USCIPI REPORT #182

3-D Motion Estimation Using a Sequence of Noisy Images

by

Gem-Sun Jason Young

May 1991

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

3–D MOTION ESTIMATION USING A SEQUENCE OF NOISY IMAGES

by

Gem–Sun Jason Young

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(Electrical Engineering)

May 1991

To my parents

Acknowledgments

I can not convey enough appreciation to Professor Rama Chellappa, my advisor and dissertation committee chairman, for making this dissertation possible. I am grateful for the financial support which he has made available during my stay in USC. His understanding of image and signal processing, statistics, mathematics, and engineering influenced this dissertation considerably. His keen insights and profound knowledge of computer vision motivated and guided this work.

I would also like to thank other members of my committee, Professor Alexander Sawchuk and Professor Michael Waterman, for taking the time to read and evaluate my work.

Dr. Ted J. Broida and Professor Rama Chellappa laid down the foundation for 3–D motion analysis using a long sequence of images. Without their pioneering contributions, this dissertation would be impossible. I am indebted to Dr. Broida for patiently teaching me the motion estimation problem during the first year of my graduate study. His friendship has been invaluable.

I am thankful to Professor Lin–Shan Lee, chairman of Department of Information Science in National Taiwan University and Mr. Ju–Yung Wang, president of Elegant Overseas Co., Ltd., Taiwan, for supporting my education overseas. I would also like to thank Chueh–Hsiang Wang, Mr. Wang's son and my undergraduate fellow classmate, for his support and friendship.

A special thanks goes to Dr. Allen Weber. His remarkable knowledge of computer systems was of immense help in my successful battle with those beasts of technology. He unselfishly gave his time to fix many system problems and over the years has become the Very Important Person of the Signal and Image Processing Institute at USC.

The SIPI staff have been very helpful. I'd like to thank Toy Mayeda, Linda Varilla, Delsa Tan and Gloria Bullock for their assistance and friendship.

Professor Jay Kuo has been a good friend and source of advice on many aspects of life. Dr. Jinchi Hao and Chin-Chao Lin are the two senior fellow graduate students in SIPI who gave me the most help on my daily life as an international student. I thank my past office mates Chih-Ling Wang, Tal Simchony, Chein-Hsun Wang, Gregory Petrisor, Venkat Venkateswar, and Dr. Yitong Zhou for teaching me many computer commands. I would also like to thank Professor Shankar Chatterjee, Dr. Anand Rangarajan, Dr. Bob Frankot, Dr. Mike Rude, Dr. Dimitrios Kalivas, Dr. Tom Hebert, Dr. Kuang-Shuh Huang, Dr. Richard Hansen, Quinfen Zheng, Kyoung Mu Lee Lily Chang, Xiao-Hong Yan, Ehoud Haberman, George Zunich, Andrew Miller, Charlie Kuznia, Clare Waterson, Ting-Hu Wu, Navid Haddadi and Ta-Kang Ku for their good company.

This research was funded in part by NSF grant IRI-87-13585.

Contents

List of Figures

List of Tables

Abstract

In this dissertation we study several aspects of three dimensional (3–D) motion and structure estimation from a sequence of noisy images. A high order motion model is proposed based on representing the constant acceleration translational motion and constant precession rotational motion in the form of a bilinear state space model using standard rectilinear states for translation and quaternions for rotation. The measurements are noisy perturbations of 3-D feature points represented in an inertial coordinate system. The structure of the moving object is estimated from the stereo image pairs prior to the estimation of motion parameters. Owing to the nonlinearity in the state model, nonlinear filters are then designed for the estimation of motion parameters. The Cramér-Rao performance bounds for motion parameter estimates are computed. A constructive proof for the uniqueness of motion parameters is given. It is shown that with uniform sampling in time, three noncollinear feature points in five consecutive binocular image pairs contain all the spatial and temporal information required for motion estimation.

If 3–D motion and structure are estimated simultaneously, we show that the standard 3×3 rotation matrix is more suitable for representing the rotational motion than quaternions. Using this representation, a new motion model for constant velocity translation and constant angular velocity rotation is given. Both monocular and binocular systems are considered. Batch and recursive algorithms are designed to search and track, respectively, the moving rigid object. Occlusion among image frames as well as between each stereo pair in the binocular case is also dealt with. A nonlinear least squares method is used to formulate the batch estimation of motion and structure parameters. Nonlinear Kalman filters are used for the recursive algorithms. Owing to parameterization of translational motion using a time–invariant global normalization factor, linear plant models

are used in the recursive filters to update the motion in closed form, and thus avoiding the time–consuming numerical integration, which commonly arises in nonlinear filtering problems.

The optical flow field is another commonly used image measurement for motion estimation. However, flow fields generated by existing algorithms are noisy and thus induce ambiguities in motion parameters. In this work, the inherent ambiguities in recovering 3-D motion information from a single optical flow field are studied using a statistical model. These ambiguities are quantified using the Cramer-Rao lower bound (CRLB), which is a lower bound for the error variances of motion parameter estimates. This performance bound is independent of the motion estimation algorithms, and can always be computed for any arbitrary 3-D motion of a rigid surface by inverting a 5×5 matrix. As a special case, the performance bound for the motion of 3-D rigid planar surfaces is studied in detail. For the general motion of an arbitrary surface, it turns out that not every pixel gives information regarding 3-D motion estimation. It is shown that the aperture problem in computing optical flow restricts the nontrivial information about the 3-D motion to a sparse set of pixels at which both components of the flow velocity are observable. Effects of two smoothing schemes on estimation accuracy are analyzed. It is shown that by introducing a smoothness constraint by fitting local patches to 3–D depths gives lower CRLB's. Surprisingly, this reduction in CRLB's is very small. Further, fitting local patches also relaxes the aperture problem since the motion information is not restricted to points at which both optical flow components are observable. In contrast, imposing smoothness on the optical flow by regularization does not lower the CRLB's. Although these results are all derived using a simple Gaussian noise model and should be interpreted with caution, they nevertheless give some new insights in analyzing inherent ambiguities.

Chapter 1

Introduction

This dissertation presents research on estimating 3–D motion and shape from a sequence of image frames. The measurement data extracted from the image frames can be either noisy image coordinates of a set of object feature points or the apparent motion field of changing brightness patterns on the image plane. The 3–D motion information of interest is the translational and rotational kinematics of the 3–D objects relative to the cameras. The 3–D shape is described using either the relative positions of object feature points or a depth map of the 3–D world, in accordance with the types of measurement data applied. The number of optical sensors is flexible. In the dissertation, we specifically address the cases of using one and two cameras to estimate 3–D motion and structure from spatial and temporal dimensions of a sequence of dynamic images.

The tasks accomplished include design and analysis of motion and structure estimation algorithms, as well as an investigation of inherent ambiguities associated with inferring 3–D motion from an image sequence which fundamentally upper bounds the performance of any feasible motion estimation algorithms. The design and analysis of algorithm are focused on the measurements obtained by tracking discrete features. It is felt that a need exists in understanding the inherent ambiguities in solving 3–D motion from 2–D images, which is generally perceived to be an ill–conditioned inverse problem [3, 4, 38, 68]. Many researchers [5, 6, 30, 42] have observed that in the two–view motion estimation problem, relatively small variations in image measurements can result in large perturbations

of motion parameters. This dissertation presents an approach which quantifies the notion of inherent ambiguities by investigating a lower bound for uncertainty in motion estimates. Interestingly, from the point of view of amount of information contained in image measurements, our results favor the discrete based approach over the continuous approach. This has focused our attention to the design and analysis of motion algorithms using the image coordinates of a sparse set of object feature points.

The objects tracked are assumed to be rigid. If more than one object with independent motion is present in the scene, it is assumed that the images have been segmented so that we can concentrate on one region corresponding to a single object. A basic assumption made is that structure and motion of the object can be described by a relatively small number of parameters. This assumption seems sufficient to model the objects without large structure change and motion with smooth kinematics. The approach is based on Broida and Chellappa's general framework [18, 20, 21, 19] for modeling object structure, motion, and imaging, estimation of model parameters, and predicting parameter estimation accuracy for any given choice of models and measurement conditions. These techniques are extended to model and estimate higher order motions, such as precession, and to address the occlusion and reappearance of feature points among image frames. Relatively long image sequences are used. A long sequence of images looks at the same object from many different directions as the object moves in the scene. This is a desired feature in recovering 3–D from its 2–D projection as the information compressed in the projection direction for one image may be recovered by observing the projection along other directions. In our simulation, even for high image noise levels motion estimates with good precision have been obtained.

An important task in motion analysis is the issue of how much information is present in images for the purpose of motion recovery. A thourgh treatment of this problem is generally difficult due to the nonlinearity introduced by the perspective projection, the image formation process in a pin–hole camera. In the special case when a sequence of noise–free stereo images are available, we

are able to solve this problem by taking advantage of the 3–D depth information contained in the stereo image pairs. In this case, a procedure is derived to recover all available motion and structure information from images of projected feature points. The derived procedure not only illustrates the amount of information present in the motion and imaging models presented in this work, but also quantifies the minimum set of required image measurements which contains all available motion information in the models. The uniqueness of 3–D motion and structure parameters is also analyzed using this motion recovery procedure. As the results are obtained by analyzing noise–free image measurements, they are restrictive because in practice we deal with noisy image sequence and hence need to estimate motion from the noise contaminated data. It is nevertheless valuable to study what can be achieved using noise free data, since this puts an upper bound on what one can do with images. Actually in the presence of measurement noise, the issue of motion information can be addressed by studying whether or not there exists a large set of motion and structure parameters which induce image measurements similar to that induced by the correct parameters. The CRLB is applied to analyze this problem. Since CRLB limits the maximum possible precision of the estimated motion parameters given by any motion estimation algorithm, incorrect motion parameters which have shorter distance measures from the true motion parameters than CRLB can not be discriminated by using the noisy measurements. As a consequence, larger CRLB implies a larger set of the incorrect motion and structure parameters indistinguishable from the correct parameters. Hence larger CRLB indicates that less information is present in the images for the purpose of motion estimation.

1.1 Applications

3–D motion estimation has wide application in several domains in sciences and engineering. Here we highlight several practical problems which require 3–D motion estimation and analysis.

Numerous applications of 3–D motion estimation and analysis have been found in 3–D object tracking, robot navigation and manipulation, autonomous

vehicle navigation, self–calibration of vision systems, and space exploration [22, 23, 37, 76]. In tracking 3–D objects, where one has to analyze and predict its motion in 3–D space, a motion estimation algorithm can be useful. The predicted motion trajectory is then useful for avoiding collision and restricting the search area of features in subsequent images [37]. Field conditions for autonomous vehicles may be very dynamic, involving rough terrain and unexpected obstacles. The navigation route thus needs to be modified according to 3–D motion information of the environment detected from image sequences. In cases when outside bombardments cause structure misalignment, on–board self–calibration of vision systems may be done using the estimated motion information [22, 23]. On the other hand, in well controlled environments such as factories or space stations, many tasks performed by mobile robots require 3–D motion estimation and object recognition to successfully navigate, move objects, and avoid obstacles. Space exploration can also take advantage of motion estimation. For instance, in docking of two moving spacecrafts, each has to estimate the translational and rotational motion of the other accurately [76]. In sum, 3–D motion estimation can be applied in any situation where a machine interacts with its environment using a vision system.

Motion estimation research also has bearings on the understanding of biological vision systems. Due to the difficulty in directly monitoring how the brain functions to see, researchers proposed input–output relationships of the biological vision system to form hypotheses regarding the biological visual process [47, 48, 70, 71]. Under these hypotheses, motion estimation algorithms are formulated and simulated on a computer system. Then the input–output relationship of the simulated algorithms is compared to that of the biological vision system. Better understanding of natural vision systems may be obtained by this comparison. For example, the startling effects on optical illusions may in the future prove to be important for understanding the functions of biological vision systems [37]. On the other hand, engineering problems related to 3–D motion estimation may be benefited from biological research [1, 2, 13, 25, 35, 36, 48, 59, 58, 61, 72]. After an engineering algorithm of motion estimation is implemented, by observing

its behaviors, biologists can in turn gain additional insights that may help in formulating more interesting hypotheses about human vision.

Another practical application is in image data compression. If image storage space is at a premium, as is often the case of high density motion picture sequences, it may be worth compressing the data images. An improved image compression scheme results when motion correlation between image frames is used. For example, if the images are of aircrafts manuveuring in space, it may be simple to store just a list of motion information, the structure of aircrafts, and other high level features. At this point, image compression becomes more of scene understanding and description, and properly belongs to the domain of 3–D motion and shape modeling, rather than that of data compression. Recently there is an increasing interest in high–definition television (HDTV) for much improved TV reception. A key task in the development of HDTV systems is to accommodate the image signals to band limited communication channels by compressing the information in dynamic image sequences [31]. Motion information may be used to reduce a large amount of redundant information in spatial and temporal domain formed by stacking up the highly correlated dynamic images [17]. By encoding motion information, 3–D surface structure and texture, it is possible to significantly compress a long sequence of images which looks at the same environment.

1.2 Motion Estimation from Stereo Image Sequences

A significant amount of work has been done in computer vision literature [3, 8, 28, 29, 41, 46, 49, 60, 66, 67, 68, 69, 75] for the estimation of 3-D motion and/or structure parameters from a monocular sequence of images. Much of the earlier work involved using two or three frames of noisy images. The nonrobustness and lack of numerical stability of these two or three view monocular image based algorithms have been noted by several researchers [5, 6, 30, 42]. Some of these problems were partly alleviated using a long sequence of noisy image frames [20, 21, 18, 76].

We develop a robust kinematic model based algorithm for 3-D motion parameter estimation from an arbitrarily long sequence of noisy stereo images. The approach uses models for the motion of rigid object, the time evolution of this motion and the observation of the object based on discrete features. The modeled motion has constant acceleration in the translational components and constant precession [76] in the rotational components. The constant precession motion is modeled as the instantaneous angular velocity of the object rotating w.r.t. a spatially fixed axis with a constant angular velocity. Constant precession motion is an example of higher order rotation than constant angular velocity motion.

When standard rectilinear states are adopted for translational motion, our motion model results in a quadratic time evolution. The rotational motion is propagated using a quaternion representation. The combined translational and rotational motion is represented in the form of a state space model with twenty two states corresponding to nine translational motion states, ten rotational states, and three states which describe the position of the rotation center. Due to the propagation of rotation, the state space model tends to be bilinear.

The measurements are modeled as noisy perturbations of 3-D coordinates of a physical point on the rigid object represented in an inertial coordinate frame. The 3-D coordinates are assumed to be obtained from the stereo pair. Although the ultimate goal is to characterize the noise distribution using the error analysis [16] of a typical stereo triangulation algorithm, for the purposes of this dissertation we assume the noise to be Gaussian. It is also assumed that matching of 3-D feature points and establishing correspondence between the matching points have been accomplished [30, 62].

The rich information contained in the stereo images allows separation of the structure estimation stage from the motion estimation stage. Structure parameters are first estimated by a recursive estimator with the variances of estimates monotonically decreasing in time. Then the structure estimates are fed into the model for motion estimation. The 3-D motion estimation problem is thus reduced to one of identifying the states of a dynamical system with bilinear state model and trilinear measurement model. Due to the nonlinearities present, an

extended Kalman filter (EKF) [43, 51] is used for the identification of motion parameters. Our simulation results show that for heavy noise situations, modifications to EKF such as the iterated extended Kalman filters (IEKF) [43, 51] may be required. Once the parameters or states in the motion model are estimated, prediction and interpolation of missing measurements can be accomplished.

One of the major questions to be answered in motion estimation problem is the uniqueness or lack thereof of motion parameters [38, 68]. It is an inverse problem that estimates the motion parameters from image measurements. The uniqueness problem is to investigate whether the motion parameters can be decided uniquely or not from the noise-free measurements. Hence uniqueness is an intrinsic property of a motion model, independent of the motion parameter estimators applied. For the two-view motion estimation problem, elegant proofs of uniqueness of motion parameters have been established for both the feature based methods [68] and optical flow approaches [38]. We consider motion estimation using an arbitrarily long sequence of images and present a uniqueness proof for the specific motion and measurement models presented here. The rotational parameters are separated from the translational parameters using spatial information in stereo image pairs. Solutions for rotational parameters are obtained. Then translational parameters are solved from linear equations. The uniqueness of motion parameters is established through a constructive proof which is a deterministic algorithm for estimating motion parameters from the noise free measurements.

The uniqueness analysis for the constant acceleration and precession motion model shows that if uniform sampling scheme is used, three noncollinear feature points in four consecutive binocular image pairs and one of the above feature points at any other measurement time are sufficient to retrieve all the motion information. Under the uniform sampling scheme the observations are made only at equally separated time instants, so the measurement model is analogous to the technique in which a stroboscope is applied to observe the rotational rate of a system. Velocity and acceleration parameters are always unique. The translational position vector and the rotation center are generally unique except in some very restricted cases–only when the orientation of the object appears to

be frozen or rotating with a constant rate under the stroboscope. However, angular velocity and precession parameters are never unique because of the same stroboscopic effect. We show that the stroboscopic effect can be studied using the notion of interframe rotation which describes the visual rotation under the stroboscope. A generalized quaternion referred to as the interframe quaternion is used to quantify the interframe rotation. Different conditions of the interframe quaternion lead to different configurations of the solution set. As these conditions can always be detected from the image measurements, the configurations of the solution set are thus decided by the stereo image measurements. Corresponding to seven conditions of the interframe quaternion, there exist seven, one nondegenerate and six degenerate, different configurations for the solution set of the rotation parameters (quaternion, angular velocity, and precession). In the nondegenerate case, the direction of precession is unique while its magnitude can assume equally spaced discrete values. Both the direction and magnitude of the angular velocity are not unique even in the nondegenerate case. There are three different configurations in the solution set for the translational position vector and the rotation center, corresponding to three conditions of the interframe quaternion. Among these, only one is nondegenerate and leads to a unique solution. As suggested by these results, we recommend the application of nonuniform sampling schemes to constrain the stroboscopic effect and thus reduce the ambiguity in rotational parameters. This is supported by the techniques used to restrict the set of possible solutions by changing the sampling rate when the rotational rate of a system is measured by a stroboscope.

As a by-product of this uniqueness analysis, a deterministic algorithm to recover motion parameters from a stereo image sequence is developed. While the determination of motion parameters from noise-free image measurements using this analysis is precise and fast, accuracy is not guaranteed in the presence of noise. Therefore stochastic filters are recommended for the estimation of motion parameters from noisy images.

As the motion problem is formulated as an estimation problem, Cramér-Rao lower bound (CRLB) [20] is a useful technique to establish a performance bound. Based on CRLB it is possible to discuss how good the IEKF is as a motion

parameter estimator for our motion and measurement models. The result shows that, while the efficiency of IEKF is satisfactory, the mean square error does not closely follow CRLB due to the degradations caused by the errors in the structure parameter estimates obtained prior to motion estimation.

1.3 Motion Estimation Using a Simple Kinematic Model

In the last section, motion estimation using a stereo image sequence has been discussed. By exploiting the depth information preserved in stereo image pairs, we have separated the motion estimation stage from the structure estimation stage. This has enabled us to avoid the nonlinearity due to the central projection of pin hole cameras. The number of unknown parameters has also been reduced in each estimation stage. These two factors make the nonlinear Kalman filter stable so that the high order motion model of constant acceleration and precession can be considered. As the structure is estimated prior to motion estimation, the description of the instantaneous orientation of the moving object becomes necessary in the motion estimation stage. In this case, quaternions have been shown more useful in representing the rotational motion than the standard parameterization of rotational motion. However, the nonlinear plant model resulting from the quaternion representation requires time–consuming numerical integration to update the covariances of the estimated motion parameters. The occlusion of feature points has not been addressed, either. We discuss another new kinematic model which combines the motion and structure estimation stages and describes rotation by the standard rotation matrix representation. This representation is much simpler than the quaternion based representation. Occlusion is dealt with. The adoption of a linear plant model in the recursive algorithm also speeds up the nonlinear filter by avoiding numerical integration. The recursive and batch formulations will be given for motion and structure estimation both for monocular and stereo image sequences.

The approach uses models for the motion of the rigid object and the observation of the object based on discrete features. The modeled motion has constant

velocity in the translational components and constant angular velocity in the rotational components. The measurements are the coordinates of the images of feature points on the image plane.

When standard rectilinear parameters are adopted for 3–D motion, the translational motion is described by six parameters, including three for position and three for velocity; the rotational motion is quantified by the three parameters for the angular velocity. The structure of the object is parameterized in a 3–D coordinate system fixed on the object. The coordinates of each feature point in this coordinate system are chosen as the three structure parameters describing the location of each feature point on the object. The combined structure and motion are thus represented by nine motion parameters, and three structure parameters per feature point.

The measurements are modeled as noisy perturbations of 2-D coordinates of the projection of feature points on the image plane. The perspective projection model relates the motion and structure parameters to the expected image measurements. Using a nonlinear least squares formulation, we minimize a criterion function which is the square of the distance between the noisy measurements and the expected measurements. Therefore in the batch algorithm, the 3-D motion and structure estimation problem is reduced to one of identifying parameters in the motion and object model. A standard conjugate gradient search routine is applied to locate the minimum of the criterion function. Once the motion and structure parameters are estimated, prediction and interpolation of missing measurements can be accomplished.

The recursive algorithm uses the standard formulation of an Extended Kalman filter (EKF), or an Iterated Extended Kalman filter (IEKF) if the image noise is heavy. The kinematic model leads to a linear plant model with a sparse plant matrix in EKF/IEKF. Hence closed form solutions to the differential equations governing the propagations of the state vector and the covariance matrix are straight forward to obtain. These closed form solutions then replace the numerical integration used to solve the differential equations, and thus speed up the filters.

In our formulation, the occlusion problem is handled by simply dropping the corresponding measurements of the missing point in the least squares criterion

function for the batch algorithm. In the recursive algorithm, we incorporate only the measurements from unoccluded points into the measurement equation to take into account the effects of occlusion.

The depth information is dealt with differently between monocular and binocular systems. As the absolute depth information is not contained in a monocular system, the motion and structure parameters are normalized by the depth of the rotation center at the starting point. As for motion estimation in a binocular system, the last section separated the structure estimation stage from the motion estimation by using stereo triangulation to recover the 3–D depth of each feature point from the stereo image pair. However, when occlusion between the stereo pair occurs, stereo triangulation is not applicable as the corresponding point on the other camera may not be available. Hence in the simple motion model discussed in this section, to consider the occlusion problem, the 3–D depth of each feature point is not estimated directly from the stereo pair. Instead, motion and structure are estimated at the same time using the motion and measurement model. Then the estimated motion and structure may be used to recover the 3–D depths of feature points.

1.4 Analysis of Inherent Ambiguities

Much work in the estimation of 3-D motion and/or structure parameters from a monocular sequence of images uses two frames of noisy images. The nonrobustness and lack of numerical stability of these two view monocular image based algorithms have been noted by several researchers [4, 9, 10, 12, 11, 45, 39, 57, 64, 71]. These documented problems make it worthwhile to analyze the inherent ambiguities in recovering 3-D motion and structure from two frames of noisy dynamic images [4].

As the estimation of motion and structure parameters is based on image measurements, the performance of the algorithms will be intrinsically restricted by the information contained in the measurements used. The lack of information about motion and structure in the image measurements causes ambiguities in the estimation of parameters. These ambiguities are inherent in the sense that

they are independent of the algorithms implemented. Analytic results of these ambiguities have been obtained using deterministic analysis for both the noise-free [38, 68], and noise-contaminated [4] image measurements.

We deal with ambiguities in the optical flow field based algorithms. The common approach for motion estimation from the image velocity field involves two steps: computation of optical flow field and estimation of motion and/or structure parameters from the computed image flow field. Horn [38] assumes that the first step is accomplished without error and then analyzes the inherent ambiguity of motion parameters from the noise-free velocity field. Adiv [4] assumes that both components of the optical flow velocity vector obtained in the first step are contaminated with noise, and then investigates the inherent ambiguity in the second step. However, the effects of the aperture problem [35, 36] involved in the first step which prevents us from obtaining both the components of the optical flow field at a majority of pixels are not addressed in [4, 38]. Also in these analyses of inherent ambiguities, only the case when no smoothness constraint is incorporated has been analyzed, in spite of the commonly held belief that imposing appropriate smoothness constraints is a useful technique to stabilize an illposed problem. Thus the effects of regularizing the illposedness in the noisy optical flow field based methods have not been considered.

These issues are studied using the Cramer-Rao inequality [18, 20, 21, 26, 63, 77, 84] in a statistical setup, which is a natural way to model an estimation process based on measurements contaminated by random noise. In our approach, algorithms recovering 3-D motion are modeled as estimators which estimate the motion parameters from the noisy flow velocity field. Due to the random nature of the noisy input flow field, the estimators give random noisy estimates of motion parameters. The performance of a motion estimation algorithm or a motion estimator can thus be studied by investigating the deviation of its estimates from the true motion parameters. A natural term to quantify this deviation of random estimates is the error variance. In this context, a good motion estimation algorithm is supposed to be an estimator which gives estimates with small error variance. However, this error variance cannot be reduced to arbitrarily small values and is lower bounded by the Cramer-Rao inequality [63]. The lower

bound for the error variance is called CRLB, and is closely tied to the particular noise distribution assumed. We show that special cases of our results support Adiv's analysis of the ambiguities. In addition, the quantitative description of the inherent ambiguities by CRLB's enables us to obtain better insights on the effects of the aperture problem, smoothing, etc on 3-D motion estimation.

In our work, we are only able to consider motion estimation algorithms which give unbiased estimates of motion parameters. The performance bound is derived as a closed form and can be computed by inverting a 5×5 positive-definite matrix. It turns out that this bound is independent of the rotational motion parameters. For any motion and surface, the performance bound can always be computed and gives a quantitative description of the inherent ambiguities of the underlying 3-D motion. We show that the aperture problem involved in the computation of optical flow fields is so severe that the velocity components at the majority of the pixels contribute no information to the performance bound. Nontrivial information about motion parameters is contained only in a sparse set of pixels at which both x- and y- components of the optical flow velocity are observable. The points at each of which only the normal component to the constant gray level contour is available do not contribute information to the calculation of CRLB's. For the special case of a rigid planar surface, we derive explicit dependence of the performance bound on several factors, such as the range of the field of view, tilt and distance of the plane, the noise level in the images, the density of pixels, and the translational motion parameters. A subset of these results support the analysis in [4]. The analysis of the motion of a planar surface gives additional insights about the motion of an arbitrary surface. For nonplanar surfaces, analytical form of CRLB's are difficult to obtain. However, the numerical values of CRLB's can always be easily computed for an arbitrary motion in front of an arbitrary surface under any noise distribution. This enables us to investigate the inherent ambiguities of motion estimation in a nonplanar environment using numerical computations. Two generally applied schemes to impose smoothness are investigated. It is shown that smoothing the optical flow field by regularization methods does not reduce the CRLB's. But the smoothness constraint incorporated by fitting local patches to 3–D smooth surfaces [74]

decreases CRLB's as more information is introduced into the estimation process by reducing the degree of freedom of the space of free parameters. Surprisingly enough, CRLB's are only reduced by a very insignificant amount by fitting local patches. Another effect of fitting local patches is that the aperture problem is relaxed as the flow components normal to a constant brightness contour lower the CRLB's. This is in contrast to the previously mentioned case when no local patches are fitted, where the normal flow components alone contribute no information in the calculation of CRLB's.

1.5 Contributions

The main contributions resulting from the research reported in this dissertation are the formulation and analysis of new model based 3–D motion estimation methods and related ambiguity problems. These contributions include the following:

- A new kinematic model for 3–D high order motion estimation is defined and analyzed. Specifically:

    1. The kinematic model of constant acceleration and precession are formulated and applied to 3–D motion estimation from a sequence of stereo images with rotation being represented using quaternions.
    2. Uniqueness results of motion parameters from noiseless data are stated and proved.
    3. The behavior of the model in a noisy environment is investigated using the Cramer–Rao lower bound.

- A simple kinematic model with rotation described by the standard 3×3 rotation matrix is developed and applied to the motion of constant translational and angular velocity. Specific contributions in this topic are:

    1. Linear plant models for both monocular and binocular image sequences are formulated to improve the performance and reduce the

computational load over the nonlinear kinematic plant model previously reported in the literature.

2. Occlusion is formulated and simulated.
3. Both batch and recursive algorithms are given for a simpler and more intuitive representation of rotational motion than the previous kinematic models.

- The inherent ambiguities in recovering 3-D motion information from a single optical flow field are studied using a statistical model. Specific contributions include:

  1. The aperture problem in a noisy environment is formulated.
  2. The effectiveness of two generally applied smoothing techniques (smoothing over the optical flow and smoothing over the 3–D depth) are analyzed.
  3. An easily computed bound is used to extensively analyze the inherent ambiguities.

1.6 Organization of Dissertation

The remainder of this dissertation is organized as follows. In Chapter 2 we present a short review of the two basic solution approaches: feature based method and optical flow technique. The existing approaches dealing with inherent ambiguities of a single optical flow field are discussed. Chapter 3 presents a kinematic model of translational motion of constant acceleration and rotational motion of constant precession. The model is applied to the estimation of 3–D motion parameters on a sequence of noisy stereo images. The uniqueness of motion parameters in the noiseless environment is stated and proved. The estimator's efficiency in a noisy environment is investigated by the Cramer–Rao lower bound. Chapter 4 deals with the motion of constant translational and angular velocity extensively. We show that the standard 3×3 rotation matrix leads to a simple time propagation of motion parameters. Linear plant models are formulated for

the recursive algorithms based on monocular as well as binocular image sequences to avoid numerical integration which propagates the motion states and the covariance matrix in a nonlinear Kalman filter. Batch solutions are also given by a conjugate gradient search. Occlusion is considered. In Chapter 5, we present a statistical model to investigate the inherent ambiguities in recovering 3-D motion information from a single optical flow field. The aperture problem in a noisy environment is formulated. The effectiveness of applying smoothness constraint on the optical flow field and on the 3–D world depths is studied. Various conditions which lead to inherent ambiguities of 3–D motion parameters are analyzed. Finally, Chapter 6 summarizes this work and suggests possible directions for future research.

Chapter 2

Background

In this chapter, we review the basic formulations and various models of 3–D motion and structure estimation that serve as a basis for this dissertation. The problem of motion estimation and analysis is to relate the model parameters representing the 3–D geometry and dynamics with the acquired image measurements, then solve the 3–D parameters from the 2–D image measurements. Approaches to this problem are reviewed in this chapter. Standard terminology is used throughout the dissertation. In this chapter, due to space restrictions, we are not able to present an exhaustive review of recent research in motion estimation and analysis; instead we list some of the important work done and provide a brief summary of the approaches based on which this dissertation is developed. Excellent surveys of this research area include [5, 7].

2.1 Basic Approaches for Solution

Existing approaches for 3–D motion estimation can be divided into two groups depending upon the measurement data used. One is based on the optical flow field [3, 41, 66, 73], while the other on the correspondence of discrete features such as points, lines and contours [46, 49, 60, 67, 68, 69, 75].

2.1.1 Optical Flow Based Methods

The concept of optical flow stems from the observation that a changing 3–D world has its corresponding image changing with time on the retina or on the image focal plane of a video camera. The change of brightness pattern of the image may be considered as the variation of a 2–D dynamic field of light. One useful quantity to describe a dynamic field is its velocity at every point in the field. This velocity field for the dynamic imagery is called the optical flow field. Although the optical flow field describes the changes of brightness pattern on the image plane, it does not necessarily reflect the motion of the 3–D world. A trivial example to illustrate this fact is the absence of any brightness change on the image plane when a uniform background moves parallel to the image plane with the lighting conditions fixed. In this example, a zero optical flow field is generated by a non–zero 3–D motion. In spite of this inconsistency, researchers usually apply the simplifying assumption that the optical flow field is identical to the projected motion field of the 3–D velocity if the images have rich textures [37]. Based on this assumption, the equations which relate the 3–D motion parameters with the optical flow field may be formulated as [37]

$$\begin{aligned} u &= (-\frac{U}{Z} - B + Cy) - x(-\frac{W}{Z} - Ay + Bx) \\ v &= (-\frac{V}{Z} - Cx + A) - y(-\frac{W}{Z} - Ay + Bx) \end{aligned} \tag{2.1}$$

where (u, v) is the optical flow field at the image point (x, y) with the associated 3–D depth Z; (U, V, W) and (A, B, C) are 3–D translation and rotation parameters, respectively. Over the image region of a rigid object, the motion parameters (U, V, W) and (A, B, C) are constant. Scenes with multiple rigid objects are segmented, and each region is associated with the corresponding constant motion parameters. Then the 3–D motion estimation is tackled by a two stage approach. Stage one is the computation of the optical flow field (u, v) from the image intensity. Stage two is the computation of the 3–D motion parameters (U, V, W) and (A, B, C) from the optical flow field (u, v) using (2.1). It is noted that stage two is a standard nonlinear parameter estimation problem. Least squares is commonly applied to deal with the noise contaminating the optical flow field.

Iterative algorithms are designed to locate the minimum in the nonlinear least squares function. As numerous methods of parameter estimation and optimization are available, stage two is manageable by existing tools. Yet the uniqueness of motion parameters and the convergence of the minimization algorithms used are still not established completely. Part of the difficulty is due to the nontrivial nonlinearity caused by the perspective projection. The problems in stage one are, however, more involved.

In stage one, the assumption of constant brightness is commonly applied to summarize the complicated lighting and reflecting conditions in a simplified model. The image of a 3–D point is assumed to have the same gray level along its trajectory on the image plane. Suppose $E(x, y, t)$ is the gray level at the image point (x, y) at time t. The assumption of constant brightness asserts

$$\frac{dE}{dt} = \frac{\partial E}{\partial x}u + \frac{\partial E}{\partial y}v + \frac{\partial E}{\partial t} = 0 \tag{2.2}$$

At each image point, only one such equation is available for solving the flow field (u, v), which has two unknowns. This underdetermination is due to the aperture problem, which claims that the point correspondence is inherently ambiguous under local computations based on the image intensity; and only one optical flow component is observable at each image point. Since the computation of the optical flow field is equivalent to the establishment of local point correspondence, the underdetermination involved in (2.2) is predicted by the aperture problem. It is necessary to incorporate more information to solve the aperture problem. Biological experiments seem to show that the human visual system perceives coherent motion of the 3–D world [15, 59, 58, 79, 80]. The correspondence is established in such a way that the motion field is smooth over the retina. Usually the correspondence of this kind differs from the true point correspondence, but offers a simple scheme to qualitatively understand the 3–D motion from 2–D imagery. (Precise quantitative inference of the 3–D world from the coherent correspondence scheme seems difficult, if not impossible, to the human vision.) Suggested by this biological scheme, smoothness constraints are imposed on the optical flow field to solve the aperture problem using regularization methods. We emphasize that although a coherent flow field is obtained by this method, it is

not the "true" optical flow field that is needed as the input for stage two. This deviation restricts stage two from obtaining robust motion estimates.

Due to the difficulty in solving the point correspondences in stage one, single stage methods are later proposed. Only recently single stage methods of estimating 3–D motion from image intensities are investigated by researchers [34, 40, 54, 55]. These single stage methods are known as "direct methods." Substitution of (2.1) into (2.2) yields

$$\frac{\partial E}{\partial x}\left[(-\frac{U}{Z} - B + Cy) - x(-\frac{W}{Z} - Ay + Bx)\right]$$

$$+ \frac{\partial E}{\partial y}\left[(-\frac{V}{Z} - Cx + A) - y(-\frac{W}{Z} - Ay + Bx)\right] + \frac{\partial E}{\partial t} = 0 \qquad (2.3)$$

This equation relates the 3–D motion parameters directly with the image intensity $E(x, y, t)$ without invoking the computation of the intermediate optical flow field. Several direct methods based on (2.3) have been developed [34, 40, 54, 55]. Two frames as well as multi–frames are used as input data. More fruitful results in this direction are possible.

2.1.2 Feature Based Approaches

As discrete salient features, including points, lines, and isolated patches are easier to identify in image frames, feature based approaches estimate the 3–D motion via tracking of the discrete features. Similar to the optical flow based methods, there are two steps in the feature based approaches. Step one extracts features and establishes the feature correspondence among image frames. Step two estimates the 3–D motion parameters based on the image coordinates of discrete features.

Although several algorithms have been developed for feature point labelling and correspondence, step one is generally recognized as a difficult problem. Small time intervals between adjacent frames are preferred in first step because the location of a feature point on the image will have limited change so that the search for corresponding features may be restricted locally. Small time intervals nevertheless lead to noisy 3–D motion estimates in the second step. To stabilize the parameter estimation process and reduce computations, large time intervals

are commonly used. Consequently more "high level" image processing techniques are needed to establish point correspondence to cope up with possible occlusion and reappearance of features when the time intervals are relatively large and a long sequence of images is used. It is the author's opinion that the research in feature based approaches is slightly ahead of optical flow methods in stage one. As discussed in the previous subsection, the first step in optical flow based methods is to establish the local correspondence for every point on the image plane. The constant brightness assumption and the smoothness constraint have commonly been imposed to obtain local correspondence. In spite of the fact that this method exploits the advantage of small time interval, it gives only local point correspondence of a coherent motion field, which is different from the true point correspondence required for motion estimation. This explains why reliable quantitative descriptions are relatively difficult to obtain, while good qualitative understanding of 3–D motion is available from the optical flow field. From this point of view, one should be skeptical about the generally accepted statement that no correspondence is needed in the optical flow based methods.

Once feature correspondence has been established in the feature based approaches, 3–D motion parameters are estimated from the image coordinates of discrete features in the second step. Using the pin–hole central projection model, the equations linking the 3–D motion and structure parameters with the image measurements can then be formulated. The 3–D motion and structure estimation problem is then reduced to identifying the unknown 3–D parameters from the image measurements. In two view motion estimation, elegant work [68] has been done to obtain the exact solution of 3–D motion parameters from the above mentioned equations when the image coordinates of feature points are noise free. However, this exact approach is very sensitive to noise. To deal with noisy data, least squares algorithms have been developed. Again, some of these modified least squares methods use several steps to exactly locate the minimum of the criterion functions. The time consuming searches are thus avoided. Basically these two–frame based algorithms are restricted to data with low noise level. The number of feature points needed is large. To cope with these limitations, model

based motion algorithms are proposed to uniformly process a long sequence of images.

2.2 Model–Based Methods

Gennery [33] uses a recursive filter to estimate the 3–D motion of a known solid object. The motion model of constant velocity and angular velocity are applied to predict the object position. Quaternions are used to parameterize the rotational motion. Both monocular and stereo images are considered. In his pioneering work, since the structure of the object is assumed known, only motion parameters need to be estimated. Broida and Chellappa [20, 21, 18] are among the first to propose a unified framework to estimate the 3–D motion and structure from an arbitrarily long sequence of noisy images. In their approach, the time evolution of the 3–D object is approximated by a kinematic model which has a fixed number of motion parameters. The motion parameters are allowed to be time–varying and physically intuitive. Suppose $\underline{x}(t)$ denotes the collection of motion and structure parameters at time t. Since $\underline{x}(t)$ models only the motion and structure of the 3–D object, it is independent of the sensor applied to monitor the motion. This feature is desirable especially in a multi–sensor system as sensor modeling is separated from modeling 3–D motion and structure.

To fix the degree of freedom involved in the kinematic model, the time–varying behavior of 3–D motion and structure parameters is described by the following first order differential equation according to the motion model used.

$$\dot{\underline{x}}(t) = \underline{f}[\underline{x}(t)] \tag{2.4}$$

A property associated with the first order differential equation of this form is that whenever the value of $\underline{x}(t_0)$ at any time instant t_0 is known, all other $\underline{x}(t)$ are uniquely decided if $\underline{f}$ is a "well–behaved" function. Therefore the degree of freedom of $\underline{x}(t)$ is actually equal to that of $\underline{x}(t_0)$ and does change with time.

The central projection model of an ideal pin–hole camera is used to relate the observed image coordinate $\underline{z}$ with the 3–D motion and structure parameters $\underline{x}$,

$$\underline{z}(t_i) = \underline{h}[\underline{x}(t_i); t_i] + \underline{n}(t_i) \tag{2.5}$$

where $\underline{n}(t_i)$ is the random noise contaminating the image coordinates $\underline{z}(t_i)$ of feature points. Note that (2.4) and (2.5) are the standard plant and measurement equations one develops prior to application of an extended Kalman filter (EKF) or iterated extended Kalman filter (IEKF). As a consequence, these recursive filters become readily applicable to estimate the 3–D motion and structure parameters. The recursive estimation is natural in the context of object tracking as newly available data are processed right away to update estimated 3–D motion without waiting for the collection of all image data.

If a closed form solution of the motion model (2.4) is available, $\underline{x}(t)$ can be expressed as a function of $\underline{x}(t_0)$,

$$\underline{x}(t) = \underline{\Phi}[\underline{x}(t_0), t - t_0] \tag{2.6}$$

Substituting this into (2.5), we obtain

$$\underline{z}(t_i) = \underline{h}\,[\underline{\Phi}[\underline{x}(t_0), t_i - t_0]; t_i] + \underline{n}(t_i)$$

Thus the nonlinear least squares problem is formulated as

$$\min_{\underline{x}(t_0)} \sum_i \{\underline{z}(t_i) - \underline{h}\,[\underline{\Phi}[\underline{x}(t_0), t_i - t_0]; t_i]\}^2 \tag{2.7}$$

After the collection of all image measurements, the motion and structure parameters $\underline{x}(t_0)$ can be estimated by locating the minimum in (2.7) using a standard conjugate gradient search. This is known as the batch formulation. In general, batch formulation is more time–consuming than the recursive formulation. However, when no reasonable initial guess is available, the batch method is more stable than the recursive one. Therefore it is suggested that for the first several frames of images, a batch algorithm be used to obtain a good estimate of 3–D

motion and structure parameters. Then the estimate is fed into a recursive filter as the initial guess to track the moving object. In this sense, the batch and recursive formulations are considered to be cooperative, rather than competitive algorithms.

The statistical model (2.5) allows the Cramer–Rao performance bound to be computed. The uncertainty in the 3–D motion estimate may then be compared with the performance bound. It is shown that under most computer simulations examined, the performance bound approximates the actual estimation accuracy very well. In this dissertation, the prediction performance bound is extended to investigate the inherent ambiguities of the optical flow field.

2.3 Inherent Ambiguities of Recovering 3–D Motion from Two Frames of Motion Pictures

An image sequence is a series of time–varying 2–D data projected from the corresponding 3–D moving world. To recover the 3–D dynamics from an array of 2–D data, a unique solution is not automatically guaranteed even if noise–free image measurements are used. Besides, the ill–posedness in recovering 3–D motion from two–frame based methods in earlier researches suggests the need for a general analysis of inherent ambiguities. As the two–view image measurements, either optical flow field or feature point coordinates, are related to the 3–D motion and structure parameters by nonlinear equations, researchers usually apply the nonlinear least squares method to recover the unknown 3–D parameters. It was observed that even a small amount of noise led to a quite different solution [5, 6, 30, 42]. This gives rise to the question whether a given image measurement could have arisen from two deffernt motions with two corresponding object structures. Due to mathematical complexity, general answers to this question are available in the literature for only two–view based algorithms. In two–view motion analysis, imaging models using noise–free measurements as well as that using noisy measurements have been investigated. In the noise–free case, the

issue of inherent ambiguities addressed is whether a given set of image measurements could have arisen from two different motions with two corresponding surface structures, or equivalently whether uniqueness of motion parameters is guaranteed from image measurements [68, 38]. Hence in the noise–free imaging model, the problem of inherent ambiguities is also known as the uniqueness problem. Similarly, in the noisy case, the question studied is whether there exists a large set of significantly incorrect motion and structure parameters which induce image measurements similar to the correct ones so that the noise perturbed image measurements can not be used to distinguish the correct solution from the incorrect ones [4].

Tsai and Huang [68] gave the proof for uniqueness of 3–D motion parameters in feature based approaches when the image measurements are noise–free and the correspondence of feature points are established without error. They first present an algorithm which solves the exact values of the 3–D motion and structure parameters up to a global scale factor from the image measurements by solving a set of linear equations followed by performing a singular value decomposition. Then via checking the requirement for uniqueness in each step of the solution process, they conclude that seven point correspondence under certain geometry is sufficient, but not necessary, for uniqueness. A sufficient condition, again not necessary, for the geometry of the seven points to guarantee uniqueness is that the points should not be traversed by two planes with one plane containing the origin, nor by a cone containing the origin.

Horn [38] presents a uniqueness analysis for optical flow methods when there is no error in image measurements and point correspondence. He considers the condition when the two frames are arbitrarily close to each other temporally, which is a limiting case of the condition considered by Tsai and Huang, whose analysis is valid for any finite time interval between the two images. A sufficient, but not necessary, condition for uniqueness is that the 3–D points should not all lie on a hyperboloid of one sheet or its degenerate forms, nor is the viewer on the surface.

Adiv [4] analyzes the inherent ambiguities in a single noisy optical flow field without concerning the effect caused by the aperture problem. He extends his

detailed analysis for planar surfaces to derive several sufficient conditions which lead to the inherent ambiguities of the optical flow field. The simulation used to demonstrate his result is based on a least squares minimization. The effect of noise on the decomposition of the environment into independently moving objects was also analyzed.

The above extensive analyses of inherent ambiguities in various simplified models give insights to stability of two–view motion estimation. With noise–free image measurements, motion estimation based on two frames is hardly ambiguous. However, if the noise is present in the images, it is felt that a long time sequence of dynamic images becomes necessary for robust 3–D motion and structure estimation. Our approach to the ambiguity problem for a long sequence of images and for a single noisy flow field with the aperture problem considered may be found in Chapter 3 and 5, respectively, of this dissertation.

Chapter 3

3-D Motion Estimation Using a Sequence of Noisy Stereo Images: Models, Estimation, and Uniqueness Results

In this chapter, we discuss a kinematic model based approach for the estimation of 3-D motion parameters from a sequence of noisy stereo images. The approach is based on representing the constant acceleration translational motion and constant precession rotational motion in the form of a bilinear state space model using standard rectilinear states for translation and quaternions for rotation. Closed form solutions of the state transition equations are obtained to propagate the quaternions. The measurements are noisy perturbations of 3-D feature points represented in an inertial coordinate system. It is assumed that the 3-D feature points are extracted from the stereo images and matched over the frames. The structure of the moving object is estimated from the stereo image pairs prior to the estimation of the motion parameters. Owing to the nonlinearity in the state model, nonlinear filters are designed for the estimation of motion parameters. Simulation results are included. The Cramér-Rao performance bounds for motion parameter estimates are computed.

A constructive proof for the uniqueness of motion parameters is given for the motion model of constant acceleration and precession, and the measurement model of binocular vision. We show that with uniform sampling in time, three noncollinear feature points in five consecutive binocular image pairs contain all the spatial and temporal information. Both nondegenerate and degenerate motions are analyzed. A deterministic algorithm to recover motion parameters from a stereo image sequence is summarized from the constructive proof.

In Section 3.1, the kinematic motion models are introduced. Nonlinear recursive filters for the estimation of motion parameters are designed in Section 3.2. The filtering algorithm is summarized in Section 3.3. Section 3.4 gives the simulation results. A uniqueness proof for the motion parameters is given in Section 3.5. Section 3.6 summarizes a deterministic algorithm to recover motion. The CRLB's are calculated in Section 3.7. Section 3.8 applies the models to the estimation of lower order motions.

3.1 Models

Three models corresponding to observation noise, object structure, and motion are introduced [20, 21, 18]. These models allow the separation of structure and motion parameter estimation stages. The structure of the object is first estimated by a recursive estimator, which gives monotonically decreasing error variances in the number of frames. The structure estimates are then incorporated into observation noise and motion models, based on which an EKF/IEKF is designed for the estimation of motion parameters [18].

3.1.1 Observation Noise Model

It is well known that the 3-D coordinates of feature points can be decided from binocular images, e.g. by stereo triangulation. Therefore the calculated 3-D coordinates of feature points from the binocular images may be taken as the measurements. Define

- inertial coordinate system $I \equiv$ arbitrary 3-D world coordinate system, as shown in Figure 3.1.

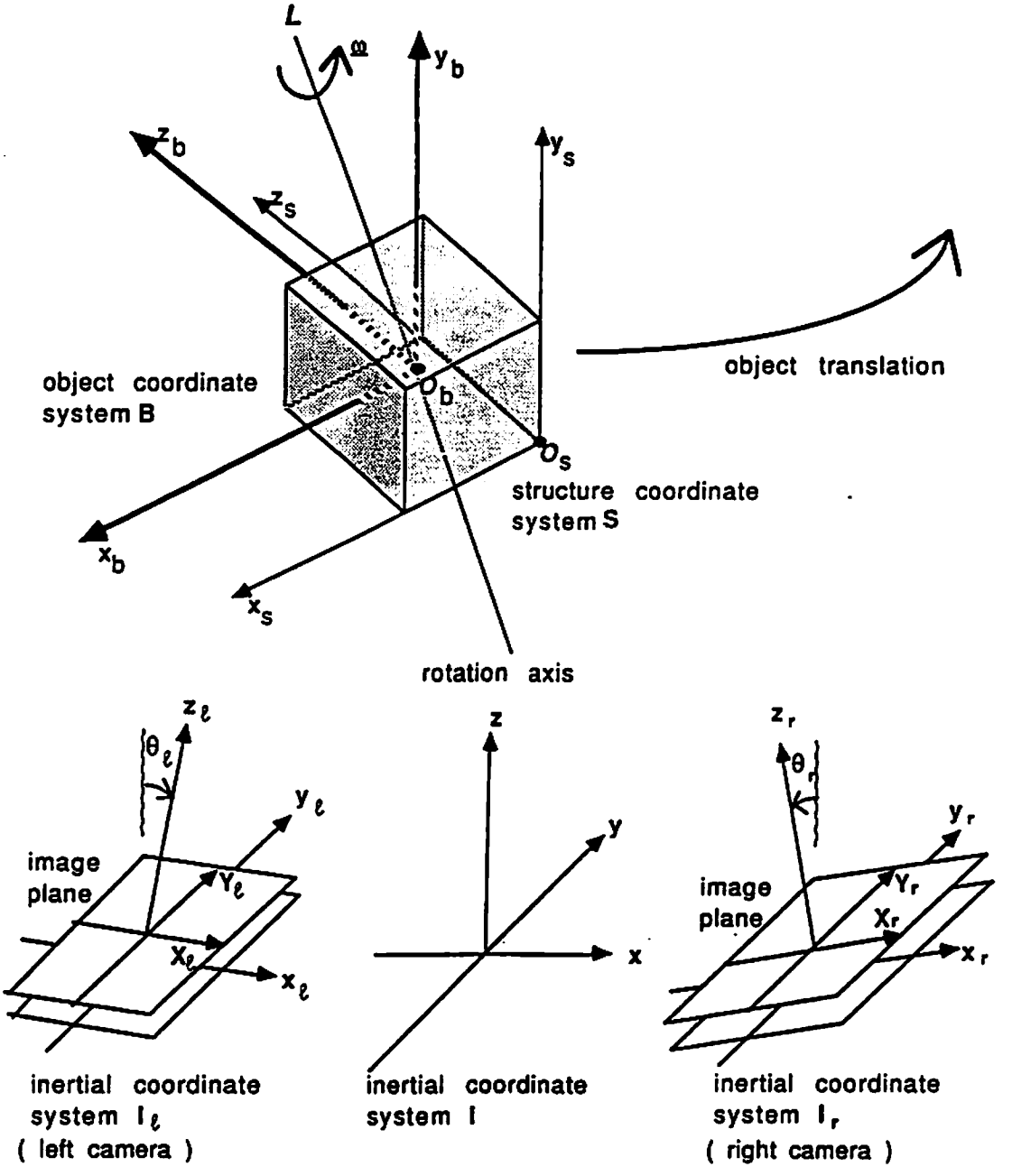


Figure 3.1: Geometry of motion and measurement models

The measurements are taken at discrete time t_i. Let

• $\underline{r}_j(t) \equiv (x_j(t), y_j(t), z_j(t))^T \equiv$ the true (noise-free) 3-D coordinates of the j^{th} feature point in I at time t;

• $\underline{r}_{nj}(t_i) \equiv (x_{nj}(t_i), y_{nj}(t_i), z_{nj}(t_i))^T \equiv$ the noisy 3-D coordinates of the j^{th} feature point in I calculated from binocular images at the measurement time t_i.

Then the observation noise model is

$$\underline{r}_{nj}(t_i) = \underline{r}_j(t_i) + \underline{n}_j(t_i) \tag{3.1}$$

where

• $\underline{n}_j(t_i) \equiv (n_{xj}(t_i), n_{yj}(t_i), n_{zj}(t_i))^T \equiv$ the noise vector, assumed to be Gaussian with zero mean.

3.1.2 Object Structure Model

In this section, the object model, which describes the structure of the rigid body by a certain structure coordinate system fixed on the rigid body, is defined. Then an admissible structure coordinate system is introduced. A recursive algorithm which estimates the structure of the rigid body is also given.

Let

• structure coordinate system $S \equiv$ any 3-D coordinate system which is fixed on the rigid body with the origin denoted by O_s;

• $\underline{s}_j \equiv (x_{sj}, y_{sj}, z_{sj})^T \equiv$ the 3-D coordinates of the j^{th} feature point in S.

These coordinates describe the structure of the object and are not a function of time because of the rigidity of the object.

To specify a structure coordinate system S, we need to define a point fixed on the rigid body to be the origin of S and three mutually orthonormal vectors fixed on the object to be the three directional vectors of the axes of S. Since the "center of mass,"

$$\underline{c}_M(t_i) \equiv \frac{1}{m}\sum_{j=1}^{m} \underline{r}_j(t_i) \tag{3.2}$$

where m is the number of feature points, is fixed on the rigid body, we may choose $\underline{c}_M$ as the origin O_s of S. The unit directional vectors along the axes of

S are chosen as

$$\begin{aligned} \underline{e}_x(t_i) &\equiv \frac{\underline{r}_2(t_i) - \underline{r}_1(t_i)}{|\underline{r}_2(t_i) - \underline{r}_1(t_i)|} \\ \underline{e}_y(t_i) &\equiv \frac{\underline{e}_x(t_i) \times [\underline{r}_3(t_i) - \underline{r}_1(t_i)]}{|\underline{e}_x(t_i) \times [\underline{r}_3(t_i) - \underline{r}_1(t_i)]|} \\ \underline{e}_z(t_i) &\equiv \underline{e}_x(t_i) \times \underline{e}_y(t_i) \end{aligned} \tag{3.3}$$

Here at least three noncollinear feature points are required. Note that $\underline{e}_x, \underline{e}_y$, and $\underline{e}_z$ are fixed on the object and form an orthonormal set. Moreover, the 3-D coordinates of the j^{th} feature point in S may be calculated as

$$\underline{s}_j = \begin{pmatrix} x_{sj} \\ y_{sj} \\ z_{sj} \end{pmatrix} = \begin{pmatrix} \underline{e}_x(t_i) \cdot \left[\underline{r}_j(t_i) - \underline{c}_M(t_i)\right] \\ \underline{e}_y(t_i) \cdot \left[\underline{r}_j(t_i) - \underline{c}_M(t_i)\right] \\ \underline{e}_z(t_i) \cdot \left[\underline{r}_j(t_i) - \underline{c}_M(t_i)\right] \end{pmatrix} \tag{3.4}$$

Since the true 3-D inertial coordinates $\underline{r}_j(t_i)$ are contaminated by noise and thus are unknown, we replace them by $\underline{r}_{nj}(t_i)$ to obtain an estimate for the center of mass,

$$\hat{\underline{c}}_M(t_i) \equiv \frac{1}{m} \sum_{j=1}^{m} \underline{r}_{nj}(t_i)$$

The same replacement in (3.3) gives $\hat{\underline{e}}_x, \hat{\underline{e}}_y$, and $\hat{\underline{e}}_z$ as the estimates of $\underline{e}_x, \underline{e}_y$, and $\underline{e}_z$, respectively. In addition, by noting that the 3-D structure coordinates $\underline{s}_j$ are not functions of time, using (3.4) and averaging the estimates from t_0 to t_i, we have the following estimates for $\underline{s}_j$ at t_i,

$$\hat{\underline{s}}_j(t_i) \equiv \begin{pmatrix} \hat{x}_{sj}(t_i) \\ \hat{y}_{sj}(t_i) \\ \hat{z}_{sj}(t_i) \end{pmatrix} \equiv \frac{1}{i+1} \sum_{k=0}^{i} \begin{pmatrix} \hat{\underline{e}}_x(t_k) \cdot \left[\underline{r}_{nj}(t_k) - \hat{\underline{c}}_M(t_k)\right] \\ \hat{\underline{e}}_y(t_k) \cdot \left[\underline{r}_{nj}(t_k) - \hat{\underline{c}}_M(t_k)\right] \\ \hat{\underline{e}}_z(t_k) \cdot \left[\underline{r}_{nj}(t_k) - \hat{\underline{c}}_M(t_k)\right] \end{pmatrix} \tag{3.5}$$

$$= \frac{i}{i+1} \begin{pmatrix} \hat{x}_{sj}(t_{i-1}) \\ \hat{y}_{sj}(t_{i-1}) \\ \hat{z}_{sj}(t_{i-1}) \end{pmatrix} + \frac{1}{i+1} \begin{pmatrix} \hat{\underline{e}}_x(t_i) \cdot \left[\underline{r}_{nj}(t_i) - \hat{\underline{c}}_M(t_i)\right] \\ \hat{\underline{e}}_y(t_i) \cdot \left[\underline{r}_{nj}(t_i) - \hat{\underline{c}}_M(t_i)\right] \\ \hat{\underline{e}}_z(t_i) \cdot \left[\underline{r}_{nj}(t_i) - \hat{\underline{c}}_M(t_i)\right] \end{pmatrix}$$

in which the third equality is the recursive form of the second one.

3.1.3 Motion Model

We assume that the underlying motion can be decomposed into a translation of constant acceleration and a rotation of constant precession [76]. The rotation center is denoted by O_b and assumed to be fixed with respect to the object. Let

- object coordinate system $B \equiv$ the 3-D coordinate system fixed on the rigid body with the origin O_b, and with its coordinate axes oriented in the same direction as the coordinate axes of S;
- $\underline{b}_j \equiv (x_{bj}, y_{bj}, z_{bj})^T \equiv$ the 3-D coordinates of the j^{th} feature point in B;
- $\underline{d} = (d_x, d_y, d_z)^T \equiv$ the 3-D coordinates of O_b in S;
- $\underline{r}(t) \equiv (x_R(t), y_R(t), z_R(t))^T \equiv$ the 3-D coordinates of O_b in I at t.

It is clear that the transformation between B and S is

$$\underline{b}_j = \underline{s}_j - \underline{d} \tag{3.6}$$

This is due to the fact that B and S have mutually parallel axes. Let $R(t)$ be the 3×3 alignment matrix that aligns the object coordinate system B with the inertial coordinate system I. Thus $R(t)$ is a rotation matrix and the transformation between I and B is

$$\underline{r}_j(t_i) = \underline{r}(t_i) + R(t_i)\underline{b}_j \tag{3.7}$$

Substitution of (3.6) into (3.7) yields the object motion model

$$\underline{r}_j(t_i) = \underline{r}(t_i) + R(t_i)(\underline{s}_j - \underline{d}) \tag{3.8}$$

where d_x, d_y, and d_z are chosen to be the three motion parameters which describe the center of rotation.

Translational Motion States

Let

- $\underline{v}(t) = (v_x(t), v_y(t), v_z(t))^T \equiv$ the velocity vector of O_b w.r.t. I and represented in I at time t;
- $\underline{a} = (a_x, a_y, a_z)^T \equiv$ the acceleration vector of O_b w.r.t. I and represented in I.

Under the assumption of constant acceleration, we have

$$\begin{cases} \dot{\underline{r}}(t) &= \underline{v}(t) \\ \dot{\underline{v}}(t) &= \underline{a}(t) \\ \dot{\underline{a}}(t) &= 0 \end{cases} \tag{3.9}$$

By direct integration, the closed form solution of (3.9) is obtained as

$$\begin{cases} \underline{r}(t_2) &= \underline{r}(t_1) + (t_2 - t_1)\underline{v}(t_1) + (t_2 - t_1)^2\underline{a}(t_1)/2 \\ \underline{v}(t_2) &= \underline{v}(t_1) + (t_2 - t_1)\underline{a}(t_1) \\ \underline{a}(t_2) &= \underline{a}(t_1) \end{cases} \tag{3.10}$$

where $x_R(t), y_R(t), z_R(t), v_x(t), v_y(t), v_z(t), a_x, a_y$, and a_z are chosen to be the 9 motion states describing the translational part of the underlying motion.

Rotational Motion States

There are about eight commonly used alternatives to represent rotation [37]. The quaternion [21, 18, 14, 32, 37, 56, 78] is one of them. Let

- $\underline{\omega}(t) = (\omega_x(t), \omega_y(t), \omega_z(t))^T \equiv$ the instantaneous angular velocity of the rigid body w.r.t. O_b with components represented in I;
- $L \equiv$ the instantaneous rotation axis on which $\underline{\omega}(t)$ lies.

The quaternion

$$\underline{q}(t) = (q_1(t), q_2(t), q_3(t), q_4(t))^T$$

is used to describe the orientation of the rigid body. As the orientation changes in time according to the angular velocity $\underline{w}(t)$, the dependence of $\underline{q}(t)$ on $\underline{w}(t)$ is [21, 18]

$$\dot{\underline{q}}(t) = \Omega\left[\underline{\omega}(t)\right] \underline{q}(t) \tag{3.11}$$

where

$$\Omega\left[\underline{\omega}(t)\right] = \frac{1}{2}\begin{pmatrix} 0 & -\omega_z(t) & \omega_y(t) & -\omega_x(t) \\ \omega_z(t) & 0 & -\omega_x(t) & -\omega_y(t) \\ -\omega_y(t) & \omega_x(t) & 0 & -\omega_z(t) \\ \omega_x(t) & \omega_y(t) & \omega_z(t) & 0 \end{pmatrix} \tag{3.12}$$

Since the orientation of the object decides how the object coordinate system B should be aligned to the inertial coordinate system I, $\underline{q}(t)$ is related to the alignment matrix $R(t)$ in (3.7). The relation between $\underline{q}$ and the "standard" rotation parameters $(n_1, n_2, n_3, \theta)^T$ of R is

$$\underline{q} = \left(n_1 \sin\frac{-\theta}{2}, n_2 \sin\frac{-\theta}{2}, n_3 \sin\frac{-\theta}{2}, \cos\frac{-\theta}{2}\right)^T$$

where $(n_1, n_2, n_3)^T$ is the unit directional vector of the rotation axis of the alignment matrix R, and θ is the rotation angle of R. The standard rotation parameters are not used as it is not easy to obtain a closed form expression for time dependence. Another equality relating $R(t)$ and $\underline{q}(t)$, after suppressing the time dependence, is [21, 18]

$$R[\underline{q}] = \begin{pmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1q_2 + q_3q_4) & 2(q_1q_3 - q_2q_4) \\ 2(q_1q_2 - q_3q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2q_3 + q_1q_4) \\ 2(q_1q_3 + q_2q_4) & 2(q_2q_3 - q_1q_4) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{pmatrix} \quad (3.13)$$

For the constant precession case, the direction of L is not fixed in the inertial coordinate system. As shown in Figure 3.2, L and the angular velocity vector $\underline{\omega}(t)$ are assumed to be rotating with a constant angular velocity $\underline{p}$ [76]. Define

- precession vector $\underline{p} = (p_x, p_y, p_z)^T \equiv$ the angular velocity of $\underline{\omega}(t)$ (or L) w.r.t. O_b and with components represented in I; Under the assumption of constant precession, $\underline{p}$ is a constant vector;
- precession axis $L_p \equiv$ the axis on which O_b and the precession vector $\underline{p}$ lie; Under the assumption of constant precession, the direction of L_p does not change with time.

Since the angular velocity vector $\underline{\omega}(t)$ rotates with the constant velocity $\underline{p}$, we have

$$\dot{\underline{\omega}}(t) = P(\underline{p})\underline{\omega}(t) \quad (3.14)$$

where

$$P(\underline{p}) = \begin{pmatrix} 0 & -p_z & p_y \\ p_z & 0 & -p_x \\ -p_y & p_x & 0 \end{pmatrix} \quad (3.15)$$

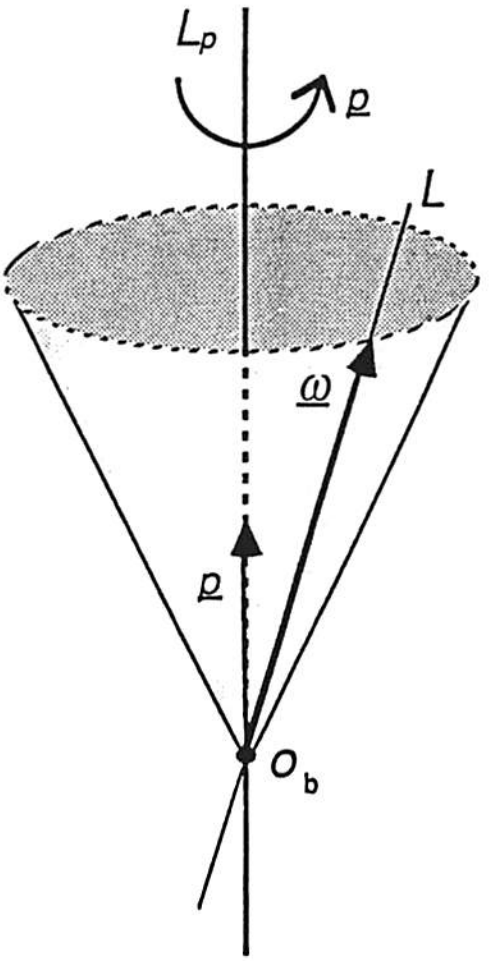


Figure 3.2: Geometry of the constant precession model.

The closed form solution of (3.14) is [27]

$$\underline{\omega}(t_2) = \exp[(t_2 - t_1)P(\underline{p})]\underline{\omega}(t_1)$$

Using standard results from Linear System Theory [27], the above can be simplified to

$$\underline{\omega}(t_2) = \Phi_1[\underline{p}; t_2 - t_1]\underline{\omega}(t_1) \tag{3.16}$$

where

$$\Phi_1[\underline{p}; \tau] = \begin{cases} I_3 & \text{if } \underline{p} = \underline{0} \\ \begin{pmatrix} l_1^2 + (1 - l_1^2)c & l_1 l_2(1-c) - l_3 s & l_1 l_3(1-c) + l_2 s \\ l_1 l_2(1-c) + l_3 s & l_2^2 + (1 - l_2^2)c & l_2 l_3(1-c) - l_1 s \\ l_1 l_3(1-c) - l_2 s & l_2 l_3(1-c) + l_1 s & l_3^2 + (1 - l_3^2)c \end{pmatrix} & \text{if } \underline{p} \neq \underline{0} \end{cases} \tag{3.17}$$

in which I_3 is the 3×3 identity matrix and

$$\begin{aligned} (l_1, l_2, l_3)^T &= (\frac{p_x}{|\underline{p}|}, \frac{p_y}{|\underline{p}|}, \frac{p_z}{|\underline{p}|})^T \\ c &= \cos(|\underline{p}|\tau) \\ s &= \sin(|\underline{p}|\tau) \end{aligned}$$

From (3.11), the closed form solution for $\underline{q}$ is [1] [76, 81, 82]

$$\underline{q}(t_2) = \Phi_2[\underline{\omega}(t_1), \underline{p}; t_2 - t_1]\underline{q}(t_1) \tag{3.18}$$

where

$$\Phi_2[\underline{\omega}(t_1), \underline{p}; t_2 - t_1] = \Phi[\underline{p}; t_2 - t_1]\Phi[\underline{\omega}(t_1) - \underline{p}; t_2 - t_1]$$

in which

$$\Phi[\underline{\lambda}; \tau] \equiv \begin{cases} I_4 \cos\frac{|\underline{\lambda}|\tau}{2} + \frac{2}{|\underline{\lambda}|}\Omega[\underline{\lambda}] \sin\frac{|\underline{\lambda}|\tau}{2} & \text{if } \underline{\lambda} \neq \underline{0} \\ I_4 & \text{if } \underline{\lambda} = \underline{0} \end{cases} \tag{3.19}$$

and Ω is defined in (3.12).

[1] (3.18) can be equivalently stated as $\underline{q}(t_2)$ is the quaternion multiplication [37] of $\underline{q}(t_1)$, $\underline{\xi}(\underline{w}(t_1) - \underline{p}; T)$, and $\underline{\xi}(\underline{p}; T)$ if the quaternion function (3.45) is used.

$q_1(t), q_2(t), q_3(t), q_4(t), \omega_x(t), \omega_y(t), \omega_z(t), \underline{p}_x, \underline{p}_y$, and $\underline{p}_z$ are chosen to be the ten rotational states.

3.2 Recursive Filter Formulation

Recursive filters [20, 18, 43, 51] are used to estimate motion parameters from noisy measurements. Both EKF and IEKF are recursive estimators of the parameters in a nonlinear system. This feature is desired in tracking moving objects, since motion parameters can be updated using newly obtained measurements without waiting for the collection of all data. The parameters to be estimated are allowed to be time-varying and are denoted by a $n \times 1$ column vector $\underline{x}(t)$, called the state vector. The time-varying behavior of the states satisfies the *plant model*

$$\dot{\underline{x}}(t) = \underline{f}\,[\underline{x}(t), \underline{u}(t); t] + G(t)\underline{w}(t) \tag{3.20}$$

where $\underline{f}$ can be a nonlinear function of $\underline{x}(t)$, $\underline{u}(t)$, and t; $\underline{u}(t)$ is the deterministic input function and $\underline{w}(t)$ a zero-mean white Gaussian noise process with covariance kernel

$$E\{\underline{w}(t)\underline{w}^T(t+\tau)\} = Q(t)\delta(\tau)$$

The measurements $\underline{z}$ at time t_i are required to depend on the states according to the *measurement model*

$$\underline{z}(t_i) = \underline{h}\,[\underline{x}(t_i); t_i] + \underline{v}_n(t_i) \tag{3.21}$$

where $\underline{h}$, again, can be a nonlinear function of $\underline{x}(t_i)$ and t_i; $\underline{v}_n$ is a white Gaussian noise sequence of mean zero and covariance kernel

$$E\{\underline{v}_n(t_i)\underline{v}_n^T(t_j)\} = \begin{cases} R_v(t_i) & , t_i = t_j \\ 0 & , t_i \neq t_j \end{cases} \tag{3.22}$$

Once plant and measurement models are specified, the EKF updates the estimate $\hat{\underline{x}}$ for the true state vector $\underline{x}$ by a series of equations given below [51]. Since the nonlinearity in $\underline{f}$ and $\underline{h}$ is linearized by first order Taylor series expansions, the

following matrix definitions are needed,

$$F \equiv \frac{\partial \underline{f}}{\partial \underline{x}} \tag{3.23}$$

$$H \equiv \frac{\partial \underline{h}}{\partial \underline{x}} \tag{3.24}$$

The estimate $\hat{\underline{x}}(t_i^+)$ for $\underline{x}(t_i^+)$ immediately after the measurements at t_i is obtained by

$$\hat{\underline{x}}(t_i^+) = \hat{\underline{x}}(t_i^-) + K(t_i)\{\underline{z}(t_i) - \underline{h}[\hat{\underline{x}}(t_i^-); t_i]\} \tag{3.25}$$

where the gain matrix is

$$K(t_i) = P(t_i^-)H^T[\hat{\underline{x}}(t_i^-); t_i]\{H[\hat{\underline{x}}(t_i^-); t_i]P(t_i^-)H^T[\hat{\underline{x}}(t_i^-); t_i] + R_v(t_i)\}^{-1} \tag{3.26}$$

The approximate covariance matrix is updated by

$$P(t_i^+) = P(t_i^-) - K(t_i)H^T[\hat{\underline{x}}(t_i^-); t_i]P(t_i^-) \tag{3.27}$$

Then the state estimate and the approximate covariance matrix are propagated from t_i^+ ($\hat{\underline{x}}(t_i^+)$ and $P(t_i^+)$) forward to the next sample time t_{i+1}^- ($\hat{\underline{x}}(t_{i+1}^-)$ and $P(t_{i+1}^-)$) according to the following differential equations

$$\dot{\hat{\underline{x}}}(t) = \underline{f}[\hat{\underline{x}}(t), \underline{u}(t); t] \tag{3.28}$$

$$\dot{P}(t) = F[\hat{\underline{x}}(t); t]P(t) + P(t)F^T[\hat{\underline{x}}(t); t] + G(t)Q(t)G^T(t) \tag{3.29}$$

The initial conditions used to solve these equations are $\hat{\underline{x}}(t_i^+)$ and $P(t_i^+)$. If (3.28) has a closed form solution, called the *state transition equation*, the computation is simplified as no numerical integration is needed to propagate $\hat{\underline{x}}(t_i^+)$ to $\hat{\underline{x}}(t_{i+1}^-)$.

If the nonlinearity in $\underline{h}$ is not satisfactorily linearized by EKF, improvement may be obtained by IEKF [51]. In IEKF, (3.25) and (3.26) are replaced by setting $\hat{\underline{x}}_0$ equal to $\hat{\underline{x}}(t_i^-)$ and doing iterations on

$$K(t_i) = P(t_i^-)H^T[\hat{\underline{x}}_k; t_i]\{H[\hat{\underline{x}}_k; t_i]P(t_i^-)H^T[\hat{\underline{x}}_k; t_i] + R_v(t_i)\}^{-1} \tag{3.30}$$

$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}(t_i^-) + K(t_i)\{\underline{z}(t_i) - \underline{h}[\hat{\underline{x}}_k; t_i] - H[\hat{\underline{x}}_k; t_i][\hat{\underline{x}}(t_i^-) - \hat{\underline{x}}_k]\} \tag{3.31}$$

for $k = 0, 1, \ldots, N-1$ and then setting $\hat{\underline{x}}(t_i^+) = \hat{\underline{x}}_N$. The iterations are stopped when the improvement in $\hat{\underline{x}}_k$ is less than a preselected threshold. The covariance matrix in (3.27) is then updated based on $\hat{\underline{x}}_N$.

The standard plant equation (3.20), measurement equation (3.21), and state transition equation for our model are formulated in the following. These are readily applicable to both EKF and IEKF.

3.2.1 Plant Equation

Let $\underline{x}(t) \equiv (\underline{r}^T(t), \underline{v}^T(t), \underline{a}^T, \underline{d}^T, \underline{q}^T(t), \underline{\omega}^T(t), \underline{p}^T)^T$ be the state vector whose components are formed by the 22 motion states described previously. The plant equation for the recursive filter is obtained from (3.9), (3.11), and (3.14) as

$$\underbrace{\frac{d}{dt}\begin{pmatrix} \underline{r}(t) \\ \underline{v}(t) \\ \underline{a} \\ \underline{d} \\ \underline{q}(t) \\ \underline{\omega}(t) \\ \underline{p} \end{pmatrix}}_{\dot{\underline{x}}(t)} = \underbrace{\begin{pmatrix} O_3 & I_3 & O_3 & O_3 & O & O_3 & O_3 \\ O_3 & O_3 & I_3 & O_3 & O & O_3 & O_3 \\ O_3 & O_3 & O_3 & O_3 & O & O_3 & O_3 \\ O_3 & O_3 & O_3 & O_3 & O & O_3 & O_3 \\ O^T & O^T & O^T & O^T & \Omega[\underline{\omega}(t)] & O^T & O^T \\ O_3 & O_3 & O_3 & O_3 & O & P(\underline{p}) & O_3 \\ O_3 & O_3 & O_3 & O_3 & O & O_3 & O_3 \end{pmatrix} \begin{pmatrix} \underline{r}(t) \\ \underline{v}(t) \\ \underline{a} \\ \underline{d} \\ \underline{q}(t) \\ \underline{\omega}(t) \\ \underline{p} \end{pmatrix}}_{\underline{f}[\underline{x}(t)]} \tag{3.32}$$

where O_3 is the 3×3 zero matrix and O 3×4 zero matrix. Note that the plant equation (3.32) has been put in the standard form (3.20) with $\underline{u}(t)$ being a zero vector, $G(t)$ a zero matrix, and $\underline{f}$ a time-invariant function.

3.2.2 Measurement Equation

If the true structure parameters $\underline{s}_j$ are replaced by the estimated structure parameters $\hat{\underline{s}}_j(t_i)$ in (3.5), from (3.1) and (3.8), the measurement equation for

EKF/IEKF is obtained as

$$\underbrace{\begin{pmatrix} x_{n1}(t_i) \\ y_{n1}(t_i) \\ z_{n1}(t_i) \\ \vdots \\ x_{nm}(t_i) \\ y_{nm}(t_i) \\ z_{nm}(t_i) \end{pmatrix}}_{\underline{z}(t_i)} = \underbrace{\left\{ \begin{pmatrix} x_R(t_i) \\ y_R(t_i) \\ z_R(t_i) \\ \vdots \\ x_R(t_i) \\ y_R(t_i) \\ z_R(t_i) \end{pmatrix} + \begin{pmatrix} & & \\ & R[\underline{q}(t_i)] & \\ & & \\ & \vdots & \\ & & \\ & R[\underline{q}(t_i)] & \\ & & \end{pmatrix} \begin{pmatrix} \hat{x}_{s1}(t_i) - d_x \\ \hat{y}_{s1}(t_i) - d_y \\ \hat{z}_{s1}(t_i) - d_z \\ \vdots \\ \hat{x}_{sm}(t_i) - d_x \\ \hat{y}_{sm}(t_i) - d_y \\ \hat{z}_{sm}(t_i) - d_z \end{pmatrix} \right\}}_{\underline{h}[\underline{x}(t_i); t_i]} + \underbrace{\begin{pmatrix} n_{x1}(t_i) \\ n_{y1}(t_i) \\ n_{z1}(t_i) \\ \vdots \\ n_{xm}(t_i) \\ n_{ym}(t_i) \\ n_{zm}(t_i) \end{pmatrix}}_{\underline{v}_n(t_i)} \tag{3.33}$$

where m is the total number of feature points. We remark that (3.33) allows the time intervals between two consecutive measurements to be nonuniform.

The noise distribution has been discussed by other authors. Blostein and Huang [16] have analyzed the noise distributions in the 3-D coordinates obtained by stereo triangulation. They concluded that the noise levels in different coordinate directions (horizontal, vertical and range) are different. The measurement model (3.33) has the flexibility to account for different noise levels if the variance of the noise in each coordinate direction is assigned a different value in the measurement noise covariance matrix $R_v(t_i)$ in (3.22). Matthies and Shafer [50] report that using 3-D Gaussian distributions to model triangulation error can lead to good performance in motion analysis. For the purpose of this dissertation we assume that the noise variables are distributed as Gaussian.

3.2.3 State Transition Equation

The closed form of the state transition equation is obtained to propagate the state vector in the recursive filters. The computation to propagate the state is simplified since no numerical integration for (3.28) is needed. From (3.10), (3.16), and (3.18), the state transition equation, which propagates $\hat{\underline{x}}(t_i^+)$ forward

to $\hat{\underline{x}}(t_{i+1}^-)$, is

$$\underbrace{\begin{pmatrix} \underline{r}(t_{i+1}^-) \\ \underline{v}(t_{i+1}^-) \\ \underline{a}(t_{i+1}^-) \\ \underline{d}(t_{i+1}^-) \\ \underline{q}(t_{i+1}^-) \\ \underline{\omega}(t_{i+1}^-) \\ \underline{p}(t_{i+1}^-) \end{pmatrix}}_{\underline{x}(t_{i+1}^-)} = \tag{3.34}$$

$$\underbrace{\begin{pmatrix} I_3 & I_3\tau & I_3\tau^2/2 & O_3 & O & O_3 & O_3 \\ O_3 & I_3 & I_3\tau & O_3 & O & O_3 & O_3 \\ O_3 & O_3 & I_3 & O_3 & O & O_3 & O_3 \\ O_3 & O_3 & O_3 & I_3 & O & O_3 & O_3 \\ O^T & O^T & O^T & O^T & \Phi_2[\underline{\omega}(t_i^+), \underline{p}; \tau] & O^T & O^T \\ O_3 & O_3 & O_3 & O_3 & O & \Phi_1[\underline{p}; \tau] & O_3 \\ O_3 & O_3 & O_3 & O_3 & O & O_3 & I_3 \end{pmatrix}}_{\equiv \Phi_c(\underline{x}(t_i^+); t_{i+1}^- - t_i^+)} \underbrace{\begin{pmatrix} \underline{r}(t_i^+) \\ \underline{v}(t_i^+) \\ \underline{a}(t_i^+) \\ \underline{d}(t_i^+) \\ \underline{q}(t_i^+) \\ \underline{\omega}(t_i^+) \\ \underline{p}(t_i^+) \end{pmatrix}}_{\underline{x}(t_i^+)}$$

where $\tau \equiv t_{i+1}^- - t_i^+$.

3.3 Filtering

The standard filtering algorithm is as follows.

1. Since the plant model (3.20) is given by (3.32) and the measurement model (3.21) by (3.33), the linearized matrices F and H can be obtained from (3.23) and (3.24).

2. The initial guess of the state vector is assigned to $\hat{\underline{x}}(t_0^-)$. The initial guess of the approximate covariance matrix is assigned to $P(t_0^-)$.

3. Set $i = 0$.

4. At time t_i, the measurement $\underline{z}(t_i)$ is taken.

5. At time t_i^+, immediately after the measurement time t_i, the estimate $\hat{\underline{x}}(t_i^+)$ for the state vector is obtained by (3.25) and (3.26). The approximate covariance matrix $P(t_i^-)$ is updated to $P(t_i^+)$ using (3.27).

   If an IEKF is used, $\hat{\underline{x}}(t_i^+)$ is obtained by the iterations defined in (3.30) and (3.31), instead of (3.25) and (3.26).

6. At time t_{i+1}^-, immediately before the measurement time t_{i+1}, the estimate $\hat{\underline{x}}(t_{i+1}^-)$ is computed by propagating $\hat{\underline{x}}(t_i^+)$ using the state transition equation (3.34). $P(t_{i+1}^-)$ is obtained by solving (3.29) using numerical integration [51].

7. Increase i by 1, then go to step 4 to incorporate new measurements.

3.4 Simulation Results

Computer simulations are used to investigate the models. The object studied is a rigid transparent echelon block, which has a 5 cm × 10 cm upper surface, a 7 cm × 10 cm lower surface, and a height of 15 cm. The simulated cameras are 100 cm away from the object. The viewing lines from the two cameras to the object form an angle of 45°. Feature points are chosen from corners of the rigid body. The synthetic images shown in Figure 3.3 are generated by the following scheme. First, the motion of edges and vertices of the object in 3-D space is generated by a computer program. Then the edges and vertices are centrally projected onto image planes of the two simulated cameras. The image planes are squares with side length 3 cm and resolution 64 × 64. As shown in Figure 3.3, the echelon block covers less than one fourth of the area of the whole image in the stereo sequence. As in [76], the effects of different noise sources such as digitization noise, corner detector error, etc. are summarized into the digitization error in our simulation, in which lower resolution simulates higher noise level. Without loss of generality, the time period between the measurements is set to 1.

The precession vector is $\underline{p} = (0, 0, 0.4)^T$. Thus the cube rotates w.r.t. a vertical line by about 22.9° per frame. Simultaneously it also rotates w.r.t.

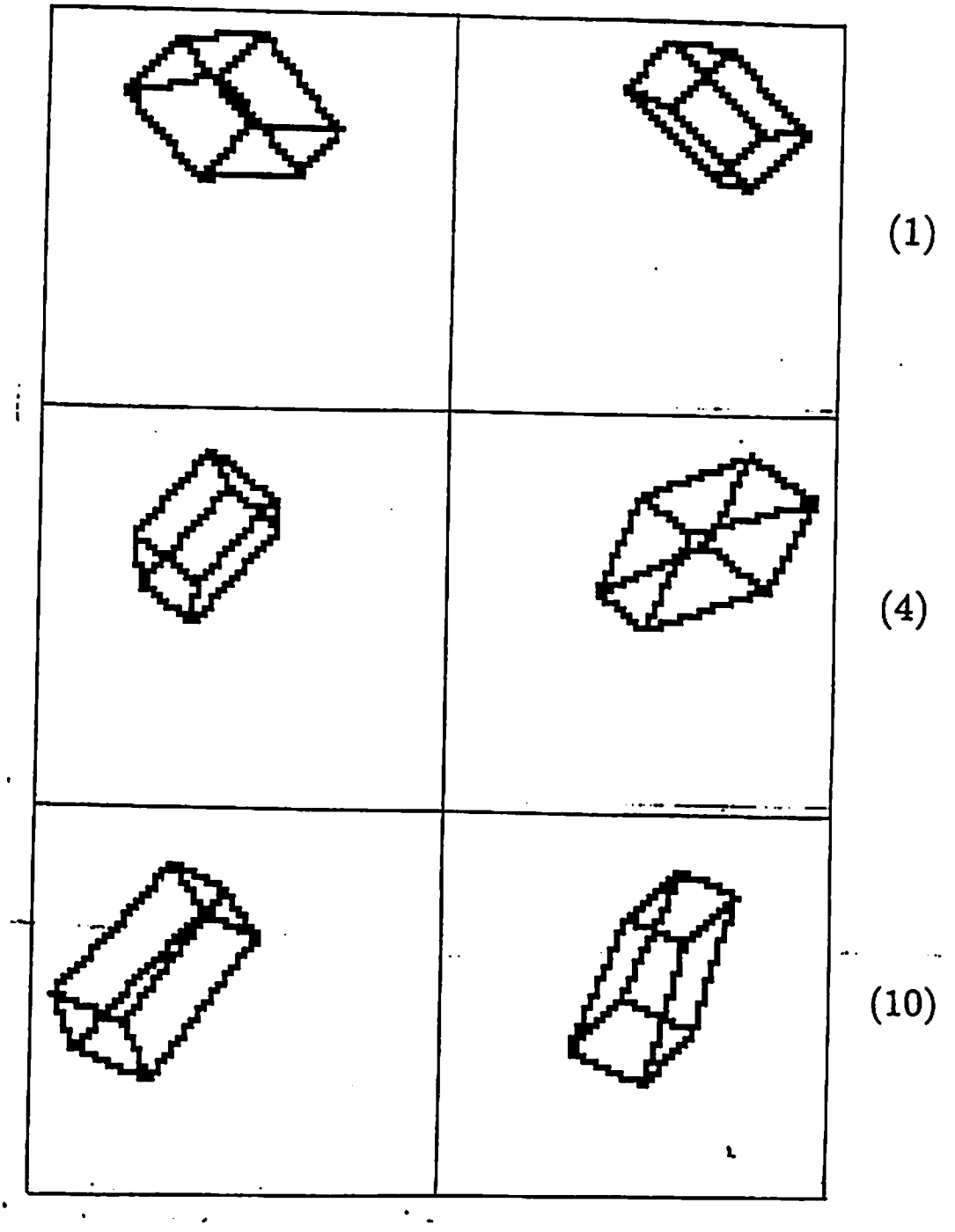


Figure 3.3: Synthetic stereo image frame sequence of a transparent echelon block. Motion model: constant acceleration and precession. Only the 1^{st}, 4^{th}, and 10^{th} stereo image pairs are shown. Left column: left view. Right column: right view. Image resolution: 64 × 64.

itself by about 8° per frame. The two-view rotation axis of the first two-view motion is $1/\sqrt{17}(1,0,4)^T$. The corresponding angular velocity is $\underline{w}(0) = (0.1144, -0.0232, 0.4747)^T$. The velocity is $\underline{v}(0) = (0.5, 0.5, 0.25)^T$ and acceleration $\underline{a} = (0.01, 0.01, 0.005)^T$. As a result of the camera set-up, the position vector $\underline{r}(0) = (-2.5, -10, 91.89)^T$ shows that the cameras are 100 cm away from the object. The other motion parameters are $\underline{q}(0) = (0.6, 0.3, 0.7, 0.2449)^T$ and $\underline{d} = (0.1, 0.1, 0.1)^T$. Only three feature points are used in our experiment.

For motion estimation, the noisy 3-D inertial coordinates of the three feature points are first computed from the digitized images by a least square algorithm. We briefly discuss this technique. The left camera is originally put at the origin of the inertial coordinate system I with its viewing line coincident with the z-axis of I. Then we move the left camera along the negative direction of the x-axis of I by the distance l. After this, the left camera is rotated w.r.t. the y-direction by the angle θ_l to have its viewing line toward the object tracked, as shown in Figure 3.1. Therefore we have

$$\begin{pmatrix} x_l \\ y_l \\ z_l \end{pmatrix} = \begin{pmatrix} \cos\theta_l & 0 & -\sin\theta_l \\ 0 & 1 & 0 \\ \sin\theta_l & 0 & \cos\theta_l \end{pmatrix} \begin{pmatrix} x+l \\ y \\ z \end{pmatrix} \tag{3.35}$$

where $(x, y, z)^T$ are the 3-D coordinates of the feature point in I, and $(x_l, y_l, z_l)^T$ are the 3-D coordinates of the feature point in the 3-D coordinate system I_l fixed on the left camera. Similarly, the right camera is moved from the origin of I along the positive direction of the x-axis of I by the distance l, then rotated w.r.t. the y-direction by the angle θ_r. Thus we have

$$\begin{pmatrix} x_r \\ y_r \\ z_r \end{pmatrix} = \begin{pmatrix} \cos\theta_r & 0 & -\sin\theta_r \\ 0 & 1 & 0 \\ \sin\theta_r & 0 & \cos\theta_r \end{pmatrix} \begin{pmatrix} x-l \\ y \\ z \end{pmatrix} \tag{3.36}$$

The central projection model is

$$\begin{cases} X_l = fx_l/z_l \\ Y_l = fy_l/z_l \end{cases} \qquad \begin{cases} X_r = fx_r/z_r \\ Y_r = fy_r/z_r \end{cases} \tag{3.37}$$

where f is the focal length and $(X_l, Y_l)^T$ the 2-D coordinates of the image point on the image plane of the left camera. Substitution of (3.35) and (3.36) into (3.37) yields

$$\underbrace{\begin{pmatrix} (\sin\theta_l)X_l/f - \cos\theta_l & 0 & (\cos\theta_l)X_l/f + \sin\theta_l \\ (\sin\theta_l)Y_l/f & -1 & (\cos\theta_l)Y_l/f \\ (\sin\theta_r)X_r/f - \cos\theta_r & 0 & (\cos\theta_r)X_r/f + \sin\theta_r \\ (\sin\theta_r)Y_r/f & -1 & (\cos\theta_r)Y_r/f \end{pmatrix}}_{\equiv H} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} -(\sin\theta_l)lX_l/f + l\cos\theta_l \\ -(\sin\theta_l)lY_l/f \\ (\sin\theta_r)lX_r/f - l\cos\theta_r \\ (\sin\theta_r)lY_r/f \end{pmatrix}}_{\equiv Z}$$

The least square solution of the 3-D inertial coordinates is

$$(x, y, z)^T = (H^T H)^{-1} H^T Z \tag{3.38}$$

The camera set-up parameters in our simulation are $l = 38.27$ cm, $\theta_l = 22.5°$, and $\theta_r = -22.5°$.

After the computation of 3-D inertial coordinates, the structure is estimated at each measurement time by the estimator in (3.5). Then an IEKF is applied to estimate the motion parameters. The initial guesses are all zeroes except that $q_4(0) = 1$, because the quaternion is of unit length.

A typical sample path of the IEKF is given in Figure 3.4, in which the ordinates are the differences between the estimated motion parameters and the ground truth. As can be seen in Figure 3.3, the object gradually moves outside the field of view. We simulate the field of view as image planes with infinite width after the fourteenth frames. Because the resolution is fixed, as the object moves away from the cameras and the object size shrinks, the noise level of the calculated 3-D inertial coordinates grows. Thus the measurement noise level after the fourteenth frame is higher than that before the fourteenth frame. As can

be seen in Figure 3.4, the estimates of translational parameters are noiser than rotational parameters. The motion parameters of higher order are less noisier, i.e. acceleration parameters are less noisier than velocity parameters, precession less noisier than angular velocity. The precession vector converges with good accuracy within eight frames. The error in the z-coordinate of $\underline{r}$ is reduced from -91.89 cm to -1.7 cm after the first frame is processed. For the sake of comparisons, the same ground truth and camera setup as in [76] have been used in our computer simulation with image resolution set to 64 $\times$ 64 in our case in contrast to 256 $\times$ 256 in their experiments.

When started from different initial guesses, the filters may converge to "wrong" motion parameters which are indistinguishable from the true parameters. This is the issue of uniqueness of motion parameters analyzed in Section 3.5. In the case of very highly deviant initial guesses, divergence is also observed, which is a well known feature for EKF and IEKF. A batch algorithm [18] can be used to start the filter with better initial guesses. Also the output of the deterministic algorithm obtained from the uniqueness analysis in Section 3.5 gives another alternative for the initial guesses.

3.5 Uniqueness Results for 3-D Motion Parameters

Theorem 6.1:

Consider the motion model defined by (3.32) and (3.33) in the absence of noise, i.e. $\underline{v}_n(t_i) = \underline{0}$. A uniform time sampling scheme is applied. That is, the stereo image pairs are allowed to be taken at $t_i \in M$ only, where

$$M \equiv \{t | t = nT, n \in Z\} \tag{3.39}$$

Assume stereo images at 5 different time instants are taken. Among these 5 image pairs, there exist 4 consecutive stereo image pairs such that 3 noncollinear feature points can be extracted and matched from each frame pair. In the remaining image pair, only one of the above feature points is used. Then the following hold.

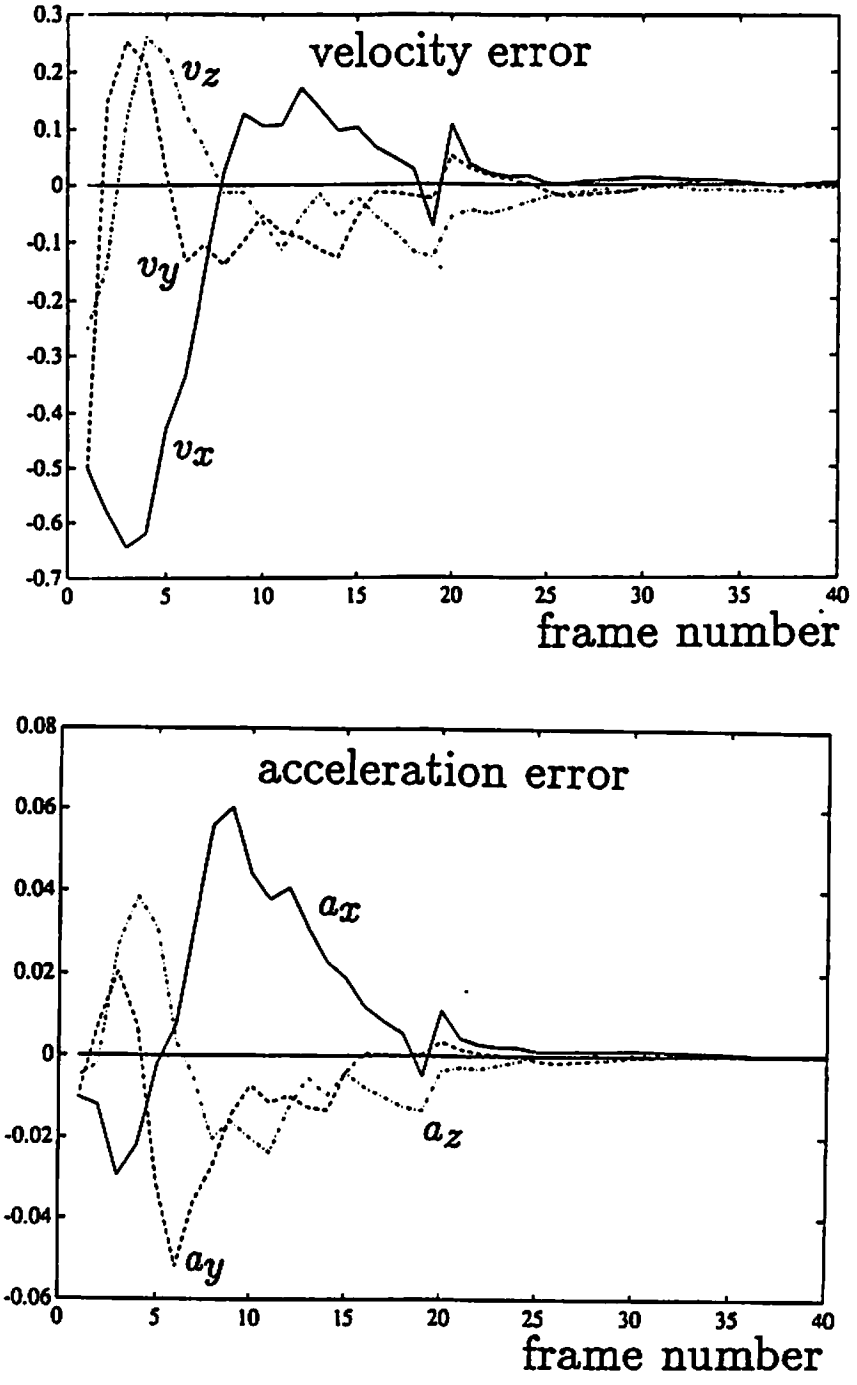


Figure 3.4: Errors in the estimates of motion parameters. A sample path of IEKF for the unknown structure, constant acceleration and precession motion model. (a) velocity error and (b) acceleration error. In Figure 3.4–8, the image resolution is 64 × 64 and 3 points/frame have been used.

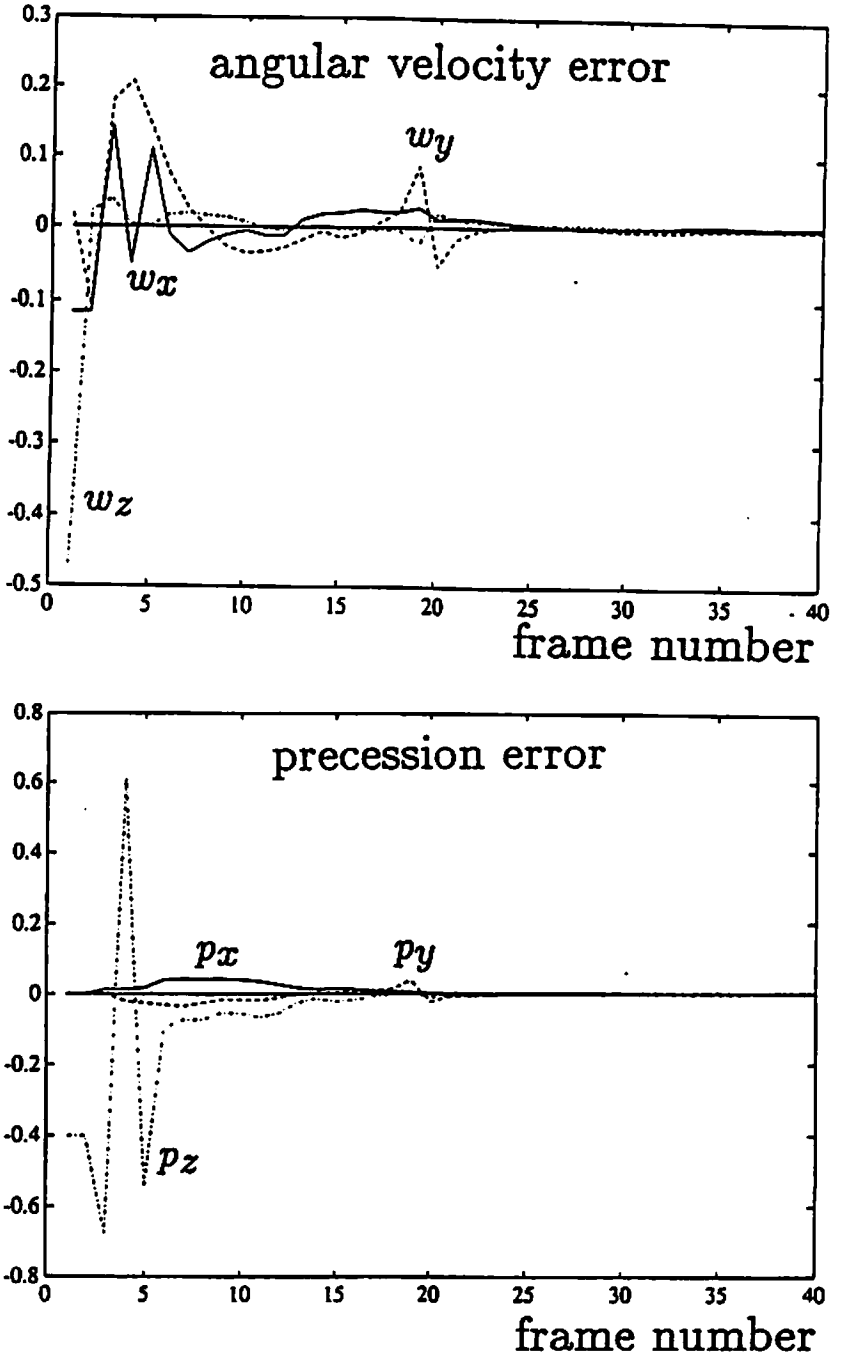


Figure 3.4: (continued) (c) angular velocity error and (d) precession error.

• The velocity $\underline{v}(t)$ and the acceleration $\underline{a}$ are unique. (proved in Section 3.5.4)

• The quaternions $\underline{q}(t_i)$, $t_i \in M$, are unique up to a sign ambiguity. (proved in Section 3.5.3)

• The solution sets of the rotation parameters $\underline{p}, \underline{w}(t)$, and $\underline{q}(t)$ have seven different configurations, only one of which is nondegenerate. (proved in Section 3.5.3)

• Let the interframe rotation refer to the rotation which aligns the orientations of the object between two consecutive measurement times. Then the above nondegenerate case for $\underline{p}, \underline{w}(t)$, and $\underline{q}(t)$ has the following properties. The interframe rotation is not any multiple of a half full turn. The direction of the interframe rotation axis is changing with time. The directions of $\underline{p}$ and $[\underline{w}(t_i) - \underline{p}]$ at any $t_i \in M$ are unique while their magnitudes can assume equally spaced positive discrete values, with spaces $2m\pi/T$, $m \in N$. The direction of $\underline{w}(t_i)$ at any $t_i \in M$ is not unique. If it is known a priori that both $\underline{p}$ and $[\underline{w}(t) - \underline{p}]$ have magnitudes less than $2\pi/T$, then both $\underline{p}$ and $\underline{w}(t)$ are unique, and $\underline{q}(t)$ is unique up to a sign ambiguity. (proved in Section 3.5.3, Case V)

• The solution sets of the position vector $\underline{r}(t)$ and rotation center $\underline{d}$ have three different configurations, only one of which is nondegenerate. (proved in Section 3.5.4)

• In the nondegenerate case for $\underline{r}(t)$ and $\underline{d}$, the interframe rotation axis is changing with time; the rotation center O_b is unique; $\underline{r}(t)$ and $\underline{d}$ are unique. (proved in Section 3.5.4, Condition I)

• In one of the degenerate cases for $\underline{r}(t)$ and $\underline{d}$, the object has the same orientation at all $t_i \in M$; the rotation is totally ambiguous (proved in Section 3.5.3, Case I); the rotation center O_b can be any point in the 3-D space; $\underline{d}$ can be any 3-D vector. (proved in Section 3.5.4, Condition II)

• In the other degenerate case for $\underline{r}(t)$ and $\underline{d}$, the interframe rotation axis is fixed; the rotation center O_b can be any point on the interframe rotation axis. (proved in Section 3.5.4, Condition III)

• The interframe rotation is described by the quaternion $\underline{\eta}(t_i)$, $t_i \in M$, [2] where

$$\underline{\eta}(t) \equiv \begin{pmatrix} \underline{\eta}_3(t) \\ \eta_4(t) \end{pmatrix} \equiv \begin{pmatrix} q_4(t) & q_3(t) & -q_2(t) & -q_1(t) \\ -q_3(t) & q_4(t) & q_1(t) & -q_2(t) \\ q_2(t) & -q_1(t) & q_4(t) & -q_3(t) \\ q_1(t) & q_2(t) & q_3(t) & q_4(t) \end{pmatrix} \underline{q}(t+T) \qquad (3.40)$$

Let us call $\underline{\eta}(t)$ the interframe quaternion. $\underline{\eta}(t_i)$, $t_i \in M$, are unique up to a sign ambiguity. (shown in Section 3.5.3)

• The configuration of a solution set of any motion parameters is uniquely decided by the interframe quaternion known up to a sign ambiguity at three consecutive measurement times. (proved in Section 3.5.3 and 3.5.4)

• If different structure coordinate systems S are used, the solution set of $\underline{r}(t), \underline{v}(t), \underline{a}, \underline{w}(t)$, and $\underline{p}$ is not changed; the quaternion is changed by a pre-multiplication of a 4×4 constant matrix which describes the alignment of the orientations of the new and the old structure coordinate systems; $\underline{d}$ is changed by an addition of a constant vector which describes the displacement between the new and the old S. (proved in Section 3.5.5)

The remainder of this section is devoted to prove Theorem 6.1.

3.5.1 Estimation of 3–D Structure

Without loss of generality, let the four consecutive measurement times in Theorem 6.1 be

$$M_4 \equiv \{t_0 = 0, t_1 = T, t_2 = 2T, t_3 = 3T\}$$

By the definition of M in (3.39), the five measurement times in Theorem 6.1 are

$$M_5 \equiv M_4 \cup \{t_4\}, \quad t_4 \in M \backslash M_4$$

[2] $\underline{\eta}(t)$ may be equivalently defined as the quaternion multiplication of the conjugate of $\underline{q}(t)$ and $\underline{q}(t+T)$.

Also define

$$M_3 \equiv \{t_0, t_1, t_2\}$$

Recall that the 3-D inertial coordinates of the j^{th} feature point ($\underline{r}_j(t_i)$, $t_i \in M_4$, and $\underline{r}_1(t_4)$) can be uniquely decided from stereo image pairs by (3.38). Then using three noncollinear feature points, the structure coordinates $\underline{s}_j$ of the j^{th} feature point in S are uniquely decided by (3.4) from $\underline{r}_j(t_0)$.

3.5.2 Preliminary Lemmas

In the following, Lemmas 3.5.1 and 3.5.2 allow the separation of the rotation parameters $\underline{p}, \underline{w}(t)$ and $\underline{q}(t)$ from the other motion parameters. Lemmas 3.5.3, 3.5.4, and 3.5.5 enable us to analyze the configurations of the solution set of rotation parameters. The proofs of these Lemmas may be found in [81, 83].

Lemma 3.5.1 At any time instant, if the 3-D coordinates $\underline{r}_j$, $j = 1, 2$, and 3, of three noncollinear feature points in the inertial coordinate system I are known, then the alignment matrix $R[\underline{q}]$ at that time instant is uniquely decided by

$$R[q] = [(\underline{r}_3 - \underline{r}_1) \;\; (\underline{r}_2 - \underline{r}_1) \;\; (\underline{r}_3 - \underline{r}_1) \times (\underline{r}_2 - \underline{r}_1)] \cdot$$

$$[(\underline{s}_3 - \underline{s}_1) \;\; (\underline{s}_2 - \underline{s}_1) \;\; (\underline{s}_3 - \underline{s}_1) \times (\underline{s}_2 - \underline{s}_1)]^{-1} \tag{3.41}$$

Lemma 3.5.2 At any time instant, if $R[\underline{q}]$ in (3.13) is unique, then $\underline{q}$ at that time instant is decided from $R[\underline{q}]$ up to a sign ambiguity.

The above Lemmas show that

$$\underline{q}(t_i), \;\; t_i \in M_4, \;\; \text{are unique up to a sign ambiguity.} \tag{3.42}$$

This with (3.40) yields that

$$\underline{\eta}(t_i), \;\; t_i \in M_3, \;\; \text{are unique up to a sign ambiguity.} \tag{3.43}$$

To emphasize the sign ambiguity, we may write

$$\underline{\eta}(t_i) = \begin{pmatrix} \underline{\eta}_3(t_i) \\ \eta_4(t_i) \end{pmatrix} = \pm \begin{pmatrix} \hat{\underline{\eta}}_3(t_i) \\ |\eta_4(t_i)| \end{pmatrix} \qquad t_i \in M_3 \tag{3.44}$$

That is, as there are two possible values of $\underline{\eta}(t_i)$ for each $t_i \in M_3$, $(\hat{\underline{\eta}}_3^T(t_i), |\eta_4(t_i)|)^T$ denotes the one which has a positive fourth component when $\eta_4(t_i) \neq 0$. For the case $\eta_4(t_i) = 0$, we arbitrarily pick up one the two possible values of $\underline{\eta}(t_i)$ and denote it by $(\hat{\underline{\eta}}_3^T(t_i), |\eta_4(t_i)|)^T$.

Lemma 3.5.3 Define

$$\underline{\psi}(t) = \underline{w}(t) - \underline{p}$$

Then the following scalars are independent of t.

$$\begin{array}{llll} |\underline{w}(t)|, & |\underline{\psi}(t)|, & \underline{w}(t) \cdot \underline{p}, & \underline{\psi}(t) \cdot \underline{p}, \\ \eta_4(t), & |\underline{\eta}_3(t)|, & |\underline{\eta}(t)| = 1, & \underline{\eta}_3(t) \cdot \underline{\eta}_3(t+\tau), \quad \underline{\eta}(t) \cdot \underline{\eta}(t+\tau). \end{array}$$

Lemma 3.5.4 Consider the quaternion function of a 3×1 vector $\underline{c}$ and scalar τ

$$\underline{\xi}(\underline{c};\tau) \equiv \begin{cases} (0,0,0,1)^T & \text{, if } \underline{c} = (0,0,0)^T \\ \left(\dfrac{\underline{c}^T}{|\underline{c}|} \sin \dfrac{-|\underline{c}|\tau}{2}, \cos \dfrac{-|\underline{c}|\tau}{2}\right)^T & \text{, otherwise} \end{cases} \tag{3.45}$$

If $\underline{\xi}$ is known up to a sign ambiguity as

$$\underline{\xi}(\underline{c};\tau) = \pm(\hat{\underline{\xi}}_3^T, |\xi_4|)^T$$

then the solutions of $\underline{c}$ are

$$\underline{c} = \begin{cases} (2m-1)\pi \hat{\underline{\xi}}_3^T/\tau & \text{,if } \xi_4 = 0 \\ 2m\pi \underline{u}/\tau & \text{,if } \xi_4 = \pm 1 \\ \left(\dfrac{2m\pi}{\tau} - \dfrac{2}{\tau} \tan^{-1} \dfrac{\hat{\xi}_3}{|\xi_4|}\right) \dfrac{\hat{\xi}_3}{|\hat{\xi}_3|} & \text{,otherwise} \end{cases} \qquad m \in Z$$

where $\underline{u}$ is any arbitrary 3×1 unit vector ($|\underline{u}| = 1$).

Lemma 3.5.5 The dynamical equation of $\underline{\eta}(t)$ is

$$\frac{d}{dt}\underline{\eta}(t) = \begin{pmatrix} P(\underline{p}) & \underline{0} \\ \underline{0}^T & 0 \end{pmatrix} \underline{\eta}(t) \tag{3.46}$$

where $P(\underline{p})$ is defined in (3.15) and $\underline{0} = (0,0,0)^T$. The solution of (3.46) is the state transition equation of $\underline{\eta}(t)$,

$$\underline{\eta}(t+\tau) = \begin{pmatrix} \Phi_1[\underline{p};\tau] & \underline{0} \\ \underline{0}^T & 1 \end{pmatrix} \underline{\eta}(t) \tag{3.47}$$

where Φ_1 is defined in (3.17), and

$$\Phi_1[\underline{p};\tau] = \exp[\tau P(\underline{p})] = R[\underline{\xi}(\underline{p};\tau)]$$

Another form of $\underline{\eta}(t)$ is

$$\underline{\eta}(t) = \tag{3.48}$$

$$\begin{cases} \underline{\xi}(\underline{w}(t);T) = \underline{\xi}(\underline{w}(t_0);T) & \text{,if } |\underline{p}| = 0 \text{ or } |\underline{\psi}(t)| = 0 \\ \begin{pmatrix} c_\psi s_p \dfrac{\underline{p}}{|\underline{p}|} + c_p s_\psi \dfrac{\underline{\psi}(t)}{|\underline{\psi}(t)|} + s_p s_\psi \dfrac{\underline{\psi}(t) \times \underline{p}}{|\underline{\psi}(t)|\,|\underline{p}|} \\ c_\psi c_p - s_p s_\psi \dfrac{\underline{\psi}(t) \cdot \underline{p}}{|\underline{\psi}(t)|\,|\underline{p}|} \end{pmatrix} & \text{,otherwise} \end{cases}$$

where

$$\begin{cases} c_p = \cos(-|\underline{p}|T/2) \\ s_p = \sin(-|\underline{p}|T/2) \end{cases} \quad \text{and} \quad \begin{cases} c_\psi = \cos(-|\underline{\psi}(t)|T/2) \\ s_\psi = \sin(-|\underline{\psi}(t)|T/2) \end{cases}$$

in which c_ψ and s_ψ do not change with t by Lemma 3.5.3.

3.5.3 Uniqueness of Rotation Parameters

In the following, we first show the solutions of these rotation parameters and their dependence on the interframe quaternion. Then the proof follows the listed solution set.

Case I

$$\exists t_i \in M_3,\ \underline{\eta}(t_i) = \pm(0,0,0,1)^T \iff \forall t,\ \underline{\eta}(t) = \pm(0,0,0,1)^T \qquad (3.49)$$

.49

solution set $S_1 \cup S_2 \cup S_3$

$$S_1 : \begin{cases} |\underline{p}| = 2m\pi/T & ,m \in N \\ |\underline{\omega}(t_0) - \underline{p}| = 2n\pi/T & ,n \in N \end{cases}$$

$$S_2 : \begin{cases} \underline{w}(t) = \underline{0} & ,t \in R \\ \underline{p} = \text{arbitrary} \end{cases}$$

$$S_3 : \begin{cases} |\underline{w}(t_0)| = 2m\pi/T & ,m \in N \\ \underline{p} = \alpha\underline{w}(t_0) & ,\alpha \in R \end{cases}$$

Remarks.

1. We use the above equations to express all solutions of $\underline{p}$ and $\underline{\omega}(t_0)$, instead of parameterizing them. For instance, the first equation in S_1 means that $\underline{p}$ can be any vector with lengths $2m\pi/T$, $m \in N$; the second shows that $\underline{\omega}(t_0) - \underline{p}$ can be any vector with lengths $2n\pi/T$, $m \in N$. This notation will be used to represent the solution sets.

2. Why do we solve for $\underline{p}$ and $\underline{\omega}(t_0)$ only? Recall that $\underline{q}(t_0)$ has been known up to a sign ambiguity by (3.42). Then the parameters of interest, $\underline{p}, \underline{\omega}(t)$, and $\underline{q}(t)$, can be obtained from $\underline{p}, \underline{\omega}(t_0)$, and $\underline{q}(t_0)$ using (3.18) and (3.16) for all t.

Proof for Case I:

To prove the solution set for this case, we consider three subcases.

Subcase I. $\underline{p} \neq \underline{0} \quad \wedge \quad \forall \alpha \in R,\ \underline{\omega}(t) \neq \alpha\underline{p}$

$\{\underline{p}, \underline{\omega}(t), \underline{\omega}(t) \times \underline{p}\}$ is a linearly independent set, and so is $\{\underline{p}, \underline{\psi}(t), \underline{\psi}(t) \times \underline{p}\}$. Since both $\underline{p}$ and $\underline{\psi}(t)$ are nonzero, (3.48) and (3) give

$$0 = c_\psi s_p = c_p s_\psi = s_\psi s_p$$

or equivalently

$$\sin(-|\underline{\psi}(t)|T/2) = \sin(-|\underline{p}|T/2) = 0$$

Hence we arrive at a subset of S_1. By backsubstitution, it is straightforward to show that every element in S_1 is a solution. In the following, whenever we reach

a subset of a solution set S_i claimed, backsubstitution will always be used to verify that all motion parameters in S_i are solutions.

Subcase II. $\underline{p} \neq \underline{0} \quad \wedge \quad \exists \alpha \in R, \ \underline{\omega}(t) = \alpha \underline{p}$

Equation (3.48) simplifies to

$$\underline{\eta}(t) = \left(\frac{\underline{p}^T}{|\underline{p}|} \sin \frac{-\alpha |\underline{p}| T}{2}, \ \cos \frac{-\alpha |\underline{p}| T}{2} \right)^T \tag{3.50}$$

The fourth components in above and (3) give

$$\alpha |\underline{p}| T = 2n\pi \qquad , n \in Z$$

The absolute values on both sides give

$$|\underline{w}(t)| T = 2m\pi, \qquad m \in \{0\} \cup N \tag{3.51}$$

The case $m = 0$ leads to a subset of S_2, the case $m \in N$ to S_3.

Subcase III. $\underline{p} = \underline{0}$

By (3.48) and (3), we have

$$\underline{\xi}(\underline{w}(t); T) = \pm (0, 0, 0, 1)^T$$

Then (3.51) holds by Lemma 3.5.4. Hence the solutions of this subcase form a subset of $S_2 \cup S_3$.

Case II

$$\begin{aligned} &\exists t_i, t_{i+1} \in M_3, \ |\eta_4(t_i)| \neq 0 \ \wedge \ |\eta_4(t_i)| \neq 1 \ \wedge \ \hat{\underline{\eta}}_3(t_i) = \hat{\underline{\eta}}_3(t_{i+1}) \qquad (3.52) \quad .52 \\ &\Longleftrightarrow \forall t, \quad \eta_4(t) \neq 0 \ \wedge \ \eta_4(t) \neq \pm 1 \ \wedge \ \underline{\eta}(t) = \underline{\eta}(t + T) \end{aligned}$$

solution set $S_4 \cup S_5 \cup S_6$

$$S_4 : \begin{cases} |\underline{p}| = 2m\pi / T & , m \in N \\ \underline{w}(t_0) - \underline{p} = \left(\dfrac{2n\pi}{T} - \dfrac{2}{T} \tan^{-1} \dfrac{|\hat{\underline{\eta}}_3(t_0)|}{|\eta_4(t_0)|} \right) \dfrac{\hat{\underline{\eta}}_3(t_0)}{|\hat{\underline{\eta}}_3(t_0)|} & , n \in Z \end{cases}$$

$$S_5: \begin{cases} \underline{w}(t_0) &= \left(\frac{2n\pi}{T} - \frac{2}{T}\tan^{-1}\frac{|\hat{\underline{\eta}}_3(t_0)|}{|\underline{\eta}_4(t_0)|}\right)\frac{\hat{\underline{\eta}}_3(t_0)}{|\hat{\underline{\eta}}_3(t_0)|} & ,n \in Z \\ \underline{p} &= \alpha\underline{\omega}(t_0) & \alpha \in R \end{cases}$$

$$S_6: \begin{cases} \underline{p} &= \left(\frac{2m\pi}{T} - \frac{2}{T}\tan^{-1}\frac{|\hat{\underline{\eta}}_3(t_0)|}{|\underline{\eta}_4(t_0)|}\right)\frac{\hat{\underline{\eta}}_3(t_0)}{|\hat{\underline{\eta}}_3(t_0)|} & ,m \in Z \\ |\underline{w}(t_0) - \underline{p}| &= 2n\pi/T & ,n \in N \end{cases}$$

Proof for Case II:

To prove the solution set for this case, we consider three subcases.

Subcase I. $\exists m \in \{0\} \cup N, \quad |\underline{p}| = 2m\pi/T$

Because

$$c_p = (-1)^m \qquad \text{and} \qquad s_p = 0$$

a comparison between (3.45) and (3.48) shows

$$\underline{\xi}(\underline{\psi}(t); T) = (-1)^m \underline{\eta}(t)$$

on which Lemma 3.5.4 can be applied to solve for $\underline{\psi}(t_0)$. Therefore the case $m = 0$ leads to a subset of S_5, the case $m \in N$ to S_4.

Subcase II. $\forall m \in \{0\} \cup N, \quad |\underline{p}| \neq 2m\pi/T \quad \wedge \quad \exists \alpha \in R, \ \underline{\omega}(t) = \alpha\underline{p}$

Similar to Subcase II in Case I, (3.50) holds. Moreover,

$$\underline{\xi}(\underline{w}(t); T) = \underline{\eta}(t) \tag{3.53}$$

by a comparison between (3.45) and (3.50). Applying Lemma 3.5.4 on (3.53) to solve for $\underline{w}(t_0)$, we arrive at a subset of S_5.

Subcase III. $\forall m \in \{0\} \cup N, \ |\underline{p}| \neq 2m\pi/T \quad \wedge \quad \forall \alpha \in R, \ \underline{\omega}(t) \neq \alpha\underline{p}$

Since $|\underline{p}|T \neq 2m\pi$, $\exp[TP(\underline{p})]$ is a nonidentity rotation matrix. Furthermore, $\underline{\eta}_3(t_i)$, $t_i \in M$, is a nonzero eigenvector of $\exp[TP(\underline{p})]$ associated with the eigenvalue 1 by (3.47) and (3). As $\underline{p}$ is the rotation axis of $\exp[TP(\underline{p})]$, $\underline{\eta}_3(t_i)$ is parallel to $\underline{p}$. This with the fact that $\{\underline{p}, \underline{\psi}(t), \underline{\psi}(t) \times \underline{p}\}$ is a linearly independent set (similarly to Subcase I in Case I) yields, by (3.48),

$$c_p s_\psi = s_p s_\psi = 0$$

or equivalently

$$c_\psi = \pm 1 \qquad \text{and} \qquad s_\psi = 0$$

Then, by (3.45) and (3.48),

$$\underline{\xi}(\underline{p}; T) = \pm \underline{\eta}(t)$$

After applying Lemma 3.5.4 to solve for $\underline{p}$, we reach a subset of S_6.

Case III

$$\begin{aligned} &\exists t_i, t_{i+1} \in M_3, \quad |\eta_4(t_i)| \neq 0 \ \wedge \ |\eta_4(t_i)| \neq 1 \ \wedge \ \hat{\underline{\eta}}_3(t_i) = -\hat{\underline{\eta}}_3(t_{i+1}) \\ &\quad \Longleftrightarrow \forall t, \quad \eta_4(t) \neq 0 \ \wedge \ \eta_4(t) \neq \pm 1 \ \wedge \ \underline{\eta}_3(t) = -\underline{\eta}_3(t+T) \end{aligned} \tag{3.54}$$

.54

solution set S_7 :

$$\begin{cases} |\underline{p}| &= (2m-1)\pi/T & , m \in N \\ \underline{p} \cdot \hat{\underline{\eta}}_3(t_0) &= 0 & \\ \underline{w}(t_0) - \underline{p} &= \dfrac{(2n-1)\pi}{T|\underline{p}|}[-|\eta_4(t_0)|\underline{p} + \underline{p} \times \hat{\underline{\eta}}_3(t_0)] & , n \in Z \end{cases}$$

Proof for Case III:

By (3.47) and (3), each $\underline{\eta}_3(t_i)$, $t_i \in M$, is a nonzero eigenvector of the rotation matrix $\exp[TP(\underline{p})]$ associated with the eigenvalue -1. Thus the rotation axis $\underline{p}$ is perpendicular to $\underline{\eta}_3(t_i)$, and the rotation angle $|\underline{p}|T$ is

$$|\underline{p}|T = (2m-1)\pi, \quad m \in N \tag{3.55}$$

If $|\underline{\psi}(t)| = 0$, then by (3.48) $\underline{\eta}(t)$ does not change with t, which is a contradiction of (3). Hence $|\underline{\psi}(t)| \neq 0$. Then (3.48) and (3.55) give

$$\underline{\eta}_3(t) = (-1)^m \left\{ c_\psi \frac{\underline{p}}{|\underline{p}|} + s_\psi \frac{\underline{\psi}(t) \times \underline{p}}{|\underline{\psi}(t)||\underline{p}|} \right\}$$

Since $\underline{p}$ is perpendicular to $\underline{\eta}_3(t_i)$, the above yields $c_\psi = 0$, or equivalently

$$|\underline{\psi}(t)| = (2n-1)\pi/T, \quad n \in N \tag{3.56}$$

Using (3.55) and (3.56), (3.48) is reduced to

$$\underline{\eta}(t) = (-1)^{m+n} \begin{pmatrix} \frac{\underline{\psi}(t) \times \underline{p}}{|\underline{\psi}(t)||\underline{p}|} \\ -\frac{\underline{\psi}(t) \cdot \underline{p}}{|\underline{\psi}(t)||\underline{p}|} \end{pmatrix}$$

This shows that $\underline{\eta}_3(t_i)$ is perpendicular to both $\underline{p}$ and $\underline{\psi}(t_i)$. Therefore we can find α and β such that

$$\underline{\psi}(t_0) = \alpha \underline{p} + \beta \underline{p} \times \underline{\eta}_3(t_0) \tag{3.57}$$

To obtain α and β, we take the inner products with $\underline{p}$ and $\underline{p} \times \underline{\eta}_3(t_0)$, respectively, on both sides of (3.57). After simplifications, we have S_7.

Case IV

$$\exists t_i, t_{i+1} \in M_3, \quad |\eta_4(t_i)| = 0 \quad \wedge \quad \hat{\underline{\eta}}_3(t_i) = \pm \hat{\underline{\eta}}_3(t_{i+1})$$
$$\Longleftrightarrow \qquad \forall t, \quad \eta_4(t) = 0 \quad \wedge \quad \underline{\eta}(t) = \pm \underline{\eta}(t+T)$$

solution set $\quad S_8 \cup S_9 \cup S_{10} \cup S_{11}$

$$S_8 : \begin{cases} |\underline{p}| = (2m-1)\pi/T & , m \in N \\ \underline{p} \cdot \hat{\underline{\eta}}_3(t_0) = 0 & \\ \underline{w}(t_0) - \underline{p} = \frac{(2n-1)\pi}{T|\underline{p}|} \underline{p} \times \hat{\underline{\eta}}_3(t_0) & , n \in Z \end{cases}$$

$$S_9 : \begin{cases} |\underline{p}| = 2m\pi/T & , m \in N \\ \underline{w}(t_0) - \underline{p} = (2n-1)\pi \hat{\underline{\eta}}_3(t_0)/T & , n \in Z \end{cases}$$

$$S_{10} : \begin{cases} \underline{w}(t_0) = (2m-1)\pi \hat{\underline{\eta}}_3(t_0)/T & , m \in Z \\ \underline{p} = \alpha \underline{w}(t_0) & , \alpha \in R \end{cases}$$

$$S_{11} : \begin{cases} \underline{p} = (2m-1)\pi \hat{\underline{\eta}}_3(t_0)/T & , m \in Z \\ |\underline{w}(t_0) - \underline{p}| = 2n\pi/T & , n \in N \end{cases}$$

Proof for Case IV:

In the case when $\underline{\eta}(t_i) = -\underline{\eta}(t_{i+1})$, $t_i \in M$, we follow the derivations in Case III to reach S_8. In the other case when $\underline{\eta}(t_i) = \underline{\eta}(t_{i+1})$, derivation similar to Case II is used to obtain S_9, S_{10}, and S_{11}.

$$\begin{array}{ll} \textbf{Case V} & \exists t_i, t_{i+1} \in M_3, \quad |\eta_4(t_i)| \neq 0 \quad \wedge \quad \hat{\underline{\eta}}_3(t_i) \neq \pm \hat{\underline{\eta}}_3(t_{i+1}) \\ \Longleftrightarrow & \forall t, \quad \eta_4(t) \neq 0 \quad \wedge \quad \underline{\eta}_3(t) \neq \pm \underline{\eta}_3(t+T) \end{array} \qquad (3.58) \quad .58$$

$$\textbf{solution set} \qquad S_{12} \cap S_{13}$$

Proof for Case V:

This is the non-degenerate case. To derive the solution set, we start from (3.48) to obtain

$$\begin{cases} \underline{\eta}_3(t_1) = R[\underline{\xi}(\underline{p}; T)]\underline{\eta}_3(t_0) \\ \underline{\eta}_3(t_2) = R[\underline{\xi}(\underline{p}; T)]\underline{\eta}_3(t_1) \end{cases} \qquad (3.59)$$

From the linear independence of the set $\{\underline{\eta}_3(t_0), \underline{\eta}_3(t_1), \underline{\eta}_3(t_0) \times \underline{\eta}_3(t_1)\}$ by (3), we have the unique solution of $R[\underline{\xi}(\underline{p}; T)]$ from (3.59),

$$R[\underline{\xi}(\underline{p}; T)] = \left[\hat{\underline{\eta}}_3(t_1) \;\; \hat{\underline{\eta}}_3(t_2) \;\; \hat{\underline{\eta}}_3(t_1) \times \hat{\underline{\eta}}_3(t_2)\right] \cdot \qquad (3.60)$$

$$\left[\hat{\underline{\eta}}_3(t_0) \;\; \hat{\underline{\eta}}_3(t_1) \;\; \hat{\underline{\eta}}_3(t_0) \times \hat{\underline{\eta}}_3(t_1)\right]^{-1}$$

Then by Lemma 3.5.2 the quaternion $\underline{\xi}(\underline{p}; T)$ can be decided up to a sign ambiguity from $R[\underline{\xi}(\underline{p}; T)]$ as

$$\underline{\xi}(\underline{p}; T) \equiv (\xi_{p1}, \xi_{p2}, \xi_{p3}, \xi_{p4}) = \pm(\hat{\underline{\xi}}_{p3}^T, |\xi_{p4}|)^T \qquad (3.61)$$

We show $\underline{\xi}(\underline{p}; T) \neq \pm(0, 0, 0, 1)^T$ by contradiction. Otherwise, (3.13) gives $R[\underline{\xi}(\underline{p}; T)] = I$ and thus, by (3.47), $\underline{\eta}(t_i)$, $t_i \in M$, does not change with t_i, which is a contradiction of (3). Lemma 3.5.4 is applied to solve (3.61) for $\underline{p}$ from $\underline{\xi}(\underline{p}; T)$, which has a sign ambiguity, as

$$S_{12}: \quad \underline{p} = \begin{cases} \left(\dfrac{2m\pi}{T} - \dfrac{2}{T}\tan^{-1}\dfrac{|\hat{\underline{\xi}}_{p3}^T|}{|\xi_{p4}|}\right)\dfrac{\hat{\underline{\xi}}_{p3}^T}{|\hat{\underline{\xi}}_{p3}^T|} & \text{,if } |\xi_{p4}| \neq 0 \\ (2m-1)\pi\hat{\underline{\xi}}_{p3}^T/T & \text{,if } |\xi_{p4}| = 0 \end{cases} \qquad m \in Z$$

As a result of (3.48), the following equality holds [3]

$$\underline{\xi}(\underline{w}(t)-\underline{p};T)=\begin{pmatrix} \xi_{p4} & -\xi_{p3} & \xi_{p2} & -\xi_{p1} \\ \xi_{p3} & \xi_{p4} & -\xi_{p1} & -\xi_{p2} \\ -\xi_{p2} & \xi_{p1} & \xi_{p4} & -\xi_{p3} \\ \xi_{p1} & \xi_{p2} & \xi_{p3} & \xi_{p4} \end{pmatrix}\underline{\eta}(t)$$

Thus $\underline{\xi}(\underline{w}(t_0)-\underline{p};T)$ is known up to a sign ambiguity because both $\underline{\xi}(\underline{p};T)$ and $\underline{\eta}(t_0)$ are known up to a sign ambiguity by (3.43) and (3.61). Hence we have

$$\underline{\xi}(\underline{w}(t_0)-\underline{p};T)=\pm(\hat{\underline{\xi}}_{w3}^{T},|\xi_{w4}|)^{T} \tag{3.62}$$

We show $\underline{\xi}(\underline{w}(t_0)-\underline{p};T)\neq\pm(0,0,0,1)^{T}$ by contradiction. Otherwise, (3.45) gives $c_\psi=\pm1$ and $s_\psi=0$, which implies that $\underline{\eta}(t)$ does not change with t by (3.48), leading to a contradiction of (3). Lemma 3.5.4 is again used to solve (3.62) for $\underline{w}(t_0)-\underline{p}$,

$$S_{13}:\ \underline{w}(t_0)-\underline{p}=\begin{cases} \left(\frac{2n\pi}{T}-\frac{2}{T}\tan^{-1}\frac{|\hat{\underline{\xi}}_{w3}^{T}|}{|\xi_{w4}|}\right)\frac{\hat{\underline{\xi}}_{w3}^{T}}{|\hat{\underline{\xi}}_{w3}^{T}|} & \text{,if } \xi_{w4}\neq0 \\ (2n-1)\pi\hat{\underline{\xi}}_{w3}^{T}/T & \text{,if } \xi_{w4}=0 \end{cases} \qquad n\in Z$$

Let us remark that if $\xi_{p4}=0$ then $\xi_{w4}\neq0$.

Case VI

$$\begin{aligned} &\exists t_i,t_{i+1}\in M_3,\ \ |\eta_4(t_i)|=0\ \wedge\ \hat{\underline{\eta}}_3(t_i)\neq\pm\hat{\underline{\eta}}_3(t_{i+1})\ \wedge\ \hat{\underline{\eta}}_3(t_i)\cdot\hat{\underline{\eta}}_3(t_{i+1})\neq0 \\ &\Longleftrightarrow\forall t,\ \ \eta_4(t)=0\ \wedge\ \underline{\eta}(t)\neq\pm\underline{\eta}(t+T)\ \wedge\ \underline{\eta}_3(t)\cdot\underline{\eta}_3(t+T)\neq0 \end{aligned}$$

$$\textbf{solution set}\qquad \bigcup_{i=1}^{2}(S_{12}\cap S_{13})_i$$

Proof for Case VI:

For each $t_i\in M_3$, since $|\eta_4(t_i)|=0$, $\hat{\underline{\eta}}_3(t_i)$ is not uniquely specified by (3.44); instead $\hat{\underline{\eta}}_3(t_i)$ can be assigned by two possible vector values, one of which is the

[3] Or equivalently, $\underline{\xi}(\underline{w}(t)-\underline{p};T)$ is the quaternion multiplication of $\underline{\eta}(t)$ and the conjugate of $\underline{\xi}(\underline{p};T)$.

negative of the other. Let us arbitrarily choose an assignment such that the following preservation of inner product, by Lemma 3.5.4, holds.

$$\underline{\hat{\eta}}_3(t_1) \cdot \underline{\hat{\eta}}_3(t_2) = \underline{\hat{\eta}}_3(t_2) \cdot \underline{\hat{\eta}}_3(t_3) \tag{3.63}$$

Then we have two solutions for $R[\underline{\xi}(\underline{p};T)]$ from (3.59),

$$R[\underline{\xi}_{(1)}(\underline{p};T)] =$$

$$\left[\ \underline{\hat{\eta}}_3(t_1)\ \ \underline{\hat{\eta}}_3(t_2)\ \ \ \underline{\hat{\eta}}_3(t_1) \times \underline{\hat{\eta}}_3(t_2)\right] \left[\underline{\hat{\eta}}_3(t_0)\ \ \ \underline{\hat{\eta}}_3(t_1)\ \ \ \underline{\hat{\eta}}_3(t_0) \times \underline{\hat{\eta}}_3(t_1)\right]^{-1}$$

$$R[\underline{\xi}_{(2)}(\underline{p};T)] =$$

$$\left[-\underline{\hat{\eta}}_3(t_1)\ \ \underline{\hat{\eta}}_3(t_2)\ -\underline{\hat{\eta}}_3(t_1) \times \underline{\hat{\eta}}_3(t_2)\right] \left[\underline{\hat{\eta}}_3(t_0)\ -\underline{\hat{\eta}}_3(t_1)\ -\underline{\hat{\eta}}_3(t_0) \times \underline{\hat{\eta}}_3(t_1)\right]^{-1}$$

Starting from each of the above solutions of $R[\underline{\xi}(\underline{p};T)]$, we can proceed by using the same procedure following (3.60) in Case V to find the corresponding solution set $S_{12} \cap S_{13}$. As $R[\underline{\xi}_{(1)}(\underline{p};T)]$ and $R[\underline{\xi}_{(2)}(\underline{p};T)]$ lead to different $S_{12} \cap S_{13}$, the solution set for this case is the union of these two $S_{12} \cap S_{13}$.

Case VII

$$\iff \begin{array}{rl} \exists t_i, t_{i+1} \in M_3, & |\eta_4(t_i)| = 0 = \underline{\hat{\eta}}_3(t_i) \cdot \underline{\hat{\eta}}_3(t_{i+1}) \\ \forall t, & \eta_4(t) = 0 = \underline{\eta}_3(t) \cdot \underline{\eta}_3(t+T) \end{array} \tag{3.64}$$

.64

$$\textbf{solution set} \qquad \bigcup_{j=1}^{4} (S_{12} \cap S_{13})_j$$

Proof for Case VII:

Same as Case VI, each $\underline{\hat{\eta}}_3(t_i)$, $t_i \in M_3$, can be assigned by two possible vector values, one being the negative of the other. Among these sign arrangements, the preservation of inner product (3.63) always holds since the inner product is zero by (3). Let us arbitrarily pick up a sign arrangement, then there are four different solutions for $R[\underline{\xi}(\underline{p};T)]$ from (3.59),

$$R[\underline{\xi}_{(1)}(\underline{p};T)] =$$

$$\left[\ \underline{\hat{\eta}}_3(t_1)\ \ \underline{\hat{\eta}}_3(t_2)\ \ \ \underline{\hat{\eta}}_3(t_1) \times \underline{\hat{\eta}}_3(t_2)\right] \left[\ \underline{\hat{\eta}}_3(t_0)\ \ \ \underline{\hat{\eta}}_3(t_1)\ \ \ \underline{\hat{\eta}}_3(t_0) \times \underline{\hat{\eta}}_3(t_1)\right]^{-1}$$

$$R[\underline{\xi}_{(2)}(\underline{p};T)] =$$

$$\left[-\hat{\underline{\eta}}_3(t_1)\ \ \hat{\underline{\eta}}_3(t_2)\ -\hat{\underline{\eta}}_3(t_1)\times\hat{\underline{\eta}}_3(t_2)\right]\left[\ \hat{\underline{\eta}}_3(t_0)\ -\hat{\underline{\eta}}_3(t_1)\ -\hat{\underline{\eta}}_3(t_0)\times\hat{\underline{\eta}}_3(t_1)\right]^{-1}$$

$$R[\underline{\xi}_{(3)}(\underline{p};T)] =$$

$$\left[\ \hat{\underline{\eta}}_3(t_1)\ \ \hat{\underline{\eta}}_3(t_2)\ \ \ \hat{\underline{\eta}}_3(t_1)\times\hat{\underline{\eta}}_3(t_2)\right]\left[-\hat{\underline{\eta}}_3(t_0)\ \ \ \hat{\underline{\eta}}_3(t_1)\ -\hat{\underline{\eta}}_3(t_0)\times\hat{\underline{\eta}}_3(t_1)\right]^{-1}$$

$$R[\underline{\xi}_{(4)}(\underline{p};T)] =$$

$$\left[-\hat{\underline{\eta}}_3(t_1)\ \ \hat{\underline{\eta}}_3(t_2)\ -\hat{\underline{\eta}}_3(t_1)\times\hat{\underline{\eta}}_3(t_2)\right]\left[-\hat{\underline{\eta}}_3(t_0)\ -\hat{\underline{\eta}}_3(t_1)\ \ \ \hat{\underline{\eta}}_3(t_0)\times\hat{\underline{\eta}}_3(t_1)\right]^{-1}$$

Same as Case VI, the above four solutions lead to four different $S_{12}\cap S_{13}$.

3.5.4 Uniqueness of Translation Parameters And Rotation Center Parameters

After the discussion of the solution sets of rotation parameters, we proceed to prove the other part of Theorem 6.1. Let us consider the following motion parameters: rotation center $\underline{d}$, position vector $\underline{r}(t)$, velocity $\underline{v}(t)$, and acceleration $\underline{a}$. The abbreviation

$$\begin{aligned}\underline{r}_0 &= \underline{r}(t_0)\\ \underline{v}_0 &= \underline{v}(t_0)\end{aligned}$$

will be used.

Linear Equation for $\underline{r}_0, \underline{v}_0, \underline{a}$ And $\underline{d}$

Assume that in Theorem 6.1 the point being tracked at every $t_i \in M_5$ is the j^{th} feature point. The 3-D structure coordinates $\underline{s}_j$ of this feature point in the structure coordinate S can be obtained uniquely at t_0 by (3.4). A rearrangement

of (3.8) and (3.10) gives

$$\underbrace{\begin{pmatrix} \underline{r}_j(t_0) - R[\underline{q}(t_0)]\underline{s}_j \\ \underline{r}_j(t_1) - R[\underline{q}(t_1)]\underline{s}_j \\ \vdots \\ \underline{r}_j(t_4) - R[\underline{q}(t_4)]\underline{s}_j \end{pmatrix}}_{\equiv Z} = \overbrace{\underbrace{\begin{pmatrix} t_0^2 I_3/2 & t_0 I_3 & I_3 & -R[\underline{q}(t_0)] \\ t_1^2 I_3/2 & t_1 I_3 & I_3 & -R[\underline{q}(t_1)] \\ \vdots & \vdots & \vdots & \vdots \\ t_4^2 I_3/2 & t_4 I_3 & I_3 & -R[\underline{q}(t_4)] \end{pmatrix}}_{\equiv H_1} \underbrace{\quad}_{\equiv H_2}}^{\equiv H} \begin{pmatrix} \underline{a} \\ \underline{v}_0 \\ \underline{r}_0 \\ \underline{d} \end{pmatrix} \tag{3.65}$$

We notice that $R[\underline{q}(t_i)]$ $t_i \in M_4$ are uniquely decided by Lemma 3.5.1. $R[\underline{q}(t_4)]$ is also unique because each case in Section 3.5.3 decides $\underline{q}(t_4)$ up to a sign ambiguity. Hence every quantity in the matrices Z and H in (3.65) is uniquely decided. As the motion parameters of interest are the solutions of (3.65), we may analyze the linear mapping H to investigate the uniqueness of the solutions.

Velocity $\underline{v}(t)$ And Acceleration $\underline{a}$ are Unique

Suppose we have a motion $\{\underline{d}, \underline{r}_0, \underline{v}_0, \underline{a}\}$ that yields the same measurements as does a motion $\{\underline{d}', \underline{r}_0', \underline{v}_0', \underline{a}'\}$. Then (3.65) yields

$$\Delta\underline{r}_0 + t_i\Delta\underline{v}_0 + t_i^2\Delta\underline{a}/2 = R[\underline{q}(t_i)]\Delta\underline{d}, \qquad t_i \in M_5 \tag{3.66}$$

where $\Delta\underline{r}_0 \equiv \underline{r}_0 - \underline{r}_0'$ etc. After taking self inner product on both sides, we have

$$\begin{aligned} &t_i^4|\Delta\underline{a}|^2 + 4t_i^3(\Delta\underline{a})^T(\Delta\underline{v}_0) + 4t_i^2\left[|\Delta\underline{v}_0|^2 + (\Delta\underline{a})^T(\Delta\underline{r}_0)\right] \\ &+ 8t_i(\Delta\underline{v}_0)^T(\Delta\underline{r}_0) + 4|\Delta\underline{r}_0|^2 = 4|\Delta\underline{d}|^2, \qquad t_i \in M_5 \end{aligned}$$

This is a polynomial of fourth order in t_i. If there are five different t_i satisfying the equality, every coefficient must be zero. Therefore

$$\Delta\underline{a} = \Delta\underline{v}_0 = \underline{0} \tag{3.67}$$

This shows that $\underline{a}$ and $\underline{v}(t_0)$ are unique. By (3.10), the velocity $\underline{v}(t)$ is unique for all t. As a result of (3.67) and (3.66), we have

$$\Delta \underline{r}_0 = R[\underline{q}(t_i)]\Delta \underline{d}, \quad t_i \in M_5 \tag{3.68}$$

Since, by (3.65), (3.66) is equivalent to

$$H\Delta \underline{x} = \underline{0}$$

where $\Delta \underline{x} \equiv (\Delta \underline{a}^T, \Delta \underline{v_0}^T, \Delta \underline{r_0}^T, \Delta \underline{d}^T)^T$, the vector $\underline{\Delta x}$ is in the null space of H, null(H). Because the first six components, $\Delta \underline{a}$ and $\Delta \underline{v}_0$, of the vectors in null(H) are proved to be zeros, we have

$$\begin{cases} \text{rank}(H_1) = 6 \quad (\text{ i.e. } H_1 \text{ has full rank. }) \\ \text{span(column vectors of } H_1 \text{)} \cap \text{span(column vectors of } H_2 \text{)} = \{\underline{0}\} \end{cases} \tag{3.69}$$

Formal Solutions of $\underline{v}_0, \underline{a}, \underline{r}_0$, and $\underline{d}$

By observing the submatrix I_3 in H_2, we have

$$\gamma \equiv \text{rank}(H_2) \geq 3$$

Then we can choose γ linearly independent column vectors of H_2 to form a full rank matrix H_2'. If $\gamma < 6$, the remaining column vectors of H_2 constitute a matrix H_2''. Also, in (3.65) let the components of $(\underline{r_0}^T, \underline{d}^T)^T$ corresponding to the column vectors of H_2' be

$$\underline{e}^T \equiv (e_1, e_2, \cdots, e_\gamma)^T$$

and the remaining components of $(\underline{r_0}^T, \underline{d}^T)^T$ be

$$\underline{f}^T \equiv (f_1, f_2, \cdots, f_{6-\gamma})^T$$

Under these arrangements, we have

$$\begin{aligned} \text{span(column vectors of } H_2'') &\subseteq \text{span(column vectors of } H_2') \\ &= \text{span(column vectors of } H_2) \end{aligned} \tag{3.70}$$

If we rewrite (3.65) as

$$\begin{aligned} Z &= (H_1 \;\; H_2' \;\; H_2'')(\underline{a}^T, \underline{v_0}^T, \underline{e}^T, \underline{f}^T)^T \\ &= H'(\underline{a}^T, \underline{v_0}^T, \underline{e}^T)^T + H_2''\underline{f} \end{aligned} \tag{3.71}$$

where

$$H' \equiv (H_1, H_2')$$

has full rank by (3.69), then from (3.69), (3.70), and (3.71), we obtain

$$\begin{cases} \underline{f} = \text{arbitrary} \\ (\underline{a}^T, \underline{v_0}^T, \underline{e}^T)^T = (H'^T H')^{-1} H'^T (Z - H_2''\underline{f}) \end{cases} \tag{3.72}$$

Uniqueness Analysis of $\underline{r}_0$ And $\underline{d}$

A result of (3.68) is

$$R[\underline{q}(t_{i+1})]R^{-1}[\underline{q}(t_i)]\Delta\underline{r}_0 = \Delta\underline{r}_0, \qquad t_i \in M_3$$

From (3.40) and (3.13), we have

$$R[\underline{\eta}(t)] = R[\underline{q}(t+T)]R^T[\underline{q}(t)] = R[\underline{q}(t+T)]R^{-1}[\underline{q}(t)] \tag{3.73}$$

Hence

$$R[\underline{\eta}(t_i)]\Delta\underline{r}_0 = \Delta\underline{r}_0 \qquad t_i \in M_3 \tag{3.74}$$

$$\textbf{Condition I} \quad \begin{aligned} &\exists t_i, t_{i+1} \in M_3, \;\; \hat{\underline{\eta}}_3(t_i) \neq \pm\hat{\underline{\eta}}_3(t_{i+1}) \\ \iff \quad &\forall t, \;\; \underline{\eta}_3(t) \neq \underline{\eta}_3(t+T) \end{aligned} \tag{3.75}$$

.75

This is the nondegenerate case. It will be shown that $\underline{r}_0$ and $\underline{d}$ are unique.

In this case, $R[\underline{\eta}(t_i)]$ is not an identity matrix since

$$\underline{\eta}(t) \neq \pm(0,0,0,1)^T \tag{3.76}$$

from the equivalent statement (3). Furthermore, (3.74) shows that $\Delta\underline{r}_0$ is an eigenvector of $R[\underline{\eta}(t_i)]$ associated with the eigenvalue 1, and so

$$\exists \alpha_i \in R, \quad \Delta\underline{r}_0 = \alpha_i \underline{\eta}_3(t_i), \quad t_i \in M_3 \tag{3.77}$$

This yields

$$\alpha_0 \underline{\eta}_3(t_0) = \alpha_1 \underline{\eta}_3(t_1) \tag{3.78}$$

Since

$$|\underline{\eta}_3(t)| = \text{constant}$$

by Lemma 3.5.3, the condition $\underline{\eta}_3(t_0) \neq \pm\underline{\eta}_3(t_1)$ in (3) implies that $\underline{\eta}_3(t_0)$ and $\underline{\eta}_3(t_1)$ are not parallel because $\underline{\eta}_3(t) \neq \underline{0}$ by (3.76). Thus the linear independence of $\underline{\eta}_3(t_0)$ and $\underline{\eta}_3(t_1)$ and (3.78) give

$$\alpha_0 = \alpha_1 = 0$$

From (3.77) and (3.68) we see

$$\Delta\underline{r}_0 = \Delta\underline{d} = \underline{0}$$

This shows that $\underline{r}_0$ and $\underline{d}$ are unique. Then it follows that $\underline{r}(t)$ are unique for all t by (3.10). Furthermore, to find the **solution set**, we can apply (3.65) to obtain

$$\textbf{Condition I} \quad S_{14} : (\underline{a}^T, \underline{v}_0{}^T, \underline{r}_0{}^T, \underline{d}^T)^T = (H^T H)^{-1} H^T Z$$

$$\begin{array}{lll} \textbf{Condition II} & \exists t_i \in M_3 & \underline{\eta}(t_i) = \pm(0,0,0,1)^T \\ & \Longleftrightarrow \qquad \forall t, & \underline{\eta}(t) = \pm((0,0,0,1)^T \end{array} \tag{3.79}$$

.79

In this degenerate case, it will be shown that the rotation center is totally ambiguous and can be anywhere in the space.

From (3) and (3.73), we have

$$I_3 = R[\underline{\eta}(t)] = R[\underline{q}(t+T)]R^{-1}[\underline{q}(t)]$$

This implies

$$R[\underline{q}(t_j+T)] = R[\underline{q}(t_j)], \quad t_j \in M$$

Since all R are equal at the measurement times, by (3.65) the rank of H_2 is equal to the rank of the 3×6 matrix

$$\begin{pmatrix} I_3 & -R[\underline{q}(t_0)] \end{pmatrix}$$

which has rank 3. Therefore we can choose

$$H_2' = \begin{pmatrix} I_3 \\ I_3 \\ \vdots \\ I_3 \end{pmatrix} \quad H_2'' = \begin{pmatrix} -R[\underline{q}(t_0)] \\ -R[\underline{q}(t_0)] \\ \vdots \\ -R[\underline{q}(t_0)] \end{pmatrix} \quad H' = (H_1, H_2')$$

$$\underline{e} = \underline{r}_0 \qquad \underline{f} = \underline{d}$$

in (3.72) to obtain the **solution set**

$$\textbf{Condition II } S_{15}: \begin{cases} \underline{d} = \text{arbitrary} \\ (\underline{a}^T, \underline{v}_0{}^T, \underline{r}_0{}^T)^T = (H'^T H')^{-1} H'^T (Z - H_2'' \underline{d}) \end{cases}$$

Hence there are three degrees of freedom involved in the six components of $(\underline{r}_0{}^T, \underline{d}^T)^T$.

$$\textbf{Condition III} \qquad \begin{array}{rl} & \exists t_i, t_{i+1} \in M_3 \quad \hat{\underline{\eta}}_3(t_i) = \pm \hat{\underline{\eta}}_3(t_{i+1}) \neq \underline{0} \\ \iff & \forall t, \quad \underline{\eta}_3(t) = \pm \underline{\eta}_3(t+T) \neq \underline{0} \end{array} \tag{3.80}$$

.80

It will be shown that this degenerate case confines the rotation center to lie on a straight line fixed on the rigid body.

As in Condition I, we have

$$\exists \alpha_i \in R, \quad \Delta \underline{r}_0 = \alpha_i \underline{\eta}_3(t_i), \quad t_i \in M_3$$

Therefore, from (3.68), it holds that

$$\exists \alpha \in R, \quad \begin{pmatrix} \Delta \underline{r}_0 \\ \Delta \underline{d} \end{pmatrix} = \alpha \begin{pmatrix} \underline{\eta}_3(t_0) \\ R^{-1}[\underline{q}(t_0)]\underline{\eta}_3(t_0) \end{pmatrix} \tag{3.81}$$

if $t_i = t_0$ is considered. This constraint shows that the null space of H_2 has dimension *at most one*. Now we show that

$$\begin{pmatrix} \underline{\eta}_3(t_0) \\ R^{-1}[\underline{q}(t_0)]\underline{\eta}_3(t_0) \end{pmatrix} \in \text{null}(H_2) \tag{3.82}$$

Premultiplying the above vector by H_2, we find that (3.82) is true if we notice that

$$\begin{aligned} & I_3\underline{\eta}_3(t_0) - R[\underline{q}(t_i)]\left(R^{-1}[\underline{q}(t_0)]\underline{\eta}_3(t_0)\right) \\ = \; & \underline{\eta}_3(t_0) - R[\underline{q}(t_i)]R^{-1}[\underline{q}(t_i - T)]R[\underline{q}(t_i - T)]\cdots \\ & R^{-1}[\underline{q}(t_1)]R[\underline{q}(t_1)]R^{-1}[\underline{q}(t_0)]\underline{\eta}_3(t_0) \\ = \; & \underline{\eta}_3(t_0) - R[\underline{\eta}(t_i - T)]\cdots R[\underline{\eta}(t_0)]\underline{\eta}_3(t_0) \\ = \; & \underline{\eta}_3(t_0) - \underline{\eta}_3(t_0) = \underline{0}, \qquad t_i \in M_5 \end{aligned}$$

in which the second equality holds by (3.73) and the third by the property that $\underline{\eta}_3(t_0)$ is an eigenvector of $R[\underline{\eta}(t_i)]$ associated with the eigenvalue 1 for all $t_i \in M$,

$$R[\underline{\eta}(t_i)]\underline{\eta}_3(t_i) = \underline{\eta}_3(t_i)$$

and

$$\underline{\eta}(t_i) = \pm\underline{\eta}(t_j), \quad \forall t_i, t_j \in M$$

by (3). Therefore the dimension of null(H_2) is one by (3.82), and thus

$$\text{rank}(H_2) = 5$$

Then we can find five linearly independent column vectors in H_2 to form a full rank matrix H_2' and obtain the **solution set** from (3.72) as

$$\textbf{Condition III } S_{16}: \begin{cases} f = \text{an arbitrary real number} \\ (\underline{a}^T, \underline{v_0}^T, e_1, e_2, \cdots, e_5)^T = (H'^T H')^{-1} H'^T (Z - H_2'' f) \end{cases}$$

Hence there is one degree of freedom involved in the six components of $(\underline{r_0}^T, \underline{d}^T)^T$. Furthermore, if $(\underline{r_0}^T, \underline{d}^T)^T$ is a solution, then

$$\forall \alpha \in R, \quad \begin{pmatrix} \underline{r_0} \\ \underline{d} \end{pmatrix} + \alpha \begin{pmatrix} \underline{\hat{\eta}_3}(t_0) \\ R^{-1}[\underline{q}(t_0)]\underline{\hat{\eta}_3}(t_0) \end{pmatrix} \tag{3.83}$$

is also a solution by (3.82) and (3.44).

Physical Interpretation of Uniqueness Results of $\underline{r_0}$ And $\underline{d}$

Condition I

As the measurements are taken at evenly separated time instants, the system is analogous to observing a rotating object under a stroboscope. In Condition I, between two consecutive snapshots of the "stroboscope", the rotation axis appears to be changing with time. Therefore, the rotation center, which is the intersection of the rotation axes at different times, can not be ambiguous.

The set of rotations which lead to Condition I is the union of the solution sets of Case V, VI, and VII in Section 3.5.3.

Condition II

Since at every snapshot, the object rotates back to the original orientation, i.e. it looks to be undergoing a translational motion only, the rotation center and the rotation axis are totally ambiguous. Every point on the rigid body can be chosen to be the rotation center O_b in Figure 3.1.

The set of rotations which cause the ambiguity of Condition II is $S_1 \cup S_2 \cup S_3$ as shown in Section 3.5.3.

Condition III

The rotation axis of the object between every two consecutive snapshots looks fixed. Thus the rotation center is confined on that unambiguous axis, but it is impossible to decide which point on that axis is the rotation center. The equality

$$\Delta \underline{r}_0 = \alpha \hat{\underline{\eta}}_3(t_0)$$

in (3.83) means that if we want to move the rotation center to get another set of motion parameters, we can only slide it along the direction of the above axis, which has the direction along $\hat{\underline{\eta}}_3(t_0)$ if expressed in the inertial coordinate system I.

The set of rotations which cause the ambiguity of Condition III is $\cup_{j=4}^{11} S_j$.

3.5.5 Effects of Changing the Structure Coordinate System S

Since the structure coordinate system S can be any 3-D coordinate system fixed on the rigid body as defined in Section 3.1.2, the choice of S is not unique. Even the admissible S defined by (3.2) and (3.3) are not unique because the number of feature points in (3.2) and the ordering of feature points in (3.3) do affect the choice of S. In this section, we discuss the effects on the solution set of motion parameters when S is changed.

Suppose that S and S' are two admissible structure coordinate systems. Let $\underline{s}_j$ and $\underline{s}'_j$ denote the 3-D coordinates of the j^{th} feature point in S and S', respectively. Similarly, we superscribe the corresponding motion parameters by $'$ if S' is adopted. If S is used, the measurement equation (3.8) and the plant equation (3.32) are

$$\underline{r}_j(t_i) = \underline{r}(t_i) + R[\underline{q}(t_i)](\underline{s}_j - \underline{d}) \tag{3.84}$$

$$\begin{cases} \dot{\underline{r}}(t) &= \underline{v}(t) \\ \dot{\underline{v}}(t) &= \underline{a} \\ \dot{\underline{q}}(t) &= \Omega[\underline{\omega}(t)]\,\underline{q}(t) \\ \dot{\underline{w}}(t) &= P(\underline{p})\underline{w}(t) \end{cases} \tag{3.85}$$

Similarly, if S' is used, we have

$$\underline{r}_j(t_i) = \underline{r}'(t_i) + R[\underline{q}'(t_i)](\underline{s}'_j - \underline{d}') \tag{3.86}$$

$$\begin{cases} \dot{\underline{r}}'(t) &= \underline{v}'(t) \\ \dot{\underline{v}}'(t) &= \underline{a}' \\ \dot{\underline{q}}'(t) &= \Omega[\underline{\omega}'(t)]\,\underline{q}'(t) \\ \dot{\underline{w}}'(t) &= P(\underline{p}')\underline{w}'(t) \end{cases} \tag{3.87}$$

Since both S and S' are two 3-D coordinate systems fixed on the rigid body, there exist a rotation matrix $R[\underline{q}_0]$ and a translation vector $\underline{t}_0$ such that for every feature point we have

$$\underline{s}'_j = R[\underline{q}_0]\underline{s}_j + \underline{t}_0 \tag{3.88}$$

Since three noncollinear feature points are used, $R[\underline{q}_0]$ and $\underline{t}_0$ are unique, and $\underline{q}_0$ is thus unique up to a sign ambiguity by Lemma 3.5.2. Define [4]

$$\begin{aligned} \underline{q}''(t) &= \underbrace{\begin{pmatrix} q_{04} & -q_{03} & q_{02} & q_{01} \\ q_{03} & q_{04} & -q_{01} & q_{02} \\ -q_{02} & q_{01} & q_{04} & q_{03} \\ -q_{01} & -q_{02} & -q_{03} & q_{04} \end{pmatrix}}_{\equiv Q_0} \underline{q}'(t) \\ \underline{d}'' &= \underline{d}' - R^{-1}[\underline{q}_0]\underline{t}_0 \end{aligned} \tag{3.89}$$

where

$$\underline{q}_0 = (q_{01}, q_{02}, q_{03}, q_{04})^T$$

Since, from (3.89) and (3.13),

$$R[\underline{q}''(t)] = R[\underline{q}'(t)]R[\underline{q}_0] \tag{3.90}$$

we have

$$\underline{r}_j(t_i) = \underline{r}'(t_i) + (R[\underline{q}'(t_i)]R[\underline{q}_0])[\underline{s}_j - (\underline{d}' - R^{-1}[\underline{q}_0]\underline{t}_0)]$$

[4] Or equivalently, $\underline{q}''(t)$ is the quaternion multiplication of $\underline{q}_0$ and $\underline{q}'(t)$.

$$= \underline{r}'(t_i) + R[\underline{q}''(t_i)](\underline{s}_j - \underline{d}'') \tag{3.91}$$

$$\begin{cases} \dot{\underline{r}}'(t) &= \underline{v}'(t) \\ \dot{\underline{v}}'(t) &= \underline{a}' \\ \dot{\underline{q}}''(t) &= \Omega[\underline{\omega}'(t)]\,\underline{q}''(t) \\ \dot{\underline{w}}'(t) &= P(\underline{p}')\underline{w}'(t) \end{cases} \tag{3.92}$$

in which the first equality holds by a substitution of (3.88) into (3.86), the second by (3.89) and (3.90), and the fifth by (3.87), (3.89), and the fact that $\Omega[\underline{\omega}'(t)]$ and Q_0 commute. Comparing (3.84) and (3.85) with (3.91) and (3.92), we find that

$$\{\underline{r}(t), \underline{v}(t), \underline{a}, \underline{d}, \underline{q}(t), \underline{w}(t), \underline{p}\}$$

and

$$\{\underline{r}'(t), \underline{v}'(t), \underline{a}', \underline{d}'', \underline{q}''(t), \underline{w}'(t), \underline{p}'\}$$

have the same solution set because they satisfy the equation sets of the same form. Therefore

$$\begin{cases} \underline{r}(t) &= \underline{r}'(t) \\ \underline{v}(t) &= \underline{v}'(t) \\ \underline{a} &= \underline{a}' \\ \underline{d} &= \underline{d}'' &= \underline{d}' - R^{-1}[\underline{q}_0]\underline{t}_0 \\ \underline{q}(t) &= \underline{q}''(t) &= Q_0\underline{q}'(t) \\ \underline{w}(t) &= \underline{w}'(t) \\ \underline{p} &= \underline{p}' \end{cases}$$

where $R^{-1}[\underline{q}_0]\underline{t}_0$ is unique and Q_0 is unique up to a sign ambiguity. This establishes the last claim in Theorem 6.1.

3.6 A Deterministic Algorithm to Recover Motion

After the completion of a constructive proof of Theorem 6.1, we summarize a deterministic algorithm which recovers motion parameters from a stereo image sequence in the absence of noise.

1. Compute the 3-D inertial coordinates $\underline{r}_j(t_i)$, $t_i \in M_4$, of each feature point and $\underline{r}_1(t_4)$ from stereo image pairs using (3.38).

2. Compute the structure coordinates $\underline{s}_j$ of each feature point using (3.4) by setting $t_i = t_0$.

3. Compute the alignment matrix $R[\underline{q}(t_i)]$, $t_i \in M_4$, using (3.41).

4. Compute the quaternions $\underline{q}(t_i)$, $t_i \in M_4$, up to a sign ambiguity from $R[\underline{q}(t_i)]$, $t_i \in M_4$. (Lemma 3.5.2)

5. Compute the interframe quaternions $\underline{\eta}(t_i)$, $t_i \in M_3$, up to a sign ambiguity by (3.40). Then put the computed values in the form of $\pm(\hat{\underline{\eta}}_3^T(t_i), |\eta_4(t_i)|)^T$, $t_i \in M_3$, as shown in (3.44).

6. Compute $\underline{w}(t_0)$ and $\underline{p}$ according to the seven different cases of $\underline{\eta}$ using the enumeration of solution sets in Case I to VII in Section 3.5.3.

7. Compute $\underline{q}(t)$ and $\underline{w}(t)$ from $\underline{q}(t_0), \underline{w}(t_0)$, and $\underline{p}$ by (3.18) and (3.16) for any t needed.

8. Compute $R[\underline{q}(t_4)]$ from $\underline{q}(t_4)$ by (3.13).

9. Compute $\underline{r}(t_0), \underline{v}(t_0), \underline{a}$, and $\underline{d}$ according the three different conditions of $\underline{\eta}$ using the enumeration of solution sets in Condition I, II, and III in Section 3.5.4.

10. Compute $\underline{r}(t)$ and $\underline{v}(t)$ from $\underline{r}(t_0), \underline{v}(t_0)$, and $\underline{a}$ by (3.10) for any t needed.

If more measurements than those in Theorem 6.1 are available in the presence of noise, a procedure such as the least squares may be applied at each step to smooth the noise. While the determination of motion parameters from noise-free image measurements using the deterministic algorithm is precise and fast, accuracy is not guaranteed in the presence of noise. Therefore the deterministic algorithm summarized in this section is recommended only for noise-free measurements or for situations when estimates of relatively low accuracy are needed, e.g. for obtaining a reasonable initial guess for recursive filters.

Weng et. al. [76] have used a deterministic algorithm to recover motion parameters using a dynamic model of constant precessional motion from noisy images. It appears that they have implicitly made the assumption that both precession and interframe rotation are strictly less than a half turn between two consecutive measurements, which is reasonable in practical problems. Then their

remarks regarding uniqueness supports our results proved in Condition I, Section 3.5.4 and Case V, Section 3.5.3. They have also proved that pure translation and rotation without precession lead to the ambiguity discussed in our Conditions II and III, respectively. These proofs are special cases of our results because pure translation is a special case of $S_1 \cup S_2 \cup S_3$, which we prove to be the only motions leading to Condition II, and rotation without precession is a special case of $\cup_{j=4}^{11} S_j$, the only motions leading to Condition III.

3.7 Performance Bounds

The estimation process formulated can be separated into two stages: establishing models to describe the underlying motion and measurements and applying an estimator to estimate the motion parameters. After the first stage and before the second stage, it is desirable to know what the maximal achievable accuracy of estimates for the model parameters is. If the maximal achievable accuracy is not satisfactory, the models need to be changed before the design of an estimator. Otherwise even optimal estimators applied thereafter will give estimates of unsatisfactory accuracy. On the other hand, once the estimator is designed, its performance can be compared to the maximal achievable accuracy to give an insight into the efficiency of the estimator. Since variances quantify the accuracy of estimates if bias is small, the maximal achievable accuracy can be quantified by the greatest lower bound of the variances of estimates. In this section, this performance bound is discussed using the Cramér-Rao inequality [63]. The performance bound is intrinsic to the models designed in the first stage, independent of the estimators applied later on, and gives an indication of the noise immunity of the motion and measurement models.

3.7.1 Fisher Information Matrix

As needed in the Cramér-Rao inequality, the Fisher information matrix, which is time varying, is calculated first.

Assume that before t, the measurements have been taken at the discrete times $t_0, \ldots, t_l$ and $t_0 < t_1 < \ldots < t_l \leq t$. Recall that $\underline{x}(t)$ is the state vector at time

t. To express $\underline{x}(t_i)$ in terms of a one-to-one function of $\underline{x}(t)$, we apply the state transition equation (3.34)

$$\underline{x}(t_i) = \Phi_c(\underline{x}(t); t_i - t)\underline{x}(t) \tag{3.93}$$

The measurement model (3.33) gives

$$\underbrace{\begin{pmatrix} \underline{z}(t_0) \\ \vdots \\ \underline{z}(t_l) \end{pmatrix}}_{\equiv \underline{Z}_t} = \underbrace{\begin{pmatrix} \underline{h}\left[\Phi_c(\underline{x}(t); t_0 - t)\underline{x}(t); t_0\right] \\ \vdots \\ \underline{h}\left[\Phi_c(\underline{x}(t); t_l - t)\underline{x}(t); t_l\right] \end{pmatrix}}_{\equiv \underline{h}_t\left[\underline{x}(t)\right]} + \underbrace{\begin{pmatrix} \underline{v}_n(t_0) \\ \vdots \\ \underline{v}_n(t_l) \end{pmatrix}}_{\equiv \underline{n}_t} \tag{3.94}$$

Therefore $\underline{Z}_t$ is the collection of all measurements before t. Since the components of $\underline{n}_t$ are Gaussian and of zero mean, the probability density function of $\underline{n}_t$ is

$$f(\underline{n}_t) = [(2\pi)^{p/2}\sqrt{\det R_n}]^{-1} \exp[-\underline{n}_t^T R_n^{-1} \underline{n}_t/2] \tag{3.95}$$

where p is the number of components of $\underline{n}_t$ and R_n the covariance matrix of $\underline{n}_t$. From (3.94) and (3.95), we have the conditional probability density function of $\underline{Z}_t$ given $\underline{x}(t)$ as

$$f(\underline{Z}_t|\underline{x}(t)) = [(2\pi)^{p/2}\sqrt{\det R_n}]^{-1} \exp[-(\underline{Z}_t - \underline{h}_t[\underline{x}(t)])^T R_n^{-1} (\underline{Z}_t - \underline{h}_t[\underline{x}(t)])/2]$$

Then the Fisher information matrix is [63]

$$\begin{aligned} J(t) &= E\left\{ \left[\frac{\partial}{\partial \underline{x}(t)} \ln f\left(\underline{Z}_t \mid \underline{x}(t)\right)\right]^T \left[\frac{\partial}{\partial \underline{x}(t)} \ln f\left(\underline{Z}_t \mid \underline{x}(t)\right)\right] \middle| \underline{x}(t) \right\} \\ &= \left(\frac{\partial \underline{h}_t[\underline{x}(t)]}{\partial \underline{x}(t)}\right)^T R_n^{-1} \left(\frac{\partial \underline{h}_t[\underline{x}(t)]}{\partial \underline{x}(t)}\right) \end{aligned} \tag{3.96}$$

where E is the expectation operator.

3.7.2 Cramér-Rao inequality

Assume that $J(t)$ is nonsingular and $\hat{\underline{x}}(t)$ is any unbiased estimator of $\underline{x}(t)$. Then the Cramér-Rao inequality is

$$E\left[(\underline{x}(t)-\hat{\underline{x}}(t))(\underline{x}(t)-\hat{\underline{x}}(t))^T\middle|\underline{x}(t)\right] \geq [J(t)]^{-1} \tag{3.97}$$

Because the inequality in (3.97) holds in the sense that the difference of the left and the right hand sides is non-negative definite, the i^{th} diagonal element of $[J(t)]^{-1}$ gives a lower bound for the variance of the i^{th} component of $\hat{x}(t)$ conditioned on $\underline{x}(t)$. These lower bounds are the Cramér-Rao lower bounds. Expressions available [51] for CRLB of biased estimators need to know the bias term explicitly as a function of parameters. In our problem, we do not have explicit expressions for the bias term. Hence we are forced to use the CRLB for unbiased estimators.

In our motion model, $J(t)$ is generally nonsingular. However, in some degenerate motions, for instance, in the case of constant angular velocity motion, $J(t)$ becomes singular. In fact, there is a close relation between the uniqueness of motion parameters and the singularity of $J(t)$. If $J(t)$ is singular, the above inequality needs to be modified. Let

- $\underline{x}'(t) \equiv$ the vector obtained from deleting the j^{th} component of $\underline{x}(t)$;
- $\hat{\underline{x}}'(t) \equiv$ any unbiased estimator of $\underline{x}'(t)$;
- $J'(t) \equiv$ the square matrix obtained from deleting the j^{th} row and column of $J(t)$.

If $J'(t)$ is nonsingular, the modified Cramér–Rao inequality is

$$E\left[(\underline{x}'(t)-\hat{\underline{x}}'(t))(\underline{x}'(t)-\hat{\underline{x}}'(t))^T\middle|\underline{x}(t)\right] \geq [J'(t)]^{-1} \tag{3.98}$$

This can be obtained by following exactly the same derivation of Cramér-Rao inequality in [63] except starting from the equality $E[(\underline{x}'-\hat{\underline{x}}')|\underline{x}]=0$. Similarly, if $\underline{x}'(t)$ and J' are obtained by deleting more than one component in $\underline{x}(t)$, (3.98) still holds. We state and prove that component deletion leads to looser bounds.

Lemma 8.1 : If neither $J(t)$ nor $J'(t)$ is singular, the CRLB obtained by (3.98) is looser (smaller) than or equal to that obtained by (3.97).

Proof :

The time dependence is being suppressed. Assume that $\underline{x}'$ is obtained from deleting k components of $\underline{x}$. Because we can always rearrange the order of the elements in $\underline{x}$, without loss of generality, it is assumed that the last k elements of $\underline{x}$ are deleted. That is, J has the form

$$J = \begin{pmatrix} J' & B \\ B^T & C \end{pmatrix},$$

where C is a nonsingular $k \times k$ matrix. Also note that if

$$J^{-1} = \begin{pmatrix} J' & B \\ B^T & C \end{pmatrix}^{-1} \equiv \begin{pmatrix} D & E \\ E^T & F \end{pmatrix},$$

where F is a $k \times k$ matrix, then [63]

$$D = J'^{-1} + J'^{-1} B F B^T J'^{-1} \tag{3.99}$$

J is positive definite, hence J^{-1} is also positive definite. F is a square submatrix of J^{-1} along the diagonal, thus F is positive definite. There exists a $k \times k$ matrix G such that $F = GG^T$ [65]. Therefore (3.99) becomes

$$D - J'^{-1} = (J'^{-1} B G)(J'^{-1} B G)^T$$

which implies that the matrix $(D - J'^{-1})$ is positive semi-definite. Thus the diagonal elements of D are no less than the corresponding ones of J'^{-1}. As a result, the lower bound given by J'^{-1} is no tighter than that by J^{-1}.

3.7.3 Simulations and Estimator Efficiency

Simulation I

For the convenience of demonstrating the theoretical results in this section, a simple motion of zero precession with known object structure is investigated using computer simulations. The estimator applied is an IEKF. The results are from 50 Monte Carlo trials. The plant model (3.32) and the state transition equation (3.34) can be simplified by setting $\underline{p} = \underline{0}$ and then removing the precession parameters from the state vector. The angular velocity is $\underline{w} = (0.2, 0.15, 0.2)^T$. Other motion parameters and camera set-up parameters are the same as those in Section 3.4. The initial guesses are zeros except that $q_4 = 1$ and $\underline{w} = (0.3, 0.1, 0.3)^T$.

The IEKF bias is shown in Figure 3.5, in which the ordinate equals the difference between the sample mean of the estimated states from 50 Monte Carlo trials and the true states. Since the bias is small, the CRLB for the unbiased estimators is expected to give a reasonable lower bound for the estimates.

In the calculation of the Fisher information matrix, the standard deviations of the added noise in (3.33) are set to $\sigma_x = 0.1129$ for $n_{xj}(t_i)$, $\sigma_y = 0.1106$ for $n_{yj}(t_i)$, and $\sigma_z = 0.2971$ for $n_{zj}(t_i)$. The above noise parameters are actually a set of sample standard deviations of errors in the 3-D inertial coordinates obtained by the stereo triangulation technique mentioned in Section 3.4 using the camera set-up parameters there.

The Fisher information matrix $J(t)$ is singular in this experiment. Any line deletion of $J(t)$ corresponding to one component in $\underline{r}(t)$ or $\underline{d}$ gives a nonsingular $J'(t)$. But the line deletions corresponding to components other than $\underline{r}(t)$ and $\underline{d}$ still give singular $J'(t)$. Singularity of $J(t)$ is closely related to the uniqueness of motion parameters.

After the first row and column of $J(t)$ are deleted, the calculated CRLB and sample mean square errors are compared in Figure 3.6. Recall that CRLB is only a lower bound, not necessary the greatest lower bound. As the CRLB and the sample mean square error are quite close, it is believed that the CRLB give a reasonably tight bound for the variances of the estimators of states and that IEKF is quite efficient.

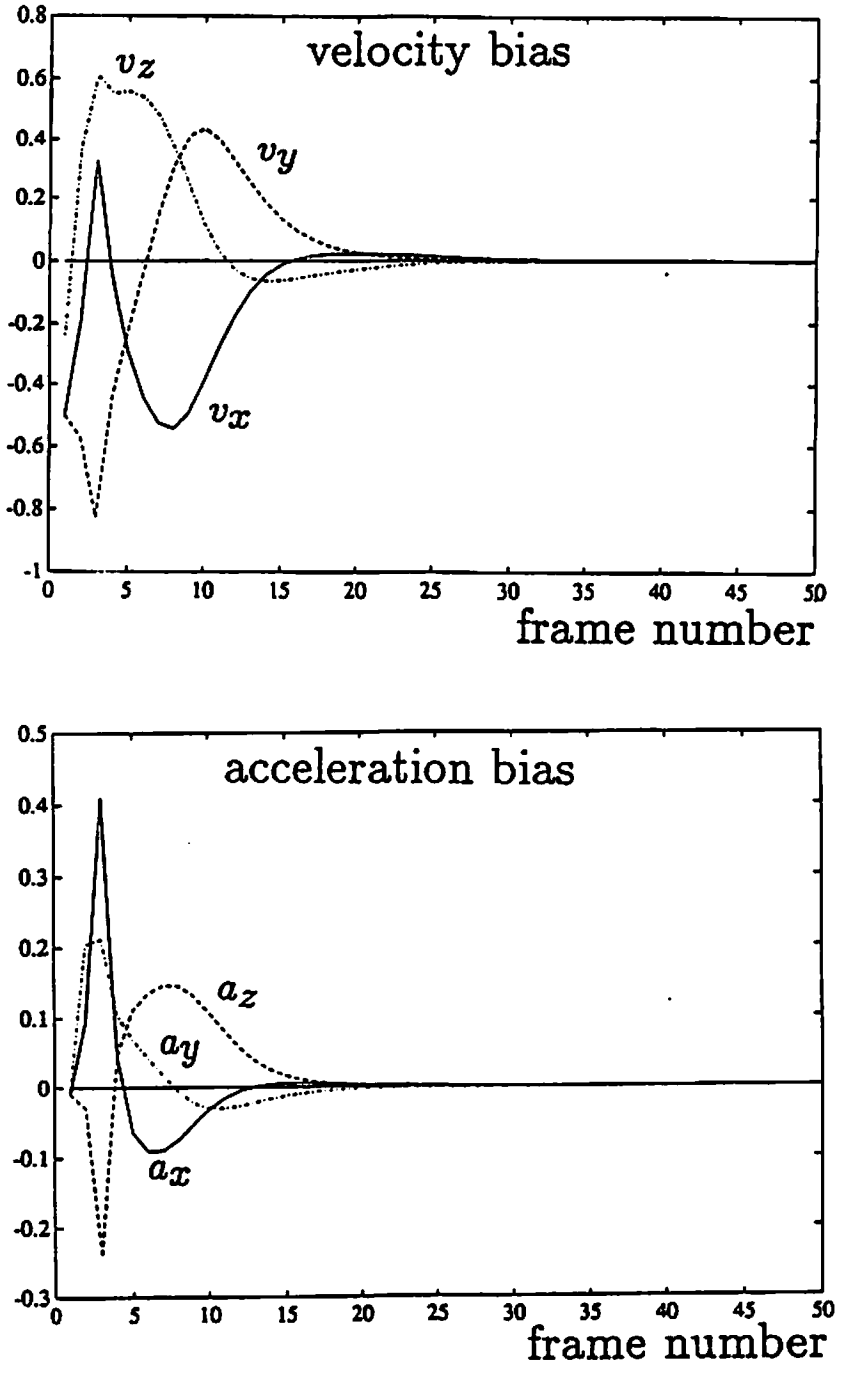
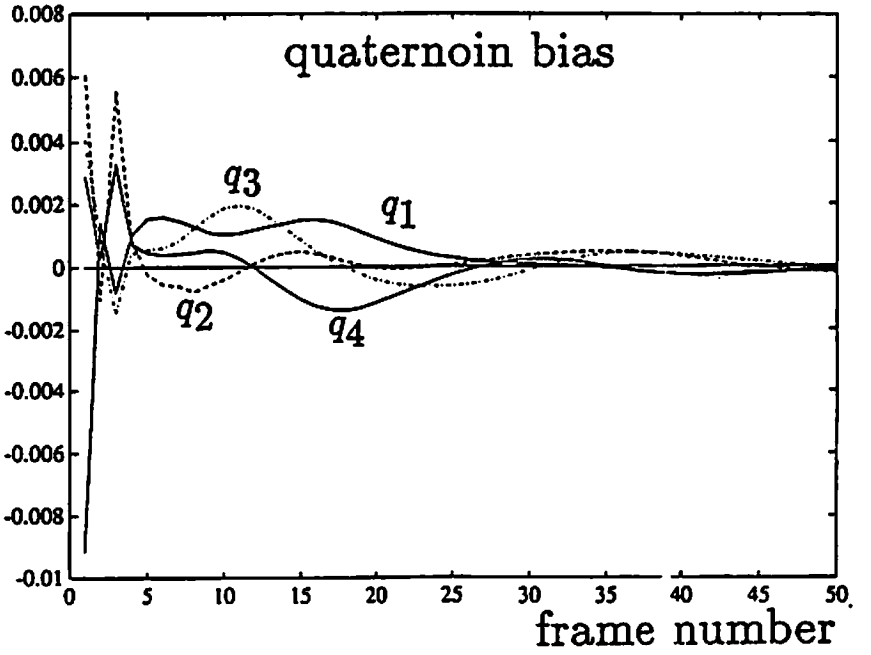


Figure 3.5: Bias in the estimates of motion parameters from 50 Monte Carlo trials for the known structure, constant acceleration and angular velocity motion model. (a) velocity bias and (b) acceleration bias.



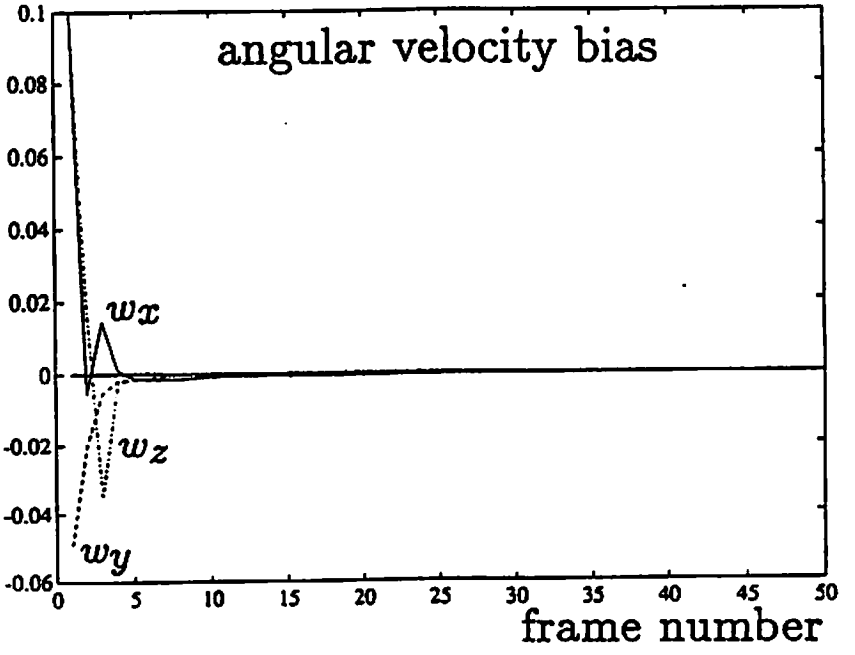


Figure 3.5: (continued) (c) quaternion bias and (d) angular velocity bias.

In Figure 3.6, the mean square errors at the 5^{th} and 6^{th} frames are lower than CRLB. This is due to the prior information contained in the initial guess of IEKF. Since the formulation of CRLB in Section 3.7.1 and 3.7.2 considers only the information in the image measurements, the calculated CRLB constitutes a performance bound when no prior information is used. However, it is well known that the initial guess is the prior information of the motion parameters in IEKF. Figure 3.6 shows that the prior information stabilizes the estimates so that the mean square error is lower than the computed CRLB in the first a couple of frames. When more frames becomes available, the prior information is outweighed by the information in the image measurements, and IEKF behaves quite close to that predicted by CRLB. If the filter is started from random initial guesses, which correspond to the case when no prior information is available, divergence is observed for largely deviant initial guesses. As in practice the totally random initial guesses are seldom used, the corresponding simulation is not considered here. Besides, the possible divergence also prevents us from running a direct Monte Carlo trial.

Simulation II

In the second simulation, we consider the motion of constant acceleration and precession. The parameters used are the same as in Section 3.4, including the ground truth, the camera set-up, and the initial guess. The bias and mean square error are obtained from 50 Monte Carlo trials. Figure 3.7 shows the IEKF bias.

CRLB is computed using (3.96) and (3.97). While CRLB thus computed is for the known structure case, it still constitutes a lower bound for the unknown structure case since the latter uses less information. The IEKF mean square error and CRLB are compared in Figure 3.8. One may see that the mean square error does not follow CRLB closely as in previous simulation, while the IEKF converges satisfactorily. This deviation results from the degradation of IEKF by the uncertainty in structure parameters estimated prior to being fed into the IEKF. If the uncertain parameters cause the filter to diverge, Schmidt-Kalman filters [43] may be applied with increased computational requirements. However, our

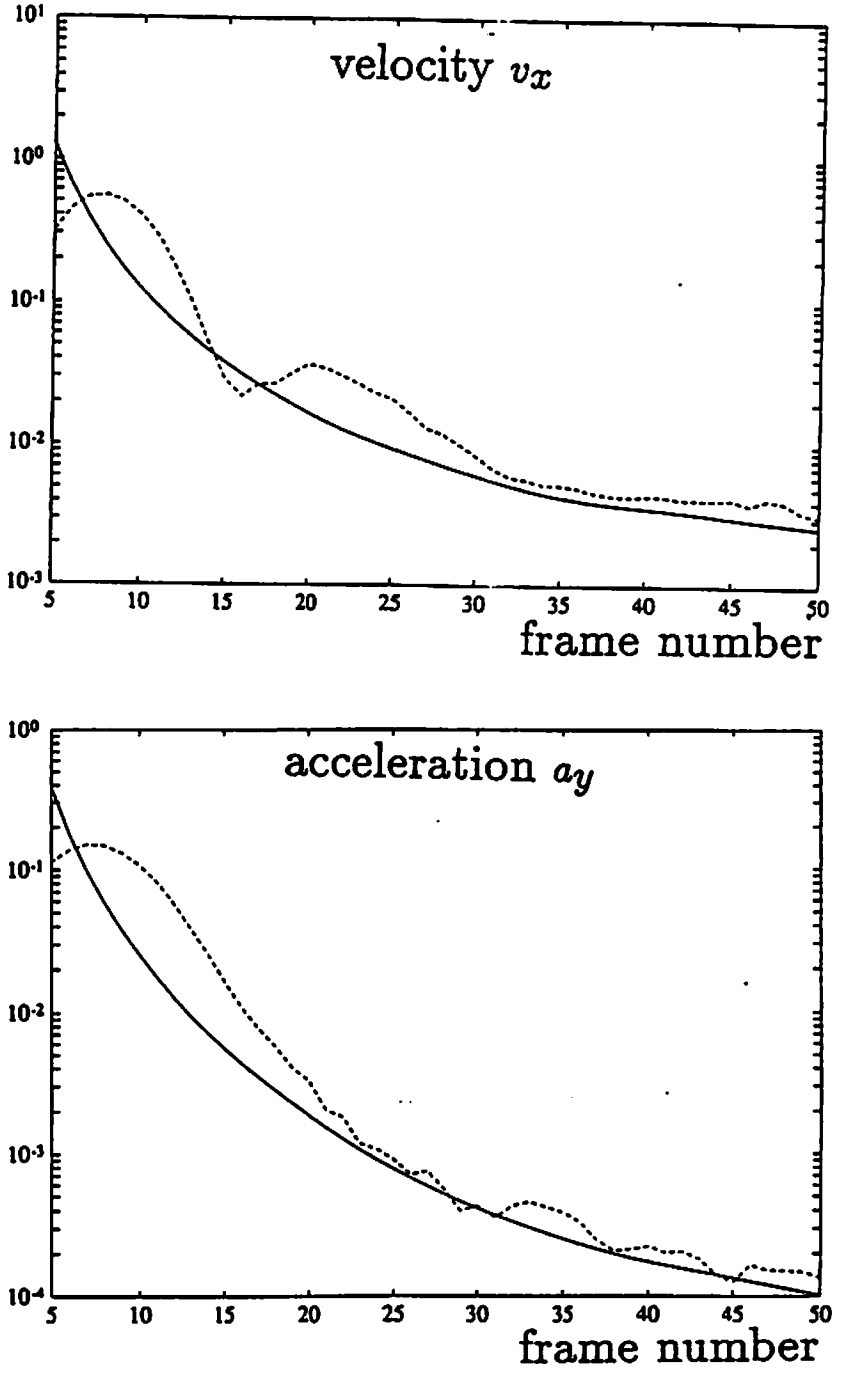
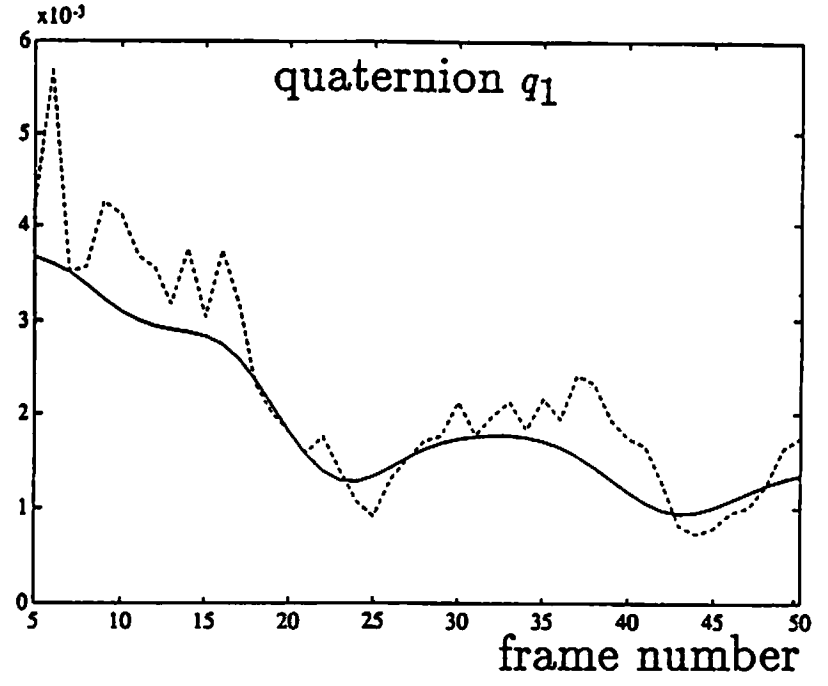


Figure 3.6: Square root of CRLB (solid) and root mean square error (dashed) from 50 Monte Carlo trials for the known structure, constant acceleration and angular velocity motion model. (a) velocity component v_x and (b) acceleration component a_y.



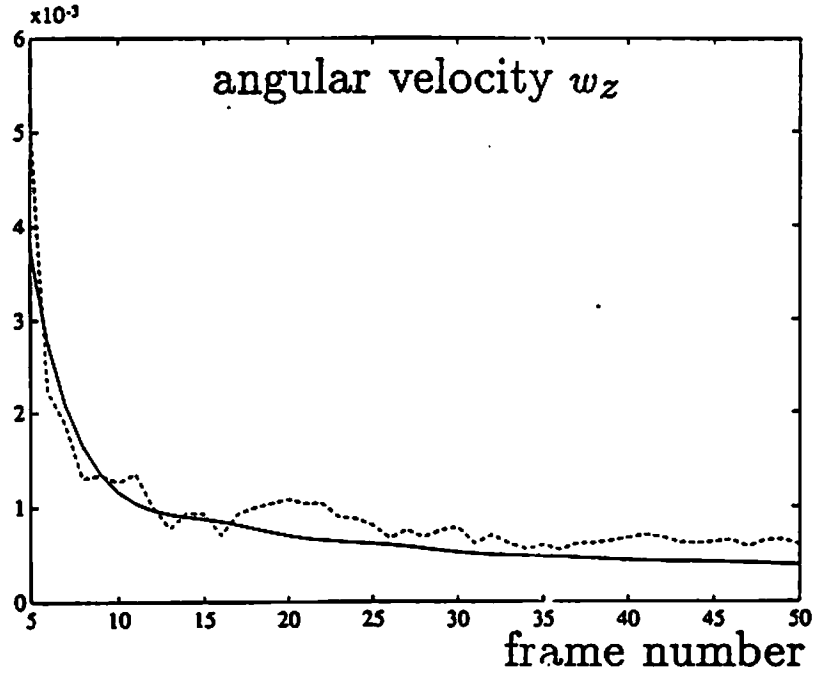


Figure 3.6: (continued) (c) quaternion component q_1 and (d) angular velocity component w_z.

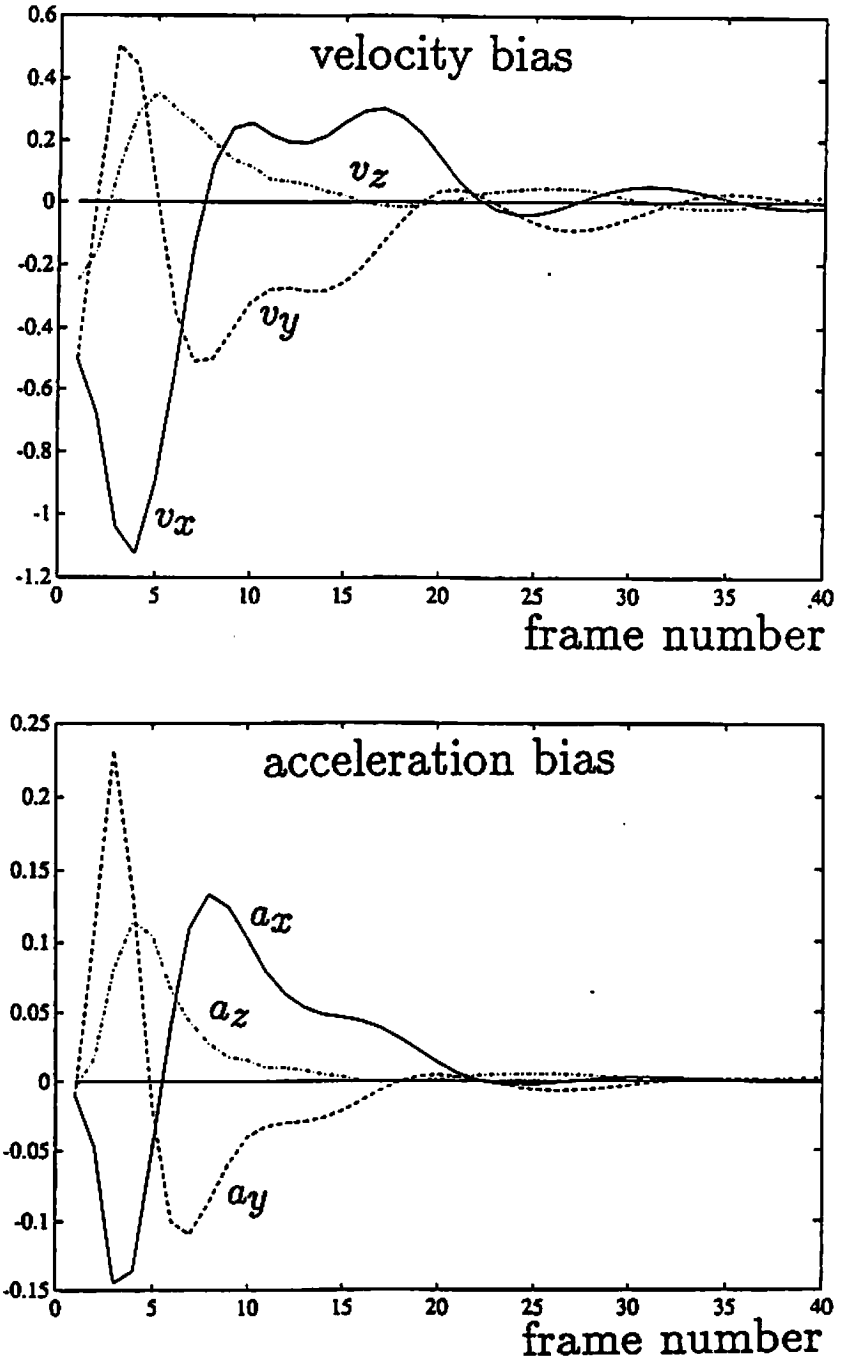


Figure 3.7: Bias in the estimates of motion parameters from 50 Monte Carlo trials for the unknown structure, constant acceleration and precession motion model. (a) velocity bias and (b) acceleration bias.

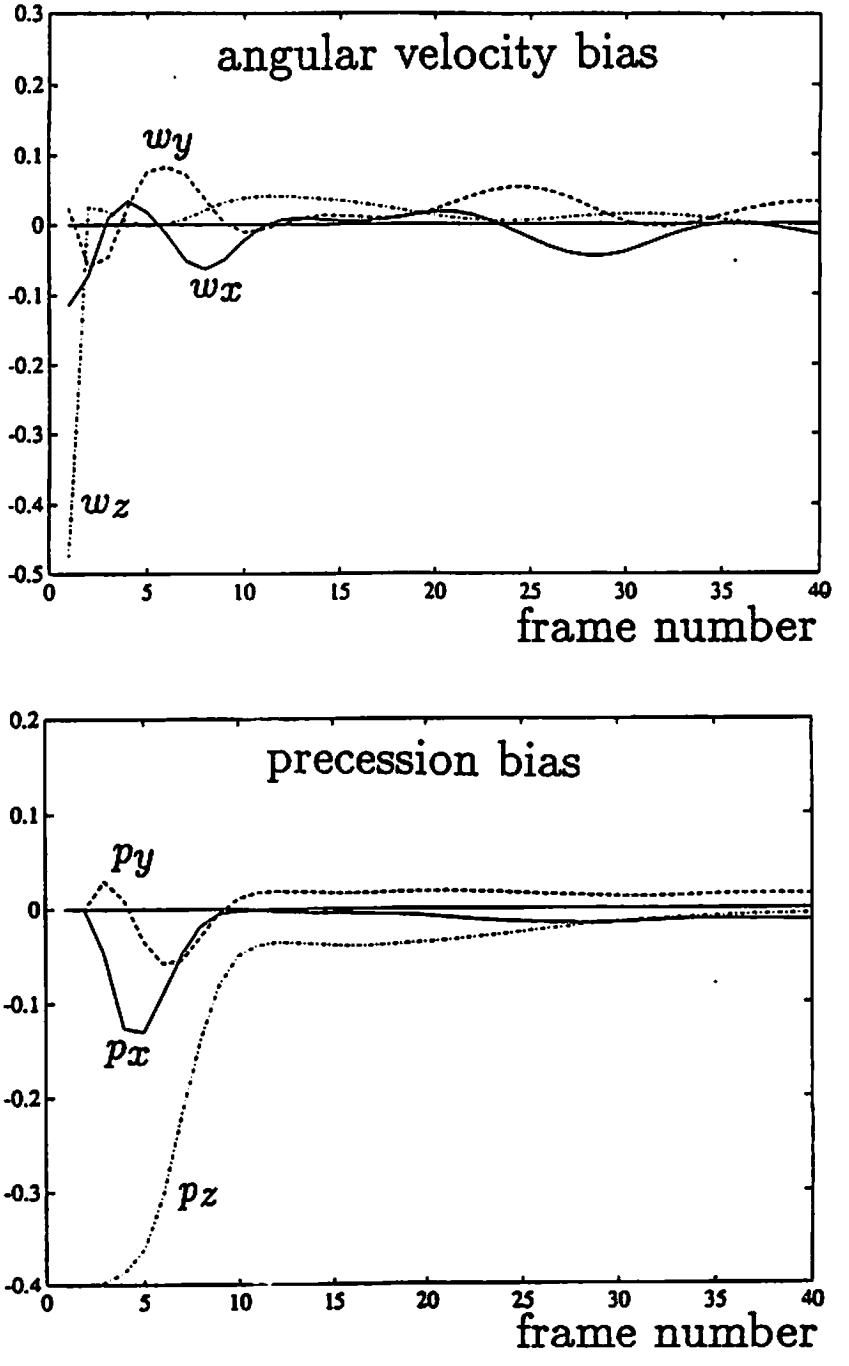


Figure 3.7: (continued) (c) angular velocity bias and (d) precession bias.

experiments show that the performance of IEKF is not degraded to a significant extent by the uncertainty in the structure parameters.

The Fisher information matrix is singular when only first four stereo pairs are used. The singularity of Fisher information matrix shows that three noncollinear feature points in four consecutive stereo pairs are not sufficient to decide all motion parameters in our model.

3.7.4 Why Are Variance and CRLB of Quaternion Not Monotonically Decreasing with Number of Frames?

Ripples are observed in both sample mean square error and CRLB curves of $\underline{q}(t_i)$ as shown in Figure 3.6. The ripples in these two curves are also consistent with each other. This phenomenon appears to be contradicting the notion that as more measurements become available, the added information should reduce the uncertainty of estimates of motion parameters, and thus both CRLB and mean square error curves should be monotonically decreasing. This paradox is explained below.

Consider the first term in the Fisher information matrix (3.96),

$$\begin{aligned}\frac{\partial}{\partial \underline{x}(t)} \ln f\left(\underline{Z}_t \mid \underline{x}(t)\right) &= \left[\frac{\partial}{\partial \underline{x}(0)} \ln f\left(\underline{Z}_t \mid \underline{x}(t)\right)\right] \frac{\partial \underline{x}(0)}{\partial \underline{x}(t)} \\ &= \left[\frac{\partial}{\partial \underline{x}(0)} \ln f\left(\underline{Z}_t \mid \underline{x}(0)\right)\right] \frac{\partial \underline{x}(0)}{\partial \underline{x}(t)}\end{aligned}$$

where the second equality holds because we can replace the condition $\underline{x}(t)$ in the conditional probability density function by $\underline{x}(0)$ dut to the one-to-one deterministic relation (3.34) between $\underline{x}(0)$ and $\underline{x}(t)$. Thus (3.96) yields

$$J^{-1}(t) = \left(\frac{\partial \underline{x}(t)}{\partial \underline{x}(0)}\right) \cdot \underbrace{E\left\{\left[\frac{\partial}{\partial \underline{x}(0)} \ln f\left(\underline{Z}_t \mid \underline{x}(0)\right)\right]^T \left[\frac{\partial}{\partial \underline{x}(0)} \ln f\left(\underline{Z}_t \mid \underline{x}(0)\right)\right] \middle| \underline{x}(0)\right\}^{-1}}_{\equiv Q^{-1}(t)} \left(\frac{\partial \underline{x}(t)}{\partial \underline{x}(0)}\right)^T \tag{3.100}$$

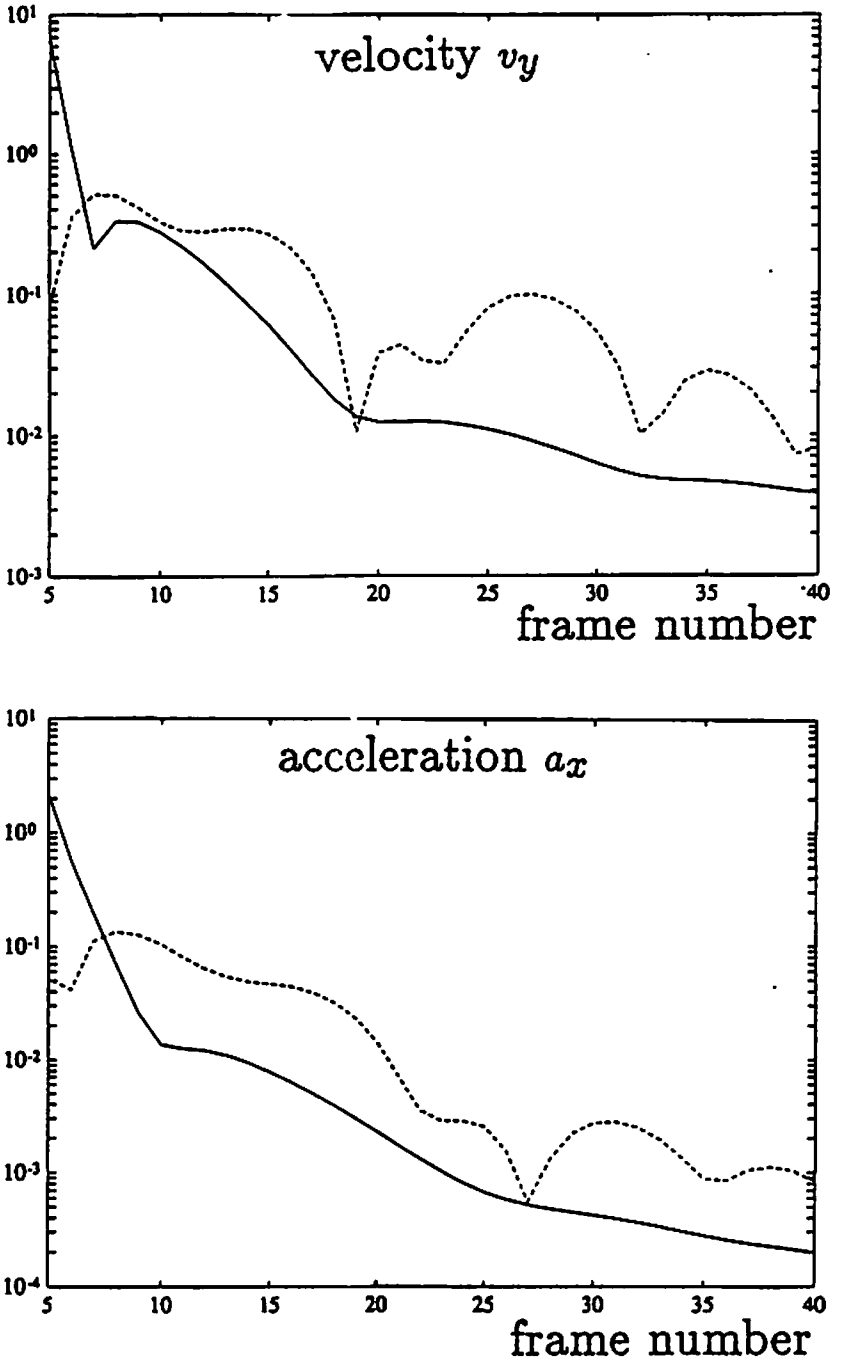


Figure 3.8: Square root of CRLB (solid) and root mean square error (dashed) from 50 Monte Carlo trials for the unknown structure, constant acceleration and precession motion model. (a) velocity component v_y and (b) acceleration component a_x.

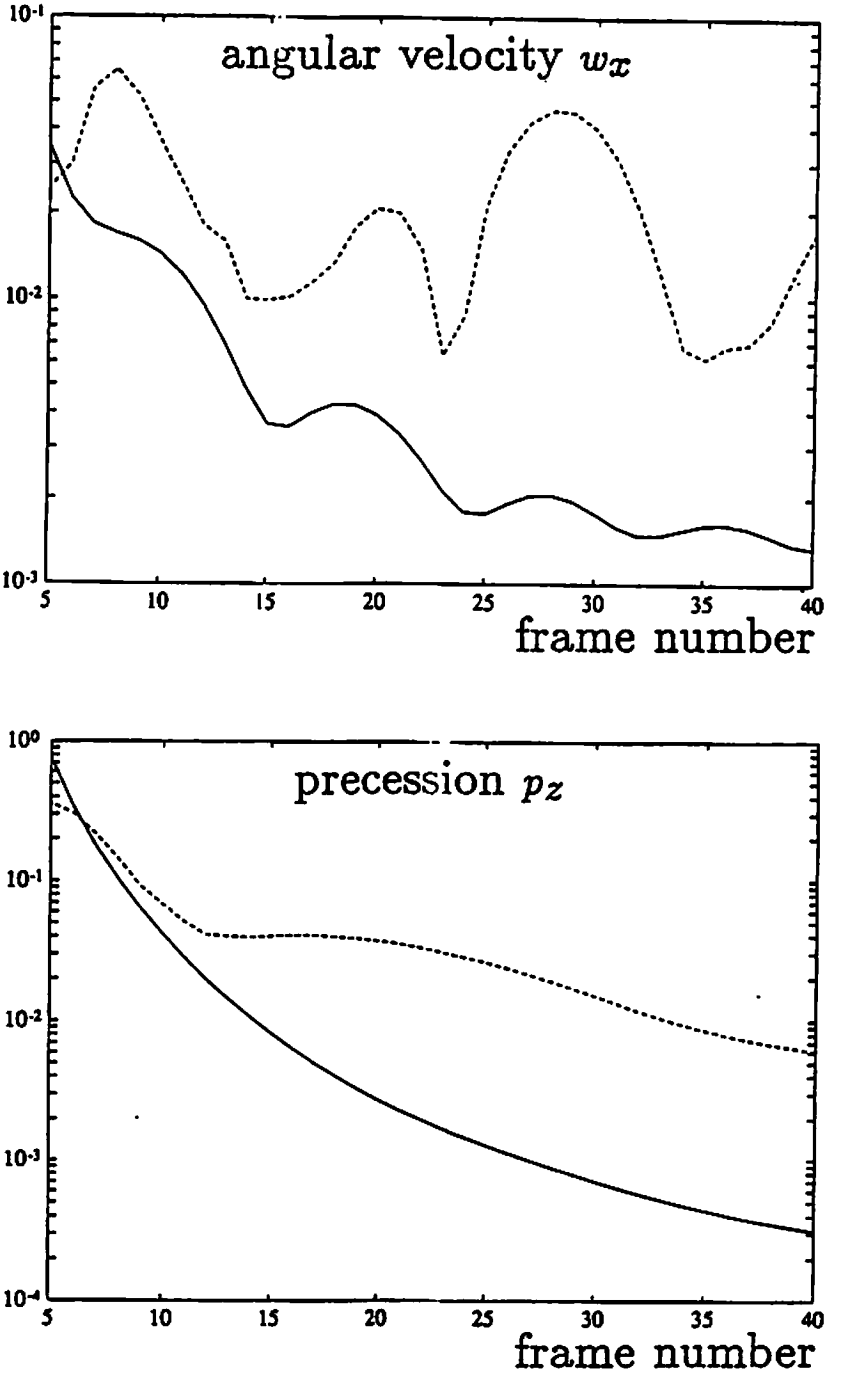


Figure 3.8: (continued) (c) angular velocity component w_x and (d) precession component p_z.

in which we have applied the fact that the Jacobian matrix is nonsingular due to the one-to-one transformation (3.34) from $\underline{x}(0)$ to $\underline{x}(t)$ and has the inverse matrix

$$\left(\frac{\partial \underline{x}(0)}{\partial \underline{x}(t)}\right)^{-1} = \frac{\partial \underline{x}(t)}{\partial \underline{x}(0)}$$

Let us note that the CRLB of $\underline{x}(t)$ based on the measurements no later than t are the diagonal terms of $J^{-1}(t)$, and the CRLB of $\underline{x}(0)$ based on the same measurements are the diagonal terms of $Q^{-1}(t)$. As more measurements become available, the added information helps the system to reduce the uncertainty in the time-invariant motion parameters $\underline{x}(0)$. Thus the diagonal terms of $Q^{-1}(t)$ decrease monotonically in t. However, by (3.100), $J^{-1}(t)$ is a time-varying congruent transformation of $Q^{-1}(t)$. The trend of the diagonal terms of $J^{-1}(t)$ in t is thus the monotonicity of the diagonal terms of $Q^{-1}(t)$ in t twisted by the time-varying congruent transformation. Therefore the CRLB of the time varying motion parameter vector $\underline{x}(t)$ is not necessarily decreasing monotonically in time.

In our experiment, the CRLB of quaternions shown in Figure 3.6 has a decreasing tendency with ripples on it. The decrease is due to the monotonicity of the diagonal terms of $Q^{-1}(t)$. The ripples can be understood as a twist of the monotonicity caused by the congruent transformation, as we will see in the following. To check the congruent transformation, we can check the transformation from $\underline{q}(0)$ to $\underline{q}(t)$ since, by (3.100), the congruent transformation is uniquely determined by the one-to-one transformation from $\underline{x}(0)$ to $\underline{x}(t)$. Because $\underline{p} = \underline{0}$ in Simulation I, (3.18) is reduced to

$$\underline{q}(t) = \Phi[\underline{w}; t]\underline{q}(0)$$

where the transformation matrix $\Phi[\underline{w}; t]$ has period $4\pi/|\underline{w}|$ by (3.19). Furthermore, within one period of $\Phi[\underline{w}, t]$, the matrix Φ in the second half period assumes the negative value of Φ in the first half period. Therefore the "scaling" from the space of $\underline{q}(0)$ into the space of $\underline{q}(t)$ has period $2\pi/|\underline{w}|$. On the other hand, the ripple period 19.5 observed in Figure 3.6 is close to $2\pi/|\underline{w}|$, where $\underline{w} = (0.2, 0.15, 0.2)^T$ (which corresponds to a rotation going through one full

turn per 19.6 frames). Hence the appearance of monotonicity with ripples in Figure 3.6 agrees with the above discussion.

3.8 Motion of Constant Velocity And Angular Velocity

If simple motions are considered, instead of applying a constant acceleration and precession model, one may find that a motion model of lower order, such as the model of constant velocity and angular velocity, is satisfactory to describe the underlying 3-D motions. As these lower order motions are special cases of the motion of constant acceleration and precession, the plant equation (3.32), measurement equation (3.33), and state transition equation (3.34) for the models of constant acceleration and precession can be readily applied to estimate the 3-D motions. However, if the degree of freedom in the motion parameters can be reduced, the following intuitively defined overestimation [5] will be improved.

$$\text{overestimation} = \frac{\text{number of measurements}}{\text{number of motion and structure parameters}}$$

In fact, one may derive the plant, measurement, and state transition equations which have the proper degree of freedom for the lower order motions directly from (3.32), (3.33), and (3.34). For instance, if the motion of constant velocity and angular velocity is considered, we can first set

$$\underline{a} = \underline{p} = 0$$

in (3.32), (3.33), and (3.34). Then remove the 7^{th}, 8^{th}, and 9^{th} rows, which correspond to $\underline{a}$, and the last 3 rows, which correspond to $\underline{p}$, from (3.32) and (3.34). The three equations thus obtained are the plant, measurement, and state transition equations in the EKF/IEKF with the proper degree of freedom for the estimation of a constant velocity and angular velocity motion.

Chapter 4

A Simple Kinematic Model Based Approach for 3–D Motion And Structure Estimation

If 3–D motion and structure are estimated simultaneously, we show in this chapter that the standard 3×3 rotation matrix is more suitable for representing rotational motion than quaternions. A new formulation for constant velocity translation and constant angular velocity rotation motion using the standard rotation matrix is given. Both monocular and binocular systems are considered. Batch and recursive algorithms are designed to search and track, respectively, the moving rigid object. Occlusion among image frames as well as between each stereo pair in the binocular case is dealt with. The approach is based on representing the constant velocity translation and constant angular velocity motion using nine rectilinear motion parameters, which are the 3–D vectors of initial position, linear velocity, and angular velocity. The structure of the moving object is represented by the coordinates of feature points in a 3–D coordinate system fixed on the object. A nonlinear least squares method is used to formulate the batch estimation of motion and structure parameters. Owing to the nonlinearity in perspective projection, a conjugate gradient method is used to search the minimum of the least squares formulation. Nonlinear Kalman filters are used in the implementation of the recursive algorithms. Linear plant models are used in the

recursive filters to propagate motion in closed form, and thus avoiding the time-consuming numerical integration step in the nonlinear filters. Images taken at nonuniform instances are also considered. Simulation results are included.

In Section 4.1, the imaging models for a monocular as well as a stereo image sequences are introduced. The kinematic model of constant translation and constant angular velocity are discussed in Section 4.2. Section 4.3 and Section 4.4 formulate the batch and recursive algorithms, respectively. Simulation results are given in Section 4.5.

4.1 Imaging Model

4.1.1 Monocular Imaging Model

As shown in Fig. 4.1, a static camera is looking at a changing environment. A 3-D reference coordinate system I is fixed with respect to the camera, with the z-axis pointing along the optical axis. The focal length is f. Hence the image plane is $z = f$. The 2-D image coordinate system on the image plane has its X- and Y- axes parallel to the x- and y- axes of I, respectively. Suppose at time t the i^{th} feature point P_i has the 3-D coordinates

$$\underline{s}_{Ii}(t) \equiv (x_i(t), y_i(t), z_i(t))^T \tag{4.1}$$

in I, and $(X_i(t), Y_i(t))^T$ on the image plane. If noise in the image coordinates are taken into account, the central projection of the pin-hole camera model results in the following imaging model [18, 20, 21],

$$\begin{cases} X_i(t) &= f\,\dfrac{x_i(t)}{z_i(t)} + n_{Xi}(t) \\ Y_i(t) &= f\,\dfrac{y_i(t)}{z_i(t)} + n_{Yi}(t) \end{cases} \tag{4.2}$$

where n.'s are the additive noise processes contaminating the measurements of locations of feature points on the image plane.

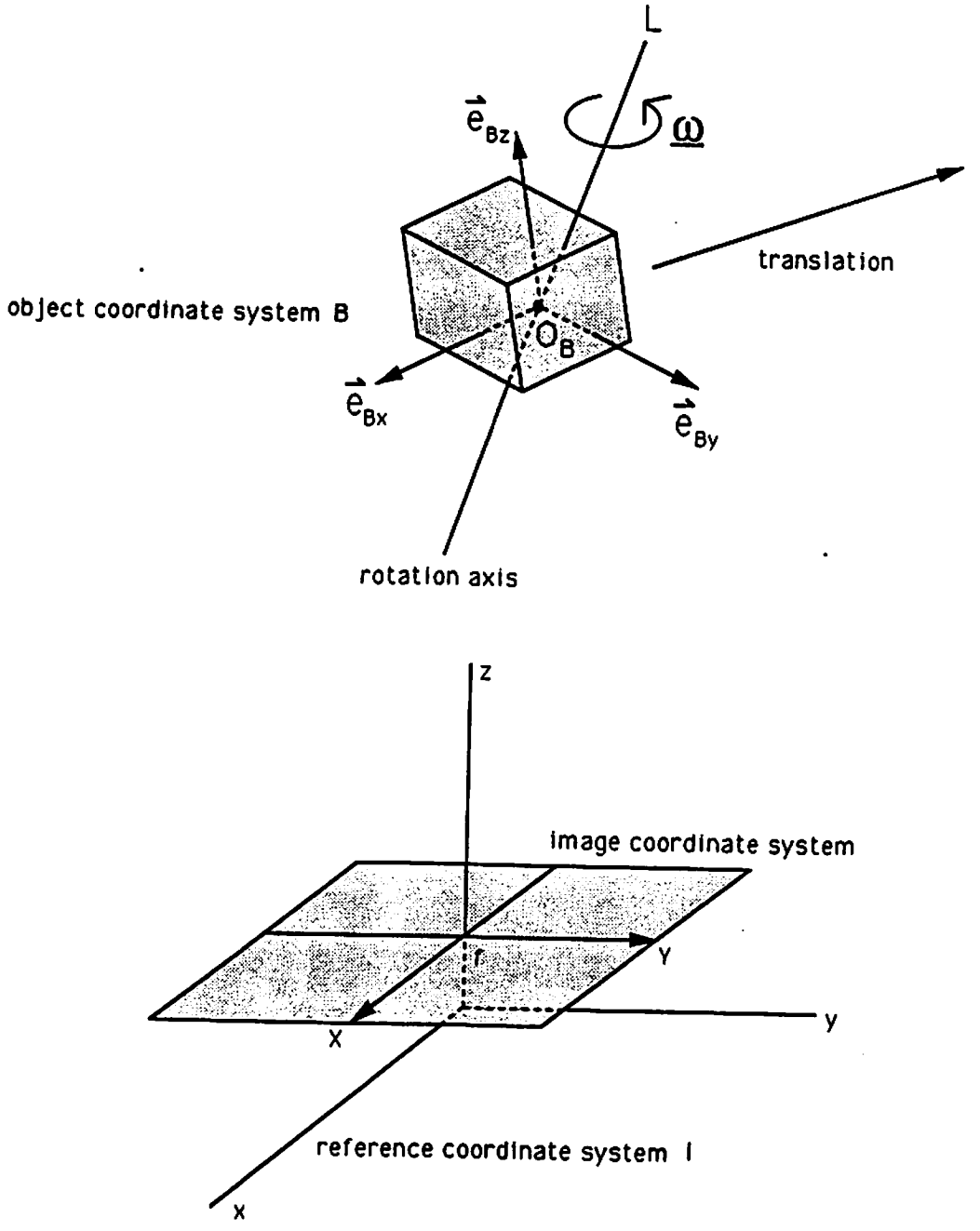


Figure 4.1: Geometry of motion and monocular measurement models

4.1.2 Binocular Imaging Model

As shown in Fig. 4.2, two cameras with parallel image planes are looking towards the same direction. A 3-D reference coordinate system I is located at the middle point of the camera pair, with its z-axes pointing along the optical axes. The baseline, which is the distance between the cameras, is $2b$. Suppose at time t the i^{th} feature point P_i has the image coordinates $(X_{li}(t), Y_{li}(t))^T$ and $(X_{ri}(t), Y_{ri}(t))^T$ on image planes of the left and the right camera, respectively. Then the central projection model leads to

$$\begin{cases} X_{li}(t) &= f\,\dfrac{x_i(t)+b}{z_i(t)} + n_{Xli}(t) \\ Y_{li}(t) &= f\,\dfrac{y_i(t)}{z_i(t)} + n_{Yli}(t) \end{cases} \tag{4.3}$$

and

$$\begin{cases} X_{ri}(t) &= f\,\dfrac{x_i(t)-b}{z_i(t)} + n_{Xri}(t) \\ Y_{ri}(t) &= f\,\dfrac{y_i(t)}{z_i(t)} + n_{Yri}(t) \end{cases} \tag{4.4}$$

where n.'s are the noise processes.

4.2 Object and Motion Model

The object being tracked is assumed to be rigid. The motion of the rigid object is decomposed into a rotation about a point O_B fixed on the object and a translation of the rotation center O_B [18, 20, 21]. Consider an object coordinate system B which is fixed on the rigid object and has the origin located at the rotation center O_B, as shown in Fig. 4.1. Denote the trajectory of O_B in the reference coordinate system I by

$$\underline{r}_{OB}(t) \equiv (r_{OBx}(t), r_{OBy}(t), r_{OBz}(t))^T \tag{4.5}$$

If the i^{th} feature point P_i has coordinates $\underline{s}_{Ii}(t)$ in I and

$$\underline{s}_{Bi} \equiv (s_{Bix}, s_{Biy}, s_{Biz})^T \tag{4.6}$$

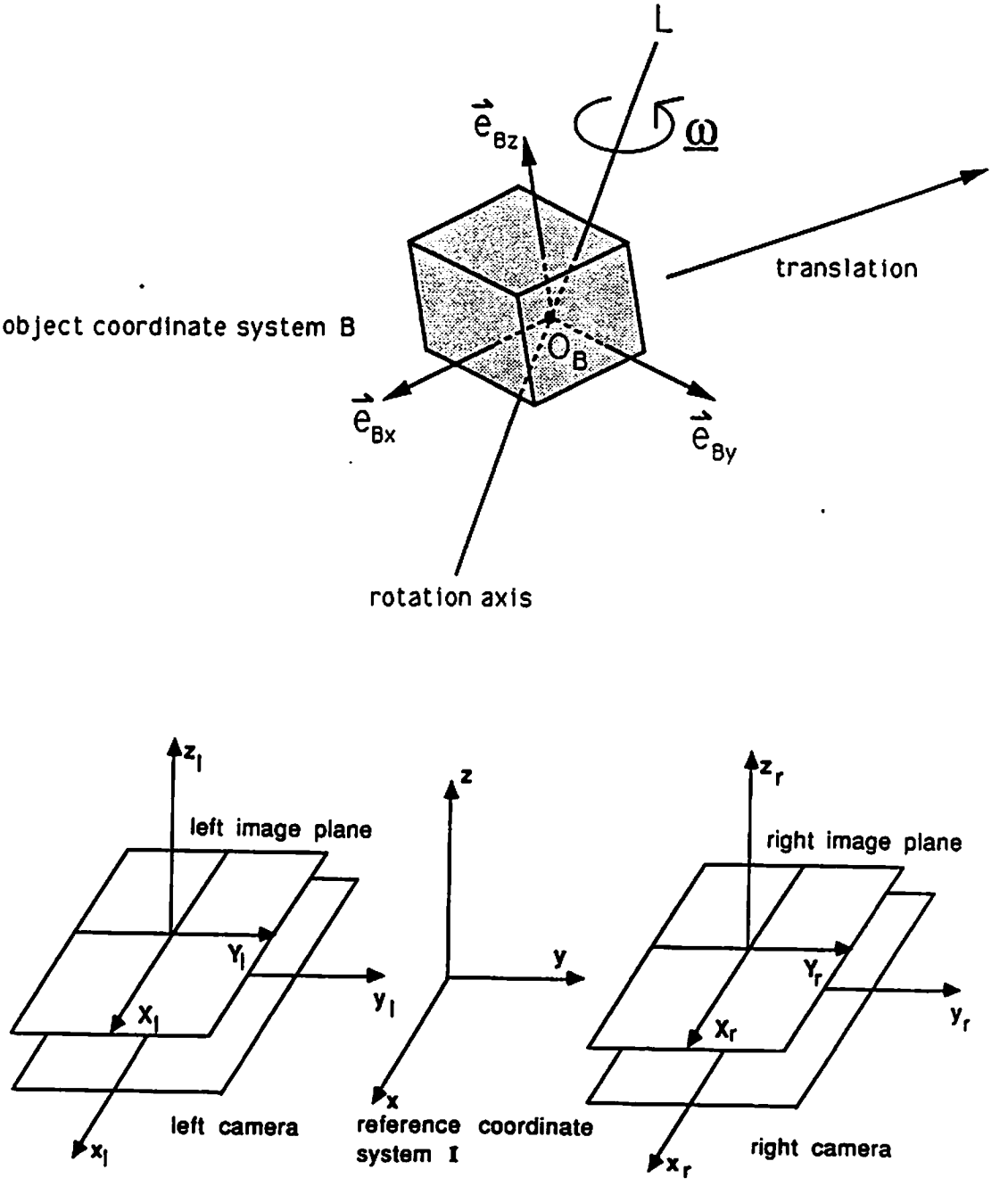


Figure 4.2: Geometry of binocular measurement model

in B, the decomposition of rotation and translation gives

$$\underline{s}_{Ii}(t) = \underline{r}_{OB}(t) + R(t)\ \underline{s}_{Bi} \tag{4.7}$$

where $R(t)$ is the 3×3 coordinate transformation matrix which aligns the object coordinate system B to the reference coordinate system I. Due the rigidity assumption, $\underline{s}_{Bi}$ is not varying in time. The time-invariant components of $\underline{s}_{Bi}$ are called the *structure parameters* as they describe the position of a feature point on the object.

4.2.1 Translational Motion Model

The translational motion is assumed to be of constant velocity. That is, it is assumed that we can find a rotation center O_B whose velocity

$$\begin{aligned} \frac{d}{dt}\,\underline{r}_{OB}(t) &\equiv \dot{\underline{r}}_{OB} \\ &\equiv (\dot{r}_{OBx}, \dot{r}_{OBy}, \dot{r}_{OBz})^T \end{aligned} \tag{4.8}$$

is constant in the reference system I. Integration of (4.8) yields

$$\underline{r}_{OB}(t) = \underline{r}_{OB}(t_0) + (t - t_0)\ \dot{\underline{r}}_{OB} \tag{4.9}$$

As this equation shows that the translational motion $\underline{r}_{OB}(t)$ is parameterized by $\underline{r}_{OB}(t_0)$ and $\dot{\underline{r}}_{OB}$, the 6 components in $\underline{r}_{OB}(t_0)$ and $\dot{\underline{r}}_{OB}$ are called the *translational motion parameters.*

	in B	in I
arrow head of $\vec{e}_{Bx}$	$(1,0,0)^T$	$\underline{r}_{OB}(t) + R(t)\ (1,0,0)^T$
arrow end of $\vec{e}_{Bx}$	$\underline{0}$	$\underline{r}_{OB}(t)$

Table 4.1: The coordinates of the arrow head and end of $\vec{e}_{Bx}$

4.2.2 Rotational Motion Model

The rotational motion is assumed to be of constant angular velocity. That is, it is assumed that the object is rotating with respect to the rotation center O_B with a constant angular velocity $\underline{w}$ expressed in I.

To parameterize the rotational motion, let us consider the unit directional vector $\vec{e}_{Bx}$ of the x–axis of the object coordinate system B, as shown in Fig. 4.1. The vector $\vec{e}_{Bx}$ may be visualized as an arrow with its end at the origin O_B of B. The coordinates of the arrow head and end of $\vec{e}_{Bx}$ are listed in Table 4.1. In Table 4.1, the coordinates in I are obtained from the coordinates in B using the coordinate transformation (4.7). Suppose the vector $\vec{e}_{Bx}$ has the 3×1 matrix $\underline{e}_{Bx}(t)$ as its matrix expression in I. Since $\underline{e}_{Bx}(t)$ is the difference between the coordinates of the arrow head and end of $\vec{e}_{Bx}$ in I, using Table 4.1, we have

$$\underline{e}_{Bx}(t) = R(t)\ (1,0,0)^T = \text{ the first column of } R(t)$$

A similar consideration on unit directional vectors along the y– and z– axes of B leads to the result that the second and third columns of $R(t)$ are $\underline{e}_{By}(t)$ and $\underline{e}_{Bz}(t)$, respectively. Therefore

$$R(t) = [\underline{e}_{Bx}(t)\ \ \underline{e}_{By}(t)\ \ \underline{e}_{Bz}(t)] \tag{4.10}$$

Now let us recall a well known fact [37] that the matrix

$$R_0[n_1, n_2, n_3\,;\,\theta]$$
$$\equiv \begin{pmatrix} n_1^2 + (1-n_1^2)c & n_1n_2(1-c) - n_3s & n_1n_3(1-c) + n_2s \\ n_1n_2(1-c) + n_3s & n_2^2 + (1-n_2^2)c & n_2n_3(1-c) - n_1s \\ n_1n_3(1-c) - n_2s & n_2n_3(1-c) + n_1s & n_3^2 + (1-n_3^2)c \end{pmatrix}$$

rotates a 3–D vector representation by an angle θ with respect to the unit directional vector $(n_1, n_2, n_3)^T$, where $c \equiv \cos\theta$ and $s \equiv \sin\theta$. In our model, the rigid body is rotating about O_B with the constant angular velocity $\underline{w} \equiv (w_x, w_y, w_z)^T$, and so is every vector fixed on the rigid body with its arrow end at O_B, e.g. $\vec{e}_{Bx}$, $\vec{e}_{By}$, and $\vec{e}_{Bz}$. As a consequence, within the time period from $t = t_0$ to t, the vectors $\vec{e}_{Bx}$, $\vec{e}_{By}$, and $\vec{e}_{Bz}$ are all rotated by the angle $|\underline{w}|(t - t_0)$ with respect to the unit directional vector

$$\left(\frac{w_x}{|\underline{w}|}, \frac{w_y}{|\underline{w}|}, \frac{w_z}{|\underline{w}|}\right)$$

Therefore

$$\begin{cases} \underline{e}_{Bx}(t) = R_0\left[\frac{w_x}{|\underline{w}|}, \frac{w_y}{|\underline{w}|}, \frac{w_z}{|\underline{w}|} \;;\; |\underline{w}|\,(t - t_0)\right] \underline{e}_{Bx}(t_0) \\ \underline{e}_{By}(t) = R_0\left[\frac{w_x}{|\underline{w}|}, \frac{w_y}{|\underline{w}|}, \frac{w_z}{|\underline{w}|} \;;\; |\underline{w}|\,(t - t_0)\right] \underline{e}_{By}(t_0) \\ \underline{e}_{Bz}(t) = R_0\left[\frac{w_x}{|\underline{w}|}, \frac{w_y}{|\underline{w}|}, \frac{w_z}{|\underline{w}|} \;;\; |\underline{w}|\,(t - t_0)\right] \underline{e}_{Bz}(t_0) \end{cases} \tag{4.11}$$

Using (4.10), (4.11) can be written as

$$R(t) = R_0\left[\frac{w_x}{|\underline{w}|}, \frac{w_y}{|\underline{w}|}, \frac{w_z}{|\underline{w}|} \;;\; |\underline{w}|(t - t_0)\right] R(t_0) \tag{4.12}$$

Let us set the orientation of B such that B and I are aligned at $t = t_0$. That is, when $t = t_0$, B and I have the same directional vectors along their x–, y–, and z–axes, respectively. Hence

$$[\underline{e}_{Bx}(t_0) \;\; \underline{e}_{By}(t_0) \;\; \underline{e}_{Bz}(t_0)] = I_3 \tag{4.13}$$

where I_3 is the 3×3 identity matrix. Using (4.10), (4.13) implies $R(t_0) = I_3$. Substituting this into (4.12), we have

$$R(t) = R_0\left[\frac{w_x}{|\underline{w}|}, \frac{w_y}{|\underline{w}|}, \frac{w_z}{|\underline{w}|} \;;\; |\underline{w}|(t - t_0)\right] \tag{4.14}$$

which shows that the rotational motion in (4.7) is parameterized by $\underline{w}$. The three components of $\underline{w}$ are thus called the *rotational parameters.* The row vectors of

$R(t)$ will be denoted by

$$R_i(\underline{w}, t-t_0) \equiv \text{ the } i^{th} \text{ row vector of } R(t) \tag{4.15}$$

4.3 Batch Formulation

Substituting (4.5), (4.8), (4.9), and (4.15) into (4.7), we have

$$\begin{pmatrix} x_i(t) \\ y_i(t) \\ z_i(t) \end{pmatrix} = \begin{pmatrix} r_{OBx}(t_0) + (t-t_0)\,\dot{r}_{OBx} + R_1(\underline{w}, t-t_0)\,\underline{s}_{Bi} \\ r_{OBy}(t_0) + (t-t_0)\,\dot{r}_{OBy} + R_2(\underline{w}, t-t_0)\,\underline{s}_{Bi} \\ r_{OBz}(t_0) + (t-t_0)\,\dot{r}_{OBz} + R_3(\underline{w}, t-t_0)\,\underline{s}_{Bi} \end{pmatrix} \tag{4.16}$$

4.3.1 Monocular Vision

It is well known that monocular vision contains the translational information only up to a global scaling factor. We take this into account by normalizing the translational and structure parameters by the z– component of $\underline{r}_{OB}(t_0)$ [18, 20, 21]. Using the notation in (4.5), the normalized translational and structure parameters are

$$\begin{aligned} \underline{r}^N_{OB}(t) &\equiv \left(r^N_{OBx}(t), r^N_{OBy}(t)\right)^T \\ &\equiv \left(\frac{r_{OBx}(t)}{r_{OBz}(t_0)}, \frac{r_{OBy}(t)}{r_{OBz}(t_0)}\right)^T \\ \dot{\underline{r}}^N_{OB} &\equiv (\dot{r}^N_{OBx}, \dot{r}^N_{OBy}, \dot{r}^N_{OBz})^T \\ &\equiv \dot{\underline{r}}_{OB}/r_{OBz}(t_0) \\ \underline{s}^N_{Bi} &\equiv (s^N_{Bix}, s^N_{Biy}, s^N_{Biz})^T \\ &\equiv \underline{s}_{Bi}/r_{OBz}(t_0) \end{aligned} \tag{4.17}$$

Using (4.17), (4.16) may be rewritten as

$$\begin{pmatrix} x_i(t) \\ y_i(t) \\ z_i(t) \end{pmatrix} = r_{OBz}(t_0) \cdot \begin{pmatrix} r^N_{OBx}(t_0) + (t-t_0)\,\dot{r}^N_{OBx} + R_1(\underline{w}, t-t_0)\,\underline{s}^N_{Bi} \\ r^N_{OBy}(t_0) + (t-t_0)\,\dot{r}^N_{OBy} + R_2(\underline{w}, t-t_0)\,\underline{s}^N_{Bi} \\ 1 + (t-t_0)\,\dot{r}^N_{OBz} + R_3(\underline{w}, t-t_0)\,\underline{s}^N_{Bi} \end{pmatrix} \tag{4.18}$$

Substituting this into (4.2), we have

$$X_i(t) = f \frac{r^N_{OBx}(t_0) + (t-t_0)\,\dot{r}^N_{OBx} + R_1(\underline{w}, t-t_0)\,\underline{s}^N_{Bi}}{1 + (t-t_0)\,\dot{r}^N_{OBz} + R_3(\underline{w}, t-t_0)\,\underline{s}^N_{Bi}} + n_{Xi}(t) \quad (4.19)$$

$$Y_i(t) = f \frac{r^N_{OBy}(t_0) + (t-t_0)\,\dot{r}^N_{OBy} + R_2(\underline{w}, t-t_0)\,\underline{s}^N_{Bi}}{1 + (t-t_0)\,\dot{r}^N_{OBz} + R_3(\underline{w}, t-t_0)\,\underline{s}^N_{Bi}} + n_{Yi}(t)$$

Let t_i, $i = 1, 2, \cdots$, be the image taken times. Suppose $n(i)$ denotes the number of unoccluded feature points appearing in the i^{th} frame. The unoccluded points on the i^{th} frame are the $j_1{}^{th}$, $j_2{}^{th}$, $\cdots$, and $j_{n(i)}{}^{th}$ feature points. Let us note that for constant translation and angular velocity case, the rotation center O_B can be any point along the rotation axis L in Fig. 4.1. Therefore there is one extra degree of freedom involved in the parameters $\underline{r}^N_{OB}(t)$, $\underline{s}^N_{B1}$, $\underline{s}^N_{B2}$, $\cdots$, $\underline{s}^N_{BM}$, where M is the number of feature points. The z- component of $\underline{s}^N_{BM}$ is set to zero to remove this extra degree of freedom. Then the unknown motion and structure parameters for motion estimation from a sequence of monocular images are chosen as

$$\underline{y}^N \equiv (\underline{r}^N_{OB}(t_0),\ \dot{\underline{r}}^N_{OB},\ \underline{w},\ \underline{s}^N_{B1},\ \underline{s}^N_{B2},\ \cdots,\ \underline{s}^N_{B(M-1)},\ s^N_{BMx},\ s^N_{BMy})$$

Notice that only the x- and y- components of $\underline{s}^N_{BM}$ are included as we have set $s^N_{BMx} = 0$. Then the least squares estimate of the motion and structure parameters $\underline{y}^N$ for (4.19) is obtained by locating the following minimum

$$\min_{\underline{y}^N} \sum_i \sum_{k=1}^{n(i)} \left\{ \left[X_{j_k}(t_i) - f \frac{r^N_{OBx}(t_0) + (t_i-t_0)\,\dot{r}^N_{OBx} + R_1(\underline{w}, t_i-t_0)\,\underline{s}^N_{Bj_k}}{1 + (t_i-t_0)\,\dot{r}^N_{OBz} + R_3(\underline{w}, t_i-t_0)\,\underline{s}^N_{Bj_k}} \right]^2 \right.$$

$$\left. + \left[Y_{j_k}(t_i) - f \frac{r^N_{OBy}(t_0) + (t_i-t_0)\,\dot{r}^N_{OBy} + R_2(\underline{w}, t_i-t_0)\,\underline{s}^N_{Bj_k}}{1 + (t_i-t_0)\,\dot{r}^N_{OBz} + R_3(\underline{w}, t_i-t_0)\,\underline{s}^N_{Bj_k}} \right]^2 \right\} \Bigg|_{s^N_{BMz} = 0} \quad (4.20)$$

Note that the above summations sum over only all unoccluded perspective projections of feature points.

4.3.2 Binocular Vision

As depth information is retained in a binocular system, the unknown motion and structure parameters are chosen to be the following unnormalized quantities,

$$\underline{y} \equiv (\underline{r}_{OB}(t_0),\ \dot{\underline{r}}_{OB},\ \underline{w},\ \underline{s}_{B1},\ \underline{s}_{B2},\ \cdots,\ \underline{s}_{B(M-1)},\ s_{BMx},\ s_{BMy})$$

Substituting (4.16) into (4.3) and (4.4), we obtain

$$\begin{aligned} X_{li}(t) &= f\,\frac{r_{OBx}(t_0)+(t-t_0)\,\dot{r}_{OBx}+R_1(\underline{w},\,t-t_0)\,\underline{s}_{Bi}+b}{r_{OBz}(t_0)+(t-t_0)\,\dot{r}_{OBz}+R_3(\underline{w},\,t-t_0)\,\underline{s}_{Bi}}+n_{Xli}(t) \\ Y_{li}(t) &= f\,\frac{r_{OBy}(t_0)+(t-t_0)\,\dot{r}_{OBy}+R_2(\underline{w},\,t-t_0)\,\underline{s}_{Bi}}{r_{OBz}(t_0)+(t-t_0)\,\dot{r}_{OBz}+R_3(\underline{w},\,t-t_0)\,\underline{s}_{Bi}}+n_{Yli}(t) \end{aligned} \quad (4.21)$$

and

$$\begin{aligned} X_{ri}(t) &= f\,\frac{r_{OBx}(t_0)+(t-t_0)\,\dot{r}_{OBx}+R_1(\underline{w},\,t-t_0)\,\underline{s}_{Bi}-b}{r_{OBz}(t_0)+(t-t_0)\,\dot{r}_{OBz}+R_3(\underline{w},\,t-t_0)\,\underline{s}_{Bi}}+n_{Xri}(t) \\ Y_{ri}(t) &= f\,\frac{r_{OBy}(t_0)+(t-t_0)\,\dot{r}_{OBy}+R_2(\underline{w},\,t-t_0)\,\underline{s}_{Bi}}{r_{OBz}(t_0)+(t-t_0)\,\dot{r}_{OBz}+R_3(\underline{w},\,t-t_0)\,\underline{s}_{Bi}}+n_{Yri}(t) \end{aligned} \quad (4.22)$$

Assume the left images are taken at t_{li}, $i=0,1,\cdots$, and the right images at t_{ri}, $i=0,1,\cdots$. $n_l(i)$ and $n_r(i)$ are the numbers of feature points appearing in the i^{th} left and right frames, respectively. In the i^{th} left image, the ${j_{l1}}^{th}$, ${j_{l2}}^{th}$, $\cdots$, ${j_{ln_l(i)}}^{th}$ feature points appear. Similarly we define $j_{r1}, j_{r2}, \cdots, j_{rn_r(i)}$ for the right image. Then the least square estimates of motion and structure parameters $\underline{y}$ for (4.21) and (4.22) are obtained by locating the following minimum

$$\min_{\underline{y}}\left\{\sum_i\sum_{k=1}^{n_l(i)}\left\{\left[X_{lj_k}(t_{li})-f\,\frac{r_{OBx}(t_0)+(t_{li}-t_0)\,\dot{r}_{OBx}+R_1(\underline{w},\,t_{li}-t_0)\,\underline{s}_{Bj_k}+b}{r_{OBz}(t_0)+(t_{li}-t_0)\,\dot{r}_{OBz}+R_3(\underline{w},\,t_{li}-t_0)\,\underline{s}_{Bj_k}}\right]^2\right.\right.$$

$$\left.\left.+\left[Y_{lj_k}(t_{li})-f\,\frac{r_{OBy}(t_0)+(t_{li}-t_0)\,\dot{r}_{OBy}+R_2(\underline{w},\,t_{li}-t_0)\,\underline{s}_{Bj_k}}{r_{OBz}(t_0)+(t_{li}-t_0)\,\dot{r}_{OBz}+R_3(\underline{w},\,t_{li}-t_0)\,\underline{s}_{Bj_k}}\right]^2\right\}\right. \quad (4.23)$$

$$+\sum_i\sum_{k=1}^{n_r(i)}\left\{\left[X_{rj_k}(t_{ri})-f\,\frac{r_{OBx}(t_0)+(t_{ri}-t_0)\,\dot{r}_{OBx}+R_1(\underline{w},\,t_{ri}-t_0)\,\underline{s}_{Bj_k}-b}{r_{OBz}(t_0)+(t_{ri}-t_0)\,\dot{r}_{OBz}+R_3(\underline{w},\,t_{ri}-t_0)\,\underline{s}_{Bj_k}}\right]^2\right.$$

$$+ \left[Y_{rj_k}(t_{ri}) - f \frac{r_{OBy}(t_0) + (t_{ri} - t_0)\, \dot{r}_{OBy} + R_2(\underline{w},\, t_{ri} - t_0)\, \underline{s}_{Bj_k}}{r_{OBz}(t_0) + (t_{ri} - t_0)\, \dot{r}_{OBz} + R_3(\underline{w},\, t_{ri} - t_0)\, \underline{s}_{Bj_k}} \right]^2 \Bigg\} \Bigg\} \Bigg|_{s_{BMz} = 0}$$

Let us remark that as t_{li} is not necessarily equal to t_{ri}, the left and right cameras are not required to take images at the same time. Since stereo triangulation requires both the left and right images to appear at the same time instant, occlusion as well as asynchronousness of the camera pair prevent us from using stereo triangulation techniques to recover 3–D depths of feature points. In our model, the 3–D depths are recovered by solving the motion and structure parameters $\underline{y}$ from (4.23) and then by substituting $\underline{y}$ into (4.16) to obtain the depths $z_i(t)$.

4.4 Recursive Formulation

In the following, a general form of EKF/IEKF that we are going to apply is reviewed. Then the EKF/IEKF formulation for our model is given in the sequel.

The motion and structure parameters to be estimated are allowed to be time-varying and denoted by an $n \times 1$ column vector $\underline{x}(t)$, called the state vector. The time-varying behavior of $\underline{x}(t)$ will be described by the *plant equation*

$$\dot{\underline{x}}(t) = F\, \underline{x}(t) \tag{4.24}$$

where F is an $n \times n$ constant matrix. Let the column vector $\underline{z}(t_i)$ be the collection of image measurements taken at time t_i. Since the image measurements $\underline{z}(t_i)$ are functions of the motion and structure parameters $\underline{x}(t)$, and contaminated by noise, we have the *measurement equation*

$$\underline{z}(t_i) = \underline{h}\,[\underline{x}(t_i); t_i] + \underline{v}(t_i) \tag{4.25}$$

where $\underline{h}$ is a nonlinear function of $\underline{x}(t_i)$ and t_i due to the nonlinearity introduced by the rotational motion and central projection; and $\underline{v}$ is assumed to be a white Gaussian noise sequence of mean zero and covariance kernel

$$E\{\underline{v}(t_i)\underline{v}^T(t_j)\} = \begin{cases} R_v(t_i) & , t_i = t_j \\ 0 & , t_i \neq t_j \end{cases} \tag{4.26}$$

Once plant and measurement equations are specified, the EKF updates the estimate $\underline{\hat{x}}$ for the true state vector $\underline{x}$ by a series of equations given below [51]. Since the nonlinearity in $\underline{h}$ is linearized by a first order Taylor series expansion, the following matrix definition is needed,

$$H \equiv \frac{\partial \underline{h}}{\partial \underline{x}} \tag{4.27}$$

The estimate $\underline{\hat{x}}(t_i^+)$ for $\underline{x}(t_i^+)$ immediately after the measurements at t_i is obtained by

$$\underline{\hat{x}}(t_i^+) = \underline{\hat{x}}(t_i^-) + K(t_i)\{\underline{z}(t_i) - \underline{h}[\underline{\hat{x}}(t_i^-); t_i]\} \tag{4.28}$$

where the gain matrix is

$$K(t_i) = P(t_i^-)H^T[\underline{\hat{x}}(t_i^-); t_i]\{H[\underline{\hat{x}}(t_i^-); t_i]P(t_i^-)H^T[\underline{\hat{x}}(t_i^-); t_i] + R_v(t_i)\}^{-1} \tag{4.29}$$

The approximate covariance matrix P is updated by

$$P(t_i^+) = P(t_i^-) - K(t_i)H^T[\underline{\hat{x}}(t_i^-); t_i]P(t_i^-) \tag{4.30}$$

Then the state estimate and the approximate covariance matrix are propagated from t_i^+ ($\underline{\hat{x}}(t_i^+)$ and $P(t_i^+)$) forward to the next sample time t_{i+1}^- ($\underline{\hat{x}}(t_{i+1}^-)$ and $P(t_{i+1}^-)$) according to the following differential equations

$$\dot{\underline{\hat{x}}}(t) = F\,\underline{\hat{x}}(t) \tag{4.31}$$

$$\dot{P}(t) = F\,P(t) + P(t)\,F^T \tag{4.32}$$

The initial conditions used to solve these equations are $\underline{\hat{x}}(t_i^+)$ and $P(t_i^+)$. If both (4.31) and (4.32) have closed form solutions, called the *state transition equation* for (4.31) and the *covariance transition equation* for (4.32), the computation is simplified as no numerical integration is needed to propagate $\underline{\hat{x}}(t_i^+)$ to $\underline{\hat{x}}(t_{i+1}^-)$ and $P(t_i^+)$ to $P(t_{i+1}^-)$.

The above is the EKF formulation. If the nonlinearity in $\underline{h}$ is not satisfactorily linearized by EKF, improvement may be obtained by IEKF [51]. In IEKF, (4.28)

and (4.29) are replaced by setting $\hat{\underline{x}}_0$ equal to $\hat{\underline{x}}(t_i^-)$ and doing iterations on

$$K(t_i) = P(t_i^-)H^T[\hat{\underline{x}}_k; t_i]\{H[\hat{\underline{x}}_k; t_i]P(t_i^-)H^T[\hat{\underline{x}}_k; t_i] + R_v(t_i)\}^{-1} \quad (4.33)$$

$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}(t_i^-) + K(t_i)\{\underline{z}(t_i) - \underline{h}[\hat{\underline{x}}_k; t_i] - H[\hat{\underline{x}}_k; t_i][\hat{\underline{x}}(t_i^-) - \hat{\underline{x}}_k]\} \quad (4.34)$$

for $k = 0, 1, \ldots, N-1$ and then setting $\hat{\underline{x}}(t_i^+) = \hat{\underline{x}}_N$. The iterations are stopped when the improvement in $\hat{\underline{x}}_k$ is less than a preselected threshold. The covariance matrix in (4.30) is then updated based on $\hat{\underline{x}}_N$.

The plant equation (4.24), measurement equation (4.25), state transition equation, and covariance transition equation for our model are formulated next. These are readily applicable to both EKF and IEKF.

4.4.1 Monocular Vision

The state vector $\underline{x}^N(t)$ is chosen as the collection of the following normalized motion and structure parameters,

$$\underline{x}^N(t) \equiv \begin{pmatrix} \underline{r}_{OB}^N(t) \\ \dot{\underline{r}}_{OB}^N \\ \underline{w} \\ \underline{s}_{B1}^N \\ \underline{s}_{B2}^N \\ \vdots \\ \underline{s}_{B(M-1)}^N \\ s_{BMx}^N \\ s_{BMy}^N \end{pmatrix}$$

As in Section 4.3.1, we have set the z- component of $\underline{s}_{BM}^N$ to zero to remove the extra degree of freedom involved in $\underline{r}_{OB}^N(t)$, $\underline{s}_{B1}^N$, $\underline{s}_{B2}^N$, $\cdots$, $\underline{s}_{BM}^N$.

Plant Equation

For rigid constant translational velocity and angular velocity motion, the time derivatives of $\dot{\underline{r}}_{OB}^N$, $\underline{w}$, $\underline{s}_{B1}^N$, $\underline{s}_{B2}^N$, $\cdots$, $\underline{s}_{BM}^N$ are all zeroes. Using this fact and (4.8),

we obtain the plant equation

$$\dot{\underline{x}}^N(t) = F^N \underline{x}^N(t) \tag{4.35}$$

where F^N is the sparse square matrix

$$F^N \equiv \left\{ F^N_{13} = F^N_{24} = 1; \quad \text{all other } F^N_{ij} = 0 \right\} \tag{4.36}$$

Measurement Equation

Substituting (4.5), (4.8), (4.15) into (4.7), we have

$$\begin{pmatrix} x_i(t) \\ y_i(t) \\ z_i(t) \end{pmatrix} = \begin{pmatrix} r_{OBx}(t) + R_1(\underline{w}, t - t_0)\, \underline{s}_{Bi} \\ r_{OBy}(t) + R_2(\underline{w}, t - t_0)\, \underline{s}_{Bi} \\ r_{OBz}(t) + R_3(\underline{w}, t - t_0)\, \underline{s}_{Bi} \end{pmatrix} \tag{4.37}$$

By (4.9) and (4.17), (4.37) becomes

$$\begin{pmatrix} x_i(t) \\ y_i(t) \\ z_i(t) \end{pmatrix} = r_{OBz}(t_0) \cdot \begin{pmatrix} r^N_{OBx}(t) + R_1(\underline{w}, t - t_0)\, \underline{s}^N_{Bi} \\ r^N_{OBy}(t) + R_2(\underline{w}, t - t_0)\, \underline{s}^N_{Bi} \\ 1 + (t - t_0)\, \dot{r}^N_{OBz} + R_3(\underline{w}, t - t_0)\, \underline{s}^N_{Bi} \end{pmatrix} \tag{4.38}$$

Substituting this into (4.2), we obtain the measurement equation

$$\underbrace{\begin{pmatrix} X_{j_1}(t_i) \\ Y_{j_1}(t_i) \\ \vdots \\ X_{j_{n(i)}}(t_i) \\ Y_{j_{n(i)}}(t_i) \end{pmatrix}}_{\underline{z}^N(t_i)} = \underbrace{\begin{pmatrix} f\, \dfrac{r^N_{OBx}(t_i) + R_1(\underline{w}, t_i - t_0)\, \underline{s}^N_{Bj_1}}{1 + (t_i - t_0)\, \dot{r}^N_{OBz} + R_3(\underline{w}, t_i - t_0)\, \underline{s}^N_{Bj_1}} \\ f\, \dfrac{r^N_{OBy}(t_i) + R_2(\underline{w}, t_i - t_0)\, \underline{s}^N_{Bj_1}}{1 + (t_i - t_0)\, \dot{r}^N_{OBz} + R_3(\underline{w}, t_i - t_0)\, \underline{s}^N_{Bj_1}} \\ \vdots \\ f\, \dfrac{r^N_{OBx}(t_i) + R_1(\underline{w}, t_i - t_0)\, \underline{s}^N_{Bj_{n(i)}}}{1 + (t_i - t_0)\, \dot{r}^N_{OBz} + R_3(\underline{w}, t_i - t_0)\, \underline{s}^N_{Bj_{n(i)}}} \\ f\, \dfrac{r^N_{OBy}(t_i) + R_2(\underline{w}, t_i - t_0)\, \underline{s}^N_{Bj_{n(i)}}}{1 + (t_i - t_0)\, \dot{r}^N_{OBz} + R_3(\underline{w}, t_i - t_0)\, \underline{s}^N_{Bj_{n(i)}}} \end{pmatrix}_{s_{BMz} = 0}}_{\underline{h}^N[\underline{x}^N(t_i);\, t_i]}$$

$$+\underbrace{\begin{pmatrix} n_{Xj_1}(t_i) \\ n_{Yj_1}(t_i) \\ \vdots \\ n_{Xj_{n(i)}}(t_i) \\ n_{Yj_{n(i)}}(t_i) \end{pmatrix}}_{\underline{v}^N(t_i)} \tag{4.39}$$

where $j_1, j_2, \cdots, j_{n(i)}$ are unoccluded feature points at time t_i, as defined in Section 4.3.1. Therefore only occluded points are incorparated in the measumement equation (4.39). Note that the measurement equation (4.39) has been put in the standard form (4.25).

State Transition Equation

$\dot{\underline{r}}^N_{OB}, \underline{w}, \underline{s}^N_{B1}, \underline{s}^N_{B2}, \cdots, \underline{s}^N_{BM}$ are constant in t for rigid constant velocity and angular velocity motion. This fact with (4.9) and (4.17) gives the state transtion equation

$$\underline{x}^N(t^-_{i+1}) = \left[I + (t^-_{i+1} - t^+_i)\ F^N\right]\ \underline{x}^N(t^+_i) \tag{4.40}$$

where I is the identity matrix.

Covariance Transition Equation

From (4.36), we see that $(F^N)^2$ is a zero matrix. Using this, a direct substitution may be used to verify that the closed form solution of (4.32) is the following covariance transition equation,

$$P^N(t^-_{i+1}) = \left[I + (t^-_{i+1} - t^+_i)\ F^N\right]\ P^N(t^+_i)\ \left[I + (t^-_{i+1} - t^+_i)\ F^N\right]^T \tag{4.41}$$

4.4.2 Binocular Vision

Since depth information is retained in the binocular images, the unnormalized motion and structure parameters constitute the state vector

$$\underline{x}(t) \equiv \begin{pmatrix} r_{OB}(t) \\ \dot{\underline{r}}_{OB} \\ \underline{w} \\ \underline{s}_{B1} \\ \underline{s}_{B2} \\ \vdots \\ \underline{s}_{B(M-1)} \\ s_{BMx} \\ s_{BMy} \end{pmatrix}$$

The plant, measurement, state transition, and covariance transition equations for the binocular case may be obtained using the same arguments as in the monocular case, as shown in the following.

Plant Equation

The plant equation is obtained from (4.8) as

$$\dot{\underline{x}}(t) = F\,\underline{x}(t) \tag{4.42}$$

where F is the sparse square matrix

$$F \equiv \{F_{14} = F_{25} = F_{36} = 1; \quad \text{all other } F_{ij} = 0\} \tag{4.43}$$

Measurement Equation

A substitution of (4.37) into (4.3) and (4.4) gives the measurement equation

$$\underbrace{\begin{pmatrix} X_{lj_{l1}}(t_{li}) \\ Y_{lj_{l1}}(t_{li}) \\ \vdots \\ X_{lj_{ln(i)}}(t_{li}) \\ Y_{lj_{ln(i)}}(t_{li}) \end{pmatrix}}_{\underline{z}(t_{li})} = \underbrace{\begin{pmatrix} f\,\dfrac{r_{OBx}(t_{li}) + R_1(\underline{w},\, t_{li}-t_0)\,\underline{s}_{Bj_{l1}} + b}{r_{OBz}(t_{li}) + R_3(\underline{w},\, t_{li}-t_0)\,\underline{s}_{Bj_{l1}}} \\ f\,\dfrac{r_{OBy}(t_{li}) + R_2(\underline{w},\, t_{li}-t_0)\,\underline{s}_{Bj_{l1}}}{r_{OBz}(t_{li}) + R_3(\underline{w},\, t_{li}-t_0)\,\underline{s}_{Bj_{l1}}} \\ \vdots \\ f\,\dfrac{r_{OBx}(t_{li}) + R_1(\underline{w},\, t_{li}-t_0)\,\underline{s}_{Bj_{ln_l(i)}} + b}{r_{OBz}(t_{li}) + R_3(\underline{w},\, t_{li}-t_0)\,\underline{s}_{Bj_{ln_l(i)}}} \\ f\,\dfrac{r_{OBy}(t_{li}) + R_2(\underline{w},\, t_{li}-t_0)\,\underline{s}_{Bj_{ln_l(i)}}}{r_{OBz}(t_{li}) + R_3(\underline{w},\, t_{li}-t_0)\,\underline{s}_{Bj_{ln_l(i)}}} \end{pmatrix}_{s_{BMz}=0}}_{\underline{h}[\underline{x}(t_{li});\, t_{li}]}$$

$$+ \underbrace{\begin{pmatrix} n_{Xlj_{l1}}(t_{li}) \\ n_{Ylj_{l1}}(t_{li}) \\ \vdots \\ n_{Xlj_{ln_l(i)}}(t_{li}) \\ n_{Ylj_{ln_l(i)}}(t_{li}) \end{pmatrix}}_{\underline{v}(t_{li})} \tag{4.44}$$

and

$$\underbrace{\begin{pmatrix} X_{rj_{r1}}(t_{ri}) \\ Y_{rj_{r1}}(t_{ri}) \\ \vdots \\ X_{rj_{rn(i)}}(t_{ri}) \\ Y_{rj_{rn(i)}}(t_{ri}) \end{pmatrix}}_{\underline{z}(t_{ri})} = \underbrace{\begin{pmatrix} f\,\dfrac{r_{OBx}(t_{ri}) + R_1(\underline{w},\, t_{ri}-t_0)\,\underline{s}_{Bj_{r1}} - b}{r_{OBz}(t_{ri}) + R_3(\underline{w},\, t_{ri}-t_0)\,\underline{s}_{Bj_{r1}}} \\ f\,\dfrac{r_{OBy}(t_{ri}) + R_2(\underline{w},\, t_{ri}-t_0)\,\underline{s}_{Bj_{r1}}}{r_{OBz}(t_{ri}) + R_3(\underline{w},\, t_{ri}-t_0)\,\underline{s}_{Bj_{r1}}} \\ \vdots \\ f\,\dfrac{r_{OBx}(t_{ri}) + R_1(\underline{w},\, t_{ri}-t_0)\,\underline{s}_{Bj_{rnr(i)}} - b}{r_{OBz}(t_{ri}) + R_3(\underline{w},\, t_{ri}-t_0)\,\underline{s}_{Bj_{rnr(i)}}} \\ f\,\dfrac{r_{OBy}(t_{ri}) + R_2(\underline{w},\, t_{ri}-t_0)\,\underline{s}_{Bj_{rnr(i)}}}{r_{OBz}(t_{ri}) + R_3(\underline{w},\, t_{ri}-t_0)\,\underline{s}_{Bj_{rnr(i)}}} \end{pmatrix}_{s_{BMz}=0}}_{\underline{h}[\underline{x}(t_{ri});\, t_{ri}]}$$

$$+\underbrace{\begin{pmatrix} n_{Xrj_{r1}}(t_{ri}) \\ n_{Yrj_{r1}}(t_{ri}) \\ \vdots \\ n_{Xrj_{rn_r(i)}}(t_{ri}) \\ n_{Yrj_{rn_r(i)}}(t_{ri}) \end{pmatrix}}_{\underline{v}(t_{ri})} \tag{4.45}$$

where j_{l1}, j_{l2}, $\cdots$, $j_{ln_l(i)}$ are the unoccluded feature points at time t_{li} for the left camera, and j_{r1}, j_{r2}, $\cdots$, $j_{rn_r(i)}$ at t_{ri} for the right camera, as defined in Section 4.3.2.

State Transition Equation

By (4.9), the state transtion equation is

$$\underline{x}(t_{i+1}^-) = \left[I + (t_{i+1}^- - t_i^+)\ F\right]\ \underline{x}(t_i^+) \tag{4.46}$$

Covariance Transition Equation

As F^2 is a zero matrix from (4.43), the covariance transition equation is

$$P(t_{i+1}^-) = \left[I + (t_{i+1}^- - t_i^+)\ F\right]\ P(t_i^+)\ \left[I + (t_{i+1}^- - t_i^+)\ F\right]^T \tag{4.47}$$

4.4.3 Filtering

After the plant, measurement, state transition, and covariance transition equations are formulated, one can apply the EKF/IEKF. Using the monocular case as an example, the standard filtering algorithm is summarized in the following.

1. Since the plant model (4.24) is given by (4.35) and the measurement model (4.25) by (4.39), the linearized matrices H^N can be obtained from (4.27).

2. The initial guess of the state vector is assigned to $\hat{\underline{x}}^N(t_1^-)$. The initial guess of the approximate covariance matrix is assigned to $P^N(t_1^-)$.

3. Set $i = 1$.

4. At time t_i, the measurement $\underline{z}^N(t_i)$ is taken.

5. At time t_i^+, immediately after the measurement time t_i, the estimate $\hat{\underline{x}}^N(t_i^+)$ for the state vector is updated from $\hat{\underline{x}}^N(t_i^-)$ by (4.28) and (4.29). The approximate covariance matrix $P^N(t_i^+)$ is updated from $P^N(t_i^-)$ using (4.30).

   If an IEKF is used, $\hat{\underline{x}}^N(t_i^+)$ is obtained using the iterations defined in (4.33) and (4.34), instead of (4.28) and (4.29).

6. At time t_{i+1}^-, immediately before the measurement time t_{i+1}, the estimate $\hat{\underline{x}}^N(t_{i+1}^-)$ is propagated from $\hat{\underline{x}}^N(t_i^+)$ using the state transition equation (4.40). $P^N(t_{i+1}^-)$ is propagated from $P^N(t_i^+)$ using the covariance transition equation (4.41).

7. Increase i by 1, then go to step 4 to incorporate new measurements.

4.5 Simulation Results

Computer simulations are used to investigate the models. A synthetic sequence of noisy images is generated by the following scheme. First, the motion of the vertices of a cube in 3-D space is generated by a computer program according to constant translational velocity and angular velocity motion model. Then the vertices are centrally projected onto the image plane of the camera. We simulate the image plane to be of finite resolution, therefore the locations of the projected vertices on the images are measured with digitization error. In this simulation, the digitization error summarizes the effects of different noise sources such as digitization noise, corner detection noise, etc. possibly encountered by a feature extraction algorithm. The case of low resolution gives feature locations of low precision and simulates a high noise level in the image measurements.

The focal length is set to 1. The object is a rigid cube with side length 4. The image plane is a square with side length 1.5 and resolution 32×32. The angular velocity is $\underline{w} = (0.2667, 0.1333, -0.2667)^T$. Hence the cube rotates about the unit directional vector $\underline{w}/|\underline{w}| = (0.6667, 0.3333, -0.6667)^T$ by $|\underline{w}| \times 360°/2\pi = 22.92°$ per unit time. The velocity is $\dot{\underline{r}}_{OB} = (-0.25, 0.25, 0.5)^T$. The initial rotation center is $\underline{r}_{OB}(0) = (1, -1, 10)^T$. Four feature points are tracked: $\underline{s}_{B1} =$

Frame number i	Unoccluded feature points				t_i
1	1	2	3	4	0.37
2	1	2	3		1.21
3		2	3	4	2.28
4	1		3	4	3.39
5		2	3		4.32
6	1	2	3	4	4.61
7	1	2	3	4	5.99
8		2	3	4	6.77
9	1	2	3		7.67
10	1	2	3	4	8.88
11	1		3		10.05
12		2	3	4	10.83
13			3	4	11.78
14	1	2		4	13.17
15			3		13.63
16			3	4	14.98
17	1	2	3	4	16.13
18	1	2		4	16.97
19		2	3	4	18.13
20	1		3	4	19.00

Table 4.2: The occlusion condition and the temporal sampling instances. Note the nonuniformity of the sampling times. The entries in the second column denote the actual feature points used in motion computation. Thus in frame 2, feature point 4 is missing due to occlusion, and so on.

$(2, 2, 4)^T$, $\underline{s}_{B2} = (-2, -2, 4)^T$, $\underline{s}_{B3} = (2, -2, 0)^T$, and $\underline{s}_{B4} = (-2, 2, 0)^T$. As a result of the above camera set-up and motion and structure parameters, the object covers less than one ninth of the area of each image.

4.5.1 Monocular Vision

The occlusion condition considered is shown in Table 4.2, in which the feature points appearing in each image frame are listed; points not listed are occluded. The time instances at which the image frames are taken are also listed in Table 4.2. Then the least squares estimates of the motion and structure parameters

$\underline{y}^N$ are obtained from (4.20). The conjugate gradient algorithm is applied to locate the minimum in (4.20). Thus the 19 free parameters with respect to which the conjugate gradient method minimizes (4.20) are $\underline{r}_{OB}^N(0)$, $\dot{\underline{r}}_{OB}^N$, $\underline{w}$, $\underline{s}_{B1}^N$, $\underline{s}_{B2}^N$, $\underline{s}_{B3}^N$, and the first two components of $\underline{s}_{B4}^N$. The initial guess for each free parameter is arbitrarily set to 0.01. The result shows that at least three frames are required for all parameters to converge. In general, more accurate motion and structure estimates can be obtained by incorporating more frames. This may be seen in Fig. 4.3, which plots the errors in estimates with respect to the number of frames. The ordinate is the difference between the estimated and the true parameters. For each number of frames, the same initial guess is used to start the conjugate gradient search. All parameters converge with good precision when the number of frames is more than 10. The errors in $\underline{r}_{OB}^N(0)$, $\underline{s}_{B1}^N$, $\cdots$, $\underline{s}_{B4}^N$ are all similar and do not decrease significantly with an increasing in frame numbers if more than 10 frames are used.

While Fig. 4.3 is a typical sample path, the bias and variance of the motion and structure estimators may be studied by Monte Carlo trials. Each trial takes the measurements with a noise realization different from the other trials and gives one sample path similar to Fig. 4.1. If enough number of trials are done, the sample bias and sample variance obtained may be good indicators of some ensemble properties of the model. In our experiment, for each trial two stochastically independent random numbers δ_x and δ_y are generated from a uniform distribution on the interval between 0 and the inter-pixel distance. Then the location of every pixel on the image plane is shifted by (δ_x, δ_y). Because the image locations of feature points are digitized to the nearest pixels, the above shift changes the noise realization while preserving the noise distribution. Other parameters, e.g. ground truth, initial guess, occlusion, and image sampling times, etc., are all same. The sample bias and root mean square error from 50 Monte Carlo trials are plotted in Fig. 4.4 and 4.5, respectively. After the 7^{th} or 8^{th} frame, the motion and structure parameters have little bias. The estimates of $\underline{r}_{OB}^N(0)$, $\underline{s}_{B1}^N$, $\cdots$, $\underline{s}_{B4}^N$ all have similar bias. The plots of root mean square error are drawn in the logarithm scale. The efficiency of the estimates improves significantly when the number of frames increases from 3 to 12.

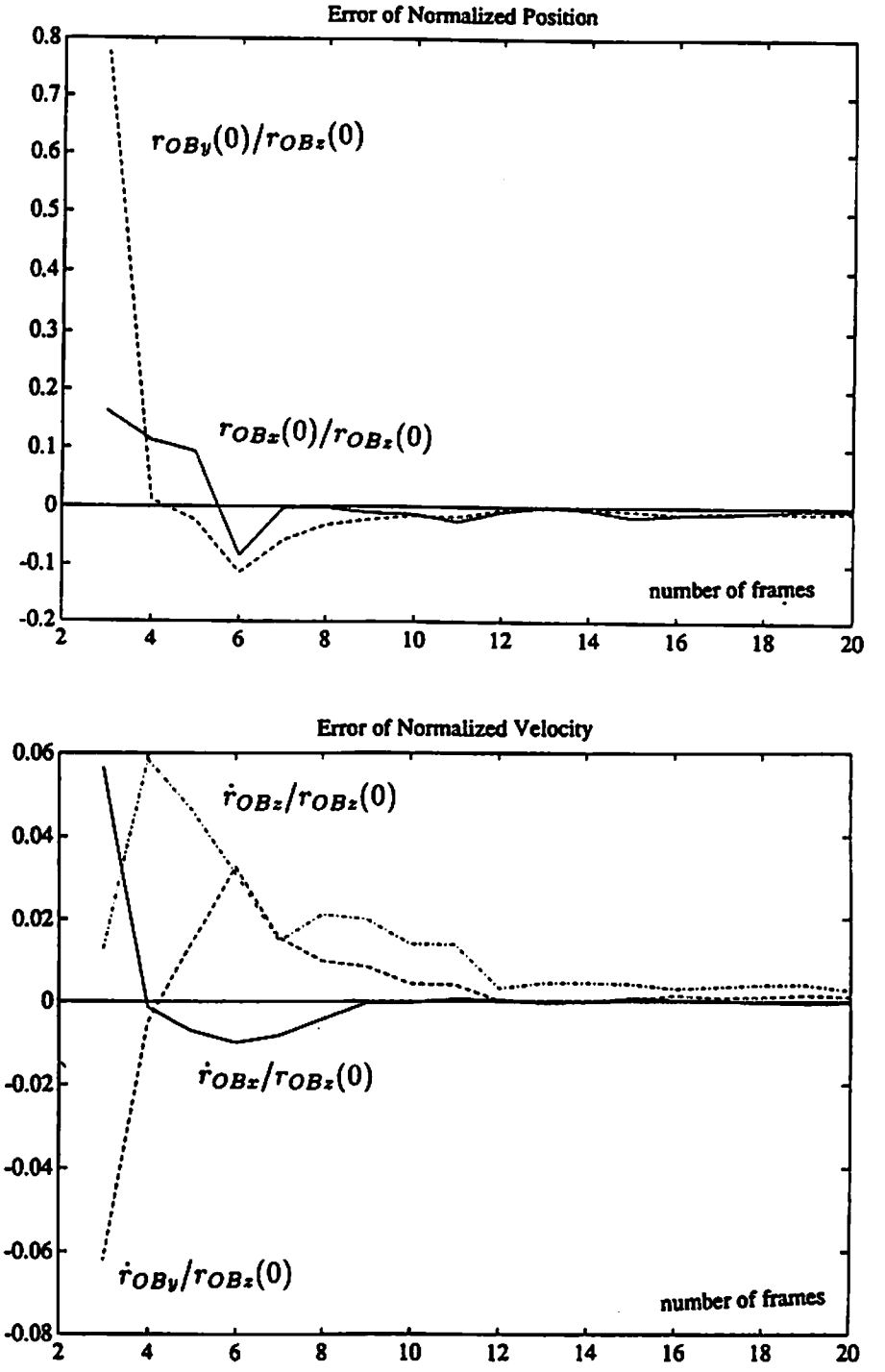


Figure 4.3: Errors in the estimates of motion and structure parameters for the monocular vision. (a) normalized translational position error and (b) normalized velocity error. The ordinate is the difference between the estimated and the true parameters.

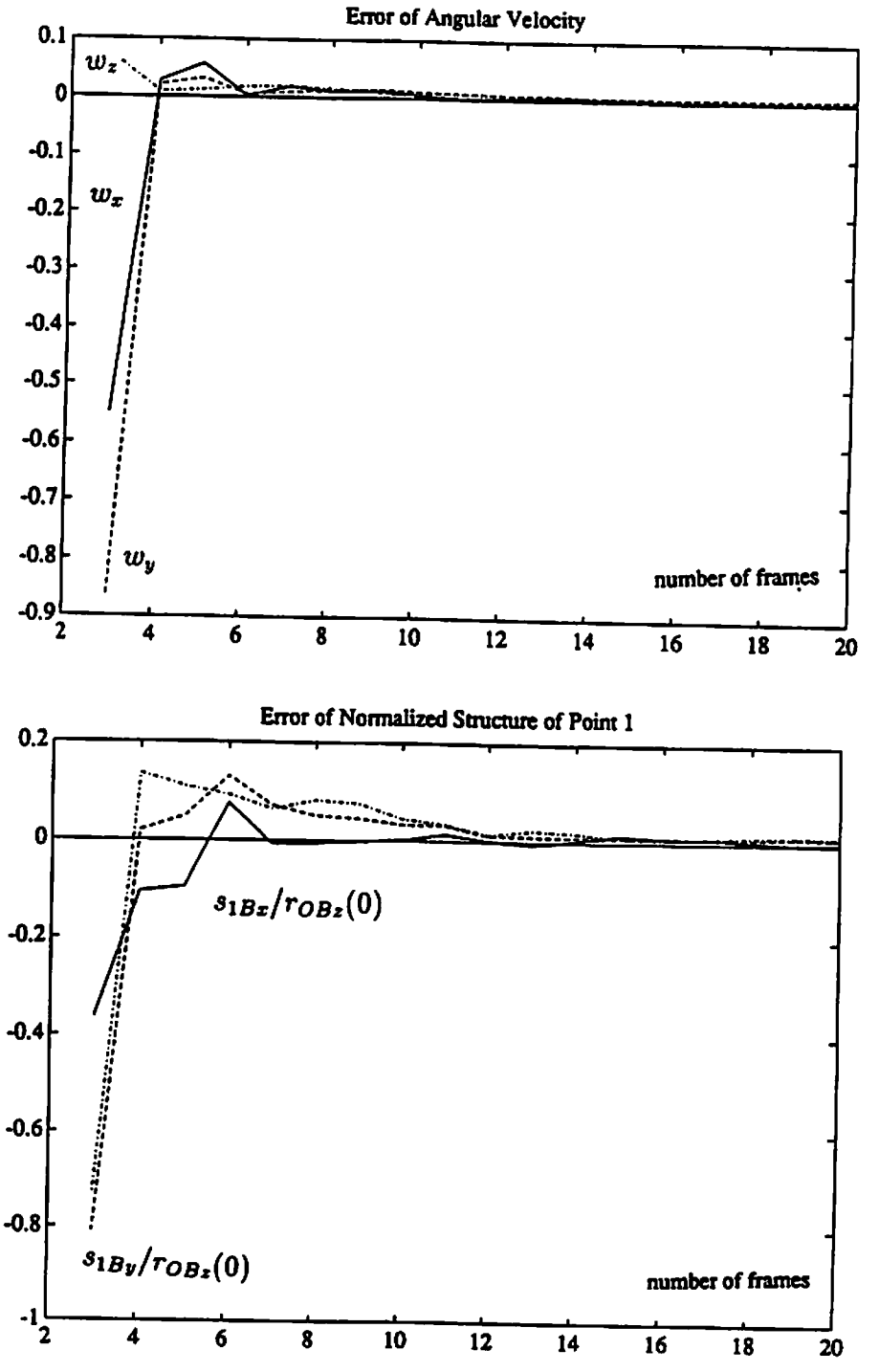


Figure 4.3: (continued) (c) angular velocity error and (d) normalized structure error.

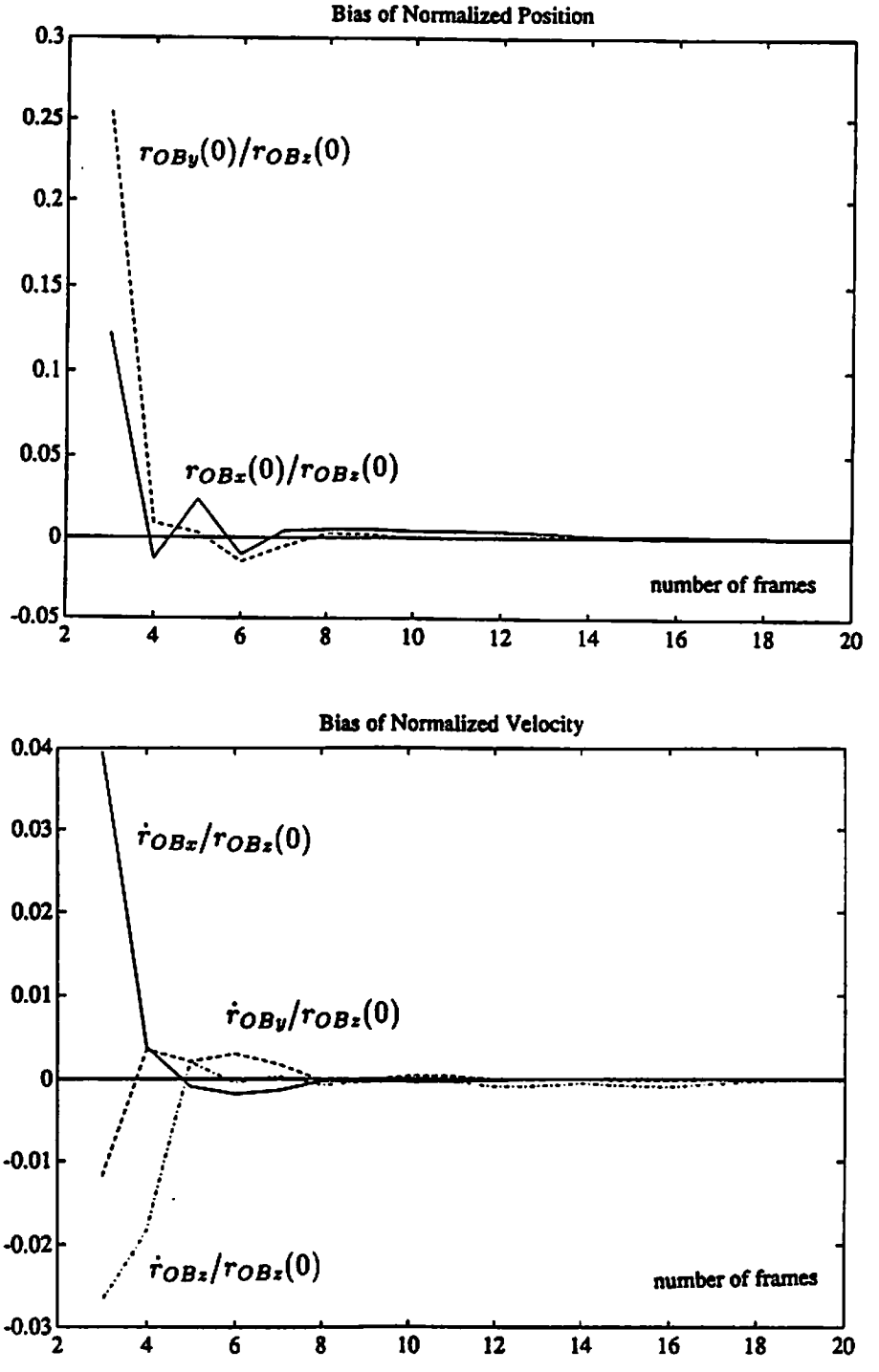


Figure 4.4: Bias in the estimates of motion and structure parameters from 50 Monte Carlo trials for the monocular vision. (a) normalized translational position bias and (b) normalized velocity bias.

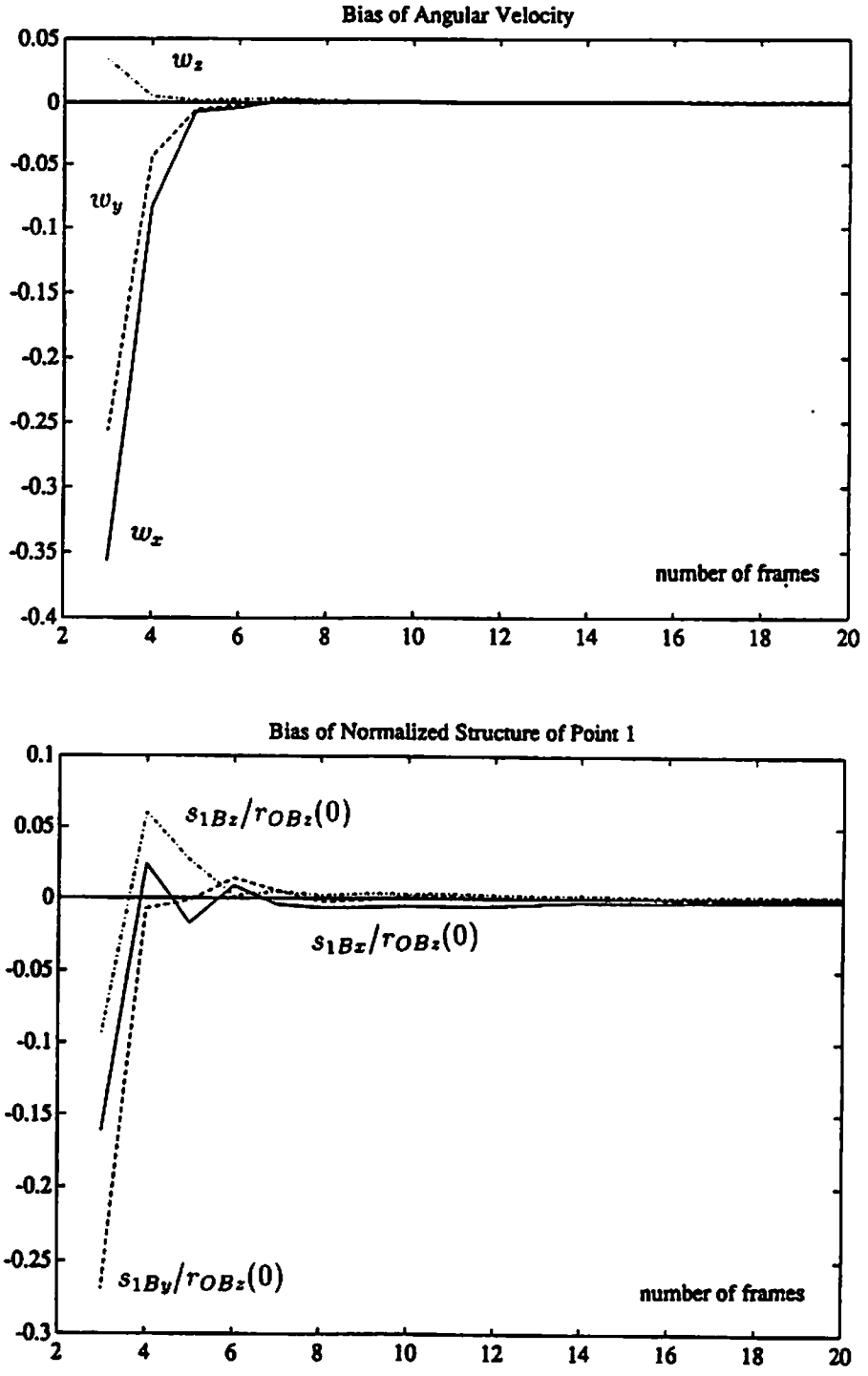


Figure 4.4: (continued) (c) angular velocity bias and (d) normalized structure bias.

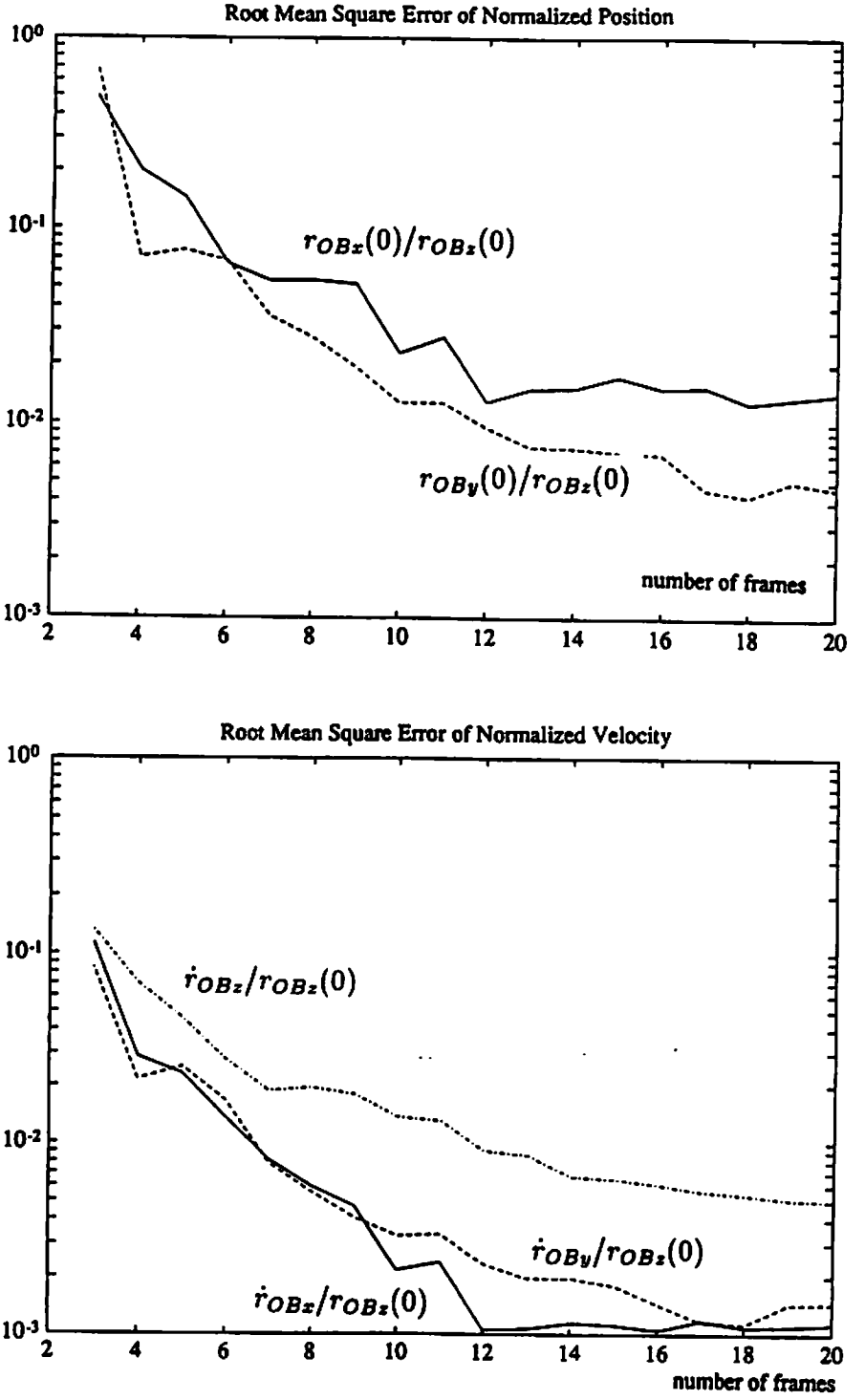


Figure 4.5: Root mean square error of motion and structure parameters from 50 Monte Carlo trials for the monocular vision. (a) normalized translational position and (b) normalized velocity.

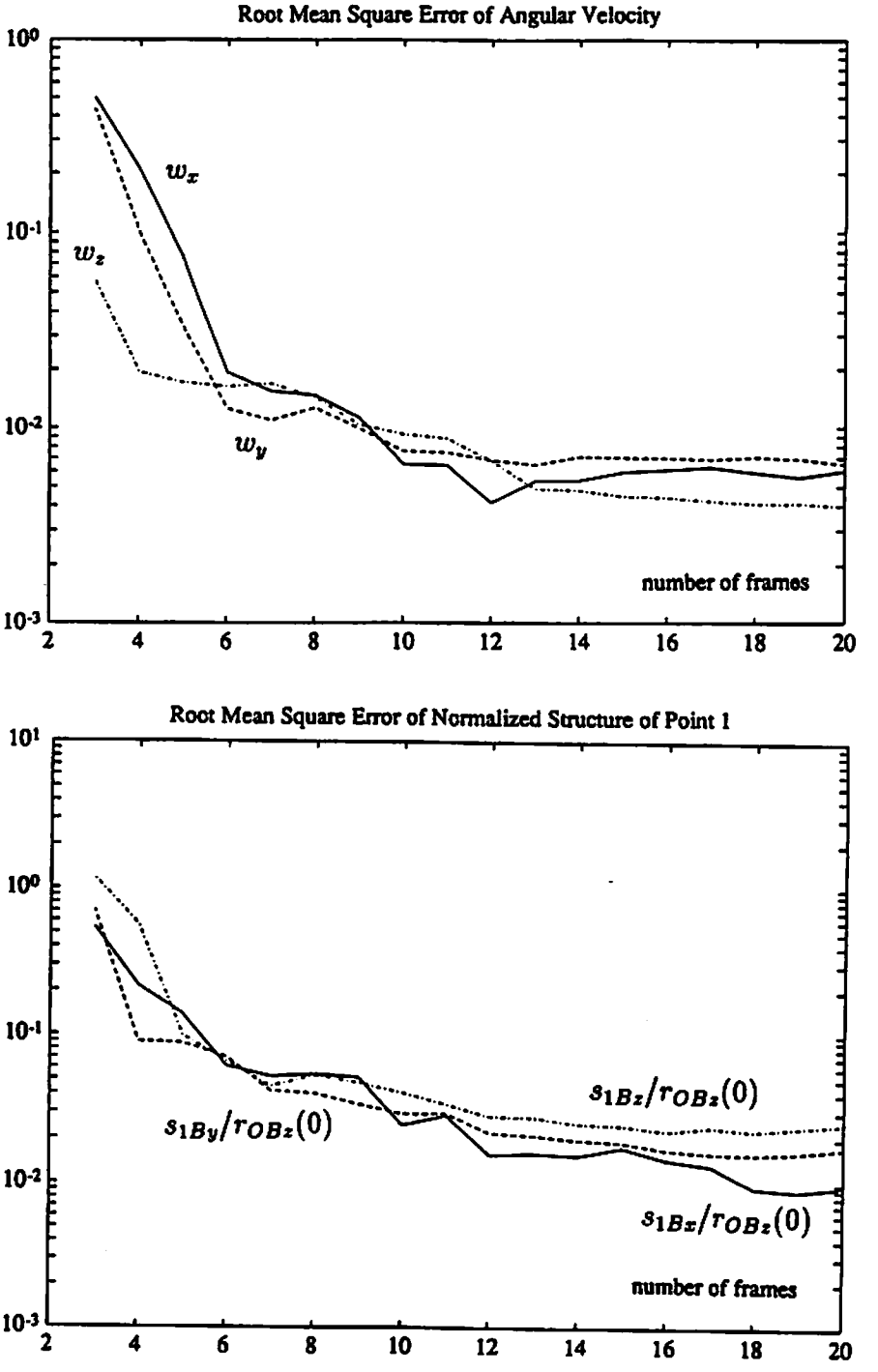


Figure 4.5: (continued) (c) angular velocity and (d) normalized structure.

4.5.2 Binocular Vision

In the simulations for binocular vision, the same camera set-up, motion and structure parameters, and initial guess in the monocular vision are used, except for the following changes. The baseline $2b$ is set to 10. The image sampling times t_{li} and the occlusion condition for the left camera is identical to those in Table 4.2. For the right camera, t_{ri} and unoccluded feature points are listed in Table 4.3. The least squares estimates of the motion and structure parameters $\underline{y}$ are obtained from (4.23) by the conjugate gradient algorithm. The 20 free parameters are $\underline{r}_{OB}(0)$, $\dot{\underline{r}}_{OB}$, $\underline{w}$, $\underline{s}_{B1}$, $\underline{s}_{B2}$, $\underline{s}_{B3}$, and the first two components of $\underline{s}_{B4}$. Fig. 4.6 shows the errors in motion and structure estimates with respect to the number of frames. The ordinate is the difference between the estimated and the true parameters. Fig. 4.7 and Fig. 4.8 show the sample bias and root mean square error, respectively, from 50 Monte Carlo trials. Similar convergence behavior to the monocular case is observed. In general, the performance in the binocular system is only slightly better than that in the monocular system. However, the monocular system does not recover the absolute depths, while the binocular system does.

The case when the image sampling times are the same for both cameras, i.e. $t_{lj} = t_{rj}$, are also simulated. The performance when the camera pair is synchronized is basically identical to that of asynchronous cameras. Let us note that an asynchronous camera pair is conceptually equivalent to a single moving camera with a controlled camera motion. Therefore this experiment demonstrates an example in which the 3–D depth of a moving object of unknown motion and structure is recovered from a monocular system with known camera motion. (The 3–D depth of the rotation center O_B is simply, by (4.9), the z– component of $\underline{r}_{OB}(t) = \underline{r}_{OB}(0) + t\,\dot{\underline{r}}_{OB}$, which is recovered from the estimated translational parameters $\underline{r}_{OB}(0)$ and $\dot{\underline{r}}_{OB}$. The depth of the i^{th} feature point is $z_i(t)$ in (4.16), which is recovered from the estimated motion and structure parameters $\underline{y}$.) We note that the two major factors which make it possible to recover the 3–D depth from the monocular vision system discussed here are the active camera motion and the motion model in Section 4.2.

Frame number i	Unoccluded feature points	t_i
1	2 3 4	0.34
2	1 2 3 4	1.22
3	1 2 4	2.21
4	1 3 4	3.40
5	2 3 4	4.30
6	4	4.66
7	1 2 3 4	5.94
8	1 2 3 4	6.72
9	1 3 4	7.66
10	1 2 3	8.82
11	1 3 4	10.11
12	1 2 3	10.89
13	1 2 4	11.71
14	3 4	13.22
15	2 3	13.68
16	2 4	15.01
17	2 3 4	16.18
18	1 2 3 4	17.02
19	1 2 4	18.08
20	2 3	19.02

Table 4.3: The occlusion condition and the temporal sampling instances for the right camera in the binocular vision. Those for the left camera are identical to the list in Table 4.2.

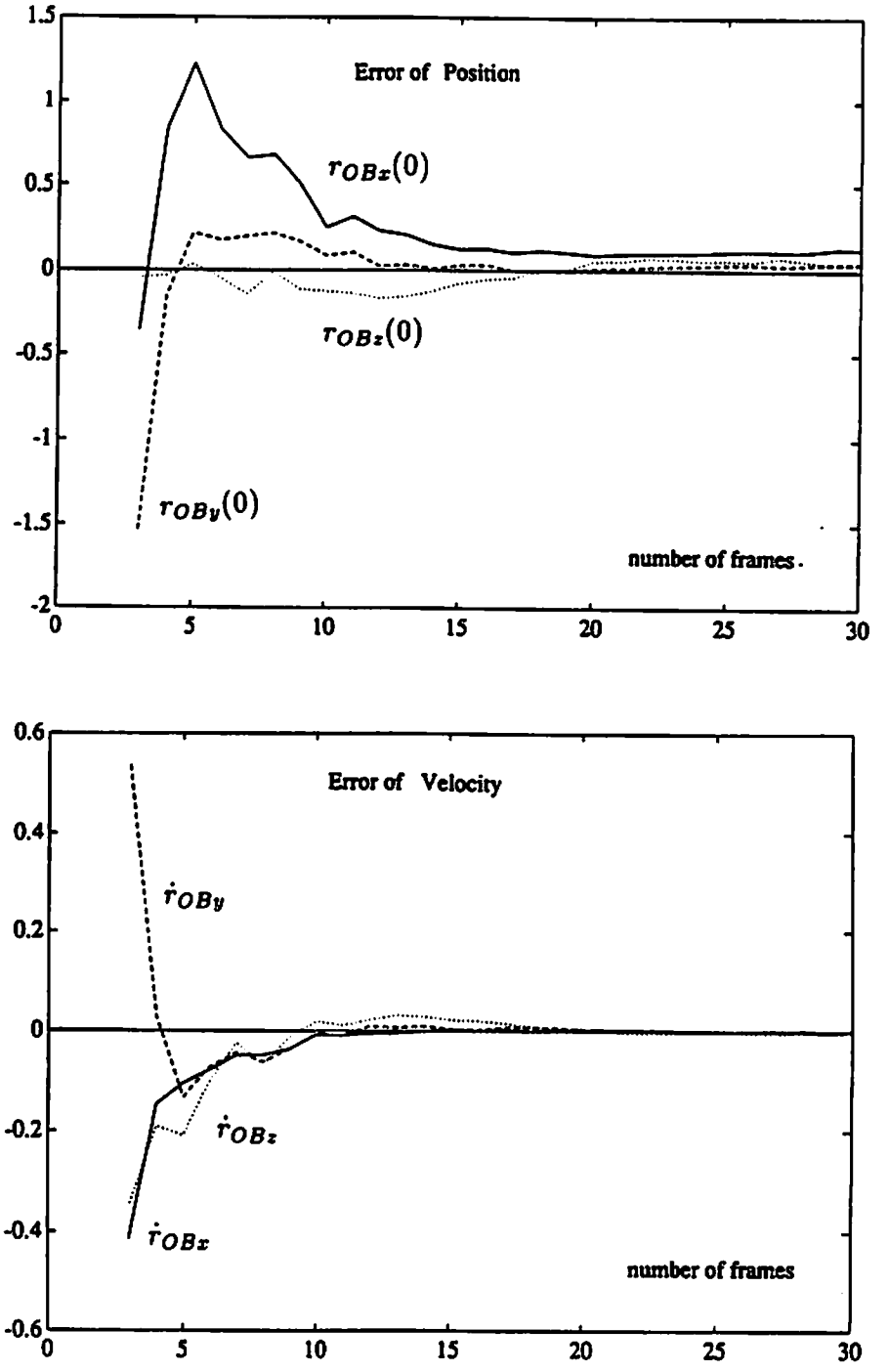
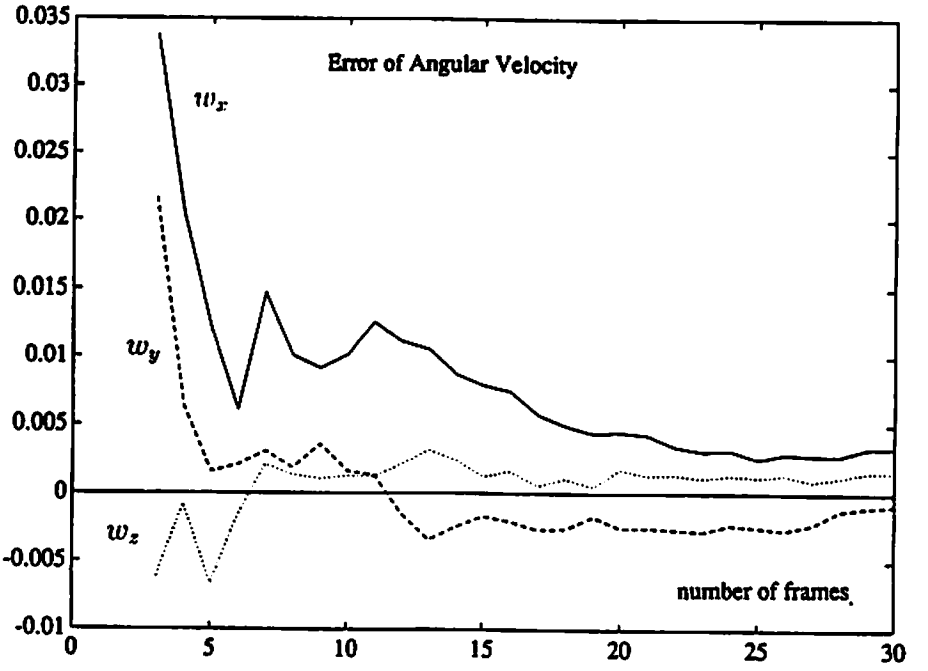


Figure 4.6: Errors in the estimates of motion and structure parameters for the binocular vision. (a) translational position error and (b) velocity error. The ordinate is the difference between the estimated and the true parameters.



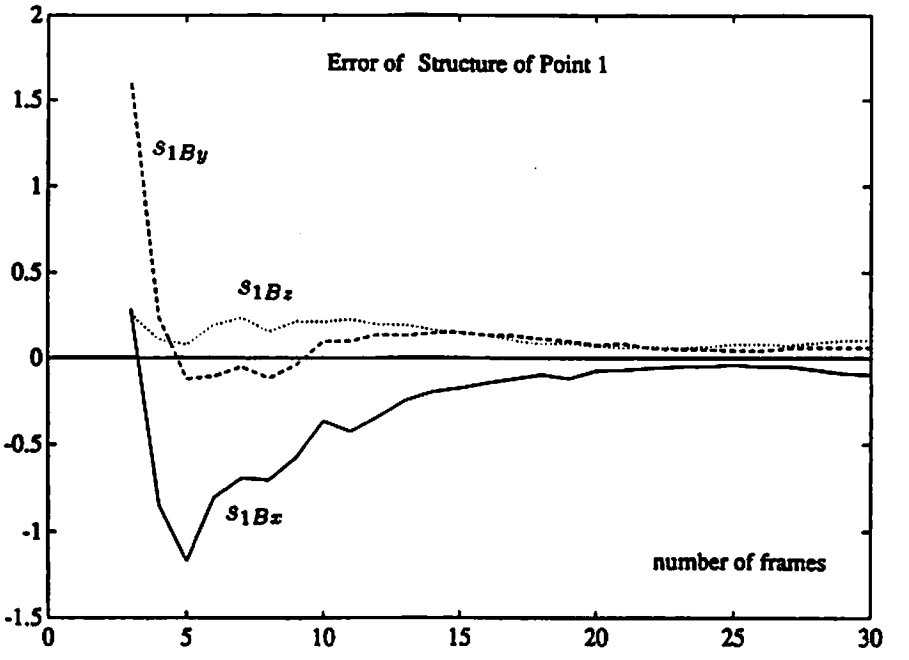


Figure 4.6: (continued) (c) angular velocity error and (d) structure error.

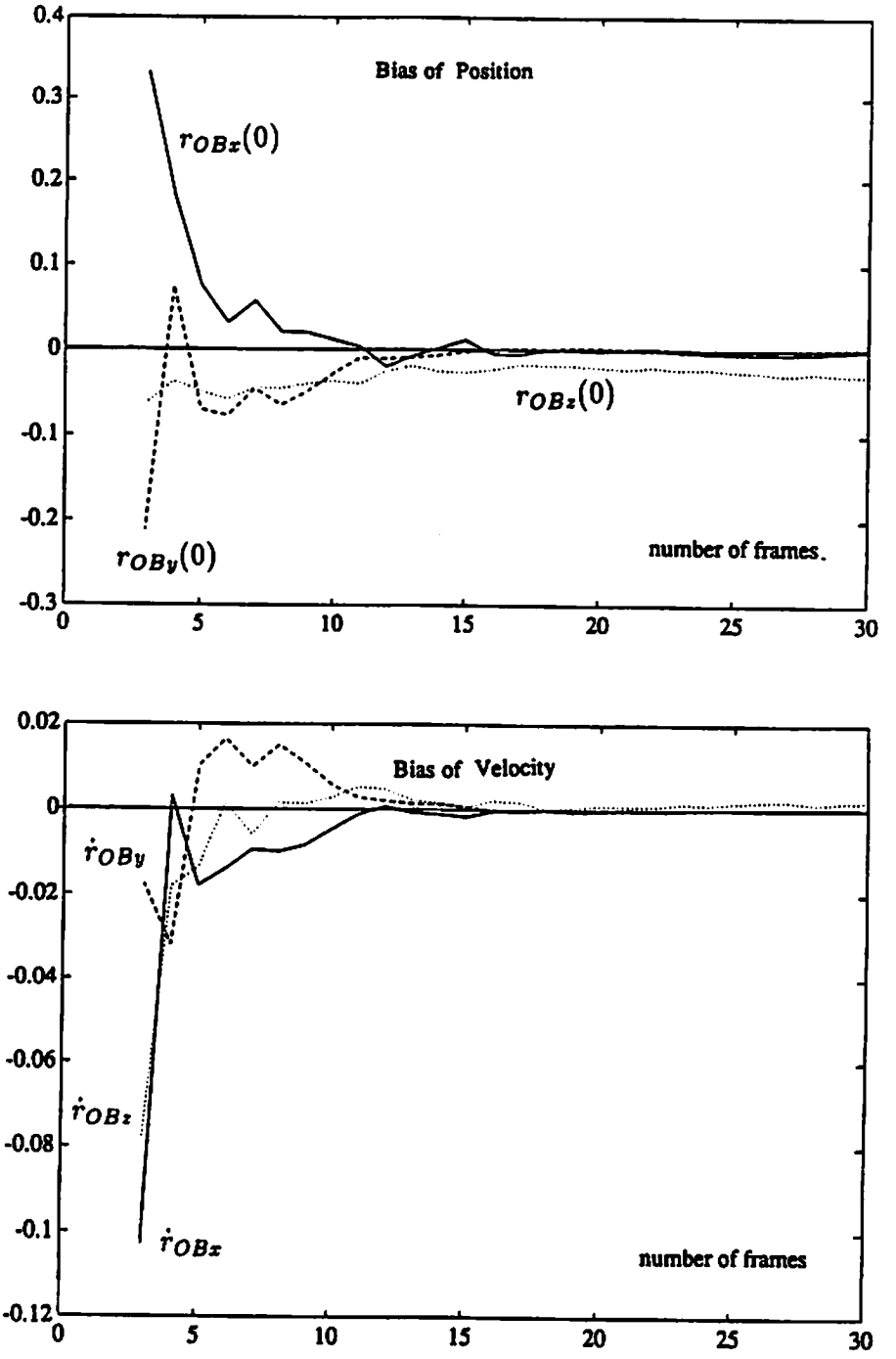


Figure 4.7: Bias in the estimates of motion and structure parameters from 50 Monte Carlo trials for the binocular vision. (a) translational position bias and (b) velocity bias.

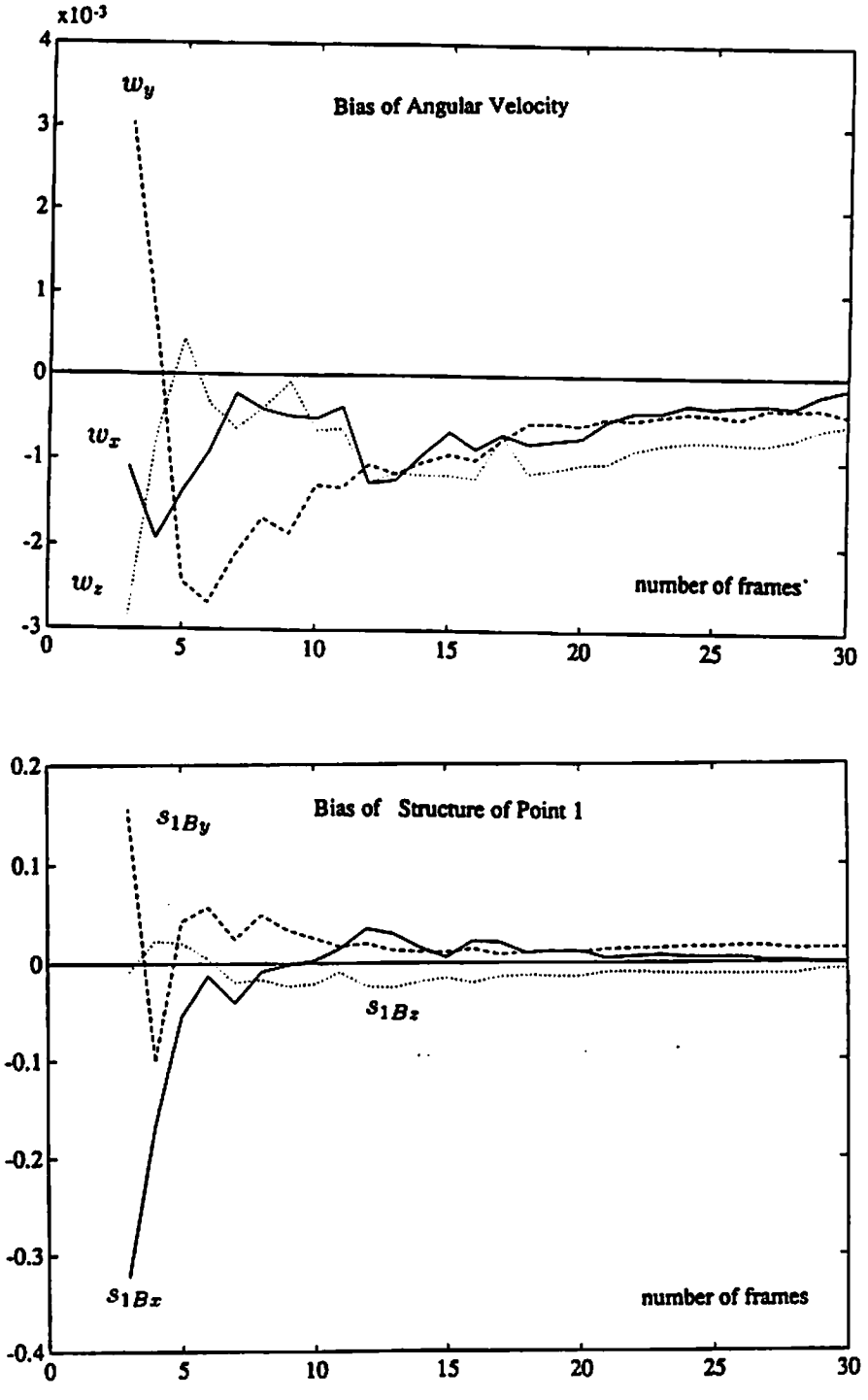


Figure 4.7: (continued) (c) angular velocity bias and (d) structure bias.

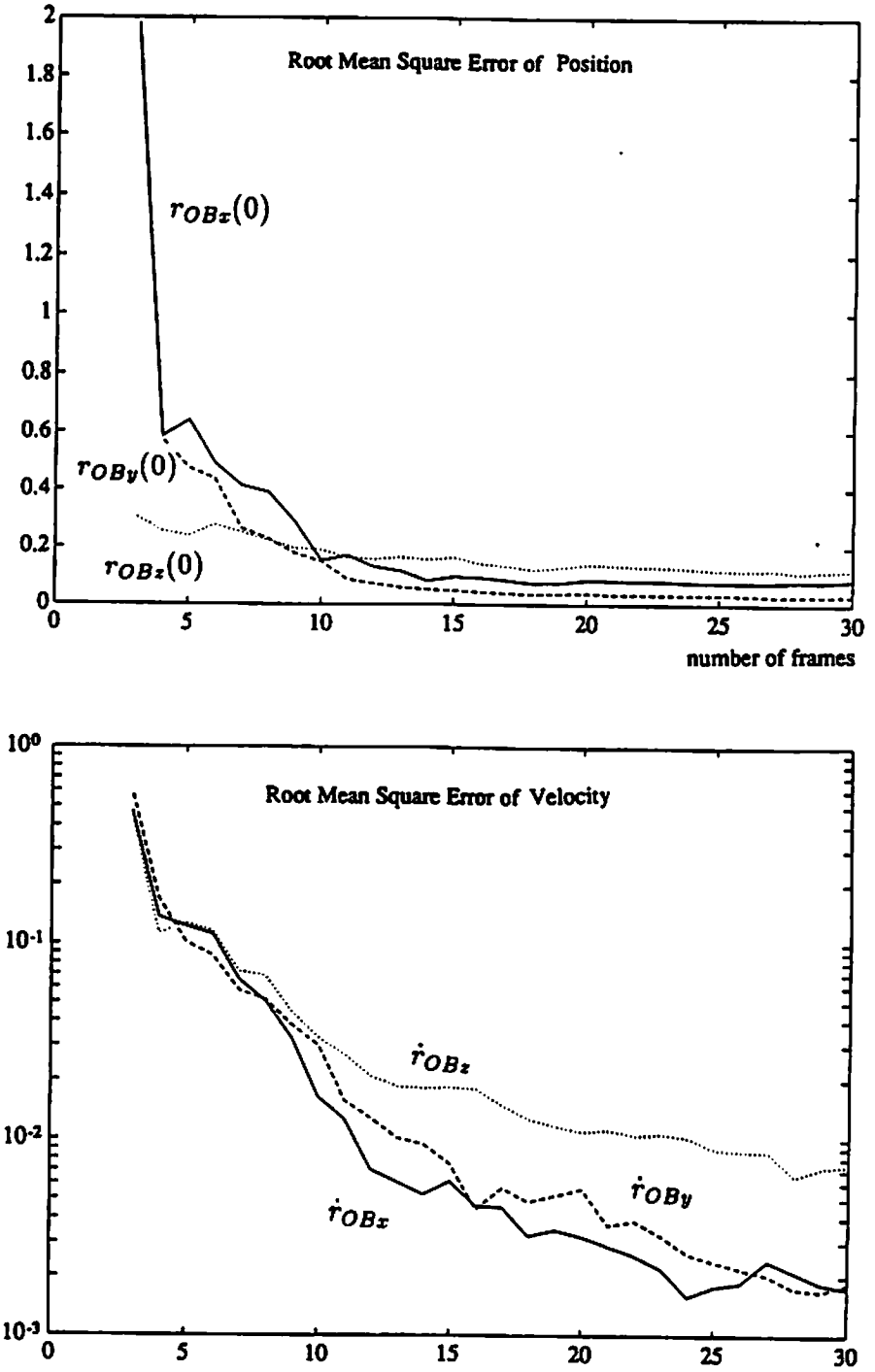
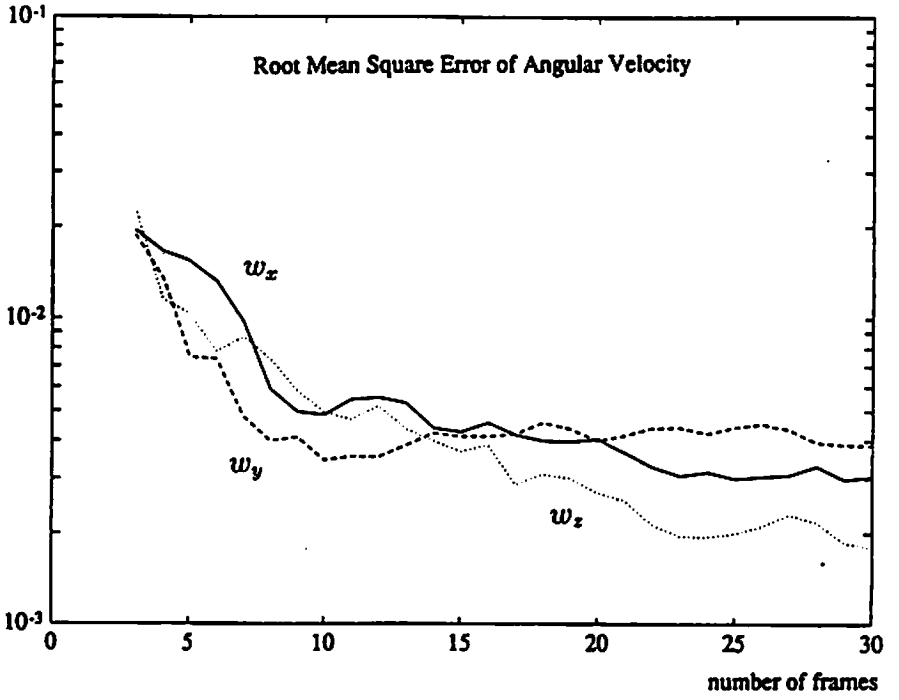


Figure 4.8: Root mean square error of motion and structure parameters from 50 Monte Carlo trials for the binocular vision. (a) translational position and (b) velocity.



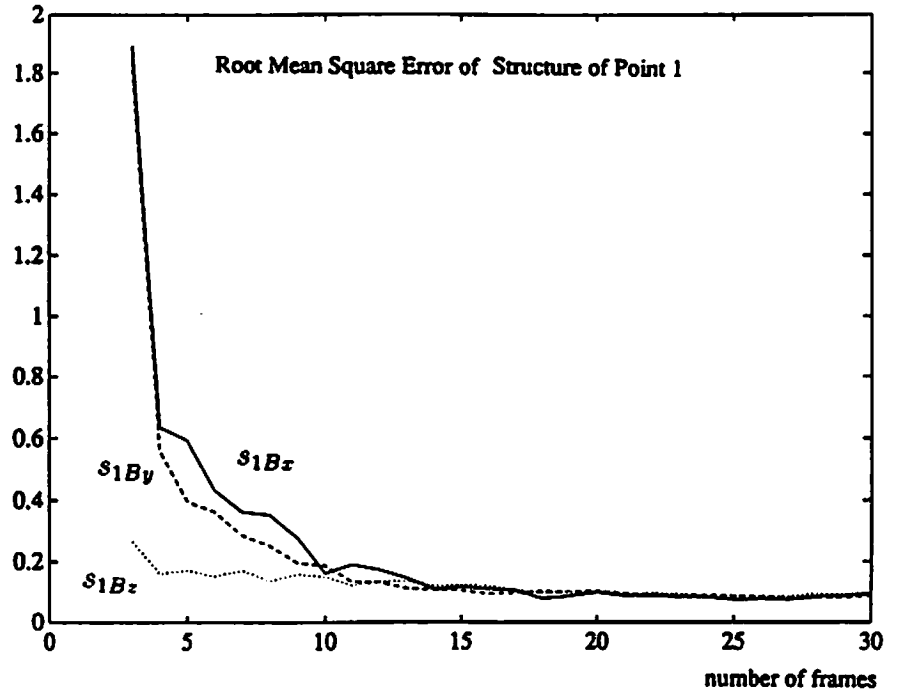


Figure 4.8: (continued) (c) angular velocity and (d) structure.

Chapter 5

Statistical Analysis of Inherent Ambiguities in Recovering 3-D Motion from A Noisy Flow Field

In this chapter, the inherent ambiguities in recovering 3-D motion information from a single optical flow field are studied using a statistical model. These ambiguities are quantified using the Cramer-Rao lower bound (CRLB), which is a lower bound for the error variances of motion parameter estimates. This performance bound is independent of the motion estimation algorithms, and can always be computed for any arbitrary 3-D motion of a rigid surface by inverting a 5×5 matrix. As a special case, the performance bound for the motion of 3-D rigid planar surfaces is studied in detail. The dependence of the bound on several factors such as, the underlying motion, surface position, surface orientation, the field of view, and the density of available pixels are derived as closed form expressions. A subset of our results support Adiv's recent analysis of the inherent ambiguities of motion parameters. For the general motion of an arbitrary surface, it turns out that not every pixel gives information regarding 3-D motion estimation. We show that the aperture problem in computing the optical flow restricts the nontrivial information about the 3-D motion to a sparse set of pixels at which both components of the flow velocity are observable. Computer simulations are used to study the dependence of the inherent ambiguities on the

underlying motion, the field of view, and the number of feature points for the motion in front of a nonplanar environment. Effects of two smoothing schemes on estimation accuracy are analyzed. It is shown that by introducing a smoothness constraint by fitting local patches to 3–D depths gives lower CRLB's. Surprizsingly, this reduction of CRLB's is very small. Further, fitting local patches also relaxes the aperture problem since the motion information is not restricted to the points at which both optical flow components are observable. In contrast, imposing smoothness on the optical flow by regularization does not lower the CRLB's. Section 5.1 derives the CRLB's of motion parameters for the general 3-D motion of a surface. In Section 5.2 we discuss the inherent motion ambiguities caused by the aperture problem. It is also shown that "solving" the aperture problem by regularization does not lower the CRLB's. The general motion of a planar surface is studied in Section 5.3. Computer simulation results for the general motion of a nonplanar surface are presented in Section 5.4. Section 5.5 investigates the effects of fitting local patches to the 3-D surface. Section 5.6 discusses some issues in the interpretation of this work.

5.1 Cramer-Rao Lower Bound for Unbiased Motion Parameter Estimators

Recovering motion parameters from an optical flow field is an estimation problem in which the deterministic unknown parameters $\underline{\theta}$ (motion and structure parameters) are estimated from the noise contaminated image measurements $\underline{\mathbf{z}}$ (the velocity components of the optical flow at every pixel). The result given by a motion estimation algorithm is a motion and structure estimator $\hat{\underline{\theta}}$, computed using the measurements $\underline{\mathbf{z}}$. Due to the random nature of the noisy image measurements $\underline{\mathbf{z}}$, any motion and structure estimator $\hat{\underline{\theta}}$ will thus have an uncertainty. (In this chapter, all random variables are printed in bold face, and estimators are denoted using $\hat{}$ symbol. Certainly, estimators are random variables.) One way to quantify the goodness of an unbiased estimator $\hat{\underline{\theta}}$ is the error covariance

matrix of $\hat{\underline{\theta}}$ conditioned on the true motion and structure parameters $\underline{\theta}$,

$$E\left\{(\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^T\middle|\underline{\theta}\right\} \tag{5.1}$$

where E is the expectation operator. Motion and structure estimators $\hat{\underline{\theta}}$ with smaller error covariance are considered to be better since less uncertainty is involved. Assume that the conditional probability density function

$$f(\underline{\mathbf{z}}|\underline{\theta})$$

of $\underline{\mathbf{z}}$ given $\underline{\theta}$ is known. Then the uncertainty of an unbiased estimator $\hat{\underline{\theta}}$ is bounded below by the Cramer-Rao inequality [63]

$$E\left\{(\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^T\middle|\underline{\theta}\right\} \geq J^{-1} \tag{5.2}$$

where J^{-1} is the inverse (assuming it exists) of the Fisher information matrix

$$J = E\left\{\left[\frac{\partial}{\partial\underline{\theta}}\ln f(\underline{\mathbf{z}}|\underline{\theta})\right]^T\left[\frac{\partial}{\partial\underline{\theta}}\ln f(\underline{\mathbf{z}}|\underline{\theta})\right]\middle|\underline{\theta}\right\} \tag{5.3}$$

The definition of the inequality in (5.2) is: for any two symmetric matrices A and B, $A \geq B$ is defined to mean that $(A-B)$ is positive semidefinite. Hence (5.2) implies that the matrix

$$E\left\{(\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^T\middle|\underline{\theta}\right\} - J^{-1}$$

is positive semidefinite. Because all positive semidefinite matrices have nonnegative diagonal terms, $E\left\{(\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^T\middle|\underline{\theta}\right\} - J^{-1}$ has nonnegative diagonal terms. Therefore

$$\begin{aligned} &\text{the } i^{th} \text{ diagonal term of } E\left\{(\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^T\middle|\underline{\theta}\right\} \\ \geq\ &\text{the } i^{th} \text{ diagonal term of } J^{-1} \end{aligned} \tag{5.4}$$

Let us denote the i^{th} component in $\underline{\hat{\theta}}$ by $\hat{\theta}_i$. Since the i^{th} diagonal term of $E\{(\underline{\hat{\theta}} - \underline{\theta})(\underline{\hat{\theta}} - \underline{\theta})^T | \underline{\theta}\}$ is the error variance of $\hat{\theta}_i$, (5.4) becomes

$$\text{error variance of } \hat{\theta}_i \quad \geq \quad \text{the } i^{th} \text{ diagonal term of } J^{-1} \quad (\equiv \text{ CRLB of } \theta_i) \tag{5.5}$$

In this context, the i^{th} diagonal term of J^{-1} is called the CRLB of θ_i, because it lower bounds the error variance of any estimator of θ_i. As the Fisher information matrix J is independent of $\underline{\hat{\theta}}$ by (5.3), the CRLB's are independent of the motion and structure estimator $\underline{\hat{\theta}}$ by (5.5). Hence no matter which unbiased estimator is applied, the error variances of the estimated parameters can never be reduced below the CRLB's. That is, CRLB's are independent of all unbiased motion estimation algorithms, and bound the error variances of the estimated motion parameters from below. Thus CRLB's give a quantitative description of the inherent uncertainty involved in motion estimation problems. On the other hand, CRLB's do depend on the 3-D motion and surface structure since J is a function of $\underline{\theta}$. Using this feature, the CRLB's may be used to study the question: what kind of 3-D motions and surfaces will cause inherent ambiguities in motion estimation? Furthermore, since J also depends on $\underline{z}$, the CRLB's depend on the type of measurements. After the CRLB's for different types of measurements are computed, the goodness of one type of measurements may be compared with the others. This comparison seems interesting especially in the context of a multi-sensor fusion system.

In particular, let us consider the case when the measurements are contaminated by independent identically distributed (IID) additive Gaussian noise. That is, we assume the measurement model

$$\underline{z} = \underline{h}(\underline{\theta}) + \underline{n} \tag{5.6}$$

where $\underline{h}$ is deterministic and the components of the noise vector $\underline{n}$ are IID Gaussian random variables with mean zero and variance σ^2. Then the Fisher information matrix (5.3) can be simplified as

$$\begin{aligned} J &= \frac{1}{\sigma^2}\left(\frac{\partial \underline{h}}{\partial \underline{\theta}}\right)^T \left(\frac{\partial \underline{h}}{\partial \underline{\theta}}\right) \\ &= \frac{1}{\sigma^2}\sum_i \left(\frac{\partial h_i}{\partial \underline{\theta}}\right)^T \left(\frac{\partial h_i}{\partial \underline{\theta}}\right) \end{aligned} \tag{5.7}$$

in which h_i is the i^{th} component of $\underline{h}$. In (5.7), the gradients $\partial h_i/\partial \underline{\theta}$ have been defined to be row vectors.

Consider the typical model used in [3, 24, 37] for estimating motion parameters from the optical flow field. Assume the scenario of a moving camera in a static environment and a 3-D coordinate system I fixed with respect to the camera, with the Z-axis pointing along the optical axis, as shown in Figure 5.1. If only rigid motions are considered, the motion can be decomposed into a translational velocity vector

$$(U', V', W')^T$$

and an angular velocity vector

$$(A, B, C)^T$$

about an axis through the origin. Let
$(X', Y', Z')^T$ = the instantaneous coordinates of a point P in the environment,
$(x, y)^T$ = the coordinates of the image of P on the image plane,
$(u, v)^T = (\dot{x}, \dot{y})^T$ = the motion flow at the point $(x, y)^T$.
Then the equations which relate the motion flow field to the structure and motion parameters are [37]

$$\begin{aligned} u &= (-\frac{U'}{Z'} - B + Cy) - x(-\frac{W'}{Z'} - Ay + Bx) \\ v &= (-\frac{V'}{Z'} - Cx + A) - y(-\frac{W'}{Z'} - Ay + Bx) \end{aligned} \tag{5.8}$$

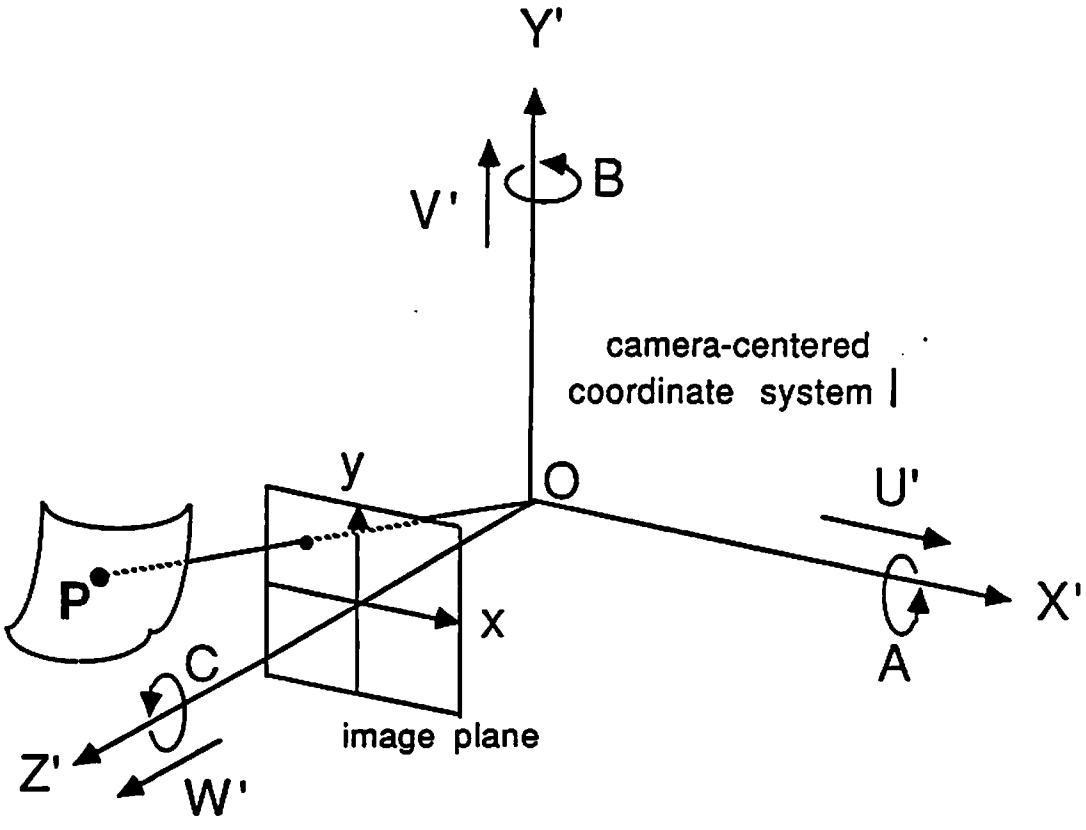


Figure 5.1: A camera-centered coordinate system (X', Y', Z'), and the corresponding image coordinates (x, y). The relative translation and rotation are (U', V', W') and (A, B, C), respectively.

Since the translational components can only be decided up to a global scale factor in a monocular system, we use the following normalization w.r.t. W' to account for this ambiguity.

$$\begin{aligned} \frac{1}{Z} &\equiv \frac{Z'}{W'} \\ U &\equiv \frac{U'}{W'} \\ V &\equiv \frac{V'}{W'} \end{aligned} \tag{5.9}$$

Z will be called the "normalized reciprocal depth." Substitution of (5.9) into (5.8) and subscribing the quantities which correspond to the i^{th} point by i, we obtain

$$\begin{aligned} u_i &= (-UZ_i - B + Cy_i) - x_i(-Z_i - Ay_i + Bx_i) \\ &\equiv h_{ui}(U, V, A, B, C, Z_i) \\ v_i &= (-VZ_i - Cx_i + A) - y_i(-Z_i - Ay_i + Bx_i) \\ &\equiv h_{vi}(U, V, A, B, C, Z_i) \end{aligned} \tag{5.10}$$

Assume that the noise in the extracted motion field obeys the following simple Gaussian model

$$\begin{cases} \mathbf{u}_i &= h_{ui}(U, V, A, B, C, Z_i) + \mathbf{n}_{ui} \\ \mathbf{v}_i &= h_{vi}(U, V, A, B, C, Z_i) + \mathbf{n}_{vi} \end{cases} \tag{5.11}$$

where $\mathbf{n}_{ui}$ and $\mathbf{n}_{vi}$ are IID zero mean Gaussian random variables with variance σ^2.

For the case when the optical flow field at M image points is available, the free parameters are the five motion parameters

$$\underline{\psi} \equiv (U, V, A, B, C)^T$$

and the M structure parameters

$$\underline{Z} \equiv (Z_1, Z_2, \cdots, Z_M)^T$$

The measurements are

$$u_i, v_i, \quad i = 1, 2, \cdots, M$$

Qualitatively, we can look at the following intuitively defined overestimation [5]

$$\begin{aligned} \text{overestimation} &\equiv \frac{\text{number of measurements}}{\text{number of parameters}} \\ &= \frac{2M}{5+M} \\ &\to 2 \qquad \text{as} \quad M \to \infty \end{aligned} \tag{5.12}$$

As the number of the image points grows, the overestimation basically remains as a constant: 2.

Quantitatively, the CRLB can be used to investigate the variances of the parameter estimates. Under the simple model (5.11), the Fisher information matrix is, by (5.7),

$$J = \frac{1}{\sigma^2} \sum_{i=1}^{M} \left[\left(\frac{\partial h_{ui}}{\partial \underline{\theta}} \right)^T \left(\frac{\partial h_{ui}}{\partial \underline{\theta}} \right) + \left(\frac{\partial h_{vi}}{\partial \underline{\theta}} \right)^T \left(\frac{\partial h_{vi}}{\partial \underline{\theta}} \right) \right] \tag{5.13}$$

where

$$\begin{aligned} \underline{\theta} &= (U, V, A, B, C, Z_1, Z_2, \cdots, Z_M)^T \\ &= (\underline{\psi}^T, \underline{Z}^T)^T \end{aligned}$$

Then the Fisher information matrix (5.13) has the partition

$$J = \begin{pmatrix} \alpha & \beta \\ \beta^T & \gamma \end{pmatrix} \tag{5.14}$$

where α is a 5×5 matrix and

$$\begin{aligned} \alpha &= \frac{1}{\sigma^2} \sum_{i=1}^{M} \left[\left(\frac{\partial h_{ui}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{ui}}{\partial \underline{\psi}} \right) + \left(\frac{\partial h_{vi}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{vi}}{\partial \underline{\psi}} \right) \right] \\ \beta &= \frac{1}{\sigma^2} \sum_{i=1}^{M} \left[\left(\frac{\partial h_{ui}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{ui}}{\partial \underline{Z}} \right) + \left(\frac{\partial h_{vi}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{vi}}{\partial \underline{Z}} \right) \right] \end{aligned} \tag{5.15}$$

$$\gamma = \frac{1}{\sigma^2}\sum_{i=1}^{M}\left[\left(\frac{\partial h_{ui}}{\partial \underline{Z}}\right)^T\left(\frac{\partial h_{ui}}{\partial \underline{Z}}\right)+\left(\frac{\partial h_{vi}}{\partial \underline{Z}}\right)^T\left(\frac{\partial h_{vi}}{\partial \underline{Z}}\right)\right]$$

From (5.10), we note that γ is a diagonal matrix with the j^{th} diagonal term

$$[\gamma]_{jj} = \frac{1}{\sigma^2}[(x_j - U)^2 + (y_j - V)^2]$$

Thus γ^{-1} is also a diagonal matrix with the j^{th} diagonal term

$$[\gamma^{-1}]_{jj} = \frac{\sigma^2}{(x_j - U)^2 + (y_j - V)^2} \tag{5.16}$$

Since the CRLB's for the motion parameters (U, V, A, B, C) are the first five diagonal terms of J^{-1}, they are the five diagonal terms of the following 5×5 matrix Q

$$J^{-1} = \begin{pmatrix} \alpha & \beta \\ \beta^T & \gamma \end{pmatrix}^{-1} \equiv \begin{pmatrix} Q & S \\ S^T & G \end{pmatrix} \tag{5.17}$$

When J is nonsingular, we have the identities [63]

$$\begin{aligned} Q &= (\alpha - \beta\gamma^{-1}\beta^T)^{-1} = \alpha^{-1} + \alpha^{-1}\beta G\beta^T\alpha^{-1} \\ G &= (\gamma - \beta^T\alpha^{-1}\beta)^{-1} = \gamma^{-1} + \gamma^{-1}\beta^T Q\beta\gamma^{-1} \\ S &= -Q\beta\gamma^{-1} = -\alpha^{-1}\beta G \end{aligned} \tag{5.18}$$

By (5.15), (5.16), and (5.18), Q can be calculated as

$$Q = (\alpha - \beta\gamma^{-1}\beta^T)^{-1} = \tag{5.19}$$

$$\sigma^2\left\{\sum_{j=1}^{M}\frac{\left[(y_j - V)\frac{\partial h_{uj}}{\partial \underline{\psi}} - (x_j - U)\frac{\partial h_{vj}}{\partial \underline{\psi}}\right]^T\left[(y_j - V)\frac{\partial h_{uj}}{\partial \underline{\psi}} - (x_j - U)\frac{\partial h_{vj}}{\partial \underline{\psi}}\right]}{(x_j - U)^2 + (y_j - V)^2}\right\}^{-1}$$

Therefore the CRLB's of the five motion parameters (U, V, A, B, C) are the diagonal terms of Q and can be computed for any 3-D motion and surface by inverting a 5×5 matrix. From (5.19), it can also be seen that the CRLB's are independent of the rotation parameters (A, B, C) and proportional to σ^2.

As large CRLB's indicate high uncertainty in estimates of motion parameters, we have the following observations for the ambiguity of parameters of a general motion in an arbitrary environment.

- The motion parameters are ambiguous if the noise level on the optical flow field is high.
- The ambiguity of motion parameters (both translational and rotational parameters) is independent of the true rotational motion.

5.2 The Aperture Problem

In the above computations, we assume that every point on the image is a "feature point" defined as the point at which both components of the image flow velocity are observable from the change of image brightness patterns. Isolated point-like features and salient points, e.g. corners, are examples of "feature points", which may be tracked and both components of their image motion determined. Due to the aperture problem [35, 36], "feature points" are usually sparse in images without very rich texture. As a consequence, for the majority of pixels, only the flow component parallel to the gradient vector of the image brightness pattern is observable. Thus at a regular point, at most one direct measurement is available. In the following, any point which is not a "feature point" will be called a "regular point." We shall show that regular points do not contribute information to CRLB's of motion parameters. Furthermore, it will be shown that imposing smoothness on the optical flow field by the regularization methods [35, 36, 41, 85] does not lower the CRLB's. In this sense, these regularization methods do not convert a feature point to a regular point, although both the optical flow components are "recovered" at each regular point. In contrast, we shall show in Section 5.5 that imposing smoothness on the 3-D depths by local quadratic surface fitting lowers the CRLB's by a very small amount.

Let us exclude the image points at which the gradient is zero, since no direct image velocity measurement is available at points within locally flat brightness regions due to the aperture problem. Consider the i^{th} regular point at which the

image brightness pattern has the gradient vector along

$$(\cos\theta_i, \sin\theta_i)^T \tag{5.20}$$

which we assume can be obtained without error. (Since only a lower performance bound is concerned, this should not cause any concern.) By the noise model (5.11), the optical flow component parallel to direction (5.20) is

$$\begin{aligned} \mathbf{w}_i &\equiv (\mathbf{u}_i, \mathbf{v}_i) \cdot (\cos\theta_i, \sin\theta_i) \\ &= h_{ui}\cos\theta_i + h_{vi}\sin\theta_i + \mathbf{n}_{ui}\cos\theta_i + \mathbf{n}_{vi}\sin\theta_i \\ &= h_i + \mathbf{n}_i \end{aligned} \tag{5.21}$$

where

$$\begin{aligned} h_i &\equiv h_{ui}\cos\theta_i + h_{vi}\sin\theta_i \\ \mathbf{n}_i &\equiv \mathbf{n}_{ui}\cos\theta_i + \mathbf{n}_{vi}\sin\theta_i \end{aligned} \tag{5.22}$$

According to (5.22), $\mathbf{n}_i$ are still IID Gaussian distributed with mean zero and variance σ^2. Let

$$\begin{aligned} M_f &= \text{number of feature points} \\ M_r &= \text{number of regular points} \end{aligned}$$

In the following, we will index the regular points by i and the feature points by j. Then the measurements are

$$\begin{aligned} &\text{regular points}: \quad \mathbf{w}_i, \quad i = 1, 2, \cdots, M_r \\ &\text{feature points}: \quad \mathbf{u}_j, \mathbf{v}_j, \quad j = 1, 2, \cdots, M_f \end{aligned}$$

and the free parameters are

$$\underline{\theta} = (U, V, A, B, C, Z_{r1}, Z_{r2}, \cdots, Z_{rM_r}, Z_{f1}, Z_{f2}, \cdots, Z_{fM_f})^T$$

In above, Z_{ri} is the normalized reciprocal depth of the i^{th} regular point and Z_{fj} is that of the j^{th} feature point. Here we can compute the overestimation as

$$\begin{aligned} & \frac{\text{number of measurements}}{\text{number of parameters}} \\ = & \frac{2M_f + M_r}{5 + (M_f + M_r)} \\ \rightarrow & \ 1 \qquad \text{if} \quad M_r \gg M_f \end{aligned}$$

In this more practical model, overestimation is only half of the previous case in (5.12). Since the number of measurements is very close to the number of parameters, the system is expected to be vulnerable to noise.

Considering the measurement model (5.21) for regular points and (5.11) for feature points, we have the Fisher information matrix, by (5.7),

$$J = \frac{1}{\sigma^2}\left\{\sum_{i=1}^{M_r}\left(\frac{\partial h_i}{\partial \underline{\theta}}\right)^T\left(\frac{\partial h_i}{\partial \underline{\theta}}\right) + \sum_{j=1}^{M_f}\left[\left(\frac{\partial h_{uj}}{\partial \underline{\theta}}\right)^T\left(\frac{\partial h_{uj}}{\partial \underline{\theta}}\right) + \left(\frac{\partial h_{vj}}{\partial \underline{\theta}}\right)^T\left(\frac{\partial h_{vj}}{\partial \underline{\theta}}\right)\right]\right\}$$

The partition of J into the form (5.14) gives

$$\alpha = \frac{1}{\sigma^2}\left\{\sum_{i=1}^{M_r}\left(\frac{\partial h_i}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_i}{\partial \underline{\psi}}\right) + \sum_{j=1}^{M_f}\left[\left(\frac{\partial h_{uj}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{uj}}{\partial \underline{\psi}}\right) + \left(\frac{\partial h_{vj}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{vj}}{\partial \underline{\psi}}\right)\right]\right\}$$

$$\beta = \frac{1}{\sigma^2}\left\{\sum_{i=1}^{M_r}\left(\frac{\partial h_i}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_i}{\partial \underline{Z}}\right) + \sum_{j=1}^{M_f}\left[\left(\frac{\partial h_{uj}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{uj}}{\partial \underline{Z}}\right) + \left(\frac{\partial h_{vj}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{vj}}{\partial \underline{Z}}\right)\right]\right\}$$

$$\gamma = \frac{1}{\sigma^2}\left\{\sum_{i=1}^{M_r}\left(\frac{\partial h_i}{\partial \underline{Z}}\right)^T\left(\frac{\partial h_i}{\partial \underline{Z}}\right) + \sum_{j=1}^{M_f}\left[\left(\frac{\partial h_{uj}}{\partial \underline{Z}}\right)^T\left(\frac{\partial h_{uj}}{\partial \underline{Z}}\right) + \left(\frac{\partial h_{vj}}{\partial \underline{Z}}\right)^T\left(\frac{\partial h_{vj}}{\partial \underline{Z}}\right)\right]\right\}$$

where the structure parameters are

$$\underline{Z} = (Z_{r1}, Z_{r2}, \cdots, Z_{rM_r}, Z_{f1}, Z_{f2}, \cdots, Z_{fM_f})^T$$

Then

$$Q = (\alpha - \beta\gamma^{-1}\beta^T)^{-1} = \tag{5.23}$$

$$\sigma^2 \left\{ \sum_{j=1}^{M_f} \frac{\left[(y_j - V)\frac{\partial h_{uj}}{\partial \underline{\psi}} - (x_j - U)\frac{\partial h_{vj}}{\partial \underline{\psi}} \right]^T \left[(y_j - V)\frac{\partial h_{uj}}{\partial \underline{\psi}} - (x_j - U)\frac{\partial h_{vj}}{\partial \underline{\psi}} \right]}{(x_j - U)^2 + (y_j - V)^2} \right\}^{-1}$$

which is independent of the regular points! Therefore only feature points contribute information to the CRLB's of motion parameters. That is,

- useful information to reduce the inherent ambiguity of motion parameters is contained only in the flow field at the feature points.

We now relate the above result to the computational theories of Hildreth [35, 36], Horn and Schunck [41], and Yuille and Grzywacz [85], which apply regularization methods to obtain both the flow components at each regular point. We shall show that these extracted components do not lower the CRLB's computed in (5.23). In this sense, the derived image flow field does not convert a regular point to a feature point, in spite of the fact that both the flow components are "recovered" at each regular point.

Hildreth [35, 36] takes the normal components of the image velocity field along previously extracted contours as direct measurements, and then fills in and smoothes the velocity vectors along the contours by minimizing a cost function obtained by using the regularization formulation. More precisely, the estimated image flow field $\hat{\underline{v}}(s)$ on the contour C is extracted by minimizing [35, 36, 85]

$$\mathbf{E}_H[\underline{v}(s)] = \int_C [\underline{v}(s) \cdot \underline{n}(s) - \mathbf{u}(s)]^2 + \lambda \int_C \frac{\partial \underline{v}(s)}{\partial s} \cdot \frac{\partial \underline{v}(s)}{\partial s} \qquad (5.24)$$

with respect to $\underline{v}(s)$, where $\underline{n}(s)$ is the unit normal vector to the curve C and $\mathbf{u}(s)$ is the measured normal velocity field as a function of the arc length s. If an upper bound for the performance of motion estimation based on the extracted field $\hat{\underline{v}}(s)$ is to be derived, we may consider the case when the unit normal vectors $\underline{n}(s)$ are measured without noise. Since $\hat{\underline{v}}(s)$ minimizes $\mathbf{E}_H$, which depends on the direct measurements $\mathbf{u}(s)$, the extracted field $\hat{\underline{v}}(s)$ is a deterministic function of $\mathbf{u}(s)$, i.e.,

$$\hat{\underline{v}}(s) = T(\{\mathbf{u}(s), \text{ all } s\}) \qquad (5.25)$$

Now let us recall the following theorem from statistical theory [44]. If Ω_1 and Ω_2 are two sets of measurements such that every element in Ω_1 is a deterministic function of all elements in Ω_2, then the Fisher information of Ω_1 is smaller than or equal to that of Ω_2. That is

$$J_{\Omega 1} \leq J_{\Omega 2}$$

where the equility holds if and only if Ω_1 is sufficient for Ω_2. Let F denote the set of measurements of both components at all feature points. Then (5.25) implies that every element in the set

$$\Omega_{extrac} \equiv F \cup \{\underline{\hat{\mathbf{v}}}(s), \quad \text{all} \quad s\}$$

is a deterministic function of all elements in the set of direct measurements

$$\Omega_{direct} \equiv F \cup \{\mathbf{u}(s), \quad \text{all} \quad s\}.$$

which is the set of the both components at each feature point and the normal component at each regular point. Consequently, by the above theorem, we have

$$J_{\Omega extrac} \leq J_{\Omega direct} \tag{5.26}$$

where $J_{\Omega extrac}$ and $J_{\Omega direct}$ are the Fisher information matrices of the motion and structure parameters based on the measurements Ω_{extrac} and Ω_{direct}, respectively. Suppose enough feature points have been incorporated so that the Fisher information matrices are nonsingular, then (5.26) gives

$$J^{-1}_{\Omega direct} \leq J^{-1}_{\Omega extrac} \tag{5.27}$$

Note that the diagonal terms of $J^{-1}_{\Omega extrac}$ are the CRLB's for motion and structure estimation based on the extracted optical flow field Ω_{extrac}. The CRLB's based on the direct measurements Ω_{direct} are the diagonal terms of $J^{-1}_{\Omega direct}$, the first five of which are the CRLB's of the five motion parameters obtained from (5.23). Therefore, from (5.27), we see that extracting the optical flow field by Hildreth's

method does not lower the CRLB's. In this sense, the extracted field does not convert a regular point to a feature point. Useful information to improve the performance bound is still contained only in the measurements from feature points. Let us remark that the equality in (5.26) holds if and only if the statistic T in (5.25) is sufficient [44]. However, the extracted field is generally not a sufficient statistic of the direct measurements because except some very restrictive motions, e.g. rigid pure translational motion parallel to the image plane, the extracted field is different from the real image velocity field and violates the rigidity assumption even when the images are noise free [35, 36]. The loss of information due to the insufficiency of T prevents algorithms based on the extracted field from attaining the CRLB's in (5.23), which are the performance bound for algorithms based on the direct measurements.

Horn and Schunck [41] extract the image flow field all over the image plane using a scheme similar to Hildreth's with the aid of constant brightness assumption. If the image intensity at position $\underline{s}$ and time t is denoted by $I(\underline{s}, t)$, the constant brightness assumption imposes the optical flow constraint

$$\nabla I \cdot \underline{v}(\underline{s}) \;=\; -\mathbf{I_t} + \mathbf{n} \tag{5.28}$$

where ∇I and $\mathbf{I_t}$ are the gradient and the temporal derivative of the image intensity, respectively, and $\mathbf{n}$ is the random noise. To obtain an upper performance bound, in (5.28) we have considered the case when the gradient direction ∇I is noiseless. Again only one component, i.e. the component in the direction of the image gradient, is available. The normal unit vector to the contour of constant gray level is

$$\underline{n}(\underline{s}) \;=\; \frac{\nabla I}{||\nabla I||} \tag{5.29}$$

We have, from (5.28),

$$\underline{n}(\underline{s}) \cdot \underline{v}(\underline{s}) \;=\; -\frac{\mathbf{I_t}}{||\nabla I||} + \mathbf{n}' \tag{5.30}$$

where the random noise term is $\mathbf{n}' \equiv \mathbf{n}/||\nabla I||$. Because $\underline{n}(\underline{s}) \cdot \underline{v}(\underline{s})$ is the normal component of $\underline{v}(\underline{s})$ to the contour of constant gray level, using (5.30) we may

define the measured normal component

$$\mathbf{u}(\underline{s}) \equiv -\frac{\mathbf{I_t}}{||\nabla I||} \tag{5.31}$$

In accordance with the optical flow constraint (5.28), Horn and Schunck extract $\hat{\underline{v}}(\underline{s})$ by minimizing an energy function [41, 85]

$$\begin{aligned}\mathbf{E}_{HS}[\underline{v}(\underline{s})] &= \int [\nabla I \cdot \underline{v}(\underline{s}) + \mathbf{I_t}]^2 + \lambda \int \nabla \underline{v}(\underline{s}) \cdot \nabla \underline{v}(\underline{s}) \\ &= \int [\underline{n}(\underline{s}) \cdot \underline{v}(\underline{s}) - \mathbf{u}(\underline{s})]^2 ||\nabla I||^2 + \lambda \int \nabla \underline{v}(\underline{s}) \cdot \nabla \underline{v}(\underline{s})\end{aligned}$$

with respect to $\underline{v}(\underline{s})$, where the second equality holds by (5.29) and (5.31). Notice the similarity between this energy function $\mathbf{E}_{HS}$ and Hildreth's $\mathbf{E}_H$ in (5.24). Using arguments similar to those following (5.24), we see that the extracted field is a deterministic function of the direct measurements, and thus does not lower the CRLB's based on the direct measurements.

Yuille and Grzywacz [85] propose to extract the flow field $\hat{\underline{v}}(\underline{s})$ by minimizing

$$\begin{aligned}\mathbf{E}_{YG}[\underline{v}(\underline{s}), V_{ia}] &= \psi_1 \sum_i ||\underline{v}(\underline{s}_i) - \underline{\mathbf{U}}(\underline{s}_i)||^2 \\ &+ \psi_2 \int_C [\underline{v}(s) \cdot \underline{n}(s) - \mathbf{u}(s)]^2 \\ &+ \psi_3 \int [\nabla I \cdot \underline{v}(\underline{s}) + \mathbf{I_t}]^2 \\ &+ \psi_4 \sum_{ia} V_{ia} ||\underline{v}(\underline{s}_i) - (\underline{\mathbf{s}}_a - \underline{\mathbf{s}}_i)||^2 \\ &+ \lambda \int \sum_{m=0}^{\infty} c_m (D^m \underline{v})^2\end{aligned} \tag{5.32}$$

with respect to $\underline{v}(\underline{s})$ and V_{ia}. The first term in (5.32) takes into account the measured flow velocity $\underline{\mathbf{U}}(\underline{s}_i)$ at the i^{th} previously matched feature point. The motion of contours and the spatial and temporal derivatives of image intensity are considered in the second and third terms, respectively. The motion of extracted, but not previously matched salient points is considered in the fourth term, in which $V_{ia} = 1$ if the i^{th} salient point in the first frame is matched to the a^{th} one in the second frame, and is 0 otherwise. The fifth term is the smoothness function. An upper performance bound may be obtained by considering the case when the

gradient direction ∇I is noiseless and all the features are matched. Then again the extracted field is a deterministic function of the direct measurements. The CRLB's based on the direct measurements are not lowered by the extracted field.

The above discussions show that, in terms of the information about the 3–D motion and structure parameters, the computational theories considered in [35, 36, 41, 85] do not convert a regular point to a feature point. The extracted flow field generally contains less information than the direct measurements, which are composed of both components at each feature point and only one component at each regular point. In a noisy environment, the aperture problem causes the regular points to contribute no information to the estimation of motion parameters. Methods in [35, 36, 41, 85] do not solve this problem in spite of their agreements with experiments on variations of the aperture problem in human vision [52, 53]. Horn, Negahdaripour, and Weldon [40, 54, 55] are among the first to point out that the extracted flow field is not a necessary step in the estimation of 3–D motion and structure. They develop "direct" motion estimation algorithms based on the direct measurements. The CRLB's computed in (5.23) constitute a performance bound on these direct algorithms since the bounds are established for the direct measurements.

With further assumptions, there are cases when the measurements at regular points do contribute to the CRLB's. So far we have only dealt with cases when the depth at one pixel is independent of the depths at other pixels. If the depths at different pixels are related to each other, e.g. when the environment is assumed to be smooth, the regular points generally contribute information to the CRLB's. Section 5.5 shows such an example. Another case is when more than two images are considered and the kinematic motion models [18, 20, 21] are imposed to establish the motion relationship at different time instants.

5.3 Motion in a Planar Environment

Analytic results of CRLB's can be obtained when the camera is moving in front of a planar environment. The study of the planar case may bring insights to the general conditions under which the motion parameter estimation is inherently

unstable [4]. We will show that a subset of the results obtained in this chapter support Adiv's analyses of inherent ambiguities of motion parameters [4]. The dependence of these inherent ambiguities on several factors such as the underlying motion, surface position, surface orientation, the field of view, and the density of available pixels will be derived as closed form expressions.

Assume a planar environment

$$Z' = pX' + qY' + s \tag{5.33}$$

Then

$$\begin{aligned} \frac{1}{Z'} &= \frac{1}{s} - \frac{p}{s}\left(\frac{X'}{Z'}\right) - \frac{q}{s}\left(\frac{Y'}{Z'}\right) \\ &= \frac{1}{s} - \frac{p}{s}x - \frac{q}{s}y \end{aligned} \tag{5.34}$$

in which the last equality holds by the central projection of the camera with the focal length normalized to 1. Using the abbreviations

$$\begin{aligned} Z_0 &\equiv W'/s \\ Z_x &\equiv -pZ_0 \\ Z_y &\equiv -qZ_0 \end{aligned} \tag{5.35}$$

and (5.9), (5.34) can be written as

$$Z = Z_x x + Z_y y + Z_0$$

Without loss of generality, let us assume in this section that every point on the image is a feature point. Let d be the inter-pixel distance on the image and Ω be the field of view. Then (5.19) can be approximated by

$$Q = \tag{5.36}$$

$$\sigma^2 \left\{ \frac{1}{d^2} \iint_{\Omega} \frac{\left[(y-V)\frac{\partial h_u}{\partial \underline{\psi}} - (x-U)\frac{\partial h_v}{\partial \underline{\psi}}\right]^T \left[(y-V)\frac{\partial h_u}{\partial \underline{\psi}} - (x-U)\frac{\partial h_v}{\partial \underline{\psi}}\right]}{(x-U)^2 + (y-V)^2} dx\, dy \right\}^{-1}$$

The double integral may be evaluated using a polar coordinate system centered at $(U,V)^T$ if the field of view Ω on the image plane is

$$\Omega = \{(x,y)^T \,|\, (x-U)^2 + (y-V)^2 \leq r^2\} \tag{5.37}$$

Then (5.36) becomes

$$Q(U,V,Z_x,Z_y,Z_0,r) = \frac{16\sigma^2}{\pi r^2 \rho_f} \times \begin{pmatrix} r^2(Z_x^2+3Z_y^2) & -2Z_xZ_yr^2 \\ +8k^2 & \\ -2Z_xZ_yr^2 & r^2(3Z_x^2+Z_y^2) \\ & +8k^2 \\ r^2(Z_xV-3Z_yU) & r^2(Z_xU-Z_yV) \\ -8UVk & -8(1+V^2)k \\ r^2(Z_xU-Z_yV) & r^2(3Z_xV-Z_yU) \\ +8(1+U^2)k & +8UVk \\ -4r^2Z_y-8Vk & 4r^2Z_x+8Uk \end{pmatrix}$$

$$\begin{pmatrix} r^2(Z_xV-3Z_yU) & r^2(Z_xU-Z_yV) & -4r^2Z_y \\ -8UVk & +8(1+U^2)k & -8Vk \\ r^2(Z_xU-Z_yV) & r^2(3Z_xV-Z_yU) & 4r^2Z_x \\ -8(1+V^2)k & +8UVk & +8Uk \\ r^2(3U^2+V^2)+ & 2r^2UV- & \\ 8[U^2V^2 & 8UV(2+ & 4U(r^2-2) \\ +(1+V^2)^2] & U^2+V^2) & \\ 2r^2UV- & r^2(U^2+3V^2)+ & \\ 8UV(2+ & 8[U^2V^2+ & 4V(r^2-2) \\ U^2+V^2) & (1+U^2)^2] & \\ 4U(r^2-2) & 4V(r^2-2) & 8(r^2+ \\ & & U^2+V^2) \end{pmatrix}^{-1} \tag{5.38}$$

where

$$k = Z_x U + Z_y V + Z_0$$

and the density of the feature points is

$$\rho_f = d^{-2}$$

From (5.38), the CRLB's, which are the diagonal terms of Q, are inversely proportional to the density of feature points ρ_f. Thus we have

- the motion parameters are ambiguous if the flow field is sparse

for a general motion in front of a planar environment. This result has been obtained in [4].

5.3.1 Motion with Perpendicular Translation And Arbitrary Rotation

In this case,

$$U = V = 0$$

The field of view (5.37) is a disk with radius r centered at the origin of the image plane coordinate system. Q in (5.38) can be simplified to

$$Q = \frac{8\sigma^2}{\pi r^4 \rho_f (Z_x^2 + Z_y^2)} \times$$

$$\begin{pmatrix} 2 & 0 & 0 & -2Z_0 & Z_y \\ 0 & 2 & 2Z_0 & 0 & -Z_x \\ 0 & 2Z_0 & 2Z_0^2 + (Z_x^2 + Z_y^2)r^2/4 & 0 & -Z_x Z_0 \\ -2Z_0 & 0 & 0 & 2Z_0^2 + (Z_x^2 + Z_y^2)r^2/4 & -Z_y Z_0 \\ Z_y & -Z_x & -Z_x Z_0 & -Z_y Z_0 & 3(Z_x^2 + Z_y^2)/4 \end{pmatrix}$$

The diagonal terms in Q are the CRLB's of the motion parameters

$$\text{CRLB}(U) = \text{CRLB}(V) = \frac{16\sigma^2}{\pi r^4 \rho_f (Z_x^2 + Z_y^2)} = \frac{16\sigma^2}{\pi r^4 \rho_f (p^2 + q^2)} \left(\frac{s}{W'}\right)^2$$

$$\text{CRLB}(A) = \text{CRLB}(B) = \frac{16\sigma^2 Z_0^2}{\pi r^4 \rho_f (Z_x^2 + Z_y^2)} + \frac{2\sigma^2}{\pi r^2 \rho_f} = \frac{16\sigma^2}{\pi r^4 \rho_f (p^2 + q^2)} + \frac{2\sigma^2}{\pi r^2 \rho_f} \tag{5.39}$$

$$\text{CRLB}(C) = \frac{6\sigma^2}{\pi r^4 \rho_f}$$

in which (5.35) has been used. Since large values of CRLB's indicate high uncertainty of motion parameters, the inherent ambiguities in motion estimation are quantified by (5.39). Adiv [4] found that the following conditions contribute to large uncertainties in recovering 3-D motion parameters. These conditions are also direct results of (5.39) because they all lead to large CRLB's.

- The field of view is small. (r is small.)
- The planar surface is at most moderately slanted. ($p^2 + q^2$ is small.)
- The object is far away. (s is large.)
- The absolute value of the translation is small. ($|W'|$ is small.)
- The resolution of the image is coarse. (σ is large or ρ_f is small.)

Furthermore, from (5.39) we have the following more specific observations about the motion considered in this subsection. The orientation (represented by p and q) of the plane affects the accuracy of U, V, A, and B, but not C. The position (represented by s) of the plane affects the accuracy of only the translational parameters, but has no effect on rotational parameters. Thus the accuracy of C is independent of the geometry (represented by p, q, and s) of the plane. If the slope of the plane is large (thus $p^2 + q^2$ is large), the accuracy of the translational parameters is high, while the accuracy of rotational parameters is essentially independent of the slope of the plane as long as the second term in (5.39) dominates. It is difficult to estimate the translational parameters if the perpendicular translation is small (i.e. $|W'|$ is small) while the accuracy of rotational parameters is independent of the perpendicular translation.

5.3.2 General Motion

In this subsection we will first inspect the transformation of CRLB's under the rotation of the 3-D coordinate system I with respect to the optical axis. Then CRLB(U)+CRLB(V), CRLB(A)+CRLB(B), and CRLB(C) can be seen as three invariants under this rotation of coordinate system, and thus give a good indication of the inherent ambiguities in motion estimation. It will be shown that the three invariants under the general motion (U, V, A, B, C) are the same as those under the motion $(\sqrt{U^2+V^2}, 0, A'', B'', C'')$. ($A''$, B'', and C'' are arbitrary because CRLB's are independent of the rotation parameters, as shown in Section 5.1.) This fact will be used to simplify the case of general motion to the case in which the translational motion is parallel to the X'–Z' plane. The latter case will be discussed in the next subsection.

Let us look at the transformation of motion and structure parameters if we rotate the 3-D coordinate system I defined in Section 5.1 by an angle θ with respect to the optical axis of the camera to obtain a new coordinate system I''. Using the abbreviation

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

we have the following transformations

$$\begin{pmatrix} X'' \\ Y'' \\ Z'' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

$$\begin{aligned} (U'', V'')^T &= R_\theta\, (U, V)^T \\ (A'', B'')^T &= R_\theta\, (A, B)^T \\ (Z_x'', Z_y'')^T &= R_\theta\, (Z_x, Z_y)^T \\ (p'', q'')^T &= R_\theta\, (p, q)^T \end{aligned}$$

$$\begin{aligned} C'' &= C \\ Z_0'' &= Z_0 \\ s'' &= s \end{aligned}$$

Using the above transformations, it may be verified from (5.38) that

$$[Q(U'', V'', Z_x'', Z_y'', Z_0'', r)]^{-1} = R\,[Q(U, V, Z_x, Z_y, Z_0, r)]^{-1} R^T$$

and thus

$$Q(U'', V'', Z_x'', Z_y'', Z_0'', r) = R\,Q(U, V, Z_x, Z_y, Z_0, r)\,R^T \tag{5.40}$$

where R is the block diagonal unitary matrix

$$R = \begin{pmatrix} R_\theta & & \\ & R_\theta & \\ & & 1 \end{pmatrix}$$

Because the CRLB's are diagonal terms of Q, from (5.40), we have

$$\begin{aligned} \mathrm{CRLB}(U'') + \mathrm{CRLB}(V'') &= \mathrm{CRLB}(U) + \mathrm{CRLB}(V) \\ \mathrm{CRLB}(A'') + \mathrm{CRLB}(B'') &= \mathrm{CRLB}(A) + \mathrm{CRLB}(B) \\ \mathrm{CRLB}(C'') &= \mathrm{CRLB}(C) \end{aligned}$$

Therefore three invariants under coordinate rotation are obtained. Specifically let

$$\begin{cases} \cos\theta = U/\sqrt{U^2 + V^2} \\ \sin\theta = V/\sqrt{U^2 + V^2} \end{cases}$$

That is, the 3-D coordinate system I is rotated until the x-axis coincides with the projection of the translational velocity on the image plane. Or equivalently, we twist the camera axis so that the projected translation on the image plane is purely along the x–axis. Then

$$\begin{cases} U'' = \sqrt{U^2 + V^2} \\ V'' = 0 \end{cases}$$

Hence the three invariants in the set of

$$\begin{cases} \text{motion parameters} & : & (U, V, A, B, C) \\ \text{structure parameters} & : & (p, q, s) \end{cases}$$

are the same as the three invariants in the set of

$$\begin{cases} \text{motion parameters} & : & (\sqrt{U^2+V^2}, 0, A'', B'', C) \\ \text{structure parameters} & : & (\frac{pU+qV}{\sqrt{U^2+V^2}}, \frac{-pV+qU}{\sqrt{U^2+V^2}}, s) \end{cases}$$

Therefore the study of general motion can be reduced to the study of the motion with $V = 0$, which will be discussed in the next section using the three invariants defined in this section.

5.3.3 Motion with $V = 0$

Consider the motion $(U, 0, A, B, C)$ with the structure parameters (Z_x, Z_y, Z_0) and the radius of field of view r. In this case the diagonal terms of $Q(U, 0, Z_x, Z_y, Z_0, r)$ may be calculated using the MACSYMA computer package to symbolically invert the 5×5 matrix in (5.38). Then the three invariants are obtained as

$$\begin{aligned} & \text{CRLB}(U) + \text{CRLB}(V) \\ = \; & \eta(s/W')^2 \left\{ (p+U)^2 \left[2r^2U^2 + (11U^4 + 24U^2 + 16) \right] \right. \\ + \; & q^2 \left[r^2(U^6 + 2U^4 + 2U^2) \right. \\ + \; & \left. \left. (3U^8 + 14U^6 + 35U^4 + 40U^2 + 16) \right] \right\} \end{aligned}$$

$$\begin{aligned} & \text{CRLB}(A) + \text{CRLB}(B) \\ = \; & \eta \left\{ (p+U)^2 \left[2r^2(p^2+q^2) + 11(Up - 12/11)^2 + 32/11 \right] \right. \\ + \; & q^2 \left[r^2(U^4 + 2U^2 + 2)(p^2+q^2) + 3(3U^2+4)(1-Up)^2 \right. \\ + \; & \left. \left. (U^2+1)[(3U^4+2U^2)p^2 + 3(U^4+U^2)q^2 + 3U^2 + 4] \right] \right\} \end{aligned} \tag{5.41}$$

$$
\begin{aligned}
& \mathrm{CRLB}(C) \\
= \; & \eta\left\{(p+U)^2\left[r^2U^2p^2+(3U^4+2U^2+3)p^2-2U(3U^2+1)p+3U^2\right]\right. \\
+ \; & q^2\left[r^2U^2p^2+(2U^4+8U^2+6)p^2+3(U^2+1)^2q^2\right. \\
- \; & \left.\left.(6U^5+2U^3-4U)p+6U^4+6U^2\right]\right\}
\end{aligned}
$$

where

$$\eta = \frac{2\sigma^2}{\pi r^4 \rho_f[(p+U)^2+q^2(U^2+1)]^2}$$

From the above equations, we conclude that

- the translational parameters are ambiguous if the object is far away. (s is large.)
- the uncertainty of rotational parameters is independent of the distance of the object.

To check the dependence of the invariants on the slope of the plane, we note that the invariants are rational functions and have the only singularity at

$$(p,q)^T = (-U,0)^T \tag{5.42}$$

The behavior around the singularity may be investigated by the change of variables

$$\begin{cases} p = -U + \rho\cos\phi \\ q = \rho\sin\phi \end{cases} \tag{5.43}$$

in (5.41) and letting ρ tend to zero. That is, we are looking for the limit of the invariants when approaching the singularity along a straight line with the directional angle ϕ in the (p,q) space. After the substitution of (5.43) into (5.41), the three invariants take the form

$$\frac{\beta\rho^2+\gamma\rho+\delta}{\alpha\rho^2}, \qquad \alpha>0, \quad \delta>0$$

Therefore as ρ tends to zero, the three invariants tend to positive infinity. This behavior of the three invariants around the singularity $(-U, 0)^T$ in the $(p, q)^T$ space is illustrated in Figure 5.2. (In this example, the motion of the planar surface $Z' = pX' + qY' + 10$ is considered. The translational motion is $(U', V', W') = (0.5, 0, 1)$. The radius of the field of view is $r = 1.5$. The density of feature points is $\rho_f = 100$. σ^2 is the noise variance of the optical flow components. As the three invariants are proportional to σ^2 by (5.41), we draw the heights in terms of the multiplicity of σ^2. The dependence of the three invariants on the slope (p, q) of the plane is shown in Figure 5.2 in the range $-10 \leq p, q \leq 10$. The scales of the $p-$ and $q-$ axes in each diagram in Figure 5.2 are the same. However, the scales of the vertical axes are different from each other in these diagrams. We mark the height of one point in each diagram to show the difference.) By (5.9) and (5.42), the singularity of the three invariants happens when

$$(p, q, -1)^T = -(U', 0, W')^T / W'$$

i.e. when the directional vector of the planar surface $(p, q, -1)^T$ is parallel to the direction of the translational motion $(U', 0, W')^T$. Thus

- the motion parameters are ambiguous if the direction of the planar environment is at most slightly different from the direction of the translational motion.

To check the dependence of the invariants on the radius of the field of view r, we note that the three invariants (5.41) are all in the form of

$$\frac{\beta r^2 + \gamma}{\alpha r^4}, \qquad \alpha > 0, \quad \beta \geq 0, \quad \gamma > 0 \tag{5.44}$$

The above is clear for CRLB(U)+CRLB(V) and CRLB(A)+CRLB(B) from (5.41). As for CRLB(C), (5.44) also holds by the proposition shown in Appendix A. Since the three invariants are the sum of two monotonically decreasing functions $\beta/(\alpha r^2)$ and $\gamma/(\alpha r^4)$, they are always decreasing with increased r. Moreover, as r tends to zero, the three invariants tend to positive infinity by (5.44). Therefore

- the motion parameters are ambiguous if the field of view is small.

The dependence of the three invariants on U is not monotonic, as illustrated in Figure 5.3. (The planar environment considered is $Z' = -0.5X' - 0.5Y' + 10$. The density of feature points is $\rho_f = 100$. The translational motion is $(U', V', W') = (U', 0, 1)$. The dependence of the three invariants on U and the radius of the field of view r is shown in Figure 5.3 in the range $-2 \leq U \leq 2$ and $0.8 \leq r \leq 3$.) Because the 3-D translational vector $(U', 0, W')^T$ has the central projection, called focus of expansion (FOE) [57],

$$(U'/W', 0/W')^T = (U, 0)^T$$

on the image plane, we have, by (5.41),

- the CRLB's of rotational parameters depends on the FOE $(U'/W', V'/W')^T$, instead of the 3-D translational vector $(U', V', W')^T$.

However, the CRLB's of translational parameters depend on FOE as well as the 3-D translational vector's perpendicular component W' to the image plane.

5.4 Motion in An Arbitrary Environment

Some insights may be obtained using simulations when analytic forms of CRLB's are not manageable. This handy feature of applying CRLB's to investigate the inherent ambiguities follows the fact that CRLB's can always be computed numerically using (5.19) once the surface and the motion are specified. Here we will consider a few examples of general motions in front of nonplanar environments using simulations.

Two simulated surfaces S_1 and S_2 are considered. The structure parameters Z' in (5.8) are plotted as a function of the image coordinates (x, y) in Fig. 5.4 for S_1 and S_2. The maximum value of Z' for S_1 is 13, minimum 5; for S_2, maximum 22.7, minimum 4.5. The dependence of the CRLB's of the five motion parameters on the translational parameters (U, V) is plotted in Fig. 5.5 and Fig. 5.6 for the motions of S_1 and S_2, respectively. The focal length is 1. The noise variance of

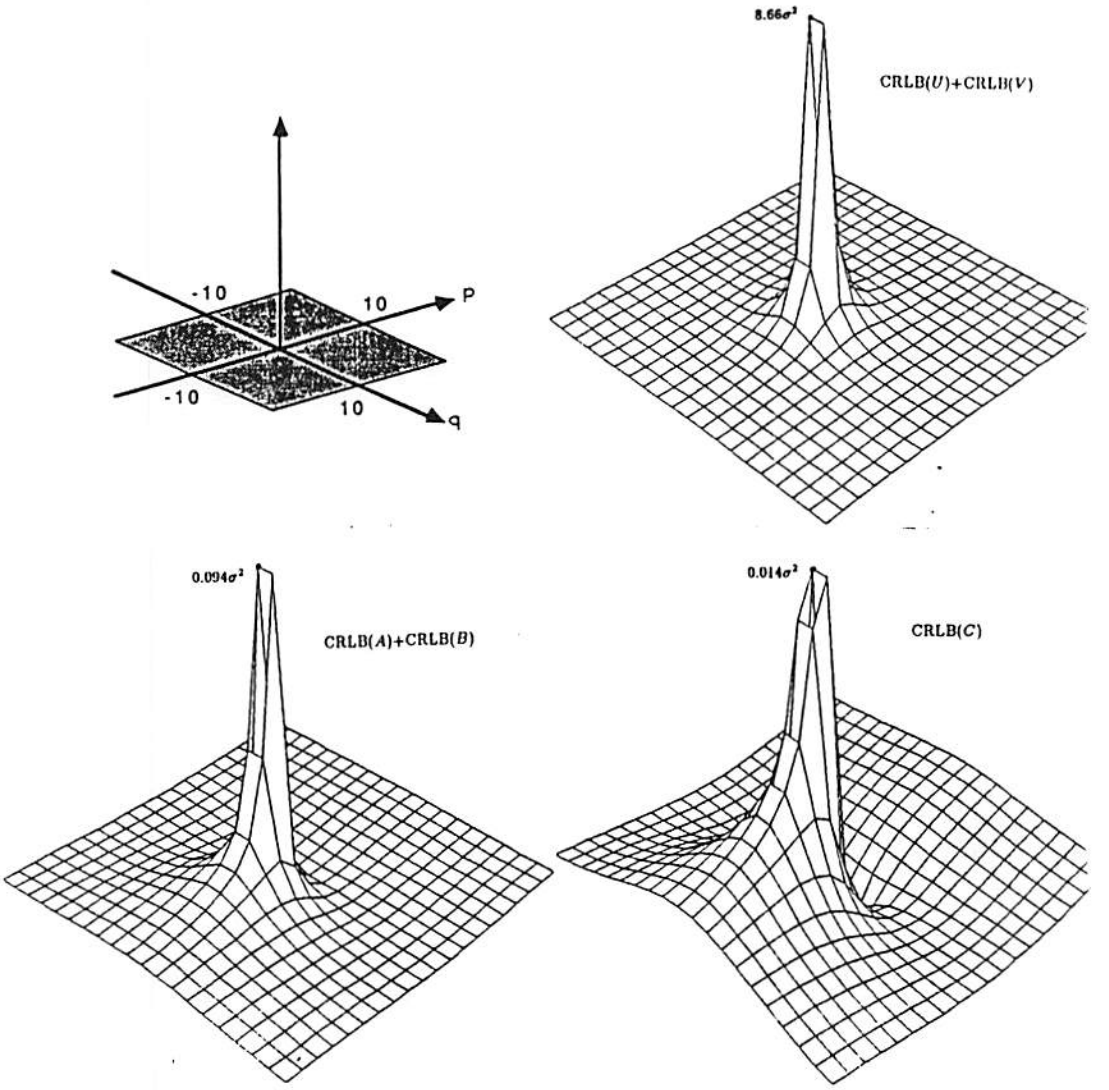


Figure 5.2: The dependence of the three invariants CRLB(U) + CRLB(V), CRLB(A) + CRLB(B), and CRLB(C) on the plane slope (p, q).

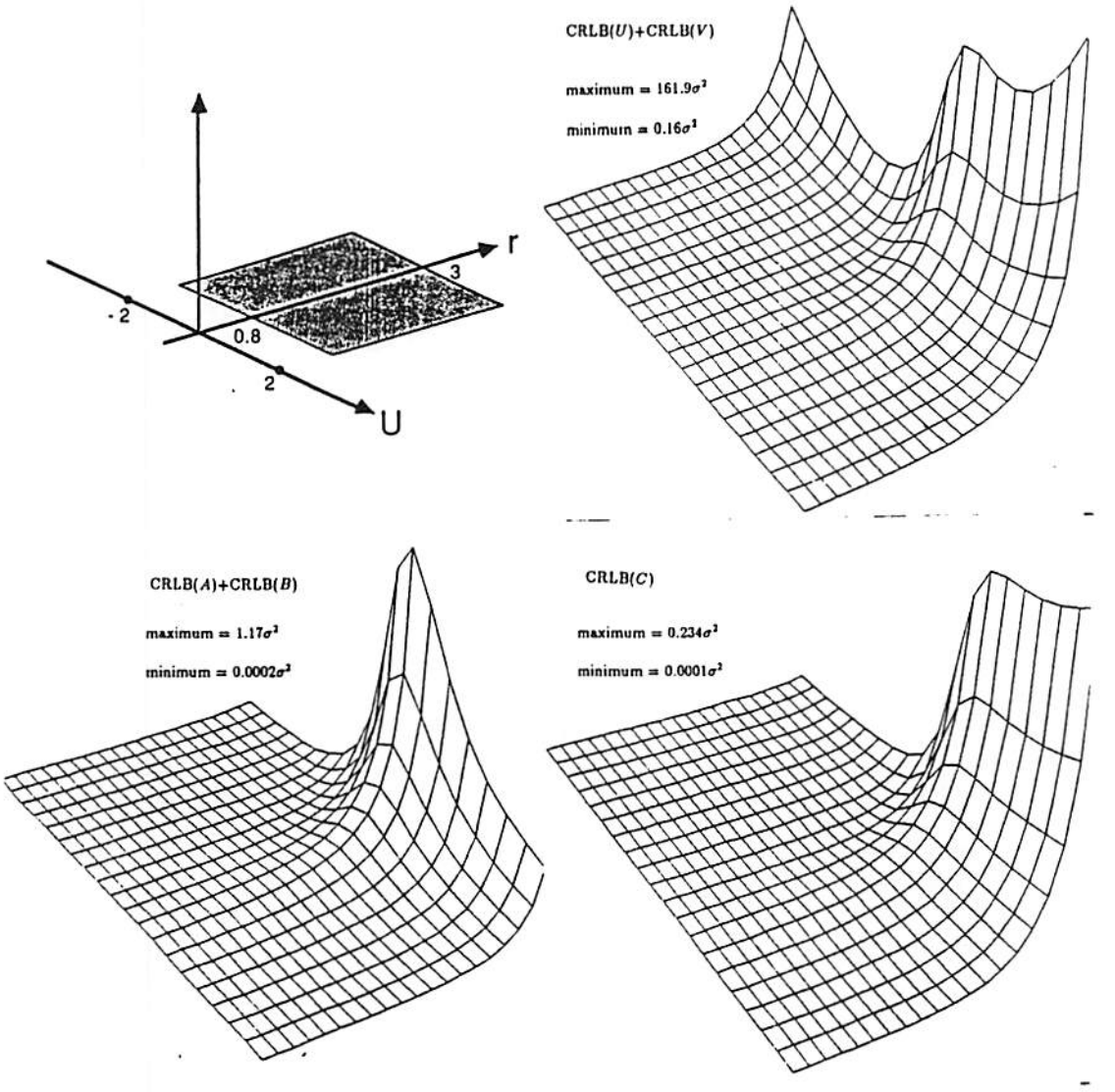


Figure 5.3: The dependence of the three invariants CRLB(U) + CRLB(V), CRLB(A) + CRLB(B), and CRLB(C) on the first translational component U and the radius of the field of view r

the flow field is σ^2. The field of view is a square with width 3. 441 feature points are uniformly spread over the field of view to form a 21×21 grid. The ranges of U and V in Fig. 5.5 and Fig. 5.6 are

$$-4.02 \leq U, V \leq 4.16$$

The maximum and minimum heights in each diagram are marked. The peaks around the zero translational motion $(U, V) = (0, 0)$ show that the motion is relatively more ambiguous for small translational motion. As the size of the translational motion grows from zero, the ambiguity of the translational parameters U and V is lowered in the beginning, but later grows toward infinity. Thus large translational motions cause large ambiguities in the translational parameters. On the other hand, large translational motions reduce the ambiguity of the rotational parameters A, B, and C.

Using (5.7), it can be shown that

- adding measurements never increases the CRLB's.

To see this, Let J and J' be the Fisher information matrices before and after the i^{th} measurement is added. From (5.7), we have

$$J' = J + \frac{1}{\sigma^2} \left(\frac{\partial h_i}{\partial \underline{\theta}} \right)^T \left(\frac{\partial h_i}{\partial \underline{\theta}} \right)$$

Then the matrix inversion lemma [63] gives

$$J^{-1} - J'^{-1} = \left[J^{-1} \left(\frac{\partial h_i}{\partial \underline{\theta}} \right)^T \right] \left\{ \left(\frac{\partial h_i}{\partial \underline{\theta}} \right) J^{-1} \left(\frac{\partial h_i}{\partial \underline{\theta}} \right)^T + \frac{1}{\sigma^2} \right\}^{-1} \left[J^{-1} \left(\frac{\partial h_i}{\partial \underline{\theta}} \right)^T \right]^T$$

Since the real number in the above curly bracket is positive by the positive-definiteness of J^{-1}, $J^{-1} - J'^{-1}$ is positive-semidefinite. Thus each diagonal term of J^{-1} is larger than or equal to the corresponding one of J'^{-1}. As CRLB's are the diagonal terms of the Fisher information matrix, we see that adding more observations will never increase the CRLB's.

In the following, we shall plot the CRLB's when the number of measurements is increased in two different ways. In the first case, the density of the feature

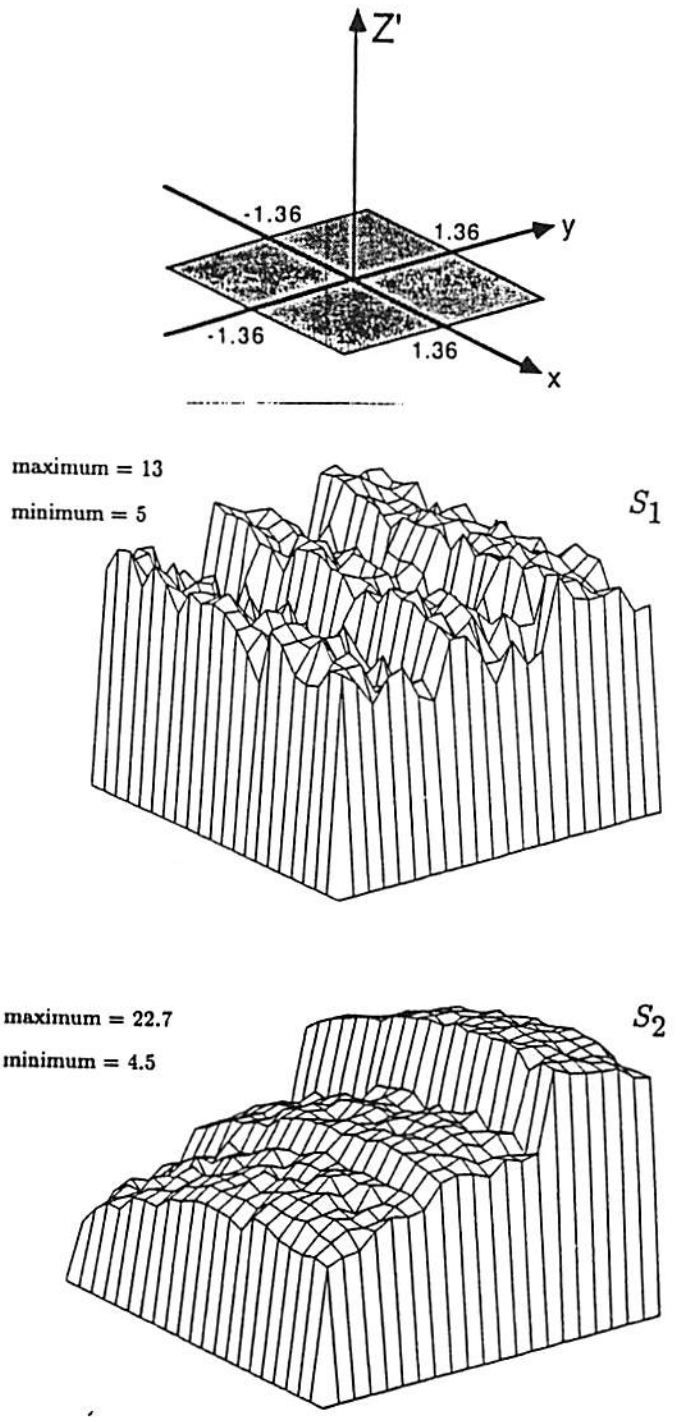


Figure 5.4: Surfaces S_1 and S_2. The structure parameters Z' are plotted as a function of the image coordinates (x, y).

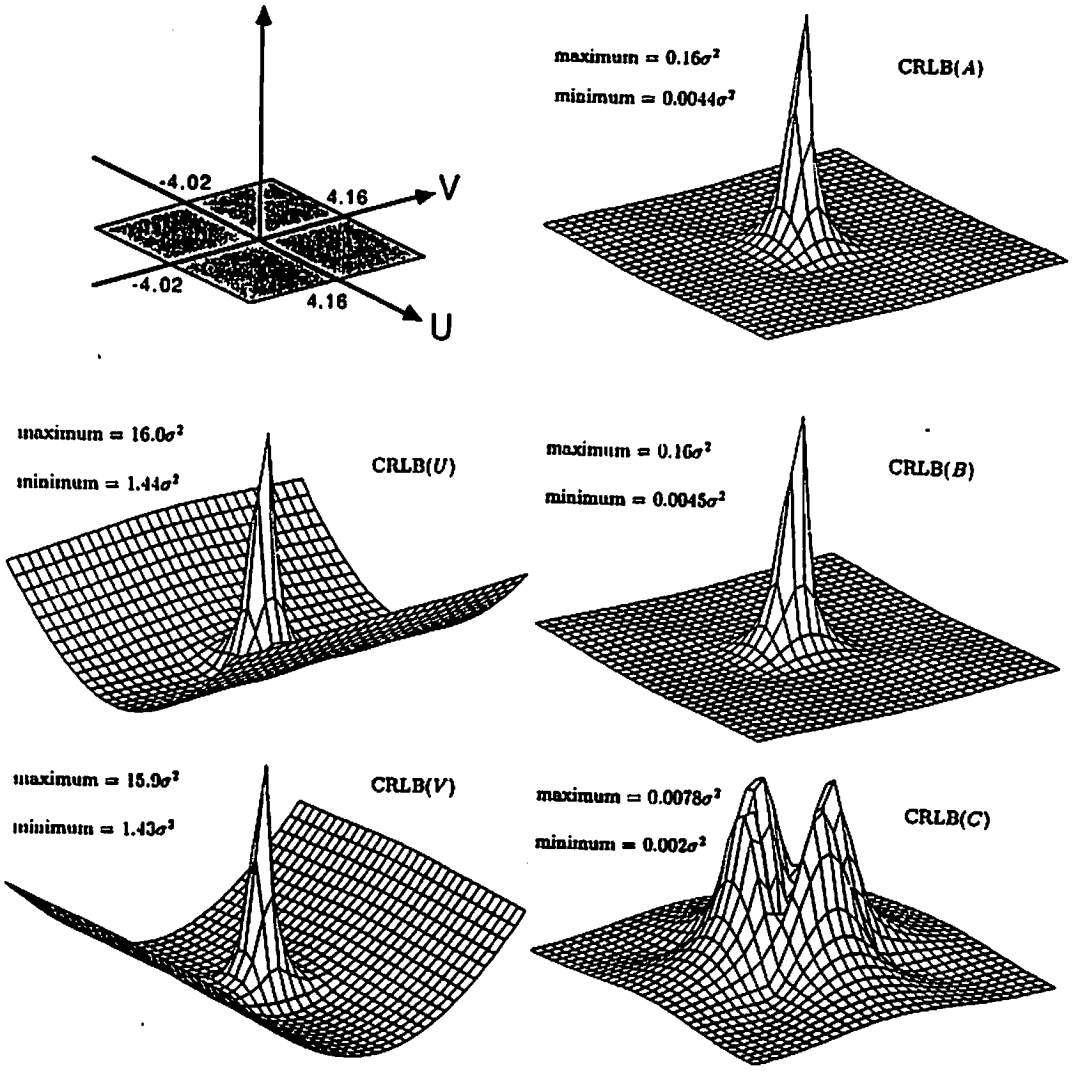


Figure 5.5: The dependence of the CRLB's of U, V, A, B, and C on the translational parameters (U, V) for the motion of the surface S_1.

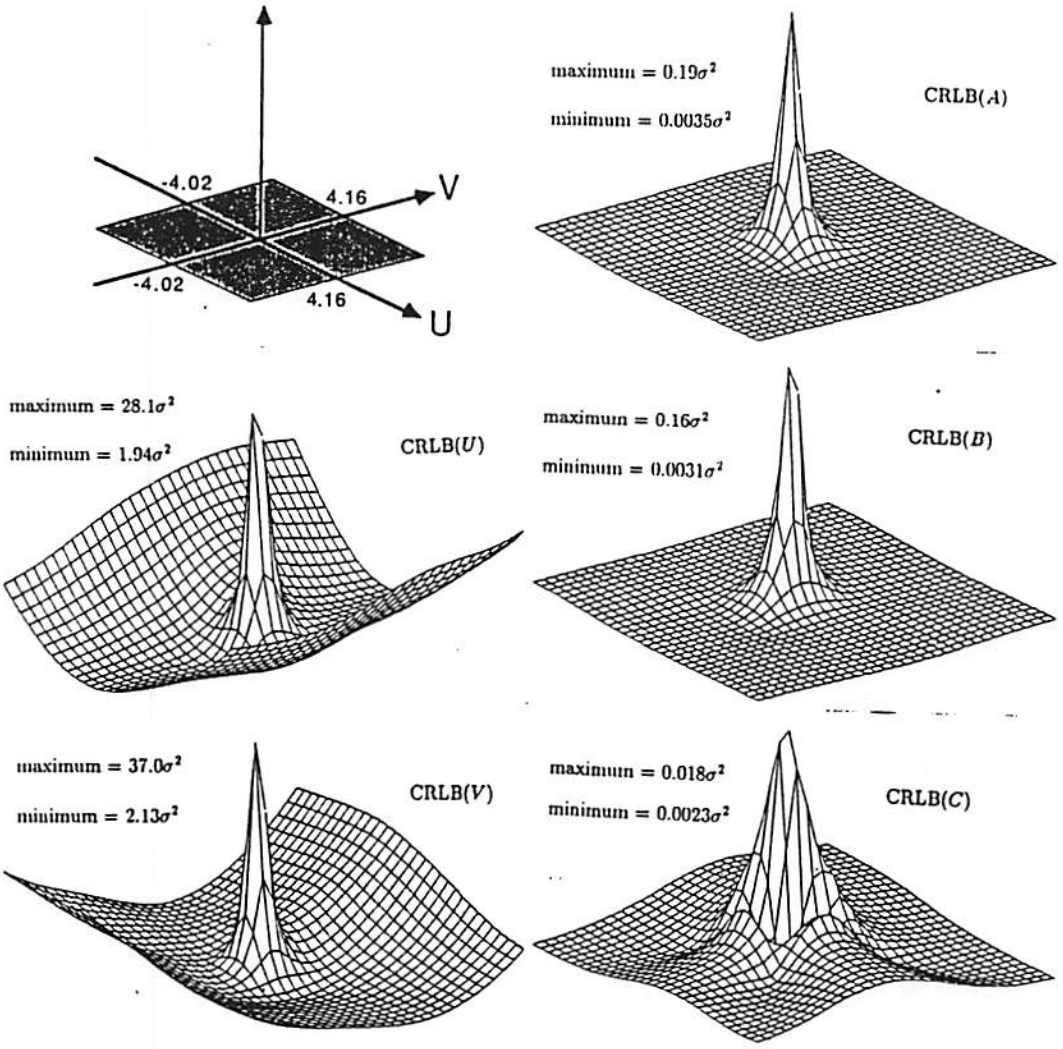


Figure 5.6: The dependence of the CRLB's of U, V, A, B, and C on the translational parameters (U, V) for the motion of the surface S_2.

points is fixed at 441/9 points per unit area, same as that in Fig. 5.5 and Fig. 5.6. The translational motion is $(U, V) = (0.61, 0.75)$. The fields of view are squares. The dependence of CRLB's on the area of the field of view is plotted in Fig. 5.7 and Fig. 5.8 for S_1 and S_2, respectively. We see that the CRLB's monotonically decrease when more feature points are used. In the second case, the field of view is fixed to be a square with width 3. The number of feature points is increased in the way that newly added feature points are placed randomly in the field of view. The dependence of CRLB's on the number of feature points is plotted in Fig. 5.9 and Fig. 5.10 for S_1 and S_2, respectively. Again, monotonically decreasing curves in the number of feature points are observed. Comparing the above two cases, we see that if the same number of feature points is used, the more the spread over the feature points are, the less the ambiguity is. More feature points always alleviate the inherent ambiguity. As the number of feature points grows, the alleviation of the inherent ambiguity becomes slower.

5.5 Motion Estimation by Fitting Local Low Order Patches to Smooth Surfaces

In the previous sections, we have computed the CRLB's of motion parameters when the depth at each pixel is independent of the depths at other pixels. If the depths vary smoothly, we may incorporate smoothness constraints to introduce more information into estimation procedures and thus obtain better accuracy in the estimated motion parameters. One way to add the smoothness constraints on the depths is by dividing the image into small windows and consider a local quadratic approximation to the surface patch in each window [74]. To see how far the CRLB's can be lowered, we shall compute the CRLB's in the optimal condition when the normalized reciprocal depths in a window are fitted by a quadratic patch *exactly*. In other cases when a quadratic patch is only approximately fitted within a window, the residual noise in the surface fitting should cause higher CRLB's than that in the optimal case. In the following, we first show that the CRLB's in the optimal case of quadratic fitting is always lower than the CRLB's when no smoothness constraint on the depth is imposed if the

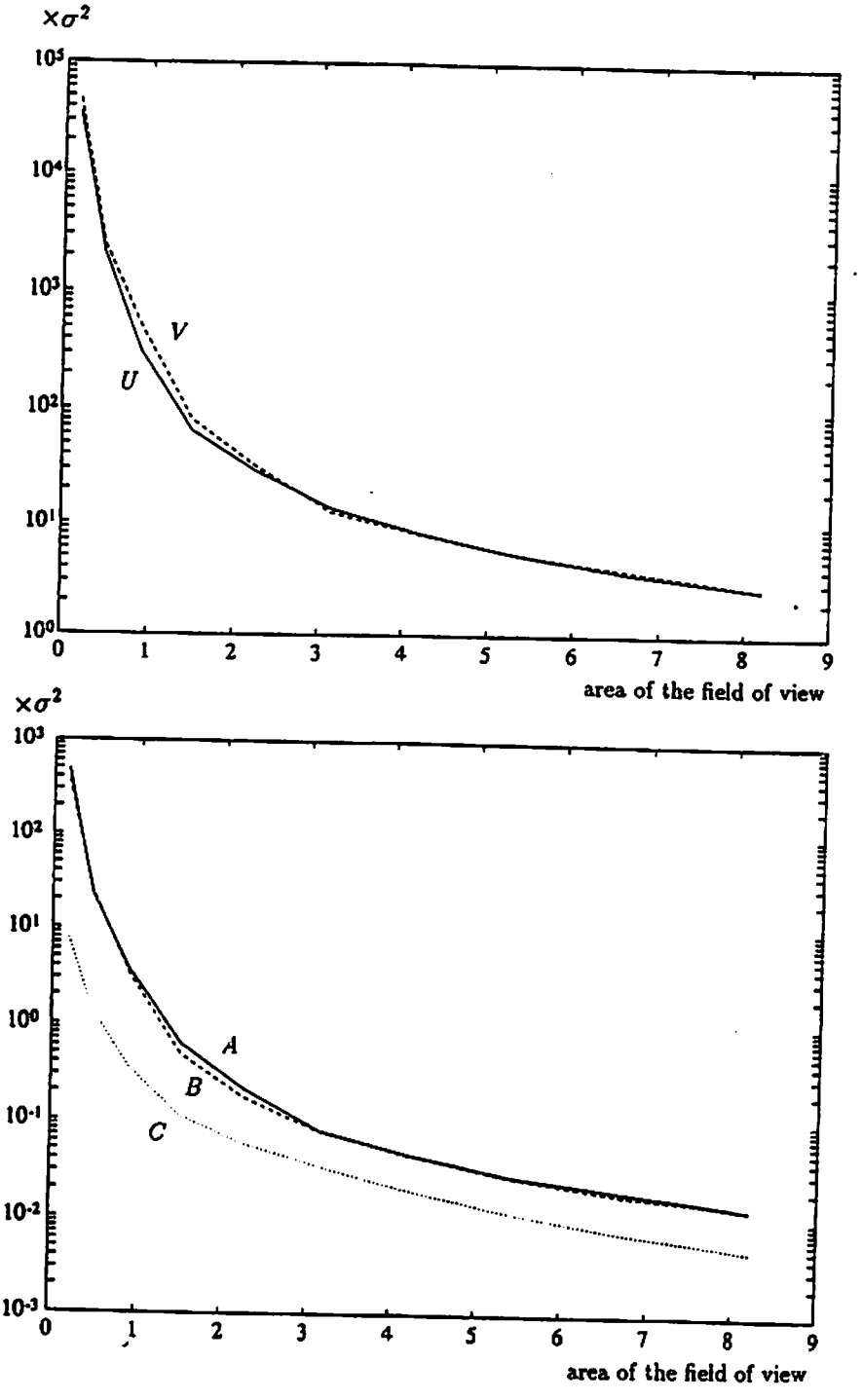


Figure 5.7: The dependence of the CRLB's of U, V, A, B, and C on the area of the field of view for the motion of the surface S_1.

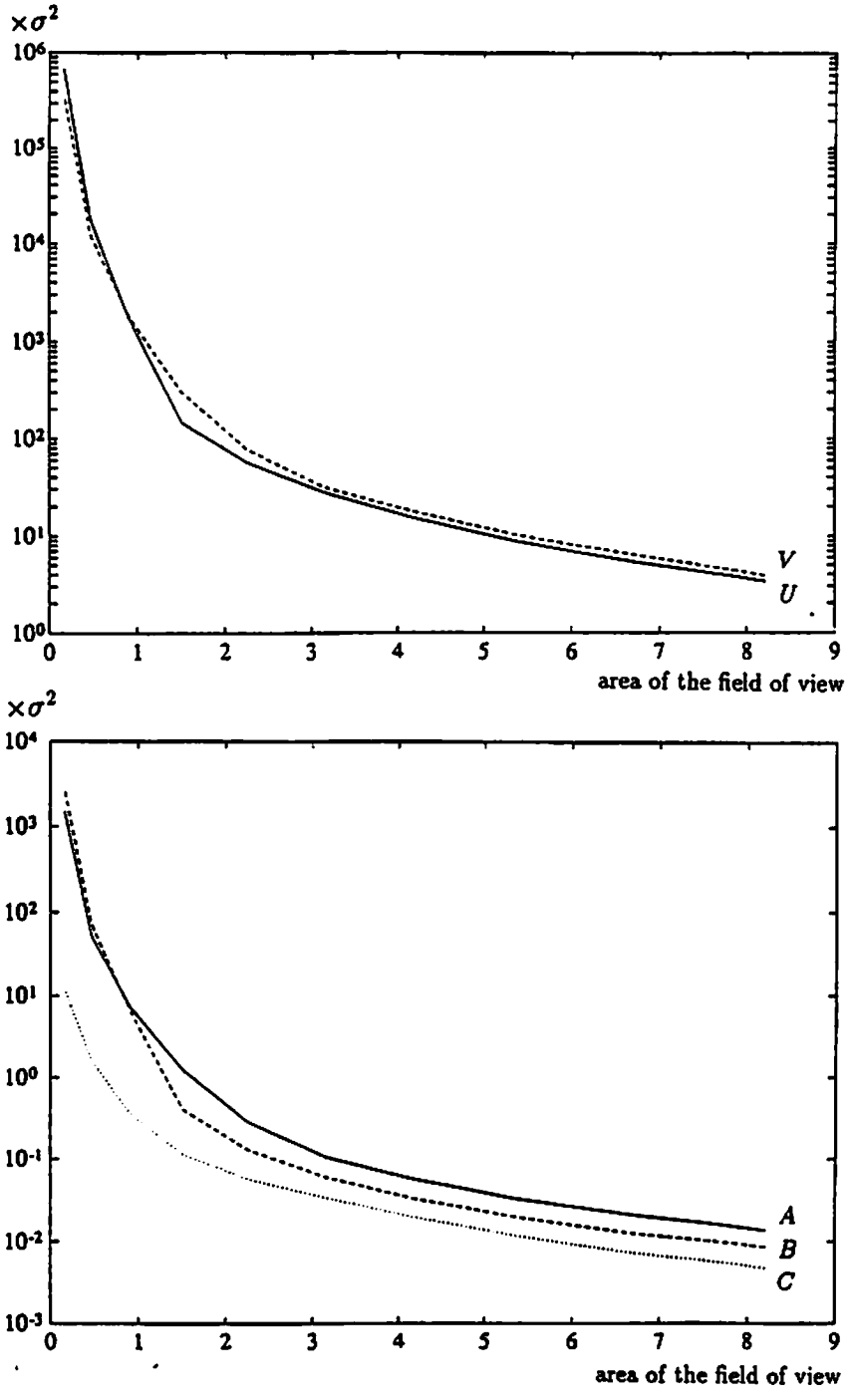


Figure 5.8: The dependence of the CRLB's of U, V, A, B, and C on the area of the field of view for the motion of the surface S_2.

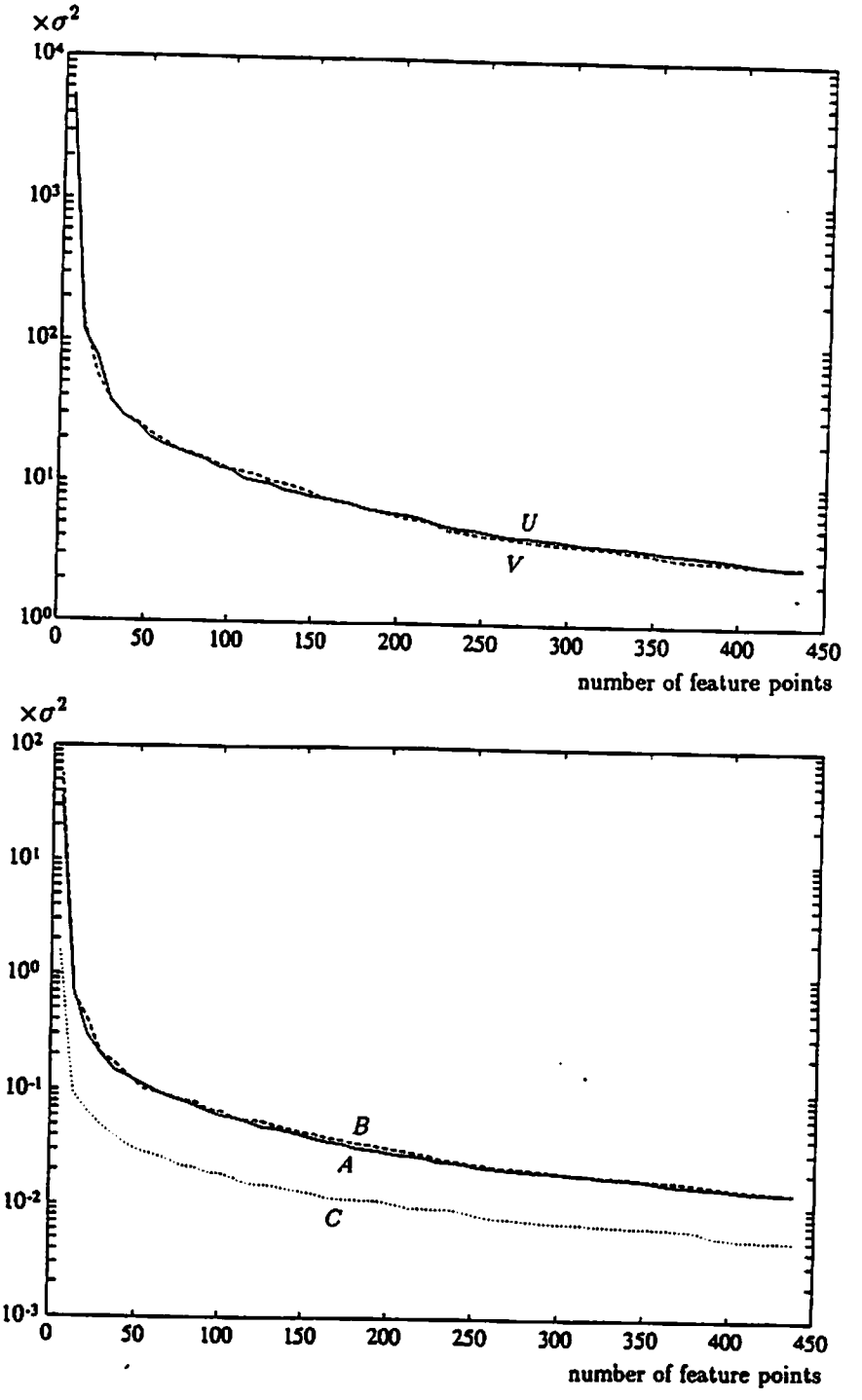


Figure 5.9: The dependence of the CRLB's of U, V, A, B, and C on the number of randomly spread feature points for the motion of the surface S_1.

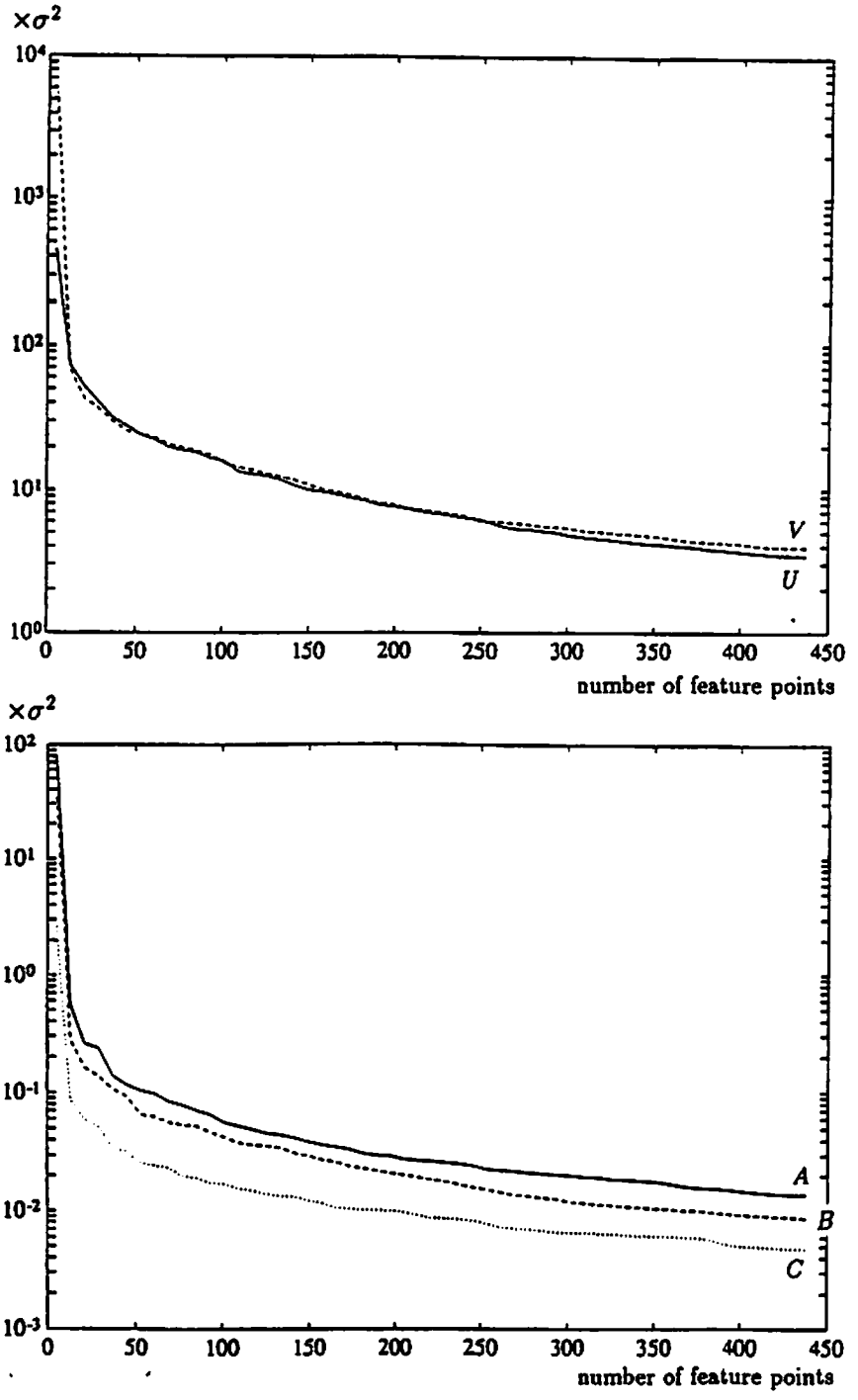


Figure 5.10: The dependence of the CRLB's of U, V, A, B, and C on the number of randomly spread feature points for the motion of the surface S_2.

window size (number of depths within one window) is larger than 6 (number of free parameters representing one quadratic patch). This is expected since reducing the degrees of freedom in the space of unknown parameters stabilizes the parameter estimation process. In fact, the reduction of CRLB's generally holds whenever local patches of any order are fitted to slowly varying surfaces as long as the number of pixels in a small window is larger than the number of free parameters per patch. We shall use quadratic patches as an example to illustrate the computation procedure. Secondly, computer simulations show that even the CRLB's in the optimal case of quadratic fitting are lowered by a very small amount compared to the CRLB's without surface fitting. Finally it will be shown that fitting local patches relaxes the aperture problem in the sense that the only flow component measured at each regular point does lower the CRLB's. This is contrasted with the case without smoothness constraint, in which CRLB's are not affected by incorporating the measurements at regular points.

Suppose that the image is partitioned into N windows, each of which contains m pixels. Let Z_{ij} be the normalized reciprocal depth of the j^{th} point in the i^{th} window. If depths varies smoothly, to obtain an upper performance bound let us consider the case when in the i^{th} window the surface patch can be fitted *exactly* by the i^{th} local quadratic patch [74],

$$Z_{ij} = a_{i0} + a_{i1}x_{ij} + a_{i2}y_{ij} + a_{i3}x_{ij}^2 + a_{i4}x_{ij}y_{ij} + a_{i5}y_{ij}^2, \qquad j = 1, 2, \cdots, m \quad (5.45)$$

where (x_{ij}, y_{ij}) denotes the position of the j^{th} pixel in the i^{th} window in the image plane coordinate system. Denote the collection of free parameters for the local patch in each window by

$$\underline{a}_i = (a_{i0}, a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i5})^T, \qquad i = 1, 2, \cdots, N$$

and the collection of all free parameters for the surface structure,

$$\underline{a} = (\underline{a}_1^T, \underline{a}_2^T, \cdots, \underline{a}_N^T)^T$$

Let us suppose that all points on the image are feature points. Then the measurements are the flow velocity components

$$u_{ij}, v_{ij}, \qquad i = 1, 2, \cdots, N; \quad j = 1, 2, \cdots, m$$

The free parameters are

$$\underline{\theta} = (U, V, A, B, C, \underline{a}^T)^T$$

Thus

$$\begin{aligned} \text{overestimation} &\equiv \frac{\text{number of measurements}}{\text{number of parameters}} \\ &= \frac{2mN}{5+6N} \\ &\to m/3 \qquad \text{as} \quad N \to \infty \\ &> 2 \qquad \text{if} \quad m > 6 \end{aligned}$$

Comparing the above with (5.12), we have better overestimation here than that without putting constraints on depths, if the window size m is greater than 6.

If the Fisher information matrix is partitioned as in (5.14), we have

$$\begin{aligned} \alpha &= \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{m}\left[\left(\frac{\partial h_{uij}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{uij}}{\partial \underline{\psi}}\right) + \left(\frac{\partial h_{vij}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{vij}}{\partial \underline{\psi}}\right)\right] \\ \beta &= \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{m}\left[\left(\frac{\partial h_{uij}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{uij}}{\partial \underline{a}}\right) + \left(\frac{\partial h_{vij}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{vij}}{\partial \underline{a}}\right)\right] \\ \gamma &= \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{m}\left[\left(\frac{\partial h_{uij}}{\partial \underline{a}}\right)^T\left(\frac{\partial h_{uij}}{\partial \underline{a}}\right) + \left(\frac{\partial h_{vij}}{\partial \underline{a}}\right)^T\left(\frac{\partial h_{vij}}{\partial \underline{a}}\right)\right] \end{aligned} \tag{5.46}$$

As a result of the above equation, γ is a block diagonal matrix (of dimension $6N \times 6N$) with the i^{th} block matrix (of dimension 6×6)

$$\gamma_i = \frac{1}{\sigma^2}\sum_{j=1}^{m}\left[\left(\frac{\partial h_{uij}}{\partial \underline{a}_i}\right)^T\left(\frac{\partial h_{uij}}{\partial \underline{a}_i}\right) + \left(\frac{\partial h_{vij}}{\partial \underline{a}_i}\right)^T\left(\frac{\partial h_{vij}}{\partial \underline{a}_i}\right)\right]$$

Moreover, β (of dimension $5 \times 6N$) has the following structure

$$\beta = (\beta_1, \beta_2, \cdots, \beta_N)$$

where the submatrix β_i (of dimension 5×6) is

$$\beta_i = \frac{1}{\sigma^2} \sum_{j=1}^{m} \left[\left(\frac{\partial h_{uij}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{uij}}{\partial \underline{a}_i} \right) + \left(\frac{\partial h_{vij}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{vij}}{\partial \underline{a}_i} \right) \right]$$

Therefore in (5.17)

$$\begin{aligned} Q &= (\alpha - \beta \gamma^{-1} \beta^T)^{-1} \\ &= (\alpha - \sum_{i=1}^{N} \beta_i \gamma_i^{-1} \beta_i^T)^{-1} \end{aligned} \tag{5.47}$$

The above equation shows that Q can be computed by inverting 5×5 and 6×6 matrices. The diagonal terms of Q then give the CRLB's of the motion parameters.

In the following, we shall compare the CRLB's of the motion parameters after and before fitting local patches. In both cases enough pixels are assumed available so that the corresponding Fisher information matrices are not singular. Let us superscribe $'$ upon the quantities involved in the calculation of CRLB's when no constraints are put on the depths. Without constraints, the Fisher information matrix can be rewritten from (5.13) as

$$J' = \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{m} \left[\left(\frac{\partial h_{uij}}{\partial \underline{\theta}'} \right)^T \left(\frac{\partial h_{uij}}{\partial \underline{\theta}'} \right) + \left(\frac{\partial h_{vij}}{\partial \underline{\theta}'} \right)^T \left(\frac{\partial h_{vij}}{\partial \underline{\theta}'} \right) \right]$$

where

$$\begin{aligned} \underline{\theta}' &= (U, V, A, B, C; Z_{11}, Z_{12}, \cdots, Z_{1m}; \cdots; Z_{N1}, Z_{N2}, \cdots, Z_{Nm})^T \\ &= (\underline{\psi}^T, \underline{Z}^T)^T \end{aligned}$$

After partitioning J' as in (5.14), we have

$$\begin{aligned}
\alpha' &= \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{m}\left[\left(\frac{\partial h_{uij}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{uij}}{\partial \underline{\psi}}\right)+\left(\frac{\partial h_{vij}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{vij}}{\partial \underline{\psi}}\right)\right] \\
\beta' &= \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{m}\left[\left(\frac{\partial h_{uij}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{uij}}{\partial \underline{Z}}\right)+\left(\frac{\partial h_{vij}}{\partial \underline{\psi}}\right)^T\left(\frac{\partial h_{vij}}{\partial \underline{Z}}\right)\right] \\
\gamma' &= \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{m}\left[\left(\frac{\partial h_{uij}}{\partial \underline{Z}}\right)^T\left(\frac{\partial h_{uij}}{\partial \underline{Z}}\right)+\left(\frac{\partial h_{vij}}{\partial \underline{Z}}\right)^T\left(\frac{\partial h_{vij}}{\partial \underline{Z}}\right)\right]
\end{aligned} \tag{5.48}$$

With the aid of the chain rule

$$\frac{\partial h_{uij}}{\partial \underline{a}} = \frac{\partial h_{uij}}{\partial \underline{Z}}\frac{\partial \underline{Z}}{\partial \underline{a}}$$

and a comparison between (5.46) and (5.48), we arrive at

$$\begin{aligned}
\alpha &= \alpha' \\
\beta &= \beta'\frac{\partial \underline{Z}}{\partial \underline{a}} \equiv \beta' Z_a \\
\gamma &= (\frac{\partial \underline{Z}}{\partial \underline{a}})^T\gamma(\frac{\partial \underline{Z}}{\partial \underline{a}}) \equiv Z_a^T\gamma' Z_a
\end{aligned} \tag{5.49}$$

in which we have used the abbreviation

$$Z_a \equiv \frac{\partial \underline{Z}}{\partial \underline{a}} \tag{5.50}$$

The rank of Z_a depends on the window shape. Sufficient conditions for Z_a to be of full rank are given in Appendix B. For regularly shaped windows, e. g. all square windows larger than 2×2, we have a full-ranked Z_a. According to (5.17) and (5.18), the CRLB's of motion parameters are the diagonal terms of

$$Q = \alpha^{-1} + \alpha^{-1}\beta G\beta^T\alpha^{-1}$$

if quadratic patches are fitted. Without constraints, the CRLB's are the diagonal terms of

$$Q' = \alpha'^{-1} + \alpha'^{-1}\beta' G'\beta'^T\alpha'^{-1}$$

By (5.49), Q can be written as

$$Q = \alpha'^{-1} + \alpha'^{-1}\beta' Z_a G Z_a^T \beta'^T \alpha'^{-1}$$

To compare these two sets of CRLB's, let us consider the matrix difference

$$Q' - Q = \alpha'^{-1}\beta'[G' - Z_a G Z_a^T]\beta'^T \alpha'^{-1} \tag{5.51}$$

G will be related to G' using (5.18) and (5.49),

$$\begin{aligned} G &= (\gamma - \beta^T \alpha^{-1} \beta)^{-1} \\ &= (Z_a^T \gamma' Z_a - Z_a^T \beta'^T \alpha'^{-1} \beta' Z_a)^{-1} \\ &= [Z_a^T (\gamma' - \beta'^T \alpha'^{-1} \beta') Z_a]^{-1} \\ &= (Z_a^T G'^{-1} Z_a)^{-1} \end{aligned}$$

Since J'^{-1} is assumed to be positive definite, so are G' and G'^{-1} by (5.17). Thus G'^{-1} can be decomposed into [65]

$$G'^{-1} = KK^T \tag{5.52}$$

in which K is nonsingular. Then the second quantity in the bracket of (5.51) can be rewritten as

$$\begin{aligned} Z_a G Z_a^T &= Z_a (Z_a^T G'^{-1} Z_a)^{-1} Z_a^T \\ &= K^{T-1} K^T Z_a (Z_a^T K K^T Z_a)^{-1} Z_a^T K K^{-1} \\ &\equiv K^{T-1} P K^{-1} \end{aligned} \tag{5.53}$$

where P is defined by

$$P \equiv K^T Z_a (Z_a^T K K^T Z_a)^{-1} Z_a^T K \tag{5.54}$$

Note that the dimensions of P and Z_a are $mN \times mN$ and $mN \times 6N$, respectively. For regular windows, e.g. all square windows larger than 2×2, the rank of Z_a is $6N$ since Z_a is of full rank. Thus (5.54) implies that the rank of P is at most

$6N$. Because $m > 6$, we see that P is singular. It is straightforward to show that P is idempotent and symmetric, i.e.

$$\begin{aligned} P^2 &= P \\ P &= P^T \end{aligned}$$

Hence P as well as $(I-P)$ are projection matrices, where I is the identity matrix. Every projection matrix has eigenvalues either zero or one, therefore $(I-P)$ is positive semi-definite and can be decomposed into

$$I - P = LL^T \tag{5.55}$$

Using (5.52), (5.53) and (5.55), we can rewrite (5.51) as

$$\begin{aligned} Q' - Q &= \alpha'^{-1}\beta' K^{T^{-1}}(I-P)K^{-1}\beta'^T\alpha'^{-1} \\ &= (\alpha'^{-1}\beta' K^{T^{-1}}L)(\alpha'^{-1}\beta' K^{T^{-1}}L)^T \end{aligned}$$

Therefore $Q' - Q$ is positive semi-definite. This implies that $Q' - Q$ has its diagonal terms no less than zero. Since the diagonal terms of Q and Q' are the CRLB's with and without fitting local patches, respectively, thus we conclude that fitting local patches to smooth surfaces never increase the CRLB's.

To briefly look at how much the CRLB's are lowered by fitting local patches, we compute the CRLB's before and after fitting quadratic surfaces for different window sizes. The CRLB's without fitting local patches to depths are computed by (5.19). The CRLB's after fitting local quadratic patches are computed by (5.47). The surfaces used correspond to the optimal case of quadratic fitting, i.e. within each window the normalized reciprocal depths are fitted exactly by a quadratic patch. The differences of the CRLB's before and after fitting quadratic patches are divided by the CRLB's before fitting. The results are shown Table 5.1. (The field of view is a square of width 3. The translational motion is $(U', V', W') = (0.64, 0.79, 1)$. The feature points are located on a uniform 60×60 grid. The focal length has been set to 1.) One may see that by fitting

	window size			
	3×3	4×4	5×5	6×6
U	0.034%	0.073%	0.11%	0.14%
V	0.029%	0.075%	0.10%	0.14%
A	0.069%	0.17%	0.25%	0.42%
B	0.072%	0.16%	0.29%	0.36%
C	0.13%	0.21%	0.34%	0.51%

Table 5.1: Percentage of CRLB's lowered by fitting local quadratic patches.

local patches, the CRLB's are lowered by only a very limited amount (less than 0.6%).

We shall show in the following that fitting local patches relaxes the aperture problem. If regular points are available and local patches are fitted, one can follow derivations similar to in Sections 5.2 and 5.5 to obtain the CRLB's from the diagonal terms of

$$Q'' = (\alpha'' - \sum_{i=1}^{N} \beta_i'' \gamma_i''^{-1} \beta_i''^T)^{-1} \tag{5.56}$$

where

$$\alpha'' = \frac{1}{\sigma^2} \sum_{i=1}^{N} \left\{ \sum_{k=1}^{M_{ir}} \left(\frac{\partial h_{ik}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{ik}}{\partial \underline{\psi}} \right) + \sum_{j=1}^{M_{if}} \left[\left(\frac{\partial h_{uij}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{uij}}{\partial \underline{\psi}} \right) + \left(\frac{\partial h_{vij}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{vij}}{\partial \underline{\psi}} \right) \right] \right\}$$

$$\beta_i'' = \frac{1}{\sigma^2} \left\{ \sum_{j=1}^{M_{ir}} \left(\frac{\partial h_{ik}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{ik}}{\partial \underline{a}_i} \right) + \sum_{j=1}^{M_{if}} \left[\left(\frac{\partial h_{uij}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{uij}}{\partial \underline{a}_i} \right) + \left(\frac{\partial h_{vij}}{\partial \underline{\psi}} \right)^T \left(\frac{\partial h_{vij}}{\partial \underline{a}_i} \right) \right] \right\}$$

$$\gamma_i'' = \frac{1}{\sigma^2} \left\{ \sum_{j=1}^{M_{ir}} \left(\frac{\partial h_{ik}}{\partial \underline{a}_i} \right)^T \left(\frac{\partial h_{ik}}{\partial \underline{a}_i} \right) + \sum_{j=1}^{M_{if}} \left[\left(\frac{\partial h_{uij}}{\partial \underline{a}_i} \right)^T \left(\frac{\partial h_{uij}}{\partial \underline{a}_i} \right) + \left(\frac{\partial h_{vij}}{\partial \underline{a}_i} \right)^T \left(\frac{\partial h_{vij}}{\partial \underline{a}_i} \right) \right] \right\}$$

in which M_{if} and M_{ir} are the numbers of feature and regular points, respectively, in the i^{th} window; regular points have been subscribed by k, feature points by j;

and

$$h_{ik} \equiv h_{uik} \cos\theta_{ik} + h_{vik} \sin\theta_{ik}$$

We have also assumed that the image brightness pattern has the gradient vector along

$$(\cos\theta_{ik}, \sin\theta_{ik})$$

at the k^{th} regular point of the i^{th} window. We find that the regular points do contribute information to CRLB's because the terms involving the regular points do not cancel out in (5.56). This fact may also be checked by a computer simulation. The surface considered is shown in Fig. 5.11. Assume that the center point of each mutually disjoint 3×3 window is a regular point at which only the flow component along the gradient of the image brightness pattern is observable. The other 8 points in each window are feature points. The CRLB's based on feature points only are computed by (5.47) and listed in the first column of Table 5.2. The CRLB's based on both feature and regular points are obtained using (5.56) and listed in the second column. Within each window, the surface is fitted exactly by a quadratic surface. The translational motion is $(U', V', W') = (0.61, 0.75, 1)$. The field of view is a square of width 3. The pixels are located on a uniform 21×21 grid. The focal length is 1. One may see from Table 5.2 that by adding the flow measurements at regular points, the CRLB's are lowered.

We summarize the analyses in Section 5.2 and 5.5 by providing the following ordering of CRLB's.

CRLB's based on the smoothed flow field extracted by regularization methods

$\geq$ CRLB's based only on feature points when *no* local patches are fitted

$=$ CRLB's based on both feature and regular points when *no* local patches are fitted

$\geq$ CRLB's based only on feature points when local patches are fitted

$\geq$ CRLB's based on both feature and regular points when local patches are fitted

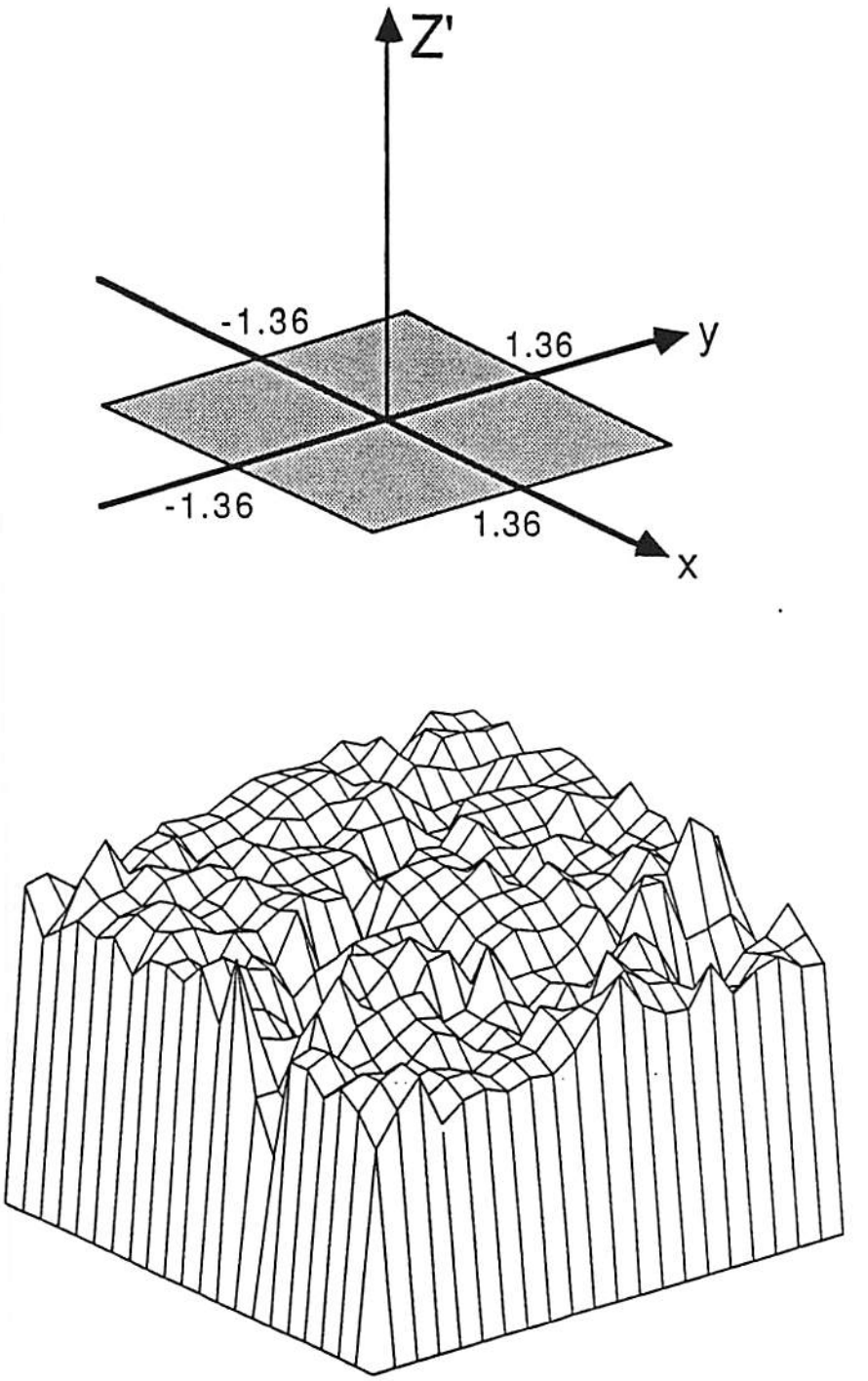


Figure 5.11: The surface considered in the simulations in Table 5.2. The structure parameters Z' are plotted as a function of the image coordinates (x, y).

	feature points only	feature and regular points
CRLB(U)	2.52	2.43
CRLB(V)	2.66	2.57
CRLB(A)	0.0111	0.0106
CRLB(B)	0.0127	0.0121
CRLB(C)	0.0050	0.0049

Table 5.2: The CRLB's without and with applying the measurements at regular points for the motion of the surface in Fig. 5.11, when local quadratic patches are fitted. Each CRLB is normalized by σ^2, the noise variance of the optical flow components.

5.6 Discussion

A simple statistical model has been proposed to describe the inherent ambiguities in recovering 3-D motion from a noisy flow field. The simplest case when the white Gaussian noise contaminates the flow field has been studied. Although the Gaussian assumption is restrictive, it is interesting to note that the results of some special cases in our simple statistical model support Adiv's analysis, in which no noise distribution is assumed. It has been shown that the proposed model explains many interesting features regarding the inherent ambiguities and is mathematically manageable. Using this model, for example, the connection between the aperture problem and the stability of motion estimation can be addressed; analytic results for the general motion of a planar surface can be obtained; and improvements of the estimation process by fitting local patches can be studied; even the general motion of a nonplanar surface can be considered using simulations. It should also be pointed out that in our work the derived performance bounds are valid only for the *unbiased* motion estimation algorithms since we do not have expressions for the bias. The results presented in this chapter should be interpreted cautiously when motion estimation algorithms which give significantly *biased* motion estimators are considered.

Using the quantitative form (5.3) for the Fisher information matrix, extension of the model to more complicated noise distributions is possible. For instance, if noise distribution other than Gaussian is considered, the Fisher information

matrix in the Cramer-Rao inequality (5.2) can be computed using (5.3). As shown in Section 5.4, although algebraic manipulation for the general motion of a nonplanar surface is not possible, numerical results can always be computed and some general properties can be inferred from the illustrations. This advantage carries over to the non Gaussian cases, as for any noise distribution, CRLB's can always be computed directly from (5.2) and (5.3). Then the dependence of CRLB's on different factors can be plotted pointwise. Observations may then be made from these plottings, similar to what has been done in Section 5.4.

Chapter 6

Conclusions And Future Research

6.1 Conclusions

Two motion models for estimation of 3–D motion and structure from a sequence of noisy images have been studied . Inherent ambiguities in recovering 3–D motion from image sequences have also been investigated.

If stereo image sequences are available, we have shown that the structure can be estimated prior to motion estimation. In this case, quaternions have been shown to be convenient to represent instantaneous orientation of the rigid body. As the structure is estimated prior to motion estimation, the simplified motion estimation stage allows the accommodation of a high order motion model including constant acceleration and precession. Nonlinear recursive filters such as EKF and IEKF have been successfully applied to implement motion tracking and estimation. It has also been shown that three noncollinear feature points at five measurement times are enough to obtain all motion information in the noise free image observations. A deterministic algorithm to recover motion parameters from a stereo image sequence has been summarized from the analysis of inherent ambiguities. The CRLB has been shown to be a useful tool to predict the performance of the EKF/IEKF prior to implementing the recursive filters.

If 3–D motion and structure are estimated simultaneously, we have shown that the standard 3×3 rotation matrix is more suitable to represent the rotational motion than quaternions. In this case, the formulations of constant velocity translation and constant angular velocity rotation model have been developed for both monocular and binocular motion estimation. The conjugate gradient method has been successfully applied to obtain a batch estimate of 3–D motion parameters from zero initial guesses. Once a good batch estimate is obtained, an EKF/IEKF may be initialized to track the object. Occlusion has been considered in the model formulation as well as in computer simulations. In the recursive approach, closed form expressions for propagating the estimated states and the covariance matrix have been used.

The performance bound quantified by using CRLB is a general tool in analyzing the inherent ambiguities of various 3–D motion estimators in a noisy environment. By using this bound, we have shown that the aperture problem with the noisy image measurement restricts the nontrivial information about 3-D motion to a sparse set of pixels at which both components of the flow velocity are observable. Various factors leading to inherent ambiguities are then derived easily. The smoothness constraint imposed in the typical computation of the flow field usually degrades the following motion estimation stage. Apparently this is due to blurring of the useful measurements from the feature points caused by smoothing at neighboring points. We also show that imposing smoothness constraint on 3–D depths improves the motion estimation stage by a very small amount. The effects of these two smoothing schemes differ from each other. Smoothing the optical flow field is similar to post processing of direct measurement data. On the other hand, smoothing the 3–D depths is similar to imposing constraints on the deterministic parameters in the statistical model.

6.2 Topics for Future Research

Several topics requiring further investigations have become apparent as a consequence of the research reported in this dissertation. In general, the field of 3–D

motion and structure estimation from video images remains open in that algorithms that can be implemented in real time are not available yet. We suggest the following research directions to extend the contributions of this dissertation.

- Testing of the proposed algorithms on real dynamic image sequences needs to be done. To obtain image sequences of an object moving along a trajectory involves precise calibration and measurements. Object positioning and camera calibration are not trivial tasks. The robustness of algorithms pioneered in this dissertation may not be appreciated until experiments on real images are conducted.

- The development of kinematic model based motion estimation algorithms using a sequence of optical flow field of normal image motion components. It has been shown that only flow components at salient points are effective in coping with noise when one single flow field is used. In order to exploit the motion at all image points, a kinematic model may be useful to establish the relationship of motions among image frames. As local computations are suitable for parallel processing, we suggest that algorithms using local computations be developed. However, local computations yield only normal components of the motion flow fields. Therefore it seems reasonable to estimate 3–D motion based on fields of normal optical flow components. Because existing algorithms for discrete feature extraction and correspondence needed in the feature based approaches involve complicated computations, the use of a sequence of optical flow fields may lead to practically interesting algorithms for real time motion estimation.

- 3–D motion estimation is in effect a feedback block embedded in a complete control loop. Usually the forward control uses the estimated motion information to influence the 3–D world. The example of passive navigation may make this point clear, in which 3–D motion information is fed back to control the vehicle. Because the control is application dependent, viewing 3–D motion estimation as a feedback subsystem may exploit special features associated with this application to improve the performance of the whole system.

- The development of multi-sensor systems which incorporate measurements from video cameras. Since the motion models are not dependent on specific sensor properties, other sensors may be easily incorporated into motion estimation process. In this case, the CRLB may be used to determine the relative information weights associated with different types of measurements. Then the system may be improved by optimizing the tradeoff between computational complexity and the information content.

Appendix A

Proposition :

If

$$(p,q)^T \neq (-U,0)^T$$

then

$$\begin{aligned}&(p+U)^2\left[(3U^4+2U^2+3)p^2-2U(3U^2+1)p+3U^2\right] \\ &+q^2\left[(2U^4+8U^2+6)p^2+3(U^2+1)^2q^2-(6U^5+2U^3-4U)p+6U^4+6U^2\right]\end{aligned} \tag{A.1}$$

is always positive.

Proof :

The quantity (A.1) may be put into the form of

$$ax^2+bx+c, \qquad x>0, \quad a>0, \quad c\geq 0 \tag{A.2}$$

where

$$\begin{aligned} x &= q^2 \\ a &= 3(U^2+1)^2 \\ b &= (U^2+1)\left[-3U^3p+U^2p^2+3U^2+2Up+3p^2\right] \qquad \text{(A.3)} \\ c &= (p+U)^2\left[(3U^4+2U^2+3)p^2+(-6U^3-2U)p+3U^2\right] \qquad \text{(A.4)}\end{aligned}$$

$c \geq 0$ because the quantity inside the brackets in (A.4) is a quadratic of p and has the discriminant

$$\Delta_p = (-6U^3 - 2U)^2 - 4(3U^4 + 2U^2 + 3)(3U^2) = -32U^2 \leq 0$$

According to the quadratic form (A.2), the proof of the proposition is done if we can show that

$$\Delta_x = b^2 - 4ac \geq 0 \qquad \Longrightarrow \qquad b \geq 0$$

Therefore let us assume henceforth

$$\Delta_x \geq 0 \tag{A.5}$$

The case

$$U = 0 \quad \text{or} \quad p = 0$$

may be excluded from our discussion since in which case (A.1) is obviously positive. Because

$$\Delta_x = b^2 - 4ac = 32U^2p^2(U^2 + 1)^2(1 - Up)(3U^2 + Up + 2)$$

we have, by (A.5),

$$(1 - Up)(3U^2 + Up + 2) \geq 0 \tag{A.6}$$

Case I. $Up < 0$

(A.3) gives

$$b = (U^2 + 1)\left[3(-Up)U^2 + U^2p^2 + (U + p)^2 + 2(U^2 + p^2)\right] \geq 0$$

Case II. $0 \leq Up \leq 1$

(A.3) yields

$$b = (U^2 + 1)\left[3(1 - Up)U^2 + U^2p^2 + 2Up + 3p^2\right] \geq 0$$

Case III. $1 < Up$

(A.6) shows

$$-(3U^2 + Up + 2) \geq 0$$

Then (A.3) becomes

$$b = (U^2 + 1)\left[Up[-(3U^2 + Up + 2)] + 2U^2p^2 + 2(U^2 + p^2) + U^2 + p^2\right] \geq 0$$

Appendix B

In this appendix, we give sufficient conditions for the matrix Z_a in (5.50) to be of full rank. It will be shown that if either configuration in Fig. B.1 is a subset of each window, Z_a has full rank.

Let the collection of the normalized reciprocal depths in the i^{th} window be denoted by

$$\underline{Z}_i \equiv (Z_{i1}, Z_{i2}, \cdots, Z_{im})^T$$

Using the abbreviation

$$Z_{ai} \equiv \frac{\partial \underline{Z}_i}{\partial \underline{a}_i} = \begin{pmatrix} 1 & x_{i1} & y_{i1} & x_{i1}^2 & x_{i1}y_{i1} & y_{i1}^2 \\ 1 & x_{i2} & y_{i2} & x_{i2}^2 & x_{i2}y_{i2} & y_{i2}^2 \\ \vdots & & & & & \vdots \\ 1 & x_{im} & y_{im} & x_{im}^2 & x_{im}y_{im} & y_{im}^2 \end{pmatrix}$$

we have

$$Z_a = \begin{pmatrix} Z_{a1} & & & \\ & Z_{a2} & & \\ & & \vdots & \\ & & & Z_{aN} \end{pmatrix}$$

Z_a is of full rank if and only if Z_{ai}, all i, are of full rank. Because Z_{ai} is a $m \times 6$ matrix with $m > 6$, a sufficient condition for Z_{ai} to be of full rank is that the