Motion Uncertainty and Field of View

Anat Levin — The Hebrew University
Richard Szeliski — Microsoft Research

May 2006

Technical Report
MSR-TR-2006-37

One of the important image and video registration goals is the accurate motion and structure estimation. On the other hand, a good motion estimation is also an important requirement for most mosaicing and novel view generation techniques. While it has been well known for a while that narrow field-of-view cameras have a hard time distinguishing between certain kinds of rotations and translations, it has been recently observed that using omni-directional cameras significantly decrease those ambiguities. In this paper we study the relationship between field of view and the amount of motion uncertainty. To this goal, we derive an analytic formula for the Fisher information matrix involved in the motion estimation task. We use this to examine the coupling between the different motion parameters and the relation between those uncertainties to the camera filed of view. In particular, we provide a formal proof to the previously observed phenomena, that for full 360^o omni-directional cameras there is *no correlation* between rotation and translation parameters estimate.

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
http://www.research.microsoft.com

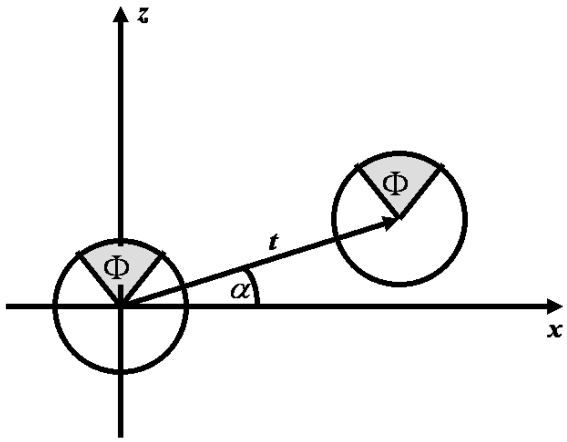

Figure 1: *Cameras setup*

1 Introduction

The simultaneous estimation of camera pose (ego-motion) and scene structure is one of the primary applications of image and video registration. While the success of every mosaicing, panoramic views generation and images based rendering applications relays on accurate motion estimation. For a while now, it has been well known that narrow field-of-view cameras have a hard time distinguishing between certain kinds of rotations and translations (5). (In the extreme case, for an orthographic camera, neither one can be recovered unambiguously with only two frames.)

Several papers have shown that using omni-directional cameras can yield very good ego-motion estimates (1, 4, 3). However, what is the exact relationship between the field of view and the amount of motion uncertainty? If we have a finite number of pixels (but a choice of optics), what is the best way to distribute these pixels spatially to yield the best ego-motion estimates?

In this paper, we perform an analysis of the "classic" two-frame (non-instantaneous) structure-from-motion problem. For our camera model, we use a spherical retina (3-D points are projected onto the unit sphere) of a fixed field of view Φ, since this allows us to vary continuously between a traditional planar sensor with a small field of view all the way to omni-directional cameras. We assume that the points are uniformly distributed over this field of view, and also assume some distribution in depth over the points. We also assume that the camera moves a unit distance in some direction that forms an angle α with the optic axis. For example, $\alpha = 0$ indicates a pure looming motion, while $\alpha = 90^\circ$ indicates a motion perpendicular to the optic axis. We also assume that all the points visible in the first frame are visible in the second.

Given such a configuration (Figure 1), we ask the following questions:

- What is the resulting covariance matrix for the motion estimate, i.e., the uncertainty in both

the rotation and the direction of motion?

- How do these uncertainties vary as a function of the field of view, assuming that we have a fixed number of pixels (i.e., that the point density is inversely proportional to the total angular area subtended by the sensor)?

- Are the motion parameters estimates strongly coupled to one another, or are they essentially independent (i.e., is the covariance matrix diagonal)? Does this relationship change with the field of view?

To answer the above questions, we derive an analytic formula for the Fisher information matrix of our motion estimation problem, and use it to compute the uncertainty (covariance matrix) as a function of viewing angle, assuming a unite distribution of points over the filed of view. This results in a formal proof to the previously observed fact, that for full 360^o omni-cameras there is no correlation between rotation and translation parameters estimate.

The rest of the paper is organized as follows. Section 2 formulates our problem. Section 3 derives the covariance matrix for the motion parameters estimation task analytically, and examines its various properties. Finally section 4 concludes the paper and discuss future research directions.

2 Problem formulation

Let us start by defining our coordinate systems and notation. Without loss of generality, we can place the first camera at the origin looking down the z-axis. The second camera is located a unit distance away from the origin, and w.l.o.g. we can place it in the x-z plane. Since the angle between the direction of view (the z axis) and the direction of motion is α, we place the second camera at $\hat{\mathbf{t}} = (\sin\alpha, 0, \cos\alpha)$ or $(s_\alpha, 0, c_\alpha)$ for short. Again, w.l.o.g. we can assume that the second camera is pointing down the z-axis, since we are only interested in computing the uncertainty in the motion estimates in the vicinity of the true solution. (If the camera were pointed in some other direction, we could pre-rotate the spherical point measurement by the current rotation estimate to get back to this canonical case.)

Let us also assume that we have n 3D points uniformly distributed over the field of view Φ. If we write each point in polar coordinates,

$$\mathbf{p} = r(\cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi) = r(c_\theta s_\phi, s_\theta s_\phi, c_\phi), \tag{1}$$

we see that the points vary over the range $\theta \in [-\pi, \pi]$ and $\phi \in [0, \Phi/2]$. How can we parameterize the distribution over the distances r? If we want to be able to easily accommodate points that lie at infinity, it is convenient to replace the distance r with its inverse *disparity* d, i.e.,

$$\mathbf{p} = d^{-1}(c_\theta s_\phi, s_\theta s_\phi, c_\phi). \tag{2}$$

We can then let d be distributed over some range $[d_{\min}, d_{\max}]$, with $d_{\min} = 0$ allowing points to go all the way out to infinity (which often occurs in outdoor situations).

Given the camera and point configuration, how do these points project onto each of the two spherical retinas? For the first camera (at the origin), we have the simple relationship

$$\mathbf{x} = d\,\mathbf{p} = (c_\theta s_\phi, s_\theta s_\phi, c_\phi). \tag{3}$$

For the second camera, we have

$$\mathbf{x}' = \mathcal{N}(\mathbf{p} - \hat{\mathbf{t}}) = \mathcal{N}(d^{-1}\mathbf{x} - \hat{\mathbf{t}}) = \mathcal{N}(\mathbf{x} - d\hat{\mathbf{t}}), \tag{4}$$

where $\mathcal{N}(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ is the *normalize* operator that converts a vector into its unit-norm direction. (Since $\mathcal{N}(s\mathbf{x}) = \mathcal{N}(\mathbf{x})$, we were able to shift the d next to the $\hat{\mathbf{t}}$ in Equation (4).)

Given a particular collection of points distributed over our viewing angle Φ and a particular translation vector $\hat{\mathbf{t}}$, we can estimate the uncertainty (covariance matrix) of our motion estimate using the Cramer-Rao lower bound (6, 4). To do this, we must first compute the Fisher Information matrix, which involves taking derivatives of our measurements $\{(\mathbf{x}_i, \mathbf{x}'_i), i = 1 \ldots n\}$ with respect to our unknowns.

But what exactly are our unknowns? The exact positions of our 3-D points $\{\mathbf{p}_i\}$ correspond to the unknown structure part of the problem. We must somehow characterize the unknown motion, even though we know for this synthetic problem what the inter-camera rotation (I) and translation ($\hat{\mathbf{t}}$) should be.

To do this, we introduce some *perturbation parameters* on the rotation and translation. We replace the rotation matrix with

$$R = (I_3 + [\omega]_\times), \tag{5}$$

where $\omega = (\omega_x, \omega_y, \omega_z)$ represent infinitessimal rotations around the x, y, and z axes and $[\mathbf{x}]_\times$ is the skew-symmetric matrix form of the cross-product with $\mathbf{x}$. (An equivalent result can be obtained by considering ω to be the *exponential twist* representation of the rotation R and taking its first-order Taylor series expansion (2).)

We represent the translation vector as a 3 parameters vector but add the constraint it has to be a unit vector.

3 Uncertainty analysis

To get closer to an analytic solution, assume that the first camera is noise-free, i.e., whatever feature us observed in the first camera is matched in the second image using a patch tracker. Then, we

have $\mathbf{p} = d^{-1}\mathbf{x}$, and the only unknown structure parameter is the disparity d for each sensed point. Furthermore, assuming $d \ll 1$, we can approximate the normalizing function of

$$\mathbf{x}'_i = \mathcal{N}(\mathbf{p}_i - \hat{\mathbf{t}}) = \mathcal{N}(d_i^{-1}\mathbf{x}_i - \hat{\mathbf{t}}) = \mathcal{N}(\mathbf{x}_i - d_i\hat{\mathbf{t}}), \tag{6}$$

as

$$\mathbf{x}'_i = \mathcal{N}(\mathbf{x}_i - d_i\hat{\mathbf{t}}) \approx \mathbf{x}_i - d_i\hat{\mathbf{t}}, \tag{7}$$

which is the first order Taylor approximation to the normalization function.

Given N pairs of corresponding image points, for a given $(\hat{\mathbf{t}}, \omega, \{d_i\})$ estimate, the expected measurement $\hat{\mathbf{x}}'_i$ after a small perturbation $\Delta t, \Delta d_i$ is

$$\hat{\mathbf{x}}'_i = (I + [\omega]_\times)\,\mathbf{x}_i - d_i\hat{\mathbf{t}} - d_i\Delta\hat{\mathbf{t}} - \hat{\mathbf{t}}\Delta d_i\,, \tag{8}$$

which results in an expected error

$$\Delta\mathbf{x}'_i = \omega \times \mathbf{x}_i - d_i\Delta\hat{\mathbf{t}} - \hat{\mathbf{t}}\Delta d_i = [-\hat{\mathbf{t}}, [x_i]_\times, -d_i I_3] \begin{bmatrix} \Delta d_i \\ \omega \\ \Delta\hat{\mathbf{t}} \end{bmatrix} \tag{9}$$

Let

$$J_i^S = \frac{\partial \mathbf{x}'_i}{\partial \mathbf{d}} = [0, \ldots, \hat{\mathbf{t}}, \ldots, 0] \tag{10}$$

be the $3 \times N$ Jacobian matrix with respect to the structure (i.e., the disparities $\mathbf{d} = \{d_1, \ldots, d_N\}$) whose i'th column equals $\hat{\mathbf{t}}$ and zeros elsewhere. Let

$$J_i^M = \frac{\partial \mathbf{x}'_i}{\partial \omega, \hat{\mathbf{t}}} = [[x_i]_\times, -d_i I_3] \tag{11}$$

be the 3×6 Jacobian w.r.t. the motion parameters, and

$$J_i = [J_i^S, J_i^M] \tag{12}$$

be the total $3 \times (N + 6)$ Jacobian.

The $(N + 6) \times (N + 6)$ Fisher information matrix A' can then be expressed as

$$A' = \sigma^{-2} \sum_i J_i^T J_i \tag{13}$$

where σ^{-2} is the inverse variance associated with each feature measurement $\mathbf{x}'_i$. To avoid the overall scale ambiguity. we constrain $\hat{\mathbf{t}}$ to be a unit norm vector, and augment A with the derivative of this constraint to obtain $A = A' + \lambda c'_t c_t$

$$c_t = \frac{1}{2}\frac{\partial \|\hat{\mathbf{t}}\|^2}{\partial \hat{\mathbf{t}}} = [0_{(N+3)\times 1}, \hat{\mathbf{t}}^T]$$

The complete Fisher information matrix has the structure

$$A = \sigma^{-2} \begin{bmatrix} B & C \\ C^T & D \end{bmatrix}, \tag{14}$$

Where

$$B = I_N, \tag{15}$$

$$C = \begin{bmatrix} -\hat{\mathbf{t}}^T[\mathbf{x}_1]_\times & d_1\hat{\mathbf{t}}^T \\ \vdots & \\ -\hat{\mathbf{t}}^T[\mathbf{x}_N]_\times & d_N\hat{\mathbf{t}}^T \end{bmatrix}, \quad \text{and} \tag{16}$$

$$D = \begin{bmatrix} \sum_i [\mathbf{x}_i]_\times^T[\mathbf{x}_i]_\times & -\sum_i d_i[\mathbf{x}_i]_\times^T \\ -\sum_i d_i[\mathbf{x}_i]_\times & \lambda\hat{\mathbf{t}}\hat{\mathbf{t}}^T + \sum_i d_i^2 I_3 \end{bmatrix}. \tag{17}$$

The ego-motion covariance matrix is then the last 6×6 block of A^{-1}.

To compute A^{-1} we use the matrix inversion lemma,

$$A^{-1} = \begin{bmatrix} B^{-1} - B^{-1}CF^T & F \\ F^T & E^{-1} \end{bmatrix} \quad \text{with} \tag{18}$$

$$E = D - C^T B^{-1} C \quad \text{and} \tag{19}$$

$$F = -B^{-1}CE^{-1}. \tag{20}$$

Since we are interested in the last 6×6 block of A^{-1}, computing E and E^{-1} is enough. In our case,

$$E \;=\; D - C^T B^{-1} C = \begin{bmatrix} E_{\omega\omega} & E_{\omega\hat{\mathbf{t}}} \\ E_{\omega\hat{\mathbf{t}}}^T & E_{\hat{\mathbf{t}}\hat{\mathbf{t}}} \end{bmatrix} \tag{21}$$

where

$$E_{\omega\omega} \;=\; \sum_i ([\mathbf{x}_i]_\times^T[\mathbf{x}_i]_\times - [\hat{\mathbf{t}}]_\times^T \mathbf{x}_i\mathbf{x}_i^T[\hat{\mathbf{t}}]_\times) \tag{22}$$

$$E_{\omega\hat{\mathbf{t}}} \;=\; -\sum_i d_i[\mathbf{x}_i]_\times^T(I_3 - \hat{\mathbf{t}}\hat{\mathbf{t}}^T) \tag{23}$$

$$E_{\hat{\mathbf{t}}\hat{\mathbf{t}}} \;=\; \lambda\hat{\mathbf{t}}\hat{\mathbf{t}}^T + \sum_i d_i^2(I_3 - \hat{\mathbf{t}}\hat{\mathbf{t}}^T) \tag{24}$$

3.1 Analytic integrals

To compute E analytically, we can replace the sums with integrals. However, since the same number of feature points (pixels) is now distributed over a larger solid angle, we need to scale the integrals by the inverse of the surface area subtended by the spherical retina, $S = 2\pi\,(1 - cos(\Phi/2))$. Also, since the feature accuracy σ is proportional to the angle subtended by each pixel, we have $\sigma^{-2} \propto S^{-1}$.

Assuming the distributions on ray orientations and on point depths are independent, the analytic expression for E involves only 1st and 2nd order moments of $\mathbf{x}$ and d. Let $\overline{d} = \int d\varphi d$, $\overline{d^2} = \int d^2\varphi d$, $\overline{\mathbf{x}} = S^{-1}\int \mathbf{x}\varphi\mathbf{x}$, $\overline{\mathbf{x}\mathbf{x}^T} = S^{-1}\int \mathbf{x}\mathbf{x}^T\varphi\mathbf{x}$, $\overline{[\mathbf{x}]_\times^T[\mathbf{x}]_\times} = S^{-1}\int [\mathbf{x}]_\times^T[\mathbf{x}]_\times\varphi\mathbf{x}$. (Note that the last two terms involves integrating only 2nd order moments of $\mathbf{x}$.)

The Fisher information matrix (i.e., the inverse of the Cramer-Rao lower bound on the covariance matrix Σ_M) is therefore

$$\Sigma_M^{-1} = \mathrm{Cov}^{-1}(\omega, \hat{\mathbf{t}}) = \\ S^{-2}\begin{bmatrix} \overline{[\mathbf{x}]_\times^T[\mathbf{x}]_\times} - [\hat{\mathbf{t}}]_\times^T\overline{\mathbf{x}\mathbf{x}^T}[\hat{\mathbf{t}}]_\times & -\overline{d}[\overline{\mathbf{x}}]_\times^T(I_3 - \hat{\mathbf{t}}\hat{\mathbf{t}}^T) \\ -\overline{d}(I_3 - \hat{\mathbf{t}}\hat{\mathbf{t}}^T)[\overline{\mathbf{x}}]_\times & \lambda\hat{\mathbf{t}}\hat{\mathbf{t}}^T + \overline{d^2}(I_3 - \hat{\mathbf{t}}\hat{\mathbf{t}}^T) \end{bmatrix} \tag{25}$$

Parameterizing $\mathbf{x}$ as

$$\mathbf{x} = (c_\theta s_\phi, s_\theta s_\phi, c_\phi), \tag{26}$$

$\theta \in [-\pi, \pi]$ and $\phi \in [0, \Phi/2]$, we can see that

$$\overline{\mathbf{x}} = [0, 0, \pi - \pi c_{\Phi/2}^2]^T \tag{27}$$

and

$$\overline{\mathbf{x}\mathbf{x}^T} = \frac{\pi}{3}\mathrm{diag}(2 - c_{\Phi/2}s_{\Phi/2}^2 - 2c_{\Phi/2}, 2 - c_{\Phi/2}s_{\Phi/2}^2 - 2c_{\Phi/2}, 2 - 2c_{\Phi/2}^3). \tag{28}$$

Examining Equation (25), we can make the following observations:

1. For any translation direction, if $\Phi = 2\pi$, $\overline{\mathbf{x}} = 0$ and therefore Σ_M^{-1} is *block diagonal* (i.e. the upper right and the lower left 3×3 blocks are zero) The covariance Σ_M is block diagonal as well, which confirms our hypothesis that 360° cameras *have* no *correlation between rotation and translation estimation*. However, there is an inner correlation between the 3 rotation parameters and the 3 translation parameters, even in a 360° camera.

2. For a translation along one of the three axes, there is no inner correlation between the three rotation parameters and similarly between the translation parameters. This is true regardless of the viewing angle.

3. Pless (4) showed numerically that if the covariance is averaged over $\hat{\mathbf{t}}$, when $\hat{\mathbf{t}}$ is distributed uniformly such that $\|\hat{\mathbf{t}}\| \leq 1$, there is no inner correlation between the rotation parameters

and there is no inner correlation between the translation parameters. This result can be derived analytically from our analysis. However, it's unclear that a specific case behaves anywhere similar to this average.

To get a more detailed understanding of the structure of the motion covariance matrix, we need to evaluate these integrals for a variety of disparity distributions and fields of view.

3.2 Numerical integrals

To generate plots of the various non-zero entries in the covariance matrix, we assume that the points depths vary uniformly in the range $[1, 100]$. We then evaluate the covariance matrix Σ_M entries as a function of Φ using numerical integration. The evaluation was carried out for three representative motion directions, i.e., a unit translation along the z axis, a unit translation along the x axis (90^o), and a 45^o motion, which is $t = [1, 0, 1]/\sqrt{2}$. Note that these results were independently verified using a numerical Monte-Carlo simulation.

For these three cases, we obtain the following structures for the covariance matrices:

$$\Sigma_{0^\circ} = \begin{bmatrix} \sigma^2_{\omega_x} & 0 & 0 & 0 & -\sigma^2_{\omega_y,t_x} & 0 \\ 0 & \sigma^2_{\omega_x} & 0 & \sigma^2_{\omega_y,t_x} & 0 & 0 \\ 0 & 0 & \sigma^2_{\omega_z} & 0 & 0 & 0 \\ 0 & \sigma^2_{\omega_y,t_x} & 0 & \sigma^2_{t_x} & 0 & 0 \\ -\sigma^2_{\omega_y,t_x} & 0 & 0 & 0 & \sigma^2_{t_x} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{t_z} \end{bmatrix}, \tag{29}$$

$$\Sigma_{45} = \begin{bmatrix} \sigma^2_{\omega_x} & 0 & \sigma^2_{\omega_x,\omega_z} & 0 & \sigma^2_{\omega_x,t_y} & 0 \\ 0 & \sigma^2_{\omega_y} & 0 & \sigma^2_{\omega_y,t_x} & 0 & -\sigma^2_{\omega_y,t_x} \\ \sigma^2_{\omega_x,\omega_z} & 0 & \sigma^2_{\omega_z} & 0 & \sigma^2_{\omega_z,t_y} & 0 \\ \sigma^2_{\omega_y,t_x} & 0 & \sigma^2_{t_x} & 0 & \sigma^2_{t_x,t_z} & \\ \sigma^2_{\omega_x,t_y} & 0 & \sigma^2_{\omega_x,t_y} & 0 & \sigma^2_{t_y} & 0 \\ 0 & -\sigma^2_{\omega_y,t_x} & 0 & \sigma^2_{t_x,t_z} & 0 & \sigma^2_{t_x} \end{bmatrix}, \tag{30}$$

$$\Sigma_{90} = \begin{bmatrix} \sigma^2_{\omega_x} & 0 & 0 & 0 & \sigma^2_{\omega_x,t_y} & 0 \\ 0 & \sigma^2_{\omega_y} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{\omega_z} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{t_x} & 0 & 0 \\ \sigma^2_{\omega_x,t_y} & 0 & 0 & 0 & \sigma^2_{t_y} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{t_y} \end{bmatrix} \tag{31}$$

Figures 2-4 plot the non zero elements of the above covariance matrices as a function of the viewing angle Φ. (In these examples, we set $\lambda = 1$, although in practice, $\lambda \to \infty$ should be used so

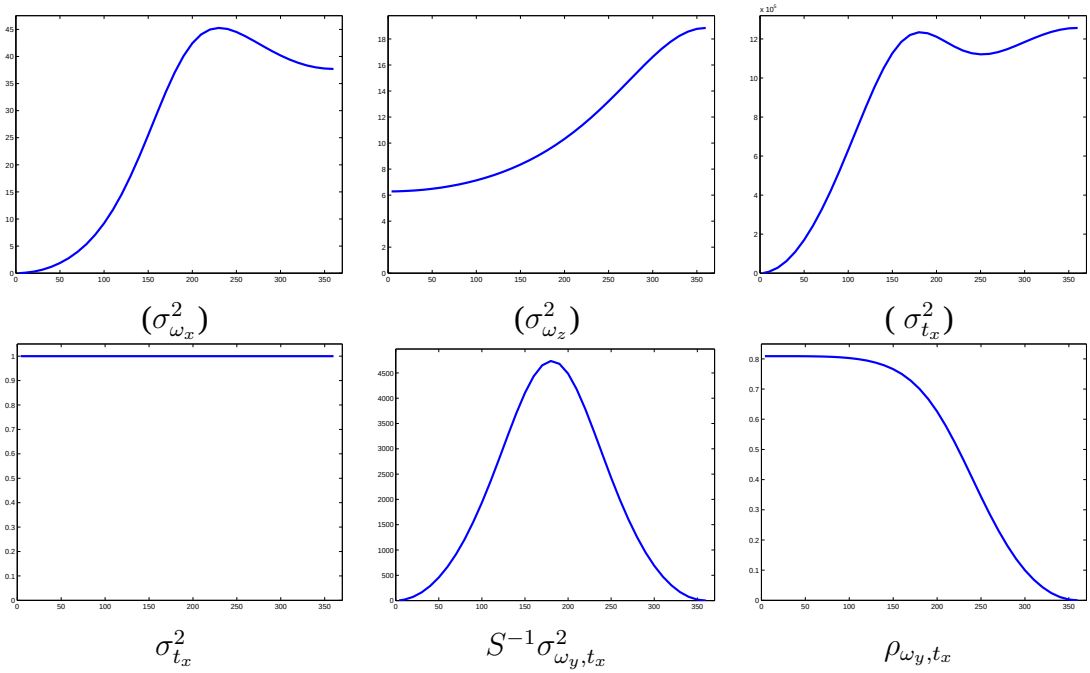

Figure 2: *variance values as a function of viewing angle, assuming forward motion*

that the translational uncertainty along the direction of motion goes to zero.) For the off-diagonal elements the correlation coefficient

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_{ii}\sigma_{jj}} \tag{32}$$

is presented as well.

Let us look at the results of the forward motion ($\alpha = 0°$) case first (Figure 2). When normalized by the surface area of the sensor, the relative variance in the pan/tilt rotation estimates $S^{-1}\sigma^2_{\omega_x}$ is relatively constant as a function of field of view, with a slight increase for hemi-spherical sensors. The absolute variance of the in-plane (optic axis) rotation estimate $\sigma^2_{\omega_z}$ is again fairly constant, with a slight increase for wider angles. The relative variance in the perpendicular translation estimates $S^{-1}\sigma^2_{t_x}$ is similarly constant, with a slight decrease for wider angles. The most interesting plot is for the correlation coefficient ρ_{ω_y,t_x} between the rotation and translation components. As expected, the correlation is very large for small fields of view, and decreases to zero for a full hemispherical sensor. It is interesting that the correlation does not start to drop off significantly until the sensor exceeds a $180°$ field of view.

4 Discussion and conclusions

In this paper we have addressed the ego motion estimation uncertainty issue and studied how does it vary as a function of the field of view. We have formally derived an analytic formula for the estimation covariance, and used it to formally prove properties observed by others, like the fact that ego-motion estimation becomes more accurate and less coupled as the filed of view incases.

One advantage of the presented analysis is that the Jacobian depends only on first image measurements. Assuming all measurement noise is in the second image, the derived formula is independent of a specific noise model . An alternative approach to using the matrix inversion lemma, is to eliminate $3D$ measurement from the estimation, by deriving the covariance directly from the epipolar constraints. This might result in a simpler formula. However, since the epipolar constraints include second image measurements, the exact covariance formula will strongly depend on a specific noise model.

Another question to be analyzed is how close does a "linear" Essential matrix technique come to the true estimate compared to a full bundle adjustment.

References

J. Gluckman and S. K. Nayar. Ego-motion and omnidirectional cameras. In *Sixth International Conference on Computer Vision (ICCV'98)*, pages 999–1005, Bombay, January 1998.

R. M. Murray, Z. X. Li, and S. S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

R. Nelson and J. Aloimonos. Finding motion parameters from spherical flow fields (or the advantages of having eyes in the back of your head). In *Biological Cybernetics 58* 58, 261–273, 1988.

R. Pless. Using many cameras as one. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'2003)*, volume II, pages 587–593, Madison, WI, June 2003.

R. Szeliski and S. B. Kang. Shape ambiguities in structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):506–512, May 1997.

H. W. Sorenson. *Parameter Estimation, Principles and Problems*. Marcel Dekker, New York, 1980.

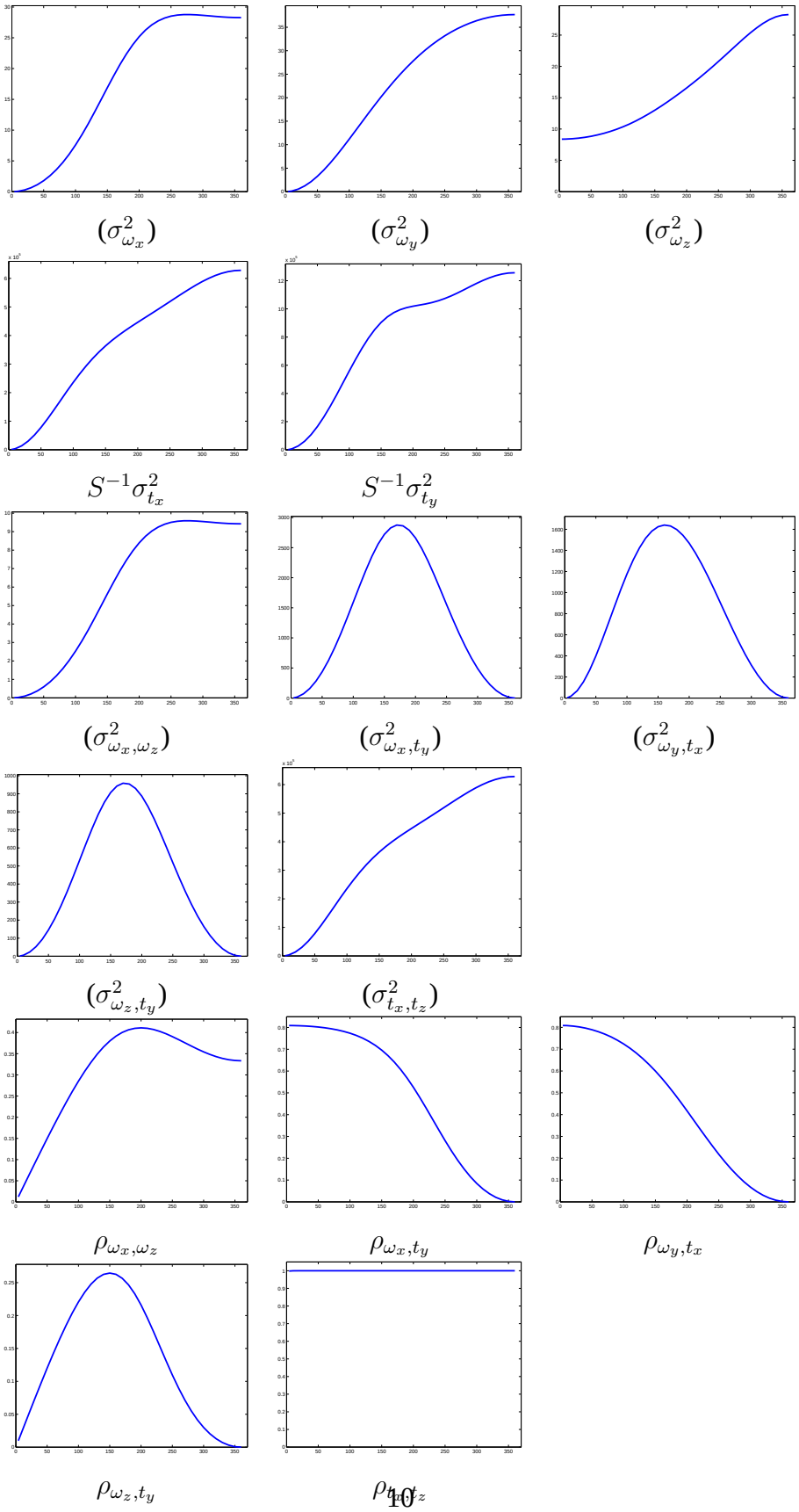

Figure 3: *variance values as a function of viewing angle, assuming* 45^o *motion*

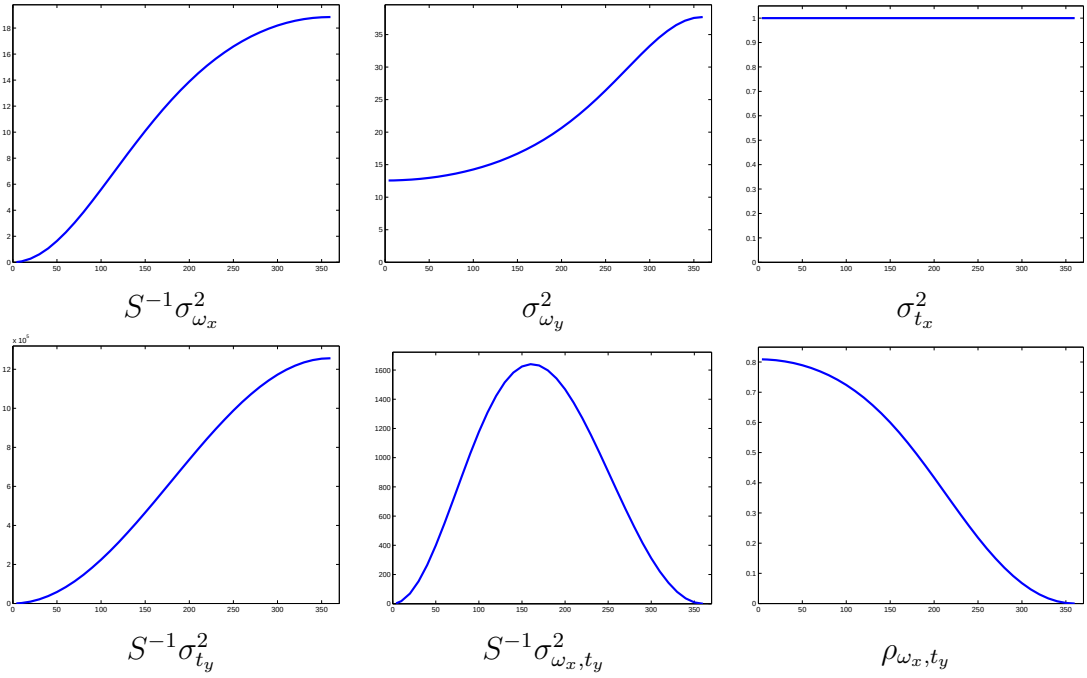

Figure 4: *variance values as a function of viewing angle, assuming* 90^o *motion*