Available online at www.sciencedirect.com
SCIENCE DIRECT
Vision Research 46 (2006) 3105–3120
Vision Research
www.elsevier.com/locate/visres

Noise causes slant underestimation in stereo and motion

Hui Ji, Cornelia Fermüller *

Computer Vision Laboratory, Center for Automation Research, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742-3275, USA

Received 28 June 2004; received in revised form 1 April 2006

Abstract

This paper discusses a problem, which is inherent in the estimation of 3D shape (surface normals) from multiple views. Noise in the image signal causes bias, which may result in substantial errors in the parameter estimation. The bias predicts the underestimation of slant found in psychophysical and computational experiments. Specifically, we analyze the estimation of 3D shape from motion and stereo using orientation disparity. For the case of stereo, we show that bias predicts the anisotropy in the perception of horizontal and vertical slant. For the case of 3D motion we demonstrate the bias by means of a new illusory display. Finally, we discuss statistically optimal strategies for the problem and suggest possible avenues for visual systems to deal with the bias.

Keywords: Shape estimation; Bias; Shape from motion; Stereo orientation disparity; Partial bias correction

1. Introduction

We are interested in the visual estimation processes through which we can understand the geometry of a scene and build three-dimensional models. More specifically, we will examine the computations that allow us to recover the 3D shape of scene surfaces. These computations are often referred to as shape from X, because cues such as motion (Spetsakis & Aloimonos, 1991; Tsai & Huang, 1984; Tomasi & Kanade, 1992), stereo (Scharstein & Szeliski, 2002), texture (Aloimonos, 1988; Bajcsy & Lieberman, 1976; Garding, 1993; Knill, 1998), shading (Ikeuchi & Horn, 1981; Nayar, Ikeuchi, & Kanade, 1990) and contours (Koenderink, 1984; Kutulakos & Dyer, 1995) encode information from which the shape of scene surfaces can be obtained.

The recovery of 3D shape is difficult. The main reason is that we have to segment the scene while we recover it. It is clear that there is an intricate interplay in the recovery and segmentation processes, which we do not fully understand yet. But we have a good understanding of the computations of the inverse geometric image formation allowing for the recovery of shape (Horn, 1986). That is, if we know where the continuous surfaces are (i.e., if we know the segmentation) and we know the surface property parameters, then we can obtain the shape.

However, computational experiments indicate that often the shape cannot be estimated correctly. For example when shape is estimated from multiple views, and even when the 3D viewing geometry is estimated correctly, the shape often is estimated incorrectly. It is known also in the psychophysical literature that human shape estimation is not veridical (Koenderink & van Doorn, 1992; Koenderink, van Doorn, & Kappers, 1995). For a variety of conditions and from a number of cues there is an underestimation of slant. Planar surface patches estimated from texture (Andersen, Braunstein, & Saidpour, 1998; Garding, 1993), contour (Perrone, 1982), stereopsis (Goutcher & Mamassian, 2002; Mitchison & McKee, 1990), and motion of various parameters (Todd & Perotti, 1999) have been found to be estimated with smaller slant, that is, closer in orientation to a front-parallel plane than they are. In this paper, we are asking whether there are computational reasons for the mis-estimation.

* Corresponding author. Fax: +1 301 414 9115.
E-mail addresses: jihui@cfar.umd.edu (H. Ji), fer@cfar.umd.edu (C. Fermüller).

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2006.04.010

In previous work, we have shown that there is a statistical problem with the estimation of image features (Fermüller & Malm, 2004; Fermüller, Pless, & Aloimonos, 2000). Here, we extend these concepts to the visual shape recovery processes. We show that there is bias and thus consistent erroneous mis-estimation in the estimation of shape. The underlying cause is the well known statistical dilemma. Since image data is noisy, in order to estimate well, we would need to obtain the statistics of the noise. However, because of the complexity of the computations it is most often not possible to accurately estimate the noise parameters. The result is that bias cannot be avoided.

We investigated the effects of bias and found that it is consistent with the empirical findings. In particular, we show in this paper that in the case of shape from motion for many 3D motions and for shape from stereo the bias causes an underestimation of slant. In another study, (Hui & Fermüller, 2003) we demonstrate that bias also causes underestimation in shape from texture. Thus, we find, that one of the reasons for inaccuracy in shape estimation is systematic estimation error, i.e., bias, which affects machine vision as well as biological vision.

1.1. The main concept and what this paper is about

The concepts underlying the statistical analysis are simple. The constraints in the recovery of shape can be formulated as linear equations in the unknown parameters. Thus we need to find the "best" solution to an over-determined equation system. We have equations in two unknown shape parameters (x, y) of the form

$$a_{1_i}x + a_{2_i}y = b_i. \tag{1}$$

In our problem (x, y) encode the two components of the surface normal vector $\boldsymbol{N} = (N_1, N_2, 1) = (x, y, 1)$. a_{1_i} and a_{2_i} are the observations, which are composed of multiple components. These components are the parameters of the image texture, that is the lines (edges) in the image and in the case of motion in addition the rotational parameters.

Consider, that we have n such equations, which we write in matrix for as

$$A\boldsymbol{x} = \boldsymbol{b} \tag{2}$$

with A an n by two matrix, $\boldsymbol{b}$ an n-dimensional vector and $\boldsymbol{x} = (x,y)$ the two-dimensional vector of unknowns. The observations a_{1_i}, a_{2_i} and b_i are always corrupted by errors. In the sequel unprimed letters are used to denote estimates, primed letters to denote the actual values, and δ's to denote errors, where $A = A' + \delta_A$ and $\boldsymbol{b} = \boldsymbol{b}' + \boldsymbol{\delta_b}$. Thus Eq. (2) can be written as

$$(A' + \delta_A)\boldsymbol{x} = (\boldsymbol{b}' + \boldsymbol{\delta_b}). \tag{3}$$

The most common choice to solving the system is by means of least squares (LS) estimation. Denoting the transpose and the inverse of a matrix by superscripts T and -1, respectively, the solution of the LS estimator $\boldsymbol{x}_{\mathrm{LS}}$ is characterized by

$$\boldsymbol{x}_{\mathrm{LS}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}. \tag{4}$$

However, it is well known, that under noisy conditions this estimator generally is biased (Fuller, 1987; van Huffel & Vandewalle, 1991).

What does this mean? Consider a problem for which you have a set of noisy measurements and you make an estimate. Then you choose another set of measurements and make another estimate. Continue many times. The expected value of your estimate is the average of your estimates in the limit. This expected value is not the true value. This is what we call statistical bias. Notice, there are no particular assumptions on the noise; it only needs to be symmetric around the true value.

Consider the simple case where all elements in δ_A and $\boldsymbol{\delta_b}$ are independent identically distributed (i.i.d.) random variables with zero mean and variance σ^2. Then under quite general conditions the expected value $E(\boldsymbol{x}_{\mathrm{LS}})$ of the estimate amounts to (Fuller, 1987)

$$E(\boldsymbol{x}_{\mathrm{LS}}) = \boldsymbol{x}' - \sigma^2\left(\lim_{n\to\infty}\left(\tfrac{1}{n}A'^{\mathrm{T}}A'\right)\right)^{-1}\boldsymbol{x}' \tag{5}$$

which implies that $\boldsymbol{x}_{\mathrm{LS}}$ is asymptotically biased (Stewart, 1990; van Huffel & Vandewalle, 1991).

The bias here is $\sigma^2(\lim_{n\to\infty}(\frac{1}{n}A'^{\mathrm{T}}A'))^{-1}\boldsymbol{x}'$. Please note, the bias does not depend on n, the number of measurements, which only shows up for the purpose of normalization, because $A'^{\mathrm{T}}A'$ is proportional to n.

An analysis of the bias term allows us to understand the errors in estimation. In general we can conclude that large variance in δ_A, an ill-conditioned A', or an $\boldsymbol{x}'$ which is oriented close to the eigenvector of the smallest eigenvalue of $A'^{\mathrm{T}}A'$ all could increase the bias and push the LS solution $\boldsymbol{x}_{\mathrm{LS}}$ away from the real solution. Generally it leads to an underestimation of vector $\boldsymbol{x}$.

The main reason for the bias are the errors in the explanatory variables A. If there are no errors in A, and there are errors only in $\boldsymbol{b}$, least squares estimation is unbiased. Basically the bias originates from the quadratic term $(A'^{\mathrm{T}}A')^{-1} = ((A' + \delta A)^{\mathrm{T}}(A' + \delta A))^{-1}$ and can be obtained from a second order expansion of Eq. (4).

To give the reader an intuition about the bias, we illustrate it by means of the line equations. Referring to Fig. 1, consider two lines $a_{1_1}x + a_{2_1}y = b_1$ and $a_{2_1}x + a_{2_2}y = b_2$ (the solid lines), which intersect in an acute angle. We want to find the intersection point. We don't have the exact lines, but we have noisy observations of these lines (the dashed lines). That is, the observed lines have orientations $(a_{1_1} + \delta_a, a_{2_1} + \delta_a)$ and $(a_{2_1} + \delta_a, a_{2_2} + \delta_a)$ and intercepts $(b_1 + \delta_b$ and $b_2 + \delta_b)$ with δ_a and δ_b i.i.d. random variables of zero mean. (However, note δ_b, the error in the intersections with the y-axis, does not contribute to the bias.) The least squares solution to the intersection of the noisy lines is found as the point closest in distance (L^2 norm) to all the lines. This is not the correct intersection, but a point closer to the origin, because the length of vector (x, y) is underestimated.

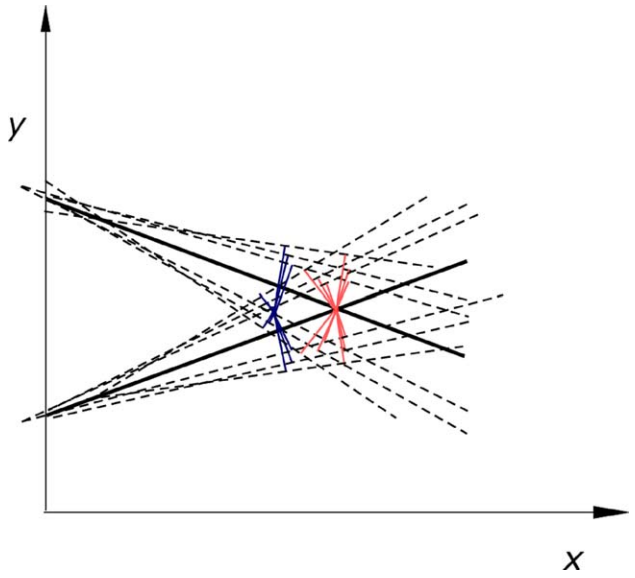

Fig. 1. Illustration of the bias by means of line equations. The solid lines denote the true constraints, and the dashed lines denote the noisy observations. The LS solution to the intersection of the noisy lines is the point with closest distance (smallest sum of squares distance) to all the dashed lines. The distances to the true solution (denoted by light gray (or red) lines) are larger than the distances to the estimated solution (denoted by dark gray (or blue) lines). Thus the estimated intersection point is the blue point, and not the correct red one. The estimate is closer to the origin, because the bias leads to underestimation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

The main assumption underlying our explanation is that our vision system uses least squares estimation. One may argue then that the bias is just an artifact of linear estimation. In Section 6 we discuss that this is not the case by taking a short digression into the statistical literature. In general classical estimation techniques will not be able to alleviate the bias significantly. The main reason is that usually there is not enough data available. If in some cases there is data to extract statistics, the best thing to do is to partially correct the bias, and this does not change the form of the bias. In Section 7 we present our hypothesis that the human visual system actually does partial correction and we show experiments that support this point of view.

The question is then, does this insight about the bias tell us anything about the way the human system estimates or about how theoretically machine vision should estimate shape? Section 8 discusses these issues. First, we could do better in the geometrical estimation problems by using color information. Three color pixels usually do not contain more information than one gray value pixel from a geometrical point of view. They do, however, contain statistical information, which we could exploit to improve the estimation. Second, in order to do good statistics, whatever we do, we do it better with larger amounts of data. We can only use large amounts of data if we have models of the scene. Thus, the bias makes a computational argument for the need to solve the estimations of image features, motion, structure, shape, and the segmentation in feedback loops. Having estimates about the 3D motion and the structure, we can segment the scene and apply our methods to larger amounts of data.

A number of previous studies have analyzed the statistics of visual processes. In particular, Kanatani (1996) discussed bias for some visual recovery processes. A few studies analyzed the statistics of structure from motion (Chowdhury & Chellappa, 2003; Daniilidis & Nagel, 1990; Maybank, 1993; Tomasi & Zhang, 1995). However, these analyses stayed at the general level of parameter estimation; no one has shown before the effects on the estimated shape.

2. Shape from multiple views

The 3D shape of a surface patch is described by the surface normal. In the literature on multiple view geometry e.g. (Backus, Banks, Ee, & Crowell, 1999; Faugeras & Luong, 2001; Hartley & Zisserman, 2000; Ma, Soatto, Kosecka, & Sastry, 2004) 3D shape is considered a by-product of the estimation of structure (the 3D coordinates of the scene). 3D shape is obtained as the spatial derivative of structure. However, one could estimate the 3D shape directly from the image texture without estimating the structure and without knowing the displacement (translation) between the cameras. Let us describe a surface patch by its local tangent plane. Estimating the structure of the patch means estimating three parameters: two parameters for the surface normal and one parameter for the depth (how far is the surface patch?). There is reference to this idea in the psychophysical stereo literature, where it is referred to as estimation from orientation disparity. The geometric constraints are explained next.

Consider two views of a planar patch separated only by translation $\boldsymbol{T}$ as in stereo (Fig. 2). Assume the patch contains a line L (an edge due to texture). The projections of this line on the two views with centers O and $\tilde{O}$ are then the two image lines $\boldsymbol{\ell}$ and $\tilde{\boldsymbol{\ell}}$. Usually a line in the image is described by an equation of the form $ax + by + c = 0$, which we can also write as $(a,b,c)\cdot(x,y,1) = 0$. This means, that the vector (a,b,c) is perpendicular to any vector $(x,y,1)$ representing the points on the line, assuming that the image is at distance one from the center O. Thus, such a line can be represented by the vector (a,b,c), and we normalize it to have a unit z-component as $\boldsymbol{\ell} = (\frac{a}{c}, \frac{b}{c}, 1)$. Geometrically, vector $\boldsymbol{\ell}$ is perpendicular to the plane through O and L. Similarly, vector $\tilde{\boldsymbol{\ell}}$ is perpendicular to the plane through $\tilde{O}$ and L. The line L is perpendicular to vector $\boldsymbol{\ell}$, and it is also perpendicular to vector $\tilde{\boldsymbol{\ell}}$. Thus the cross-product vector $\boldsymbol{\ell} \times \tilde{\boldsymbol{\ell}}$ is parallel to the line L. Since the surface normal $\boldsymbol{N}$ is perpendicular to the patch, which includes the line L, we obtain

$$(\boldsymbol{\ell} \times \tilde{\boldsymbol{\ell}}) \cdot \boldsymbol{N} = 0 \tag{6}$$

where "$\cdot$" denotes the scalar product. This is the linear equation that we use to estimate $\boldsymbol{N}$.

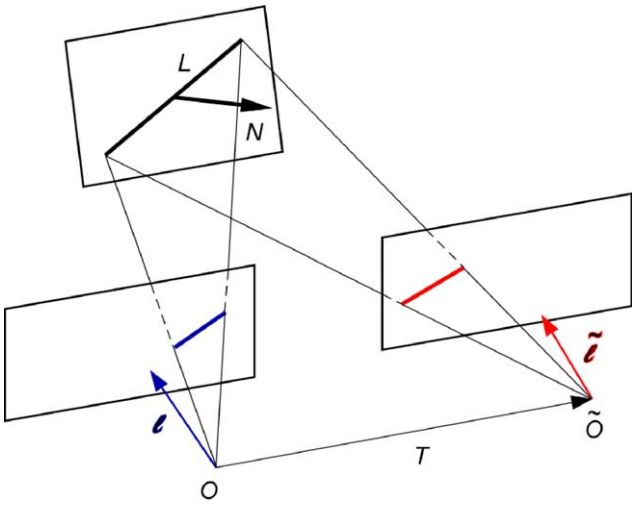

Fig. 2. Orientation disparity constraint in stereo. A line L in space is projected on the two views as $\boldsymbol{\ell}$ and $\tilde{\boldsymbol{\ell}}$. The representation for $\boldsymbol{\ell}$ and $\tilde{\boldsymbol{\ell}}$ are the vectors normal to the planes defined by the line in space and each of the centers O and $\tilde{O}$. Since $\boldsymbol{\ell}$ and $\tilde{\boldsymbol{\ell}}$ are both perpendicular to the line L, it follows that $\boldsymbol{\ell}\times\tilde{\boldsymbol{\ell}}$ is parallel the line L. $\boldsymbol{N}$ is normal to the plane containing L. Thus we have $(\boldsymbol{\ell}\times\tilde{\boldsymbol{\ell}})\cdot\boldsymbol{N}=0$.

2.1. Analysis of Eq. (6)

To analyze the bias and predict parametric influences we need to relate the image lines to the line in 3D. To facilitate the analysis in the following discussion, we present 3D lines by their Plücker coordinates. A line L in 3D space, which is a four-dimensional object, is then represented by the two vectors $(\boldsymbol{L}_{\mathrm{d}},\boldsymbol{L}_{\mathrm{m}})$. $\boldsymbol{L}_{\mathrm{d}}$ is a unit vector parallel to the line L in space, and thus it denotes its orientation. $\boldsymbol{L}_{\mathrm{m}}$, which is called the moment of the line, is a vector perpendicular to the plane through L and the origin O with value the distance of L from O. $\boldsymbol{L}_{\mathrm{m}}$ is parallel to $\boldsymbol{\ell}$ and perpendicular to $\boldsymbol{L}_{\mathrm{d}}$ (Fig. 3). Then the line coordinates in the two views are related as (see Appendix A)

$$\widetilde{\boldsymbol{L}_{\mathrm{d}}}=\boldsymbol{L}_{\mathrm{d}}$$
$$\widetilde{\boldsymbol{L}_{\mathrm{m}}}=\boldsymbol{L}_{\mathrm{m}}+\boldsymbol{T}\times\boldsymbol{L}_{\mathrm{d}}.$$

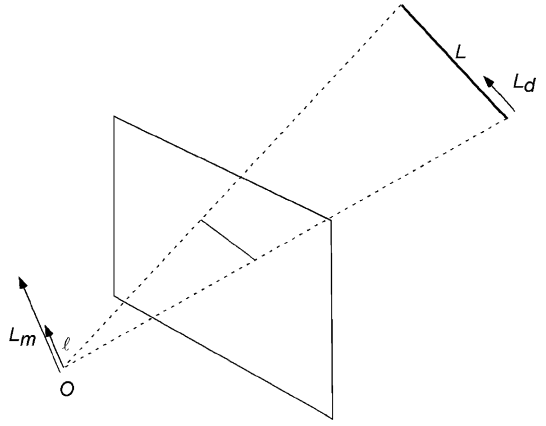

Fig. 3. Plücker representation of a line L as $(\boldsymbol{L}_{\mathrm{d}},\boldsymbol{L}_{\mathrm{m}})$ and its image $\boldsymbol{\ell}$.

Thus,

$$\boldsymbol{\ell}\times\tilde{\boldsymbol{\ell}}=\frac{\boldsymbol{L}_{\mathrm{m}}}{\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}}}\times\frac{\widetilde{\boldsymbol{L}_{\mathrm{m}}}}{\hat{z}\cdot\widetilde{\boldsymbol{L}_{\mathrm{m}}}}=\frac{\boldsymbol{L}_{\mathrm{m}}}{\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}}}\times\frac{(\boldsymbol{L}_{\mathrm{m}}+\boldsymbol{T}\times\boldsymbol{L}_{\mathrm{d}})}{\hat{z}\cdot\widetilde{\boldsymbol{L}_{\mathrm{m}}}}$$
$$=-\frac{(\boldsymbol{L}_{\mathrm{m}}\cdot\boldsymbol{T})\boldsymbol{L}_{\mathrm{d}}}{(\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}})(\hat{z}\cdot\widetilde{\boldsymbol{L}_{\mathrm{m}}})}.$$

For comparing the different configurations, we will simply use the projective relation (with $\sim$ denoting equality up to a scale factor)

$$\boldsymbol{\ell}\times\tilde{\boldsymbol{\ell}}\sim(\boldsymbol{T}\cdot\boldsymbol{L}_{\mathrm{m}})\boldsymbol{L}_{\mathrm{d}}. \tag{7}$$

3. Shape from stereo

Let us write the estimation Eq. (6) as

$$\boldsymbol{e}\cdot\boldsymbol{N}=0,$$

with $\boldsymbol{e}=\boldsymbol{\ell}\times\tilde{\boldsymbol{\ell}}$. Let $\boldsymbol{N}=(N_1,N_2,1)$ then be the surface normal, and let $\boldsymbol{\ell}_i=\{(a_i,b_i,1)\}$ denote the lines in the left image and $\tilde{\boldsymbol{\ell}}_i=\{(\tilde{a}_i,\tilde{b}_i,1)\}$ the corresponding lines in the right image. Substituting these coordinates into Eq. (6), we obtain for every observed line an equation in the two parameters (N_1,N_2) of the form

$$(e_{1_i},e_{2_i})\cdot(N_1,N_2)=-e_{3_i}, \tag{8}$$

where

$$\begin{cases} e_{1_i}=b_i-\tilde{b}_i,\\ e_{2_i}=-a_i+\tilde{a}_i,\\ e_{3_i}=a_i\tilde{b}_i-\tilde{a}_i b_i. \end{cases}$$

The line measurements (a_i,b_i) are always corrupted by noise in practice. Let the noise $\delta a_i=a_i-a_i'$, $\delta b_i=b_i-b_i'$ and $\delta\tilde{a}_i=\tilde{a}_i-\tilde{a}_i'$, $\delta\tilde{b}_i=\tilde{b}_i-\tilde{b}_i'$ be independent random variables with zero mean and covariance δ^2. Thus e_{k_i}, $k=1,2,3$ are also corrupted by noise. Then we have

$$(e_{1_i}'+\delta e_{1_i})N_1+(e_{2_i}'+\delta e_{2_i})N_2=-(e_{3_i}'+\delta e_{3_i}). \tag{9}$$

Let E, E' and δE denote the $n\times 2$ matrices incorporating the n measurements e_{1_i} and e_{2_i} and G, G' and δG denote the $n\times 1$ matrices incorporating the e_{3_i}. Then the estimation of $\boldsymbol{x}=(N_1,N_2)$ is obtained by solving the equation

$$\begin{aligned} &E\boldsymbol{x}=G \quad \text{or}\\ &(E'+\delta E)\boldsymbol{x}=G'+\delta G. \end{aligned} \tag{10}$$

Assuming that the errors are much smaller then the real values, we develop the LS solution of $\boldsymbol{x}$ in a second order Taylor expansion and obtain as an approximation for the estimate of $\boldsymbol{x}$ (see Appendix B):

$$E(\boldsymbol{x})=\boldsymbol{x}'-2n\delta^2M'^{-1}\boldsymbol{x}'$$
$$=(I-2n\delta^2M'^{-1})\boldsymbol{x}' \quad \text{with} \quad M'=E'^{\mathrm{T}}E'. \tag{11}$$

Since M' is a positive definite matrix, so is M'^{-1}. Considering the perturbation $2n\delta^2M'^{-1}$ being small, we have $||I-2n\delta^2M'^{-1}||^2<1$. Then we conclude that

$$||E(\boldsymbol{x})||=||(I-2n\delta^2M'x^{-1}\boldsymbol{x}')||<||I\boldsymbol{x}'||=||\boldsymbol{x}'|| \tag{12}$$

i.e., generally vector $\boldsymbol{x}$ is underestimated, and the degree of underestimation highly depends on the structure of matrix M'.

3.1. The effects on slant

The slant σ is the angle between the surface normal and the negative z-axis (0° slant corresponds to a plane parallel to the image plane, 90° to a plane that contains the optical axis) and the tilt τ is the angle between the direction of the projection of the surface normal onto the image plane and the x-axis (see Fig. 4). Using these coordinates $\boldsymbol{N} = (\cos\tau\tan\sigma, \sin\tau\tan\sigma, 1)$.

We know from the previous section (Eq. (12)) that $\|\boldsymbol{x}\|$ is underestimated. Since $\sigma = \cos^{-1}(1 + \|\boldsymbol{x}\|)^{-\frac{1}{2}}$ is a strictly increasing function of $\|\boldsymbol{x}\|$, by linear approximation, the slant σ is also underestimated. That is,

$$E(\sigma) < \sigma' \tag{13}$$

i.e., the expected value of the estimated slant is smaller than the actual value.

The degree of underestimation can be found by analyzing matrix M', or more specifically, the inverse of matrix M'. The inverse of a matrix can be written as its adjoint over its determinant, i.e., $M'^{-1} = \frac{\mathrm{adj}(M')}{\det(M)}$, which shows that large bias results from a small determinant and an $\boldsymbol{x}'$ close to the smaller eigenvalue of M'.

3.2. Anisotropy in the perception of stereoscopic slant

An interesting phenomenon in stereoscopic vision is the anisotropy in the perception of slanted (or tilted) planes. A surface slanted about the horizontal axis is estimated much easier and more accurately than a surface slanted about the vertical axis (Cagnello & Rogers, 1993; Mitchison & McKee, 1990; Ryan & Gillam, 1994). In both cases there is an underestimation of slant, but it is much larger for slant about the vertical. Cagnello and Rogers (1993) argued that this effect is due to orientation disparity, which generally (assuming the texture lines to be mostly vertical and horizontal) is smaller for surfaces slanting about the vertical. However, as shown in (Mitchison & McKee, 1990), the effect also exists, even though in weaker form, when the texture is made up of lines oriented at 45°. For such a configuration the orientation disparity in the two differently slanted planes should be the same. From this result, the authors argue that orientation disparity can not be the cause. We now show that bias in orientation disparity can account for this anisotropy.

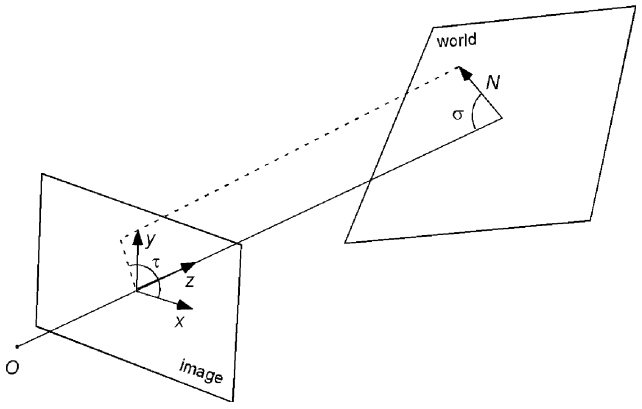

Fig. 4. The slant of a surface is the angle σ between the negative z-axis and the surface normal, the tilt τ is the angle between the the projection of the surface normal onto the image plane and the x-axis.

3.3. Analysis of stereoscopic slant

Consider a fronto-parallel plane which is textured with two sets of orthogonal line segments of orientation θ and $\frac{\pi}{2} + \theta$ (see Fig. 5a). When we slant this plane with angle σ about the vertical axis (Fig. 5b) the corresponding surface normal $\boldsymbol{N}_\mathrm{v} = (\boldsymbol{x}_\mathrm{v}, 1) = (\tan\sigma, 0, 1)$. When we slant the fronto-parallel plane about the horizontal axis (Fig. 5c), the surface normal is $\boldsymbol{N}_\mathrm{h} = (\boldsymbol{x}_\mathrm{h}, 1) = (0, \tan\sigma, 1)$. Substituting the components of $\boldsymbol{e}$ into Eq. (11) and relating $\boldsymbol{x}$ to the slant σ we obtain an approximation of the estimated slant, which, however, is valid only for smaller slants (till about 45°). The expected slant of the vertical and horizontal tilted plane, σ_v and σ_h amounts to (see Appendix C):

$$E(\sigma_\mathrm{v}) = \sigma - \delta^2 \frac{\sin 2\sigma}{2} \frac{nE(\sum e'^2_{2_\mathrm{v}})}{E(\det(M'_\mathrm{v}))} = \sigma - \delta^2 \frac{\sin 2\sigma}{2} C_\mathrm{v} \tag{14}$$

$$E(\sigma_\mathrm{h}) = \sigma - \delta^2 \frac{n \sin 2\sigma}{2} \frac{E(\sum e'^2_{1_\mathrm{h}})}{E(\det(M'_\mathrm{h}))} = \sigma - \delta^2 \frac{\sin 2\sigma}{2} C_\mathrm{h}. \tag{15}$$

In the equations above the effects of $M'^{-1}\boldsymbol{x}' = \frac{\mathrm{adj}(M')\boldsymbol{x}'}{\det(M')}$ show up as the terms C_v and C_h. The terms $E(\sum e'^2_{2_\mathrm{v}})$ and $E(\sum e'^1_{2_\mathrm{h}})$ originate from $\mathrm{adj}(M')\boldsymbol{x}'$, and the terms $E(\det(M'_\mathrm{v}))$ and $E(\det(M'_\mathrm{h}))$ are the determinants of M'_v and M'_h. C_v and C_h determine the degree of underestimation. The larger they are the more the underestimation will be, and their ratio is a measure of the relative error.

Denoting the displacement between the cameras as t in Appendix C we derive:

$$C_\mathrm{v} = \frac{1}{t^2} \frac{(\sin^4\theta + \cos^4\theta)}{\cos^4\sigma(\sin^2\theta\cos^2\theta)} \tag{16}$$

$$C_\mathrm{h} = \frac{2}{t^2\cos^2\sigma}, \tag{17}$$

and thus their ratio amounts to:

$$\frac{C_\mathrm{v}}{C_\mathrm{h}} = \frac{1}{\cos^2\sigma} \frac{\tan^2\theta + \cot^2\theta}{2} > 1. \tag{18}$$

Let us get an intuition for the above equations. The differences in the bias can be understood from the relation

$$\boldsymbol{e} \sim (\boldsymbol{T} \cdot \boldsymbol{L}_\mathrm{m})\boldsymbol{L}_\mathrm{d}.$$

The two parameters involved in Eq. (18) are σ, the slant (or rotation) of the surface and θ, which defines the orientation

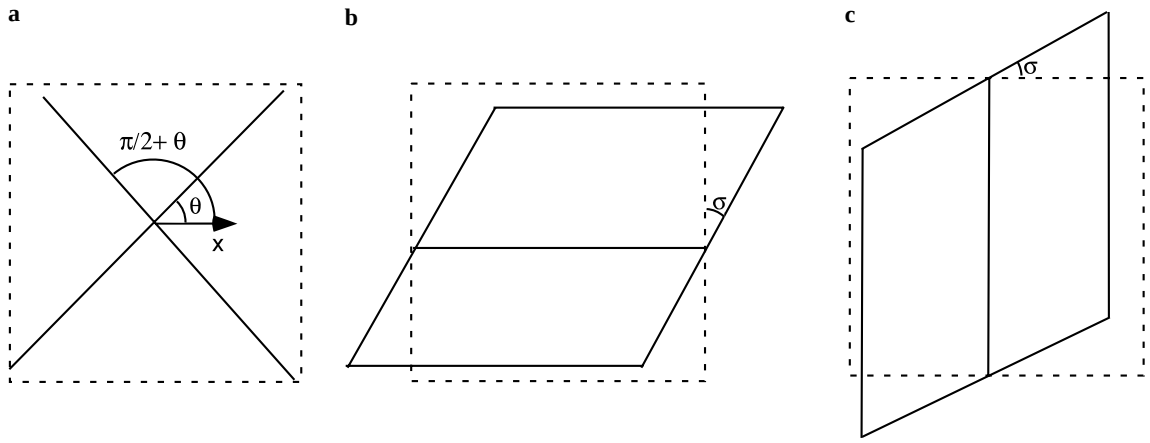

Fig. 5. (a) In the fronto-parallel setting the orientation of the line elements on the plane is θ. The plane is rotated (slanted) by angle σ about the (b) vertical axis and (c) about the horizontal axis.

of the image lines (Fig. 5). The effect of σ enters through the determinants. A slant about the vertical axis shortens the x component of $\boldsymbol{L}_{\mathrm{m}}$ and $\boldsymbol{L}_{\mathrm{d}}$ by a multiplicative factor of $\cos(\sigma)$. A slant about the horizontal axis effects in the same way the y component of $\boldsymbol{L}_{\mathrm{m}}$ and $\boldsymbol{L}_{\mathrm{d}}$. Since the translation is parallel to the x-axis, i.e., $\boldsymbol{T} = (t, 0, 0)$ and the product $(\boldsymbol{T} \cdot \boldsymbol{L}_{\mathrm{m}})$ has only the x-component of $\boldsymbol{L}_{\mathrm{m}}$, there is more shortening effect in the determinant of the vertically slanted plane, and the ratio of $\frac{C_{\mathrm{v}}}{C_{\mathrm{h}}}$ is $\frac{1}{\cos^2\sigma}$.

The effect of θ enters due to $\mathrm{adj}(M')\boldsymbol{x}'$: the components involving θ are the same for both matrix M_{v} and M_{h}. (Only the components due to slant are different). Because of the product $\boldsymbol{T} \cdot \boldsymbol{L}_{\mathrm{m}}$, for both matrices there is a smaller θ component parallel to the x-axis than parallel to the y-axis (smaller $e_1'^2$ than $e_2'^2$. In other words, the smaller eigenvalue component is closer to the x axis than to the y axis). $\boldsymbol{x}_{\mathrm{v}} = (\tan\sigma, 0)$ is parallel to the x-component and $\boldsymbol{x}_{\mathrm{h}} = (0, \tan\sigma)$ is parallel to the y-component, and thus there is larger bias for vertical slant. The ratio of the two terms amounts to $\frac{\tan^2\theta + \cot^2\theta}{2}$.

Now, using Eq. (18) we can predict the perception of slanted planes (Cagnello & Rogers, 1993; Mitchison & McKee, 1990). For a pattern with vertical and horizontal line segments ($\theta \approx 90°$) the ratio $\frac{C_{\mathrm{v}}}{C_{\mathrm{h}}} \gg 1$. This predicts the estimation for the plane slanted about the vertical to be significantly worse than for the plane slanted about the horizontal. When the line segments are at 45° the bias is still larger for the slant about the vertical since $\frac{C_{\mathrm{v}}}{C_{\mathrm{h}}} = \frac{1}{\cos^2\sigma} > 1$. Thus the perception for the slant is still predicted to be more erroneous, but with much less anisotropy.

3.4. Experiments and predictions

We created stereograms on a computer display as follows: a plane textured with lines in two orthogonal directions was slanted in space about the vertical and horizontal axes with the slant in the range of 0° to 55°, and its images were created by projection. In order to keep the number of lines constant (between four and five lines in one direction) we zoomed in. We tested two line orientations, a pair with 45° and 135° and a pair with 30° and 120°. Three observers were shown first a random sequence of horizontally, then a sequence of vertically slanted planes and asked to adjust a cardboard on the desk next to the screen to denote the perceived slant. Figs. 6 and 7 show

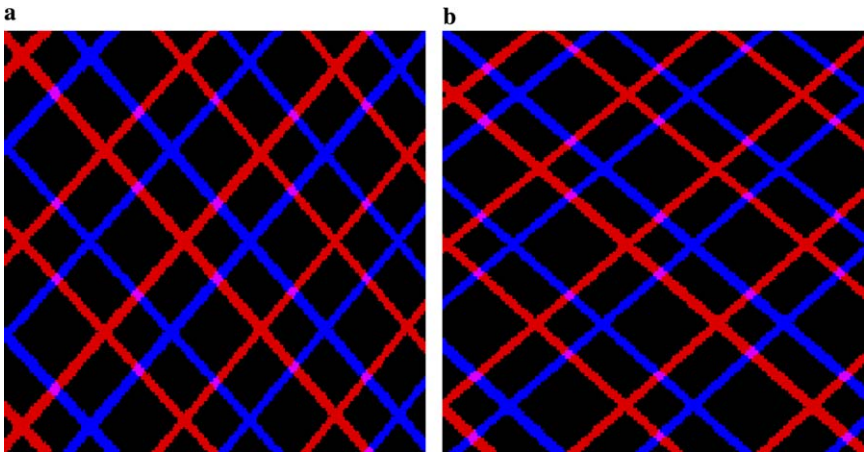

Fig. 6. Stereoscopic images of a plane slanted by 30° about the vertical (a) and horizontal (b) axes. View the images with red-blue glasses with the red on the right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

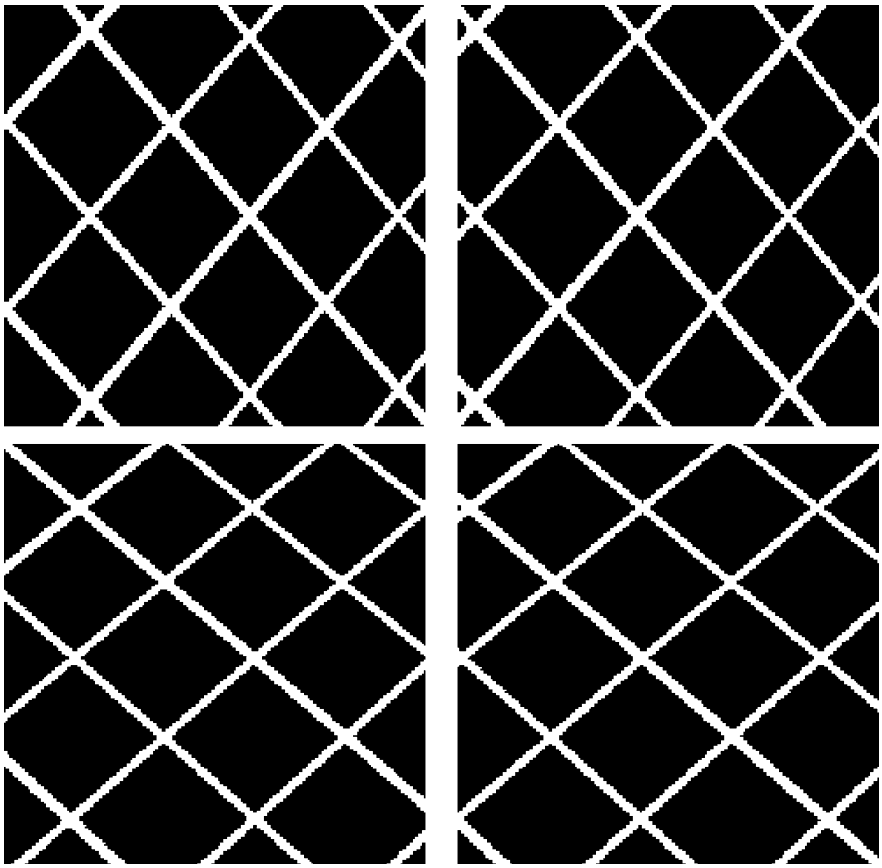

Fig. 7. Free fusion versions of Fig. 6 for uncrossed viewing.

the stereograms for 30° slant and the lines oriented at 45° (and 135°).

Fig. 8 plots the measurement along with our predictions. The data points shown are the mean values over all all trials and all three subjects. The standard deviation was about 20%. Mostly one of our subjects was much more accurate in his estimates than the the other two. Fig. 9 shows the data of the individual subjects for one configuration.

As can be seen from Fig. 8, the predictions model the data very well within this range of angles. We should note that we do not attempt to model larger slants. First, at larger angles texture is very compressed, and thus additional information about the texture distortion and the vanishing lines becomes important. We think that this information has more influence on the perception than the bias. Second, our equations are approximations which are valid only for smaller angles. In particular, the Taylor expansion in Eq. (C.4) is not justified for larger angles. Our measurements indicate the same general behavior as found in (Cagnello & Rogers, 1993; Mitchison & McKee, 1990). To allow for comparison we show data of (Mitchison & McKee,

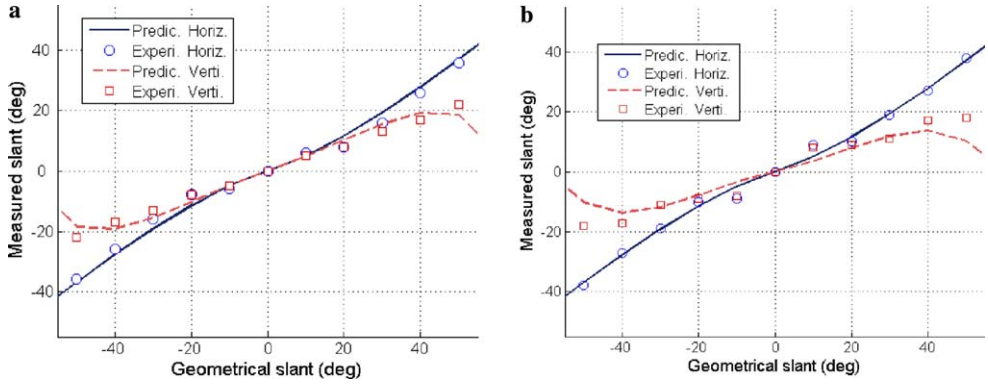

Fig. 8. Our experiments: Predictions and measurements for textures oriented at (a) 45° and (b) 30°.

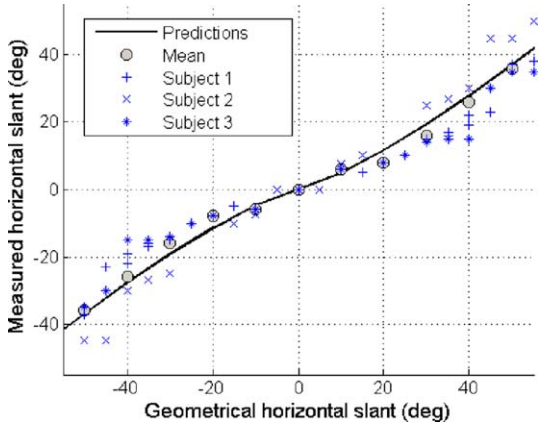

Fig. 9. Measurements from the three subjects for a plane slanted about the horizontal axis and textured with lines of 45° orientation.

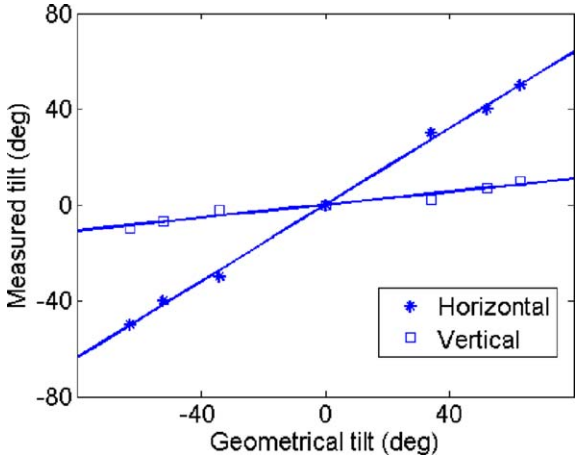

Fig. 10. Slant perceived by one of the subjects in (Mitchison & McKee, 1990) for 45° oriented texture lines.

1990) for one of their subjects. Referring to Fig. 10, one can see that Mitchison and McKee found a worse estimation for vertically slanted planes, and they found a nearly linear measurement function.

4. Shape from motion

In the case of differential motion, when the camera (or eye) moves with instantaneous translational velocity $\boldsymbol{T}$ and rotational velocity $\boldsymbol{\omega}$, Eq. (6) takes a similar form. If $\boldsymbol{N}$ is the normal of a plane containing a line with image $\boldsymbol{\ell}$ and temporal derivative $\dot{\boldsymbol{\ell}}$ the estimation Eq. (6) becomes (see Appendix A)

$$(\boldsymbol{\ell} \times (\dot{\boldsymbol{\ell}} - \omega \times \boldsymbol{\ell})) \cdot \boldsymbol{N} = 0 \quad \text{or}$$
$$\boldsymbol{e} \cdot \boldsymbol{N} = 0 \tag{19}$$

with $\boldsymbol{e} = \boldsymbol{\ell} \times (\dot{\boldsymbol{\ell}} - \omega \times \boldsymbol{\ell})$ (see Appendix A).

Let $\{\boldsymbol{\ell}_i = (a_i, b_i, 1)\}$ denote the lines on the plane, and $\{\dot{\boldsymbol{\ell}}_i = (\dot{a}_i, \dot{b}_i, 0)\}$ denote the motion parameters of the lines $\boldsymbol{\ell}_i$. Then in the prime equations

$$(e_{1_i}, e_{2_i}) \cdot (N_1, N_2) = -e_{3_i}, \tag{20}$$

the parameters are

$$\begin{cases} e_{1_i} = -\dot{b}_i + (-(1 + b_i^2)\omega_1 + a_i b_i \omega_2 + a_i \omega_3), \\ e_{2_i} = \dot{a}_i + (a_i b_i \omega_1 - (1 + a_i^2)\omega_2 + b_i \omega_3), \\ e_{3_i} = -(\dot{a}_i b_i - \dot{b}_i a_i) + (a_i \omega_1 + b_i \omega_2 - (a_i^2 + b_i^2)\omega_3). \end{cases}$$

There is noise in the measurements of the line locations and the measurements of the line movement. Let the error random variables be $\delta a_i = a_i - a_{i'}$ and $\delta b_i = b_i - b_{i'}$ with expected value 0 and variance δ_1^2, and $\delta \dot{a}_i = \dot{a}_i - \dot{a}'_i$ and $\delta \dot{b}_i = \dot{b}_i - \dot{b}'_i$ with expected value 0 and variance δ_2^2. Then from the second order Taylor expansion of the LS solution we obtain the expected value of $\boldsymbol{x} = (N_1, N_2)$ (see Appendix B) as

$$E(\boldsymbol{x}) = \boldsymbol{x}' - M'^{-1}(\delta_2^2 D' + \delta_1^2 F')\boldsymbol{x}' - M'^{-1}\delta_1^2 H' \tag{21}$$

where

$$D' = \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix}, \quad H' = \omega_3 \sum_i^n \begin{pmatrix} \omega_1(6b_i'^2 + c'_i + 3) \\ \omega_2(6a_i'^2 + c'_i + 3) \end{pmatrix},$$
$$F' = \sum_i^n \begin{pmatrix} 6b_i'^2\omega_1^2 + c'_i\omega_2^2 + \omega_3^2 + 2\omega_1^2 & 2c'_i\omega_1\omega_2 + 2\omega_1\omega_2 \\ 2c'_i\omega_1\omega_2 + 2\omega_1\omega_2 & c'_i\omega_1^2 + 6a_i'^2\omega_2^2 + \omega_3^2 + 2\omega_2^2 \end{pmatrix}$$

with $c'_i = a_i'^2 + b_i'^2$.

For the case when rotation around the z-axis can be ignored (i.e, $\omega_3 = 0$) Eq. (21) simplifies to

$$E(\boldsymbol{x}) = (I - M'^{-1}(\delta_2^2 D' + \delta_1^2 F'))\boldsymbol{x}' = (I - \delta_A)\boldsymbol{x}'. \tag{22}$$

D' and F' are positive definite matrices and the perturbations δ_1 and δ_2 are small. Thus δ_A is also a positive definite matrix, and by the same arguments as in the case of stereo, the slant can shown to be underestimated.

To show the degree of underestimation, next we will analyze the determinant of matrix M'; the smaller the determinant the larger the underestimation. The velocity of rotation also contributes to the magnitude of the bias as can be seen from matrix F'; larger velocity more bias.

4.1. Predictions and illusory display

To say more about the dependence of slant estimation on the texture distribution we use the relation (Eq. (A.7))

$$\boldsymbol{e} \sim (\boldsymbol{t} \cdot \boldsymbol{\ell})\boldsymbol{L}_{\mathrm{d}}.$$

Let us consider a slanted plane with a texture of two major directional components. Let the directional components be $\boldsymbol{L}_{d_1} = (\cos\tau_1 \sin\sigma_1, \sin\tau_1 \sin\sigma_1, cos\sigma_1)$ and $\boldsymbol{L}_{d_2} = (\cos\tau_2 \sin\sigma_2, \sin\tau_2 \sin\sigma_2, \cos\sigma_2)$. That is σ_1 and σ_2 are the angles between the texture lines on the world and the negative z-axis, and τ_1 and τ_2 are the angles between the projections of the texture lines on the image plane and the x-axis (see Fig. 11). The determinant $det(M)$ of M amounts to (Appendix D)

$$\det(M) = \sum (\boldsymbol{t} \cdot \boldsymbol{\ell}_{1_i})^2 \sum (\boldsymbol{t} \cdot \boldsymbol{\ell}_{2_i})^2 (\sin\sigma_1 \sin\sigma_2 \sin(\tau_1 - \tau_2))^2. \tag{23}$$

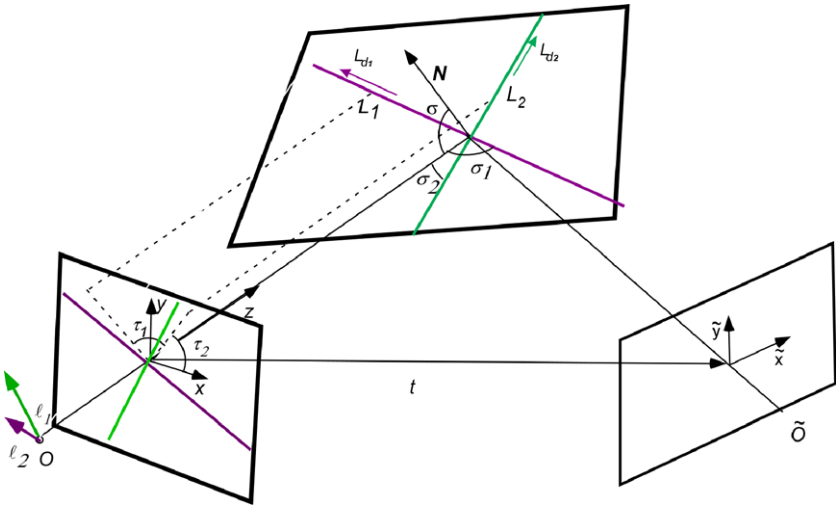

Fig. 11. A plane with a texture of two orientations $\boldsymbol{L}_{d_1}$ and $\boldsymbol{L}_{d_2}$ is imaged under motion. σ_1 and σ_2 are the angles between the texture lines on the world and the negative z-axis, and τ_1 and τ_2 are the angles between the projections of the texture lines on the image plane and the x-axis.

In this equation the term $\sum(t \cdot \boldsymbol{\ell}_{1_i})^2 \sum(t \cdot \boldsymbol{\ell}_{2_i})^2$ is due to the products $\boldsymbol{T} \cdot \boldsymbol{\ell}_1$ and $\boldsymbol{T} \cdot \boldsymbol{\ell}_2$ in the $\boldsymbol{e}$'s and the term (sin σ_1 sin σ_2 sin $(\tau_1 - \tau_2)$)^2 is due to $\boldsymbol{L}_{d_1}$ and $\boldsymbol{L}_{d_2}$.

Using our model we can predict the findings from experiments in the literature. In (Todd & Perotti, 1999) it has been observed that an increase in the slant of a rotating surface causes increased underestimation of the slant. This can be understood from the change of $\boldsymbol{L}_{\mathrm{d}}$ in Eq. (23). Intuitively, larger slant causes an increase in the z- and decrease in the x- and y-component of $\boldsymbol{L}_{\mathrm{d}}$ and thus a smaller det(M). By our formula in Eq. (23) this is manifested in the factor sin (σ_1) sin (σ_2), where σ_1 and σ_2 are the the angles between the directions of the line in space and the negative z-axis. Unless, they are 0°, these values decrease with an increase of the slant of the plane, and this leads to a smaller det(M). Hence, we get a larger error towards underestimation of the slant.

To demonstrate the predictive power of the model we created two illusory displays. In the first one, the scene consists of a plane with two textures, one in the upper half, the other in the lower half. Fig. 12a shows the plane when it is parallel to the screen. The texture in the upper part consists of two line clusters with slope 8° and 98°. The lower part has two lines clusters with slope 45° and 135°. A video was created for the camera orbiting the sphere along a great circle in the yz-plane as shown in Fig. 12b—that is, the camera translates and rotates such that it keeps fixating at the center. At the beginning of the motion, the slant of the plane with respect to the camera is 15°, at the end it is 45°. The image sequence can be seen in (Fermüller, 2003) (www.cfar.umd.edu/users/fer/optical/Newsite/shape/video.avi). As can be experienced, it creates the perception of the plane being segmented into two parts, with the upper part having a much smaller slant. This is predicted by the biases in the different textures. For the upper texture the bias is much larger, thus producing larger underestimation of the slant, and the underestimation gets worse as the slant increases. The difference in the values of the bias can be understood from the difference in the values for $(\boldsymbol{T} \cdot \boldsymbol{\ell})$ (the term $\sum(\boldsymbol{t} \cdot \boldsymbol{\ell}_{1_i})^2 \sum(t \cdot \boldsymbol{\ell}_{2_i})^2$ in Eq. (23)). $\boldsymbol{T}$ is nearly parallel to the y-axis, and thus $\boldsymbol{T} \cdot \boldsymbol{\ell}$ is close to zero for vertical texture lines, making the determinant very small. In a second display (www.cfar.umd.edu/users/fer/optical/Newsite/shape/

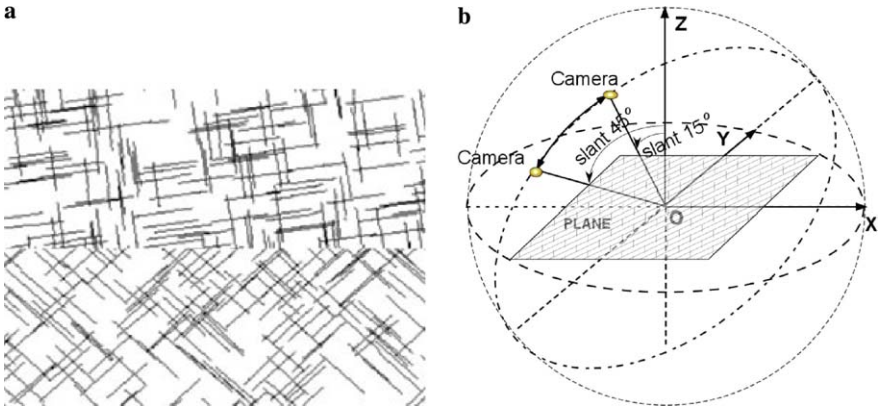

Fig. 12. (a) The plane in view. (b) Scene geometry in the shape from motion demonstration.

Table 1
Influence of viewing geometry and texture on the size of the bias

Large bias	Small bias
Two view geometry (motion and stereo)	
Small $\|\boldsymbol{T}\|$	Large $\|\boldsymbol{T}\|$
Large $\|R\|$	Small $\|R\|$
Large slant	Small slant
Small $(\boldsymbol{T}\cdot\boldsymbol{\ell}_1)$ $(\boldsymbol{T}\cdot\boldsymbol{\ell}_2)$	Large $(\boldsymbol{T}\cdot\boldsymbol{\ell}_1)$ $(\boldsymbol{T}\cdot\boldsymbol{\ell}_2)$
Stereo	
Horizontal and vertical texture lines	Texture lines at 45°
Surface slanted about vertical axis	Surface slanted about horizontal axis

video4.avi) the plane is divided into multiple segments with two alternating textures. In every other segment there is large bias, and this gives rise to the perception of the plane folding as if it were a staircase.

Before going on, let us note that the underestimation of slant is not due to the particular constraints we employed. We may instead compute structure from normal flow (the component of optical flow perpendicular to edges) by fitting a plane to the flow data. The surface normal vector obtained this way (from the structure estimates) will have a qualitatively similar behavior.

5. Summary of the parametric influences on the bias

Table 1 summarizes the findings of the last two sections. It displays the effects of the viewing geometry parameters and the texture on the bias for general 3D motion between the cameras and for stereo in particular. $\boldsymbol{T}$ and R denote the translation and rotation, and $|\boldsymbol{T}|$ and $|R|$ their absolute value. $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2$ are are two orthogonal image lines expressed in projective coordinates. Notice that any texture may be represented by two orthogonal directions corresponding to the eigenvectors of the matrix M'.

6. Why is estimation so difficult

The statistical model used to describe the data in our equation $A\boldsymbol{x} = \boldsymbol{b}$ is the errors-in-variable model, which is defined as

Definition 1 (*Errors-in-variable model*).

$$\boldsymbol{b}' = A'\boldsymbol{x} + \boldsymbol{\epsilon},$$
$$\boldsymbol{b} = \boldsymbol{b}' + \boldsymbol{\delta_b},$$
$$A = A' + \delta_A,$$

x' are the true but unknown parameters. $A = (A_{i,j})$ and $\boldsymbol{b} = (b_i)$ are observations of the true but unknown values A' and $\boldsymbol{b}'$. $\delta_{A_{i,j}}, \delta b_j$ are the measurement errors and $\boldsymbol{\epsilon}$ is the modeling error. This is the error due to our model assumptions.

It is well known that for this model least squares (LS) estimation is biased. The main reason is that it does not consider errors in the explanatory variables, that is the δ_A. Let us then investigate the theoretical question. Are there better ways to estimate? Are there better statistical estimators that do not suffer from bias? The answer is that in general we cannot avoid bias. We may be able to reduce the bias, if there is enough data available to obtain reasonable error statistics. In this case the bias will still be of the same form, only smaller. Simply weighting the data to make it uniform is statistically not justified as it would increase the variance in the direction of lesser data significantly. Choosing a nonlinear estimator (such as Total Least Squares), which would give a different form of bias, does not appear to give better results because of the noise to be expected. A short discussion summarizing the main arguments is given next. Technical details are presented elsewhere (Fermüller, Shulman, & Aloimonos, 2001; Hui & Fermüller, 2003).

The so-called corrected least squares (CLS) estimator is the classical technique to address bias. If the statistics of the noise, that is the covariance matrix of δ_A, is known, an asymptotically unbiased linear estimator could be constructed. The problem is that for small amounts of data, accurate estimation of the variance of the noise is a very difficult problem, and has high variance itself, and this leads to higher variance for the CLS estimation. It is well known that the scale of the error variance is difficult to obtain in practice.

In the computational vision literature more attention has been given to the nonlinear technique of total least square (TLS), which deals with the errors in A and $\boldsymbol{b}$ symmetrically and only requires the ratio of the error variances. If all the errors $\delta_{A_{i,j}}$ and δb_j are identical and independent, or their ratio can be obtained, then TLS estimation is asymptotically unbiased. Estimation of the ratio of errors is not easy either. However, the main problem for TLS is modeling error (or also called system error Fuller (1987)). Theoretically one can use multiple tests to obtain the measurement errors, like re-measuring or re-sampling; but unless the exact parameters of the model are known, one cannot test for the modeling error. The noise we expect is actually much more complicated than simple i.i.d. additive noise. It is correlated, and this would cause further problems for TLS, causing convergence problems for the corresponding nonlinear nonconvex objective function to be minimized (Ng & Solo, 2001). TLS is attractive in the sense that it has an obvious geometrical explanation, but it does not appear advantageous for the vision applications discussed. Its improvement over usual least squares in the statistical sense would be offset with more complicated error models or if mis-modeling the error.

We can classify the errors into two categories: measurement noise and modeling error. In the problem at hand the measurements are the line parameters $\{a_i, b_i\}$, and the image motion parameters of the lines $\{\dot{a}_i, \dot{b}_i\}$. We can expect measurement errors due to sensor noise which effects the measurements of image intensity $I(x, y, t)$. It seems reasonable to approximate the sensor noise as i.i.d. But we have to consider dependencies when the images

are smoothed. Other errors in measurement are due to bad fitting, when estimating the line parameters with edge detectors and discretization due to the edge detectors and difference operators computing the derivatives.

Modeling errors are due to erroneous assumptions. When computing the motion of lines, we assume that the image intensity is constant between frames. Significant errors occur at specular components. We use first order expansions when deriving velocities. Thus, errors are expected for large local velocities. Furthermore, the modeling of the scene as consisting of planar patches is an approximation to the actual surface of the scene.

Sensor noise may be considered i.i.d. and is easier to deal with (Lydia & Victor, 2001; Nagel, 1995). But other errors could be more significant, and they are more elaborate, making the statistics rather complicated. It is too difficult to estimate the statistics of the combined noise, which is necessary to apply the classical techniques.

There is another technique widely known in Economics which theoretically may be well suited for vision, the technique of instrumental variables (IV technique), which deals with the errors in the explanatory variables but does not require the error variance as a priori. This technique uses additional variables, the instrumental variables, which could be additional measurements of the explanatory variables. Let these variables be called W. If the errors in the measurements of the two methods can be treated as independent, an asymptotically unbiased estimator (Fuller, 1987) can be created, whose variance is close to the variance of the CLS estimator, by solving the replaced equation system

$$(W^{\mathrm{T}}A)\boldsymbol{x} = (W^{\mathrm{T}})\boldsymbol{b}, \tag{24}$$

with standard least squares estimation. But even, if the errors are not fully independent, but not completely related, the technique can help reduce the bias.

Possible ways to obtain instrumental variables are by taking multiple measurements of the explanatory variables, for example, by using multiple edge detections, fitting schemes, or difference operators.

Using color image sequences we could create even better instrumental variables. We may use one color channel as instrumental variables to another color channel. It is quite reasonable to assume that the sensor noise components in the different color channels are independent. The approximation errors in the image gradients would not be completely independent since there is a similarity in the structure of the color intensity functions. This means, we could not completely remove the bias from approximation error, but we could partially correct the bias caused by this error. We cannot correct the bias from the modeling error. But this is the advantage of this technique despite the presence of modeling error, it still can deal with the other errors.

We have implemented and tested a number of the classical techniques, including CLS, TLS and the IV technique and found that we could improve the solution only in a minor way (Hui & Fermüller, 2003). The best results were obtained with the IV technique.

7. Is our vision system doing the best?

Bias is only one component of estimation, the other is variance; and there is a trade-off between the two. Generally an estimator correcting for bias increases the variance while decreasing the bias. In statistics, the performance of an estimator is evaluated by a risk function. Usually the mean squared error (MSE) is used as performance criterion. It is the expected value of the square of the difference between the estimated and the true value. In the case of a scalar variable, if x' is used to denote the actual value, $\hat{x}$ to denote the estimate, and $E(\cdot)$ to denote the expected values, the MSE is defined as

$$\begin{aligned}\mathrm{MSE}(\hat{x}) &= E((\hat{x}-x')^2) = (E(\hat{x})-x')^2 + E(x'-E(\hat{x}))^2 \\ &= \mathrm{bias}^2(\hat{x}) + \mathrm{cov}(\hat{x}),\end{aligned} \tag{25}$$

that is, as the sum of the square of the bias (denoted as $\mathrm{bias}(\hat{x})$) and the variance (denoted as $\mathrm{cov}(\hat{x})$). In the case of $\boldsymbol{x}$ being a vector, Eq. (25) generalizes to

$$\mathrm{MSE}(\hat{\boldsymbol{x}}) = \mathrm{bias}^{\mathrm{T}}(\hat{\boldsymbol{x}})\mathrm{bias}(\hat{\boldsymbol{x}}) + \mathrm{cov}(\hat{\boldsymbol{x}}) \tag{26}$$

with $\mathrm{cov}(\hat{\boldsymbol{x}})$ denoting the trace of the covariance matrix of $\hat{\boldsymbol{x}}$.

Let us assume we know the variance of the error as well as the covariance of the LS estimate exactly. In Appendix E, we derive an expression for the best linear estimator. This would be a partial correction using CLS. How much to correct depends on the covariance of the LS estimate. The larger the covariance, the less the correction.

Thus, theoretically the best we could do is to partially correct. If there is a sufficient amount of data we should be able to somewhat correct the bias. A good choice for doing so, would be a conservative, that is slight correction using CLS or the IV method. Such a correction would lead to an estimation with bias qualitatively of the same form as LS but smaller in value.

What about the human vision system? We assume that it is doing the best it can. If it has enough data, it should be able to perform some correction. Two observations make us believe that it does.

Partially corrected estimation would explain why the illusory perception in many optical illusions weakens after extended viewing, in particular when subjects are asked to fixate (Helmholtz, 1962; Yo & Wilson, 1992). In these cases, we can assume that the noise parameters stay fixed, and the visual system can reasonably well estimate them.

We can draw conclusions by varying the covariance (of the estimator) in a pattern. Two patterns created by line segments of same orientation but different density give rise to the same bias, but different covariance. The smaller the covariance of the estimator, the more the correction and thus the less the bias should be. Thus, a pattern with higher density and smaller covariance should be estimated better. We tested different textures and found our perception to be consistent with the hypothesis. An example is shown in Fig. 13. Patterns of larger density and thus smaller covariance of the estimator appear to result in better estimation.

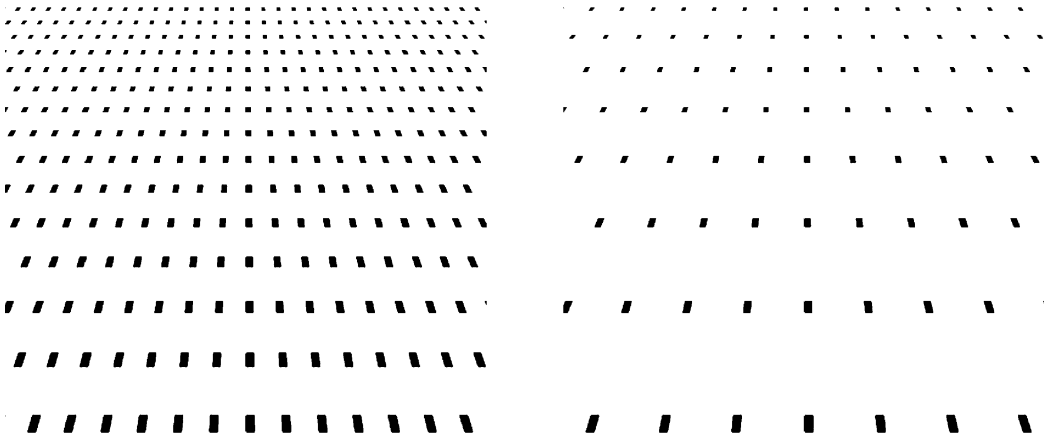

Fig. 13. Influence of texture density on estimation. The plane is slanted with an angle of 45°. The denser pattern appears to be estimated closer to the veridical than the sparser pattern.

8. Discussion: Structure in the feed-back loop

We have analyzed the effects of noise on the estimation of shape and found that there is bias. We showed that this bias predicts the underestimation of slant, which is known from computational and psychophysical experiments. Our analysis was based on LS estimation, but we showed that other estimators suffer from bias too, and for the most appropriate estimators the bias is of the same form as for LS estimation. The main reason for the inability to correct for the bias lies in the difficulty to obtain good estimates of the statistics of the data.

Vision scientists have long realized that a large part of the visual processes are carried out by feeding information from higher processing areas to lower processing areas. The Computer Vision literature creating algorithms, however, has not embraced this view yet. That is nobody will argue that recognition requires top-down process, but the problems of reconstruction discussed here have been studied in a pure feed-forward fashion. This brings us to the question: is there a need for computational feed-back? It is clear that the problems of model parameter estimation and segmentation are computationally antagonistic to each other. The idea was that it is possible to set up a large minimization which includes discontinuity localization and parameter estimation that will solve the problems. However, the approach has not been proven to be successful. It appears that the information one can extract from the signal without having knowledge of the scene model is not sufficient to perform good reconstruction.

The fact that we have to deal with bias makes a clear argument for feed-back processes. We know, in whatever way we estimate, we need a lot of data to do it well. But in order to use large amounts of data we need models underlying our data. For the reconstruction problems this means models of the scene structure or shape, and models of the texture, the shading and other cues. Thus we need feed-back from higher processes, which compute 3D geometry, to lower processes. After having obtained from first computations using very local data a preliminary segmentation and estimates of the viewing geometry, more global data (over larger spatio-temporal extent) should be used in a feed-back loop to re-segment and better estimate.

Acknowledgment

The support of this research by the National Science Foundation under Grant IIS-00-8-1365 is gratefully acknowledged.

Appendix A. Shape from lines in multiple views: The constraint

Consider the general stereo configuration of two cameras displaced by a rigid motion with translation $\boldsymbol{T}$ and rotation R. Let the scene be a textured plane with surface normal $\boldsymbol{N}$. The texture is described by the lines on the plane. A line L in 3D space is a four-dimensional object and can be elegantly described by Plücker coordinates. Let $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ be two points with unit distance and $\boldsymbol{P}$ any point on $L = (\boldsymbol{L}_\mathrm{d}, \boldsymbol{L}_\mathrm{m})$. Then

$$\begin{cases} \boldsymbol{L}_\mathrm{d} = \boldsymbol{P}_1 - \boldsymbol{P}_2; \\ \boldsymbol{L}_\mathrm{m} = \boldsymbol{P} \times \boldsymbol{L}_\mathrm{d} = \boldsymbol{P}_2 \times \boldsymbol{P}_1. \end{cases} \tag{A.1}$$

$\boldsymbol{L}_\mathrm{d}$ denotes the direction of the line in space, and $\boldsymbol{L}_\mathrm{m}$ its moment. Geometrically $\boldsymbol{L}_\mathrm{m}$ is a vector perpendicular to the plane through L and the coordinate center O with value the distance of L from O (Fig. 3). $\boldsymbol{L}_\mathrm{d}$ and $\boldsymbol{L}_\mathrm{m}$ are perpendicular, that is $\boldsymbol{L}_\mathrm{d} \cdot \boldsymbol{L}_\mathrm{m} = 0$. The projection ℓ of the 3D line L on the image is just $\boldsymbol{L}_\mathrm{m}$ normalized, i.e to have the third coordinate 1, it is $\ell = \frac{\boldsymbol{L}_\mathrm{m}}{\hat{z} \cdot \boldsymbol{L}_\mathrm{m}}$, where $\hat{z}$ is a unit vector parallel to the z-axis.

Since points in the two views are related as $\tilde{\boldsymbol{P}} = R\boldsymbol{P} + \boldsymbol{T}$, the line parameters in the two views are related as

$$\begin{cases} \widetilde{\boldsymbol{L}_\mathrm{d}} = \tilde{\boldsymbol{P}}_1 - \tilde{\boldsymbol{P}}_2 = R(\boldsymbol{P}_1 - \boldsymbol{P}_2) = R\boldsymbol{L}_\mathrm{d}; \\ \widetilde{\boldsymbol{L}_\mathrm{m}} = \tilde{\boldsymbol{P}}_2 \times \tilde{\boldsymbol{P}}_1 = (R\boldsymbol{P}_2 + \boldsymbol{T}) \times (R\boldsymbol{P}_1 + \boldsymbol{T}) = R\boldsymbol{L}_\mathrm{m} + \boldsymbol{T} \times R\boldsymbol{L}_\mathrm{d}. \end{cases} \tag{A.2}$$

Thus, the orientation of the line can be obtained as

$$\ell \times (R^{\mathrm{T}}\tilde{\ell}) = \frac{\boldsymbol{L}_{\mathrm{m}}}{\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}}} \times \frac{\boldsymbol{L}_{\mathrm{m}} + R^{\mathrm{T}}\boldsymbol{T}\times\boldsymbol{L}_{\mathrm{d}}}{\hat{z}\cdot\widetilde{\boldsymbol{L}_{\mathrm{m}}}} = \frac{-(\boldsymbol{L}_{\mathrm{m}}\cdot R^{\mathrm{T}}\boldsymbol{T})\boldsymbol{L}_{\mathrm{d}}}{(\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}})(\hat{z}\cdot\widetilde{\boldsymbol{L}_{\mathrm{m}}})} = \frac{-(\ell\cdot R^{\mathrm{T}}\boldsymbol{T})}{\hat{z}\cdot\widetilde{\boldsymbol{L}_{\mathrm{m}}}}\boldsymbol{L}_{\mathrm{d}}. \tag{A.3}$$

Since $\boldsymbol{L}_{\mathrm{d}}$ is perpendicular to the surface normal N we have that

$$(\ell \times R^{\mathrm{T}}\tilde{\ell})\cdot\boldsymbol{N} = 0. \tag{A.4}$$

In the case of differential motion, where the motion of a point in space has velocity $\dot{\boldsymbol{P}} = \boldsymbol{T} + \omega\times\boldsymbol{P}$, we have that

$$\begin{cases} \dot{\boldsymbol{L}}_{\mathrm{d}} = \dot{\boldsymbol{P}}_1 - \dot{\boldsymbol{P}}_2 = \omega\times(\boldsymbol{P}_1 - \boldsymbol{P}_2) = \omega\times\boldsymbol{L}_{\mathrm{d}} \\ \dot{\boldsymbol{L}}_{\mathrm{m}} = \dot{\boldsymbol{P}}_2\times\boldsymbol{P}_1 + \boldsymbol{P}_2\times\dot{\boldsymbol{P}}_1 = \boldsymbol{t}\times\boldsymbol{L}_{\mathrm{d}} + \omega\times\boldsymbol{L}_{\mathrm{m}}. \end{cases} \tag{A.5}$$

Hence

$$\dot{\ell} = \frac{\dot{\boldsymbol{L}}_{\mathrm{m}}}{(\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}})} - \frac{(\dot{\boldsymbol{L}}_{\mathrm{m}}\cdot\hat{z})}{(\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}})}\frac{\boldsymbol{L}_{\mathrm{m}}}{(\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}})} = \frac{1}{(\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}})}\boldsymbol{t}\times\boldsymbol{L}_{\mathrm{d}} + \omega\times\ell - \frac{(\dot{\boldsymbol{L}}_{\mathrm{m}}\cdot\hat{z})}{(\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}})}\ell, \tag{A.6}$$

and the constraint in (2.1) takes the form

$$\ell\times(\dot{\ell} - \omega\times\ell) = -\frac{\boldsymbol{t}\cdot\ell}{\hat{z}\cdot\boldsymbol{L}_{\mathrm{m}}}\boldsymbol{L}_{\mathrm{d}}. \tag{A.7}$$

Thus if the 3D line is on the plane with normal vector $\boldsymbol{N}$, its image ℓ must obey

$$\boldsymbol{N}\cdot(\ell\times(\dot{\ell} - \omega\times\ell)) = 0. \tag{A.8}$$

To clarify the use of these equations; slant is estimated using Eqs. (A.4) and (A.8). The relations (A.3) and (A.7) are used to analyze the bias.

Appendix B. Expected value of least squares solution

Consider the equation system

$$(e_{1_i}, e_{2_i})\cdot(N_1, N_2) = -e_{3_i}, \tag{B.1}$$

with corrupted measurements

$$e_{k_i} = e'_{k_i} + \delta e_{k_i}, \quad k = 1,2,3.$$

The least square solution amounts to $\boldsymbol{x} = (E^{\mathrm{T}}E)^{-1}E^{\mathrm{T}}G$, where E and G are the $n\times 2$ and $n\times 1$ matrices $E = E' + \delta_E = (e_{1_i}, e_{2_i})_n$, $G = G' + \delta_G = (-e_{3_i})_n$. Assuming that the errors are much smaller than the real values, we develop the LS solution of $\boldsymbol{x}$ in a second order Taylor expansion. Since the noise terms are considered i.i.d. with mean 0, we obtain as an approximation for the expected value $E(\cdot)$:

$$E(\boldsymbol{x}) \approx \boldsymbol{x}' + \sum_i \sum_{\delta t_i\in\delta_V} \frac{\partial^2 E(\boldsymbol{x})}{\partial\delta t_i^2}\bigg|_{\delta t_i = 0} \frac{E(\delta t_i^2)}{2} \tag{B.2}$$

where in the case of motion δ_{v} is the set of all variables $\{\delta_{a'_i}, \delta_{b'_i}, \delta_{\dot{a}'_i}, \delta_{\dot{b}'_i}\}$, and in the case of stereo δ_{v} is the set $\{\delta_{a'_i}, \delta_{b'_i}, \delta_{\tilde{a}'_i}, \delta_{\tilde{b}'_i}\}$. Let M' denote $E'^{\mathrm{T}}E'$. Using the fact that for any arbitrary matrix Q

$$\frac{-\partial Q^{-1}}{\partial x} = Q^{-1}\frac{\partial Q}{\partial x}Q^{-1} \tag{B.3}$$

the expected value of $\boldsymbol{x}$ is approximated as

$$E(x) = x' - \sum_i \sum_{t_i\in V} \frac{\delta t_i^2}{2}\left(M'^{-1}\begin{pmatrix} \frac{\partial^2 e'^2_{1_i}}{\partial t_i^2} & \frac{\partial^2 e'_{1_i}e'_{2_i}}{\partial t_i^2} \\ \frac{\partial^2 e'_{1_i}e'_{2_i}}{\partial t_i^2} & \frac{\partial^2 e'^2_{2_i}}{\partial t_i^2} \end{pmatrix} x' + M'^{-1}\begin{pmatrix} \frac{\partial^2 e'_{1_i}e'_{3_i}}{\partial t_i^2} \\ \frac{\partial^2 e'_{2_i}e'_{3_i}}{\partial t_i^2} \end{pmatrix}\right), \tag{B.4}$$

with V the set $\{a'_i, b'_i, \dot{a}'_i, \dot{b}'_i\}$ or $\{a'_i, b'_i, \tilde{a}'_i, \tilde{b}'_i\}$.

B.1. Motion analysis

Let the variances of δa_i and δb_i be δ_1^2, and the variance of $\delta\dot{a}_i$ and $\delta\dot{b}_i$ be δ_2^2. We then substitute for e_i in (B.4) from (20) and write out the derivatives piecemeal:

$$\begin{aligned}
\frac{\delta^2 e'^2_{1_i}}{\delta a'^2_i} + \frac{\delta^2 e'^2_{1_i}}{\delta b'^2_i} &= 2(b'_i\omega_2 + \omega_3)^2 + 2(a'_i\omega_2 - 2b'_i\omega_1)^2 \\
&\quad + 4\omega_1(\dot{b}'_i + (1+b'_i)^2\omega_1 - a'_ib'_i\omega_2 - a'_i\omega_3), \\
\frac{\delta^2 e'_{1_i}e'_{2_i}}{\delta a'^2_i} + \frac{\delta^2 e'_{1_i}e'_{2_i}}{\delta b'^2_i} &= 2(b'_i\omega_2 + \omega_3)(b'_i\omega_1 - 2a'_i\omega_2) \\
&\quad + 2\omega_2(\dot{b}'_i + (1+b'^2_i)\omega_1 - a'_ib'_i\omega_2 - a'_i\omega_3) \\
&\quad - 2\omega_1(\dot{a}'_i + a'_ib'_i\omega_1 - (1+a'^{2_i})\omega_2 + b'_i\omega_3) \\
&\quad + 2(a'_i\omega_2 - 2b'_i\omega_1)(a'_i\omega_1 + \omega_3), \\
\frac{\delta^2 e'^2_{2_i}}{\delta a'^2_i} + \frac{\delta^2 e'^2_{2_i}}{\delta b'^2_i} &= 2(b'_i\omega_1 - 2a'_i\omega_2)^2 - 4(\dot{a}'_i + a'_ib'_i\omega_1 \\
&\quad - (1+a'^2_i)\omega_2 + b'_i\omega_3)\omega_2 + 2(a'_i\omega_1 + \omega_3)^2, \\
\frac{\delta^2 e'_{1_i}e'_{3_i}}{\delta a'^2_i} + \frac{\delta^2 e'_{1_i}e'_{3_i}}{\delta b'^2_i} &= 2(b'_i\omega_2 + \omega_3)(\dot{b}'_i + \omega_1 - 2a'_i\omega_3) \\
&\quad + 4\omega_3(\dot{b}'_i + (1+b'^2_i)\omega_1 - a'_ib'_i\omega_2 - a'_i\omega_3) \\
&\quad - 2\omega_1(\dot{b}'_ia'_i - \dot{a}'_ib'_i + a'_i\omega_1 + b'_i\omega_2 \\
&\quad - (a'^2_i + b'^2_i)\omega_3) + 2(-2b'_i\omega_1 + a'_i\omega_2) \\
&\quad \times(-\dot{a}'_i + \omega_2 - 2b'_i\omega_3), \\
\frac{\delta^2 e'_{2_i}e'_{3_i}}{\delta a'^2_i} + \frac{\delta^2 e'_{2_i}e'_{3_i}}{\delta b'^2_i} &= 2\omega_2(\dot{a}'_ib'_i - \dot{b}'_ia'_i - a'_i\omega_1 - b'_i\omega_2 \\
&\quad + (a'^2_i + b'^2_i)\omega_3) + 2(b'_i\omega_1 - 2a'_i\omega_2) \\
&\quad \times(\dot{b}'_i + \omega_1 - 2a'_i\omega_3) - 4\omega_3(\dot{a}'_i + a'_ib'_i\omega_1 \\
&\quad - (1+a'^2_i)\omega_2 + b'_i\omega_3) \\
&\quad + 2(a'_i\omega_1 + \omega_3)(-\dot{a}'_i\omega_2 - 2b'_i\omega_3),
\end{aligned}$$

$$\frac{\delta^2 e'^2_{1_i}}{\delta\dot{a}'^2_i} + \frac{\delta^2 e'^2_{1_i}}{\delta\dot{b}'^2_i} = 2, \qquad \frac{\delta^2 e'^2_{2_i}}{\delta a'^2_i} + \frac{\delta^2 e'^2_{2_i}}{\delta\dot{b}'^2_i} = 2,$$

$$\frac{\delta^2 e'_{1_i}e'_{3_i}}{\delta\dot{a}'^2_i} + \frac{\delta^2 e'_{1_i}e'_{3_i}}{\delta\dot{b}'^2_i} = -2a'_i \qquad \frac{\delta^2 e'_{2_i}e'_{3_i}}{\delta\dot{a}'^2_i} + \frac{\delta^2 e'_{2_i}e'_{3_i}}{\delta\dot{b}'^2_i} = -2b'_i \tag{B.5}$$

and all other second order derivatives are 0. For the simplicity of expression, we consider a_i and b_i to be independent random variables which are symmetric with respect to the center of the image coordinate system; in other words, $E(a_i^k) = E(b_i^k) = 0, k = 1, 3$. The $\dot{a}_i$ and $\dot{b}_i$ are very small and set to 0. Then with enough equations, the expected value for the LS solution of $\boldsymbol{x}$ is well approximated by

$$E(\boldsymbol{x}) = \boldsymbol{x}' - M'^{-1}(\delta_2^2 D' + \delta_1^2 F')\boldsymbol{x}' - M'^{-1}\delta_1^2 H' \tag{B.6}$$

where

$$D' = \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix}, \quad H' = \omega_3 \sum_i^n \begin{pmatrix} \omega_1(6b_i'^2 + c_i' + 3) \\ \omega_2(6a_i'^2 + c_i' + 3) \end{pmatrix}, \tag{B.7}$$

$$F' = \sum_i^n \begin{pmatrix} 6b_i'^2\omega_1^2 + c_i'\omega_2^2 + \omega_3^2 + 2\omega_1^2 & 2c_i'\omega_1\omega_2 + 2\omega_1\omega_2 \\ 2c_i'\omega_1\omega_2 + 2\omega_1\omega_2 & c_i'\omega_1^2 + 6a_i'^2\omega_2^2 + \omega_3^2 + 2\omega_2^2 \end{pmatrix} \tag{B.8}$$

with $c_i' = a_i'^2 + b_i'^2$.

B.2. Stereo

Substituting in (B.4) from (8) and setting the variance $E(\delta a_i^2) = E(\delta b_i^2) = E(\delta \tilde{a}_i^2) = E(\delta \tilde{b}_i^2) = \delta^2$ we obtain for the derivatives:

$$\sum_{\delta t_i} \frac{\delta^2 e'_{1_i} e'_{2_i}}{\delta t_i^2} = 0, \quad \sum_{\delta t_i} \frac{\delta^2 e'_{1_i} e'_{3_i}}{\delta t_i^2} = -2(a + \tilde{a}),$$
$$\sum_{\delta t_i} \frac{\delta^2 e'_{2_i} e'_{3_i}}{\delta t_i^2} = -2(b + \tilde{b}), \quad \sum_{\delta t_i} \frac{\delta^2 e'^2_{1_i}}{\delta t_i^2} = 4, \quad \sum_{\delta t_i} \frac{\delta^2 e'^2_{2_i}}{\delta t_i^2} = 4. \tag{B.9}$$

To simplify, we align the image center such that $E(a_i + \tilde{a}_i) = 0$ and $E(b_i + \tilde{b}_i) = 0$, and we obtain for the expected value of $\boldsymbol{x}$

$$E(\boldsymbol{x}) = \boldsymbol{x}' - 2n\delta^2 M'^{-1}\boldsymbol{x}'. \tag{B.10}$$

Appendix C. Stereo: Slant about the vertical and horizontal

Let us use $\boldsymbol{N}_\mathrm{v} = (\boldsymbol{x}_\mathrm{v}, 1) = (\tan\sigma, 0, 1)$ and $\boldsymbol{N}_\mathrm{h} = (\boldsymbol{x}_\mathrm{h}, 1) = (0,\ \tan\sigma, 1)$ to denote the surface normals in the planes slanted about the vertical and the horizontal axes, respectively. Denoting the real slant as σ we derive from Eq. (8) the expected value of $\boldsymbol{x} = (N_1, N_2)$ as

$$E(\boldsymbol{x}) = \begin{pmatrix} N_1' \\ N_2' \end{pmatrix} - \delta^2 \begin{pmatrix} \frac{1}{n}\sum e'^2_{1_i} & \frac{1}{n}\sum e'_{1_i} e'_{2_i} \\ \frac{1}{n}\sum e'_{1_i} e'_{2_i} & \frac{1}{n}\sum e'^2_{2_i} \end{pmatrix}^{-1} \begin{pmatrix} N_1' \\ N_2' \end{pmatrix}$$
$$= \begin{pmatrix} N_1' - n\delta^2 |E(M')|^{-1}(E(\sum e_2'^2)N_1' - E(\sum e_1' e_2')N_2') \\ N_2' - n\delta^2 |E(M')|^{-1}(E(\sum e_1'^2)N_2' - E(\sum e_1' e_2')N_1') \end{pmatrix}. \tag{C.1}$$

Then for the two settings above, omitting terms of $O(\delta^4)$, we have

$$E(||\boldsymbol{x}_\mathrm{v}||^2) = (1 - \epsilon_\mathrm{v})||\boldsymbol{x}'||^2 = \left(1 - \left(2n\delta^2 \frac{E(\sum e'^2_{2_\mathrm{v}})}{|E(M'_\mathrm{v})|}\right)\right)||\boldsymbol{x}'||^2, \tag{C.2}$$

$$E(||\boldsymbol{x}_\mathrm{h}||^2) = (1 - \epsilon_\mathrm{h})||\boldsymbol{x}'||^2 = (1 - (2n\delta^2 \frac{E(\sum e'^2_{1_\mathrm{h}})}{|E(M'_\mathrm{h})|}))||\boldsymbol{x}'||^2. \tag{C.3}$$

Consider the following Taylor expansion:

$$\begin{aligned} E(\sigma) &= \cos^{-1}(((1 - \epsilon)||\boldsymbol{x}'||^2 + 1)^{-\frac{1}{2}}) \\ &= \cos^{-1}(((1 - \epsilon)\tan^2\sigma + 1)^{-\frac{1}{2}}) \\ &= \sigma - (\tfrac{1}{4}\sin 2\sigma)\epsilon + O(\epsilon^2), \end{aligned} \tag{C.4}$$

which is a reasonable approximation only for smaller angles σ. For larger angles, ϵ, which is a function of the translation and the lines, becomes large. We then have

$$E(\sigma_\mathrm{v}) = \sigma - n\delta^2 \frac{\sin 2\sigma}{2} \frac{E(\sum e'^2_{2_\mathrm{v}})}{|E(M'_\mathrm{v})|} = \sigma - n\delta^2 \frac{\sin 2\sigma}{2} C_\mathrm{v} \tag{C.5}$$

$$E(\sigma_\mathrm{h}) = \sigma - n\delta^2 \frac{\sin 2\sigma}{2} \frac{E(\sum e'^2_{1_\mathrm{h}})}{|E(M'_\mathrm{h})|} = \sigma - n\delta^2 \frac{\sin 2\sigma}{2} C_\mathrm{h}. \tag{C.6}$$

Denote the lines in the front-parallel view by

$$\hat{\boldsymbol{L}}_{\mathrm{m}_i} = (-\sin\theta, \cos\theta, k_i), \qquad \hat{\boldsymbol{L}}_{\mathrm{d}_i} = (\cos\theta, \sin\theta, 0),$$
$$\hat{\boldsymbol{L}}_{\mathrm{m}_j} = (-\cos\theta, -\sin\theta, k_i), \qquad \hat{\boldsymbol{L}}_{\mathrm{d}_j} = (-\sin\theta, \cos\theta, 0). \tag{C.7}$$

Then through rotation we obtain the parameters in the vertically and horizontally tilted plane as

$$\begin{aligned} \boldsymbol{L}_{\mathrm{m_v}} &= R_y(\sigma)\hat{\boldsymbol{L}}_\mathrm{m}, \quad \boldsymbol{L}_{\mathrm{d_v}} = R_y(\sigma)\hat{\boldsymbol{L}}_\mathrm{d}, \\ \boldsymbol{L}_{\mathrm{m_h}} &= R_x(\sigma)\hat{\boldsymbol{L}}_\mathrm{m}, \quad \boldsymbol{L}_{\mathrm{d_h}} = R_x(\sigma)\hat{\boldsymbol{L}}_\mathrm{d} \end{aligned} \tag{C.8}$$

where

$$R_y(\sigma) = \begin{pmatrix} \cos\sigma & 0 & -\sin\sigma \\ 0 & 1 & 0 \\ \sin\sigma & 0 & \cos\sigma \end{pmatrix} \quad \text{and}$$
$$R_x(\sigma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\sigma & -\sin\sigma \\ 0 & \sin\sigma & \cos\sigma \end{pmatrix}.$$

We have

$$\boldsymbol{e} \sim (\boldsymbol{T} \cdot \boldsymbol{L}_\mathrm{m})\boldsymbol{L}_\mathrm{d}.$$

Let $\boldsymbol{T} = (t, 0, 0)$. Then under the assumption that the images have small fields of view, that is the magnitude of the k_is is small, we have that

$$\begin{aligned} &\frac{1}{n}E(M'_v) \sim t^2\cos^2\sigma \\ &\quad \times \begin{pmatrix} 2\cos^2\sigma\sin^2\theta\cos^2\theta & \cos\sigma\sin\theta\cos\theta(\sin^2\theta - \cos^2\theta) \\ \cos\sigma\sin\theta\cos\theta(\sin^2\theta - \cos^2\theta) & \sin^4\theta + \cos^4\theta \end{pmatrix} \\ &\frac{1}{n}E(M'_h) \sim t^2 \\ &\quad \times \begin{pmatrix} 2\sin^2\theta\cos^2\theta & \cos\sigma\sin\theta\cos\theta(\sin^2\theta - \cos^2\theta) \\ \cos\sigma\sin\theta\cos\theta(\sin^2\theta - \cos^2\theta) & \cos^2\sigma(\sin^4\theta + \cos^4\theta) \end{pmatrix}. \end{aligned} \tag{C.9}$$

Therefore

$$C_{\mathrm{v}} = n\frac{E(\sum e_{2_{\mathrm{v}}}'^2)}{|E(M_{\mathrm{v}}')|} = \frac{1}{t^2}\frac{(\sin^4\theta + \cos^4\theta)}{\cos^4\sigma(\sin^2\theta\cos^2\theta)}, \tag{C.10}$$

$$C_{\mathrm{h}} = n\frac{E(\sum e_{1_{\mathrm{h}}}'^2)}{|E(M_{\mathrm{h}}')|} = \frac{2}{t^2\cos^2\sigma}, \tag{C.11}$$

and

$$\frac{C_{\mathrm{v}}}{C_{\mathrm{h}}} = \frac{\frac{E(\sum e_{2_{\mathrm{v}}}'^2)}{|E(M_{\mathrm{v}}')|}}{\frac{E(\sum e_{1_{\mathrm{h}}}'^2)}{|E(M_{\mathrm{h}}')|}} = \frac{1}{\cos^2\sigma}\frac{\sin^4\theta + \cos^4\theta}{2\sin^2\theta\cos^2\theta} = \frac{1}{\cos^2\sigma}\frac{\tan^2\theta + \cot^2\theta}{2} > 1. \tag{C.12}$$

Appendix D. Matrix M for motion

Consider a slanted plane with a texture of two major directional components. Let the directional components be $\boldsymbol{L}_{d_1} = (\cos\tau_1\sin\sigma_1, \sin\tau_1\sin\sigma_1, \cos\sigma_1)$ and $\boldsymbol{L}_{d_2} = (\cos\tau_2\sin\sigma_2, \sin\tau_2\sin\sigma_2, \cos\sigma_2)$. From (A.7) we have that

$$\boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \sim (\boldsymbol{t}\cdot\boldsymbol{\ell})\boldsymbol{L}_{\mathrm{d}}.$$

Thus

$$M = E^{\mathrm{T}}E = \begin{pmatrix} \sum e_{1_i}^2 & \sum e_{1_i}e_{2_i} \\ \sum e_{1_i}e_{2_i} & \sum e_{2_i}^2 \end{pmatrix} = \sum(\boldsymbol{t}\cdot\boldsymbol{\ell}_{1_i})^2\sin^2\sigma_1\begin{pmatrix} \cos^2\tau_1 & \sin\tau_1\cos\tau_1 \\ \sin\tau_1\cos\tau_1 & \sin^2\tau_1 \end{pmatrix} + \sum(\boldsymbol{t}\cdot\boldsymbol{\ell}_{2_i})^2\sin^2\sigma_2\begin{pmatrix} \cos^2\tau_2 & \sin\tau_2\cos\tau_2 \\ \sin\tau_2\cos\tau_2 & \sin^2\tau_2 \end{pmatrix} \tag{D.1}$$

and the determinant $\det(M)$ of M amounts to

$$\det(M) = \sum(\boldsymbol{t}\cdot\boldsymbol{\ell}_{1_i})^2\sum(\boldsymbol{t}\cdot\boldsymbol{\ell}_{2_i})^2(\sin\sigma_1\sin\sigma_2\sin(\tau_1-\tau_2))^2. \tag{D.2}$$

Appendix E. Bias correction

Consider the equation system $A\boldsymbol{x} = \boldsymbol{b}$. Let the errors be described by the errors in variable model. The bias of the LS estimator $\boldsymbol{x}_{\mathrm{LS}}$ amounts to

$$\mathrm{bias}(\boldsymbol{x}_{\mathrm{LS}}) = \lim_{n\to\infty} E(\boldsymbol{x}_{\mathrm{LS}} - \boldsymbol{x}') = -\sigma^2(\lim_{n\to\infty}(\frac{1}{n}A'^{\mathrm{T}}A'))^{-1}\boldsymbol{x}'. \tag{E.1}$$

Assuming the variance of the error δA is known, the bias can be removed with the CLS estimator, which amounts to

$$\boldsymbol{x}_{\mathrm{CLS}} = (A^{\mathrm{T}}A - n\sigma^2 I)^{-1}(A^{\mathrm{T}}\boldsymbol{b}), \tag{E.2}$$

and can be rewritten as

$$\boldsymbol{x}_{\mathrm{CLS}} = (I - n\sigma^2(A^{\mathrm{T}}A)^{-1})^{-1}\boldsymbol{x}_{\mathrm{LS}}. \tag{E.3}$$

Thus the covariance matrix $C(\cdot)$ of $\boldsymbol{x}_{\mathrm{CLS}}$ amounts to

$$C(\boldsymbol{x}_{\mathrm{CLS}}) = (I - n\sigma^2(A^{\mathrm{T}}A)^{-1})^{-1}C(\boldsymbol{x}_{\mathrm{LS}})(I - n\sigma^2(A^{\mathrm{T}}A)^{-1})^{-1}. \tag{E.4}$$

The trace of the covariance matrix, which we denote as $\mathrm{cov}(\cdot)$, amounts to:

$$\mathrm{cov}(\boldsymbol{x}_{\mathrm{CLS}}) = \beta\cdot\mathrm{cov}(\boldsymbol{x}_{\mathrm{LS}}) \tag{E.5}$$

with $\beta < 1$.

Let us now investigate what the theoretically best linear estimator should be. We have to adjust the corrected least squares estimator, such that we achieve the smallest MSE (as defined in Eq. (26)). Let $\boldsymbol{x}_\alpha = \alpha\boldsymbol{x}_{\mathrm{CLS}} + (1-\alpha)\boldsymbol{x}_{\mathrm{LS}}$ denote the adjusted CLS estimator.

Then we have that

$$\begin{aligned}\mathrm{MSE}(\boldsymbol{x}_\alpha) &= \alpha^2\mathrm{cov}(\boldsymbol{x}_{\mathrm{CLS}}) + (1-\alpha)^2(\mathrm{bias}^2(\boldsymbol{x}_{\mathrm{LS}}) + \mathrm{cov}(\boldsymbol{x}_{\mathrm{LS}})) \\ &= \alpha^2\mathrm{cov}(\boldsymbol{x}_{\mathrm{LS}})\beta + (1-\alpha)^2(\mathrm{bias}^2(\boldsymbol{x}_{\mathrm{LS}}) + \mathrm{cov}(\boldsymbol{x}_{\mathrm{LS}}))\end{aligned} \tag{E.6}$$

which is a quadratic expression in α. Thus, the MSE minimum is achieved for

$$\begin{aligned}\alpha &= \frac{\mathrm{bias}^2(\boldsymbol{x}_{\mathrm{LS}}) + \mathrm{cov}(\boldsymbol{x}_{\mathrm{LS}})}{\mathrm{bias}^2(\boldsymbol{x}_{\mathrm{LS}}) + \mathrm{cov}(\boldsymbol{x}_{\mathrm{LS}})(1+\beta)} \\ &= 1 - \frac{\mathrm{cov}(\boldsymbol{x}_{\mathrm{LS}})\beta}{\mathrm{bias}^2(\boldsymbol{x}_{\mathrm{LS}}) + \mathrm{cov}(\boldsymbol{x}_{\mathrm{LS}})(1+\beta)} \\ &= 1 - \frac{\beta}{1+\beta+\frac{\mathrm{bias}^2(\boldsymbol{x}_{\mathrm{LS}})}{\mathrm{cov}(\boldsymbol{x}_{\mathrm{LS}})}}\end{aligned} \tag{E.7}$$

This shows that according to the MSE criterion a less bias corrected $\boldsymbol{x}_\alpha$ is better than a bias corrected $\boldsymbol{x}_{\mathrm{CLS}}$. The larger the covariance of the LS estimates, the less the correction should be.

References

Aloimonos, J. (1988). Shape from texture. *Biological Cybernetics, 58*, 345–360.

Andersen, G. J., Braunstein, M. L., & Saidpour, A. (1998). The perception of depth and slant from texture in three-dimensional scenes. *Perception, 27*, 2635–2656.

Backus, B., Banks, M., Ee, R., & Crowell, J. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research, 39*, 1143–1170.

Bajcsy, R., & Lieberman, L. (1976). Texture gradient as a depth cue. *Computer Graphics and Image Processing, 5*, 52–67.

Cagnello, R., & Rogers, B. J. (1993). Anisotropies in the perception of stereoscopic surfaces: the role of orientation disparity. *Vision Research, 33*(16), 2189–2201.

Chowdhury, A. & Chellappa, R. (2003). Statistical error propagation in 3d modeling from monocular video. In: IEEE Workshop on Statistical Analysis in Computer Vision.

Daniilidis, K., & Nagel, H.-H. (1990). Analytical results on error sensitivity of motion estimation from two views. *Image and Vision Computing, 8*, 297–303.

Faugeras, O., & Luong, Q.-T. (2001). *The geometry of multiple images*. Cambridge, MA: MIT Press.

Fermüller, C. (2003). http://www.optical-illusions.org.
Fermüller, C., & Malm, H. (2004). Uncertainty in visual processes predicts geometrical optical illusions. *Vision Research, 44*, 727–749.
Fermüller, C., Pless, R., & Aloimonos, Y. (2000). The Ouchi illusion as an artifact of biased flow estimation. *Vision Research, 40*, 77–96.
Fermüller, C., Shulman, D., & Aloimonos, Y. (2001). The statistics of optical flow. *Computer Vision and Image Understanding, 82*, 1–32.
Fuller, W. (1987). *Measurement error models*. New York: Wiley.
Garding, J. (1993). Shape from texture and contour by weak isotropy. *Journal of Artificial Intelligence, 64*, 243–297.
Goutcher, R., & Mamassian, P. (2002). A ground plane preference for stereoscopic slant. In: Poster, European Conference on Vision Perception.
Hartley, R., & Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.
Helmholtz, H. L. F. V. (1962). *Treatise on physiological optics* (Vol. III). New York: Dover, Translated from the Third German Edition by J. P. C. Southall.
Horn, B. K. P. (1986). *Robot vision*. New York: McGraw Hill.
Hui, J., & Fermüller, C. (2003). Uncertainty in 3d shape estimation. In: Proceedings of the Third International Workshop on Statistical and Computational Theories of Vision.
Ikeuchi, K., & Horn, B. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence, 17*(1–3), 141–184.
Kanatani, K. (1996). *Statistical optimization for geometric computation: Theory and practice*. Amsterdam: Elsevier.
Knill, D. C. (1998). Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Research, 38*, 1655–1682.
Koenderink, J. (1984). What does the occluding contour tell us about solid shape? *Perception, 13*, 321–330.
Koenderink, J., & van Doorn, A. (1992). Surface perception in picture. *Perception and Psychophysics, 52*(5), 487–496.
Koenderink, J., van Doorn, A., & Kappers, A. M. L. (1995). Depth relief. *Perception, 24*, 115–126.
Kutulakos, K., & Dyer, C. (1995). Global surface reconstruction by purposive control of observer motion. *Artificial Intelligence, 78*(1-2), 147–177.
Lydia, N., & Victor, S. (2001). Errors-in-variables modeling in optical flow estimation. *IEEE Trans. on Image Processing, 10*, 1528–1540.
Ma, Y., Soatto, S., Kosecka, J., & Sastry, S. S. (2004). *An invitation to 3-D vision*. Springer.
Maybank, S. J. (1993). *Theory of reconstruction from image motion*. Berlin: Springer.
Mitchison, G., & McKee, S. (1990). Mechanisms underlying the anisotropy of stereoscopic tilt perception. *Vision Research, 30*, 1781–1791.
Nagel, H. (1995). Optical flow estimation and the interaction between measurement errors at adjacent pixel positions. *International Journal of Computer Vision, 15*, 271–288.
Nayar, S. K., Ikeuchi, K., & Kanade, T. (1990). Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Journal of Robotics and Automation, 6*(4), 418–431.
Ng, L., & Solo, V. (2001). Errors-in-variables modeling in optical flow estimation. *IEEE Transactions on Image Processing, 10*, 1528–1540.
Perrone, J. (1982). Slant underestimation: a general model. *Perception, 11*, 641–654.
Ryan, C., & Gillam, B. (1994). Cue conflict and stereoscopic surface slant about horizontal and vertical axes. *Perception, 23*, 645–658.
Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision, 47*(1/2/3), 7–42.
Spetsakis, M. E., & Aloimonos, J. (1991). A multi-frame approach to visual motion perception. *International Journal of Computer Vision, 6*, 245–255.
Stewart, G. W. (1990). Stochastic pertubation theory. *SIAM Review, 32*, 576–610.
Todd, J., & Perotti, V. J. (1999). The visual perception of surface orientation from optical motion. *Perception and Psychophysics, 61*, 1577–1589.
Tomasi, C., & Kanade, T. (1992). Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision, 8*, 137–154.
Tomasi, C., & Zhang, J. (1995). Is structure from motion worth pursuing. In *Proceedings of the seventh international symposium on robotics research* (pp. 391–400). Springer Verlag.
Tsai, R. Y., & Huang, T. S. (1984). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 13–27.
van Huffel, S., & Vandewalle, J. (1991). The total least squares problem: computational aspects and analysis. SIAM.
Yo, C., & Wilson, H. R. (1992). Moving 2D patterns capture the perceived direction of both lower and higher spatial frequencies. *Vision Research, 32*, 1263–1270.