Journal of Mathematical Imaging and Vision, 5, 77–90 (1995).

Point Reconstruction from Noisy Images

MAGNUS ANDERSSON AND DEMETRIOS BETSIS* email:magnus.andersson@bion.kth.se
Computational Vision and Active Perception Laboratory (CVAP), Department of Numerical Analysis and Computing Science, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Abstract. In this paper we treat the problem of determining optimally (in the least-squares sense) the 3D coordinates of a point, given its noisy images formed by any number of cameras of known geometry. The optimality criterion is determined by the covariance matrices associated with the images of the point. The covariance matrices are not restricted to be positive definite but are allowed to be singular. Thus, image points constrained to lie along straight lines can be handled as well. Estimation of the covariance of the reconstructed point is provided.

The often appearing two-camera stereo case is treated in detail. It is shown in this case that, under reasonable conditions, the main step of the reconstruction reduces to finding the unique zero of a sixth degree polynomial in the interval $(0, 1)$.

Keywords. stereo, reconstruction, least-squares estimate, covariance matrix

1 Introduction

The reconstruction problem, when errors are present, has been studied by several authors. Blostein and Huang [1] and Rodrigues and Aggarwal [2] studied the effects of digitization errors in the images. However, these studies were limited to the special case of parallel camera geometry and were concerned mainly with the error analysis. Among recent works, the work of Deriche, Vaillant and Faugeras [3] should be mentioned. In their work reconstruction from line segments and end-points of line segments is studied. However, the covariances of the measurements are used only to determine the approximate covariance for the solution point, but do not enter into the functional to be minimized. Another approach, that has been proposed, is to use Kalman filtering, see e.g. Ayache and Faugeras [4]. However, a Kalman filter approach linearizes the problem and will therefore, in general, yield an approximate solution.

In this paper we have made the following assumptions:

- The geometries (location of the optical center and orientation of the axes) of N cameras are **known.**
- The N images of a point $p \in R^3$ have been identified and matched.
- The error of every image point is **normally distributed** with zero mean and known covariance matrix.

Under the above assumptions the following problems are solved:

- Compute the **least squares estimate** of the point p.
- Compute an estimate of its covariance matrix.

Since trinocular systems are common and furthermore in Photogrammetric applications points may appear in as many as 6 pictures, we have chosen to treat the problem for any number of images, N. However, the important case $N = 2$ is treated in detail. It is shown that the main step of the solution is reduced to finding the unique (under reasonable conditions) zero in the interval $(0, 1)$ of a polynomial in one variable.

The errors in the position of the images are assumed to be normally distributed with zero mean.

The solution presented has the following features. First, the least squares estimation of p is formulated as an **unconstrained** minimization problem, where the functional to be minimized is determined by the metrics associated with

*The authors are listed in random order.

the covariance matrices of the measured points. The covariance matrices are allowed to be singular, and thus linear constraints on the location of an images point can be incorporated in its covariance matrix. Secondly, the above formulation makes it possible, by using the implicit function theorem, to calculate an approximate covariance matrix for the solution point.

The paper is organized as follows:

- Section 2, Preliminaries and Notation, explains the basics.
- Section 3, Noise and Norms, is a general discussion about the metric defined by a symmetric positive (semi)definite matrix and the formulas it gives rise to.
- Section 4, Reconstruction, treats in detail the problem at hand. The general case is studied in subsection 4.1 and the case with rank deficient covariance matrices in 4.2. In 4.3 a plausible choice of the initial approximation necessary to solve numerically the problem is presented. The special two-camera case is studied in detail in 4.4. A short discussion about the plausibility of the solution is found in 4.5.

2 Preliminaries and Notation

We assume that images are formed according to the *pinhole camera* model with the image plane in front of the optical center, which is denoted by O. Points and their coordinates in R^3, with the exception of the optical centers, will be denoted by lower case letters, e.g. $p = (x, y, z)$, whereas their images and their coordinates in the projective plane(s) will be denoted by upper case letters, e.g. $P = (X, Y)$. The presence of at least two cameras makes it necessary to use subscripts or superscripts to indicate the camera or the coordinate system involved. If the optical center of a camera is located at a point with world coordinates T, the translation vector, and if the orientation of the optical axis of the camera (z-axis) and the x- and y-axes of the projective plane, is described by a 3×3 rotation matrix R then the transformations from the world (global) coordinate system, p, to the camera coordinate system, p^c, and vice versa are

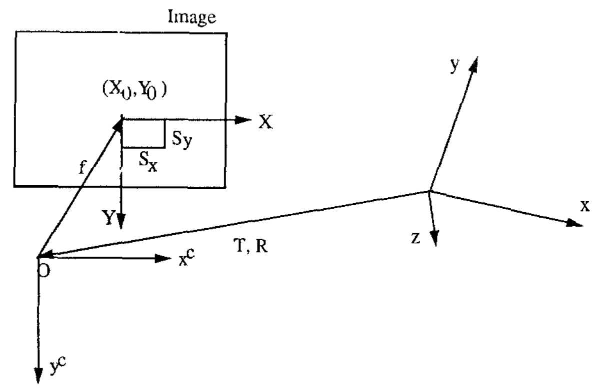

Fig. 1. The camera model.

described by

$$p^c = R^T(p - T) \qquad p = Rp^c + T \tag{1}$$

where the superscript T denotes transposition.

For a real camera the optical axis does not necessarily intersect the image plane at the origin of the image coordinate system, but at a point $P_0 = (X_0, Y_0)$. Finally the size of a pixel, in world coordinates, can vary in the X and Y directions. The size of a pixel in the X direction is S_X and the size of a pixel in the Y direction is S_Y. The distance from the *optical center*, O, to the image plane, the *focal length* is denoted by f.

Using

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} \frac{S_X}{f}(X - X_0) \\ \frac{S_Y}{f}(Y - Y_0) \end{pmatrix} \tag{2}$$

the image coordinates are transformed into a standard pinhole camera systems with $f = S_X = S_Y = 1$ and $X_0 = Y_0 = 0$, which will be used in the sequel.

3 Noise and Norms

It is assumed that the error associated with the image points obeys a *Gaussian distribution* with zero mean. This means that an image point, P, can be described by an *average* and a *covariance matrix*. The actual derivation of the covariance matrix can be a difficult problem, but here we assume that the covariance matrices have already been derived. The problem of reconstruction from image point correspondences when the error associated with each image point is the quantization error and therefore not normally distributed, has been treated elsewhere [1] [2]. However, there are many situations where the errors in the coordinates of image points come from other sources. In such cases, the distribution of the errors can often be approximated by a *Gaussian distribution.* Moreover, the treatment presented here, allows singular covariance matrices, i.e. image points that are constrained to lie on lines or even to be considered as exact!

3.1 *The Covariance Matrix*

The covariance matrix C plays a crucial role in the following. This matrix is *symmetric positive definite* or *semidefinite.* Semidefinite covariance matrices incorporate linear constraints and vice versa linear constraints can be incorporated in covariance matrices, see e.g. the monograph of Arthur Albert [5]. If p is a stochastic variable with mean value $\overline{p}$ and a semidefinite covariance matrix C then p is constrained to vary inside the linear variety $\{\overline{p} + \mathcal{R}(C)\}$, where $\mathcal{R}(C)$ denotes the range of C. E.g. the stochastic variable P

$$P = \begin{pmatrix} X \\ Y \end{pmatrix}, \text{ with } \overline{P} = \begin{pmatrix} A \\ B \end{pmatrix}$$

$$\text{and rank-one } C_P = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \quad (3)$$

implies that P is constrained to vary along the line $(\pm\sqrt{c}\ \sqrt{a})P = \pm\sqrt{c}A + \sqrt{a}B$ where the negative sign is chosen if $b \geq 0$, and the variance of P is $\sigma_P^2 = a + c$.

Conversely, linear constraints can be incorporated in covariance matrices. E.g. a point $P \in R^2$ constrained to lie on a line $a^T P = c$, $\|a\|_2 = 1$, with variance σ_P^2 along this line can be considered as a stochastic variable in the whole of R^2 with covariance matrix given by $C_P = \sigma_P^2(I - aa^T)$.

Trivially, a quantity known exactly can be considered as a stochastic variable with covariance matrix zero.

If y is a function of a stochastic variable x, i.e. $y = f(x)$, then the covariance matrix of y, can be approximated by

$$C_y \approx F C_x F^T \quad (4)$$

where F is the *Jacobian matrix* of f. The formula is exact if f is linear.

3.2 *Norms and Distances*

If the $n \times n$ covariance matrix C is of full rank i.e. positive definite, it defines an inner product and thereby a norm and a distance by

$$\langle x, y\rangle_C = x^T C^{-1} y, \quad \|x\|_C^2 = x^T C^{-1} x,$$
$$\varrho_C(x, y) = \|x - y\|_C \quad (5)$$

If the covariance matrix is rank deficient $(\dim(\mathcal{R}(C)) < n)$, we can still define a function, which strictly speaking is not a norm but it will serve the purpose of measuring distances, by using the pseudoinverse $C^\dagger$ instead of C^{-1}. By abusing the notation and the term we shall call it a norm in R^n.

$$\|x\|_C^2 = x^T C^\dagger x \quad \text{if } x \in \mathcal{R}(C)$$
$$= \infty \quad \text{otherwise} \quad (6)$$

It is well-known that, under the assumption $a^T C a \neq 0$, the point x on the hyperlane $a^T x = c$ nearest, in the C-norm, to a given point y is given by

$$x = y - \frac{a^T y - c}{a^T C a} C a \quad (7)$$

and that

$$\|x - y\|_C^2 = \frac{(a^T y - c)^2}{a^T C a} \quad (8)$$

Note that the inverse (or the pseudoinverse) of C does not enter the formulas. They are valid for definite C as well as semidefinite C. The only restriction is that $a^T C a \neq 0$.

4 Reconstruction

In this section we will determine the functional to be minimized, first in the case of full rank covariance matrices, section 4.1, and then describe the necessary modifications in order to deal with singular covariance matrices, section 4.2. A method to find an initial approximation to the solution is presented in section 4.3. In the two camera, the solution is simplified into a minimization of a functional in one variable, section 4.4.

4.1 Full Rank Covariance Matrices

A point $p = (x, y, z)^T$ is projected through a standard pinhole camera with extrinsic parameters R, T to (X_p, Y_p), which is given by

$$\begin{pmatrix} X_p \\ Y_p \\ 1 \end{pmatrix} = \frac{1}{\rho^T(p-T)} R^T(p-T) \tag{9}$$

where ρ is the third column of R.

Given the N *measured* images of p, the N lines passing through the image points and the corresponding optical centers will not, in general, pass through the same point. Our assumption about Gaussian distribution implies that the best p is the one that minimizes the sum of the squares of the corrections of the images of the point (measured in the norms induced by the associated covariance matrices).

The natural way of defining a distance between the measured point $M = (A, B)$ and the corrected point $P = (X_p, Y_p)$, which is the projection of the solution point, is by using the covariance matrix of M, C_M, i.e.

$$d^2 = \|P - M\|^2_{C_M} \tag{10}$$

In order to determine the least squares solution p, the sum of the squares of the distances must be minimized. By using the expression from equation (9) it can be shown, after some manipulations, that the squared distance can be written as

$$d^2 = \frac{(p-T)^T Q(p-T)}{[\rho^T(p-T)]^2} = \frac{(p-T)^T Q(p-T)}{(p-T)^T W(p-T)} \tag{11}$$

where $W = \rho\rho^T$ is a rank-one 3×3 matrix and where Q is a 3×3 symmetric positive semidefinite matrix, the elements of which depend on the parameters of the camera, the coordinates A, B and the covariance matrix C_M.

$$Q = R \begin{bmatrix} G & a \\ a^T & b \end{bmatrix} R^T, \quad \text{where} \quad G = C_M^{-1},$$

$$a = -G \begin{bmatrix} A \\ B \end{bmatrix}, \quad b = [A \quad B] G \begin{bmatrix} A \\ B \end{bmatrix} \tag{12}$$

Q is obviously semidefinite since the distance must be zero, when the projection of p coincides with the image point.

The functional to be minimized is

$$f(p) = \sum_{i=1}^{N} \frac{(p-T_i)^T Q_i (p-T_i)}{(p-T_i)^T W_i (p-T_i)} = \sum_{i=1}^{N} d_i^2 \tag{13}$$

where the subscript i denotes the quantities associated with the i:th camera system, and the minimization is to be done over all of R^3. In spite of the fact that the behavior of each summand of f is perfectly understood, minimizing f is not at all trivial but can be done by traditional optimization techniques, e.g. Gauss-Newton direction methods.

Since the individual components of f are smooth functions, except for p lying in a plane through the optical center parallel to the projective plane, a necessary condition at the minimum point p is

$$\frac{\partial f}{\partial p} = \sum_{i=1}^{N} \frac{2}{(p-T_i)^T W_i (p-T_i)} \times [Q_i(p-T_i) - d_i^2 W_i(p-T_i)] = 0 \tag{14}$$

The above formula implicitly defines the solution p as a function of the measured points in the images. Let $Z = (A_1, B_1, A_2, \ldots, B_N)^T$, where (A_1, B_1), $(A_2, B_2), \ldots$ are the measured image points. Then, by applying the implicit function theorem we get the derivative of p with respect to Z as

$$\frac{\partial p}{\partial Z} = -\left[\frac{\partial^2 f}{\partial p\, \partial p}\right]^{-1} \left[\frac{\partial^2 f}{\partial p\, \partial Z}\right] \tag{15}$$

where $\partial p/\partial Z$ is a $3 \times 2N$ matrix, $\partial^2 f/\partial p \partial p$ is a 3×3 matrix (the Hessian of f with respect to p) and $\partial^2 f/\partial p \partial Z$ is a $3 \times 2N$ matrix.

Computation of $\partial p/\partial Z$ at the solution point provides an approximate covariance matrix of p by means of (4).

$$C_p = \left[\frac{\partial p}{\partial Z}\right] C_Z \left[\frac{\partial p}{\partial Z}\right]^T$$

$$= \left[\frac{\partial p}{\partial Z}\right] \begin{bmatrix} C_1 & 0 & 0 & \cdots \\ 0 & C_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \\ \cdots & 0 & 0 & C_N \end{bmatrix} \left[\frac{\partial p}{\partial Z}\right]^T \quad (16)$$

4.2 Rank Deficient Covariance Matrices

If an image point has a rank-deficient covariance matrix this implies that the point must lie on a line (rank one) or is fully known (rank 0). Rank-deficient covariance matrices might not be very common in practice, and are included here only to make the analysis complete.

When one or more of the covariance matrices are rank deficient the functional to be minimized (13) must be suitably modified. Besides that, the minimization has to be carried out under constraints. However, these constraints are linear. Therefore, an easily derived linear transformation yields an unconstrained minimization problem in fewer variables.

There are three different cases:

- The covariance matrix C associated with the given image point is of full rank. In this case the corresponding summand of f in (13) remains unchanged and no constraint is issued.
- The covariance matrix is a rank-one matrix. In this case the corresponding summand of f in (13) must be modified. The inverse of the covariance matrix must be replaced by the pseudoinverse which has a particularly simple form. In our case

$$C = \begin{bmatrix} a & b \\ b & c \end{bmatrix}, \quad C \text{ is rank-one}$$

$$\Rightarrow C^{\dagger} = \frac{1}{(a+c)^2} C \quad (17)$$

In addition, a constraint must be added. In view of (3) and with the notation used there, it is easily seen that p must satisfy

$$q^T p = q^T T$$

$$\text{where} \quad q = R \begin{pmatrix} \pm\sqrt{c} \\ \sqrt{a} \\ \mp\sqrt{c}A - \sqrt{a}B \end{pmatrix}. \quad (18)$$

The minus sign in the first component, $\pm\sqrt{c}$, (matched with the plus sign in the third component), is chosen when $b \geq 0$. R, T are the extrinsic parameters of the pertinent camera.

- The covariance matrix is zero. This is an unusual case and it is considered only for the sake of completeness. The corresponding summand of f is eliminated and two constraints are added. The constraints describe that the point p must lie on the line joining the optical center and the image point. The constraints can be chosen as any two planes (not identical) containing the image point and the optical center. An easy choice

$$q_1^T p = q_1^T T \quad q_1 = R \begin{pmatrix} 0 \\ 1 \\ -B \end{pmatrix},$$

$$\text{and} \quad q_2^T p = q_2^T T \quad q_2 = R \begin{pmatrix} 1 \\ 0 \\ -A \end{pmatrix} \quad (19)$$

Each camera will yield 0, 1 or 2 constraints of the type $q^T p = c$. The resulting m constraints can be collected to

$$Qp = h, \quad \text{where} \quad Q \quad \text{is } m \times 3 \quad (20)$$

The singular value decomposition of Q gives

$$Q = U\Sigma V^T \Rightarrow \Sigma V^T p = U^T h = h' \quad (21)$$

$$\Sigma p' = h', \quad \text{where} \quad p' = V^T p \quad (22)$$

Constraints on p are thus transformed into constraints on p'. For each nonzero σ_i the corresponding component of $p' = (x', y', z')$ has a fixed value, $p'_i = h'_i/\sigma_i$. For the σ_i values that are zero the corresponding component of p' is free to vary, unless h'_i is not zero, in which case the constraints are inconsistent. As a result of this transformation, we know how many (the rank of the matrix Q) of the first components of p' are fixed (and their values) and how many of the last components of p' vary freely over all R. So p' can be divided into $p' = (p'_k, p'_v)$ where p'_k are constant(s) and p'_v are variables. Every such p' gives rise to a point $p = Vp'$, the projection of which into any of the cameras will lie at a finite distance from the corresponding measured point, i.e.

$$d^2 = \|P - M\|^2_{C_M} = (P - M)^T C^\dagger_M (P - M) \tag{23}$$

The equation (11) can now be written as

$$\begin{aligned} d^2 &= \frac{(p - T)^T Q (p - T)}{(p - T)^T W (p - T)} \\ &= \frac{(Vp' - T)^T Q (Vp' - T)}{(Vp' - T)^T W (Vp' - T)} \\ &= \frac{(p' - T')^T V^T Q V (p' - T')}{(p' - T')^T V^T W V (p' - T')} \\ &= \frac{(p' - T')^T Q' (p' - T')}{(p' - T')^T W' (p' - T')} \end{aligned} \tag{24}$$

where $T' = V^T T$ and $W' = V^T W V$. Furthermore

$$Q' = V^T R \begin{bmatrix} C^\dagger_M & -C^\dagger_M M \\ -(C^\dagger_M M)^T & M^T C^\dagger_M M \end{bmatrix} R^T V \tag{25}$$

The functional to be minimized is

$$g(p'_v) = \sum_{i=1}^{N} \frac{(p' - T'_i)^T Q'_i (p' - T'_i)}{(p' - T'_i)^T W'_i (p' - T'_i)} = \sum_{i=1}^{N} d_i^2 \tag{26}$$

The minimization is unrestricted and we can proceed exactly as before. At the minimum point the derivatives with respect to p'_v should be zero and we have

$$\frac{\partial g}{\partial p'_v} = 0 \tag{27}$$

The equation (27) implicitly defines p'_v as a function of M, and exactly as before the covariance of p'_v can be computed. To get the covariance of p' we simply border the matrix by zeroes to get a 3×3 covariance matrix. This gives

$$C_{p'} = \begin{bmatrix} 0 & 0 \\ 0 & C_{p'_v} \end{bmatrix} \tag{28}$$

and since $p = Vp'$ we have

$$C_p = V C_{p'} V^T \tag{29}$$

4.3 Initial Approximation for the Point Reconstruction Problem

A plausible initial approximation to the reconstructed point is the point p which is closest to the rays defined by the optical centers of each camera and the corresponding image points. These rays, l_i, can be written

$$l_i = O_i + \beta(P_i - O_i) = s_i + \alpha t_i, \quad \alpha, \beta \in R \tag{30}$$

where

$$s_i^T t_i = 0, \qquad t_i^T t_i = 1 \tag{31}$$

This representation, which is unique up to the sign of t_i, makes the solution easier. Given a point p in R^3, the squared minimum distance from this point to the line, $s + \alpha t$ is

$$\begin{aligned} d^2 &= \|p - s\|^2 - (t^T p)^2 \\ &= p^T p - 2 s^T p + s^T s - (t^T p)^2 \end{aligned} \tag{32}$$

Given N cameras (and therefore N lines) the problem is to find the point p minimizing the sum of the squared distances to these lines, i.e.

$$\begin{aligned} &\min_{p \in R^3} \sum_{i=1}^{N} d_i^2 \\ &= \min_{p \in R^3} \sum_{i=1}^{N} p^T p - 2 s_i^T p + s_i^T s_i - (t_i^T p)^2 \end{aligned} \tag{33}$$

At the minimum the derivative with respect to p should be zero, i.e.

$$\frac{\partial \sum_{i=1}^{N} d_i^2}{\partial p} = \sum_{i=1}^{N} (2p - 2s_i - 2(t_i^T p) t_i) = 0 \tag{34}$$

which can be rewritten as

$$p - \frac{TT^T}{N}p = \left(I - \frac{TT^T}{N}\right)p = \frac{1}{N}\sum_{i=1}^{N} s_i = \bar{s},$$
$$T = [t_1 t_2 \cdots t_N] \qquad (35)$$

The matrix

$$A = I - \frac{1}{N}TT^T \qquad (36)$$

is nonsingular unless

$$t_1 = \pm t_2 = \cdots = \pm t_N \qquad (37)$$

which implies that all rays l_i are parallel. This can be seen as follows: TT^T is a symmetric positive semidefine 3×3 matrix. As such its eigenvalues are 3 nonnegative real numbers, the sum of which is

$$\sum_{i=1}^{N} \lambda_i(TT^T) = \mathrm{tr}(TT^T) = \mathrm{tr}(T^T T)$$
$$= \sum_{i=1}^{N} t_i^T t_i = N. \qquad (38)$$

Since the eigenvalues of A are

$$1 - \frac{1}{N}\lambda_i(TT^T), \qquad i = 1,2,3 \qquad (39)$$

the matrix A is singular if and only if

$$\lambda_1(TT^T) = N, \quad \lambda_2(TT^T) = \lambda_3(TT^T) = 0, \ (40)$$

i.e. TT^T is a rank one matrix. Since $\mathrm{rank}(B) = \mathrm{rank}(BB^T) = \mathrm{rank}(B^T B)$ for all matrices B, T must be a rank one matrix. Since the t_i:s are unit vectors, equation (37) follows. An initial approximation to the reconstructed point is therefore given by

$$p = \left(I - \frac{TT^T}{N}\right)^{-1} \bar{s} \qquad (41)$$

except when all t_i are parallel. In the latter case it is easy to prove that t_1 spans the nullspace of A. This together with the fact that $\bar{s}$ is orthogonal to t_1, gives the initial approximation

$$p = \bar{s} + \gamma t_1, \quad \forall \gamma \in R \qquad (42)$$

However, this is an unusual case, and should normally not occur.

4.4 *Special Solution for the Two-Camera Case*

Very often the reconstruction problem is solved for two-cameras, i.e. $N = 2$. Formula (14) would yield a system of 3 polynomial equations of degree 6 in 3 variables, the coordinates of p. However, in this special case, the problem can be solved as follows: Consider the one parameter family $\{\pi_\lambda, \lambda \in R\}$ of epipolar planes, i.e. planes containing the optical centers. The parameterization is chosen so that π_0 corresponds to the plane of the family containing the left image point and π_1 to the plane containing the right image point. For each epipolar plane it is possible to find a best reconstruction point as follows. Each epipolar plane defines two epipolar lines, i.e the intersection lines of the epipolar plane with the two image planes. For each of the two epipolar lines, the point on the line that is closest to the image point, as well as its distance from it can be determined. The two points define two rays, the intersection of which is the best reconstructed point for the chosen epipolar plane. This implies that it suffices to determine the epipolar plane on which the best reconstruction point lies, i.e. to determine the best λ (a one-variable problem) instead of determining the best p (a three-variable problem). It turns out the functional to be minimized has the form

$$f(\lambda) = \frac{(1-\lambda)^2}{r(\lambda)} + \frac{\lambda^2}{q(\lambda)} = f_L(\lambda) + f_R(\lambda) \quad (43)$$

where r and q are second degree polynomials, the coefficients of which depend on the image coordinates, the covariance matrices and the camera geometry. Minimizing f leads to finding the roots of a sixth degree polynomial. It seems obvious that the best reconstructed point will lie on an epipolar plane passing between the two image points, i.e. $0 \le \lambda \le 1$. However, due to strange camera geometries and/or special covariance matrices the solution can actually lie outside this interval. Below it will be shown that under reasonable conditions there is a unique zero in the above interval.

4.4.1 Deriving the Functional. Instead of first finding the epipolar lines and then calculating the distances on the image planes the formulas

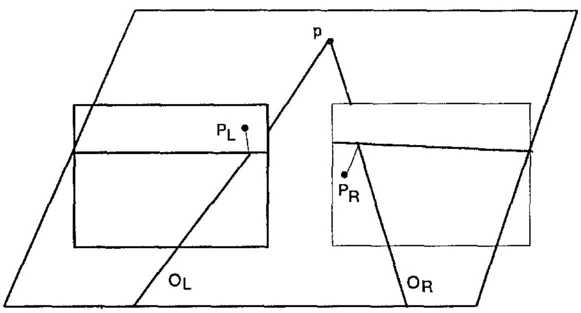

Fig. 2. Determination of the best point on a given epipolar plane. Note that due to the distance measure using the covariance matrices the distances are not measured orthogonal to the epipolar lines in the Euclidean sense, but in the metric induced by the covariance matrices.

(7) and (8) allow us to treat the problem directly in R^3.

The left image point p_L can be considered as a point in R^3 with coordinates expressed in the left camera system $(X_L, Y_L, 1)^T$. The associated 2×2 covariance matrix C_L must be suitably modified to a 3×3 matrix, by bordering it with zeroes. In the left camera coordinate system the point and its covariance matrix are given by

$$p_L^L = p_L = \begin{pmatrix} X_L \\ Y_L \\ 1 \end{pmatrix} \qquad G_L = \begin{bmatrix} C_L & 0 \\ 0 & 0 \end{bmatrix}$$

Similarly, in the right camera coordinate system the point p_R and its covariance matrix are given by

$$p_R^R = p_R = \begin{pmatrix} X_R \\ Y_R \\ 1 \end{pmatrix} \qquad G_R = \begin{bmatrix} C_R & 0 \\ 0 & 0 \end{bmatrix}$$

Choosing, for definiteness, to work in the left camera system, the coordinates of the point p_R^L are given by

$$p_R^L = M^T \begin{pmatrix} X_R \\ Y_R \\ 1 \end{pmatrix} + R_L^T(T_R - T_L) \quad \text{where} \quad M = R_R^T R_L$$

and its associated covariance matrix must be modified to $M^T G_R M$.

An epipolar plane, π_λ, contains O_L, i.e. the origin of the left camera coordinate system, the optical center of the right camera O_R, with coordinates $R_L^T(T_R - T_L)$, and a point lying on the line passing through p_L and p_R, say $p_\lambda = \lambda p_L^L + (1 - \lambda) p_R^L$ for some $\lambda \in R$. The normal of π_λ is then $a_\lambda = p_\lambda \times O_R$. The distance from p_L to π_λ is given by equation (8). The numerator is given by

$$\begin{aligned}(a_\lambda^T p_L^L)^2 &= [p_\lambda, O_R, p_L^L]^2 \\ &= [\lambda p_L^L + (1-\lambda) p_R^L, O_R, p_L^L]^2 \\ &= (1-\lambda)^2 [p_R^L, O_R, p_L^L]^2\end{aligned}$$

Similarly, the numerator of the distance from p_R to π_λ is

$$(a_\lambda^T p_R^L)^2 = \lambda^2 [p_R^L, O_R, p_L^L]^2$$

The coordinates of the normal a_λ are given by

$$\begin{aligned}a_\lambda &= [\lambda p_L + (1-\lambda)(M^T p_R + R_L^T(T_R - T_L))] \\ &\quad \times R_L^T(T_R - T_L) \qquad (44) \\ &= [\lambda p_L + (1-\lambda) M^T p_R] \times R_L^T(T_R - T_L) \quad (45) \\ &= R_L^T[(\lambda R_L p_L + (1-\lambda) R_R p_R) \\ &\quad \times (T_R - T_L)] = R_L^T b_\lambda\end{aligned}$$

where

$$b_\lambda = (\lambda R_L p_L + (1-\lambda) R_R p_R) \times (T_R - T_L)$$

Thus the objective is to find that λ, that minimizes the functional

$$f(\lambda) = [p_L, p_R, O_R]^2 \times \left(\frac{(1-\lambda)^2}{b_\lambda^T R_L G_L R_L^T b_\lambda} + \frac{\lambda^2}{b_\lambda^T R_R G_R R_R^T b_\lambda} \right)$$

If the image points lie in a epipolar plane then

$$[p_L, p_R, O_R] = 0 \quad \text{i.e.} \quad f(\lambda) = 0$$

and there is noting to minimize. The point p can be found as the intersection of the two coplanar rays defined by the image points and the corresponding optical centers. If $[p_L, p_R, O_R] \neq 0$ then we have to minimize

$$\begin{aligned}f(\lambda) &= \frac{(1-\lambda)^2}{b_\lambda^T R_L G_L R_L^T b_\lambda} + \frac{\lambda^2}{b_\lambda^T R_R G_R R_R^T b_\lambda} \\ &= \frac{(1-\lambda)^2}{r(\lambda)} + \frac{\lambda^2}{q(\lambda)} = f_L(\lambda) + f_R(\lambda) \quad (46)\end{aligned}$$

Thus the main step of the reconstruction problem is solved by minimizing a one-variable functional. Once λ is determined, the points on π_λ closest to the image points are determined by (7), and the intersection of the two rays can be trivially determined.

The quantities involved in the computation of the denominators, beside the covariance matrices, are

$$\alpha = M^T \begin{pmatrix} X_R \\ Y_R \\ 1 \end{pmatrix}, \; \beta = M \begin{pmatrix} X_L \\ Y_L \\ 1 \end{pmatrix},$$
$$\gamma = R_L^T(T_R - T_L), \; \delta = R_R^T(T_R - T_L) \quad (47)$$

We have that the denominator of f_L is

$$r(\lambda) = (\alpha \times \gamma + \lambda(p_L \times \gamma - \alpha \times \gamma))^T \times G_L(\alpha \times \gamma + \lambda(p_L \times \gamma - \alpha \times \gamma)) \quad (48)$$

a second degree polynomial, the coefficients of which can be expressed in terms of the following constants

$$\begin{aligned} A &= (\alpha \times \gamma)^T G_L(\alpha \times \gamma) \\ B &= (p_L \times \gamma)^T G_L(\alpha \times \gamma) \\ C &= (p_L \times \gamma)^T G_L(p_L \times \gamma). \end{aligned} \quad (49)$$

We get

$$f_L(\lambda) = \frac{(1-\lambda)^2}{r(\lambda)} = \frac{(1-\lambda)^2}{A + 2(B-A)\lambda + (C - 2B + A)\lambda^2} \quad (50)$$

Note that A and C are the G_L-norms (in the extended meaning) of the vectors $\alpha \times \gamma$ and $p_L \times \gamma$ respectively, and that B is their G_L-inner-product. It is trivial to see that the Cauchy-Schwartz inequality remains valid in this extended meaning of norms and inner-products. The Cauchy-Schwartz inequality gives $B^2 \leq AC$, which simply states that the discriminant of the denominator is nonpositive, i.e. the denominator is nonnegative.

Similarly the right term of the objective functional is

$$f_R(\lambda) = \frac{\lambda^2}{q(\lambda)} = \frac{\lambda^2}{D + 2(E-D)\lambda + (F - 2E + D)\lambda^2} \quad (51)$$

where the constants appearing in the denominator are given by

$$\begin{aligned} D &= (p_R \times \delta)^T G_R(p_R \times \delta) \\ E &= (p_R \times \delta)^T G_R(\beta \times \delta) \\ F &= (\beta \times \delta)^T G_R(\beta \times \delta) \end{aligned} \quad (52)$$

The critical points of $f(\lambda)$ are the zeroes of

$$f'(\lambda) = \frac{-(1-\lambda)q(\lambda)^2(2r(\lambda)+(1-\lambda)r'(\lambda))+\lambda r(\lambda)^2(2q(\lambda)-\lambda q'(\lambda))}{r(\lambda)^2 q(\lambda)^2} \quad (53)$$

Computing the zeroes of f' can be carried out by using any polynomial solver applied on the numerator (which is at most of degree 6). There are some special geometric configurations that are worth noting.

1. The normal to an image plane is orthogonal to the line connecting the optical centers. For definiteness, let us suppose that the normal of the left image plane (the 3rd column of R_L) is orthogonal to $T_R - T_L$. This implies that $\gamma_3 = 0$. In this case f_L reduces to a perfect square, and the polynomial the zero of which must be computed is at most of 5th degree. The same conclusion is reached if $\delta_3 = 0$. If both $\gamma_3 = \delta_3 = 0$ then both f_L and f_R are perfect squares and the degree of the polynomial reduces to 4.
2. Both normals are orthogonal to the line connecting the optical centers and parallel to each other. In this case we have in addition that $m_{33} = 1$ and the other elements of the 3rd row and column of M are zeroes. Thus we have in addition $\alpha_3 = \beta_3 = 1$ and therefore both denominators reduce to constants and the minimization is trivial.

To find all the roots of the polynomial can be time consuming and therefore it will be shown in the next section that, *under reasonable conditions*, the zero that provides the *global* minimum of f is a zero in the interval (0, 1), which, moreover, is the **unique** zero in this interval.

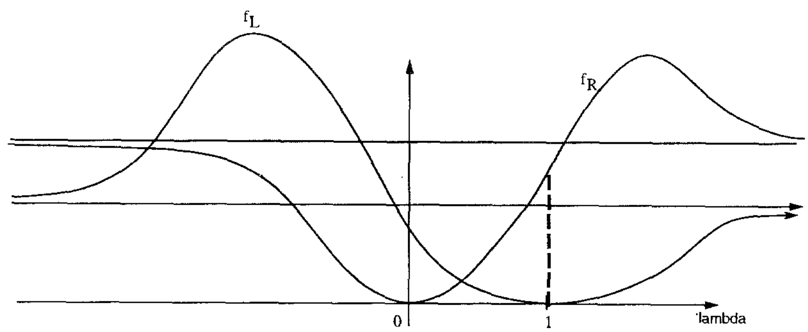

Fig. 3. General sketch of the distance functions. Note that the maxima can occur on either side of the minima.

4.4.2 Sufficient Conditions for a Global Minimum of f in $(0,1)$. Instead of treating f we shall treat the two summands f_L and f_R separately. Figure 3 provides a qualitative sketch of f_L and f_R.

We have

$$f_L(0) = \frac{1}{A} \quad f_R(1) = \frac{1}{F}$$
$$f_L(\infty) = \lim_{\lambda\to\infty} f_L(\lambda) = \frac{1}{C - 2B + A}$$
$$f_R(\infty) = \lim_{\lambda\to\infty} f_R(\lambda) = \frac{1}{F - 2E + D}$$
$$f_L(0) < f_L(\infty) \Leftrightarrow 2B - C > 0$$
$$f_R(1) < f_R(\infty) \Leftrightarrow 2E - D > 0$$

By using elementary arguments we shall prove that the two conditions, $2B - C > 0$ and $2E - D > 0$, confine the global minimum of f to the interval (0, 1). We also have

$$f'_L(\lambda) = \frac{2}{r(\lambda)^2}(\lambda - 1)(B + (C - B)\lambda)$$

where $\lambda = 1$ is a zero and $\lambda = \frac{B}{B-C}$ is either a zero or a pole.

$$f'_R(\lambda) = \frac{2}{q(\lambda)^2}(\lambda(D + (E - D)\lambda))$$

where $\lambda = 0$ is a zero and $\lambda = \frac{D}{D-E}$ is either a zero of a pole.

In addition we have

$$f''_L(\lambda) = \frac{2s(\lambda)}{r(\lambda)^3}$$

where

$$\begin{aligned} s(\lambda) = {} & (4B^2 - 2AB - AC) \\ & + 6(AB + BC - 2B^2)\lambda \\ & - 3(2B - C)(C - 2B + A)\lambda^2 \\ & + 2(B - C)(C - 2B + A)\lambda^3 \end{aligned} \tag{54}$$

$$f''_L(1) = \tfrac{2}{C} \qquad f''_L(0) = \frac{2(4B^2-2AB-AC)}{A^3}$$
$$f''_R(\lambda) = \frac{2(D^2-3D(F-2E+D)\lambda^2-2(E-D)(F-2E+D)\lambda^3)}{q(\lambda)^3}$$
$$f''_R(0) = \tfrac{2}{D} \qquad f''_R(1) = \frac{2(4E^2-2EF-DF)}{F^3}$$

i.e. $f''_L(1)$ and $f''_R(0)$ are positive, whereas the signs of $f''_L(0)$ and $f''_R(1)$ coincide with those of $4B^2 - 2AB - AC$ and $4E^2 - 2EF - DF$ respectively.

Let us examine f_L. We have

PROPOSITION 1. *The condition*

$$2B - C > 0 \tag{55}$$

implies that f achieves its global minimum at a point $\lambda^* \geq 0$. Moreover, the condition

$$4B^2 - 2AB - AC > 0 \tag{56}$$

implies that $f''_L(\lambda) > 0$ for all $\lambda \in (0,1)$.

Proof. We have the following three cases:

1. The discriminant of r negative, i.e. $B^2 - AC < 0$.
This implies that f_L is bounded. The derivative of f_L has two zeroes. The zero at $\lambda = \frac{B}{B-C}$ corresponds to the maximum of

f_L, whereas the zero at $\lambda = 1$ corresponds to the minimum. We also have

$$\lim_{\lambda \to \pm\infty} f_L'(\lambda) = 0 \tag{57}$$

The condition $2B - C > 0$ implies that $B > \frac{C}{2}$ and this places the number $\frac{B}{B-C}$, where f_L achieves its maximum, outside the interval $[-1, 1]$.
Moreover, B must be positive and therefore

$$f_L'(0) = -\frac{2B}{A^2} < 0$$

Let us examine the two possible locations of the maximum of f_L and their implications.

- $\lambda = \frac{B}{B-C} > 1$.

  In the interval $(-\infty, 0), f_L$ is strictly decreasing from $f_L(\infty)$ to $f_L(0)$. Therefore, $f_L(\lambda) > f_L(0)$ for all $\lambda \in (-\infty, 0)$. Since f_R is nonnegative and $f_R(0) = 0$ we get that $f(\lambda) > f(0)$ for all $\lambda \in (-\infty, 0)$, i.e. f achieves its global minimum at some point in the interval $(0, \infty)$. The point $\lambda = 0$ is excluded since $f'(0) = -\frac{2B}{A} < 0$.

  By combining (57) with the two zeroes of f_L', it is obvious that the three zeroes $z_1 < z_2 < z_3$ of f_L'' are located as follows

  $$-\infty < z_1 < 1 < z_2 < \frac{B}{B-C} < z_3 < \infty$$

  Since $f''(1) > 0$ the condition $f_L''(0) > 0$, i.e. $4B^2 - 2AB - AC > 0$, guarantees that $z_1 < 0$ and that $f_L''(\lambda) > 0$ for all $\lambda \in (0, 1)$.

- $\lambda = \frac{B}{B-C} < -1$.

  In the interval $(-\infty, 0), f_L$ is strictly increasing from $f_L(\infty)$ to achieve its maximum at $\frac{B}{B-C}$ and then strictly decreases to $f_L(0)$. Therefore, $f_L(\lambda) > f_L(0)$ for all $\lambda \in (-\infty, 0)$. With the same reasoning as above we see that f achieves its global minimum at some point in the interval $(0, \infty)$.

  The three zeroes $z_1 < z_2 < z_3$ of f'' are located as follows

  $$-\infty < z_1 < \frac{B}{B-C} < z_2 < 1 < z_3 < \infty$$

  Since $f''(1) > 0$ the condition $f_L''(0) > 0$ guarantees that $z_2 < 0$ and that $f_L''(\lambda) > 0$ for all $\lambda \in (0, 1)$.

2. The discriminant equal to zero, i.e. $B^2 - AC = 0$ and $A \neq C$.

In this case the denominator of f_L is a perfect square which does not reduce to a constant. The point $\frac{B}{B-C}$ is a pole of f_L and its derivatives. We have

$$f_L(\lambda) = \frac{(1-\lambda)^2}{(C - 2B + A)\left(\lambda - \frac{B}{B-C}\right)^2}$$

$$f_L'(\lambda) = \frac{2C}{(C-B)(C-2B+A)} \times \frac{\lambda - 1}{\left(\lambda - \frac{B}{B-C}\right)^3}$$

$$f_L''(\lambda) = \frac{-4C}{(C-B)(C-2B+A)} \times \frac{\lambda - \frac{2B-3C}{2B-2C}}{\left(\lambda - \frac{B}{B-C}\right)^4}$$

The reasoning, concerning the location of the global minimum of f is analogous to that of the previous case. The word maximum is replaced by infinity. There is only one zero of f_L''. It is obvious that the second condition guarantees that f_L'' is positive throughout the whole interval $[0, 1]$.

3. The discriminant equal to zero, i.e. $B^2 - AC = 0$ and $A = C$.

   In this case the denominator reduces to a constant. No discussion required. □

Examining f_R yields in an almost identical way

PROPOSITION 2. *The condition*

$$2E - D > 0 \tag{58}$$

implies that f achieves its global minimum at a point $\lambda^* \leq 1$. Moreover, the condition

$$4E^2 - 2EF - DF > 0 \tag{59}$$

implies that $f_R''(\lambda) > 0$ for all $\lambda \in (0, 1)$.

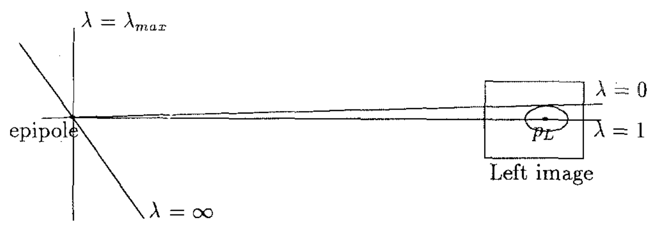

Fig. 4. The normal situation. The distance between the point p_L and the epipolar line for $\lambda = 0$ is much smaller than the distance to the epipolar line for $\lambda = \infty$ and much smaller than the distance to the epipole. Note that the distances are measured using the covariance matrices. This implies that all points on the drawn ellipse lie at the same distance from the point p_L, which is the center of the ellipse.

Combining the two propositions yields that

1. The conditions $2B - C > 0$ and $2E - D > 0$ imply that f achieves its global minimum at a point $\lambda^* \in (0, 1)$.
2. The conditions $4B^2 - 2AB - AC > 0$ and $4E^2 - 2EF - DF > 0$ imply that the second derivatives of f_L and f_R are positive throughout $(0, 1)$, and therefore their sum f is strictly convex in $(0, 1)$. This implies that it has only one minimum in this interval. λ^* can be found by computing the zero of a polynomial of degree at most 6, which is a very simple numerical problem.

Remark. The conditions 55–56 and 58–59, are *sufficient*, but not necessary. Therefore, even if these conditions are not satisfied, there might still be a unique zero in the interval $(0, 1)$ that gives the minimum.

4.5 Justifying the Term "Reasonable Conditions"

The above conditions are easy to use as the quantities involved will be calculated anyhow. When valid, the only iterative part of the algorithm is the computation of λ^*, within the interval $(0, 1)$. The rest of the quantities are given by explicit formulas. However, in order to be helpful the conditions must be satisfied in situations that occur in practice. The following discussion is meant to show that this is the case. To do so we will illustrate three different situations graphically. After this we will show an example of a camera setting where the conditions are not fulfilled for some points.

In Figure 4, the normal situation is illustrated for the left camera. The epipole is outside the image and the distance between the epipolar planes corresponding to the two image points ($\lambda = 0$ and $\lambda = 1$) is small. Furthermore the covariance matrices for the measurements give rise to a distance measure where the ellipses representing points at equal distance from P_L are not too elongated. If the situation is the same in the right image, the solution is given by a unique λ in the interval $(0, 1)$. In Figure 5, a more unusual situation is illustrated. Due to the very elongated ellipses, corresponding to points of equal distance from the measured points, it is possible that the distance to the epipolar line corresponding to $\lambda = 0$ is longer than the distance to the line corresponding to $\lambda = \infty$. Therefore there is a possibility that the minimum is outside the interval $(0, 1)$. In the third example, Figure 6, it is not the shape of the ellipses that gives rise to a strange behavior, but the closeness of the epipolar point to P_L. This is due to the geometry of the cameras and is further illustrated in the following example.

4.5.1 An Example of a Camera Setting. We present here an example of a specific camera configuration where the given conditions fail for certain points. This neither implies that the solution is outside the interval $(0, 1)$, nor that there are more roots than one in this interval. As mentioned before the conditions are sufficient but not necessary. When two cameras are used for stereo it is common to arrange them in a symmetric way, either with their optical axes parallel or with a symmetric vergence. Such a setting will be used as an example, see Fig-

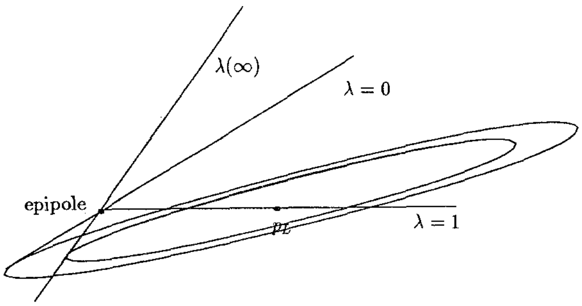

Fig. 5. A very unusual situation. Covariance matrices giving rise to elongated ellipses with special orientation can give a distance from p_L to the epipolar line for $\lambda = \infty$ smaller than the distance to the line for $\lambda = 0$.

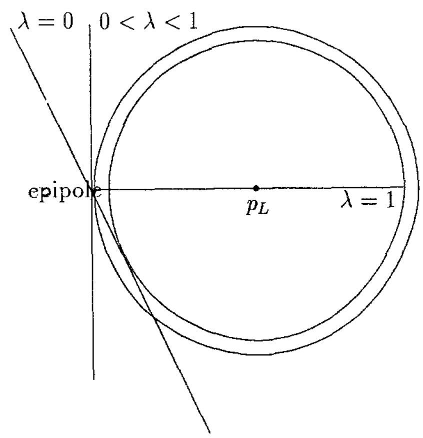

Fig. 6. Special camera configurations influences the distances and can make the maximum value of f_L occur inside the interval $(0, 1)$, despite that the covariance matrices gives points on equal distance from p_L as circles.

ure 7. In the example the vergence angle is 90°. This unusually large angle is chosen in order to make it possible to illustrate a situation where the "reasonable conditions" are not fulfilled for all points.

We will study a plane through the center of the images, with a rather wide opening angle of 45°. If small errors for the measured points are assumed[1] then the vectors p_L, p_R, and $T_R - T_L$ will be (almost) in the same plane. Referring to equation (47) and (48) this means that the quantities $\alpha \times \gamma$ and $p_L \times \gamma$ will be parallel vectors and can be written

$$\alpha \times \gamma \approx k_1 \widehat{n}, \quad p_L \times \gamma \approx k_2 \widehat{n} \qquad (60)$$

where $\widehat{n}$ is a unit vector. Nothing that α is P_R

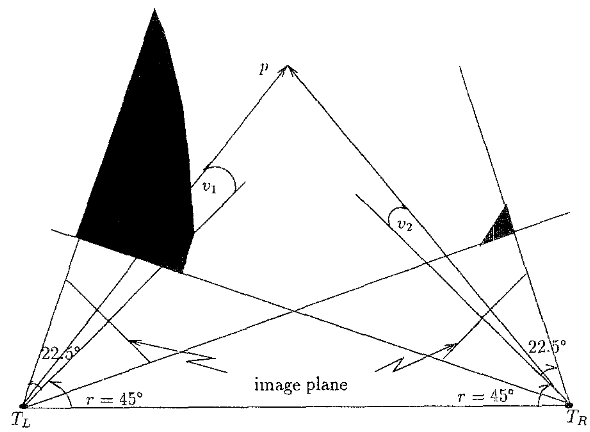

Fig. 7. A two camera setting with the cameras at right angles. The plane depicted, viewed from above, is a plane containing the optical centers and approximately passing through the centers of the images. The dark areas represent 3D points, the projections of which fail to satisfy the "reasonable conditions", concerning the left camera. The small black area to the right corresponds to points for which the first criterion, $2B - C > 0$ fails. The large black area to the left contains points for which the second derivative criterion is not fulfilled. See text for more details.

expressed in the left coordinate system and using the notation from Figure 7 we have

$$\begin{aligned}\alpha \times \gamma &= |p_R||T_R - T_L| \sin(r + v_2)\widehat{n} \\ &= \frac{1}{\cos(v_2)}|T_R - T_L| \sin(r + v_2)\widehat{n} \\ &= \frac{\sin(r + v_2)}{\cos(v_2)}|T_R - T_L|\widehat{n} \qquad (61)\end{aligned}$$

$$\begin{aligned}p_L \times \gamma &= |p_L||T_R - T_L| \sin(r + v_1)\widehat{n} \\ &= \frac{\sin(r + v_1)}{\cos(v_1)}|T_R - T_L|\widehat{n} \qquad (62)\end{aligned}$$

The constants A, B and C, see equation (49), can now be written as

$$\begin{aligned}A &= \frac{\sin(r + v_2)}{\cos(v_2)}|T_R - T_L|\widehat{n}^T G_L \\ &\quad \times \frac{\sin(r + v_2)}{\cos(v_2)}|T_R - T_L|\widehat{n} \\ &= \frac{\sin(r + v_2)^2}{\cos(v_2)^2}|T_R - T_L|^2 \widehat{n}^T G_L \widehat{n} \qquad (63)\end{aligned}$$

$$B = \frac{\sin(r + v_1)\sin(r + v_2)}{\cos(v_1)\cos(v_2)}$$

$$\times |T_R - T_L|^2 \hat{n}^T G_L \hat{n} \quad (64)$$

$$C = \frac{\sin(r + v_1)^2}{\cos(v_1)^2} |T_R - T_L|^2 \hat{n}^T G_L \hat{n} \quad (65)$$

This implies that

$$2B - C > 0$$
$$\Leftrightarrow 2\frac{\sin(r + v_2)}{\cos(v_2)} - \frac{\sin(r + v_1)}{\cos(v_1)} > 0 \quad (66)$$

In the same way

$$4B^2 - 2AB - AC > 0$$
$$\Leftrightarrow 3\frac{\sin(r + v_1)}{\cos(v_1)} - 2\frac{\sin(r + v_2)}{\cos(v_2)} > 0 \quad (67)$$

In Figure 7 the small area to the right corresponds to the part of 3D space where the first condition (55) is not fulfilled whereas the larger area to the left corresponds to the part of space (= combinations of points in the images) where the second condition (55) is not fulfilled. The same reasoning applied to the right image will give a similar result, but with left and right interchanged.

If one combines the result for the left and the right camera it is clear that the "reasonable conditions" are valid for most combination of points in the two images. In more common situations, with a vergence angle $\approx 50°$ or less (= the optical axes are more parallel) the reasonable conditions are valid for all image points.

Acknowledgments

The authors wish to thank the Swedish National Board for Industrial and Technical Development, NUTEK, for providing full financial support during the period this report was in progress. We wish also to thank Prof. Jan-Olof Eklundh for expert guidance of our work. Special thanks are due to Ph.D. Niklas Nodström for stimulating discussions.

Notes

1. An error of 10 pixels in an ordinary image corresponds to an angle error of about 1°, and will therefore effect the discussion only to a limited extent!

References

1. S.D. Blostein and T.S. Huang, "Error analysis in stereo determination of 3-D point positions," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 9, no. 6, pp. 752–765, 1987.
2. J.J. Rodrigues and J.K. Aggarwal, "Quantization error in stereo imaging," in *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, Ann Arbour, USA, June 1988, pp. 153–158.
3. R. Deriche, R. Vaillant, and O.D. Faugeras, "From noisy edge points to 3D reconstruction of a scene: A robust approach and its uncertainty analysis," in *Proceedings of the 7th Scandinavian Conference on Image Analysis*, Aalborg, Denmark, August 1991, pp. 225–232.
4. N. Ayache and O.D. Faugeras, "Maintaining representations of the environment of a mobile robot," *IEEE Trans. on Robotics and Automation*, vol. 5, no. 6, pp. 804–818, 1989.
5. A. Albert, *Regression and the Moore-Penrose Pseudoinverse*, Academic Press: New York, 1972.

Magnus Andersson received a MSc degree in Engineering Physics in 1983, and a PhD degree in Computer Science in 1994, both from the Royal Institute of Technology in Stockholm. He has been working at the Computational Vision and Active Perception Laboratory (CVAP) since 1985, mainly with tracking methods in computer vision.

Demetrios Betsis recieved a MSc in Mathematics and Statistics from Stockholm University in 1971, and a PhD in Numerical Analysis from the same university in 1985.