Uncertainty in Shape Estimation from Multiple Views

Abstract

At the apogee of visual recovery are shape models of the scene. Cues such as motion, texture, shading, and contours encode information about the scene surfaces. The information is obtained from image features through estimation processes using constraints that model the inverse image formation. There are many difficulties involved. This paper shows, there is another previously not recognized difficulty of a statistical nature. Because there is noise in all the input parameters, we cannot avoid bias. We first introduce a new constraint for shape from motion which relates image lines and rotation to shape. Then we demonstrate the bias for this constraint. Because the human visual system has to cope with it as well, it makes errors. This explains the underestimation of slant found in computational and psychophysical experiments, and demonstrated here for some illusory displays. We discuss the advantages and disadvantages of the best known estimators, and the optimal procedures for the problem at hand. Finally, we show experiments that confirm the theoretical analysis.

1. Introduction

A large body of work in the Computer Vision literature has been devoted to visual recovery. The term refers to computations which extract physical and geometrical properties of the scene by inverting the (optical and geometrical) image formation process. It includes the low level processes concerned with the detection of features and higher level processes, such as the computation of the light sources, the motion, structure and shape of the scene as well as the material properties of surfaces. The computations involved are estimation processes.

It has been shown in previous work that there is a serious problem with the estimation of image features, that is image points, lines and image motion [1]. There is bias and thus consistent erroneous estimation of image features. The underlying cause is the well known statistical dilemma. Image data is noisy, but because of the complexity of visual computations it is most often not possible to accurately estimate the noise parameters. As a result, bias cannot be avoided.

Since all visual recovery processes are estimations, they all should be effected by this problem. In other words, bias, that is consistent misestimation, really is a principle of visual estimation processes.

In the following section we will model the estimation of shape from motion. But the findings apply to shape computations from any cue. Because the bias is inevitable, also the human visual system suffers from it. Psychophysical studies have found an underestimation of slant for many shape computations. We show that linear estimation models predict these findings. We also create some displays that let you observe this illusory perception.

Shape from motion, or in general shape from multiple views is an active research area, and many research groups are involved in extracting 3D models on the basis of multiple view geometry [2, 3]. The theory proceeds by first solving for camera geometry (where are the cameras?). After the cameras are placed, the structure of the scene is obtained by extending the lines from the camera centers to corresponding features; their intersections provide points in 3D space which make up a 3D model.

Thus, the structure of the scene requires both the translation and the rotation between views. But the structure can be viewed as consisting of two components: (a) the shape, i.e. the normals to the surface and (b) the (scaled) depth. It is quite simple to show that shape depends only on the rotation between two views, while depth depends also on the translation. This new constraint allows us to provide a simple exposition of the uncertainty in the estimation of shape from motion and deduce the underestimation of slant.

1.1 The statistical problem

The constraints used in visual processes often can be formulated as equations linear in the unknowns. Thus the problem is reduced to finding the “best” solution to an overdetermined equation system of the form $A'u' = b'$ where $A' \in R^{N \times K}$ and $b' \in R^{N \times 1}$ and $N \geq K$. The observation $A'$ and $b'$ are always corrupted by the errors, and in addition there is system error. We are dealing with what is called the errors-in-variable (EIV) model in statistical regression, which is defined as:

Definition 1 (Errors-In-Variable Model)

\[
\begin{align*}
    b &= A u + \epsilon \\
    b' &= b + \delta_b \\
    A' &= A + \delta_A
\end{align*}
\]

$u$ are the true but unknown parameters. $A'$ and $b'$ are observations of the true but unknown values $A$ and $b$. $\delta_A, \delta_b$
are the measurement errors and $\epsilon$ is the system error which exists if $A$ and $b$ are not perfect related.

The most common choice to solving the system is by means of LS (least squares) estimation. However, it is well known, that the LS estimator $\hat{u}_t$, whose solution is characterized by $u_t = (A^T A)^{-1} A^T b$, generally is biased [6].

Consider the simple case where all elements in $\delta_A$ and $\delta_b$ are i.i.d random variables with zero mean and variance $\sigma^2$. Then

$$\lim_{n \to \infty} E(u_t - u) = -\sigma^2 \left( \lim_{n \to \infty} \frac{1}{n} A^T A \right)^{-1} u,$$

which implies that $u_t$ is asymptotically biased. Large variance in $\delta_A$, ill-conditioned $A$ or an $u$ which is oriented close to the eigenvector of the smallest singular value of $A$ all could increase the bias and push the LS solution $u_t$ away from the real solution. Generally it leads to an underestimation of the parameters.

In section 3 we discuss alternative estimators. We discuss the underlying problem and the most promising approaches. However, even the best estimator cannot be expected to fully correct the bias.

## 2 Shape from Motion

### 2.1 Formulation of the constraint

Consider the scene to be a textured plane with surface normal $n$. The texture is described by the lines on the plane.

A line $L$ in 3d space is described by Plücker coordinates $L = (L_d, L_m)$, where

$$\begin{cases} L_d = P_1 - P_2; \\ L_m = L_d \times P = P_2 \times P_1. \end{cases}$$

for any points $P, P_1, P_2$ on the line. $L_d$ denotes the direction of the line in space, and $L_m$ its moment. The length of $L_m$ is the distance from the origin to the line. $L_d$ and $L_m$ are perpendicular, that is $L_d \cdot L_m = 0$. The projection of the 3D line $L$ on the image is just $L_m$ and normalized to have the third coordinate 1, it is:

$$\ell = \frac{1}{z \cdot L_m} L_m.$$

The camera undergoes a rigid motion described by a translation vector $T$ and a rotation matrix $R$. Let superscripts 0 and 1 denote quantities at time instances $t^0$ and $t^1$, that is $P^1 = T + R P^0$. Then

$$\begin{cases} L_d^1 = R L_d^0; \\ L_m^1 = R (L_m^0 + T \times L_d^0). \end{cases}$$

$\ell$ is perpendicular to the plane through the line in space and the origin. Thus, the orientation of the line can be obtained from its images at two time instances and the rotation, that is

$$\ell^0 \times (R^T \ell^1) = \frac{L_m^0}{z \cdot L_m^0} \times \frac{R^T (L_m^0 + L_d^0 \times T)}{z \cdot L_m^0} = \frac{(L_m^0 \times T) L_m^0}{(z \cdot L_m^0)(z \cdot L_m^0)} = k L_m^0$$

We model here differential motion, that is a point in space has velocity $P = t + \omega \times P$, in which case we have that

$$\begin{cases} \dot{L}_d = P_1 - P_2 = \omega \times (P_1 - P_2) = \omega \times L_d \\ \dot{L}_m = P_2 \times P_1 + P_2 \times \dot{P}_1 = t \times L_d + \omega \times L_m. \end{cases}$$

Hence

$$\dot{\ell} = \frac{\dot{L}_m}{z \cdot L_m} - \frac{\dot{z} L_m}{z \cdot L_m} L_m = \frac{1}{z \cdot L_m} t \times L_d + \omega \times \ell + \frac{\dot{\ell}}{z \cdot L_m} L_d,$$

and the constraint in (3) takes the form

$$\ell \times (\dot{\ell} - \omega \times \ell) = 0 \quad \text{or} \quad \frac{1}{z \cdot L_m} \ell \times \dot{L}_d.$$  

Thus, if the 3D line is on the plane with normal vector $n$, its image $\ell$ must obey the following constraint

$$n \cdot (\ell \times (\dot{\ell} - \omega \times \ell)) = 0 \quad \text{or} \quad n \cdot e = 0$$

with $e = (\ell \times (\dot{\ell} - \omega \times \ell))$

### 2.2 Error analysis

Let $n = (n_1, n_2, 1)$ be the surface normal, and let $\{ \ell_i = (a_i, b_i, 1) \}$ denote the lines on the plane, and $\{ \ell_i = (\dot{a}_i, \dot{b}_i, 0) \}$ denote the motion parameters of the lines $\ell_i$. We estimate the orientation of the plane using LS estimation.

From (6) we know, that $n$ in the ideal case should satisfy equation,

$$(e_1, e_2, 1) \cdot \left( \begin{array}{c} n_1 \\ n_2 \\ 2 \end{array} \right) = -e_3.$$  

where

$$\begin{cases} e_1 = -\dot{b}_i + (1 + b_i^2) \omega_1 + a_i b_i \omega_2 + a_i \omega_3 \\ e_2 = \dot{a}_i + (a_i b_i \omega_1 - (1 + a_i^2) \omega_2 + b_i \omega_3) \\ e_3 = (\dot{a}_i b_i - \dot{b}_i a_i) + (a_i \omega_1 + b_i \omega_2 - (a_i^2 + b_i^2) \omega_3). \end{cases}$$

There is noise in the measurements of the line locations and the measurements of line motion. For simplicity of notation, let us ignore here the error in the estimates of the rotation parameters. Throughout the paper let primed letters denote estimates, unprimed letters denote real values, and $\delta$’s denote the errors. That is, $\delta a_i = a_i^\prime - a_i$ and $\delta b_i = b_i^\prime - b_i$ with expected value 0, variance $\delta_i^2$; $\delta a_i = a_i^\prime - a_i$ and $\delta b_i = b_i^\prime - b_i$ with expected value 0, variance $\delta_i^2$. The errors $\delta a_i$ and $\delta b_i$ are not perfectly related.
\[ \delta b_i = b_i' - b_i \text{ with expected value 0 and variance } \delta^2. \] 

Then we have

\[ (e_1 + \delta e_1) n_1' + (e_2 + \delta e_2) n_2' = -(e_3 + \delta e_3). \]

Let \( E \) and \( \delta E \) denote the \( N \times 2 \) matrices and \( G \) denote the \( N \times 1 \) matrix as follows,

\[ E = (e_{1i}, e_{2i})_n, \quad \delta E = (\delta e_{1i}, \delta e_{2i})_n, \]
\[ G = (-e_{3i})_n, \quad \delta G = (-\delta e_{3i})_n. \]

Then the estimation \( u' = (n_1', n_2') \) is obtained by solving the equation,

\[ (E + \delta E)^T (E + \delta E) u' = (E + \delta E)^T (G + \delta G). \]

Let \( M \) denote \( E^T E \). Assuming that the errors are much smaller than the real values, we develop the LS solution of \( u' \) is well approximated by

\[ E(u') = u - M^{-1}(\delta^2 D + \delta^2 F)u = M^{-1}\delta^2 H, \quad (7) \]

where

\[ D = \begin{pmatrix} N & 0 \\ 0 & N \end{pmatrix}; \quad H = \omega_3 \sum_i \begin{pmatrix} \omega_1 b_i^2 & 0 \\ 0 & a_i^2 \omega_2 \end{pmatrix}, \]
\[ F = \sum_i \begin{pmatrix} 4b_i^2 \omega_1^2 + c_i \omega_2^2 + \omega_3^2 & c_i \omega_1 \omega_2 \\ c_i \omega_1 \omega_2 & c_i \omega_1^2 + 4a_i^2 \omega_2^2 + \omega_3^2 \end{pmatrix} \]

where \( c_i = a_i^2 + b_i^2 \).

### 2.3 The effects on slant

The slant \( \sigma \) is the angle between the surface normal and the negative \( Z \)-axis \((0^\circ) \) slant corresponds to a plane parallel to the image plane, \( 90^\circ \) slant corresponds to a plane that contains the optical axis) and the tilt \( \tau \) is the angle between the direction of the projection of the surface normal onto the \( XY \)-plane and the \( X \)-axis. Using these coordinates \( \begin{pmatrix} a \\ b \end{pmatrix} = (\cos \tau \sin \sigma, \sin \tau \sin \sigma, \cos \sigma) \).

For the case when rotation around the \( Z \)-axis can be ignored \((i.e., \omega_3 = 0)\) equation \((7)\) simplifies to

\[ E(u') = (I - \delta_A)u = (I - M^{-1}(\delta^2 D + \delta^2 F))u. \]

Since \( D \) and \( F \) are positive defined matrices, so is \( \delta_A \). And usually the perturbation \( \delta \) is small. Then the eigenvalues of \((I - \delta_A)\) are between zero and one, which leads to the Rayleigh quotient inequality:

\[ \frac{E(u')^T E(u')}{u'^T u} \leq ||I - \delta_A|| < 1. \]

Since \( \sigma = \cos^{-1}(1 + u'^T u) \) is a strictly increasing function, by linear approximation, we have

\[ E(u') < \sigma, \]

which shows that slant is underestimated. The degree of underestimation highly depends on the structure of matrix \( M \); the inverse of \( M \) is involved in equation \((7)\). Thus, the smaller the determinant of matrix \( M \), the larger the bias in the estimation. The velocity of rotation also contributes to the magnitude of the bias as can be seen from matrix \( F \); larger velocity more bias.

We can say more about the dependence of slant estimation on the texture distribution. Recall from equation \((4)\) that

\[ e = \frac{(t \cdot \ell)}{L_m} L_d. \]

Let us consider a slanted plane whose texture only has two major directional components. Let the directional components be \( L_{d1} = (\cos \tau_1 \sin \sigma_1, \sin \tau_1 \sin \sigma_1, \cos \sigma_1) \) and \( L_{d2} = (\cos \tau_2 \sin \sigma_2, \sin \tau_2 \sin \sigma_2, \cos \sigma_2) \). Then we have

\[ M = E^T E = \left( \sum_i c_{1i}^2 \right) \left( \sum_i c_{2i}^2 \right) \]
\[ = \sum_i \left( \frac{T}{\ell} \cdot \frac{\ell}{L_m} \right)^2 \sin^2 \sigma_1 \left( \cos^2 \tau_1 \sin \tau_1 \cos \tau_1 \right) \]
\[ + \sum_i \left( \frac{T}{\ell} \cdot \frac{\ell}{L_m} \right)^2 \sin^2 \sigma_2 \left( \cos^2 \tau_2 \sin \tau_2 \cos \tau_2 \right) \]

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

\[ \det(M) = \left[ \left( \frac{1}{N} \sum_i \left( \frac{T}{\ell} \cdot \frac{\ell}{L_m} \right)^2 \right) \sin \sigma_1 \sin \sigma_2 \sin(\tau_1 - \tau_2) \right]^2. \]

The smaller \( \det(M) \), the larger the underestimation. Using our model we can predict the findings from experiments in the psychological literature on the perception of shape from motion \([12 \quad 13]\). For example, it has been observed in \([12]\), that an increase in the slant of a rotating surface causes increased underestimation of the slant. By our formula, it’s easy to see that \( \det(M) \) has a factor \( \sin(\sigma_1) \sin(\sigma_2) \), where \( \sigma_1 \) and \( \sigma_2 \) are the angles between the directions of the line in space and the negative \( Z \)-axis. Unless, they are 0 degree, 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 an illusory display. The scene consists of a plane with two textures, one in the upper half, the other in the lower.
Figure 1: Scene geometry of the shape from motion demonstration

Figure 3 shows the plane when it is parallel to the screen. The texture in the upper part consists of two line clusters with slope $0^\circ$ and $90^\circ$. The lower part has two line clusters with slope $45^\circ$ and $135^\circ$. Figure 4 shows the rotated plane for the case of $\sigma = 60^\circ$ and $\tau = 5^\circ$. The plane is located at the center of a sphere. A video was created for the camera orbiting the sphere along a great circle in the YZ plane as shown in Figure 1—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 $\tau$, at the end it is $\eta$, the tilt is $\theta$. The image sequence can be seen in the video supplied. As can be experienced, it creates the perception of the plane to be 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 greater underestimation of slant. The ratio of the determinants of the upper and lower texture is a good measure. For the given scene it takes values between $\eta'$ (for $\tau$ slant) and $\eta$ (for $\eta$ slant). Hence more underestimation of the slant for the upper part.

Two other properties can be observed that have been discussed above. First, the underestimation gets worse as the slant increases. Second, the misperception is stronger for larger velocities. This is because of the increase of $F$ in equation (7), but also because of the increase of the error in the variance $\delta^2$. That is, there will be larger errors in the temporal derivatives of the line parameters as the speed increases.

Figure 2: (a) Plane with $90^\circ$ slant and $0^\circ$ tilt (b) Plane with $60^\circ$ slant and $5^\circ$ tilt

3 \ Statistical Alternatives

The statistical model that describes the data in visual estimation processes is the errors-in-variable model (Definition 1). Least squares (LS) estimation does not consider errors in the explanatory variables, that is $\delta_A$. The obvious question thus arises: Are there better alternatives that reduce the bias? In this section we discuss well known approaches from the literature and what the difficulties are in applying them to the vision problem. We conclude that the most promising approach is the technique of instrumental variables.

3.1 CLS and TLS estimation

The classical way to correct for the bias is by means of the so-called CLS (corrected least squares) estimator. The CLS estimator is a moment-based method in the sense that:

$$\lim_{n \to \infty} E(A^T A' - n \sigma^2 I) = A^T A.$$ 

Hence an asymptotically unbiased estimation for $u$ is obtained by solving

$$u_c = (A^T A' - n \sigma^2 I)^{-1}(A^T b').$$

CLS estimation requires the knowledge about the variance $\sigma^2$ of the error.

The other popular technique is TLS (total least square) estimation. The basic idea is to deal with the errors in $A'$ and $b'$ symmetrically, instead of just considering $b'$. The TLS estimator seeks to

$$\text{minimize}_{A,b} ||[A', b'] - [\hat{A}, \hat{b}]||_F^2 \quad \text{subject to} \quad \hat{b} \in R(\hat{A}),$$

where $|| \cdot ||_F$ is the Frobenius norm and $R(\hat{A})$ denotes the space spanned by $\hat{A}$'s columns. The TLS estimation $u_t$ is obtained by solving $Au_t = \hat{b}$. If all errors in $\delta_A, \delta_b$ are i.i.d. random variable, then TLS estimation is asymptotically unbiased. If the errors are not i.i.d., then the covariance matrix $\Lambda$ has to be known up to a scale factor in order to whiten the data. That is, we need to transform $(A', b')$ to $(A^*, b^*) = (A', b')\Lambda^{-1}$, for which TLS estimation would be asymptotically unbiased.
TLS recently has attracted a lot of attention. At first sight TLS appears more attractive than CLS. First, it deals with errors in \( A \) and \( b \) symmetrically, which makes better estimation. Second, it only requires the ratio of the error variances. However asymptotically both estimators converge to 0. And obtaining the ratio of the variances of \( \delta_A \) and \( \delta_b \), is actually at least as hard as obtaining the variance of \( \delta_A \). This is particularly true, in the presence of system error. And an incorrect value of the ratio often results in an unacceptably large over correction for the bias in TLS estimation. The importance of system error in linear regression has been expressed by many statisticians. System error is different from measurement error, like re-measuring or re-sampling; but unless we know the exact parameters of the model, we couldn’t test the system error. Thus, in the presence of significant system error, unless reliable estimation of both measurement error and system error is possible, TLS is an inappropriate technique for correcting the bias.

CLS has better robustness on model assumptions than TLS. Moreover system error doesn’t influence it much. If we can reliably obtain the variance of \( \delta_A \), it is the right estimator to choose. The problem is that accurate estimation of the variance of the error is a challenging task if the sample sizes are small. For small amounts of data the estimation of the variance has high variance itself. Consequently this would lead to higher variance for CLS.

### 3.2 Bias correction techniques using re-sampling

Popular re-sampling techniques which have been discussed for bias correction are bootstrap and Jacknife. Suppose we have a sample \( x = (x_1, x_2, \ldots, x_n) \), and define an estimator \( \hat{\theta} = s(x) \) on it. We wish to estimate the bias and variance of \( \theta \). The Jacknife technique focuses on the samples that leave out one observation at a time:

\[
x_{(i)} = (x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n),
\]

which are called the \( i \)-th Jacknife samples. Then for each Jacknife sample, we can have an estimation \( \hat{\theta}_i = s(x_{(i)}) \), which is called a Jacknife replica of \( \hat{\theta} \). The bias estimation of \( \hat{\theta} \) is defined as

\[
\text{bias}_{\text{Jack}} = (n-1)(\frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i - \hat{\theta}).
\]

Instead of re-sampling by deleting individual observations from the original sample set, the bootstrap technique randomly re-samples on the given data set \( x \). Since the bias estimation by Jacknife is a linear approximation to the bias estimation by bootstrap, we only discuss Jacknife here.

The idea behind Jacknife is simple. Expand the estimation of \( \hat{\theta} \) as follows:

\[
E(\hat{\theta}) = \theta + \text{bias}(\hat{\theta}) = \theta + C_0 + \frac{1}{n} C_1 + \frac{1}{n^2} C_2 + O\left(\frac{1}{n^3}\right),
\]

where the \( C_i \)'s are fixed values, not related to the observations \( x_i \). Then we have that

\[
E(\text{bias}_{\text{Jack}}) = (n-1)(\frac{1}{n} \sum_{i=1}^{n} E(\hat{\theta}_i) - E(\hat{\theta}))
\]

\[
= (n-1)(\frac{1}{n} \sum_{i=1}^{n} (\theta' + C_0 + \frac{1}{n} - \frac{1}{n-1} C_1
\]

\[
+ \frac{1}{(n-1)^2} C_2 + O\left(\frac{1}{n^3}\right) - E(\hat{\theta}))
\]

\[
= \frac{1}{n} C_1 + O\left(\frac{1}{n^2}\right).
\]

This shows, that for the case of an asymptotically unbiased estimator \( \hat{\theta} = s(x) \) (that is when \( C_0 = 0 \) the estimated bias differs from the actual bias by an error \( O\left(\frac{1}{n}\right)\)). Hence the bias of the bias-corrected estimation would be \( O\left(\frac{1}{n}\right)\), which is an improvement of one order of \( n \). However, for the case of an asymptotically biased estimator, improving the bias with Jacknife wouldn’t help, as the constant term \( C_0 \) wouldn’t be corrected. The same conclusion holds for the bootstrap technique. In summary, re-sampling techniques wouldn’t help correcting the bias for the EIV model. Instead these techniques are useful for estimating the variance in order to provide confidence intervals ([111]).

### 3.3 Balance between bias and variance

There is a trade-off between bias and variance. It’s generally the case that the estimator correcting for bias increases the variance while decreasing the bias, even when the variance of the error is known precisely. Both TLS and CLS have larger variance than the LS estimator.

The mean squared error (MSE) is often used as a criterion for the performance of an estimator. The MSE is defined as:

\[
MSE(u') = E((u' - u)(u' - u)^T)
\]

\[
= (E(u') - u)(E(u') - u)^T + \text{cov}(u'),
\]

that is, it is the sum of the square of the bias plus the variance. Based on this criterion of minimizing MSE, there are compromised estimators which outperform both uncorrected biased estimators and corrected unbiased estimators. One way to go is to adjust the CLS estimator and use an estimator \( u_{\alpha} = \alpha u_e \), where \( \alpha < 1 \). It can be shown there exists an \( \alpha \) which minimizes the MSE. Since we can write

\[
u_e = (I - n\sigma^2(A^TA')^{-1})u_t
\]
and
\[
\text{cov}(u_{c}) = (I - n\sigma^2(A^TA')^{-1})^{-2}\text{cov}(u_{i}),
\]
we have that
\[
MSE(u_{a}) = \alpha^2\text{cov}(u_{c}) + (\alpha - 1)^2uu^T + \alpha^2(I - n\sigma^2(A^TA')^{-1})^{-2}\text{cov}(u_{i}) + (\alpha - 1)^2uu^T,
\]
which is a quadratic expression. For the one parameter case the MSE takes the minimum when
\[
\alpha = \frac{u^2}{u^2 + \text{cov}(u_{i})^2(1 - \sigma^2E(A^TA'))}.
\]
Thus to achieve MSE superiority a less bias corrected \(u_{a}\) is preferred to a complete bias corrected \(u_{c}\).

### 3.4 Bias correction without a priori knowledge of the error

All the estimation techniques mentioned so far require knowledge of the error covariance. However, the noise parameters in images are hard to obtain. Direct estimation of the error variances is not a wise decision, because it is highly unstable and requires a large sample size. There is a method – the technique of instrumental variables which takes care of the errors in the explanatory variables and does not require the error variance as a priori.

We denote the \(K\)-dimensional row vector of \(A'\) by \(A'_{c}\). Let \(W_{i}\) be a \(q\)-dimensional row vector with \(q \geq K\). Let \(W_{i}\) be such that

i) \(E(W_{i}^T(\delta_{A_{i}}, \delta_{b_{i}})) = (0, 0)\);

ii) The rank of \((\sum_{i=1}^{N} W_{i}^TW)^{-1} \sum_{i=1}^{N} W_{i}^TA'_{i}\) is full, then \(W = (W_{i})\) are called the instrumental variables of \(A'\). For example, we could have two methods to measure \(A'\). If the error in the measurements of the two methods can be treated as independent, then the measurements of one method could be considered as the instrumental variables of the measurements of the other method.

Then we have an unbiased estimator of \(u\) by solving the replaced equations system
\[
(W^TA')u = W^TB',
\]
which by standard least square method amounts to
\[
u = (A'^TWW^TA')^{-1}(A'^TWW^TB'),
\]
This estimator [10] is asymptotically unbiased [10] with the variance close to the variance of the CLS estimator.

We found this method to be most successful in dealing with our problem. The technique of instrumental variables is highly robust for improperly specified error models. If we take multiple measurements of the explanatory variables as the instrumental variables, the worst that can happen is that the different measuring method have the exact measurement error, in which case the method reduces to LS estimation.

### 3.5 The error in the data

The measurements used for shape from motion are the line parameters \(\{a_{i}, b_{i}\}\), and the image motion parameters of the lines, \(\{\tilde{a}_{i}, \tilde{b}_{i}\}\). We can expect four types of noise:

1) **Sensor noise**: affects 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 dependences when the images are smoothed.

2) **Fitting error**: Estimating the line parameters \(a_{i}, b_{i}\) amounts to edge detection. Clearly there are errors in this process. Longer edges are associated with smaller errors and shorter edges with larger errors.

3) **Discretization Error**: Derivatives are computed using difference operators, which have truncation errors associated with them. The magnitude of the error depends on the smoothness and the frequency of the texture.

4) **System error**: When computing the motion of lines, we assume that the image intensity is constant between frames. Significant errors occur whenever there are 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. The error depends on the smoothness of the surface.

Among the errors above, the sensor noise has been considered in a number of papers in structure from motion ([4],[7]). Other errors have hardly been mentioned or have been simply ignored. But actually other errors could contribute much more to the error than the sensor noise. Furthermore, the sensor characteristics may stay fixed. But the other noise components do not. They change with the lighting conditions, the physical properties of the scene being viewed, and the orientation of the viewer in 3D space.

Considering all the errors, what we are dealing with is a heteroscedastic EIV. That is, the errors \(\delta_{A_{i}}\) and \(\delta_{b_{i}}\) cannot be assumed to be independent and identical. CLS and TLS couldn’t be applied for correcting the bias, while the technique of instrumental variables still can handle this model. One may argue that we could simplify the error model to a simple EIV. The problem is still there. TLS would be very dangerous to apply due to its high sensitivity to the error model. CLS seems reasonable, but the heteroscedastic nature of the error make the estimation of the variance unreliable. Re-sampling techniques like bootstrap are not going to work for correcting the bias. The instrumental variable technique promises to be the best solution. It could be realized by using multiple line edge detections, fitting schemes, and difference operators to obtain multiple measurements for the explanatory variables. We will not be able to achieve complete independence of the measurement errors. But we only can make the estimation better, not worse. In the worst case we will have the same estimation as LS.
4 Experiments

We compared the different regression methods for the simpler linear problem of estimating constant optical flow from synthetic data. The basic constraint \( A'u' = b' \) is the optical flow constraint equation

\[
I_x u + I_y v = -I_t,
\]

where \( I_x, I_y \) and \( I_t \) are the spatial and temporal image derivatives and \((u, v)\) are the components of the optical flow.

We compared seven regression methods. TLS estimation was implemented by simple assuming all errors to be i.i.d. CLS was implemented by assuming the errors in \( I_x \) and \( I_y \) to be i.i.d.. The variance of the errors was estimated by the SVD method, that is by taking the smallest singular value of the matrix \([A; b]\) as the estimation of the variance. We implemented two types of bootstrap methods. In one method, the samples \((I_x, I_y, I_t)\) were bootstrapped, in the other the residuals \( I_t - I_x u - I_y v \). We also implemented a simple version of the instrumental variable method by using three differently sized gaussian filters to obtain three samples for each image gradient.

We used a large number of textures and motions. The table below shows the result for one motion with three scenes: a square (1), a diamond (2), and a parallelogram (3) with texture information only from the edges. And another motion for a square textured with a sine wave pattern.

The experiments demonstrate that LS tends to underestimate the parameters, and it imposes serious problem for the “singular” texture structures \((1-3)\); but for the sine wave pattern sequence the bias is small, since there is no major dominant gradient direction. TLS tends to overestimate the parameters, that is, it tends to over-correct the bias. CLS hardly corrects any bias. The reason could be either that the estimation of the variance is highly untrustable, or that the assumption that the errors in \( I_x \) and \( I_y \) are independent is not right. The performance of Bootstrap and Jacknife is close to CLS. The bias hardly gets corrected. The instrumental variable method seems a little bit better than the other methods, but it still didn’t correct the bias much. But it’s not like TLS which often over corrects the bias. The implementation of the instrumental variable method is still brute. A better implementation could possibly give better bias correction.

The tables blow list the error in length and the error in angle between the estimation and the ground truth.

<p>| Experiments with optical flow: ((\rho, \theta) = (0.9, 45)). | (\rho = 0.9428) and (\theta = 45) |
|---|---|---|---|---|---|---|---|</p>
<table>
<thead>
<tr>
<th>LS</th>
<th>TLS</th>
<th>CLS</th>
<th>Boot 1</th>
<th>Boot 2</th>
<th>Jack</th>
<th>Inst.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0331</td>
<td>-0.0304</td>
<td>-0.0335</td>
<td>-0.0323</td>
<td>-0.0322</td>
<td>-0.0321</td>
</tr>
<tr>
<td>2</td>
<td>0.0159</td>
<td>0.1014</td>
<td>0.0149</td>
<td>0.0346</td>
<td>0.0149</td>
<td>0.0365</td>
</tr>
<tr>
<td>3</td>
<td>-0.2949</td>
<td>-0.1145</td>
<td>-0.2952</td>
<td>-0.2901</td>
<td>-0.2900</td>
<td>-0.2804</td>
</tr>
</tbody>
</table>

5 Conclusion

This paper has analyzed the statistics of shape estimation. We discussed that bias is a serious problem for any visual estimation process. We analyzed the bias by means of a new constraint for shape from motion, and we showed that the bias predicts the underestimation of slant, which is known from computational and psychophysical experiments.

References
