Necessary Conditions to Attain Performance Bounds on Structure and Motion Estimates of Rigid Objects *

Margrit Betke
Image and Video Computing Group
Computer Science Department
Boston University

Eran Naftali and Nicholas C. Makris
Acoustics Group
Department of Ocean Engineering
Massachusetts Institute of Technology

Abstract

Analytic conditions that are necessary for the maximum likelihood estimate to become asymptotically unbiased and attain minimum variance are derived for estimation problems in computer vision. In particular, problems of estimating the parameters that describe the 3D structure of rigid objects or their motion are investigated. It is common practice to compute Cramer-Rao lower bounds (CRLB) to approximate the mean-square error in parameter estimation problems, but the CRLB is not guaranteed to be a tight bound and typically underestimates the true mean-square error. The necessary conditions for the Cramer-Rao lower bound to be a good approximation of the mean-square error are derived. The tightness of the bound depends on the noise level, the number of pixels on the surface of the object, and the texture of the surface. We examine our analytical results experimentally using polyhedral objects that consist of planar surface patches with various textures that move in 3D space. We provide necessary conditions for the CRLB to be attained that depend on the size, texture, and noise level of the surface patch.

1 Introduction

In many practical problems in computer vision, nonlinear inversions are required to estimate parameters from measured data [15, 18, 11]. These parameters, for example, may be used to model the three-dimensional (3D) structure of an object, its shape, its surface reflectance properties, or its motion in space. The nonlinear inversion of random data often leads to estimates that are biased and do not attain the minimum variance possible, namely the Cramer-Rao lower bound (CRLB). The maximum likelihood estimator (MLE) [1] is widely used because if an asymptotically unbiased and minimum variance estimator exists as the signal-to-noise ratio (SNR) becomes high, it is guaranteed to be the MLE. Since exact expressions for the bias, variance and error correlation of the MLE are often difficult or impractical to derive analytically, it has become popular in computer vision, pattern recognition, image processing, and other disciplines to instead compute limiting bounds such as the CRLB since these are usually much easier to obtain. The CRLB, however, typically provides an unrealistically optimistic approximation to the MLE variance in many nonlinear inverse problems.

Recently, general analytic conditions for the maximum likelihood estimate to become asymptotically unbiased and attain minimum variance have been derived [20, 29]. In this paper, we apply these general statistical results to the classical computer vision problems of 3D motion and structure estimation for rigid objects. The "Cramer-Rao lower bound [..] plays an essential role" [19] in computer vision and has been widely used in the literature to address object motion and structure estimation problems [3, 4, 5, 6, 7, 8, 9, 13, 19, 20, 21, 22, 23, 25, 26, 27, 28]. Here we derive analytical expressions that are necessary for the Cramer-Rao lower bound to be a good approximation to the mean-square error. Since the structure of real-world objects can be approximated by a collection planar surfaces, we focus on the problem of estimating the pose and motion of planar surfaces. We show analytically and experimentally that the tightness of the bound depends on the noise level, surface texture, and the number of pixels comprising the surface. In particular, we find that parameters describing surfaces with little texture are more difficult to estimate than those describing surfaces with sharply varying brightness levels. For surfaces with little texture, the mean-square error for position estimation cannot be approximated well by the Cramer-Rao lower bound, even in low signal-to-noise.

The approach is to apply the tools of higher order asymptotic inference, which rely heavily on tensor anal-

*The support of the Office of Naval Research, the National Science Foundation (IIS-0093367) and the Whitaker Foundation is acknowledged. Email: betke@cs.bu.edu, http://www.cs.bu.edu/faculty/betke.

0-7695-1272-0/01 $10.00 © 2001 IEEE

ysis to expand the MLE as a series in inverse orders of SNR. From this series, analytic expressions for the first-order bias, second-order covariance, and second-order error correlation of a general MLE are presented in terms of the joint moments of derivatives of the log-likelihood function with respect to the parameters to be estimated. Since the first-order error correlation is shown to be the CRLB, which is only valid for unbiased estimates, the second-order error correlation provides a tighter error bound on the MLE than the CRLB that is applicable in relatively low SNR even when the MLE is biased to first order. These expressions are then used to determine general analytic requirements on SNR and sample size that are necessary for a MLE to become asymptotically unbiased and attain minimum variance. This is done by showing when the first-order bias becomes negligible compared to the true value of the parameter and when the second-order covariance term becomes negligible compared to the CRLB. The first-order bias is evaluated for general multivariate Gaussian data. The second-order covariance and error correlation terms are evaluated for the special case of additive Gaussian noise with parameter-independent covariance.

2 Uncertainty Models in Computer Vision

In this section, we discuss various approaches to formalize statistical estimation problems in computer vision that differ in the way that uncertainty is modeled. Our theoretical results are applicable to all these models and to any probability distribution. We illustrate them experimentally in one example.

2.1 Geometric Uncertainty

For the "geometric estimation from noisy data" [19], parameterizations are developed to model object shape. The shape parameters are then estimated in the presence of noise in observations that are 2D image or 3D world coordinates of object points in the scene.

The approach has been used to estimate the parameters that describe circles [25], lines, and conics [19]. The observed data, i.e., points on these curves, are assumed to be corrupted by additive zero-mean Gaussian noise. Cramer-Rao lower bounds have been derived to approximate the mean-square estimation error [19, 25].

The geometric uncertainty framework has also been used to address 3D motion parameter estimation [7, 27, 8, 26, 16]. The observed data are 3D coordinates of object points that are obtained from the analysis of a sequence of monocular [7, 8, 26, 16] and stereo [27] images. It is assumed that the correspondence of the object points has been established, but that the coordinate positions are corrupted by zero-mean Gaussian noise. CRLBs for estimating rotation, translation, velocity, and acceleration parameters have been derived in the above references.

2.2 Noise in Brightness Measurements

In this uncertainty model, the noisy observations are the measured image brightness values. The noise is due to the camera system, for example, introduced by camera defocus, electronic shot noise of CCD cameras, or small mechanical vibrations. Experiments with camera systems have shown that this noise is Gaussian [6, 23]. This Gaussian distribution is expected as a consequence of the central limit theorem. In particular, let vector $\mathbf{I}$ represent a $K_x \times K_y$ image $I(x, y)$ where the rows of the image are concatenated into a column vector. Each component I_k of $\mathbf{I}$ contains an independent intensity measurement $I(x, y)$ for $1 \leq k \leq K$, where $K = K_x K_y$ is the number of pixels.

Experiments have shown that the standard deviation does not depend on brightness mean or pixel position. The noise, therefore, is additive and signal independent. The measured average skew and kurtosis are very close to the corresponding Gaussian values at each pixel. The sample covariance $\mathbf{C}$ of brightness between image pixels indicates that the brightness measurements are statistically independent across the pixels, i.e., $\mathbf{C} = \sigma^2 \mathbf{1}$, where $\mathbf{1}$ is the identity matrix. The probability density for $\mathbf{I}$ is therefore approximately $P(\mathbf{I}) = 1/(2\pi\sigma^2)^{K/2} \exp(-1/2\sigma^2 \sum_{k=1}^{K} (I_k - \mu_k)^2)$ for $0 \leq I_k < \infty$, where the variance σ is constant and the mean μ_k varies throughout the image.

This uncertainty model has been used to measure the extent to which a planar object can be resolved under affine parameterization and CRLBs have been derived [6]. For the depth-from-defocus problem, where depth is estimated from observed brightness values that are blurred due to camera defocus, CRLBs have also been derived [23].

2.3 Algorithmic Uncertainty

In many computer vision estimation problems, observations are not the measured brightness values directly, but instead are obtained from (preprocessing) algorithms that may introduce noise. These algorithms address, for example, edge detection, optical flow, stereo correspondence, and image reconstruction. It is widely assumed that the observations are corrupted by additive zero-mean Gaussian noise [2, 9, 12, 13, 14, 17, 21, 22, 28].

Cramer-Rao lower bounds are derived for estimation of the motion parameters of simulated rigid objects given noisy optical flow input [2, 13, 12, 21, 28]. The mean-square error of Canny's algorithm has been shown to be higher by a factor of two than the lower limit established by the Cramer-Rao bound [17]. In medical imaging, CRLBs on the error variances of image reconstruction for computed tomography, ultrasound tomography, and magnetic resonance have been derived [10, 3]. CRLBs have been derived for estimation of quadric surfaces given corresponding image patches viewed by a stereo system, where the correspondence algorithm introduce zero-mean Gaussian noise [9]. CRLBs have also been derived for estimation of tilt and slant angles [22] and elongation and orientation [14] of planar textured surfaces.

2.4 Foreground/Background Uncertainty

In order to recognize objects in a scene, which is the overall goal of many estimation techniques, transformed object templates are often matched to the scene image. Recent work proposes to model object and scene variations by learning probability distributions of the foreground and background with sampling methods [24]. Bayesian learning of posterior distributions is based on estimating prior distributions, which are often assumed to be uniform, and the assumption that large sample sets can be obtained. However, sample sizes are commonly too small to provide the distribution reliably.

Our approach is based on classical rather than Bayesian estimation theory. The sample-size issues we discuss here are therefore unrelated to the sampling issues in [24]. To compute the ML estimate, an analytical form of the distribution is required. In scenes with zero-mean backgrounds and fixed object sizes, the sum-squared difference or correlation measures maximize the likelihood; in more general scenes, they approximate the MLE [20].

3 Projection Model

Commonly, the central projection model is used to describe the imaging geometry [15]. The origin of a 3D coordinate system is placed at the projection center. The x and y-coordinate axes are aligned with the image plane, the optical axis serves as the z-coordinate axis, and the distance between projection center and image is given by focal length f. Spatial point coordinates $\mathbf{p} = (X, Y, Z)^T \in R^3$ are mapped into image coordinates $h(\mathbf{p}) = (x, y)^T \in R^2$. If the depth Z of a scene point $\mathbf{p} = (X, Y, Z)^T$ is known, the scene point can be related to image point $(x, y)^T$ by the perspective projection: $\mathbf{p} = (X, Y, Z)^T = (x, y, f)^T \frac{Z}{f} = (\frac{x}{f}, \frac{y}{f}, 1)^T Z$. If a calibrated stereo system is available, the Z-coordinates of scene points can be recovered [15].

4 Pose and Motion of Rigid Objects

We are interested in the recognition of rigid objects and their motion, in particular, objects with surfaces that can be well approximated by a collection of polygonal faces. The *pose* of such a polygonal face can be described by the position of its centroid and the orientation of a unit vector normal to the surface. We assume that each planar surface has an initial position and orientation with respect to the camera. The initial view of the object serves as the object's "surface model." To recognize the object in an arbitrary scene, the transformation of the surface model into the scene surface must be determined.

Similarly, for the motion recognition problem, we assume that a rigid object is given in an initial position and orientation. To recognize the motion of the object, the transformation of the initial object pose to the new pose must be determined.
The six-parameter perspective transformation

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix} \qquad (1)$$

maps an object point (x_0, y_0) into a new object point (x, y), where vector $(a_{13}, a_{23}, 1)$ describes the translation of the object and parameters $a_{11}, a_{12}, a_{21}, a_{22}$ the linear distortion of the object. The perspective transformation is used to model to the motion of the object or differences between object and model pose. It is convenient to describe the linear distortion of the object in terms of its rotation α, contractions s_x, s_y, and skew ξ, and its position in terms of its translation $(t_x, t_y, 1)$. The affine transformation can then be written as $(x, y, 1)^T =$

$$\begin{pmatrix} s_x \cos(\alpha) & s_y \sin(\alpha) & t_x \\ -s_x \sin(\alpha + \xi) & s_y \cos(\alpha + \xi) & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix}. \qquad (2)$$

5 Object Pose and Motion Estimation

The problem of recognizing a rigid object's pose or motion in a scene can be described as the problem of estimating the parameter vector $\boldsymbol{\theta} = (a_{11}, ..., a_{23})$ given the image data $\mathbf{I}$. Using the noise model introduced in Section 2.2, the likelihood function for $\boldsymbol{\theta}$, given the image data $\mathbf{I}$, is

$$P(\mathbf{I}|\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{K/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{K} (I_k - m_k(\boldsymbol{\theta}))^2 \right) \qquad (3)$$

where the mean $m_k(\boldsymbol{\theta})$ explicitly depends on the parameters to be estimated, while the noise variance σ^2

is independent of the parameter set. For n independent and identically distributed snapshots of the image, the probability density is $\prod_{i=1}^{n} P(\mathbf{I}|\boldsymbol{\theta})$. Note K is the length of image vector $\mathbf{I}$ and differs from n.

5.1 Definitions and Notation

The log-likelihood function $\ell(\boldsymbol{\theta})$ is defined as $\ell(\boldsymbol{\theta}) = \ln(\prod_{i=1}^{n} P(\mathbf{I}|\boldsymbol{\theta}))$ when evaluated at measured brightness values $\mathbf{I}$. Let the number of parameters be m and the rth coordinate of $\boldsymbol{\theta}$ be denoted by θ^r. The partial derivative of the log-likelihood function with respect to θ^r is then denoted by $\ell_r = \partial\ell/\partial\theta^r$, for $1 \le r \le m$.

We use the letter ν for joint moments of the log-likelihood derivatives. More specifically, if $R_1 = r_{11}\ldots r_{1n_1}, \ldots, R_m = r_{m1}\ldots r_{mn_m}$ are sets of coordinate indices, the joint moments of the log-likelihood derivatives are $v_{R_1,\ldots,R_m} = \mathrm{E}[\ell_{R_1}\ldots\ell_{R_m}]$. Here n_m indicates the order of the derivative of the mth parameter. For example, assume $R_1 = rs$ and $R_2 = t$. Then $\nu_{R_1} = \nu_{rs} = \mathrm{E}[\ell_{rs}] = \mathrm{E}[\frac{\partial^2\ell}{\partial r\partial s}]$ and $\nu_{R_1,R_2} = \nu_{rs,t} = \mathrm{E}[\ell_{rs}\ell_t] = \mathrm{E}[\frac{\partial^2\ell}{\partial r\partial s}\frac{\partial\ell}{\partial t}]$.

For two indices r and s, the Fisher information is defined by $J_{rs} = E[\ell_r\ell_s]$. To describe the components of the inverse Fisher information matrix $\mathbf{J}^{-1}$, we lift the indices, i.e., $J^{rs} = [\mathbf{J}^{-1}]_{rs}$. In general, lifting the indices produces quantities that are denoted by $\nu^{R_1,\ldots,R_m} = J^{r_{11}s_{11}}\ldots J^{r_{mn_m}s_{mn_m}}\nu_{s_{11}\ldots s_{1n_1},\ldots,s_{m1}\ldots s_{mn_m}}$.

We use the Einstein summation convention, which says that the product $\mathbf{a}^T\mathbf{b} = \sum a_i b_i$ of two vectors $\mathbf{a}$ and $\mathbf{b}$ can be written as $a^i b_i$. So whenever an index occurs twice in a term, once in the subscript, once in the superscript, summation over the index is implied.

5.2 Asymptotic expansions of bias, error correlation, and covariance of the MLE

The maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}$ is defined to be the value of $\boldsymbol{\theta}$ that maximizes the log-likelihood function $\ell(\boldsymbol{\theta})$ for the measured image data $\mathbf{I}$. It can be expressed as an asymptotic expansion around $\boldsymbol{\theta}$ in increasing orders of inverse sample size n^{-1} or equivalently $(\mathrm{SNR})^{-1}$. Typically we have $n = 1$, a single snapshot, with high SNR. Note that SNR is always linearly proportional to n but not necessarily to K. The generic component ℓ_r can be expanded around $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}} = \ell_r + \ell_{rs}(\hat{\theta}-\theta)^s + 1/2\,\ell_{rst}(\hat{\theta}-\theta)^s\hat{\theta}-\theta)^t + 1/6\,\ell_{rstu}(\hat{\theta}-\theta)^s\hat{\theta}-\theta)^t\hat{\theta}-\theta)^u + \ldots, \quad (4)$$

where $(\hat{\theta}-\theta)^s = \hat{\theta}^s - \theta^s$, etc. [1]. Inversion of Eq. 4 yields

$$(\hat{\theta}-\theta)^r = J^{rs}\ell_s\nabla + \tfrac{1}{2}\nu^{rst}\ell_s\ell_t + J^{rs}J^{tu}H_{st}\ell_u\nabla + \tfrac{1}{6}(\nu^{rstu} + 3\nu^{rs\nu}J_{\nu w}\nu^{wtu})\ell_s\ell_t\ell_u + \nu^{rsu}J^{t\nu}H_{u\nu}\ell_s\ell_t + \tfrac{1}{2}J^{rs}\nu^{tu\nu}H_{st}\ell_u\ell_\nu + \tfrac{1}{2}J^{rs}J^{tu}J^{\nu w}H_{st\nu}\ell_u\ell_w + J^{rs}J^{tu}J^{\nu w}H_{st}H_{u\nu}\ell_w\nabla + \ldots \quad (7)$$

where $H_R = \ell_R - \nu_R$ and the symbol ∇ indicates a drop of asymptotic magnitude of order $n^{-1/2}$ under ordinary repeated sampling, or equivalently a drop of $SNR^{-1/2}$. More specifically, the first term $J^{rs}\ell_s$ is of order $O_p(n^{-1/2})$, which means a polynomial of order $n^{-1/2}$. The second term $\frac{1}{2}\nu^{rst}\ell_s\ell_t + J^{rs}J^{tu}H_{st}\ell_u$ is of order $O_p(n^{-1})$, and the remaining terms that are shown are of order $O_p(n^{-3/2})$. From Equation (7), Naftali and Makris [20] derive general analytic expressions for the bias, error correlation, and covariance of the maximum likelihood estimate to the second order. The first-order bias of the maximum likelihood estimate $\hat{\theta}$ is $b(\hat{\theta}^r) =$

$$\mathrm{E}[(\hat{\theta}-\theta)^r] = \frac{1}{2}J^{rs}J^{tu}(v_{stu}+2v_{st,u})\nabla + O_p(n^{-3/2}) \quad (8)$$

as shown in [1, 20]. A necessary condition for the MLE to become asymptotically unbiased is for this first-order bias to become much smaller than the true value of the parameter θ^r.

The error correlation of the MLE is shown to be [20] $\mathrm{Cor}\left(\hat{\theta}^r, \hat{\theta}^a\right) =$

$$\mathrm{E}[(\hat{\theta}-\theta)^r(\hat{\theta}-\theta)^a] = J^{rs}\ \nabla\nabla + S\ \nabla\nabla, \quad (9)$$

where the second order term S is defined as $2J^{mb}J^{nc}\nu_{lnm}(J^{rs}J^{la} + J^{as}J^{lr})\nu_{s,b,c}(n^1) + \frac{1}{2}J^{cd}J^{ef}(J^{rs}J^{ab} + J^{as}J^{rb})\nu_{bce,d,f,s}(n^2) + J^{tu}(J^{rs}J^{ab}J^{cd} + J^{rd}J^{ab}J^{cs} + J^{ad}J^{rb}J^{cs})\nu_{st,u,bc,d}(n^2) + J^{bm}J^{cq}J^{tp}\nu_{lmn}\nu_{opq}(\frac{1}{4}J^{rl}J^{ao}J^{sn} + \frac{1}{2}J^{rs}J^{al}J^{on} + \frac{1}{2}J^{as}J^{rl}J^{on})\nu_{s,t,b,c}(n^2) + \frac{1}{2}J^{sm}\nu_{lmn}(J^{tn}J^{cd}(J^{rl}J^{ab} + J^{al}J^{rb}) + 2J^{bn}J^{cd}(J^{rl}J^{at} + J^{al}J^{rt}) + J^{cl}J^{tn}(J^{rd}J^{ab} + J^{ad}J^{rb}))\ \nu_{s,t,bc,d}(n^2) + \frac{1}{6}J^{mb}J^{nc}J^{od}\nu_{lmno}(J^{rs}J^{la} + J^{as}J^{rl})\nu_{s,b,c,d}(n^2) + 4J^{bm}(J^{rs}J^{al} + J^{as}J^{rl})\nu_{s,m,lb}(n^1)$, where notation such as $\nu_{bce,d,f,s}(n^2)$ means the n^2 order terms of the joint moment $\nu_{bce,d,f,s}$. Note that the first term J^{rs} is of order $O_p(n^{-1})$, while the remaining terms are of order $O_p(n^{-2})$.

The covariance of the MLE is $\mathrm{Cov}\left(\hat{\theta}^r, \hat{\theta}^s\right) =$

$$\mathrm{E}[(\hat{\theta}^r - \mathrm{E}[\theta^r])(\hat{\theta}^s - \mathrm{E}[\theta^s])] = J^{rs}\ \nabla\nabla + S - Q \quad (10)$$

where $Q = \frac{1}{4}J^{rs}J^{tu}J^{aw}J^{yz}(\nu_{stu}\nu_{wyz} + 2\nu_{stu}\nu_{wy,z}(n^1) + 2\nu_{st,u}\nu_{wyz}(n^1) + 4\nu_{st,u}\nu_{wy,z}(n^1))\ \nabla\nabla$.

The first term J^{rs} is of order $O_p(n^{-1})$ and the remaining terms are of order $O_p(n^{-2})$.

5.3 Cramer-Rao Lower Bound

The first order term J^{rs} in Eqs. 9 and 10 is the Cramer-Rao lower bound [20]. A necessary condition for the MLE to attain minimum variance is for the second-order term in Eq. 10 to become much smaller than the first-order term, the CRLB. For Gaussian data, $J_{rs} = n/\sigma^2 \sum\sum_{x,y=1}^{K}(\frac{\partial m(x,y;\boldsymbol{\theta})}{\partial\theta_r}\frac{\partial m(x,y;\boldsymbol{\theta})}{\partial\theta_s})$.

The pixels in the background that are located next to the object's occluding contour will generally be affected by small changes in $\boldsymbol{\theta}$. Intuitively, it is easier to estimate $\boldsymbol{\theta}$ if there is large contrast between object and background. To provide an error bound that does not vary with the unknown object background, it is practical to define the parameter estimation problem only for set O^- of pixels comprising the expected object $q(x,y;\boldsymbol{\theta})$ and not its boundary. With this definition, the brightness contrast between object and unknown background does not contribute to the object's Fisher information. The Fisher information matrix then becomes

$$J_{ij} = \frac{1}{\sigma^2}\sum_{(x,y)\in O^-}\left(\frac{\partial q(x,y;\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial q(x,y;\boldsymbol{\theta})}{\partial\theta_j}\right). \quad (11)$$

We will use the letter n to indicate the number of pixels in the set O^-.

5.4 Asymptotic Bias, Error Correlation and Covariance of the MLE for Gaussian Data

The problem of recognizing an object's pose or motion in a scene is described as the problem of estimating a parameter vector $\boldsymbol{\theta}$ in additive Gaussian noise with parameter-independent covariance (Eq. 3). The first-order bias for multivariate Gaussian data is $b(\hat{\theta}^r) = \sum_{s=1}^{m}\sum_{t=1}^{m}\sum_{u=1}^{m} \frac{2}{n} [J^{-1}]_{rs}[J^{-1}]_{tu}(\frac{1}{2}\mathrm{tr}(\mathbf{C}^{-1}\frac{\partial^2\mathbf{C}}{\partial\theta^s\partial\theta^t}\mathbf{C}^{-1}\frac{\partial\mathbf{C}}{\partial\theta^u}) - \frac{1}{2}\sum_{s,t}\mathrm{tr}(\mathbf{C}^{-1}\frac{\partial^2\mathbf{C}}{\partial\theta^s\partial\theta^u}\mathbf{C}^{-1}\frac{\partial\mathbf{C}}{\partial\theta^t})^T + (\frac{\partial^2\boldsymbol{\mu}}{\partial\theta^s\partial\theta^t})^T\mathbf{C}^{-1}(\frac{\partial\boldsymbol{\mu}}{\partial\theta^u}) - \sum_{s,t}(\frac{\partial^2\boldsymbol{\mu}}{\partial\theta^s\partial\theta^u})^T\mathbf{C}^{-1}(\frac{\partial\boldsymbol{\mu}}{\partial\theta^t}) + (\frac{\partial\boldsymbol{\mu}}{\partial\theta^s})^T\mathbf{C}^{-1}(\frac{\partial\mathbf{C}}{\partial\theta^u})\mathbf{C}^{-1}(\frac{\partial\mathbf{C}}{\partial\theta^t}) - \sum_{s,t}(\frac{\partial\boldsymbol{\mu}}{\partial\theta^s})^T\mathbf{C}^{-1}(\frac{\partial\mathbf{C}}{\partial\theta^t})\mathbf{C}^{-1}(\frac{\partial\boldsymbol{\mu}}{\partial\theta^u}))$.
The second-order bias, error correlation, and covariance have been evaluated similarly for Gaussian data [20, 29].

5.5 Single Parameter Case

For the case of image data in additive Gaussian noise that depends on a single parameter θ, the bias, mean-square error, and variance of the MLE $\hat{\theta}$ can be derived from Eqs. 8–10. The bias $b(\hat{\theta})$ is zero for the scalar case. The mean-square error and variance are therefore the same and reduce to $\mathrm{var}(\hat{\theta}) = \mathrm{MSE}(\hat{\theta}) - b^2(\hat{\theta}) =$

$$\mathrm{MSE}(\hat{\theta}) = \frac{\sigma^2}{S_1}\nabla\nabla + \left(\frac{\sigma^2}{S_1}\right)^2\frac{S_4}{S_1}\nabla\nabla + O_p(n^{-3}), \quad (12)$$

where
$S_1 = \sum_{(x,y)\in O^-}\left(\frac{\partial q(x,y;\theta)}{\partial\theta}\right)^2, S_4 = \sum_{(x,y)\in O^-}\left(\frac{\partial^2 q(x,y;\theta)}{\partial\theta^2}\right)^2.$

5.6 Position Estimation

Consider the example that the true position $(\theta_1,\theta_2) = (x_0,y_0)$ of the object q is unknown. Let us estimate the coordinates x_0 and y_0 separately. The maximum likelihood estimate of parameter x_0 corresponds to the peak output of a matched filter for a signal in additive Gaussian noise. To derive the joint moments of the log-likelihood function for x_0, we only need to evaluate the first and second derivatives of object q with respect to x_0 and substitute them into the sums

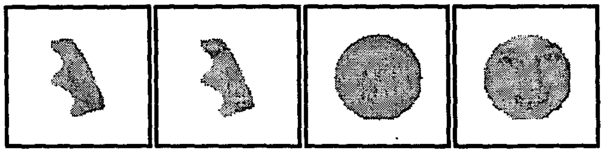
$$S_1 = \sum_{(x,y)\in O^-}\left(\frac{\partial q(x-x_0)}{\partial x_0}\right)^2 = \sum_{(x,y)\in O^-}\left(\frac{\partial q(x)}{\partial x}\right)^2, \quad (13)$$

$$S_4 = \sum_{(x,y)\in O^-}\left(\frac{\partial^2 q(x-x_0)}{\partial x_0^2}\right)^2 = \sum_{(x,y)\in O^-}\left(\frac{\partial^2 q(x)}{\partial x^2}\right)^2. \quad (14)$$

The mean square error of the MLE in position coordinate x_0 is then evaluated by Eq. 12.

6 Experiments and Discussion

We consider the example of position estimation as described in Section 5.6. We use Eq. 12 to compute the mean-square error in the estimate of position coordinate x_0 for various planar surfaces. The first set of surfaces are shown in Fig. 1 with a (1) smooth texture, and (2) face texture. The Cramer-Rao lower bounds are shown as functions of image noise in Fig. 2.



Surface A.1 Surface A.2 Surface B.1 Surface B.2

Figure 1: Images of two planar surfaces in space. On the left, surfaces A.1 and B.1 have smoothly varying brightness patterns. On the right, a face pattern is superimposed on surfaces A.3 and B.3.

Each graph in Fig. 2 is plotted as a function of the standard deviation of the image noise. The standard deviation shown as a fraction of the maximum dynamic range of the brightness. The errors lie in the subpixel range. Consumer video cameras typically take images that contain at most $2^8 = 256$ brightness levels and have a high signal-to-noise ratio. For example, with a low standard deviation of $\sigma = 3$ brightness levels, which is $3/255 = 1.2\%$ of the maximal dynamic range, the CRLBs on position error for the textured surfaces in Fig. 1 are shown in Table 1 below. The errors in position estimation are given in pixel widths.

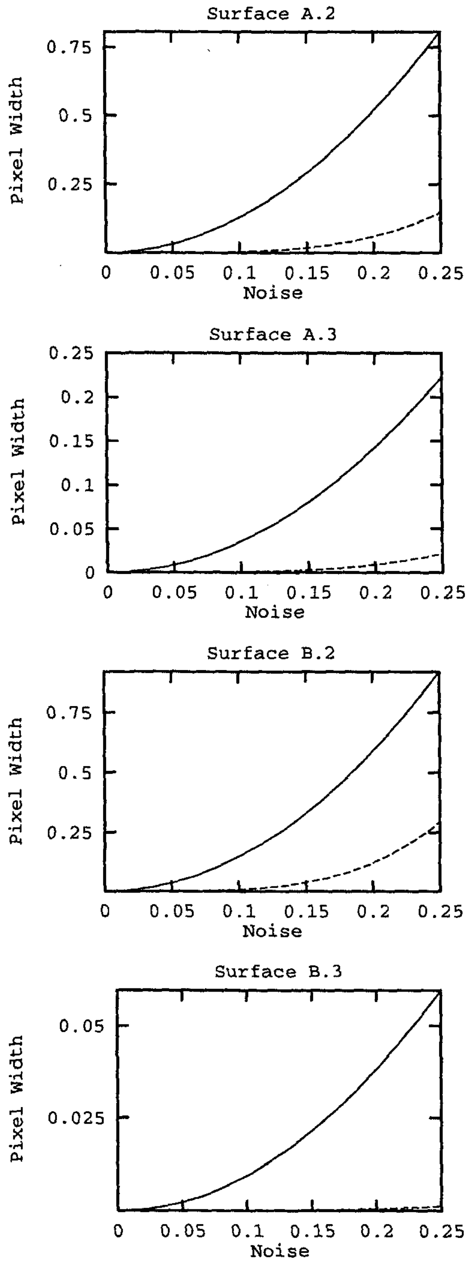


Figure 2: The CRLB on position error (solid lines) and the second-order error term (dashed lines) are shown as a functions of Gaussian additive noise. The standard deviation of the noise is plotted as the fraction of the maximum dynamic range of the brightness.

Table 1: CRLB on Position Error with $\sigma = 3$ Gray Levels

Surface	Smooth Texture	Face Texture	Number of Pixels K
A	0.00178	0.00049	795
B	0.00202	0.00013	2821

Since the same texture is superimposed on surfaces A.3 and B.3, the mean-square error depends on the number of pixels that comprise the respective surfaces, and is therefore larger for surface A.3 than for surface B.3. A comparison of the results for the same surface shape but different textures illustrates the dependence of the CRLB bound on texture. The intuitive argument that it is easier to estimate the position of a surface with a strongly varying texture than a surface with small texture changes is demonstrated by our quantitative results.

A necessary condition for the maximum likelihood estimate of position coordinate x_0 to attain minimum variance is for the second-order term $(\sigma^2/S_1)^2 \, S_4/S_1$ in Eq. 12 to become much smaller than the first-order term σ^2/S_1, the Cramer-Rao lower bound. We compute the ratio

$$\frac{(\sigma^2/S_1)^2 \, S_4/S_1}{\sigma^2/S_1} = \frac{\sigma^2 S_4}{S_1} \qquad (15)$$

of the second- to the first-order terms. Note that this ratio is small when $\sigma << \sqrt{S_1/S_4}$. The sums S_1 and S_4 depend on the number of pixels on the surface and the brightness changes on the surface.

Figure 2 shows the second-order term as a function of noise level for the four surfaces (dashed lines). Our experiments show for these surfaces that the second-order term can be neglected for noise levels that are typical for consumer video cameras. However, the second-order term cannot be neglected for surfaces with little texture, such as the transformed surfaces shown in Figs. 3 and 4. The surfaces shown in these images have been created with a small number of digitization levels. Surfaces C-1, D-1, C-2 and D-2 have been created with 3 and 6 brightness levels, respectively. For Surface D-1 the second-order error term is larger than the CRLB for noise a level of $\sigma \geq 3$ which is typical for a consumer video camera.

7 Conclusions

We have used higher order asymptotic statistical theory to derive necessary analytic conditions for the Cramer-Rao lower bound to become a good approximation to the mean-square error of parameter estimates for the 3D structure and motion of rigid objects.

The accuracy of the approximation depends on the noise level, the number of pixels comprising the surface, and the texture of the surface. Our analytical

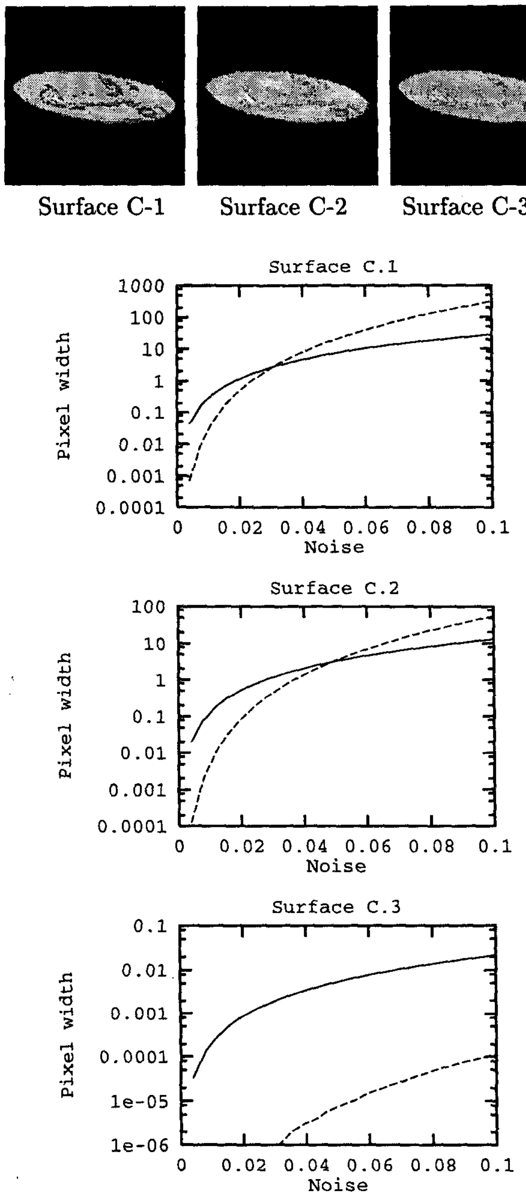

Surface C-1 Surface C-2 Surface C-3

Figure 3: Surfaces C-1 and C-2 have textures for which the second-order term cannot be neglected. For Surface C-1, the second-order term is larger than the CRLB for noise levels $\sigma \geq 8$, which corresponds to 3% of the maximum dynamic range. For Surface C-2, the second-order term is larger than the CRLB for noise levels $\sigma \geq 12$, which corresponds to 4.6% of the maximum dynamic range.

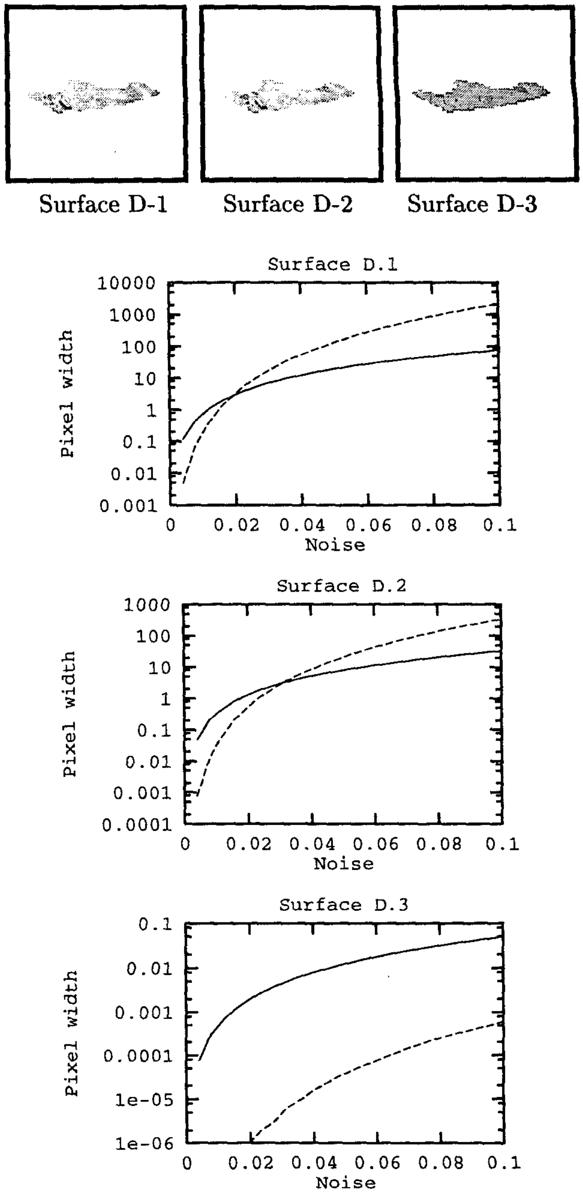

Surface D-1 Surface D-2 Surface D-3

Figure 4: Surfaces D-1 and D-2 also have textures for which the second-order term cannot be neglected. For Surface D-1, the second-order term is larger than the CRLB for noise levels $\sigma \geq 3$, which corresponds to 1.2% of the maximum dynamic range. For Surface D-2, the second-order term is larger than the CRLB for noise levels $\sigma \geq 8$, which corresponds to 3% of the maximum dynamic range.

results are applied in experiments using surfaces with smoothly and sharply varying brightness levels. For surfaces with smoothly varying textures and surfaces with a small number of digitization levels, the CRLB does not approximate the mean-square error for position estimation well. We provide necessary conditions for the CRLB to be attained that depend on the size, texture, and noise level of the surface patch.

References

[1] O. Barndorf-Nielsen and D. Cox. *Inference and Asymptotics.* Chapman & Hall, UK, 1994.

[2] Y. Barniv. Error analysis of combined optical-flow and stereo passive ranging. *IEEE Trans Aerosp Electron Syst*, 28(4):978–989, 1992.

[3] S. Basu and Y. Bresler. Feasibility of tomography with unknown view angles. *IEEE Trans Image Process*, 9(6):1107–1122, 2000.

[4] M. Betke and N. C. Makris. Fast object recognition in noisy images using simulated annealing. In *ICCV*, pages 523–530, 1995.

[5] M. Betke and N. C. Makris. Information-conserving object recognition. In *ICCV*, pages 145–152, 1998.

[6] M. Betke and N. C. Makris. Recognition, resolution and complexity of objects subject to affine transformation. *Int J Comput Vis*, 44(1), September 2001.

[7] T. Broida and R. Chellappa. Performance bounds for estimating to three-dimensional motion parameters from a sequence of noisy images. *J Opt Soc Am*, 6(6):879–889, 1989.

[8] T. Broida and R. Chellappa. Estimating the kinematics and struture of a rigid object from the sequence of monocular images. *IEEE Trans Pattern Anal Mach Intell*, 13(6):497–513, 1991.

[9] B. Cernuschi-Frias, D. Cooper, Y.-P. Hung, and P. Belhumeur. Toward a model-based Bayesian theory for estimating and recognizing parameterized 3-D objects using two or more images taken from different positions. *IEEE Trans Pattern Anal Mach Intell*, 11:540–564, 1989.

[10] A. Devaney and G. Tsihrintzis. Maximum likelihood estimation of object location in diffraction tomography. *IEEE Trans Signal Process*, 39(3):672–682, 1991.

[11] R. Duda, R. Hart, and D. Stork. *Pattern Classification, 2nd Ed.* John Wiley & Sons, NY, 2001.

[12] A. M. Earnshaw and S. D. Blostein. The performance of camera translation direction estimators from optical flow: analysis, comparison, and theoretical limits. *IEEE Trans Pattern Anal Mach Intell*, 18(9):927–932, 1996.

[13] T. Endoh, T. Toriu, and N. Tagawa. The maximum likelihood estimator is not "optimal" on 3-D motion estimation from noisy optical flow. In *ICIP*, pages 247–251, 1994.

[14] K. Eom. Segmentation of monochrome and color textures using moving average modeling approach. *Image Vis Comput*, 17(3–4):233–244, 1999.

[15] O. Faugeras. *Three-Dimensional Computer Vision – A Geometric Viewpoint.* MIT Press, 1993.

[16] M. Jayakumar and R. Banavar. Risk-sensitive filters for recursive estimation of motion from images. *IEEE Trans Pattern Anal Mach Intell*, 20(6):659–666, 1998.

[17] R. Kakarala and A. Hero. On achievable accuracy in edge localization. *IEEE Trans Pattern Anal and Mach Intell*, 14(7):777–781, 1992.

[18] K. Kanatani. *Statistical optimization for geometric computation: Theory and practice.* Elsevier Science, Amsterdam, Netherlands, 1996.

[19] K. Kanatani. Cramer-Rao lower bounds for curve fitting. *CVGIP*, 60(2):93–99, 1998.

[20] E. Naftali and N. C. Makris. Necessary conditions for a maximum likelihood estimate to become asymptotically unbiased and attain the Cramer-Rao lower bound – Part 1: General approach with an application to time-delay and Doppler shift estimation. *J Acoust Soc Am.* In press for Oct or Nov 2001.

[21] O. Nestares, D. Fleet, and D. Heeger. Likelihood functions and confidence bounds for total-least squares problem. In *CVPR*, pages 523–530, 2000.

[22] H. Permuter and J. Francos. Estimating the orientation of planar surfaces: algorithms and bounds. *IEEE Trans Inf Theory*, 46(5):1908–1920, 2000.

[23] A. Rajagopalan and S. Chaudhuri. Performance analysis of maximum likelihood estimator for recovery of depth from defocused images and optimal selection of camera parameters. *Int J Comput Vis*, 30(3):175–190, 1998.

[24] J. Sullivan, A. Blake, M. Isard, and J. MacCormick. Object localization by Bayesian correlation. In *CVPR*, pages 1068–1075, 1999.

[25] S. Thomas and Y. Chan. Cramer-Rao lower bounds for estimation of a circular arc center and its radius. *CVGIP*, 57(6):527–532, 1995.

[26] J. Weng, Y. Cui, and N. Ahuja. Transitory image sequences, asymptotic properties, and estimation of motion and structure. *IEEE Trans Pattern Anal Mach Intell*, 19(5):451–464, 1997.

[27] G.-S. Young and R. Chellappa. 3-D motion estimation using a sequence of noisy stereo images, models, estimation, and uniqueness results. *IEEE Trans Pattern Anal Mach Intell*, 12:735–759, 1990.

[28] G.-S. Young and R. Chellappa. Statistical analysis of inherent ambiguities in recovering 3-D motion from a noisy flow field. *IEEE Trans Pattern Anal Mach Intell*, 14(10):995–1013, 1992.

[29] M. Zanolin, E. Naftali, and N. C. Makris. Second order bias of a multivariate Gaussian maximum likelihood estimate with a chain rule for higher moments. *J Roy Stat Soc B*, 2001. Submitted.