la imatge u al llarg d'una línia recta que parteix del punt principal u_0 de la càmera i arriba al punt $u' = \frac{f'}{f}u + (1 - \frac{f'}{f})u_0$. A efectes pràctics, això es pot explicar com multiplicar la matriu de calibratge que correspon a la primera imatge de la seqüència pel factor f'/f . Aquest factor es pot aproximar amb l'escala ρ de l'afinitat que relaciona dues vistes. Assumint

⁴Hi ha algorismes que aprofiten aquesta autoregulació de la distància focal [59].

⁵Si no es coneix la posa 3D, poden projectar-se poses diferents de l'objecte de la mateixa manera en la imatge. Per exemple, una rotació R_x més una translació d'aproximació pot formar la mateixa projecció en la imatge que una rotació R_y més un moviment d'allunyament.

7. Estimació del moviment en el context de la visió activa

una raó d'aspecte unitària, ρ es pot recuperar a partir de la matriu fonamental afí [62]

$$F_A = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix}$$

com

$$\rho^2 = \frac{c^2 + d^2}{a^2 + b^2}$$

on $\rho > 0$ per definició.

Tradicionalment, F_A ha estat calculada, a partir de correspondències de punts en la imatge, com el vector singular $\mathbf{N} = (a, b, c, d)^T$ corresponent al valor singular més petit d'una matriu construïda amb les correspondències de punts normalitzades. Com a mínim es necessiten *quatre* correspondències de punts no coplanars.

En canvi, amb la parametrització de la deformació basada en la matriu de forma (2.12) que s'ha introduït, es pot calcular l'escala actual de la deformació afí en relació amb el contorn inicial com una funció del valor singular més gran λ_1 en la descomposició de SVD de $\mathbf{M}\mathbf{M}^T$. La translació escalada en l'eix Z es pot calcular com (2.19)

$$\frac{T_z}{Z_0} = \frac{f'}{f} \frac{1}{\sqrt{\lambda_1}} - 1. \quad (7.1)$$

Per tal d'obtenir una mesura de l'escala de la deformació que ha patit la projecció del contorn respecte de la imatge original cal eliminar l'efecte del canvi de distància focal. Així, independentment de com s'hagi anat canviant el *zoom*, s'obté una mesura de la diferència en escala entre el contorn actual i l'original. Imposant $f'/f = 1$ s'obté

$$e = \frac{1}{\sqrt{\lambda_1}} - 1, \quad (7.2)$$

com a funció d'error que s'ha d'utilitzar en l'algorisme de control del *zoom*. No és directament l'escala de l'afinitat, però està linealment relacionada amb l'homotècia corresponent. Observeu que, en l'estimació del moviment propi d'un robot,⁶ aquest valor ja es computa, per tant generar les demandes per al *zoom* no suposa cap sobrecost per al procés d'estimació del moviment. Ara, però, es té la possibilitat d'aprofitar els avantatges d'una càmera amb *zoom*. És fàcil veure, tal com s'ha derivat, que es tracta

⁶Comparat amb [136], només es recupera el moviment propi (codificat a partir del vector de forma), i no l'estructura o la projecció 3D. A més, tal com està dissenyat l'algorisme de seguiment, tampoc cal separar el fons de l'escena de l'objecte.

d'una mesura 2D. Canviant els valors de la distància focal amb aquesta funció d'error no cal cap calibratge de la càmera ni tampoc cap estimació de la distància inicial Z_0 . A més, tal com va determinar Tordoff [135], la idea de recobrar una propietat de la projecció global en comptes d'una propietat dels punts individuals és un avantatge en condicions sorolloses. Això és justament el que s'obté utilitzant el model de càmera de perspectiva dèbil i la introducció de l'espai de forma afí.

7.2.2 Implicacions per a l'algorisme de recuperació del moviment

A partir de la mesura d'error proposada anteriorment, s'han dissenyat diversos algorismes de control. Aquests algorismes s'han assajat en una càmera Sony EVI-D30, que disposa de *zoom* i d'un protocol de comunicació amb el controlador del *zoom* anomenat VISCA. En aquest protocol es pot fer una petició al *zoom* per posició o bé per velocitat. En les proves realitzades, l'algorisme de control per velocitat no proporcionava cap avantatge sobre els de control per posició. Això és a causa que el protocol Visca de comunicació només proporciona la possibilitat de demanar una sèrie limitada de velocitats (codificades en nivells que van de l'1 al 8). Quan l'algorisme de control demana una velocitat intermèdia, per exemple 2.35, no es pot transmetre aquesta petició, sinó que cal arrodonir-la, en l'exemple a 2 o 3; depenent de l'arrodoniment, això provoca que el control dissenyat sigui o bé poc reactiu i lent, o bé massa reactiu i per tant provoqui sobrepuigs. Això introdueix inestabilitats en aquest tipus d'algorismes de control.

Els requisits de precisió, en termes de la mida de la projecció del contorn de l'objectiu en la imatge, no són gaire estrictes (l'objectiu d'allargar el seguiment mantenint la mida de l'objecte dins d'uns límits no és molt restrictiu). És per això que un algorisme de control per posició només amb component proporcional a és suficient. Perquè funcioni cal posar en correspondència les translacions euclídiades amb els canvis de la distància focal. Aquesta correspondència es pot trobar amb un procés molt simple al començament de la seqüència. Després d'inicialitzar el contorn definint-ne els punts de control, es provoca un impuls a l'entrada del control per posició del *zoom* i s'obté la relació entre el canvi de *zoom* i l'error computat per la funció d'error. Cal fer notar que per dur a terme aquest procés no cal cap calibratge de la càmera ni tampoc cal disposar d'informació sobre l'escena.

7. Estimació del moviment en el context de la visió activa

Atès que s'està utilitzant una càmera no calibrada no es coneix el valor de la distància focal, i per tant, tampoc la proporció entre les distàncies focals f'/f . Tanmateix, es pot utilitzar la proporció entre les diferents posicions del *zoom* per generar les demandes al mecanisme de *zoom* de la càmera. S'assumeix que una funció lineal relaciona la distància focal i la demanda de *zoom*. Aquesta és una bona aproximació quan les posicions de *zoom* no estan en la zona de màxima distància focal [87]. Com a conseqüència, a la pràctica només cal restringir el rang de valors possibles del *zoom* eliminant les posicions extremes.

Quan l'algorisme de seguiment es combina amb una càmera amb *zoom*, la seqüència d'imatges que prové d'un moviment del robot es veu òbviament alterada, i es produeix una seqüència d'imatges diferent a la que es produiria amb una càmera de focus fix. Si es controla el *zoom* amb l'expressió (7.2) es fa que el *zoom* corregeixi la mida de la projecció compensant les translacions al llarg de l'eix òptic de la càmera, es pot veure fàcilment que la translació T_z computada depèn tan sols de la raó $\frac{f'}{f}$.

7.2.3 Resultats experimentals

L'esquema que s'ha seguit en els experiments és el següent. De manera semblant als experiments presentats en capítols anteriors, el contorn actiu s'inicialitza manualment. Això defineix una matriu de forma (2.12) que s'utilitza com a referència per a tots els moviments. Per a cada iteració de l'algorisme, s'adquireix una imatge nova. Un filtre Kalman calcula la deformació afí que millor encaixa el contorn original amb el contorn actual. A partir d'aquesta deformació afí es pot extreure tota la informació exigida: la demanda per al *zoom* i el moviment 3D.

Es fa un primer experiment de referència amb una càmera de distància focal fixa. L'objectiu és verificar que la funció d'error proposada és invariant a les rotacions, és a dir, que utilitzant aquesta funció no es demandaran canvis de *zoom* quan tan sols es produeixin rotacions de la càmera. Per fer-ho, cal dissenyar un experiment en què la càmera giri al voltant de l'objecte sense canviar la distància Z_0 entre càmera i objecte. Gràcies al fet que el moviment que es computa en realitat és relatiu, la seqüència d'imatges resultant és equivalent tant si es mou la càmera i es deixa l'objectiu fix, com si es deixa la càmera fixa i el que es mou és l'objectiu. Per a aquest experiment s'utilitza un braç robòtic amb un objectiu artificial a l'element terminal (figura 7.2) i

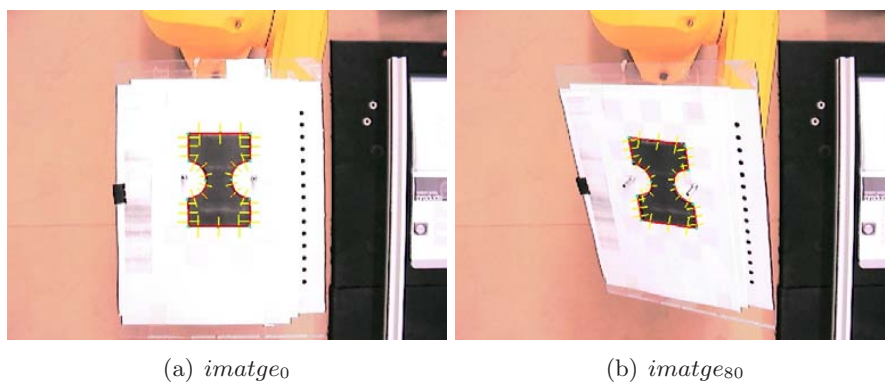


Figura 7.2: (a) Primera imatge i (b) imatge rotada al màxim per a l'experiment de rotació.

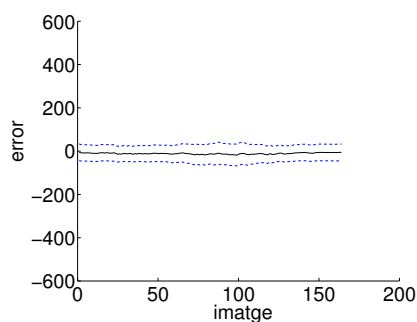


Figura 7.3: Valors d'error computats per ser utilitzats per generar demandes de *zoom* en un experiment que implica només rotació de la càmera. Es pot veure que la funció és robusta a rotacions ja que no s'obté cap error.

una càmera fixa. S'escull un objectiu fàcil de seguir, però l'algorisme de seguiment és el mateix que s'ha fet servir en altres ocasions en entorns no preparats [2, 17]. Amb aquesta configuració, un cop calibrada la posició entre la càmera i l'element terminal del robot, es pot assegurar una rotació pura de l'objectiu.

S'han realitzat diverses rotacions de l'objecte amb angles diferents. Aquí es mostra un dels experiments més representatius. L'experiment implica una rotació de l'objecte sobre un eix que passa pel centre de l'objecte a 45° d'inclinació sobre el pla frontoparal·lel al pla de la imatge de la càmera. La seqüència comença amb l'objecte frontoparal·lel (rotació de 0°), primer es gira fins a 40° i finalment es retorna a la posició inicial. La figura 7.3 mostra els valors computats per a la funció d'error (7.2) dissenyada per controlar el *zoom*. Com es pot veure s'obté un valor proper de zero fins i tot al mig de la seqüència, quan l'angle de rotació és de 40° . Així, tal com s'esperava, es pot dir que

7. Estimació del moviment en el context de la visió activa

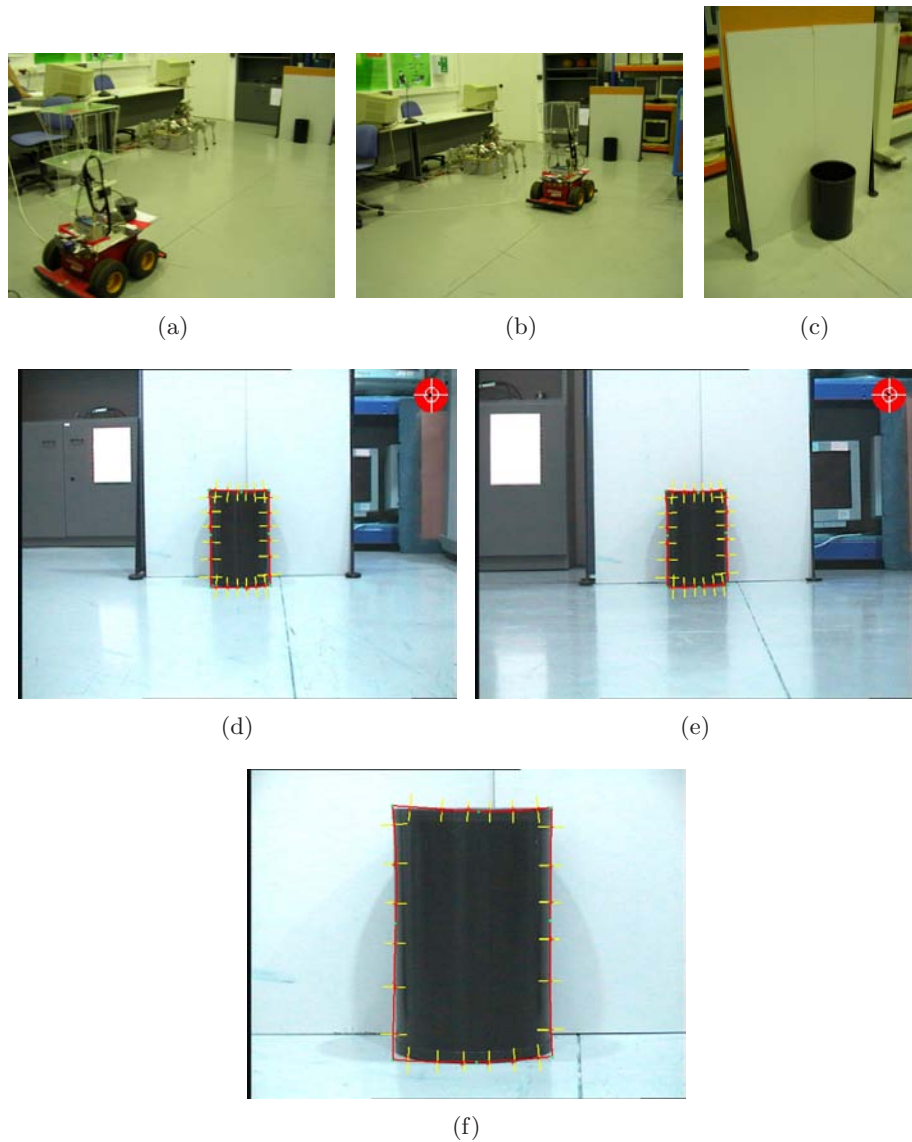


Figura 7.4: Imatges que il·lustren l'experiment realitzat. (a) Robot en la posició inicial i (b) robot en la posició final, després de fer una translació. (c) Imatge detallada de l'objectiu emprat: una paperera. (d) Imatge inicial de la paperera amb un contorn actiu ajustat a la projecció del contorn del pot. (e) Imatge adquirida en la posició final després que el *zoom* hagi canviat amb l'algorisme de control proposat. (f) Imatge que s'hauria obtingut en la posició final si no s'hagués controlat el *zoom*.

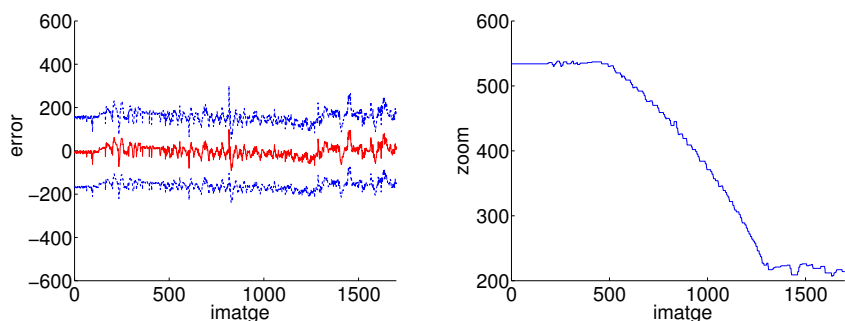
la funció d'error és robusta a rotacions.

L'experiment següent es fa amb un robot mòbil Pioneer AT. Se l'ha proveït d'una càmera EVI-D31, que està equipada amb una lent motoritzada que permet fer *zoom*

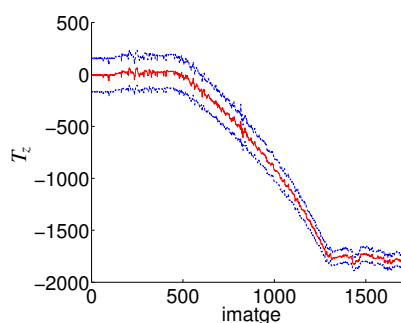
i amb un dispositiu que permet inclinar i fer girar la càmera. Per a aquest experiment, l'orientació de la càmera es manté a un valor constant i només actua el *zoom*. L'algorisme de control és per posició, en un rang que va de 0 a 1024. Tal com s'ha esmentat anteriorment, a l'inici es fa un calibratge molt senzill canviant la posició de *zoom* i mesurant la resposta de la funció d'error. Així es pot computar fàcilment un valor aproximat pel guany proporcional. La figura 7.4(a) mostra el robot en la posició inicial. El vehicle dibuixa una trajectòria lineal enfocant l'objectiu. La localització final del robot es pot veure a la figura 7.4(b). L'objectiu utilitzat és una paperera cilíndrica. La figura 7.4(c) és una vista detallada de l'objectiu i la figura 7.4(d) mostra el pot d'escombraries des del punt de vista del robot en l'instant inicial, amb un contorn actiu inicialitzat al voltant de la projecció del contorn de l'objectiu a la imatge. Mentre el robot s'està movent, per a cada imatge adquirida l'algorisme de seguiment calcula la deformació afí del contorn actual respecte de l'inicial, i computa el moviment propi. A velocitat d'adquisició, amb la implementació actual de l'algorisme a 20 fps, el sistema és capaç de generar una demanda per al *zoom*. La figura 7.4(e) mostra l'última imatge adquirida pel robot en la posició final i mostra que el control de *zoom* ha anul·lat eficaçment el moviment d'aproximació. La figura 7.4(f) mostra la imatge que resultaria si el control de *zoom* es desactivés. Com es pot veure, la projecció de l'objectiu és molt més gran, i si el robot continués aproximant-se l'objectiu aviat es projectaria fora del pla de la imatge.

La figura 7.5(a) mostra la funció d'error computada durant l'experiment anterior. Es pot observar que està centrada en l'entorn de 0, i que els errors es mantenen sempre petits. A la figura 7.5(b) es poden observar les posicions de *zoom* que resulten de les peticions calculades per l'algorisme de control. Atès que la trajectòria és d'aproximació, els valors del *zoom* decreixen per obtenir distàncies focals cada vegada més curtes. La translació del robot es mostra a la figura 7.5(c). Per obtenir informació mètrica caldria mesurar la distància inicial entre la càmera i l'objectiu, i caldria calibrar la càmera i trobar la distància focal real. Com que s'està utilitzant una càmera amb *zoom*, caldria conèixer la relació entre totes les posicions del *zoom* i la distància focal corresponent.

7. Estimació del moviment en el context de la visió activa



(a) La funció d'error al centre i la variància per sobre i per sota (b) Posició de *zoom* demanada per l'algorisme



(c) T_z reconstruïda al centre i variància per sobre i per sota

Figura 7.5: Resultats de l'experiment amb el control per posició del *zoom*.

7.3 Moviment d'orientació de la càmera

En el paradigma de la visió activa apareixen dues qüestions principals a què cal donar resposta. Primer, cal determinar on es vol fixar l'atenció. Per donar resposta a aquesta pregunta primer cal saber quina és la tasca que cal accomplir.⁷ En el context presentat la resposta és immediata: cal seguir el contorn de l'objecte que es fa servir per deduir el moviment. Segon, cal descriure els detalls del controlador que ha de permetre fixar aquesta atenció en el sistema robòtic. Sovint, la solució adoptada per al primer problema guia la solució del segon.

Per detectar l'objectiu en la imatge s'han desenvolupat diverses estratègies. Detectar un objectiu que es mou amb una càmera estàcionària no és un tasca difícil. Es pot

⁷Seguint la idea, exposada al capítol 6, que cal entendre la visió activa dels del punt de vista de la tasca [8].

aconseguir amb tècniques senzilles i ben conegudes, com la clàssica d'imatge diferència, o bé amb tècniques una mica més elaborades d'imatge diferència amb finestres d'interès que es poden moure, i de les quals fins i tot es pot canviar la mida [35]. La tasca esdevé més difícil quan és l'observador qui es mou, ja que aquest provoca un moviment en tota la imatge. Una de les possibles solucions és facilitar al màxim el procés de segmentació fent servir fortes restriccions sobre la forma de l'objectiu (cantonades [112], línies [113], etc.), la seva mida [108] o el moviment de la càmera [113].

Una altra de les solucions adoptades és la de suposar que entre imatges consecutives la velocitat del moviment és constant. Aquesta suposició és correcta quan es tracta d'una càmera de vigilància, com les usades a [135], però quan es tracta d'una càmera muntada sobre un vehicle mòbil sovint apareixen moviments indesitjats produïts per vibracions, terrenys irregulars o petites col·lisions [49]. Per tal d'eliminar aquestes pertorbacions es pot fer servir una tècnica d'estabilització d'imatges [13, 58, 86]. Les tècniques d'estabilització tenen en compte tota la imatge i l'objectiu que aquí es planteja és mantenir al centre de la imatge el punt que estava al centre a l'inici. En general, en estabilització no es té coneixement del punt d'interès, i el que s'intenta és modificar les imatges preses amb una camera en moviment perquè sembli estacionària. Tradicionalment, el que es fa primer és estimar el moviment de la càmera i fer-lo servir més tard per rectificar la seqüència d'imatges. Com s'ha vist en els capítols precedents, i en part gràcies a les ambigüitats presents, no és fàcil estimar les rotacions quan el moviment és general.

Les tècniques de zona d'interès⁸ generalment s'apliquen a un objecte determinat. En aquest context, el que es pretén amb la visió activa és mantenir aquest objecte centrat en la imatge [110]. Una de les aportacions més interessants en aquesta línia la van fer Reid i Murray [112] implementant la idea de visió foveal, amb diferent resolució al centre i a la perifèria de la imatge. Ho van dur a terme reduint la resolució de la imatge a les zones perifèriques. En canvi, Kuniyoshi *et. al.* [82] van construir una lent que produïa directament visió foveal. Altres autors han proposat fer servir dues càmeres acoplades mecànicament [138] amb lents diferents, de manera que una té visió detallada i l'altra, amb una lent de distància focal menor, té visió general.

⁸En anglès *fixation*.

7.3.1 Estratègies de control de l'orientació de la càmera

Un cop determinat el focus d'atenció cal fer-ne el seguiment. La solució més senzilla per mantenir la projecció de l'objecte centrada en la imatge és calcular el centroide dels punts que conformen la projecció. Amb la posició del centroide es pot calcular la distància al centre de la imatge,⁹ tant en l'eix X com en l'eix Y . Encara que sembli simple, aquesta aproximació s'ha fet servir extensament [30, 35, 108, 138]. Aquesta és clarament una mesura 2D en la imatge. Alternativament, també es pot utilitzar l'estimació de la posició 3D $X = (x, y, z)$. En aquest cas, la funció d'error d'un controlador pot ser directament proporcional a $\arctan\left(\frac{x}{z}\right)$ en el cas del capcineig, i $\arctan\left(\frac{y}{z}\right)$ per a la guinyada [10].

Es pot abordar el problema des d'un punt de vista diferent observant que els algorismes de control dels motors del sistema d'orientació sovint es dissenyen amb valors de velocitat [35, 114]. Per a aquests tipus d'algorismes el convenient és fer mesures de velocitat, i per això generalment estan associats a algorismes de flux òptic. Una altra solució possible és estimar la posició 2D del punt d'interès i fer servir un model de moviment per estimar el moviment de la càmera i així poder regular la posició de la càmera amb una llei de control des del punt de vista del servocontrol visual [34]. A part dels controladors clàssics, també s'han explorat controladors més complexos, per exemple controladors basats en lògica difusa [59].

Si es considera un sistema amb dues càmeres es pot construir un model del moviment [105], però cal un bon calibratge, tant dels paràmetres intrínsecs com extrínsecs del sistema. En el marc d'un sistema multicàmera amb diferent resolució, també s'han proposat diverses solucions, com per exemple la utilització de dues càmeres fixes per generar les demandes d'orientació d'una tercera càmera que és activa, o bé obtenir la informació d'orientació amb una sola càmera fixa [124]. D'aquests algorismes, n'hi ha que no consideren la informació de calibratge. En aquest cas, el moviment de la càmera generalment es determina amb un procés previ de calibratge qualitatiu, que posa en correspondència els moviments d'orientació amb moviments en la imatge, que depenen, naturalment, de la distància entre la càmera i l'objecte que s'està visualitzant. Com es

⁹En realitat, a qualsevol altre punt de la imatge. El centre és el punt més lògic si l'objectiu és mantenir centrada la projecció dins de la imatge.

mostrarà més endavant, en la solució que es proposa també s'inclou una fase prèvia de calibratge que, com es podrà comprovar, és extremadament senzilla.

Rivlin i Rotstein [114] van observar que amb un controlador lineal es pot implementar una estratègia de seguiment suau que és vàlida en segons quins casos, però que és inadequada en algunes situacions pràctiques. Un dels problemes més comuns és que, mentre en un sistema biològic el control és continu, en un sistema mecànic el control ha de ser necessàriament discret. A més, hi ha el problema afegit dels retards provocats per l'anàlisi de les imatges. És ben conegut que si aquests retards no són suficientment petits es produeixen problemes d'estabilitat en el controlador [9]. Una de les maneres d'augmentar l'estabilitat del sistema és obtenir un model dinàmic suficientment acurat [107], però moltes vegades el cost de trobar aquest model és molt elevat.

L'algorisme de control que es proposa, de manera semblant al proposat anteriorment per al control del *zoom*, utilitza els valors t_x i t_y calculats de l'afinitat (2.11)

$$\mathbf{Q}' - \mathbf{Q} = \mathbf{W}\mathbf{S},$$

on (2.13)

$$\mathbf{S} = (t_x, t_y, M_{11} - 1, M_{22} - 1, M_{21}, M_{12})^\top,$$

que parametrizen el moviment del contorn en la imatge, que representa el pas intermedi abans de calcular la translació de la càmera respecte de l'objecte que es fa servir d'objectiu. Si es fa servir aquesta estimació, sense afegir càrrega de còmput a l'algorisme, s'obté una bona mesura de l'error de posicionament del contorn respecte del centre de la imatge.¹⁰ Es pot veure fàcilment que aquesta és una mesura 2D sobre la imatge.

Es pot fer un senzill experiment per comprovar que aquesta funció d'error és menys sensible al soroll que la tradicionalment usada del centroide. En un experiment simulat es defineix el polígon de control d'un contorn actiu sobre un pla de la imatge, també simulat, i així es pot inicialitzar una matriu de forma. S'afegeix soroll, gaussià i de mitjana zero, en cadascuna de les coordenades dels punts i es calcula, d'una banda, el centroide del núvol de punts, i de l'altra, els vectors de forma que modelen el moviment

¹⁰De fet, respecte de la posició inicial del contorn en la seqüència d'imatges. A efectes pràctics, s'ha implementat una funció que, un cop inicialitzat un contorn, el centra en la imatge abans de començar el procés de seguiment.

7. Estimació del moviment en el context de la visió activa

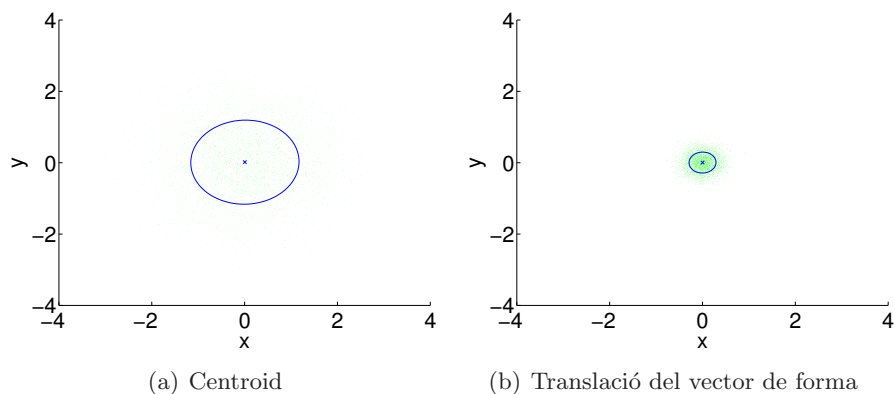


Figura 7.6: Comparació entre el càlcul del centroide i els valors t_x t_y calculats a partir de la deformació d'un contorn. Pel mateix nivell de soroll aplicat a les dades, les estimacions amb el vector de forma són molt menys sorolloses.

generat pel soroll sobre del contorn. El resultat es pot observar a la figura 7.6. Com ja s'ha esmentat al capítol 2, el fet de trobar el vector de forma es pot veure com un procés de minimització que ajusta els punts de control a l'espai definit per la matriu de forma. Així, és natural que part del soroll es cancel·li i el resultat sigui molt menys sorollós en el cas de les translacions estimades amb el vector de forma en comparació amb el càlcul del centroide.

De manera semblant a la solució proposada per al control del *zoom*, un controlador proporcional és suficient per acomplir l'objectiu de moure els mecanismes de capcineig i guinyada quan els requisits de precisió no són elevats [10]. Es coneix bé que només amb un controlador lineal no es pot arribar a una solució òptima [114], i és per això que per a aplicacions que demanen més precisió s'han proposat diversos mètodes que intenten estimar el moviment de la càmera, generalment amb un filtre de Kalman.

El controlador que es proposa (per posició segons una mesura en el pla de la imatge) no seria adequat si també calgués considerar els moviments sacàdics, i caldria implementar un segon controlador per seguir aquests moviments i dissenyar un tercer algorisme de control que decidís quin dels controladors és l'adequat en cada moment (de manera semblant a les propostes de Reid *et. al.* [112] i de Rivlin i Rotstein [114]).

7.3.2 El *zoom* preserva la resolució dels moviments en el pla frontoparal·lel

Cal veure que gràcies al canvi en el *zoom* la resolució en el pla frontoparal·lel es manté constant i l'error i la incertesa de T_x , T_y i R_z ja no depenen de Z_0 .

Després d'una translació lateral del robot es pot calcular el vector de forma que la codifica. Les components t_x i t_y corresponen a les translacions laterals i, tal com s'ha vist, serveixen de funció d'error per a l'algorisme que controla l'orientació de la càmera. Aquest algorisme s'encarrega d'orientar la càmera per tornar a centrar la projecció de l'objecte en la imatge, i per tant, reduir la component del vector de forma que resulta després de l'orientació. Clarament, per a la mateixa translació lateral del robot, es calcula un valor diferent amb l'algorisme de moviment segons que l'objecte estigui proper o estigui allunyat. Observant (2.20 i 2.20) es pot veure que les translacions laterals que s'obtenen estan escalades per la distància inicial Z_0 entre la càmera i l'objecte. A la figura 7.7(a) es pot veure gràficament que una translació lateral quan es veu un objecte proper provoca un desplaçament lateral de la projecció de l'objecte menor que no pas la d'un objecte llunyà.

Observant (2.20 i 2.20) també es pot veure que el canvi de distància focal que es produeix quan es fa *zoom* no afecta el càlcul de les translacions, ja que només depenen de f_0 . Quan es canvia el valor del *zoom* per mantenir la mida de la projecció del contorn en la imatge s'està compensant la diferència entre la distància inicial Z_0 i l'actual. Això fa que una mateixa translació lateral del robot correspongui sempre a la mateixa translació lateral de la projecció de l'objecte en el pla de la imatge (figura 7.7(b)). És a dir, l'efecte de fer *zoom* és que les components t_x i t_y del vector de forma tenen sempre el mateix valor, independentment de la distància de la càmera a l'objecte. En canvi, el moviment d'orientació que ha de fer la càmera és diferent segons l'objecte estigui a prop o estigui lluny.

En general, tal com ja va notar Murray [105], quan es canvia la distància focal de la càmera (per exemple, perquè es fa *zoom*) es canvia el guany del sistema de control de l'orientació, que podria derivar en una resposta massa reactiva o poc reactiva, tant del sistema de control del capcineig com del de la guinyada. La solució ha estat tradicionalment obtenir els paràmetres de calibratge per poder sintonitzar els guanys del sistema de control [64]. Gràcies al fet que els requisits en termes de precisió a l'hora

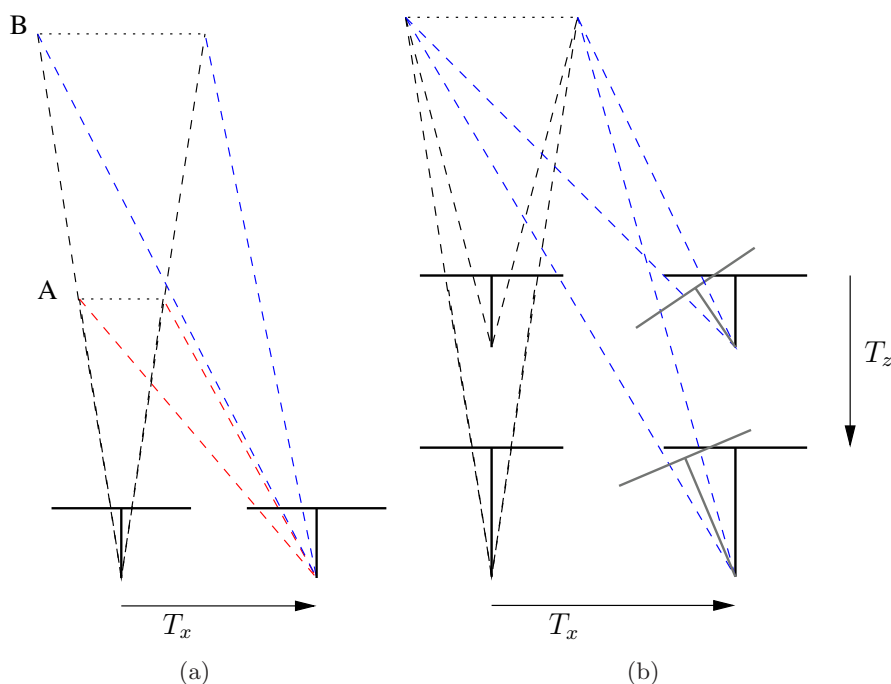


Figura 7.7: (a) La mateixa projecció de dos objectes A i B després d'una translació lateral és diferent segons la distància Z_0 entre la càmera i l'objecte. (b) Quan es canvia el *zoom* per compensar la translació T_z , la mateixa translació lateral T_x del robot correspon a la mateixa projecció en el pla de la imatge. En canvi, l'orientació que cal realitzar per centrar la projecció en la imatge és diferent.

de mantenir l'objecte centrat no són molt estrictes, un senzill calibratge a l'inici de la seqüència és suficient per sintonitzar el control adequadament.

7.3.3 Implicacions per a l'algorisme de moviment

Hi ha diversos mètodes per transformar algorismes dissenyats per a càmeres passives i fer-los vàlids en sistemes de càmera activa. Per no haver de reformular els algorismes s'ha proposat el mètode del pla frontal [113] (també conegut com a càmera virtual [117]). Es pren la direcció de la càmera quan es troba en la posició canònica (sense moviment d'orientació) i es construeix un pla imaginari, perpendicular a la direcció de visió de la càmera i a una distància determinada. Quan la càmera es mou, la direcció de visualització talla aquest pla per un punt. Es pot demostrar que el conjunt centre de rotació-pla frontal d'una càmera activa, i el conjunt centre òptic-pla de la imatge d'una càmera passiva són matemàticament equivalents.

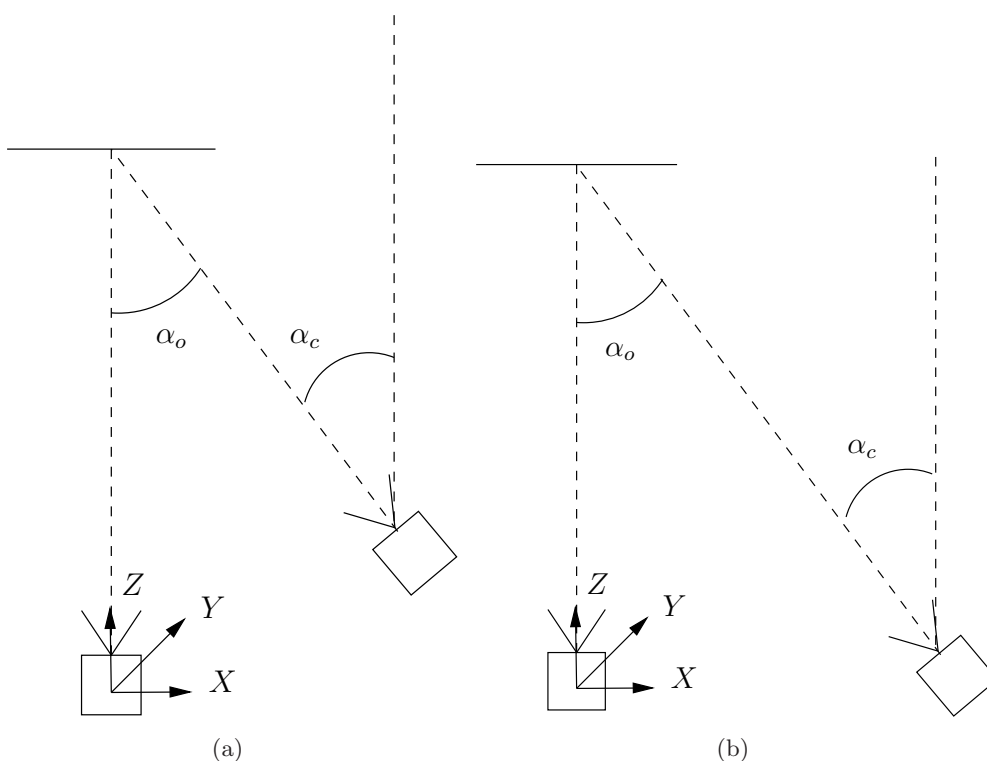


Figura 7.8: Tant el moviment d'una càmera orbitant al voltant d'un objecte (a), com el moviment en línia recta paral·lel a l'objecte (b) fan que la càmera, per mantenir la projecció hagi de rotar un angle α_c . Es pot observar que, des del punt de vista de la càmera, és com si l'objecte hagués rotat $-\alpha_c$.

Es pot estendre el concepte de pla frontal a una esfera, que és la base del mètode de l'esfera d'atenció visual¹¹ [40]. Es pot imaginar una esfera de radi conegut i amb centre al centre de projecció de la càmera. Totes les coordenades es poden referenciar en l'esfera, amb la qual cosa s'obté una representació esfèrica del món. La idea de transformar coordenades planars en coordenades esfèriques no és nova [51] i darrerament s'està fent servir extensament en el context de càmeres omnidireccionals [47].

Observant la geometria del problema (figura 7.8), es pot veure que gràcies a la col·locació dels sistemes de referència, un moviment de guinyada de la càmera provocada per una translació del robot es pot expressar com una rotació en l'eix Y de l'objecte. Per simetria, un moviment de capcineig es pot expressar com una rotació en l'eix X . Gràcies al fet que la formulació del moviment es fa amb el sistema de coordenades a

¹¹En anglès, *gazesphere*.

7. Estimació del moviment en el context de la visió activa

l'objecte resulta fàcil veure que, un cop s'ha recuperat el moviment a partir del vector de forma, només cal afegir el moviment d'orientació realitzat amb els mecanismes de capcineig i guinyada. El problema és que es coneix el signe del moviment induït a la càmera, però no es coneix el signe del moviment de rotació deduït amb el vector de forma (ambigüitat *Necker reversal* [62]). Per tant, sense informació addicional no és possible determinar si per trobar el moviment resultant cal sumar o sostroure les dues rotacions.

El que convé en aquest punt del treball és definir el sistema de coordenades amb origen en la càmera, en lloc de tenir l'origen en l'objecte. Aquesta és una formulació més natural en robòtica mòbil i permet, d'una banda, comparar i fusionar el moviment amb l'obtingut amb altres sensors o tècniques (capítol 2), i de l'altra, localitzar el robot en l'entorn. Les fórmules per representar el moviment calculat a partir d'un vector de forma amb el sistema de referència centrat en la càmera ja han estat presentades al capítol 2, en concret les translacions:

$$\begin{aligned}\frac{T_x}{Z_0} &= \frac{t_x - (u' - u)}{f\sqrt{\lambda_1}} - R_{13}, \\ \frac{T_y}{Z_0} &= \frac{t_y - (v' - v)}{f\sqrt{\lambda_1}} - R_{23}, \\ \frac{T_z}{Z_0} &= \frac{f'}{f} \frac{1}{\sqrt{\lambda_1}} - R_{33}.\end{aligned}$$

En aquest cas, la precisió en la computació de les rotacions R_x i R_y adquireix gran importància, ja que, com es pot observar, se'n fa servir el valor per computar les translacions. Les expressions per calcular les rotacions, tal com ja va notar Martínez [97], són les mateixes independentment de si el sistema de referència està centrat en la càmera o en l'objecte.

Tal com ja s'ha vist en experiments de capítols anteriors, les rotacions R_x i R_y del robot al voltant de l'objectiu no es poden recuperar correctament fins que no són d'aproximadament 10° com a mínim. Això és degut al fet que les deformacions que pateix el contorn quan el robot rota al voltant de l'objectiu són molt petites, i generalment queden per sota de la precisió del sistema de visió i amagades pel soroll en l'adquisició de la imatge. Quan es computa el moviment amb el sistema de coordenades fixat a la càmera, les components de rotació R_x i R_y són petites, i s'obté un estimació dolenta de les translacions. La incertesa calculada en aquests casos (capítol 5) és coherent amb

7.3 Moviment d'orientació de la càmera

aquesta observació, ja que és gran quan no hi ha rotació i disminueix a mesura que el valor de la rotació creix.

Per conèixer l'orientació dels mecanismes de capcineig i guinyada s'ha fet servir tradicionalment la seva odometria [40, 113]. Una de les aproximacions més comunes quan es treballa amb un dispositiu d'orientació de la càmera és la de suposar que la rotació de la càmera es produeix sobre uns eixos que pertanyen al pla de la imatge [33, 41, 65, 129] i que són ortogonals. En aquest cas, la projecció d'un punt en la càmera es pot modelar de forma matricial com

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{P}\mathbf{R}_c\mathbf{R}_g \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (7.3)$$

on $[XYZ]^T$ és un punt en coordenades de món, \mathbf{R}_c és la rotació del sistema de capcineig, \mathbf{R}_g és la rotació del sistema de guinyada i \mathbf{P} és la matriu de projecció de la càmera. S'ha demostrat [64] que quan la distància entre la càmera i l'objecte és gran, els errors provocats per aquesta simplificació són negligibles.

En realitat, a causa dels desplaçaments mecànics entre els motors que fan el moviment de capcineig i guinyada, els eixos de rotació (figura 7.9(a)) no tan sols no són al pla de la imatge sinó que ni tan sols es creuen [38]. En aquest cas, es pot proposar un model més complex que tingui en compte les translacions entre els eixos de rotació i els possibles desalineaments. Si es modela el dispositiu d'orientació amb eixos arbitraris a l'espai, la projecció es pot expressar com

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{P}(\mathbf{R}_c(\mathbf{R}_g + \mathbf{t}_t) + \mathbf{t}_p) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (7.4)$$

on les coordenades al pla de la imatge es calculen normalitzant el resultat [62] (x/z , y/z).

Hi ha diversos mètodes per trobar els valors de les matrius de transformació entre els diferents sistemes de referència (7.9(b)), generalment desenvolupats per a sistemes de visió activa estèreo [40] i per a sistemes multicàmara [38]. Els patrons petits són adequats per a càmeres passives [149] i volums de treball reduïts, i generalment no ho són per a càmeres actives. El que cal per obtenir un bon calibratge és mostrejar al

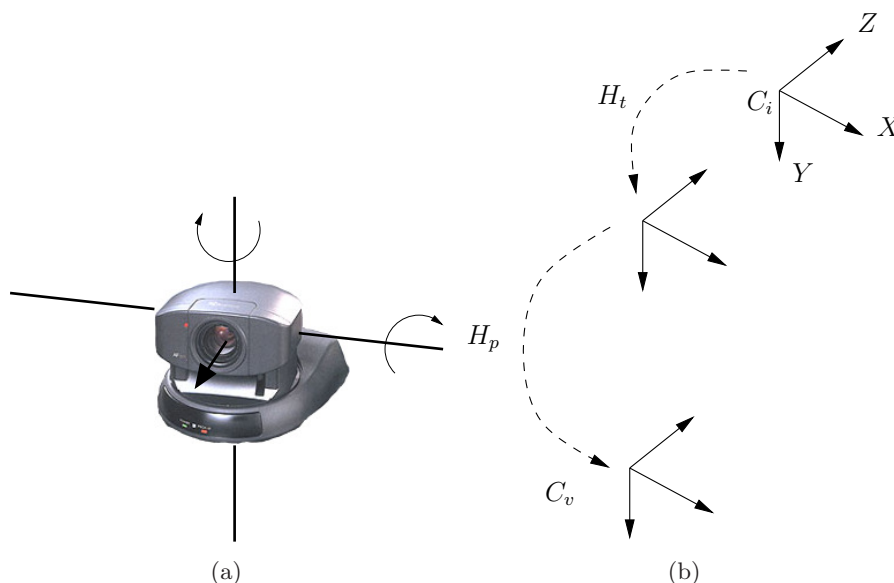


Figura 7.9: (a) Esquema dels eixos de rotació dels mecanismes d'orientació (b) Esquema dels diferents sistemes de coordenades del sistema de càmera activa

màxim l'espai de treball, de manera que s'han proposat mètodes que mouen el patró de calibració de manera precisa per l'entorn, d'altres que utilitzen un patró de calibració molt grani d'altres que movent un objecte conegut (o un punt lluminós) per l'espai aconseguen posar en correspondència totes les imatges.

El sistema que es proposa es planteja sense coneixement dels paràmetres de calibratge de la càmera, i conseqüentment, tampoc no se suposen coneguts els de calibratge del sistema d'orientació, de manera que s'utilitza el model simplificat.

7.3.4 Experiments

L'objectiu de l'experiment que es presenta és validar l'algorisme de control proposat per als mecanismes de capcineig i guinyada quan també es canvia el valor del *zoom* de la càmera. De la mateixa manera, també es vol validar, d'una manera qualitativa, la proposta que s'ha realitzat per incorporar el moviment d'orientació de la càmera a l'algorisme de recuperació del moviment propi. Aquest experiment s'ha dut a terme en les mateixes condicions que el presentat anteriorment (figura 7.4), però en aquest cas, a més del *zoom*, també es controla la posició dels mecanismes de capcineig i guinyada. La trajectòria també és en línia recta, però, a diferència de l'experiment anterior, l'objectiu

7.3 Moviment d'orientació de la càmera

no és davant del robot, sinó que s'ha desplaçat a l'esquerra del robot. D'aquesta manera, el moviment del robot provoca un desplaçament lateral de la projecció de l'objectiu en la imatge, i si l'orientació de la càmera no es corregeix la càmera surt ràpidament del pla de la imatge. Tal com es pot observar a la figura 7.10, la translació entre la posició inicial (figura 7.10(a)) i la final (figura 7.10(b)) provoca un clar canvi d'orientació de la càmera.

Les especificacions del sistema d'orientació de la càmera donades pel fabricant són, per al capcineig, un moviment de $\pm 100^\circ$ a $80^\circ/sec$, i per a la guinyada, un moviment de $\pm 25^\circ$ a $50^\circ/sec$ ¹². Tanmateix, les consignes que cal enviar al controlador per canviar l'orientació no estan expressades en graus: per al capcineig, el moviment està expressat en el rang [0000 - 035D] cap a la dreta i [FFFF - FCA4] cap a l'esquerra, i per a la guinyada, en el rang [0000 - 011E] cap amunt i [FFFF - FEE3] cap avall. Això fa que la resolució dels dos sistemes d'orientació sigui diferent, i que també sigui diferent segons el sentit de gir: 8.6 passos/o per al capcineig a la dreta i 8.6 passos/o per al capcineig a l'esquerra, i 11.44 passos/o per a la guinyada amunt i 11.4 passos/o per a la guinyada avall.

Tal com ja s'ha exposat a la secció 7.3.2, es pot observar que davant d'un moviment de capcineig (o de guinyada) de la càmera, la projecció d'un objecte proper es desplaçarà sobre la imatge molt més que la projecció d'un objecte llunyà. Una possible solució seria estimar la distància inicial entre la càmera i l'objecte amb un procediment com el descrit al capítol anterior. La solució adoptada en aquest experiment és semblant a l'emprada a la secció 7.2.3: un cop inicialitzat el contorn actiu se centra en la imatge, i llavors es provoca un moviment controlat de la càmera que allunya la projecció del centre de la imatge. Mesurant l'efecte del moviment provocat sobre la projecció, es pot establir prou acuradament un factor que serveix per adaptar els paràmetres del controlador.

Els valors retornats pel sistema d'odometria de la càmera per a cada imatge es poden veure a la figura 7.11. Es pot observar que durant l'experiment els valors de posició dels tres mecanismes de la càmera activa han canviat. Es pot diferenciar clarament la fase d'aproximació del robot, quan el valor del *zoom* disminueix; l'aturada del robot,

¹²A la pràctica, aquestes velocitats màximes haurien d'estar restringides segons quin sigui el temps d'adquisició de la càmera per tal d'obtenir imatges nítides (*motion blur*). En aquest experiment no cal fer aquesta consideració ja que el robot es mou lentament.

7. Estimació del moviment en el context de la visió activa

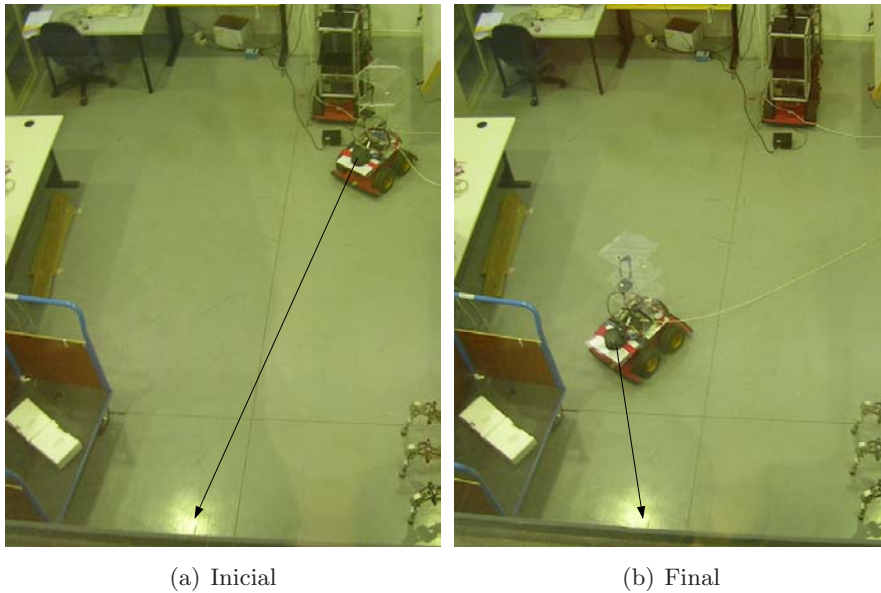


Figura 7.10: Posició inicial i final del robot utilitzat en l'experiment en què es controla el *zoom*, el capcineig i la guinyada.

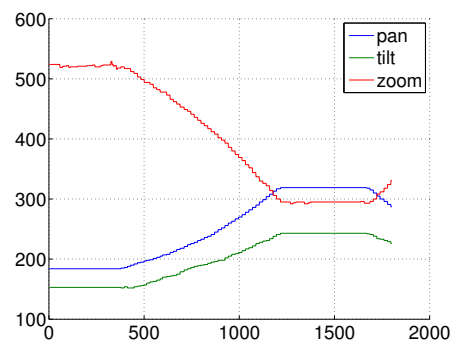


Figura 7.11: Posicions dels mecanismes de *zoom*, capcineig i guinyada.

7.3 Moviment d'orientació de la càmera

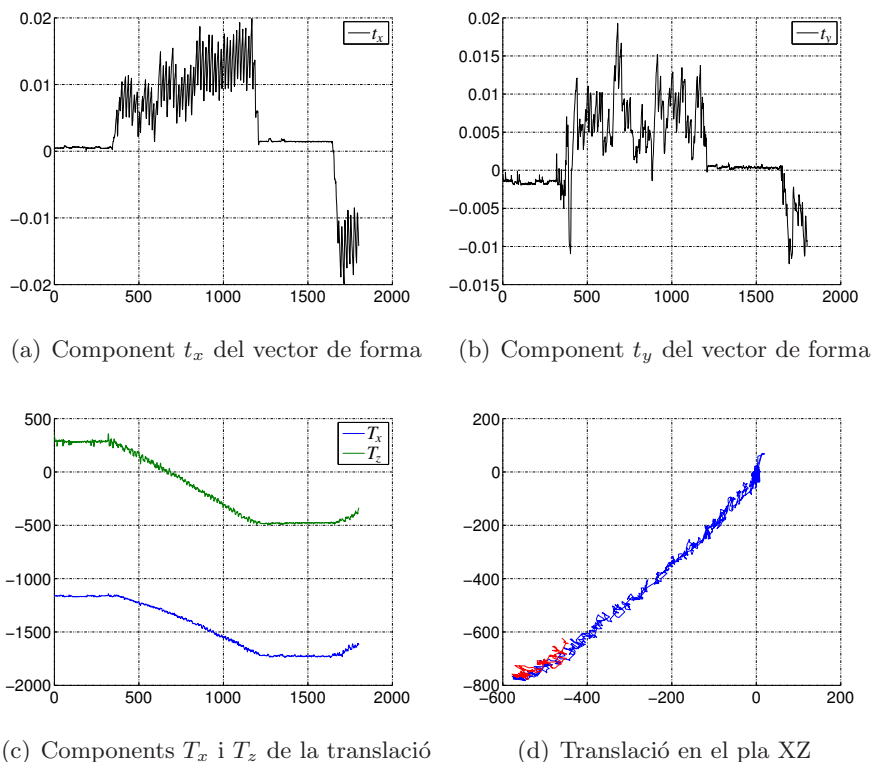


Figura 7.12: Trajectòria computada.

en què els valors no canvien, i un petit retrocés al final. El moviment del robot provoca el moviment dels tres mecanismes. Encara que en posició del controlador, capcineig i guinyada es mouen amb uns valors semblants, el moviment real és molt diferent ja que la resolució en graus per cada controlador és diferent.

Amb els valors retornats pel mecanisme de càmera activa es pot modificar l'algorisme de recuperació del moviment propi per tal que es tingui en compte el canvi d'orientació de la càmera a més del canvi del *zoom*. L'algorisme detallat es pot veure a l'algorisme 7.1. En aquest experiment, de la mateixa manera que en l'experiment anterior, no es coneixen els paràmetres de calibratge de la càmera i, per tant, tampoc no es coneix la relació real entre la distància focal inicial i la nova distància focal que es té cada vegada que canvia el *zoom*. En canvi, el que sí que es coneix és el valor en posició del controlador del *zoom*. Per a aquest experiment es fa servir la raó entre les diferents posicions com a aproximació de la raó de les diferents distàncies focals. Es pot veure fàcilment que la relació entre aquestes dues mesures és lineal si es considera que

7. Estimació del moviment en el context de la visió activa

Entrada : $\mathbf{S}, Z_0, f, \theta_{capcineig}, \phi_{guinyada}$, i opcionalment f', u, u', v, v'

Sortida : posa 3D = $\{T_x, T_y, T_z, \phi, \theta, \psi\}$

2 *si no hi ha paràmetres opcionals* **aleshores**

3 es considera una càmera sense *zoom*, de manera que s'inicialitza

$$u = u' = v = v' = 0 \quad f' = f$$

4 **fi**

6 Recompondre \mathbf{M} a partir de la part vectorial de \mathbf{S} com

$$\mathbf{M} = \begin{bmatrix} S_3 + 1 & S_6 \\ S_5 & S_4 + 1 \end{bmatrix}$$

7 Trobar els vectors propis λ_1 i λ_2 i el valor propi \mathbf{v}_1 a partir de la descomposició en valors singulars de $\mathbf{M}\mathbf{M}^T$

8 La rotació θ_x es troba a partir dels valors propis:

$$\theta_x = \text{acos}\left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right)$$

9 La rotació ϕ_z es calcula a partir de les components del vector propi \mathbf{v}_1 :

$$\phi_z = \text{atan2}(v_{12}, v_{11})$$

10 Per calcular ψ_z primer cal recompondre les matrius de rotació

$$\mathbf{Rpsi} = \frac{1}{\lambda_1} \mathbf{R}_z |_2(-\phi) * \begin{bmatrix} 1 & 0 \\ 0 & 1/\cos(\theta) \end{bmatrix} * \mathbf{M}$$

i ψ_z es calcula amb:

$$\psi_z = \text{atan2}(\text{Rpsi}_{21}, \text{Rpsi}_{11})$$

11 Trobar la matriu de rotació \mathbf{R} a partir de les rotacions $(\phi_z, \theta_x, \psi_z)$ al voltant dels eixos ZXX

12 Descompondre \mathbf{R} per trobar les rotacions (ϕ, θ, ψ) al voltant dels eixos ZYX

13 *si* $\phi > 10$ **aleshores** $\phi = \phi^\circ + \phi_{guinyada}$

sino $\phi = \phi_{guinyada}$

14 *si* $\theta > 10$ **aleshores** $\theta = \theta^\circ + \theta_{capcineig}$

sino $\theta = \theta_{capcineig}$;

15 Les translacions es calculen aplicant:

$$T_x = \left(\frac{S_1 - (u - u')}{f\sqrt{\lambda_1}} - R_{13} \right) Z_0$$

$$T_y = \left(\frac{S_2 - (v - v')}{f\sqrt{\lambda_1}} - R_{23} \right) Z_0$$

$$T_z = \left(\frac{f'}{f\sqrt{\lambda_1}} - R_{33} \right) Z_0$$

Algorisme 7.1 : Algorisme per computar la posa a partir del vector de forma tenint en compte els angles de capcineig i guinyada.

7.3 Moviment d'orientació de la càmera

un determinat increment en posicions de *zoom* correspon sempre a un mateix increment de distància focal.

Tal com s'ha proposat a la secció 7.3.3, la informació de capcineig i guinyada s'incorpora a l'algorisme observant que el canvi d'orientació de la càmera es pot explicar com si es tractés d'un canvi d'orientació de l'objecte. És per això que, un cop computades les orientacions a partir del vector de forma (algorisme 7.1, línies 8 i 10) se'ls afegeix el valor dels angles de capcineig i guinyada, abans de fer-les servir per computar les components de translació del moviment (línia 15). A efectes pràctics i per facilitar la implementació d'aquest experiment, s'ha decidit no incorporar a l'estimació del moviment les rotacions en els eixos X i Y que siguin menors de 10° (línies 13 i 14).

Sense el control de l'orientació de la càmera, el moviment del robot faria que la projecció del contorn de l'objecte sortís ràpidament del camp de visió de la càmera. L'algorisme de control emprat en aquest experiment manté la projecció centrada i el contorn es pot seguir durant tot l'experiment gràcies al moviment d'orientació induït en la càmera. Com a resultat, tal com es pot veure a les figures 7.12(a) i 7.12(b), les components del vector de forma t_x i t_y que modelen les translacions laterals són molt petites durant tota la seqüència (en comparació, per exemple, amb les obtingudes al capítol 2) i es mantenen sempre al voltant de 0.

La translació real s'obté quan es calcula el moviment 3D i s'incorpora la rotació del capcineig i la guinyada. A la figura 7.12(c) es poden veure les components T_x i T_z per separat, i a la figura 7.12(d), la trajectòria que aquestes translacions defineixen sobre el pla de moviment (segons el sistema de coordenades definit, el moviment es produeix en el pla XZ). Es comprova que la trajectòria resultant s'aproxima força a una línia recta, i que les translacions són coherents entre si, ja que T_z té un valor més gran que T_x . També es pot observar clarament la darrera fase de l'experiment, representada en color vermell, en què el robot fa un moviment de retrocés. El moviment recuperat per aquest cas està correctament sobreposat amb el moviment anterior.

Cal recordar que el punt principal en el model estenopeic canvia per a cada nova distància focal de la càmera. Aquest canvi afecta la computació de les translacions T_x i T_y , com es pot veure en les equacions 2.16-2.21. En aquest experiment no s'ha calibrat la càmera, de manera que s'ha considerat que en cada nova imatge el punt principal (u_i, v_i) coincideix amb el punt principal (u_0, v_0) corresponent a la primera

7. Estimació del moviment en el context de la visió activa

imatge.¹³ Aquesta és una de les raons que expliquen per què la trajectòria obtinguda per la translació T_x no és totalment recta.

Finalment, es mostra la incertesa calculada per a cada pas de la seqüència de moviment (figura 7.13). Per a l'experiment realitzat, amb el sistema de coordenades situat amb origen a la càmera, les incerteses corresponents es poden observar a la figura 7.13(a) per a les components de translació, i a la figura 7.13(b) per a les components de rotació. Coherentment amb els resultats del capítol 5, la incertesa de la component T_z disminueix a mesura que el robot s'acosta a l'objectiu, i torna a augmentar quan se'n separa. A diferència del que s'esperava, però, les incerteses en les components de moviment T_x i T_y no són menyspreables, i s'observa que augmenten a mesura que ho fa la rotació. Això es pot explicar pel canvi de posició del sistema de referència. Les translacions es computen fent servir les rotacions, i és lògic que la incertesa en les rotacions faci augmentar la incertesa de les translacions. Observant les incerteses que resultarien si es definís el sistema de coordenades a l'objectiu (figures 7.13(c) i 7.13(d)), es pot veure que els resultats són coherents amb els resultats del capítol 5.

7.4 Conclusions

En aquest capítol s'han proposat diverses estratègies per controlar els mecanismes d'una càmera activa. D'una banda, el control sobre el *zoom* permet mantenir la mida de la projecció de l'objectiu a mesura que el robot s'aproxima o s'allunya, de manera que l'espai de moviments que es poden estimar queda incrementat. De l'altra, el control del mecanisme d'orientació (capcineig i guinyada) permet utilitzar objectius que no estiguin davant del robot, ja que la càmera es va orientant per seguir l'objectiu i va compensant el moviment del robot. S'ha fet un esforç per abordar el problema tenint en compte els canvis d'orientació i, a la vegada, els canvis de *zoom*.

Les funcions d'error que serveixen per dissenyar els controladors utilitzen resultats parcials de l'algorisme de recuperació del moviment propi, i per tant no s'afegeix cap sobrecost. Això és important perquè cal que l'algorisme resultant de recuperació del moviment propi amb càmera activa funcioni en temps real.

¹³De fet, en no calibrar la càmera, es considera implícitament que el punt principal coincideix amb el centre de la imatge.

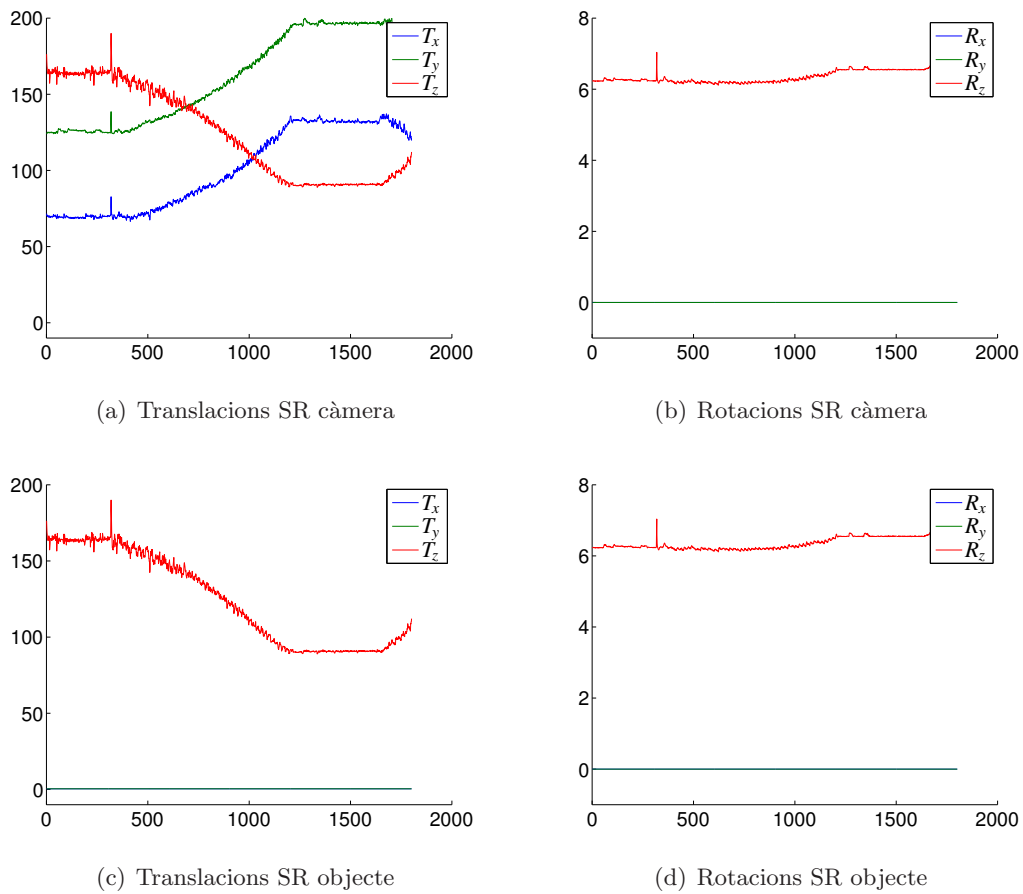


Figura 7.13: Incertesa de cadascuna de les components de la posa. (a) Components de translació i (b) de rotació quan el sistema de referència està definit a la càmera. (c) Components de translació i (d) de rotació quan el sistema de referència està definit a l'objecte. La incertesa de T_z disminueix perquè el moviment és d'aproximació a l'objecte, i les incerteses T_x i T_y augmenten quan el sistema de referència està a la càmera perquè en el seu càlcul hi influeix la incertesa calculada per a les rotacions.

7. Estimació del moviment en el context de la visió activa

Els experiments realitzats han servit per demostrar que els algorismes de control proposat són vàlids, ja que permeten mantenir la projecció de l'objectiu amb una mida constant i centrada en la imatge. S'ha demostrat que el control es pot portar a terme sense coneixement dels paràmetres de calibratge de la càmera i sense coneixement de l'entorn. Tot i aquesta falta d'informació, en els experiments s'ha estimat, de manera qualitativa, el moviment realitzat. En el cas del *zoom*, s'ha considerat que la raó entre distàncies focals es pot aproximar per la raó entre les posicions del controlador del *zoom* i, més important, que la distància entre la càmera i l'objectiu és un paràmetre fix. En el cas de la guinyada i el capcineig, s'ha assumit que el punt principal de la càmera no canvia quan es fa *zoom*, i també s'ha considerat que els eixos de rotació del mecanisme d'orientació es tallen sobre el pla de la imatge.

S'ha vist que el fet d'utilitzar una càmera activa no canvia els algorismes proposats per al càlcul de la incertesa (capítol 5), excepte que el moviment ara cal calcular-lo amb el nou algorisme proposat en aquest capítol. En el darrer experiment es pot veure que la incertesa en les components T_x i T_y quan el sistema de coordenades té l'origen en la càmera està influïda per la incertesa en les components de rotació. Aquest és un comportament lògic si es té en compte que per expressar el moviment en sistema de coordenades de càmera cal utilitzar les rotacions calculades.

Es pot observar que el fet de conèixer el signe de la rotació, principalment del mecanisme de guinyada, no permet eliminar l'ambigüïtat *Necker reversal* quan el moviment del robot és general. Això es deu al fet que la rotació realitzada amb la càmera i la realitzada amb el robot (i mesurada amb el vector de forma) són en general complementàries, i per tant, sense informació addicional, no es pot conèixer el signe d'aquesta darrera.

No s'ha abordat el tema de determinar quina és la mida òptima de la projecció de l'objecte en la imatge per fer-ne el millor seguiment. En general, els algorismes proposats són independents del mètode que es tria per obtenir la parametrització del moviment en l'espai de forma i el fet de determinar la mida més adequada és un paràmetre dependent de l'algorisme concret que s'utilitza per fer el seguiment.

Seria interessant fer un estudi més precís de com influeixen els errors de suposar el model estenopec quan es té una càmera amb *zoom*. D'una banda, l'algorisme de recuperació del moviment propi considera que la distància entre la càmera i l'objecte

en la imatge inicial és una constant. S'ha vist al capítol 6.2 que això és només una aproximació. En realitat s'introdueix un lleuger escalat addicional a la computació de les translacions, i sobretot a T_z . Caldria fer una avaluació experimental, semblant a la proposada a [135], per verificar que els efectes són realment negligibles.

També caldria avaluar, amb un senzill experiment, com varia el punt principal a mesura que es canvia la distància focal de la càmera amb el *zoom*. Ja que s'introdueix un biaix que depèn de la distància focal, caldria fer un experiment per avaluar l'error que es comet suposant que no canvia. De tota manera, aquest moviment és el més fàcil de calibrar.

Capítol 8

Conclusions i treball futur

Lo pastor se meravellà per què Fèlix lo apellave «sènyer» en aquell temps, e en lo començament no-l hac apellat «sènyer», com lo pastor fos aquell mateix en lo començament, quan se atrobà ab Fèlix, que era en la fi de lur pertiment. Dementre que el pastor en axí se meravellave, ell conech que honor mils se cové ab sàvies peraules que ab vils vestiments.

El pastor es va sorprendre de ser anomenat «senyor» per Fèlix precisament llavors i no pas des del començament; ell era la mateixa persona, tant en el moment de trobar-se Fèlix, com ara. Mentre el pastor així se sorprenia, s'adonava que l'honor s'adiu més amb sàvies paraules que amb vils vestiments.

Ramon Llull, *Llibre del cel* (del *Llibre de meravelles*). Edició de Ll. Baulenas.

En aquesta tesi s'ha partit d'un algorisme que parametriza el moviment d'una càmera en funció de deformacions en la imatge captada, i se n'ha demostrat la utilitat per estimar el moviment d'un robot. D'una banda, s'ha fet un estudi de l'error i s'ha proposat un algorisme per al càlcul en temps real de la incertesa de la posa. De l'altra, s'ha introduït el marc necessari per incorporar visió activa i millorar així l'abast d'utilització de l'algorisme. S'ha volgut explorar la possibilitat de fer servir contorns actius, que ofereixen alguns avantatges davant de les característiques típiques més àmpliament utilitzades. No obstant això, també s'ha mostrat que és possible utilitzar la parametrizació proposada en el cas de correspondències de punts.

Una de les principals assumpcions és la de treballar amb un model simplificat de càmera: el model de perspectiva dèbil. *A priori* això limita el tipus d'escenes on es pot aplicar l'algorisme a objectes planars¹ i a condicions de visualització que no provoquin

¹En general, a objectes amb poc relleu en comparació amb la distància a la càmera.

8. Conclusions i treball futur

efectes perspectius en la projecció de l'objecte. A la pràctica, això no ha comportat una limitació, ja que ha estat possible fer tots els experiments, que han inclòs distàncies que han abastat més de set metres (capítols 2.3.1 i 2.3.2) fins a un metre (capítols 6 i 7). Fins i tot, al capítol 3, ha calgut forçar les condicions de visualització utilitzant una distància de 0.5 metres per poder observar errors en les mesures.²

El moviment general requereix emprar un espai de forma de sis graus de llibertat (gdll) per codificar correctament tots els possibles moviments. No obstant això, en el cas que l'espai de treball del robot o els moviments que és capaç de fer estiguin restringits es poden emprar diferents codificacions de l'espai de forma. Al llarg de tots els capítols s'ha mostrat la utilitat d'emprar aquestes diferents codificacions: al capítol 2 s'ha emprat un espai de tres gdll per a un robot planar, suposant que la càmera està alineada amb el pla de moviment, i un espai de quatre gdll quan la càmera no està alineada amb el pla de moviment; al capítol 3 s'ha emprat un espai de cinc gdll quan es coneix que l'objecte està frontoparal·lel al pla de la càmera, i al capítol 6 s'ha utilitzat un espai de tres gdll per modelar una càmera que només fa *zoom*.

Seguidament, es resumeixen les principals aportacions i s'indiquen algunes línies d'investigació futura.

8.1 Contribucions

Les aportacions més importants fetes en aquest treball de tesi es poden detallar de la manera següent:

- Per primera vegada s'ha **avaluat l'algorisme** d'estimació del moviment propi comparant-lo amb un mètode de posicionament **làser** (capítol 2.3.1). Els resultats han estat molt prometedors i han mostrat el potencial de l'algorisme, sobretot en el cas de les translacions.
- S'ha mostrat com es poden obtenir les expressions de la transformació inversa de l'algorisme proposat, és a dir, com es pot expressar la deformació en la imatge en funció del moviment de la càmera. D'una banda, aquestes expressions han

²Naturalment, aquestes distàncies cal considerar-les sempre en conjunció amb el sistema de lents emprat en cada experiment, com s'ha observat al capítol 6.3.

permès fusionar la informació del seguiment de contorns amb la informació d'un **sensor inercial**, i s'ha vist com es pot incorporar informació d'altres sensors dins l'algorisme de seguiment (capítol 2.3.2). De l'altra, aquestes expressions serviran per poder formular en un futur l'algorisme de seguiment en termes de l'*Unscented Kalman Filter*.

- S'han estudiat en profunditat les propietats de la representació en espai de forma i s'ha proposat un nou algorisme per al càlcul de la **direcció epipolar afí** (capítol 3). Aquest algorisme suposa condicions de visualització de perspectiva dèbil i assumeix que el moviment de la càmera està lliure de ciclorrotació. Amb la sèrie d'experiments duts a terme, s'ha vist que la direcció epipolar, més enllà del seu interès natural de relacionar dues vistes, és una bona mesura per avaluar la qualitat de l'afinitat computada amb el contorn actiu.
- Un dels objectius principals ha estat caracteritzar l'error que es comet. Primer s'ha fet un estudi des del punt de vista analític, i després des del punt de vista estadístic. Amb l'enfocament analític s'han obtingut resultats qualitativament bons, però valors d'incertesa clarament subestimats (capítol 4). Llavors s'ha presentat un estudi estadístic per calcular la **incertesa en l'espai de posa** (capítol 5). No s'ha continuat el camí analític perquè, tal com s'ha vist, l'enfocament estadístic obre vies interessants de treball futur. S'ha vist estadísticament que les translacions T_x i T_y i la rotació R_z es recuperen amb força precisió, mentre que l'error que es comet en la translació T_z és més gran. Pel que fa a les rotacions R_x i R_y , quan són petites no és possible recuperar-ne correctament els valors i conseqüentment la incertesa calculada és molt gran. A mesura que aquestes rotacions augmenten la incertesa disminueix i el valor recuperat es va fent més precís. Experimentalment s'ha determinat que a partir dels 10° ja es poden considerar aquestes rotacions. També s'ha determinat que les components R_x i R_y estan correlacionades amb la translació T_z , i s'ha proposat una explicació per a aquest fet.
- Amb l'objectiu de mostrar l'aplicabilitat de l'algorisme en situacions reals, s'ha proposat un nou algorisme per calcular, en **temps real**, la incertesa associada a cada posa. Els experiments han mostrat la seva vàlua, tant per a grans translacions com per a grans rotacions (capítol 5.5.2).

8. Conclusions i treball futur

- S'ha proposat un nou algorisme, basat en un calibratge geomètric, per recuperar la **distància inicial** entre l'objecte fixat i la càmera (capítol 6). Al nostre parer, aquesta és una aportació molt interessant, ja que elimina un dels dos impediments que fan que la translació que es pot recuperar amb l'algorisme estigui afectada per un factor d'escala. Per tal d'eliminar el darrer impediment, i poder obtenir així informació mètrica, cal saber també els paràmetres de calibratge de la càmera: la distància focal i la posició del punt principal quan varia el *zoom*. Això es podria aconseguir aplicant alguna de les tècniques existents de calibratge o autocalibratge.
- L'aplicació de l'algorisme en situacions reals està limitada, bàsicament, pel fet que cal mantenir la projecció dins del pla de la imatge. S'ha proposat l'ús d'una **càmera activa** (capítol 7), que ha d'incorporar tant un *zoom* com un mecanisme de **capcineig i guinyada**. S'ha proposat un nou algorisme per controlar aquests tres graus de llibertat addicionals. D'una banda, el control del *zoom* permet mantenir constant la mida de la projecció, de manera que els moviments d'aproximació i allunyament que es poden fer sense perdre de vista l'objectiu són molt més grans. De l'altra, el control de l'orientació de la càmera permet mantenir centrada la projecció en la imatge, cosa que permet una llibertat de moviments molt més gran del robot. Aquest algorisme aprofita resultats parcials de l'algorisme de recuperació del moviment propi, de manera que no s'afegeix càrrega de còmput addicional. També s'ha mostrat com es pot incorporar el moviment d'orientació de la càmera al càlcul del moviment del robot.

8.2 Treball futur

Més enllà del treball dut a terme queden un reguitzell de problemes interessants que caldria considerar:

- En tots els experiments realitzats la inicialització del contorn ha vingut donada manualment per un operador. Els treballs previs sobre inicialització automàtica de contorns actius són prometedors, però en escenes generals produeixen un nombre massa gran de contorns [26]. Caldria trobar un algorisme que permetés ava-

luar cadascun d'aquests contorns i retornés la inicialització del contorn més ben condicionat per fer-ne el seguiment.

- El treball amb el sensor inercial ha mostrat que visió i inercial són dos sensors que es complementen molt naturalment. Amb l'objectiu de trobar millors estimacions del moviment, el sensor inercial hauria d'ajudar, principalment, a discernir el signe de les rotacions R_x i R_y i a determinar-ne el valor, sobretot en rotacions petites. Una possibilitat és incorporar un segon pas de mesura en el filtre de Kalman amb la informació del sensor inercial expressada en espai de forma.
- En el context de la recuperació de la direcció epipolar afí, en un futur seria interessant investigar en quines circumstàncies es pot recuperar la informació addicional sobre el moviment de la càmera i l'estructura de l'escena a partir de dues (o més) vistes no calibrades d'un objecte planar sota perspectiva dèbil. S'ha esmentat al capítol 3 que si es coneixen la raó d'aspecte i el biaix (per exemple, els paràmetres intrínsecs determinant l'estructura mètrica en la imatge), llavors la direcció de l'eix de rotació també es pot recuperar. En la mateixa línia, es pot tractar de recuperar l'orientació del pla de l'escena quan es tracten situacions degenerades, com per exemple quan aquesta orientació coincideix amb el pla de la imatge, o quan els dos plans tenen una direcció comuna.
- La motivació més important d'aquest treball és trobar la millor estimació de la posa del robot. Actualment, el filtre de Kalman s'encarrega de trobar la millor parametrització del moviment en espai de forma, i a partir d'aquesta parametrització s'obtenen els paràmetres de la posa. No es pot expressar el vector d'estats del filtre de Kalman en espai de posa perquè la transformació d'espai de forma a espai de posa no és derivable, i per tant no se'n pot trobar el jacobià. Al capítol 5 s'ha vist com es pot propagar la incertesa a través d'aquesta transformació utilitzant la *Unscented Transformation*. En un futur seria interessant estendre aquesta idea i utilitzar un *Unscented Kalman Filter* per poder expressar el vector d'estats del filtre directament en l'espai de posa. Per fer-ho caldrà utilitzar les expressions de la transformació inversa, que ja han estat derivades al capítol 2.3.2.
- Amb el model estenopec teòricament no es poden fer servir algorismes que suposin que el centre òptic és fix. Per al control de la càmera activa aquesta apreciació

8. Conclusions i treball futur

no té importància, ja que s'utilitza un algorisme en llaç tancat. Pel que fa a l'algorisme de deducció del moviment, canviant el *zoom* ímplicitament es canvia l'escala. Com que la deducció que s'ha fet fins ara no considera una càmera calibrada, la translació de fet ja està afectada per un factor d'escala. Caldria experimentar amb una càmera calibrada els efectes de l'aproximació feta, i eventualment estudiar les fórmules de deducció del moviment amb un model de càmera diferent, que sigui més apropiat.

- Finalment, a l'inici de la tesi va semblar interessant explorar la via de fer el seguiment de múltiples contorns. Malauradament, els treballs que s'ha dut a terme no han estat fructífers. D'una banda, s'ha vist que mantenir un objecte dins del camp visual amb una càmera passiva implica uns moviments possibles molt restringits, gairebé només aproximacions i allunyaments. Si cal tenir més d'un contorn al llarg de tota la seqüència, els moviments són massa restringits. De l'altra, per tal de poder aprofitar la informació de dos contorns cal conèixer la relació entre la profunditat de cadascun respecte a la càmera. Si continuem suposant aquesta informació desconeguda no es pot obtenir cap benefici del seguiment de dos contorns.

En el seu lloc, seguint el treball de Cham i Cipolla [27], sembla interessant en el futur explorar la via de l'estereovisió. D'una banda, la redundància de fer el seguiment en cadascuna de les càmeres i fusionar el resultat ha de conduir a obtenir una estimació més robusta de l'afinitat, i en conseqüència, de l'estimació del moviment del robot. De l'altra, un sistema estèreo proporciona informació de profunditat i permet avaluar més fàcilment la planaritat dels objectes. Fins i tot es podria relaxar aquesta condició utilitzant les equacions de Blake [17] per a objectes no planars.

Chapter 8

Conclusions and future work

Lo pastor se meravellà per què Fèlix lo apellave «sènyer» en aquell temps, e en lo començament no-l hac apellat «sènyer», com lo pastor fos aquell mateix en lo començament, quan se atrobà ab Fèlix, que era en la fi de lur pertiment. Dementre que el pastor en axí se meravellave, ell conech que honor mils se cové ab sàvies peraules que ab vils vestiments.

The shepherd was surprised of being called «Sir» by Fèlix precisely then and not from the beginning; he was the same person, in the moment of finding Fèlix, as well as now. While the shepherd surprised himself like this, he noticed that the honor matches with wise words more than with vile dressings.

Ramon Llull. *Llibre del cel* (del *Llibre de meravelles*). Edició de Ll. Baulenas.

This thesis focuses on an algorithm that parameterizes the motion of a camera by the induced deformations in the images, and its utility for estimating the motion of a robot. On the one hand, a study of the error has been carried out and an algorithm has been proposed for the calculation of the pose uncertainty in real time. On the other hand, the necessary framework has been introduced to incorporate active vision, improving the utility of the algorithm. The possibility of using active contours has been explored, which offers some advantages in front of the most used point-based features. However, it has also been shown that it is possible to use the parametrization proposed in the case of point correspondences.

One of the main assumptions is to work with the weak-perspective camera model. A priori this limits the type of scenes where the algorithm can be applied, namely planar objects¹ and viewing conditions where perspective effects in the projection of

¹In general, objects with little relief in comparison with the distance to the camera.

8. Conclusions and future work

the object do not appear. In practice this has not supposed a limitation, since it has been possible to carry out all experiments, with distances that go from more than 7 meters (Chapters 2.3.1 and 2.3.2) until 1 meter (Chapters 6 and 7). Even, in Chapter 3, it has been necessary to force the viewing conditions to a distance of 0.5 meters in order to observe errors in the measurements².

In order to correctly represent all the possible motions a shape space of 6 degrees of freedom (dof) is required. However, in the case that the working space of the robot or the possible motions are restricted, it is possible to use different shape space codifications. Along all chapters the utility of using these different codings has been shown: in Chapter 2 we used a 3-dof space for a planar robot assuming that the camera optical axis is parallel to the motion plane, and a 4-dof space when the camera is not parallel; in Chapter 3 we used a 5-dof space when it is known that the object is frontoparallel to the imaging plane of the camera; and in Chapter 6 we used a 3-dof space to model a camera which only zooms.

Next the main contributions are summarized, and some lines of future research are indicated.

8.1 Contributions

The most important contributions of this thesis can be detailed as follows:

- For the first time the egomotion estimation algorithm has been **evaluated** and has been compared with a **laser** positioning method (Chapter 2.3.1). The results have been very promising and have shown the potentiality of the algorithm, especially in the case of translations.
- It has been shown how to calculate the expressions of the inverse transformation of the proposed algorithm, that is, expressing the deformation in the image depending on the camera motion. These expressions have allowed to merge the information of the contour tracking with the information of an **inertial sensor**,

²Naturally, it is necessary to consider these distances in conjunction with the system of lenses employed in each experiment, as observed in Chapter 6.3.

thus showing how information of other sensors can be incorporated into the tracking algorithm (Chapter 2.3.2). In the future these expressions will allow us to formulate the tracking algorithm in terms of the *Unscented Kalman Filter*.

- The properties of the representation in shape space have been studied in depth and a new algorithm has been proposed for the calculation of the **affine epipolar direction** (Chapter 3). This algorithm supposes conditions of weak perspective visualization and assumes that the motion of the camera is free of cyclorotation. With the series of experiments carried out we concluded that the epipolar direction, beyond its natural interest of relating two views, is a good measure to evaluate the quality of the affinity computed from the active contour.
- One of the main goals has been to characterize the error obtained. A study has been carried out, first from the analytical point of view and afterwards from the statistical point of view. With the analytical approach some qualitatively good results have been obtained, but uncertainty values are clearly subestimated (Chapter 4). Then, a statistical approach has been presented to compute the **uncertainty in pose space** (Chapter 5). We have not gone further with the analytic approach because, as it has been seen, the statistical approach opens several ways of interesting future work. It has been demonstrated statistically that translations T_x , T_y and the rotation R_z are recovered with great accuracy, while the error committed in the translation T_z is bigger. Regarding the rotations R_x and R_y , it is not possible to retrieve their values correctly when they are small and, consequently, the calculated uncertainty is very large. If these rotations increase, then the uncertainty diminishes and the retrieved values are more precise. It has been determined experimentally that these rotations can be considered starting from 10° . It has also been determined that the components R_x and R_y are correlated with the translation T_z , and an explanation for this fact has been proposed.
- With the goal of showing the applicability of the algorithm in real situations, a new algorithm has been proposed to calculate, in **real-time**, the uncertainty associated to each pose. The experiments have shown its good performance, for large translations as well as for large rotations (Chapter 5.5.2).

8. Conclusions and future work

- A new algorithm, based on geometric calibration, has been proposed to retrieve the **initial distance** between the fixed object and the camera (Chapter 6). In our opinion, this is a very interesting contribution since it serves to calculate one of the parameters needed to recover metric translations. In order to eliminate the remaining parameters it is necessary to know the camera calibration parameters: the focal length and the principal point location in relation to the zoom position. This could be achieved with one of the existing techniques of calibration or self-calibration.
- The application of the algorithm in real situations is limited, principally, by the need to keep the object projection into the image frame. The use of an **active camera** has been proposed (Chapter 7), including **zoom** and a **pan-and-tilt** mechanism. A new algorithm has been developed for controlling these three additional degrees of freedom. On the one hand, the control of the zoom allows to keep the size of the projection constant, so that the movements of approach and receding can become larger without losing the target. On the other hand, the control of the orientation of the camera allows to keep the projection centered in the image, which allows a much greater freedom in the motions of the robot. This new algorithm uses partial results from the egomotion algorithm, so that no additional computing cost is added. It has also been shown how the movement of orientation of the camera can be incorporated in the calculation of the robot egomotion.

8.2 Future work

Beyond the performed work, there remain some interesting problems to consider:

- In all the experiments carried out, the initialization of the contour has been given manually by an operator. Previous works about automatic initialization of active contours are promising, but for general scenes they produce a too large number of contours [26]. It would be interesting to find an algorithm that evaluates each of these contours and returns the initialization of the best fitted contour.

- The work using the inertial sensor together with vision has shown that the two sensors complement one another very naturally. With the goal of finding better motion estimations, the inertial sensor should help, mainly, to distinguish the sign of the R_x and R_y rotations, and to determine their values, especially for small rotations. A possibility is to incorporate a second step of measure into the Kalman filter with the information of the inertial sensor expressed in shape space.
- In the context of the recovery of the affine epipolar direction, in the future it would be interesting to investigate under which circumstances additional information about camera motion and scene structure can be recovered from two (or more) uncalibrated views of a planar object under weak-perspective viewing conditions. It has been mentioned in Chapter 3 that, if the aspect ratio and skew (i.e., the intrinsic parameters determining the metric structure in the image) are known, then the direction of the axis of rotation can also be recovered. Along the same line, we will tackle the recovery of the orientation of the scene plane, as well as what occurs in degenerate situations in which such orientation is the same as that of the image plane, or when both planes have a common direction.
- The main motivation of this work is to find the best estimation of the egomotion of a robot. At present, the Kalman filter is in charge of determining the best motion parametrization in shape space, and from this parametrization the pose parameters are obtained. The state vector of the Kalman filter cannot be expressed in pose space because the function from shape space to pose space is not derivable, and therefore its Jacobian can not be found. In Chapter 5 it has been shown how the uncertainty can be propagated through this function using the *Unscented Transformation*. In the future, it would be interesting to extend this idea and to use an *Unscented Kalman Filter* to be able to express the state vector of the filter directly in the space of pose. To do so, it will be necessary to use the expressions of the inverse function, which have already been derived in Chapter 2.3.2.
- Theoretically, the pinhole model can not be used in algorithms that assume that the optical center is fixed. For the control of the active camera this does not matter, since the proposed algorithm uses a closed loop. Regarding the egomotion

8. Conclusions and future work

algorithm, changes in the zoom are implicitly changing also the scale. Since the algorithm used until now does not consider a calibrated camera, the translation is affected by a scale factor. It would be necessary to explore the effects of this approach by using a calibrated camera, and eventually to study the egomotion equations with a different and more appropriate camera model.

- Finally, at the beginning of the thesis it seemed interesting to explore the tracking of multiple contours. Unfortunately, the works that have been carried out have not been fruitful. We have seen that maintaining the projection of an object in the field of view with a passive camera implies very restricted robot motions. For the same reason, if it is necessary to view more than one contour along all the sequence the motions are too much restricted. Furthermore, in order to be able to use the information from two contours, it is necessary to know the ratio between the depths of each one with respect to the camera. If we continue supposing that this information is unknown, no benefit can be obtained of the tracking of two contours.

Instead, following the work of Cham and Cipolla [27], it seems more interesting to explore in the future the path of stereovision. The redundancy of tracking with two cameras and merging the results has to yield a more robust estimation of the shape vector, and consequently, of the robot motion. Also, a stereo system provides information of depth and allows to evaluate the planarity of the objects more easily. One could even consider the possibility of relaxing the planar object restriction using the equations of Blake [17] for non planar objects.

Apèndix A

Rotacions expressades en angles d'Euler

Una rotació general es pot expressar com una seqüència de rotacions parcials al voltant d'eixos del sistema de coordenades triat. Es coneixen com a angles d'Euler les seqüències de rotacions al voltant dels eixos ZXZ i ZYZ. Les seqüències involucrant els tres eixos coordenats, com per exemple XYZ o YXZ, es coneixen formalment com a angles de Cardan, o bé de Tyan-Brait, tot i que, atesa la confusió històrica en la convenció que cal emprar, moltes vegades s'anomenen també angles d'Euler. Aquestes rotacions poden estar referides al sistema de referència actual o a l'inicial. En el primer cas, la composició de les rotacions successives s'obté postmultiplicant les matrius de rotació simples en l'ordre donat. En el segon, s'obté premultiplicant-les. Suposem un sistema de coordenades XYZ i una rotació XYZ $R = R_x(\phi)R_y(\theta)R_z(\psi)$, és equivalent:

- Eixos de rotació fixos
 - Rotació del sistema XYZ segons l'eix X amb un valor de ϕ
 - Rotació segons l'eix Y original amb un valor de θ
 - Rotació segons l'eix Z original amb un valor de ψ
- Eixos de rotació mòbils
 - Rotació del sistema XYZ segons l'eix Z amb un valor de ψ
 - Rotació segons l'eix Y actual amb un valor de θ
 - Rotació segons l'eix X actual amb un valor de ϕ

Apèndix B

Propagació del soroll als valors i vectors propis

Si es té una matriu amb una pertorbació petita i es vol conèixer la pertorbació corresponent en els seus vectors i valors propis, es té el teorema següent [144, 145]:

Teorema B.1. *Sigui $A = [a_{ij}]$ una matriu $n \times n$ simètrica, i H una matriu ortonormal tal que*

$$H^{-1}AH = \text{diag} \{ \lambda_1, \lambda_2, \dots, \lambda_n \}$$

on $\text{diag} \{ \lambda_1, \lambda_2, \dots, \lambda_n \}$ denota la matriu diagonal amb els elements corresponents a la diagonal. Els seus valors propis es poden ordenar de manera decreixent. Sense pèrdua de generalitat, es considera el valor propi λ_1 . Assumint que λ_1 és un valor propi simple, llavors

$$\lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_n.$$

Es denota

$$H = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_n].$$

Sigui \mathbf{X} un valor propi de A associat amb λ_1 . Llavors \mathbf{X} és un vector a $\text{span} \{ \mathbf{h}_1 \}$, l'espai lineal cobert per \mathbf{h}_1 . Sigui $\mathbf{X}(\varepsilon)$ el vector propi de la matriu pertorbada $A(\varepsilon) = A + \Delta_A$, associat amb el valor propi pertorbat $\lambda_1(\varepsilon)$. $\mathbf{X}(\varepsilon)$ es pot escriure com

$$\mathbf{X}(\varepsilon) = \mathbf{X} + \boldsymbol{\delta}_x$$

amb $\boldsymbol{\delta}_x \in \text{span} \{ \mathbf{h}_2, \mathbf{h}_3, \dots, \mathbf{h}_n \}$. Si ε és el màxim valor absolut dels elements a $\Delta_A = [\delta_{a_{ij}}]$, es té que

$$\Delta_A = \varepsilon B$$

on $B = [b_{ij}]$, i $b_{ij} = \delta_{ij}/\varepsilon$. Per això $|b_{ij}| \leq 1, 1 \leq i \leq n, 1 \leq j \leq n$. Llavors, per qualsevol valor prou petit ε , la pertorbació de λ_1 es pot expressar amb una sèrie convergent de vectors a ε :

$$\delta_{\lambda_1} \triangleq \lambda_1(\varepsilon) - \lambda_1 = p_1\varepsilon + p_2\varepsilon^2 + p_3\varepsilon^3 \dots \quad (\text{B.1})$$

B. Propagació del soroll als valors i vectors propis

i el vector de pertorbació δ_x es pot expressar amb una sèrie convergent de vectors en l'espai $\text{span}\{\mathbf{h}_2, \mathbf{h}_3, \dots, \mathbf{h}_n\}$. En altres paraules, sigui $H_2 = [\mathbf{h}_2 \mathbf{h}_3 \dots \mathbf{h}_n]$, per a qualsevol ε , suficientment petit i positiu hi ha vectors $(n-1)$ -dimensionals $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3, \dots$ tals que

$$\delta_x = \varepsilon H_2 \mathbf{g}_1 + \varepsilon^2 H_2 \mathbf{g}_2 + \varepsilon^3 H_2 \mathbf{g}_3 + \dots \quad (\text{B.2})$$

El terme lineal (en ε) a (B.1) és

$$p_1 \varepsilon = \mathbf{h}_1^T \Delta_A \mathbf{h}_1.$$

El terme lineal (en ε) a (B.2) és

$$\varepsilon H_2 \mathbf{g}_1 = H \Delta_1 H^T \Delta_A \mathbf{X}$$

on

$$\Delta_1 = \text{diag}\{0, (\lambda_1 - \lambda_2)^{-1}, \dots, (\lambda_1 - \lambda_n)^{-1}\} \quad (\text{B.3})$$

És a dir, suprimint el segon i els termes posteriors (considerant pertorbació de primer ordre), per al valor propi

$$\delta_{\lambda_1} \cong \mathbf{h}_1^T \Delta_A \mathbf{h}_1,$$

i per al vector propi

$$\delta_x \cong H \Delta_1 H^T \Delta_A \mathbf{X}.$$

Per trobar la pertorbació de primer ordre d'un altre vector propi, per exemple \mathbf{X}_2 associat al valor propi simple λ_2 , només cal modificar la matriu Δ_2 :

$$\Delta_2 = \text{diag}\{(\lambda_2 - \lambda_1)^{-1}, 0, \dots, (\lambda_2 - \lambda_n)^{-1}\}$$

Apèndix C

Resultats del càlcul de la covariància en el moviment

Com ja s'ha esmentat anteriorment a la secció 5.4.2, un dels objectius proposats és caracteritzar les covariàncies corresponents a l'estimació del moviment propi calculat amb l'algorisme proposat i, per tant, també identificar les possibles correlacions entre els components del moviment. Com en els experiments mostrats en aquella secció, s'afegeix soroll gaussià amb mitjana zero i desviació $\sigma = 0.5$ a la projecció de l'objecte. Per tal de visualitzar les matrius de covariàncies 6×6 obtingudes, es representa les components dos a dos i es dibuixa el valor mitjà i l'el·lipse de l'error al 50%. Fixeu-vos que per representar totes les submatrius 2×2 possibles calen quinze combinacions 2D. També es comparen els resultats de Monte Carlo i UT. La UT parteix de la covariància obtinguda en espai de forma per la simulació de Monte Carlo per seleccionar els punts sigma amb l'esquema simètric, descrit a la secció 5.5.1.

C.1 Càmera en la posició inicial

El primer experiment d'aquesta secció es fa al voltant de la posició inicial, sense cap moviment de la càmera. La figura C.1(a) mostra que no hi ha correlació entre els components de translació. Coherentment amb els resultats de la secció 5.4.1 (figura 5.4), es pot observar que quan la projecció de l'objecte està centrada en la imatge les components T_x i T_y es recuperen de forma precisa, amb $\bar{x} \approx 0$ i $\bar{y} \approx 0$, i un error petit. Com era d'esperar, el càlcul de la translació en profunditat T_z és menys precís. L'estimació

C. Resultats del càlcul de la covariància en el moviment

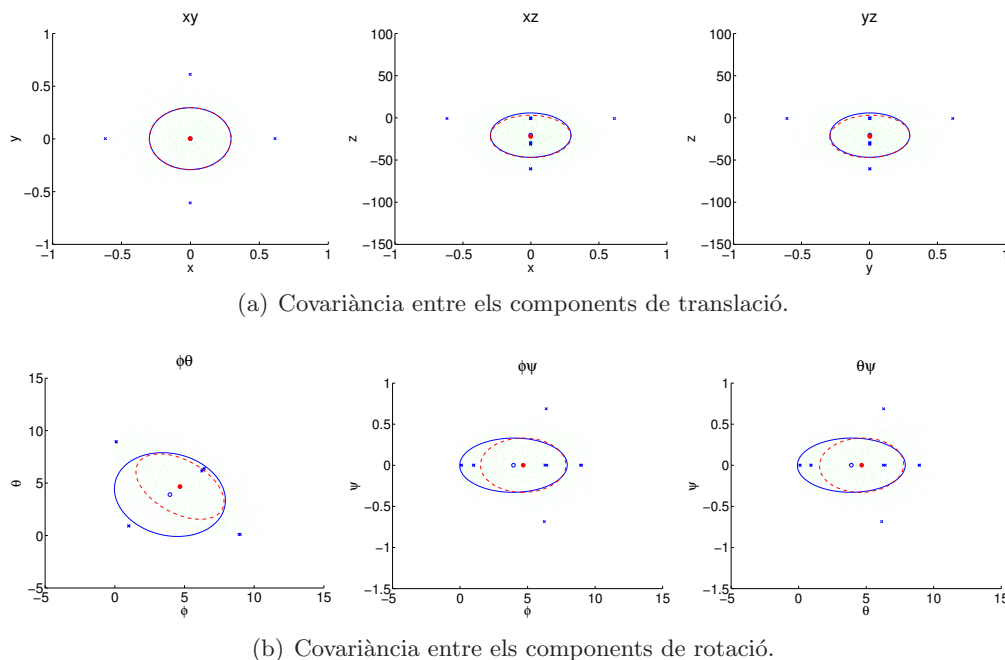


Figura C.1: Representació de les submatrius de covariància 2×2 per a pertorbacions al voltant de la posició inicial de la càmera: (a) correspon a covariàncies entre components de translació, i (b) a covariàncies entre components de rotació. Es mostren els resultats de Monte Carlo (punts verds per a les projeccions de cada mostra, i un cercle sòlid i una el·lipse discontinua per a la mitjana i la covariància) i els resultats de la UT (creus per a punts sigma i un cercle buit i una el·lipse sòlida per a la mitjana i la covariància).

dels estadístics amb la UT és molt semblant a l'obtinguda amb la simulació de Monte Carlo.

La figura C.1(b) mostra les correlacions entre les components de rotació. Com ja s'ha esmentat anteriorment, a la secció 5.3, a causa de la representació de la rotació utilitzada no es poden recuperar els valors correctes de les rotacions R_x i R_y quan aquests valors són gairebé zero i la rotació es calcula com una mitjana. Atès que no es permeten valors negatius per a aquestes rotacions, en les representacions de les covariàncies apareix una línia recta en el 0 que actua a mode de frontera. La rotació R_z es recupera correctament. Les estimacions de la covariància obtingudes amb la UT no es corresponen exactament amb les obtingudes amb l'algorisme de Monte Carlo a causa del biaix introduït en les rotacions, i s'ha de verificar més endavant si l'aproximació de la UT és vàlida per a aquests casos.

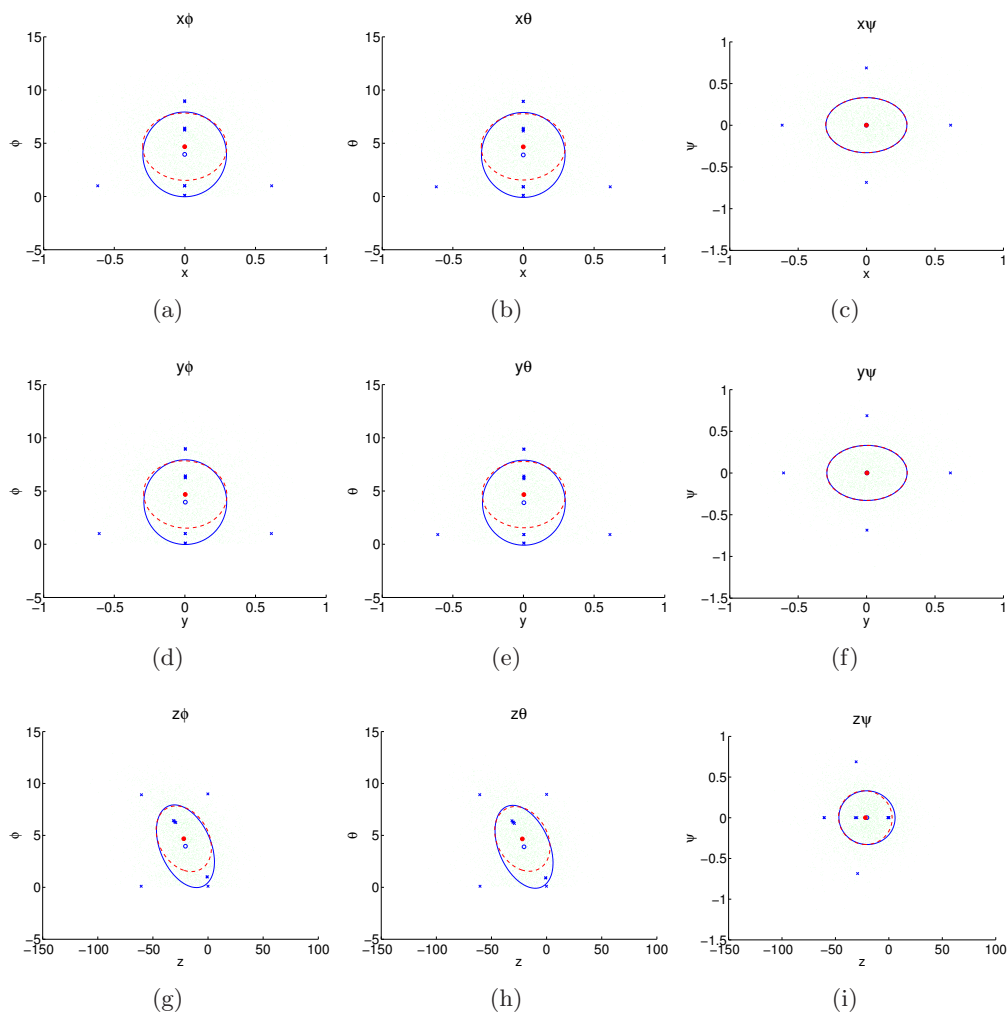


Figura C.2: Representació de submatrius de covariància 2×2 entre translacions i rotacions per a pertorbacions al voltant de la posició inicial de la càmera. Es mostren els resultats de Monte Carlo (punts verds per a les projeccions de cada mostra, i un cercle sòlid i una el·lipse discontinua per a la mitjana i la covariància) i els resultats de la UT (creus per a punts sigma, i un cercle buit i una el·lipse sòlida per a la mitjana i la covariància).

A la figura C.2 es poden veure les matrius de covariàncies 2×2 restants, que inclouen un component de translació i un de rotació. Es pot veure clarament l'efecte del biaix per a les rotacions aproximadament zero en totes les figures que impliquen les rotacions R_x i R_y . Anàlogament a la figura anterior, els casos restants es calculen correctament amb la UT.

C. Resultats del càlcul de la covariància en el moviment

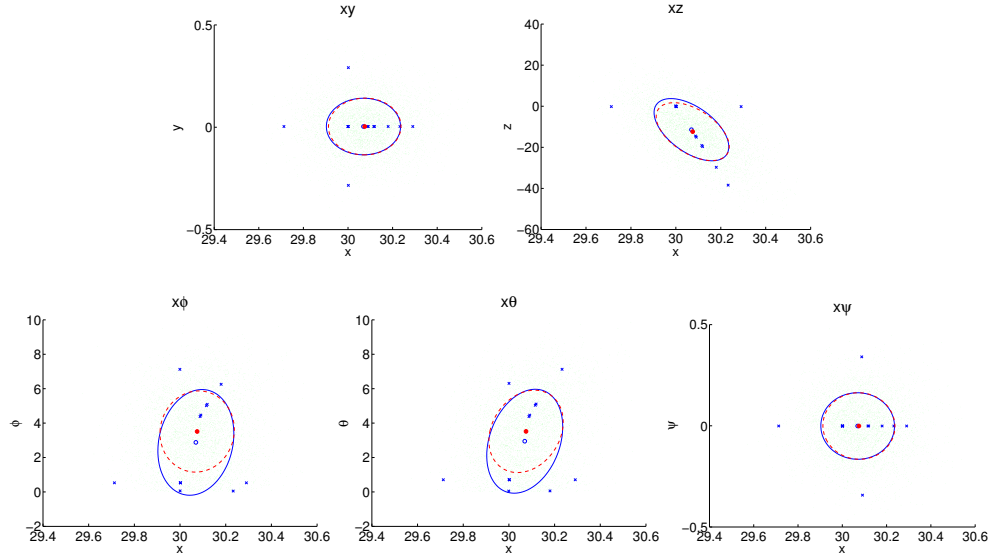


Figura C.3: Representació de les submatrius de covariància 2×2 de translacions i rotacions per a pertorbacions al voltant de $T_x = 30$ mm. Es mostren els resultats de Monte Carlo (punts verds per a les projeccions de cada mostra, i un cercle sòlid i una el·lipse discontinua per a la mitjana i la covariància) i els resultats de la UT (creus per a punts sigma, i un cercle buit i una el·lipse sòlida per a la mitjana i la covariància). Els resultats són similars per a una translació de $T_y = 30$ mm.

Es pot observar una correlació entre l'estimació de la rotació ϕ al voltant de l'eix X i la translació al llarg de l'eix Z , que s'ha explicat anteriorment a la secció 5.4.2. Apareix una relació creuada similar entre la rotació θ al voltant de l'eix Y i T_z .

C.2 Moviments simples de la càmera

Ara es volen calcular les covariàncies quan es produeix un moviment de la càmera al llarg i al voltant de cadascun dels eixos de coordenades. La informació significativa apareix en l'eix de moviment escollit per a cada experiment, així que de les quinze submatrius 2×2 només cal mostrar les cinc submatrius que impliquen aquest eix.

Primer es fa una translació de 300 mm al llarg de l'eix X . A la figura C.3 es poden veure els resultats obtinguts. La incertesa és congruent amb els valors presentats a la figura C.1(a). La translació T_x es recupera correctament amb mitjana $\bar{x} \approx 0$ i incertesa petita. A la figura que relaciona $T_x - T_z$ sembla que hi hagi present una correlació entre les dues variables, però observant l'escala de la figura es pot concloure que la

C.2 Moviments simples de la càmera

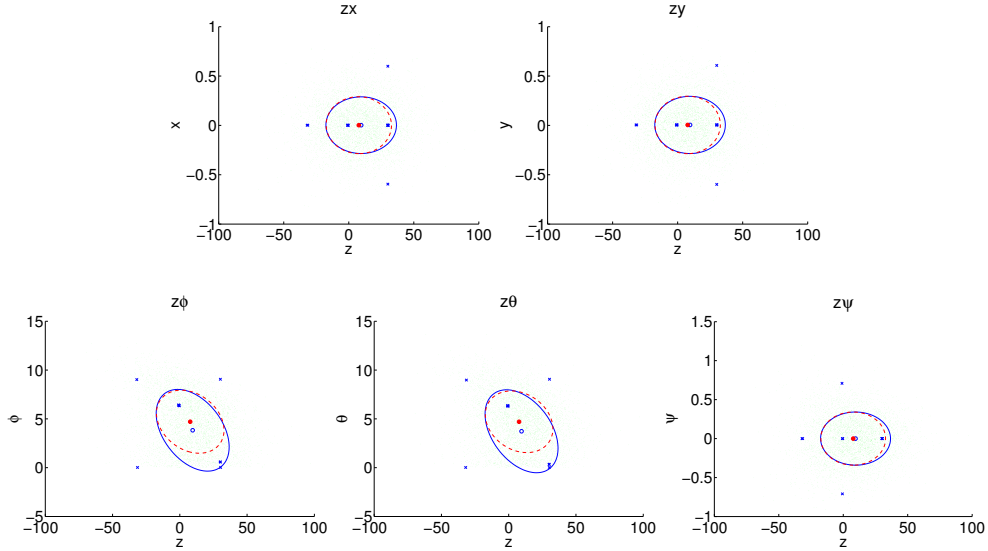


Figura C.4: Representació de les submatrius de covariància 2×2 de translacions i rotacions per a pertorbacions al voltant de $T_x = 30$ mm. Es mostren els resultats de Monte Carlo (punts verds per a les projeccions de cada mostra, i un cercle sòlid i una el·lipse discontinua per a la mitjana i la covariància) i els resultats de la UT (creus per a punts sigma, i un cercle buit i una el·lipse sòlida per a la mitjana i la covariància).

correlació és realment inapreciable. La figura que relaciona les components $R_x - R_y$ mostra l'efecte del biaix descrit anteriorment, i per tant, tampoc no es pot extreure cap conclusió. T_x i R_z no presenten gens de correlació, com es pot veure en l'última gràfica. La diferència entre els estadístics calculats amb la UT respecte dels calculats amb Monte Carlo és insignificant, excepte en les components R_x i R_y (aquest efecte també es pot observar a les figures prèvies C.1(b) i C.2).

Els resultats i les consideracions presentades pel moviment T_x són també vàlids en el cas d'una translació T_y .

Quan la translació es fa al llarg de l'eix Z (figura C.4), les translacions T_x i T_y es calculen correctament, i la seva incertesa es manté petita, però el valor T_z se subestima. Com era d'esperar, el biaix en les rotacions R_x i R_y també és present, ja que són rotacions properes a zero. La correlació observada en l'experiment de moviment nul entre $T_z - R_x$ i $T_z - R_y$ també s'observa aquí. La UT, altre cop, ajusta correctament la mitjana i la covariància als valors obtinguts amb Monte Carlo.

L'experiment següent es fa orbitant la càmera al voltant de l'objecte sobre l'eix

C. Resultats del càlcul de la covariància en el moviment

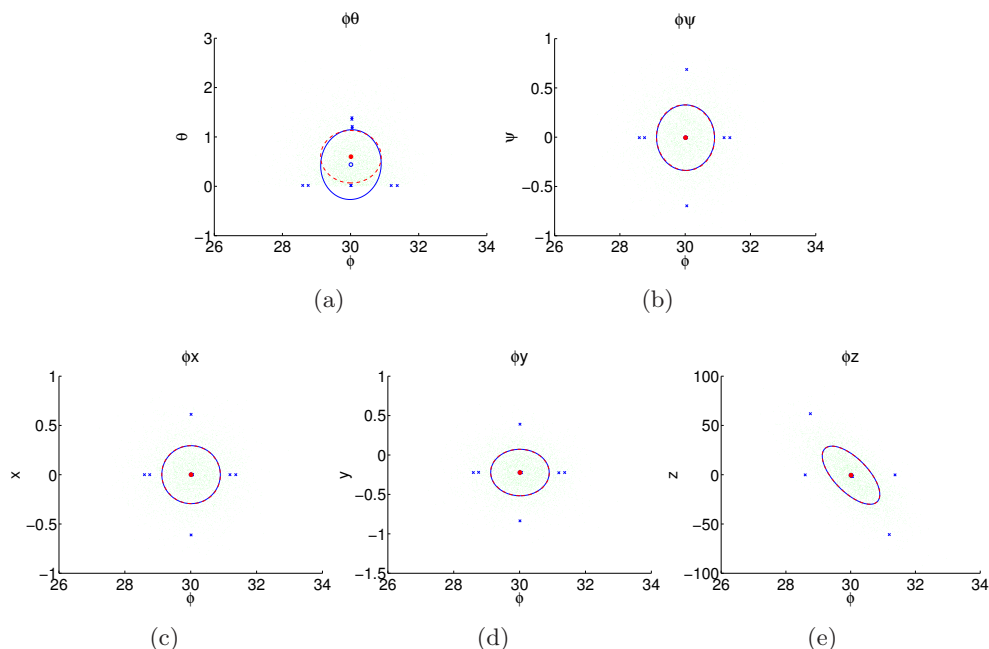


Figura C.5: Representació de les submatrius de covariància 2×2 de translacions i rotacions per a pertorbacions al voltant de $R_x = 30$ mm. Es mostren els resultats de Monte Carlo (punts verds per a les projeccions de cada mostra, i un cercle sòlid i una el·lipse discontinua per a la mitjana i la covariància) i els resultats de la UT (creus per a punts sigma, i un cercle buit i una el·lipse sòlida per a la mitjana i la covariància). Els resultats són similars per a una translació de $R_y = 30$ mm.

X (figura C.5). De manera congruent amb els resultats previs (comparant amb la figura 5.4(b) quan la rotació està al voltant de 30°), la incertesa en els valors R_x és menor que en els experiments sense rotació (Fig. C.1). El biaix en la component R_x ha desaparegut i en canvi encara està present en la component R_y . R_z es recupera correctament i els valors d'incertesa no han variat. Respecte de les translacions, T_x es recupera correctament però T_y es calcula amb un lleuger error. Aquest efecte ja s'ha esmentat anteriorment a la secció 5.4.2 i s'ha explicat mitjançant el procés de formació de la imatge amb el model de perspectiva dèbil (figura 5.7(b)).

L'última gràfica de la figura C.5 mostra la correlació entre R_x i T_z presentada anteriorment a la secció 5.4.2. Es pot observar que el biaix introduït quan les rotacions R_x i R_y estan sobrevalorades no és present. Això és perquè les rotacions de valor gran es calculen millor i, consegüentment, no cal subestimar T_z per compensar els errors. Un altre cop els estadístics obtinguts amb la UT són molt semblants als obtinguts amb

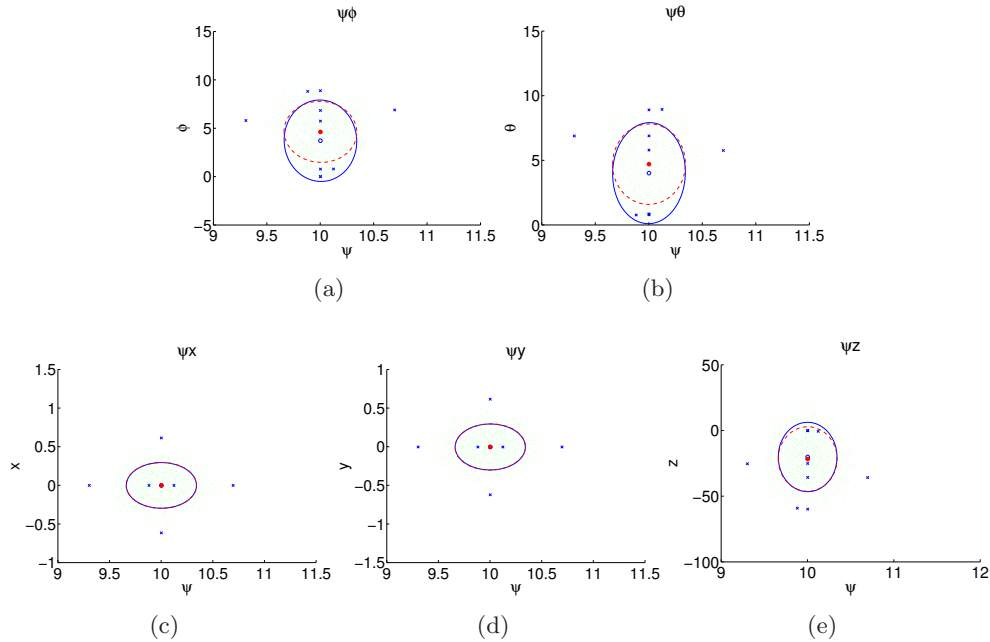


Figura C.6: Representació de les submatrius de covariància 2×2 de translacions i rotacions per a pertorbacions al voltant de $R_z = 10$ mm. Es mostren els resultats de Monte Carlo (punts verds per a les projeccions de cada mostra, i un cercle sòlid i una el·lipse discontinua per a la mitjana i la covariància) i els resultats de la UT (creus per a punts sigma, i un cercle buit i una el·lipse sòlida per a la mitjana i la covariància).

la simulació de Monte Carlo.

Les consideracions anteriors són aplicables a les rotacions sobre l'eix Y .

En l'últim experiment es produeix una rotació de la càmera al voltant l'eix òptic Z . Els resultats es mostren a la figura C.6. El valor d'aquesta rotació es recupera de forma precisa, i l'error es manté petit. No s'observa cap correlació amb les rotacions R_x o R_y , però el biaix típic en aquestes variables també està present. Les translacions es recuperen com en els experiments previs, i no s'observa cap correlació amb R_z .

Els experiments mostrats són congruents amb els resultats previs, tant de simulació com amb els fets en ambients reals. En tots els experiments s'ha observat que la covariància aproximada amb la UT és molt similar a l'obtinguda amb la simulació de Monte Carlo. Es pot concloure que la UT es pot utilitzar per calcular la covariància del moviment. S'ha confirmat la correlació entre les components T_z i R_x o R_y , i les petites translacions que es computen quan es fan algunes rotacions.

Apèndix D

Mètodes per seleccionar punts sigma

S'han proposat altres mètodes de selecció de punts sigma, però els beneficis que tenen per a la nostra aplicació actual no són clars. S'esmenten aquí per completar l'exposició, amb referències en què es pot trobar informació més detallada.

Si només cal capturar la mitjana i la covariància, es pot utilitzar l'anomenat conjunt *símplex*¹ de punts sigma [77]. Aquest és un conjunt de $N_x + 1$ punts que minimitza l'error dels tres primers moments estadístics (mitjana, covariància i biaix). Es pot interpretar com si tots els punts sigma estiguessin en una hiperesfera de radi $2^{N_x/2}$. Aquesta aproximació té un desavantatge important: fins i tot en dimensions relativament baixes hi ha problemes potencials d'estabilitat numèrica.

Per superar aquest problema va aparèixer l'esquema Símplex Esfèric [75]. En aquest cas, se seleccionen $N_x + 2$ punts, i tots els punts tenen el mateix pes, excepte el punt zero. Es pot interpretar com si tots els punts fossin en una hiperesfera de radi $\sqrt{N_x}(1 - W^0)$ amb els pesos dels punts sigma (excepte del punt zero) proporcionals a $(1 - W^0)/(N_x + 1)$. Tanmateix, depenent del valor del tercer moment estadístic (biaix) de la distribució, és preferible no utilitzar aquesta reducció. Si el biaix de la distribució veritable és zero, un conjunt Simètric de punts sigma serà més acurat a l'hora de capturar els estadístics. Generalment, el biaix d'un conjunt Símplex de punts no és zero, i en el pitjor cas (quan

¹En geometria, un símplex és l'envolvent convex d'un conjunt de $(n + 1)$ punts en un espai euclidià de dimensió n o més alt (per exemple, un conjunt de punts tal que cap m -pla no en conté més de $(m + 1)$).

D. Mètodes per seleccionar punts sigma

el biaix del conjunt Símplex no s'alinea amb el de la distribució veritable) els resultats són més pobres que els obtinguts amb un conjunt Simètric [74].

Si se sap que la distribució és gaussiana, els seus primers quatre moments (mitjana, covariància, biaix i curtosi) es poden capturar exactament amb una UT de quart ordre² [74, 76] utilitzant un conjunt de $2N_x^2 + 1$ punts sigma. No es pot capturar el sisè moment ni moments d'ordre més alt, però els punts s'escullen per tal de minimitzar-ne els errors. Es pot interpretar com posar els punts sigma en les cares d'un hipercub, així com a les cantonades.

La *Unscented Transformation* escalada [77] és un mètode per reduir arbitràriament qualsevol conjunt donat de punts sigma de tal manera que els primers dos moments estadístics de la distribució es conserven. Quan es té coneixement dels estadístics d'ordre més alt es fa servir per incorporar-les parcialment a l'algorisme. Normalment s'utilitza combinat amb l'esquema Simètric [133]. Un paràmetre (α) escala els pesos i un altre paràmetre (β) escala la contribució del punt zero a la covariància.

²De manera anàloga a la UT de segon ordre, s'anomena de quart ordre, ja que es pretenen conservar els primers quatre moments estadístics en fer la transformació.

Producció científica

Aquesta tesi ha permès, fins al moment, la següent producció científica:

Àmbit internacional

Publicació	Capítol relacionat
G. Alenyà, E. Martínez, and C. Torras. Fusing visual tracking with inertial sensing to recover robot egomotion. <i>In INERVIS. Procs of International Conference on Advanced Robotics</i> , volume 3, pages 1891-1898, 2003.	2
G. Alenyà, E. Martínez, and C. Torras. Fusing visual and inertial sensing to recover robot egomotion. <i>Journal of Robotic Systems</i> , 21:23-32, 2004.	2
G. Alenyà, J. Escoda, A.B. Martínez, and C. Torras. Using laser and vision to locate a robot in an industrial environment: A practical experience. <i>In Proceedings of the IEEE International Conference on Robotics and Automation</i> , pages 3539-3544, Barcelona, April 2005.	2
M. Alberich-Carramiñana, G. Alenyà, J. Andrade-Cetto, E. Martínez, and C. Torras. Affine epipolar direction from two views of a planar contour. <i>In Proceedings of the Advanced Concepts for Intelligent Vision Systems Conference</i> , LNCS 4179, pages 944-955, Antwerp, September 2006.	3

Producció científica

Publicació	Capítol relacionat
G. Alenyà and C. Torras. Robot egomotion from the deformation of active contours. In V. Kordic, A. Lazinica, and M. Merdan, editors, <i>Mobile Robots</i> . pro literatur Verlag, 2007.	2, 5
G. Alenyà, M. Alberich and C. Torras. Depth from the visual motion of a planar target induced by zooming. In <i>Proceedings of the IEEE International Conference on Robotics and Automation</i> , pages 4727-4732, Rome, April, 2007.	6
G. Alenyà and C. Torras. Zoom control to compensate camera translation within a robot egomotion estimation approach. In <i>Sixth International Workshop on Robot Motion and Control (Ro-MoCo)</i> , LNCIS 360, pages 81-88, Bukowy Dworek, Poland, June 2007.	7

Àmbit nacional

Publicació	Capítol relacionat
G. Alenyà, E. Martínez, C. Torras. Estimació del moviment d'un robot observant contorns actius. 1es Jornades UPC de Recerca en Automàtica, Visió i Robòtica, pages 319-324, Barcelona, 2004.	2
G. Alenyà, C. Torras. Anàlisi estadístic de la propagació de l'error en estimar el moviment 3D a partir d'imatges. 2es Jornades UPC de Recerca en Automàtica, Visió i Robòtica, pages 161-168, Barcelona, 2006.	5

Enviats a revistes per a la seva revisió

Publicació	Capítol relacionat
G. Alenyà and C. Torras. Camera motion estimation by tracking contour deformation: Precision analysis. Enviat el 2 de novembre de 2006.	5
M. Alberich-Carramiñana, G. Alenyà, J. Andrade-Cetto, E. Martínez, and C. Torras. Recovering Epipolar Direction from Two Affine Views of a Planar Object. Enviat l'1 de març de 2007.	3
G. Alenyà, M. Alberich-Carramiñana and C. Torras. Depth from zooming on arbitrary targets. Enviat el 18 de juliol de 2007.	6

Bibliografia

- [1] M. Alberich-Carramiñana, G. Alenyà, J. Andrade-Cetto, E. Martínez, and C. Torras. Affine epipolar direction from two views of a planar contour. In *Advanced Concepts for Intelligent Vision Systems*, volume 4179 of *Lecture Notes in Computer Science*, pages 944–955, Antwerp, September 2006.
- [2] G. Alenyà, J. Escoda, A.B.Martínez, and C. Torras. Using laser and vision to locate a robot in an industrial environment: A practical experience. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3539–3544, Barcelona, April 2005.
- [3] G. Alenyà, E. Martínez, and C. Torras. Fusing visual tracking with inertial sensing to recover robot egomotion. In *INERVIS. Procs of International Conference on Advanced Robotics*, volume 3, pages 1891–1898, 2003.
- [4] G. Alenyà, E. Martínez, and C. Torras. Fusing visual and inertial sensing to recover robot egomotion. *Journal of Robotic Systems*, 21:23–32, 2004.
- [5] G. Alenyà and C. Torras. Depth from the visual motion of a planar target induced by zooming. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 4727–4732, Rome, April 2007.
- [6] G. Alenyà and C. Torras. Robot egomotion from the deformation of active contours. In Sascha Kolski, editor, *Mobile Robots, Perception and Navigation*, pages 1–18. pro Literatur Verlag, 2007.
- [7] G. Alenyà and C. Torras. Zoom control to compensate camera translation within a robot egomotion estimation approach. In *Sixth International Workshop on Robot*

BIBLIOGRAFIA

- Motion and Control*, volume 360 of *Lecture Notes in Control and Information Sciences*, pages 81–88, 2007.
- [8] John Y. Aloimonos. Perspective approximations. *Image and Vision Computing*, 8(3):179–192, 1990.
- [9] K.J. Astrom and T. Hagglund. *PID Controllers: Theory, Design, and Tuning*. Instrument Society of America, 2nd. edition, 1995.
- [10] R. Atienza and A. Zelinsky. Interactive skills using active gaze tracking. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 188–195. ACM Press, 2003.
- [11] M. Baba, N. Asada, A. Oda, and T. Migita. A thin lens based camera model for depth estimation from blur and translation by zooming. In *Proceedings of the 15th International Conference on Vision Interface*, pages 274–281, Calgary, May 2002.
- [12] R Bajcsy. Active perception. *Proceedings of the IEEE*, 76:966–1005, 1988.
- [13] S. B. Balakirsky and R. Chellappa. Performance characterization of image stabilization algorithms. *Real Time Imaging*, 2(5):297–313, October 1996.
- [14] A. Bartoli, P. Sturm, and R. Horaud. Structure and motion from two uncalibrated views using points on planes. In *Proc. 3rd. Intl. Conf. on 3D Digital Imaging and Modeling*, pages 83–90, Canada, 2001.
- [15] P.A. Beardsley, A. Zisserman, and D.W. Murray. Sequential updating of projective and affine structure from motion. *International Journal of Computer Vision*, 23(3):235–259, 1997.
- [16] M. Betke and L. Gurvits. Mobile robot localization using landmarks. *IEEE Transactions on Robotics and Automation*, 13(2):251–263, 1997.
- [17] A. Blake and M. Isard. *Active contours*. Springer, 1998.
- [18] A. Blake, M. Isard, and D. Reynard. Learning to track the visual motion of contours. *Artificial Intelligence*, 78:101–134, 1995.

- [19] J. Borenstein, H.R. Everett, L. Feng, and D. Wehe. Mobile robot positioning - sensors and techniques. *Journal of Robotic Systems*, 14(4):231–249, 1997.
- [20] M. Born and E. Wolf. *Principles of Optics*. Cambridge University Press, 7th edition, 1999.
- [21] S.S. Brandt. Conditional solutions for the affine reconstruction of n-views. *Image and Vision Computing*, 23(7):619–630, July 2005.
- [22] M. Brooks, W. Chojnacki, D. Gawley, and A. van den Hengel. What value covariance information in estimating vision parameters? In *International Conference on Computer Vision*, pages 302–308, 2001.
- [23] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1992.
- [24] J.W. Bruce and P.J. Giblin. *Curves and Singularities*. Cambridge, 1992.
- [25] F. Caballero, L. Merino, J. Ferruz, and A. Ollero. A visual odometer without 3d reconstruction for aerial vehicles. applications to building inspection. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 4673–4678, Barcelona, April 2005.
- [26] T. Cham and R. Cipolla. Automated b-spline curve representation incorporating mdl and error-minimizing control point insertion strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(1), 1999.
- [27] T. Cham and R. Cipolla. Stereo coupled active contours. In *Proceedings of the 13th IEEE Conference on Computer Vision and Pattern Recognition*, pages 1094 – 1099, Fort Collins, June 1999.
- [28] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley. On the fitting of surfaces to data with covariances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1294–1303, 2000.
- [29] O. Chum, T. Werner, and J. Matas. Two-view geometry estimation unaffected by a dominant plane. In *Proceedings of the 19th IEEE Conference on Computer Vision and Pattern Recognition*, pages 772–780, San Diego, June 2005.

BIBLIOGRAFIA

- [30] J. J. Clark and N. J. Ferrier. *Attentive visual servoing*, pages 137–154. MIT Press, 1993.
- [31] J. Clarke and A. Zisserman. Detection and tracking of independent motion. *Image and Vision Computing*, 14:565–572, Aug. 1996.
- [32] J.C. Clarke. Modelling uncertainty: A primer. Technical Report 2161/98, University of Oxford. Dept. Engineering science, 1998.
- [33] R.T. Collins and Y. Tsin. Calibration of an outdoor active camera system. In *Proceedings of the 13th IEEE Conference on Computer Vision and Pattern Recognition*, pages 528–534, Fort Collins, June 1999.
- [34] A. Cretual and F. Chaumette. Dynamic stabilization of a pan and tilt camera for submarine image visualization. *Computer Vision and Image Understanding*, 79(19):47–65, 2000.
- [35] A. Cretual and F. Chaumette. Application of motion-based visual servoing to target tracking. *International Journal of Robotics Research*, 20:878–890, 2001.
- [36] A. Criminisi. *Accurate visual metrology from single and multiple uncalibrated images*. Springer, 2001.
- [37] G. Csurka, C. Zeller, Z. Zhang, and O.D. Faugeras. Characterizing the uncertainty of the fundamental matrix. *Computer Vision and Image Understanding*, 68(1):18–36, 1997.
- [38] J. Davis and X. Chen. Calibrating pan-tilt cameras in wide-area surveillance networks. In *Proceedings of the IEEE International Conference on Computer Vision*, Nice, October 2003.
- [39] A. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1403–1410, Nice, October 2003.
- [40] A. J. Davison. *Mobile Robot Navigation Using Active Vision*. PhD thesis, University of Oxford, 1998.

- [41] L. de Agapito, R. I. Hartley, and E. Hayman. Linear self-calibration of a rotating and zooming camera. In *Proceedings of the 13th IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, June 1999.
- [42] L. de Agapito, E. Hayman, and I. Reid. Self-calibration of rotating and zooming cameras. *International Journal of Computer Vision*, 45(2), November 2001.
- [43] C. Delherm, Jean-Marc Lavest, Michel Dhome, and Jean-Thierry Laprest. Dense reconstruction by zooming. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision*, volume 1065 of *Lecture Notes in Computer Science*, pages 427–438, London, UK, April 1996. Springer-Verlag.
- [44] D. Demirdjian, A. Zisserman, and R. Horaud. Stereo autocalibration from one plane. In D. Vernon, editor, *Proceedings of the 6th European Conference on Computer Vision*, volume 1842, 1843 of *Lecture Notes in Computer Science*, pages 625–639, Dublin, June 2000. Springer-Verlag.
- [45] J. Denzler, M. Zobel, and H. Niemann. Information theoretic focal length selection for real-time active 3d object tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 1, pages 400–407, Nice, October 2003.
- [46] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice*. Springer, 2001.
- [47] C. Drocourt, L. Delahoche, 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*, volume 2, pages 1329–1334, Detroit, May 1999.
- [48] T. Drummond and R. Cipolla. Application of lie algebras to visual servoing. *International Journal of Computer Vision*, 37(1):21–41, 2000.
- [49] Z. Duric and A. Rosenfeld. Shooting a smooth video with a shaky camera. *Machine Vision Applications*, 13(5):303–313, 2003.

BIBLIOGRAFIA

- [50] A. Earnshaw and S. Blostein. A perturbation analysis and performance bound for the estimation of 3-d camera translation direction. *Canadian journal of electrical and computer engineering*, 23(1):25–37, 1998.
- [51] O. Faugeras. *Three-Dimensional Computer Vision. A Geometric Viewpoint*. The MIT Press, Cambridge, 1993.
- [52] J.A. Fayman, O. Sudarsky, and E. Rivlin. Zoom tracking. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 4, pages 2783–2788, Leuven, May 1998.
- [53] J.A. Fayman, O. Sudarsky, E. Rivlin, and M. Rudzsky. Zoom tracking and its applications. *Machine Vision and Applications*, 13(1):25 – 37, 2001.
- [54] 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*, 24:381–385, 1981.
- [55] J. Foley, A. van Dam, S. Feiner, and F. Hughes. *Computer Graphics. Principles and Practice*. Addison-Wesley Publishing Company, 1996.
- [56] D.A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice-Hall, 2003.
- [57] D. Fox, S. Thrun, W. Burgard, and F. Dellaert. Particle filters for mobile robot localization. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte-carlo Methods in Practice*, Statistics for Engineering and Information Science, chapter 19, pages 401–428. Springer, 2001.
- [58] B. Georgescu and P. Meer. Point matching under large image deformations and illumination changes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):674–688, 2004.
- [59] S. Ghidary, Y. Nakata, T. Takamori, and M. Hattori. Human detection and localization at indoor environment by home robot. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, Nashville, October 2000.

- [60] N. Gonçalves and H. Araújo. Analysis and comparison of two methods for the estimation of 3d motion parameters. *Robotics and Autonomous Systems*, 45(1):23–50, 2003.
- [61] K.D. Gremban, C.E. Thorpe, and T. Kanade. Geometric camera calibration using systems of linear equations. In *Proceedings of the Image Understanding Workshop*, pages 820–825, Cambridge, 1988.
- [62] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition, 2004.
- [63] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, June 1997.
- [64] E. Hayman. *The Use of Zoom within Active Vision*. PhD thesis, Oxford, 2000.
- [65] E. Hayman and D.W. Murray. The effects of translational misalignment when self-calibrating rotating and zooming cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):1015–1020, August 2003.
- [66] E. Hayman, T. Thorhallson, and D.W Murray. Zoom-invariant tracking using points and lines in affine views. an application of the affine multifocal tensors. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 269–277, Corfu, September 1999.
- [67] R. Horaud, F. Dornaika, B. Lamiroy, and S. Christy. Object pose: the link between weak perspective, paraperspective, and full perspective. *International Journal of Computer Vision*, 22(2), 1997.
- [68] R. P. Horaud, D. Knossow, and M. Michaelis. Camera cooperation for achieving visual attention. *Machine Vision Applications*, 16(6):331–342, February 2006.
- [69] Berthold KP Horn. *Robot vision*. MIT Press, 1986.
- [70] G. Jang, S. Kim, W. Lee, and I. Kweon. Color landmark-based self-localization for indoor mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1037–1042, Washington, May 2002.

BIBLIOGRAFIA

- [71] P. Jensfelt, D. Kragic, J. Folkesson, and M. Bjorkman. A framework for vision based bearing only 3d slam. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1944–1950, Orlando, May 2006.
- [72] 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*, 45(3):477–482, March 2000.
- [73] S. J. Julier and J. K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In I. Kadar, editor, *Proceedings of the 11th SPIE International Symposium on Aerospace/Defense Sensing, Simulation, and Controls*, pages 182–193, Orlando, April 1997. International Society for Optical Engineering.
- [74] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, March 2004.
- [75] SJ. Julier. The spherical simplex unscented transformation. In *Proceedings of the American Control Conference*, Denver, June 2003.
- [76] SJ. Julier and JK. Uhlmann. A consistent, unbiased method for converting between polar and cartesian coordinate systems. In *SPIE Proceedings*, 1997.
- [77] SJ. Julier and JK. Uhlmann. Reduced sigma point filters for the propagation of means and covariances through nonlinear. In *Proceedings of the American Control Conference*, pages 887–892, Anchorage, Alaska, May 2002.
- [78] S. Kalkan, F. Wörgötter, and N. Krüger. Statistical analysis of local 3d structure in 2d images. In *Proceedings of the 20th IEEE Conference on Computer Vision and Pattern Recognition*, pages 1114–1121, New York, June 2006.
- [79] J. Y. Kaminski and A. Shashua. On calibration and reconstruction from planar curves. In D. Vernon, editor, *Proceedings of the 6th European Conference on Computer Vision*, volume 1842, 1843 of *Lecture Notes in Computer Science*, pages 678–694, Dublin, June 2000. Springer-Verlag.
- [80] Y. Kanazawa and K. Kanatani. Do we really have to consider covariance matrices for image features? In *Proceedings of the IEEE International Conference on Computer Vision*, pages 301–306, Vancouver, BC, Canada, July 2001.

-
- [81] J. Koenderink and A. J. van Doorn. Affine structure from motion. *J. Opt. Soc. Am. A*, 8(2):377–385, 1991.
- [82] K. Kuniyoshi, N. Kita, K. Sugimoto, S. Nakamura, and T. Suehiro. A foveated wide angle lens for active vision. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2982–2988, Nagoya, May 1995.
- [83] J.M. Lavest, C. Delherm, B. Peuchot, and N. Daucher. Implicit reconstruction by zooming. *Computer Vision and Image Understanding*, 66(3):301–315, June 1997.
- [84] J.M. Lavest, G. Rives, and M. Dhome. Three-dimensional reconstruction by zooming. *IEEE Transactions on Robotics and Automation*, 9:196–206, 1993.
- [85] T. Lefebvre, H. Bruyninckx, and J. De Schutter. Comment on “ A new method for the nonlinear transformation of means and covariances in filters and estimators”. *IEEE Transactions on Automatic Control*, 47(8):1406–1408, 2002.
- [86] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1465–1479, 2006.
- [87] M. Li and J.-M. Lavest. Some aspects of zoom-lens camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(11):1105–1110, November 1996.
- [88] Y. Liu, T.S. Huang, and O.D. Faugeras. Determination of camera location from 2d to 3d lines and point correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):28–37, 1990.
- [89] H.C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. *Nature*, 293(11):133–135, 1981.
- [90] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [91] J. Ma and S. I. Olsen. Depth from zooming. *Journal of the Optical Society of America A-Optics Image Science and Vision*, 7(10):1883–1890, oct 1990.

BIBLIOGRAFIA

- [92] Y. Ma, S. Soatto, J. Koseck?, and S. S. Sastry. *An invitation to 3-D vision*. Springer, 2004.
- [93] E. Malis and R. Cipolla. Camera self-calibration from unknown planar structures enforcing the multiview constraints between collineations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1268–1272, September 2002.
- [94] D. Marr. *Vision: a computational investigation into the human representation and processing of visual information*. Freeman, 1982.
- [95] H. A. Martins, J. R. Birk, and R. B. Kelley. Camera models based on data from two calibration planes. *Computer Graphics Image Processing*, 17(2):173–180, 1981.
- [96] H. Martinsson, A. Bartoli, F. Gaspard, and J.-M. Lavest. Handling missing data in the computation of 3d affine transformations. In *Proc. of EMMCVPR*, number 3757 in LNCS, pages 90–106. Springer, 2005.
- [97] E. Martínez. *Recovery of 3D structure and motion from the deformation of an active contour in a sequence of monocular images*. PhD thesis, Universitat Politècnica de Catalunya, 2000.
- [98] E. Martínez and C. Torras. Qualitative vision for the guidance of legged robots in unstructured environments. *Pattern Recognition*, 34:1585–1599, 2001.
- [99] E. Martínez and C. Torras. Contour-based 3d motion recovery while zooming. *Robotics and Autonomous Systems*, 44:219–227, 2003.
- [100] A. M. Mathai. *Jacobians of matrix transformations and functions of matrix argument*. World Scientific Publishers, 1997.
- [101] P. F. McLauchlan and D. W. Murray. A unifying framework for structure and motion recovery from image sequences. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 314–320, Massachusetts, June 1995.
- [102] C. Micheloni and G.L. Foresti. Zoom on target while tracking. In *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 117–120, September 2005.

-
- [103] D. D. Morris, K. Kanatani, and T. Kanade. Uncertainty modeling for optimal structure from motion. In *Vision Algorithms Theory and Practice*. Springer LNCS, September 1999.
- [104] D. W. Murray, F. Du, P. F. McLauchlan, I.D. Reid, P. M. Sharkey, and M. Brady. *Design of stereo heads*, pages 155–172. MIT Press, 1993.
- [105] David W. Murray, Kevin J. Bradshaw, Philip F. McLauchlan, Ian D. Reid, and Paul M. Sharkey. Driving saccade to pursuit using image motion. *International Journal of Computer Vision*, 16(3):205–228, 1995.
- [106] Z. Myles and N. da Vitoria Lobo. Recovering affine motion and defocus blur simultaneously. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6):652–658, June 1998.
- [107] Kevin Nickels and Seth Hutchinson. Weighting observations: The use of kinematic models in object tracking. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1677–1682, Leuven, May 1998.
- [108] P. Nordlund and T. Uhlin. Closing the loop: detection and pursuit of a moving object by a moving observer. *Image and Vision Computing*, 14(4):265–275, 1996.
- [109] T. Papadopoulo and I. A. Loukakis. Estimating the jacobian of the svd: theory and applications. Technical Report RR-3961, INRIA, 2000.
- [110] K. Pauwels, M. Lappe, and M.M. Hulle. Fixation as a mechanism for stabilization of short image sequences. *International Journal of Computer Vision*, 72(1):67–78, 2007.
- [111] P. Pomiers. Integration policy for critical multilayered distributed robotics applications. In *IEEE Intelligent Vehicle Symposium*, 2002.
- [112] I.D. Reid and D.W. Murray. Active tracking of foveated feature clusters using affine structure. *International Journal of Computer Vision*, 18(1):41–60, April 1996.
- [113] I.D. Reid, D.W. Murray, and K.J. Bradshaw. Towards active exploration of static and dynamic scene geometry. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 718–723, San Diego, May 1994.

BIBLIOGRAFIA

- [114] E. Rivlin and H. Rotstein. Control of a camera for active vision: Foveal vision, smooth tracking and saccade. *International Journal of Computer Vision*, 39(2):81–96, 2000.
- [115] V. Rodin and A. Ayache. Axial stereovision: Modelization and comparison between two calibration methods. In *Proceedings of the 1stIEEE International Conference on Image Processing*, pages 725–729, Austin, Texas, November 1994.
- [116] W. Rossmann. *Lie groups: an introduction through linear groups*. Oxford University Press, 2002.
- [117] S. Rowe and A. Blake. Statistical background modelling for tracking with a virtual camera. In *Proceedings of the British Machine Vision Conference*, pages 423–432, Birmingham, 1995.
- [118] S. Dutta Roy, S. Chaudhury, and S. Banerjee. Active recognition through next view planning: A survey. *Pattern Recognition*, 37(3):429–446, 2004.
- [119] D. Scharstein and A. Briggs. Real-time recognition of self-similar landmarks. *Image and Vision Computing*, 19(11):763–772, 2001.
- [120] Y.Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision*, 39(2):141–162, 2000.
- [121] B. Schiele and J. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36:31–50, 2000.
- [122] L. Sciavicco and B. Siciliano. *Modeling and Control of Robot Manipulators*. Springer-Verlag, London, 2000.
- [123] S. Se, D. Lowe, and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International Journal of Robotics Research*, 21(8):735–758, August 2002.
- [124] A. W. Senior, A. Hampapur, and M. Lu. Acquiring multi-scale images by pan-tilt-zoom control and automatic multi-camera calibration. In *Proceedings of the Seventh IEEE Workshops on Application of Computer Vision*, pages 433–438, 2005.

- [125] L. S. Shapiro, A. Zisserman, and M. Brady. 3D motion recovery via affine epipolar geometry. *International Journal of Computer Vision*, 16(2):147–182, 1995.
- [126] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the 9th IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, Seattle, June 1994.
- [127] R. Sim and G. Dudek. Learning environmental features for pose estimation. *Image and Vision Computing*, 17:445–460, 2001.
- [128] R. Sim and N. Roy. Global A-optimal robot exploration in SLAM. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 673–678, Barcelona, April 2005.
- [129] S. Sinha and M. Pollefeys. Towards calibrating a pan-tilt-zoom cameras network. In *5th OMNIVIS. Workshop on omnidirectional vision, camera networks, and non-classical cameras*, Prague, May 2004.
- [130] J. Sola, A. Monin, M. Devy, and T. Lemaire. Undelayed initialization in bearing only SLAM. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, August 2005.
- [131] G. W. Stewart. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Review*, 15(4):727–764, 1973.
- [132] P. Sturm and S.J. Maybank. On plane-based camera calibration: a general algorithm, singularities, applications. In *Proceedings of the 13th IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 432–437, Fort Collins, June 1999.
- [133] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, Cambridge, 2005.
- [134] C. Tomasi and Takeo Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, November 1992.

BIBLIOGRAFIA

- [135] B. Tordoff. *Active control of zoom for computer vision*. PhD thesis, University of Oxford, 2002.
- [136] B. Tordoff and D. Murray. Reactive control of zoom while fixating using perspective and affine cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):98–112, January 2004.
- [137] P. Torr and D. Murray. Outlier detection and motion segmentation. *Sensor Fusion VI, SPIE*, 2059:432–443, 1993.
- [138] A. Ude, C. Gaskett, and G. Cheng. Foveated vision systems with two cameras per eye. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3457–3462, Orlando, May 2006.
- [139] S. Utcke. Grouping based on projective geometry constraints and uncertainty. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 739–746, Bombay, January 1998.
- [140] R. van der Merwe, N. de Freitas, A. Doucet, and E. Wan. The unscented particle filter. In *Advances in Neural Information Processing Systems 13*, Nov 2001.
- [141] R. van der Merwe and E. Wan. The square-root unscented kalman filter for state and parameter-estimation. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, Utah, May 2001.
- [142] R. van der Merwe and E. Wan. Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Hong Kong, April 2003.
- [143] E.A. Wan and R Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pages 153–158, 2000.
- [144] 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*, 2(5):451–475, May 1989.

- [145] J. Weng, T. S. Huang, and N. Ahuja. *Motion and structure from image sequences*. Springer-Verlag, 1993.
- [146] J. H. Wilkinson. Modern error analysis. *SIAM Review*, 13(4):548–568, 1971.
- [147] R. G. Willson and S. A. Shafer. What is the center of the image? *Journal of the Optical Society of America A-Optics Image Science and Vision*, 11:2946–2955, 1994.
- [148] Z. Zhang. Determining the epipolar geometry and its uncertainty: a review. *International Journal of Computer Vision*, 27(2):161–195, 1998.
- [149] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330– 1334, 2000.