[12] The Optimal Combination of Multiple Sensors Including Stereo Vision

John Porrill, Stephen B Pollard and John E W Mayhew

AI Vision Research Unit
University of Sheffield, Sheffield S10 2TN, UK

Reprinted, with permission of Butterworth Scientific Ltd, from *Image and Vision Computing*, 1987, **5**, 174-180.

Abstract

The statistical combination of information from multiple sources is considered. The particular needs of the target application, stereo vision, require that the formulation be adequate to deal with highly correlated errors and constraints, and that it deal naturally with geometrical data.

1. Introduction

Stereo viewing of an object supplies partial information that can have unacceptably large errors in depth. To build up a more accurate and complete model of the object we would like to view it from many positions and match and combine these views. In a preliminary experiment using artificially generated stereo images the views were matched[1], transformed into the same frame, and combined by a weighted least squares method. Since the data was artificial we had access to the true camera motion when performing the transformation, and the merging behaved correctly. However when the estimate of camera motion supplied by the matcher was used, the resulting model, rather than gradually firming up, gently disintegrated; the errors in the model were correlated with the error in estimated camera motion and accumulated rather than cancelled. In attempting to overcome this we found that it could fruitfully be regarded as a multiple sensor problem (one stereo sensor for each view) the object being to combine these sources of information when their relative calibration is imperfectly known. Although our main interest is stereo vision we shall try to discuss this problem in full generality, using stereo vision only as an illustration.

The combination of information from multiple sensors including vision-like sensors is of immediate practical importance in robotics where one can find many examples of ad hoc combination rules. For example the Intelligent Mobile Platform described by Crowley[2] combines maps of its surroundings obtained from an ultrasonic range sensor into a 'composite local model' by averaging.

Recent treatments have attempted to find a solution which is optimal in some sense. For example the refinement of stereo data by a stationary Kalman filter has been described by Faugeras et al.[3], and Durrant-Whyte[4] has considered the problem of statistical combination of sensors, each of which has its own reference frame, while retaining geometrical consistency between frames. The treatment given here draws largely from these two sources.

2. Statistical Combination

What are the benefits of statistical combination of information? Firstly there is a reduction in error proportional to the square root of the number of similar observations. This is welcome and in many applications is the whole aim of data combination, but in stereo vision, where we might have a typical error in an observed angle of five degrees say, the one hundred observations needed to reduce the error to a respectable half a degree are not likely to be available.

Even though stereo can be very inaccurate the errors are highly anisotropic: lateral errors are much smaller than depth errors. If two stereo views from different known positions are available we can largely eliminate this anisotropy and give much better localisation in depth. This illustrates a general point that the combination of information from sensors with complementary anisotropies can be very beneficial. We can represent the error of each sensor by a confidence ellipsoid and we require statistical combination laws to perform the book-keeping task of 'intersecting' these ellipsoids. One common way of simplifying the mathematics of statistical combination is to majorise the error covariance matrices by diagonal matrices since they are much easier to handle. This replaces error ellipsoids by their smallest enveloping spheres. Since we believe that error anisotropy can be on our side, we cannot make use of this method.

Another source of increased accuracy can be the imposition of constraints. For example if we see that three lines almost intersect we may hypothesise the existence of a trihedral vertex. A statistical model allows us to test this hypothesis and if it is accepted to optimally adjust the data to satisfy the constraint and accurately specify the vertex. The data could thus be adjusted to achieve consistency with a symbolic (region, edge, vertex graph) description of the data.

3. Gauss-Markov Estimation Theory

The simplest and best-developed theory of statistical combination of measurements deals with minimum variance estimators for linear measurement equations, these are maximum likelihood estimators when the noise is Gaussian. Though the robustness of such methods is open to question (we will discuss this later) they are mathematically very attractive. If we try to impose linear constraints (which are essentially exact measurements) the error covariance matrix becomes singular, and the classical Gauss-Markov theorem fails though a generalisation is available to deal with this case. Since this gives an essentially complete theoretical basis for the treatment of linear measurements, we will summarise the results here, based on an elegant treatment of the theory as an application of the Moore-Penrose pseudo-inverse by Albert[5]. Morrison[6] also deals with the subject of linear models and gives more statistical background.

4. Linear Measurement Equations and Linear Constraints

The state to be estimated is a vector $\mathbf{x} \in \mathbf{R}^m$, the measurement and noise are vectors $\mathbf{z}, \mathbf{u} \in \mathbf{R}^n$, and the measurement equation is assumed to be linear

$$\mathbf{z} = H\mathbf{x} + \mathbf{u}$$

where the 'plant matrix' H is $n \times m$. The noise $\mathbf{u}$ is taken to be Gaussian with zero mean and covariance matrix R. This formulation is more general than it looks since if we also have a prior estimate $\mathbf{x}_0$ with prior covariance S_0 and $\mathbf{x}$ is subject to linear constraints $C\mathbf{x} = \mathbf{c}$ we can form a composite equation of exactly the same form

$$\tilde{\mathbf{z}} = \begin{bmatrix} \mathbf{z} \\ \mathbf{x}_0 \\ \mathbf{c} \end{bmatrix} = \begin{bmatrix} H \\ 1 \\ C \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{bmatrix} = \tilde{H}\mathbf{x} + \tilde{\mathbf{u}}$$

where the covariance of the composite measurement error $\tilde{\mathbf{u}}$ is

$$\mathrm{Cov}[\tilde{\mathbf{u}}] = \tilde{R} = \begin{bmatrix} R & 0 & 0 \\ 0 & S_0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

From now on we assume we are dealing with this general situation and drop the tildes. It is easy to see that the matrix R will be singular when constraints (essentially exact measurements) are present. Since R is symmetric and positive and it has a unique positive symmetric square root

$$V = R^{1/2} \qquad R = V^2.$$

To manipulate such singular matrices we will need a generalisation of the concept of the inverse M^{-1} of a matrix M to general matrices. This is given by the Moore-Penrose pseudo-inverse

$$M^+ = \lim_{\delta \to 0} (M^t M + \delta^2 1)^{-1} M^t$$

for details of its properties and many uses see Albert[5]. A discussion of the role of other generalised inverses in estimation theory can be found in Mitra[6].

We include the possibility that the measurement $\mathbf{z}$ may not be sufficient (even when exact) to determine the state $\mathbf{x}$ completely. When this is so there are still linear functionals of $\mathbf{x}$ which can be estimated from the given measurements, the estimable linear functionals.

5. Definition

The functional $\mathbf{y} = L\mathbf{x}$ is said to be *estimable* if it has an unbiased linear estimate $\hat{\mathbf{y}} = M\mathbf{z}$ (that is an estimate such that $E[\hat{\mathbf{y}}] = \mathbf{y}$) for some matrix M.

Theorem

The $\mathbf{y}$ above is estimable if and only if

$$LH^+H = L$$

that is, if and only if

$$\mathrm{Range}(L^t) \subseteq \mathrm{Range}(H^t)$$

More simply, the rows of L must be linear combinations of the rows of H.

In particular taking $L = 1$, $\mathbf{x}$ itself is estimable only if H has full row-rank.

The generalised Gauss-Markov theorem now gives us the best estimate of estimable $\mathbf{y}$'s.

Theorem

Let $\bar{V} = V(1 - HH^+)$ and $G = H^+(1 - (\bar{V}^+\bar{V})^t)$ and let $\hat{\mathbf{x}} = G\mathbf{z}$ then

a) $E[\hat{\mathbf{x}}] = H^+H\mathbf{x}$ (this is the 'estimable part' of $\mathbf{x}$).

b) If $\mathbf{y}$ as above is estimable, then its best linear unbiased estimator is $\hat{\mathbf{y}} = L\hat{\mathbf{x}}$

c) If the error covariance $R = V^2$ is non-singular

$$\hat{\mathbf{x}} = (H^t R^{-1} H)^+ H^t R^{-1} \mathbf{z}$$

This last can be recognised as the usual formula for the weighted least squares estimate of $\mathbf{x}$ with a pseudo-inversion replacing the inversion.

Suppose we now wish to find the estimate of $\mathbf{y}$ given that $\mathbf{x}$ satisfies a further set of constraints $C\mathbf{x} = \mathbf{c}$. We could extend the measurement equation further so as to include these constraints as extra measurements but we can also write down extensions to the results above which solve this problem directly.

Theorem

Let $M = H(1 - C^+C)$. Then $\mathbf{y}$ is estimable given the additional constraints if and only if $LM^+M = L$ that is, if and only if $\mathrm{Range}(L^t) \subseteq \mathrm{Range}(M^t)$

All $\mathbf{y}$'s which were estimable without the constraints are still estimable, and some more may become estimable (for example $C\mathbf{x}$ is obviously estimable as $\mathbf{c}$ whatever its status before).

Theorem

Let M be as above and let

$$\bar{V} = V(1 - MM^+)$$

$$G = (1 - C^+C)\, H^+ \,(1 - (\bar{V}^+\bar{V})^t)$$

and

$$\hat{\mathbf{x}} = C^+\mathbf{c} + G(\mathbf{z} - HC^+\mathbf{c})$$

then

a) $E[\hat{\mathbf{x}}] = C^+\mathbf{c} + M^+M\mathbf{x}$

b) If $\mathbf{y}$ is estimable its best linear unbiased estimator is given by $\hat{\mathbf{y}} = L\hat{\mathbf{x}}$.

If we want to test the hypothesis that the constraint holds given the previous measurement $\mathbf{z}$ we first calculate the estimators $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}_C$ for $\mathbf{x}$ before and after imposing the constraint. Each estimate tries to minimise the weighted sum square error, but the second minimisation is constrained, so its residual will be larger. The increase in residual

$$\delta\varepsilon = (\mathbf{z} - H\hat{\mathbf{x}}_C)^t R^+ (\mathbf{z} - H\hat{\mathbf{x}}_C) - (\mathbf{z} - H\hat{\mathbf{x}})^t R^+ (\mathbf{z} - H\hat{\mathbf{x}})$$

measures the distortion of the data required to impose the constraint. A maximum likelihood test of the constraint can be performed by testing $\delta\varepsilon$ as a central χ^2 variable with $\mathrm{Rank}(HH^+ - MM^+)$ degrees of freedom (this is the number of independent constraints imposed). In general

we will not be testing alternative hypotheses against another, if we were we would need to discuss the power of this test, details can be found in Morrison[6].

6. Recursive Least Squares

The theory as described above covers all our needs, but is very unwieldy computationally and conceptually. A great simplification is achieved by considering the recursive computation of estimators and residuals when single scalar measurements and constraints are added sequentially. Of course this has practical application since data acquisition is often serial in nature.

Suppose we have already performed some measurements and let A be the matrix which projects vectors perpendicular to the estimable subspace of the state space $\mathbf{R}^n$ (so if $\mathbf{y}$ is estimable then $A\mathbf{y} = 0$) if we have performed measurements sufficient in principle to determine $\mathbf{x}$ then $A = 0$. Let $\hat{\mathbf{x}}$ be our present estimate of the estimable part of $\mathbf{x}$ and let the covariance of this estimate be $\hat{S}$. Let the residual so far be ε.

Now suppose we make a single scalar measurement

$$z = \mathbf{h}^t\mathbf{x} + u \qquad \mathrm{var}[u] = \sigma^2$$

with measurement noise uncorrelated with the noise in our previous measurements. Then we can give update rules for A, $\hat{S}$, $\hat{\mathbf{x}}$, and ε. There are two cases: firstly if the new measurement direction $\mathbf{h}$ has already been sampled *i.e.* if $A\mathbf{h} = 0$ then:

$$A' = A$$

$$\hat{S}' = \hat{S} - \frac{(\hat{S}\mathbf{h})^t(\hat{S}\mathbf{h})}{\sigma^2 + \mathbf{h}^t\hat{S}\,\mathbf{h}}$$

$$\mathbf{k} = \frac{\hat{S}\mathbf{h}}{\sigma^2 + \mathbf{h}^t\hat{S}\,\mathbf{h}}$$

$$\varepsilon' = \varepsilon + \frac{(z - \mathbf{h}^t\mathbf{x})^2}{\sigma^2 + \mathbf{h}^t\hat{S}\,\mathbf{h}}$$

if the measurement is in a new direction

$$A' = A - \frac{(A\mathbf{h})(A\mathbf{h})^t}{\mathbf{h}^t A\,\mathbf{h}}$$

$$\hat{S}' = \hat{S} - \frac{(\hat{S}\mathbf{h})(A\mathbf{h})^t + (A\mathbf{h})(\hat{S}\mathbf{h})^t}{\mathbf{h}^tA\mathbf{h}} + \frac{\sigma^2 + \mathbf{h}^t\hat{S}\mathbf{h}}{(\mathbf{h}^tA\mathbf{h})^2}(A\mathbf{h})(A\mathbf{h})^t$$

$$\mathbf{k} = \frac{A\mathbf{h}}{\mathbf{h}^t A\,\mathbf{h}}$$

$$\varepsilon' = \varepsilon$$

in either of these cases the new estimate is given as

$$\hat{\mathbf{x}}' = \hat{\mathbf{x}} + \mathbf{k}\,(z - \mathbf{h}^t\hat{\mathbf{x}})$$

In this formulation the equations form a stationary Kalman filter for sequential scalar measurements. As an alternative to storing and updating the matrix A if one knows that certain components of $\mathbf{x}$ have not been measured one can assign them very large variances initially.

A problem with the recursive application of this update rule is that a sequence of successive measurement must be mutually independent, this is a major restriction. One case in which this is not a problem is when we need to impose a sequence of constraints, since constraints can always be treated as independent if they are not redundant or inconsistent. The imposition of a single scalar constraint (that is a measurement with $\sigma = 0$) when we have sufficient previous measurements that $\mathbf{x}$ itself is estimable (so $A = 0$) is remarkably simple, the update rule is

$$\hat{S}' = \hat{S} - \frac{(\hat{S}\mathbf{h})\,(\hat{S}\mathbf{h})^t}{\mathbf{h}^t\,\hat{S}\,\mathbf{h}}$$

$$\mathbf{k} = \frac{\hat{S}\mathbf{h}}{\mathbf{h}^t\,\hat{S}\,\mathbf{h}}$$

$$\hat{\mathbf{x}}' = \hat{\mathbf{x}} + \mathbf{k}\,(z - \mathbf{h}^t\,\mathbf{x})$$

$$\varepsilon' = \varepsilon + \frac{(z - \mathbf{h}^t\,\mathbf{x})^2}{\mathbf{h}^t\,\hat{S}\,\mathbf{h}}$$

The maximum likelihood test of the constraint treats $\varepsilon' - \varepsilon$ as χ^2 on one degree of freedom for Gaussian errors.

7. Measurement Primitives

To allow a measurement process to return only vectors in $\mathbf{R}^n$ is conceptually limiting. Often the type of process we want to consider is more naturally regarded as measuring and returning a more complex geometrical primitive. For example a simple touch sensor might 'measure' a plane by returning a point on it; an inaccurately calibrated stereo rig might attempt to measure the projective transformation between its own visual world and the real world. Of course such primitives, however complex, can always be encoded into vectors in $\mathbf{R}^n$, the point is that this should be done in a way that reflects their natural geometric structure.

Suppose we are dealing with a class of geometric primitives forming a smooth manifold M (most of the entities we are interested in satisfy this condition: points, lines, planes, circles, rotations, projective transformations ...). An error prone measurement of such a primitive can be thought of as sampling from a random process on M. How are the statistics of such processes to be specified? If we choose local coordinates 'sensibly' in a neighbourhood U of some point of M, that is, a chart

$$\phi : U \to \mathbf{R}^n$$

assigning coordinates $\phi(\pi) = \mathbf{x} = (x^1, x^2, \cdots, x^n)^t$ to points π of M we can say that a process is 'approximately normal' with mean a given primitive $\mu \in M$ if the probability distribution of $\mathbf{x}$ has the approximate form

$$p(\mathbf{x}) \propto \exp(-\frac{(\mathbf{x}-\mathbf{m})^t\,R^+\,(\mathbf{x}-\mathbf{m})}{2})$$

where $\mathbf{m} = \phi(\mu)$, and the probability density function is non-negligible only inside U. Under changes of coordinates that are approximately affine over the support of the probability distribution (using the summation convention)

$$x^{i'} = x^{i'}_0 + A^{i'}_j\,x^j + \cdots$$

(A is the Jacobian matrix of the transformation) the mean transforms like a vector and the covariance matrix $R^{ij} = E[x^ix^j]$ like a second rank tensor

$$m^{i'} = m^i + x^{i'}_0 \qquad R^{i'j'} = A^{i'}_k\,A^{j'}_l\,R^{kl}$$

To discuss what is meant by a 'sensible' choice of coordinates would lead into very deep waters indeed. Stated simply, the coordinates chosen should not be too distorted with respect to the underlying symmetries of the manifold

(all the manifolds in which we are interested have such symmetries) at the scale of description required. For example, if we are describing a point on the unit sphere polar coordinates are fine near the equator, and singular at the poles. We can describe a probability distribution with peak near to a pole adequately in terms of these coordinates only if it has very small variance.

A problem with local coordinates is that they are often inconvenient for the description of geometrical relationships. For example the condition that two unit vectors be perpendicular is not a pretty sight when written out in polar coordinates. However we often have an alternative global description (in this case as a vector $\mathbf{n}$ with $\mathbf{n}\cdot\mathbf{n} = 1$) in which such relationships are easily expressed. The problem with such descriptions is that they tend to involve non-linear constraints (the condition $\mathbf{n}\cdot\mathbf{n} = 1$ above, for example). This second description (which we will call the global description) will in general be an an embedding of M in a higher dimensional space $\mathbf{R}^N$.

$$\iota : M \to \mathbf{R}^N : \pi \to \xi = (\xi^1, \xi^2, \ldots, \xi^N)^t$$

(*e.g.* for the two dimensional manifold of unit vectors it is the natural embedding as the sphere in $\mathbf{R}^3$). We want to choose local coordinates in such a way that transformation between the local and global descriptions is simple, In general this will require that we deal with affine approximations to the transformation, but we this is no great loss, since our statistics can only deal with affine transformations. These will take the form

$$\xi = \xi_0 + F_x\,\mathbf{x} \qquad \mathbf{x} = f_\xi\,(\xi - \xi_0)$$

where F_x and f_ξ are $N\times n$ and $n\times N$ matrices respectively, and ξ_0 represents the origin of the local coordinate system. The local coordinates should be attached to the primitive rigidly, in the sense that a rigid motion in space preserves the form of the matrices F_x and f_ξ. The systematic use of these representations can simplify and mechanise the problem of deriving measurement equations and constraints.

8. Sensors and Constraints

For our purposes a sensor observes a measurement primitive ξ, its internal state (motion since last view, miscalibration, *etc.*) is described by another primitive σ, and it returns a third primitive ζ as measurement. If the measurement were exact the relationship between the three would typically have the form

$$\mathbf{h}(\zeta, \sigma, \xi) = 0$$

Choose local coordinates about the actual measurement ζ. On the assumption of small approximately normal errors we know that the true measurement ζ is related to the actual measurement ζ_0 by

$$\zeta = \zeta_0 + F_z\,\mathbf{z} \qquad \mathbf{z} = N(0, R)$$

If we have estimates σ_0 and ξ_0 of the other two quantities and introduce local coordinates about these estimates we can linearise the equation (with obvious choice of notation) as

$$\tilde{\mathbf{z}} = -\mathbf{h}(\zeta_0, \sigma_0, \xi_0) = \left[\frac{\partial \mathbf{h}}{\partial \sigma}\cdot F_s\,,\ \frac{\partial \mathbf{h}}{\partial \xi}\cdot F_x\right]\begin{bmatrix}\mathbf{s}\\ \mathbf{x}\end{bmatrix} + \frac{\partial \mathbf{h}}{\partial \zeta}\cdot F_z\,\mathbf{z} = \tilde{H}\tilde{\mathbf{x}} + \tilde{\mathbf{u}}$$

This is a linear measurement equation for the perturbations $\mathbf{x}$, $\mathbf{s}$ where the measurement error has covariance

$$\tilde{R} = \mathrm{cov}[\tilde{\mathbf{u}}] = \frac{\partial \mathbf{h}}{\partial \zeta}\,F_z\,R\,F_z^t\,\frac{\partial \mathbf{h}}{\partial \zeta}^t$$

This equation can now be treated by Gauss-Markov theory. However if we wish to perform each component $\tilde{z}_i$ of the measurement sequentially they must be independent, that is, the matrix $\tilde{R}$ must be diagonal. This is unlikely to be the case. As stated previously we do not wish to majorise $\tilde{R}$ by a diagonal matrix, since this loses information. One solution is to find the eigenvectors $\mathbf{e}_i$ of $\tilde{R}$ and apply the the scalar measurements

$$\mathbf{e}_i^t\,\tilde{\mathbf{z}} = \mathbf{e}_i^t\,H\tilde{\mathbf{x}} + \mathbf{e}_i^t\,\tilde{\mathbf{u}}$$

sequentially since these are independent.

A second approach is to extend our state space to include the measurement $\mathbf{z}$ and to regard the equation above as a set of constraints on the augmented state vector. If we ensure that the set of constraints is not redundant then they can be applied sequentially.

In order to justify the above linearisations we must be able to obtain good initial estimates of the state and calibration. In particular the total sensor information available must be sufficient in the absence of error to determine all the unknowns which have non-linear descriptions or measurement equations; the purpose of the statistical method is to optimise this estimate.

9. Sensor Combination

The above analysis allows us to outline a framework for sensor combination that is conceptually very attractive. We have a world containing a list of primitives with estimated positions $(\xi_1, \xi_2, \cdots)$ and a set of independent sensors whose calibrations with respect to the world frame are estimated as $(\sigma_1, \sigma_2, \cdots)$. The world frame can be chosen for convenience, as the frame of our most important sensor for example. The state vector is the list $(\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{x}_1, \mathbf{x}_2, \cdots)$ of corrections which must be made to the calibrations and world primitives. Each of these corrections is referred to an intrinsic frame attached to the primitive, initially each will be zero.

If a measurement of the form of the last section is made, the error in the measurement is adjoined to the state space and the covariance matrix of the measurement adjoined to the state covariance. The measurement equation is then applied as a series of constraints. If the measurement itself is a primitive of interest it can be kept in the state vector, otherwise it is dropped. At any stage we can apply the corrections specified in the state vector to the primitives, this does not require any change in the state covariance, since such changes would be second order small quantities. If we want to change our world frame, we need only change our global representations of the primitives, the corrections and the state covariance being attached to the intrinsic frame of the primitives.

We will now describe the above process in some detail in a case study of the multiple stereoview problem.

10. The Stereo Sensor

Our immediate application of the above analysis is to stereo data. This is provided by a 'sensor' which is a

combination of error prone processing stages: acquisition of a pair of images, edge detection to sub-pixel acuity by a Canny operator[8], stereo edge matching by the PMF algorithm[9], and finally plane, line and circle fitting to produce the geometrical descriptive base (GDB)[10].

11. Description of Stereo Primitives

The most reliable part of the GDB is a list of straight edges found in the scene (this is the basic input to our matcher). The endpoints of these edges are not very informative due to unpredictability in segmentation, so the measurement primitive which we output is essentially a straight line in three space. We need a convenient global description of these primitives; the most convenient is as a pair of vectors $(\mathbf{p}_0, \mathbf{v}_0)$ where $\mathbf{p}_0$ is the position of a point on the line and $\mathbf{v}_0$ is its direction vector. We now choose vectors $\mathbf{v}_1$, $\mathbf{v}_2$ such that the $\mathbf{v}_i$ form a basis, and add these to our description of the line in the GDB, this then forms an intrinsic reference frame which will be carried with the line throughout its history.

Any nearby line can be described by a position vector and a direction vector

$$\mathbf{p} = \mathbf{p}_0 + p_1\mathbf{v}_1 + p_2\mathbf{v}_2$$

$$\mathbf{v} = \mathbf{v}_0 + v_1\mathbf{v}_1 + v_2\mathbf{v}_2$$

(note that $\mathbf{v}$ is unit to first order) and we use (p_1, p_2, v_1, v_2) as local coordinates on the line manifold. The transformations between local and global descriptions are thus

$$\begin{bmatrix}\mathbf{p}\\ \mathbf{v}\end{bmatrix} = \begin{bmatrix}\mathbf{p}_0\\ \mathbf{v}_0\end{bmatrix} + \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & 0 & 0\\ 0 & 0 & \mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\begin{bmatrix}p_1\\ p_2\\ v_1\\ v_2\end{bmatrix}$$

$$\begin{bmatrix}p_1\\ p_2\\ v_1\\ v_2\end{bmatrix} = \begin{bmatrix}\mathbf{v}_1^t & 0\\ \mathbf{v}_2^t & 0\\ 0 & \mathbf{v}_1^t\\ 0 & \mathbf{v}_2^t\end{bmatrix}\begin{bmatrix}\mathbf{p}-\mathbf{p}_0\\ \mathbf{v}-\mathbf{v}_0\end{bmatrix}$$

12. Error Estimation

We must now estimate the error with which our system localises a line. For such a complex sensor a complete description of the error process is impossible. We will describe two ways to approximate it.

First we will consider an idealised situation where we are observing not lines, but sets of n collinear points. These are matched without error between left and right images.

The imaging process is assumed to produce equal uncorrelated errors of variance σ^2 in their left and right image X-coordinates and the left image Y coordinate, so that fitting a line by orthogonal regression in (X_L, X_R, Y_L)-space (disparity space) is optimal[10]. This then produces a centroid position error of variance σ^2/n and an angular error of variance $12\sigma^2/nl^2$ where l is the length of the line in disparity space.

In terms of local line coordinates in disparity space the error covariance is thus (1_2 is the 2×2 unit matrix)

$$S_{\text{disp}} = \frac{\sigma^2}{n}\begin{bmatrix}1_2 & 0\\ 0 & \frac{12}{l^2}1_2\end{bmatrix}$$

We must transform this result to world coordinates. If the Jacobian matrix of the map from disparity space to the world is J Then

$$\begin{bmatrix}\mathbf{p}-\mathbf{p}_0\\ (\mathbf{v}-\mathbf{v}_0)|J\mathbf{v}_0|\end{bmatrix} = \begin{bmatrix}J & 0\\ 0 & J\end{bmatrix}\begin{bmatrix}\mathbf{p}-\mathbf{p}_0\\ \mathbf{v}-\mathbf{v}_0\end{bmatrix}_{\text{disp}}$$

Using the transformation above the error covariance of the description in the world is thus

$$S = \begin{bmatrix}\Sigma & 0\\ 0 & \frac{12}{l^2|J\mathbf{v}_0|^2}\Sigma\end{bmatrix}$$

where

$$\Sigma_{ij} = \frac{\sigma^2}{n}\mathbf{v}_i^t JJ^t\mathbf{v}_j \qquad i,j = 1, 2$$

if $\mathbf{p}_0$ is taken as the projection of the disparity space centroid into the world.

The above idealisation is very unrealistic for two main reasons. Firstly the stereo matching of continuous lines mixes the horizontal errors with the vertical errors; for lines with making angles θ with the horizontal which are close to zero depth values are highly inaccurate (when $\theta = 0$ matching is impossible). A crude way of compensating for this is to multiply J by a matrix producing an expansion factor of $1/\sin\theta$ in depth before using the above formulae. Secondly the points detected on continuous lines are not randomly scattered about the line, but wander slowly from one side of the line to the other. This can be compensated for by replacing n by a smaller effective number of points on the line which counts these wanderings. Though crude, this model then captures most of the essential information about stereo errors.

As an alternative to the calculation of an *a priori* covariance we can try to estimate the covariance from our data. Fit a line $(\mathbf{p}_0, \mathbf{v}_0)$ to the data, the true line being The true line is $(\mathbf{p}, \mathbf{v})$ with local coordinates as above. Suppose one of the data points (x, y, z) projects to $(X, Y) = (x/z, y/z)$ in the left image. Let $\mathbf{a} = (X, Y, 1)^t$. The condition that this image point lies on the projection of the true line into the image is $\mathbf{p}\cdot(\mathbf{a}\times\mathbf{v}) = 0$ which linearises to

$$\mathbf{p}_0\cdot(\mathbf{a}\times\mathbf{v}_0) + p_1\,\mathbf{v}_1\cdot(\mathbf{a}\times\mathbf{v}_0) + p_2\,\mathbf{v}_2\cdot(\mathbf{a}\times\mathbf{v}_0) + v_1\,\mathbf{p}_0\cdot(\mathbf{a}\times\mathbf{v}_1)$$

If we assume no prior knowledge, then by walking down the string of points imposing this constraint in each image, we can simultaneously correct our initial line and build up its covariance matrix.

13. Geometrical Constraints

We would like a module embodying the theory above to be available to a geometrical reasoning system as a knowledge source aiding in the interpretation of a single stereo view. As the geometrical reasoning system hypotheses geometrical relationships (orthogonality, coincidence of lines etc.) the module is able to assign likelihoods to these hypotheses, and if they are accepted, to correct the data to be consistent with the current under-

standing of the geometry.

There are a very large number of possible relationships we might like to test and impose (for example that three lines form a rectangular trihedral vertex). Rather than try to include all useful constraints we can implement a vocabulary of pairwise constraints (orthogonality, incidence properties) in terms of which the more complex constraints can be expressed.

For example the condition on the state vector that two lines with directions v, v′ be orthogonal is

$$(\mathbf{v}_0 + \nu_1\mathbf{v}_1 + \nu_2\mathbf{v}_2)\cdot(\mathbf{v}_0' + \nu_1'\mathbf{v}_1' + \nu_2'\mathbf{v}_2') = 0$$

which linearises to

$$\mathbf{v}_0\cdot\mathbf{v}_0' + \nu_1\mathbf{v}_1\cdot\mathbf{v}_0' + \nu_2\mathbf{v}_2\cdot\mathbf{v}_0' + \nu_1'\mathbf{v}_0\cdot\mathbf{v}_1' + \nu_2'\mathbf{v}_0\cdot\mathbf{v}_2' = 0$$

The condition that the two lines intersect is

$$(\mathbf{p} - \mathbf{p}')\cdot(\mathbf{v}\times\mathbf{v}') = 0$$

which linearises to

$$(\mathbf{p}_0 - \mathbf{p}_0')\cdot(\mathbf{v}_0\times\mathbf{v}_0')$$
$$+ p_1\mathbf{v}_1\cdot(\mathbf{v}_0\times\mathbf{v}_0') + p_2\mathbf{v}_2\cdot(\mathbf{v}_0\times\mathbf{v}_0') - p_1'\mathbf{v}_1'\cdot(\mathbf{v}_0\times\mathbf{v}_0') - p_2'\mathbf{v}_2'\cdot(\mathbf{v}_0\times\mathbf{v}_0')$$
$$+ \nu_1(\mathbf{p}_0 - \mathbf{p}_0')\cdot(\mathbf{v}_1\times\mathbf{v}_0') + \nu_2(\mathbf{p}_0 - \mathbf{p}_0')\cdot(\mathbf{v}_2\times\mathbf{v}_0')$$
$$+ \nu_1'(\mathbf{p}_0 - \mathbf{p}_0')\cdot(\mathbf{v}_0\times\mathbf{v}_1') + \nu_2'(\mathbf{p}_0 - \mathbf{p}_0')\cdot(\mathbf{v}_0\times\mathbf{v}_2') = 0$$

These constraints would be sufficient to impose such composite hypotheses as, for example, that three lines form a rectangular trihedral vertex.

14. Multiple Views

In the case of multiple stereo views the primitive describing sensor calibration is the motion of the stereo rig relative to the background frame. The most convenient global description of such a rotation is as a rotation matrix R and a translation vector $\mathbf{t}$. Nearby rotations have the form

$$\mathbf{x} \to R\mathbf{x} + \mathbf{t} + \boldsymbol{\omega}\times\mathbf{x} + \boldsymbol{\tau}$$

and the pair of vectors $(\boldsymbol{\omega}, \boldsymbol{\tau})$ can be used as local coordinates for this type of primitive.

Suppose we want to work in the frame of our latest stereo view, and we have already built up a model with an associated covariance matrix from previous views. The matcher relates the old view to this and calculates an approximate transformation $(R_0, \mathbf{t}_0)$ between the two views, this is used to transform all the elements of the old model into the new frame. This transformation is inaccurate, so the 'calibration error' between views $(\boldsymbol{\omega}, \boldsymbol{\tau})$ is adjoined to the state vector of the model. For each pair of matched lines $(\mathbf{p},\mathbf{v}) \to (\mathbf{p}',\mathbf{v}')$ the constraint that they are related by the rigid motion $(R, \mathbf{t})$ is firstly that $\mathbf{p}$ moves onto the line $(\mathbf{p}',\mathbf{v}')$

$$(\mathbf{p}' - R\mathbf{p} - \mathbf{t}) - (\mathbf{p}' - R\mathbf{p} - \mathbf{t})\cdot\hat{\mathbf{v}}'\,\hat{\mathbf{v}}' = 0$$

(where $\hat{\mathbf{a}} = \mathbf{a}/|\mathbf{a}|$), and secondly that the direction vectors coincide

$$\hat{\mathbf{v}}' - R\hat{\mathbf{v}} = 0$$

Each of these represents only two independent constraints, which can be extracted by taking the scalar product with $\mathbf{v}_1'$ and $\mathbf{v}_2'$. The result linearises to

$$(\mathbf{p}_0' - \mathbf{p}_0)\cdot\mathbf{v}_1' + p_1' - p_1\mathbf{v}_1\cdot\mathbf{v}_1' - p_2\mathbf{v}_2\cdot\mathbf{v}_1'$$
$$- \nu_1'(\mathbf{p}_0' - \mathbf{p}_0)\cdot\mathbf{v}_0 - (\mathbf{p}_0\times\mathbf{v}_1')\cdot\boldsymbol{\omega} - \mathbf{v}_1'\cdot\boldsymbol{\tau} = 0$$

$$(\mathbf{p}_0' - \mathbf{p}_0)\cdot\mathbf{v}_2' + p_2' - p_1\mathbf{v}_1\cdot\mathbf{v}_2' - p_2\mathbf{v}_2\cdot\mathbf{v}_2'$$
$$- \nu_2'(\mathbf{p}_0' - \mathbf{p}_0)\cdot\mathbf{v}_0 - (\mathbf{p}_0\times\mathbf{v}_2')\cdot\boldsymbol{\omega} - \mathbf{v}_2'\cdot\boldsymbol{\tau} = 0$$

$$-\mathbf{v}_0\cdot\mathbf{v}_1' + \nu_1' - \nu_1\mathbf{v}_1\cdot\mathbf{v}_1' - \nu_2\mathbf{v}_2\cdot\mathbf{v}_1' - (\mathbf{v}_0\times\mathbf{v}_1')\cdot\boldsymbol{\omega} = 0$$
$$-\mathbf{v}_0\cdot\mathbf{v}_2' + \nu_2' - \nu_1\mathbf{v}_1\cdot\mathbf{v}_2' - \nu_2\mathbf{v}_2\cdot\mathbf{v}_2' - (\mathbf{v}_0\times\mathbf{v}_2')\cdot\boldsymbol{\omega} = 0$$

The correction to the new view of the line is adjoined (with initial value zero) to the state vector, and the constraints imposed. We can choose to keep either of the resulting descriptions, since we want to work in the frame of the latest view we keep the new description.

15. Discussion

It will be clear that the above scheme requires heavy computations and the storage of large amounts of data. In this form it will serve as a test-bed for investigating the importance of various factors in sensor combination. Its major advantages are that it loses no information, can deal with general geometrical structures in a natural way, and it is very convenient to introduce new types of measurement or constraint. A major complication is our insistence that many of the correlations (off diagonal terms in the covariance matrix) are of great importance; for example sensor calibration errors are correlated with all the measurements from that sensor, and their omission would make nonsense of any results, this is particularly important in stereo vision, where accurate calibration is often difficult.

A practical algorithm would need to avoid much of the calculation however. One way to approach this is discussed in Durrant-Whyte[4], where the calculation of the optimal estimator is arranged in such a way that it is clear when correlations are no longer contributing to the solution.

Another possible source of error is the linearisation we must make to keep the analysis tractable. We are dealing with a special case of the extended Kalman filter which is often very successful (see Gelb[13] for some examples), but is not guaranteed to be stable or optimal. Basically we must try to ensure that the linear approximation is always justified if we are to be safe. One can deal directly with the non-linear problem[11] but these methods are much more complex and less flexible. It is also worth noting that since we apply constraints only to linearised order, exact consistency is not achieved. We must relinearise about the new solution (keeping the old covariances of course) and recurse to an exact solution if required.

The above problem leads on to the question of robustness in the statistical sense. Gauss-Markov theory requires the assumption that a minimum variance estimator is sensible. This can break down very badly if the data has non-negligible probabilities for 'large' errors, so-called flyers. Robust statistical methods protect against these flyers (essentially samples from a second random process about which we have little information) in many ways. One way is to minimise some other quantity than the variance. Such methods can lead to much heavier computation. A

more direct method is to try to catch the flyers before they are processed, and there are principled ways of doing this when one has fairly large samples. We have a problem with sequential processing however, since by the time one detects an early flyer it will already have been merged into the model. In stereo vision extremal boundaries are one source of this problem. Not corresponding to real entities, they move between views, but for closely space views this movement may be of the same order of magnitude as the allowable errors in matching. There will have to be heuristics built in to a matcher to protect a module of the form we envisage from these problems. (Information on robust statistical methods can be found in Rey[12]).

Finally we present an example of the algorithm at work. Figure 1 shows a typical view of the widget generated by the WINSOM body modeller. Straight edges have been extracted from eight views circling above the object, matched and filtered by length and frequency of occurence, then combined to produce the rough skeleton model in Figure 2. The pairs of edges of this skeleton are then checked for closeness to intersection and perpendicularity (the two examples of constraints given above) and when this is significant the constraint is imposed optimally. There is one iteration about the new solution. The result is Figure 3 (where edges have been extended up to vertices). As an example of the rapidity of convergence let θ be the angle between the lines 7 and 10, and δ be their distance of closest approach. These lines are detected as a possible perpendicular vertex. The values of θ and δ at each iteration are:

Iteration	θ	δ
0	89.5	0.05
1	89.9994	0.001
2	89.99999	0.00002

References

1 **Pollard, S. B., Porrill, J., Mayhew, J. E. W., Frisby, J. P.,** Matching Geometrical Descriptions in 3-Space, *AIVRU Memo 022, 1986.*

2 **Crowley, J. L.,** Navigation for an Intelligent Mobile Robot, *IEEE J. Rob. Aut., 31-41, 1985.*

3 **Faugeras, O. D., Ayache N., Faverjon, B.,** Building Visual Maps by Combining Noisy Stereo Measurements, presented at *IEEE Robotics conference, San Francisco 1986.*

4 **Durrant-Whyte, H. F.,** Consistent Integration and Propagation of Disparate Sensor Observations, *Thesis, University of Pennsylvania, 1985*

5 **Albert, A.,** *Regression and the Moore-Penrose Pseudo-Inverse, Academic Press, New York, 1972.*

6 **Morrison, D. F.,** Multivariate Statistical Methods, *McGraw-Hill, 1976.*

7 **Mitra, S. K.,** Generalised inverses of matrices and applications to linear models, *Handbook of Statistics, North-Holland, 1980.*

8 **Canny, J. F.,** Finding Edges and Lines in Images, *MIT AI memo, 720, 1983.*

9 **Pollard, S. B., Mayhew, J. E. W., Frisby, J. P.,** PMF: A Stereo Correspondence Algorithm Using a Disparity Gradient Limit, *Perception, 14, 449-470, 1985.*

10 **Porrill, J., Pridmore, T. P., Mayhew, J. E. W., Frisby, J. P.,** Fitting Planes, Lines and Circles to Stereo Disparity Data, *AIVRU memo 017, 1986.*

11 **Tarantola, B., Valette, B.,** Generalised Nonlinear Inverse Problems Solved Using the Least Squares Criterion, *Rev. Geophys. & Space Phys, 20, 2, 219-232, 1982.*

12 **Rey, W. J.,** Introduction to Robust and Quasi-Robust Statistical Methods, *Springer-Verlag, 1983.*

13 **Gelb, A. (ed.),** Applied Optimal Estimation, *M. I. T. Press, 1974.*

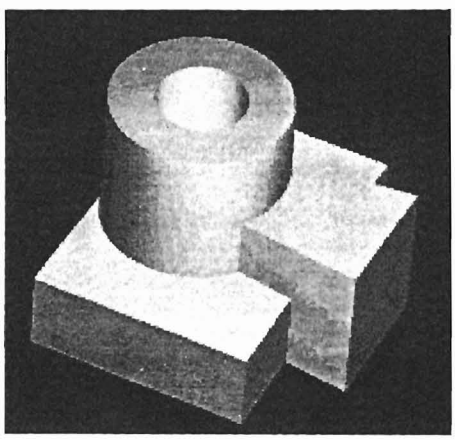

Figure 1. View of the widget.

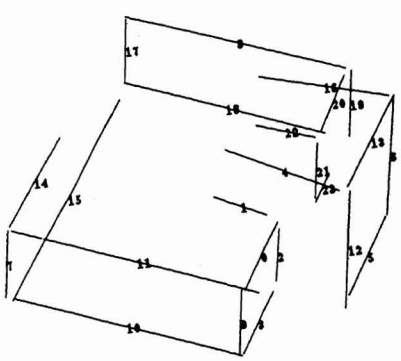

Figure 2. Skeleton model of the widget

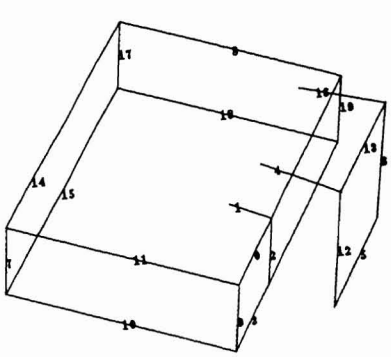

Figure 3. Widget after one Geomstat iteration