Towards robust metric reconstruction via a dynamic uncalibrated stereo head

M J Brooks [a] L de Agapito [b] D Q Huynh [c] and L Baumela [d]

[a] *Centre for Sensor Signal and Information Processing, Signal Processing Research Institute, Technology Park, Adelaide, SA 5095, Australia*

[b] *Robotics Research Group, Department of Engineering Science, University of Oxford, 19 Parks Road, Oxford, OX1 3PJ, United Kingdom*

[c] *Department of Computer Science, The University of Western Australia, Nedlands, Perth, WA 6907, Australia*

[d] *Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Spain*

Email correspondence: mjb@cs.adelaide.edu.au

Abstract

We consider the problem of metrically reconstructing a scene viewed by a moving stereo head. The head comprises two cameras with coplanar optical axes arranged on a lateral rig, each camera being free to vary its angle of vergence. Under various constraints, we derive novel explicit forms for the epipolar equation, and show that a static stereo head constitutes a degenerate camera configuration for carrying out self-calibration. The situation is retrieved by consideration of a stereo head undergoing ground plane motion, and new closed-form solutions for self-calibration are derived. An error analysis reveals that reconstruction is adversely affected by inward-facing camera vergence angles that are similar in value, and by a principal point location whose horizontal component is in error. It is also shown that the adoption of domain-specific robust techniques for computation of the fundamental matrix can significantly improve the quality of scene reconstruction. Experiments conducted with dynamic stereo head images confirm that avoidance of near-degenerate configurations and use of robustness techniques are essential if reliable reconstructions are in future to be attained.

Key words: Self-calibration, metric reconstruction, stereo head, degeneracy, epipolar equation, fundamental matrix, ego-motion.

1 Introduction

Attention has recently been devoted to the problem of how relative orientation and camera calibration parameters may automatically be determined from stereo and time-varying images (see, for example, [4,7,10,18,19,22,23,25,27]). Analysis of the epipolar equation and the fundamental matrix has revealed that at most 7 imaging parameters may be determined from corresponding points arising in a pair of images. Faugeras et al. [7] reported that 5 intrinsic parameters can be recovered if 3 stereo views of the scene are obtained (see also [18]). Using a simpler camera model, Hartley [10] showed that 2 independent focal lengths, in addition to 5 relative orientation parameters, could be estimated from a stereo pair by a method involving singular value decomposition of the fundamental matrix.

In this paper, we revisit the issue of self-calibration and metric reconstruction in the special context of a stereo head, perhaps the most commonly adopted binocular camera configuration in robotics. Here two cameras are arranged analogously to a pair of eyes, being placed on a lateral rig, each free to vary its angle of vergence. The optical axes of the cameras are therefore coplanar. In section 2, we derive explicit, analytical forms for the fundamental matrix, proving in section 3 that a static stereo head constitutes a degenerate camera configuration since some ability to perform self-calibration is inevitably lost. As such, this work presents one of the first studies of the feasibility of self-calibration.

It emerges that there are configurations of the stereo head for which self-calibration of this kind may not be carried out. Thus, an isosceles configuration, in which left and right cameras verge inwards an identical amount, is shown for the first time to be such a degenerate configuration. Cameras having parallel optical axes are a more familiar degenerate configuration.

These degeneracies might be thought of as curios of only theoretical interest, given that a stereo head is in practice unlikely to satisfy precisely any of these singularities. However, our implementations and error analyses suggest that near-singular configurations of the stereo head give rise to significant errors in self-calibration and thus reconstruction. A main goal of this paper is therefore to lay bare these sensitive configurations and to derive some necessary conditions for attainment of good self-calibration and reconstruction outcomes.

In section 4 we consider a special kind of motion of the stereo head in which the baseline and optical axes remain confined to a plane. New closed-form solutions for self-calibration are thereby obtained, inspired by an earlier technique of Zhang et al. [25]. Our direct approach may be seen as a contribution to active vision systems [20,16,2,5,6]. Key factors in the approach taken here are the

consideration of explicit, analytical forms of the fundamental matrix, and the use of the vergence angles in the parameterisation of the problem.

We next give some necessary conditions for attaining improved self-calibration and reconstruction results. It is shown that better estimates of the fundamental matrix may be obtained when a domain-specific parameterisation and associated robust computational techniques are adopted. Improved self-calibration and 3D reconstructions may be obtained by avoidance of near-degenerate configurations (see section 6). A key factor here is an error analysis of principal point location error versus self-calibration accuracy. It is demonstrated that errors can be mitigated by judicious avoidance of near-singular head configurations.

Finally, in sections 7 and 8, we show via both synthetic and real-world experiments that techniques arising out of these analyses permit relatively robust reconstructions to be computed either from an image pair obtained from static stereo head, or from two pairs of images obtained from a stereo head undergoing ground-plane motion. Our work extends considerably the explicit analysis of the epipolar geometry presented in Brooks et al. [3].

Research most closely related to that presented here is the self-calibration dynamic stereo work of Zhang et al. [25] and the special techniques for computation of the stereo-head fundamental matrix due to Li et al. [15]. However, the former does not seek closed-form solutions amenable to error analysis, while the latter is not concerned with self-calibration or reconstruction. The metric calibration work of Zisserman et al. [27] is also related. Here, they show that metric structure may be computed to a two fold ambiguity from a single general motion of a fixed stereo rig. This approach involves the identification of certain geometrical objects which are invariant to Euclidean transformations in 3-D space, namely the plane at infinity and the absolute conic. However, none of these approaches investigate degenerate configurations or conduct error analyses of the self-calibration process.

2 Stereo head assembly

2.1 The epipolar equation

First we recall the epipolar geometry which underpins our analysis. We shall adopt a notation similar, but not identical, to that of Faugeras et al. [7]; see the Appendix for a summary of the differences. Let $\mathbf{m}$ and $\mathbf{m}'$ denote corresponding points, in homogeneous coordinates, in the left and right images,

respectively. We may express the epipolar equation as

$$\mathbf{m}^T \mathbf{F}\, \mathbf{m}' = 0, \tag{1}$$

where $\mathbf{F}$ is the fundamental matrix [7,17], given by

$$\mathbf{F} = \mathbf{A}^T\, \mathbf{T}\, \mathbf{R} \mathbf{A}'. \tag{2}$$

Here, $\mathbf{R}$ embodies the pure rotation that renders the left image parallel with the right image, $\mathbf{T}$ is a skew-symmetric matrix formed from the baseline vector connecting the left and right optical centres, and $\mathbf{A}$ and $\mathbf{A}'$ are the intrinsic parameter matrices of the left and right cameras.

Consider now the special case of a stereo head in which a pair of cameras is mounted on a lateral rig. The cameras are free to vary their angles of vergence. The y-axes of the two images are parallel, and are orthogonal to the baseline vector, as depicted in Figure 1. The optical axes, (Cc and $C'c'$), and the baseline, $\mathbf{t}$ are therefore coplanar. The matrices $\mathbf{R}$, $\mathbf{T}$ and $\mathbf{A}$ now take the

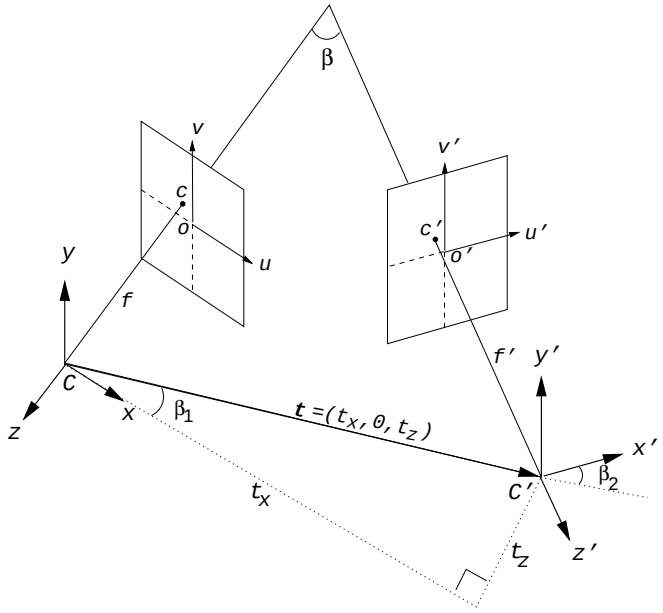


Fig. 1. Stereo head configuration.

forms

$$\mathbf{R} = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}, \qquad \mathbf{T} = \begin{pmatrix} 0 & -t_z & 0 \\ t_z & 0 & -t_x \\ 0 & t_x & 0 \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ 0 & 0 & -f \end{pmatrix}.$$

Here, β is the angular rotation about the y-axis that renders the left image parallel with the right image; $\mathbf{T}$ is formed out of the baseline vector $\mathbf{t} =$

$(t_x, 0, t_z)^T$; and, for each camera, the focal length and the principal point are the only unknown intrinsic parameters, denoted by f and (u_0, v_0), with the image coordinate system axes assumed to be orthogonal and similarly scaled.

In view of equation (2), the fundamental matrix is now given by

$$\mathbf{F} = \begin{pmatrix} 0 & -t_z & t_z v_0' \\ \phi & 0 & -u_0'\phi - f'\sigma \\ -v_0\phi & u_0 t_z - f t_x & u_0' v_0 \phi + v_0'(f t_x - u_0 t_z) + f' v_0 \sigma \end{pmatrix}, \qquad (3)$$

where $\phi = t_z \cos\beta + t_x \sin\beta$, and $\sigma = -t_x \cos\beta + t_z \sin\beta$.

As is well known, absolute dimensions of depth cannot be determined solely from knowledge of corresponding points and the associated fundamental matrix. Accordingly, without loss of generality, we set the baseline length to unity, and note that the direction of the baseline vector is now effectively described by 1 parameter. There are therefore 8 unknowns encoded within $\mathbf{F}$, these being $\beta, u_0, v_0, f, u_0', v_0', f'$ and either t_x or t_z.

2.2 Vergence-angle parameterisation

The form of $\mathbf{F}$ is simplified if an adjustment is made to the parameterisation by incorporating the left and right vergence angles β_1 and β_2, where

$$t_x = \cos\beta_1, \qquad t_z = -\sin\beta_1, \qquad \beta = \beta_1 + \beta_2. \qquad (4)$$

Here, β_1 and β_2 specify the extent to which the left and right optical axes

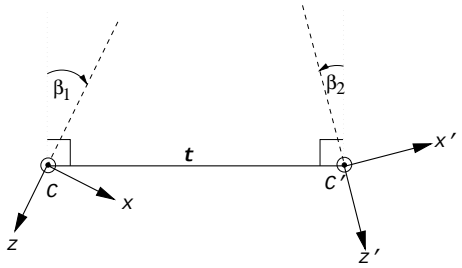


Fig. 2. Plan view of stereo head showing vergence angles.

point inwards from the direction 'straight-ahead'. (Note, therefore, that the left and right vergence angles are measured in an opposite sense. See Figure 2.) Relative orientation in this situation is now determined by the pair β_1, β_2, instead of β, t_x.

Equation (3) is then expressed in more economical form as

$$\mathbf{F} = \begin{pmatrix} 0 & \sin\beta_1 & -v_0' \sin\beta_1 \\ \sin\beta_2 & 0 & (-u_0' \sin\beta_2 + f' \cos\beta_2) \\ -v_0 \sin\beta_2 & (-u_0 \sin\beta_1 - f \cos\beta_1) & \begin{array}{l} v_0(u_0' \sin\beta_2 - f' \cos\beta_2) \\ +v_0'(u_0 \sin\beta_1 + f \cos\beta_1) \end{array} \end{pmatrix} \quad (5)$$

in which the 8 unknowns are $\beta_1, \beta_2, u_0, v_0, f, u_0', v_0', f'$.

In the event that sufficiently-many corresponding points can be located in the two images, it may be possible to obtain a numerical estimate, $\mathbf{F}_{est}$, of the matrix $\mathbf{F}$. Let

$$\mathbf{F}_{est} = \begin{pmatrix} \delta_1 & \delta_2 & \delta_3 \\ \delta_4 & \delta_5 & \delta_6 \\ \delta_7 & \delta_8 & \delta_9 \end{pmatrix}. \quad (6)$$

Noting that $\mathbf{F}_{est}$ may only be determined up to a scale factor, we may form the equation

$$\mathbf{F} = \lambda \mathbf{F}_{est}. \quad (7)$$

Here, the unknown λ aligns the scales of the two matrices. We can now obtain 7 equations by linking respective elements of the matrices. However, these equations are not all independent, since $F_{33} = F_{23}F_{31}/F_{21} + F_{32}F_{13}/F_{12}$. Thus we may obtain up to 6 independent equations, one of which will be utilised in eliminating λ. We therefore observe that, of the 8 unknowns encoded within $\mathbf{F}$, at most 5 may be determined provided the remaining 3 are known. Further constraints are therefore needed if we are to solve for the various parameters.

3 Self-calibration of a static stereo head

We now consider how equation (7) may be solved under additional assumptions. In doing so, we shall gain insight into the feasibility of self-calibration of a static rig. Our aim is to develop closed-form expressions for the various parameters.

Case 1: u_0, v_0, u'_0, v'_0 known

If the locations of the principal points are known, we may without loss of generality offset the coordinates of image points. This simple form of image rectification then permits the setting of $u_0 = v_0 = u'_0 = v'_0 = 0$ in (7), giving

$$\mathbf{F} = \begin{pmatrix} 0 & \sin\beta_1 & 0 \\ \sin\beta_2 & 0 & f'\cos\beta_2 \\ 0 & -f\cos\beta_1 & 0 \end{pmatrix} = \lambda \begin{pmatrix} 0 & \delta_2 & 0 \\ \delta_4 & 0 & \delta_6 \\ 0 & \delta_8 & 0 \end{pmatrix}. \tag{8}$$

This case is relevant either to a pair of static stereo cameras having independent, unknown focal lengths, or to a single mobile camera in which the focal length may be varied. We now have 5 unknown parameters, including λ, but are able to generate only 4 independent equations. Thus, whereas in general we may obtain via self-calibration 2 focal lengths and 5 relative orientation parameters, we are unable to fix any of the unknown parameters in this special situation. The camera configuration is therefore degenerate for self-calibration.

Case 2: u_0, v_0, u'_0, v'_0 known; $f = f'$

Here we assume that the left and right focal lengths are equal, and seek only the 3 unknown parameters f, β_1, β_2. This situation corresponds to a single camera with fixed intrinsic parameters moving in such a way that its optical axis and x-axis remain always parallel to the ground plane. We may equivalently regard this as a study of stereo head images due to a pair of identical cameras. Equation (7) now reduces to

$$\mathbf{F} = \begin{pmatrix} 0 & \sin\beta_1 & 0 \\ \sin\beta_2 & 0 & f\cos\beta_2 \\ 0 & -f\cos\beta_1 & 0 \end{pmatrix} = \lambda \begin{pmatrix} 0 & \delta_2 & 0 \\ \delta_4 & 0 & \delta_6 \\ 0 & \delta_8 & 0 \end{pmatrix}, \tag{9}$$

yielding 4 independent equations. All 3 imaging parameters can now be determined, viz:

$$f = \sqrt{(\delta_8^2 - \delta_6^2)/(\delta_4^2 - \delta_2^2)}, \quad \tan\beta_1 = -f\,\delta_2/\delta_8, \quad \tan\beta_2 = f\,\delta_4/\delta_6. \tag{10}$$

Note here that f is computed in the same units as the coordinates of the corresponding points used to estimate the fundamental matrix.

Case 3: u_0, v_0, u'_0, v'_0 known; $\beta_1 = \beta_2 = \beta/2$

We now seek to determine β, f and f', given the very special situation in

which the vergence angles are equal, with the principal axes of the cameras and the baseline forming an isosceles triangle. Our equation is now

$$\mathbf{F} = \begin{pmatrix} 0 & \sin\beta/2 & 0 \\ \sin\beta/2 & 0 & f'\cos\beta/2 \\ 0 & -f\cos\beta/2 & 0 \end{pmatrix} = \lambda \begin{pmatrix} 0 & \delta_2 & 0 \\ \delta_4 & 0 & \delta_6 \\ 0 & \delta_8 & 0 \end{pmatrix}. \tag{11}$$

Noting that $F_{12} = F_{21}$, we see that none of the unknown parameters may be determined without more information being provided. Remarkably, if the focal lengths are known to be equal, it remains impossible to recover any of the parameters. Note, however, that the ratio of the focal lengths may be determined.

Case 4: $(u_0, v_0, f) = (u_0', v_0', f')$

Here we assume left and right cameras have identical focal length and principal point location. This also corresponds to a mobile camera moving horizontally. The 5 parameters $\beta_1, \beta_2, u_0, v_0, f$ are now free, equation (7) reducing to:

$$\begin{pmatrix} 0 & \sin\beta_1 & -v_0\sin\beta_1 \\ \sin\beta_2 & 0 & -u_0\sin\beta_2 + f\cos\beta_2 \\ -v_0\sin\beta_2 & \begin{matrix} -u_0\sin\beta_1 \\ -f\cos\beta_1 \end{matrix} & \begin{matrix} u_0v_0(\sin\beta_1+\sin\beta_2) \\ +v_0f(\cos\beta_1-\cos\beta_2) \end{matrix} \end{pmatrix} = \lambda \begin{pmatrix} 0 & \delta_2 & \delta_3 \\ \delta_4 & 0 & \delta_6 \\ \delta_7 & \delta_8 & \delta_9 \end{pmatrix} \tag{12}$$

Here we note the following constraints:

(1) $F_{11} = F_{22} = 0$
(2) $F_{31} = F_{21}F_{13}/F_{12}$
(3) $F_{33} = (F_{23} + F_{32})F_{13}/F_{12}$

The analytic form of $\mathbf{F}$ has zero determinant. Given that the self-calibration process operates on an unscaled estimate of the fundamental matrix, we may infer that equation (12) offers the potential for at most 4 of the 5 imaging parameters to be obtained, assuming that the other is known. Interestingly, we observe that:

(1) If at least one vergence angle is non-zero, v_0 may immediately be obtained from $\mathbf{F}$.

(2) In general, at most an additional 3 parameters may then be obtained providing the remaining one is known.
(3) Analysis of a slightly more general fundamental matrix reveals that the aspect ratio can never be determined since the relevant parameter fails to appear in the matrix.
(4) If δ_2 and δ_4 are observed to be equal, this corresponding to a symmetric vergence angle configuration, then the principal point location may be determined without knowledge of any other parameter.

Note that these results are fundamental and are not specific to our particular approach. These observations may be confirmed by inspection of the well known Kruppa equations.

4 Self-calibration of a stereo head undergoing ground-plane motion

Having seen that a static stereo head, with coplanar optical axes, is a degenerate configuration for self-calibration, we now assess the consequences of moving the head. Specifically, we permit:

(1) Motion of the head such that the optical axes and x-axes of the cameras are confined to the ground plane. This therefore captures the situation in which an upright robot head undergoes ground plane motion.
(2) Independent vergence angles of the head that may vary with the motion.
(3) Each camera to have an unknown but fixed focal length.

The following analysis adopts a technique of Zhang et al. [25] in which various fundamental matrices are utilised.

4.1 Formulating the fundamental matrices

Let the rig move from an initial position to a final position. Let the left-right pair of images in the initial position be termed I_1 and I_2, and let the left-right images in the final position be termed I_3 and I_4 (see Figure 3). The left camera is thus responsible for the successive images I_1 and I_3. Assume that the determining of corresponding points has led to estimates for the fundamental matrices linking the following image pairs: (I_1, I_2), (I_3, I_4), (I_1, I_3), (I_2, I_4). Let the associated analytical fundamental matrices be termed $\mathbf{F}^{12}$, $\mathbf{F}^{34}$, $\mathbf{F}^{13}$, $\mathbf{F}^{24}$. As before, we aim to solve for the parameters embedded within these matrices by exploiting the fact that the analytical and the estimated forms of the fundamental matrix are directly proportional. Note that, in this regard, the approach pursued in Zhang et al. [25] is quite different in that a least-

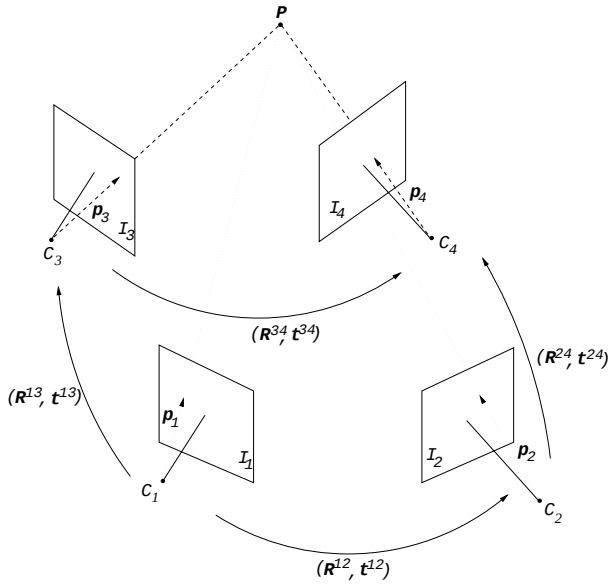


Fig. 3. Motion of the stereo head.

squares approach is used to solve a more general problem in which motion is not confined to the plane (although, unlike here, the relative orientation of the head is assumed fixed).

Recalling equation (8), the initial position of the rig gives rise to the fundamental matrix, $\mathbf{F}^{12}$, given by

$$\mathbf{F}^{12} = \begin{pmatrix} 0 & \sin\beta_1^{12} & 0 \\ \sin\beta_2^{12} & 0 & f'\cos\beta_2^{12} \\ 0 & -f\cos\beta_1^{12} & 0 \end{pmatrix}. \tag{13}$$

Assuming that the focal lengths of the respective cameras remain fixed, and that the vergence angles are free to shift, we obtain the following fundamental matrix, $\mathbf{F}^{34}$, capturing the epipolar relationship between the left and right images of the rig in its final position:

$$\mathbf{F}^{34} = \begin{pmatrix} 0 & \sin\beta_1^{34} & 0 \\ \sin\beta_2^{34} & 0 & f'\cos\beta_2^{34} \\ 0 & -f\cos\beta_1^{34} & 0 \end{pmatrix}. \tag{14}$$

Similarly, the fundamental matrix relating the image pair (I_1, I_3) is given by

$$\mathbf{F}^{13} = \begin{pmatrix} 0 & \sin\beta_1^{13} & 0 \\ \sin\beta_2^{13} & 0 & f\cos\beta_2^{13} \\ 0 & -f\cos\beta_1^{13} & 0 \end{pmatrix}, \tag{15}$$

with the fundamental matrix for the image pair (I_2, I_4) being

$$\mathbf{F}^{24} = \begin{pmatrix} 0 & \sin\beta_1^{24} & 0 \\ \sin\beta_2^{24} & 0 & f'\cos\beta_2^{24} \\ 0 & -f'\cos\beta_1^{24} & 0 \end{pmatrix}. \tag{16}$$

Here we note that the focal lengths of the respective cameras remain unchanged in the movement of the rig from its initial to final position. We observe that, under the above parameterisation, image I_1 undergoes a rotation of $(\beta_1^{13}+\beta_2^{13})$, relative to its own local coordinate system, in becoming oriented in parallel with image I_3.

Note that, in common with Zhang et al. [25], we make no use of the cross matrices $\mathbf{F}^{14}$ and $\mathbf{F}^{23}$ in our work. In our special situation, these matrices are theoretically redundant as they may be derived from the other fundamental matrices (via self-calibration results). However, their incorporation into our method might be expected to further improve robustness.

4.2 Solving the fundamental matrix equations

It is now necessary to further enhance our notation so as to be able to deal simultaneously with various fundamental matrices. Let the numerical estimate, $\mathbf{F}_{est}^{ij}$, of fundamental matrix $\mathbf{F}^{ij}$ be represented as

$$\mathbf{F}_{est}^{ij} = \begin{pmatrix} \delta_1^{ij} & \delta_2^{ij} & \delta_3^{ij} \\ \delta_4^{ij} & \delta_5^{ij} & \delta_6^{ij} \\ \delta_7^{ij} & \delta_8^{ij} & \delta_9^{ij} \end{pmatrix}, \tag{17}$$

and let $\mu_k^{ij} = (\delta_k^{ij})^2$. Let a right bar and superscript indicate the fundamental matrix from which the elements derive. Thus, for example, note that

$$\mu_k + \mu_l|^{ij} = (\delta_k^{ij})^2 + (\delta_l^{ij})^2. \tag{18}$$

In view of the earlier analysis, resulting in (10), we may immediately infer that

$$f = \sqrt{\frac{\mu_8 - \mu_6}{\mu_4 - \mu_2}}\Bigg|^{13}, \qquad f' = \sqrt{\frac{\mu_8 - \mu_6}{\mu_4 - \mu_2}}\Bigg|^{24}$$

$$\tan\beta_1^{13} = -f\frac{\delta_2}{\delta_8}\Bigg|^{13}, \ \tan\beta_2^{13} = f\frac{\delta_4}{\delta_6}\Bigg|^{13}, \ \tan\beta_1^{24} = -f'\frac{\delta_2}{\delta_8}\Bigg|^{24}, \ \tan\beta_2^{24} = f'\frac{\delta_4}{\delta_6}\Bigg|^{24}$$

It then follows that

$$\tan\beta_1^{12} = -f\frac{\delta_2}{\delta_8}\Big|^{12}, \ \tan\beta_2^{12} = f'\frac{\delta_4}{\delta_6}\Big|^{12}, \ \tan\beta_1^{34} = -f\frac{\delta_2}{\delta_8}\Big|^{34}, \ \tan\beta_2^{34} = f'\frac{\delta_4}{\delta_6}\Big|^{34}$$

We therefore have closed-form solutions for the 2 focal lengths and the rotations between images. Implicit in the above are the directions of the various translations between perspective centres. Note that we have not made explicit use of the rotational constraint $\beta^{12} + \beta^{24} - \beta^{34} - \beta^{13} = 0$. The incorporation of this constraint might be expected to improve robustness of the estimates. However, this constraint is implicit in the baseline constraint equation which takes into account the rigidity of the rig and which we now solve to compute the relative magnitudes of the baseline vectors.

4.3 Solving the baseline constraint equation

We have yet to completely determine the relative orientation of all image pairs as we have still to compute the relative magnitudes of the baselines. (As noted earlier, it is not possible to compute absolute scale of the baselines only from corresponding points.) These relative magnitudes will complete the description of the motion of the head.

Let the magnitude of the head's baseline vector in the initial position be unity. The baseline vector, $\mathbf{t}^{12}$, may therefore be written as

$$\mathbf{t}^{12} = (\cos\beta_1^{12}, 0, -\sin\beta_1^{12})^T. \tag{19}$$

Letting L^{ij} denote the length of the baseline vector $\mathbf{t}^{ij}$, we may immediately write down the remaining baseline vectors as

$$\mathbf{t}^{13} = L^{13}(\cos\beta_1^{13}, 0, -\sin\beta_1^{13})^T \tag{20}$$
$$\mathbf{t}^{24} = L^{24}(\cos\beta_1^{24}, 0, -\sin\beta_1^{24})^T \tag{21}$$
$$\mathbf{t}^{34} = L^{34}(\cos\beta_1^{34}, 0, -\sin\beta_1^{34})^T. \tag{22}$$

Our task is now to determine the lengths L^{13}, L^{24}, L^{34}. Returning to Figure 3, we observe (after [25]) the baseline constraint equation

$$\mathbf{R}^{12}\,\mathbf{t}^{24} = \mathbf{t}^{13} - \mathbf{t}^{12} + \mathbf{R}^{13}\mathbf{t}^{34}. \tag{23}$$

Expanding this, we have

$$L^{24}\begin{pmatrix}\cos(\beta^{12}+\beta_1^{24})\\ 0\\ -\sin(\beta^{12}+\beta_1^{24})\end{pmatrix}=\begin{pmatrix}L^{13}\cos\beta_1^{13}-\cos\beta_1^{12}+L^{34}\cos(\beta_1^{34}+\beta^{13})\\ 0\\ -L^{13}\sin\beta_1^{13}+\sin\beta_1^{12}-L^{34}\cos(\beta_1^{34}+\beta^{13})\end{pmatrix}.$$

Here β^{13} rotates I_1 parallel to I_3 and is such that $\beta^{13}=\beta_1^{13}+\beta_2^{13}$. Recall that $\beta^{12}=\beta_1^{12}+\beta_2^{12}$. Clearly, the 3 unknown lengths may not be determined from the above equation. But, on the assumption that the baseline length of the rig remains constant, so that L^{34} also has unit length, we may readily infer that

$$L^{13}=\left(\sin(\omega-\beta^{13})-\sin\phi\right)/\sin\tau, \tag{24}$$

where $\omega=\beta^{12}+\beta_1^{24}-\beta_1^{34}$, $\phi=\beta_2^{12}+\beta_1^{24}$ and $\tau=\beta_1^{13}-\beta^{12}-\beta_1^{24}$. The formula for L^{24} then follows directly from the baseline constraint equation. We have therefore described the motion of the rig.

4.4 Incorporating camera tilt

Our analysis here is a generalisation of that considered in the previous section in that the head may now tilt up or down, by a rotation about the baseline. We note that the baseline remains confined to a plane, and that the optical axes of the two cameras are at all times coplanar, but are not confined to the same plane in consecutive head positions. Critically, in the analysis presented here, either the initial or final position of the head should have zero tilt.

We now consider how a rotation of θ about the baseline maps the left image to a new position. A rotation of θ about the baseline is equivalent, in the left image's coordinate system, to three composite rotations: a rotation of β_1^{34} about the y-axis, followed by a rotation of θ about the x-axis, and then a rotation of $-\beta_1^{34}$ about the y-axis. In addition to this tilting, the previous rotation in the plane may still take place. The fundamental matrix may therefore be expressed as $\mathbf{F}^{13}=\mathbf{A}^T\,\mathbf{T}\,\mathbf{R}\mathbf{A}$, where

$$\mathbf{T}=\begin{pmatrix}0 & \sin\beta_1^{13} & 0\\ -\sin\beta_1^{13} & 0 & -\cos\beta_1^{13}\\ 0 & \cos\beta_1^{13} & 0\end{pmatrix},\qquad \mathbf{A}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&-f\end{pmatrix}, \tag{25}$$

and $\mathbf{R}=R_y(-(\beta_1^{13}+\beta_2^{13}))R_y(-\beta_1^{34})R_x(\theta)R_y(\beta_1^{34})$. Here, we adopt the convention that $R_m(\psi)$ signifies a rotation of ψ about the m-axis.

A rather complex fundamental matrix results which, after considerable manipulation, may be simplified to give $\mathbf{F}^{13}$ as

$$\begin{pmatrix} -\sin\beta_1^{13}\sin\beta_1^{34}\sin\theta & \sin\beta_1^{13}\cos\theta & f\sin\beta_1^{13}\cos\beta_1^{34}\sin\theta \\ \cos\beta_1^{34}\sin(\beta_1^{34}+\beta_2^{13}) & -\sin\theta\cos(\beta_1^{34}+\beta_2^{13}) & f\sin\beta_1^{34}\sin(\beta_1^{34}+\beta_2^{13}) \\ -\sin\beta_1^{34}\cos\theta\cos(\beta_1^{34}+\beta_2^{13}) & & +f\cos\beta_1^{34}\cos\theta\cos(\beta_1^{34}+\beta_2^{13}) \\ f\cos\beta_1^{13}\sin\beta_1^{34}\sin\theta & -f\cos\beta_1^{13}\cos\theta & -f^2\cos\beta_1^{13}\cos\beta_1^{34}\sin\theta \end{pmatrix}.$$

The following equations may now be derived:

$$\tan\beta_1^{13} = -f\frac{\delta_2}{\delta_8}\Bigg|^{13}, \qquad \tan\beta_1^{34} = -f\frac{\delta_1}{\delta_3}\Bigg|^{13}, \qquad \tan\theta = -\frac{\delta_7}{\delta_8\sin\beta_1^{34}}\Bigg|^{13}$$

$$f^2 = \frac{-\delta_3\delta_6}{\delta_2\delta_5+\delta_1\delta_4}\Bigg|^{13}, \quad \sec^2(\beta_1^{34}+\beta_2^{13}) = \sin^2\theta\left(1+\frac{\mu_4+\mu_6/f^2}{\mu_5}\right)\Bigg|^{13}$$

We therefore have closed-form solutions for the 5 unknowns f, θ, β_1^{34}, β_1^{13}, β_2^{13}.

Consideration of the fundamental matrix $\mathbf{F}^{24}$ yields symmetric formulae for the right camera vergence angles. The analysis is completed when we note that the previous formulae for baseline lengths are precisely applicable here, since the moving baseline has remained confined to a plane.

5 Robust computation of the fundamental matrix

Accurate self-calibration is highly dependent on a good numerical estimate of the fundamental matrix. Accordingly, in this section we show how the fundamental matrix may best be computed in this particular domain.

5.1 The 6-point algorithm

Given that the fundamental matrix $\mathbf{F}$ and corresponding points $\mathbf{m} = (x, y, 1)$ and $\mathbf{m}' = (x', y', 1)$ in the initial and final images are related via the equation $\mathbf{m}^T\mathbf{F}\,\mathbf{m}' = 0$, a fundamental matrix $\mathbf{F}$ of the form given in (12) can be computed using $n \geq 6$ correspondences to solve the following linear system

$$\mathbf{A}\mathbf{x} = \mathbf{0}, \tag{26}$$

where $\mathbf{A}$ is an $n \times 7$ coefficient matrix given by

$$\mathbf{A} = \begin{pmatrix} x_1 y'_1 & x_1 & y_1 x'_1 & y_1 & x'_1 & y'_1 & 1 \\ & & \vdots & & & & \\ x_n y'_n & x_n & y_n x'_n & y_n & x'_n & y'_n & 1 \end{pmatrix} \tag{27}$$

and $\mathbf{x}$ is a 7-vector which contains the coefficients of the fundamental matrix given by

$$\mathbf{x} = (\delta_2, \delta_3, \delta_4, \delta_6, \delta_7, \delta_8, \delta_9)^T. \tag{28}$$

Note that under this parameterisation $\mathbf{F}$ is constrained to have its first two diagonal entries equal to zero.

Equation (26) can be solved using the linear algorithm, with the aid of singular value decomposition (SVD). However, the linear algorithm is known not to be very stable (see [11]) because it fails to impose the essential constraint that $\text{rank}(\mathbf{F}) = 2$, and because the measure to be minimised is not geometrically meaningful.

Hartley [11] showed that results can be improved by normalizing the coordinates of image points, instead of using pixel units directly, so that their new values are on average equal to one. This is achieved by translating each image origin to the centroid of image points and by performing isotropic scaling on the image coordinates so that the average distance of corresponding points to the new image origin is $\sqrt{2}$.

In our case, the image origin was translated to the estimated location of the principal point prior to the isotropic scaling of image coordinates. This we found was the best approach for the particular data we encountered. In addition, *row scaling* of $\mathbf{A}$ was carried out. This is based on the observation that each row of $\mathbf{A}$ in (26) is, ideally, orthogonal to $\mathbf{x}$. Since the rows are essentially independent, each row of $\mathbf{A}$ may be scaled to a unit vector. We are unaware of previous reporting of this technique.

Zhang [26] compared the performance of the linear method with the performances of the M-estimator, LMedS and RANSAC techniques ([24], [8], [21]) and confirmed that, when applied in conjunction with the aforementioned operations, the linear method generates a fundamental matrix of comparable quality.

Given that the location of the principal point is known, the question arises whether it is better to use the parameterisation of $\mathbf{F}$ given in (9) imposing 5 zeros, or the parameterisation of $\mathbf{F}$ given in (12) imposing 2 zeros. Our implementations suggest that it is far better to impose only 2 zeros, using a

6-point algorithm. The reason for this would seem to be that, if the principal point is not accurately known, the 5-zeros assumption is invalid, and the 3-point algorithm causes the non-zero entries to incur compensatory distortion. Such distortions will obviously have a deleterious effect upon subsequent self-calibration.

5.2 Enforcing the rank-2 constraint

The main drawback of the linear method is that the rank-2 constraint on the fundamental matrix is not enforced. Hartley [11] imposed this constraint a posteriori by replacing the matrix $\mathbf{F}$ obtained via SVD by the matrix $\mathbf{F}'$ which minimizes the Frobenius norm $\|\mathbf{F}-\mathbf{F}'\|$ subject to the constraint $\det(\mathbf{F}')=0$.

However, this parameterisation is not suitable in our case as it fails to preserve the property that there should be two zero entries in the fundamental matrix (as in (12)), a property that is essential to our direct approach to self-calibration. Instead, we impose the rank-2 constraint explicitly within a non-linear optimisation framework. Specifically we minimize the *epipolar distance*; that is, the distance of image points to their corresponding epipolar lines, a geometrically meaningful measure.

To be able to deal with the outliers caused by false matches or by poor localisation of image features, we employ a robust M-estimator method. Our procedure for estimating $\mathbf{F}$ may now be summarised as follows:

(1) Obtain an initial estimate of $\mathbf{F}$ using the 6-point linear algorithm, image-origin translation, isotropic scaling and row scaling.
(2) Refine $\mathbf{F}$ by using an M-estimator technique to solve the following non-linear minimisation

$$\min_{\mathbf{F}} \sum_i w_i r_i^2$$

subject to the constraints that $\|\mathbf{x}\|^2 = 1$ and $\det(\mathbf{F}) = 0$. Here, the residual

$$r_i = \sqrt{\frac{1}{l_{i,1}^2+l_{i,2}^2}+\frac{1}{l_{i,1}'^2+l_{i,2}'^2}}\left(\mathbf{m}_i^T\mathbf{F}\mathbf{m}_i'\right)$$

is referred to as the *epipolar distance* [17] and represents the geometric distance of the points $\mathbf{m}_i$ and $\mathbf{m}_i'$ to their associated epipolar lines defined

by:

$$(l_{i,1}, l_{i,2}, l_{i,3}) = \mathbf{F}\mathbf{m}_i,$$
$$(l'_{i,1}, l'_{i,2}, l'_{i,3}) = \mathbf{F}^T\mathbf{m}'_i.$$

The chosen weighting factor w_i is that proposed by Huber [12] and later used by Luong et al. [17], and is given by

$$w_i = \begin{cases} 1 & \text{if } |r_i| \leq \sigma \\ \sigma/|r_i| & \text{if } \sigma < |r_i| \leq 3\sigma \\ 0 & \text{otherwise} \end{cases}$$

where σ is the standard deviation of the residuals r_i.

We have thus described various techniques used to obtain an improved estimate of the fundamental matrix. One possibility not investigated is that of using constraint (2) from Case 4 in section 3 which can be expected to further improve the estimation of the fundamental matrix.

We next analyse how self-calibration is adversely affected by error in the location of the principal point.

6 Error analysis of self-calibration

In order to apply the self-calibration methods previously described, it is necessary to obtain an estimate of the fundamental matrix from corresponding image points that are expressed with respect to the correct principal point location. However, it is well known that in practice the coordinates of the principal point are very difficult to estimate with precision, even if classical camera calibration methods are used [25]. We now assess the consequences of such imprecision.

Assume that we deal with a stereo pair due to a single camera in motion, as considered in Case 4 of section 3. Accepting the inevitability of error in the location of the principal point, let this error be given by $(\Delta u_0, \Delta v_0)$. Incorporating this discrepancy into the analytical formulation of the fundamental

matrix (with recourse to equation 12) yields

$$\begin{pmatrix} 0 & \sin\beta_1 & -\Delta v_0 \sin\beta_1 \\ \sin\beta_2 & 0 & -\Delta u_0 \sin\beta_2 + f\cos\beta_2 \\ -\Delta v_0 \sin\beta_2 & \begin{matrix} -\Delta u_0 \sin\beta_1 \\ -f\cos\beta_1 \end{matrix} & \begin{matrix} \Delta u_0 \Delta v_0 (\sin\beta_1 + \sin\beta_2) \\ +\Delta v_0 f(\cos\beta_1 - \cos\beta_2) \end{matrix} \end{pmatrix} = \lambda \begin{pmatrix} 0 & \delta_2 & \delta_3 \\ \delta_4 & 0 & \delta_6 \\ \delta_7 & \delta_8 & \delta_9 \end{pmatrix} \quad (29)$$

If we continue to utilize the direct formulae given in (10) to solve for f, β_1 and β_2, the estimated values for these parameters are likely to be in error.

Accordingly, let f, β_1 and β_2 denote the true focal length and virtual vergence angles of the moving camera, and let $\hat{f}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ denote the corresponding estimates.

6.1 Focal length estimation errors

Assume that the fundamental matrix has the form given in (29). Application to this matrix of the direct formula for the focal length given in (10) yields

$$\hat{f}^2 = f^2 - \Delta u_0^2 + \frac{2f\Delta u_0(\sin\beta_1\cos\beta_1 + \sin\beta_2\cos\beta_2)}{\cos^2\beta_1 - \cos^2\beta_2}. \quad (30)$$

Let $\nu = \beta_1 - \beta_2$, then substituting ν into (30) gives

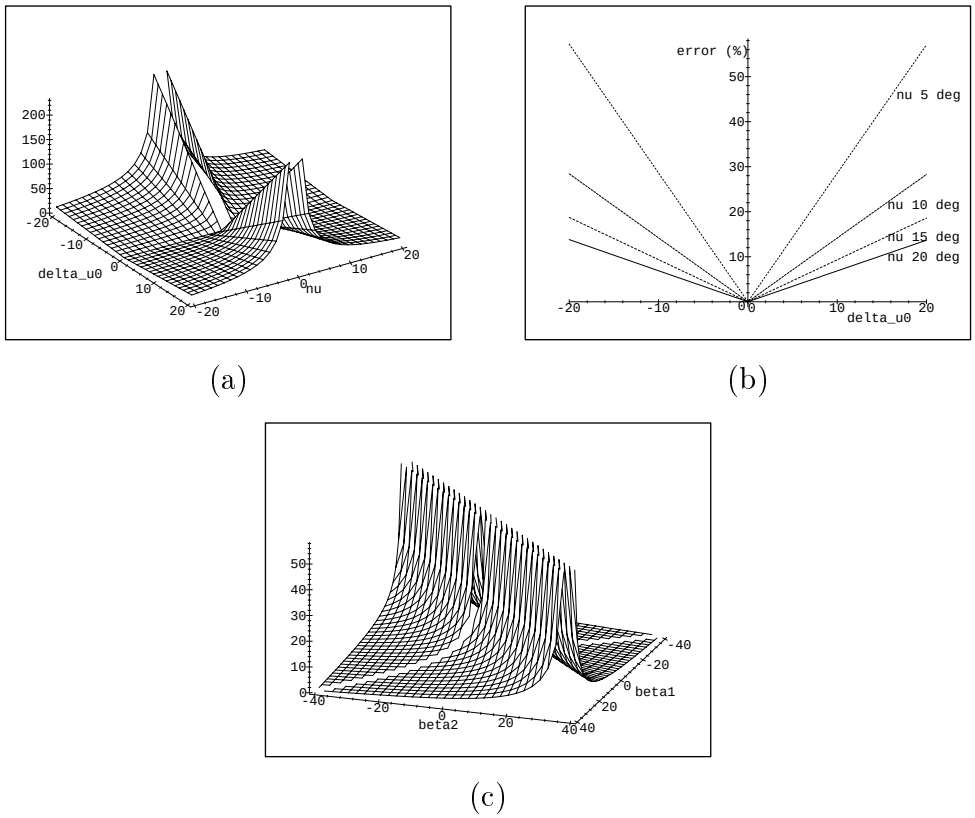
$$\hat{f}^2 = f^2 - \Delta u_0^2 - 2f\Delta u_0 \cot(\nu). \quad (31)$$

Note that when $\Delta u_0 = 0$, we obtain via self-calibration a perfect estimate, $\hat{f}$. Interestingly, note also that $\hat{f}$ is independent of the perturbation in the vertical coordinate of the principal point Δv_0. The relative error of $\hat{f}^2$ is given by

$$|\hat{f}^2 - f^2|/f^2 = |\Delta u_0^2 + 2f\Delta u_0 \cot(\nu)|/f^2,$$

and is depicted in Figure 4(a) as a function of Δu_0 and ν.

Immediately we observe that it is essential to ensure that the values of the vergence angles are dissimilar if we are to obtain a good estimate of the focal length. This is a more dominant factor than Δu_0 departing from zero. Clearly, the relative error of $\hat{f}^2$ increases without limit as ν approaches zero. This corresponds to an isosceles triangle configuration of the cameras of the



(a) (b)

(c)

Fig. 4. (a) Relative error (as %) of $\hat{f}^2$ versus Δu_0 and ν. (b) Relative error of $\hat{f}^2$ (as %) versus Δu_0 with $\nu = 5,\ 10,\ 15,\ 20\,\text{deg}$, in the order of decreasing magnitude of gradient of the lines. (c) Relative error of $\hat{f}^2$ (as %) versus β_1 and β_2 with $\Delta u_0 = 10$ pixels. Note that $\hat{f}$ is not computable at $\beta_1 = \pm\beta_2$. In each case, f has been fixed at 800 pixels in length.

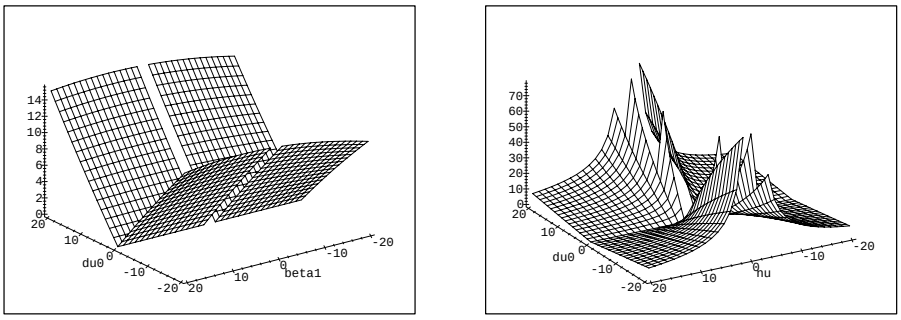
stereo head that was previously shown to be a degenerate singularity for self-calibration [3]. We also observe that accurate self-calibration will typically require that vergence angles differ by several degrees. Given f equal to 800 pixels and $\nu \leq 20\,\text{deg}$, then the relative error of $\hat{f}^2$ is shown in Figure 4(b) to vary approximately linearly when Δu_0 varies in the range $[-20...20]$ pixels. As an example, a difference in vergence angles of at least 5 degrees would be necessary to ensure that the focal length is obtained with a relative error under 10 % if the location of the principal point is known to within a horizontal distance of 5 pixels.

Figure 4(c) shows the relative error of $\hat{f}^2$ as a function of β_1 and β_2. As may be observed, an isosceles camera setup is not the only degenerate configuration for self-calibration. Focal length is not computable when $\beta \equiv \beta_1 + \beta_2 = 0$ as inspection of (30) reveals. This familiar degeneracy arises precisely when the optical axes in the successive camera positions are parallel, corresponding to a pure translation between camera views.

Using the direct formula for β_1 given in (10), the estimated vergence angle $\hat{\beta}_1$ becomes

$$\hat{\beta}_1 = \tan^{-1}\left(-\frac{\hat{f}\sin\beta_1}{\Delta u_0 \sin\beta_1 + f\cos\beta_1}\right). \tag{32}$$

The relative error $\left|(\hat{\beta}_1 - \beta_1)/\beta_1\right|$ is not given explicitly here, but is plotted in Figure 5. It is clear from Figure 5(b) that if the true value β_1 is held fixed while ν varies, and that $\Delta u_0 \neq 0$, the relative error of $\hat{\beta}_1$ grows without limit as ν approaches zero. However, if ν is maintained at 10 deg or more, then the relative error of $\hat{\beta}_1$ varies approximately linearly with principal point error Δu_0 (Figure 5(a)). Results analogous to these apply to $\hat{\beta}_2$.



(a) (b)

Fig. 5. (a) Relative error (as %) of β_1 versus Δu_0 and β_1 with ν fixed at 10 deg. (b) Relative error (as %) of β_1 versus Δu_0 and ν with β_1 fixed at 10 deg.

7 Experimental results with synthetic data

With the aid of synthetic tests, we now analyse the sensitivity of our self-calibration method to noise in the location of image points. Some consequences of deviations of the stereo head from the assumed model are also given.

A cloud of 35 points was randomly generated within a cubic volume of side 2400mm lying approximately 600mm in front of the stereo head. These points were then projected onto each of the 4 image planes arising in the two positions of the stereo head. The location of each image point was then perturbed in a random direction by a distance governed by a Gaussian distribution with zero mean and standard deviation, σ, expressed in pixel units. Such a distribution results in an expected value for the perturbation distance of approximately 0.8 σ. As a matter of interest, in the many tests carried out here, the highest perturbation distance was found to be 3.7 σ.

Left and right focal lengths were set at 6mm and 8mm, with a fixed baseline length of 300mm. Vergence angles were 15 deg and 17 deg in the initial position, and 18 deg and 22 deg in the final position. The motion of the head was such that the upward tilt was 10 deg, rotation of the baseline in the plane was 12 deg, with the length of the translation vector mapping the left camera from initial to final position being approximately the same as the baseline length of the head. Image sizes were 1000×1000 pixels.

Experiments were conducted with σ varying from 0.0 to 1.2 in steps of 0.1. For each value of σ, self-calibration was run 20 times (each time operating on a different set of images) and the root-mean-square (rms) error of each parameter was computed. Figure 6 gives a brief summary of how self-calibration is affected by increasing noise, in the case considered. Errors (rms) in lengths and tilt rotation are given as percentages of the true values, while errors (rms) in the vergence angles are expressed in degrees.

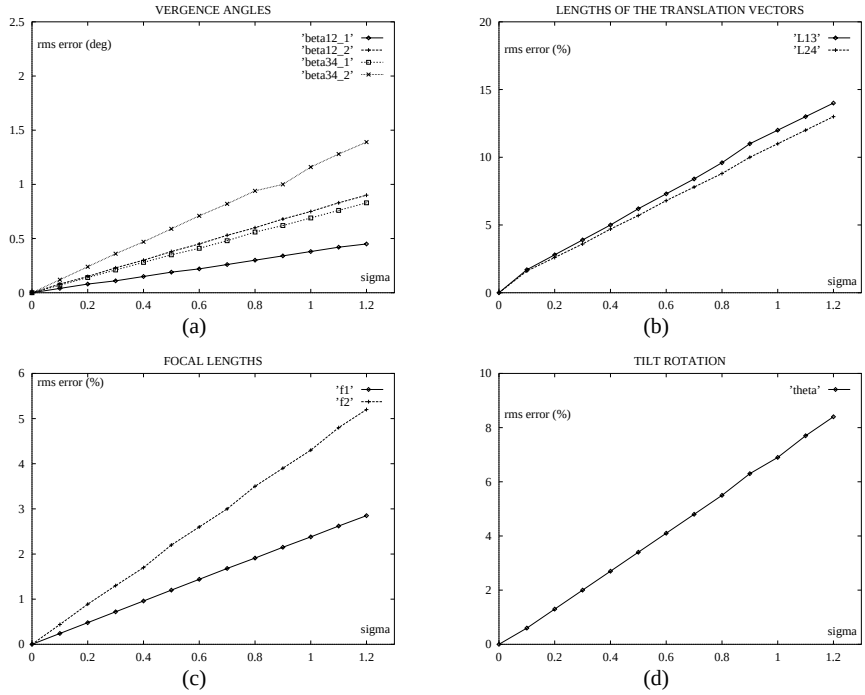


Fig. 6. Sensitivity of the self-calibration method to noise in the location of image points: results of synthetic tests.

We can see from the figures that errors in the estimates of the various parameters vary approximately linearly with the extent of the introduced noise, over the range considered. The lengths of the translation vectors L^{13} and L^{24} are the parameters most affected by noise, with relative errors of up to 14% occurring with noise $\sigma = 1.2$.

At this high noise level, the rms error of the right camera's focal length is 5.2%, the error in the estimated tilt rotation is 8.5%, and the maximum rms

error of the vergence angles is 1.4 deg. Comparable results were obtained for similar head movements.

Given that points can routinely be located by automated techniques with an accuracy of better than $\sigma = 0.5$, the above results suggest that this approach holds promise. Note that optimisation as a post-process of the estimates has not been attempted.

The errors induced in the estimation of the calibration parameters when the configuration of the stereo head departs from the assumed model of coplanar optical axes was also analysed via synthetic tests (see [1]). A difference in the elevation of left and right cameras of up to 15mm, given a baseline length of 300mm, was found to cause a relative rms error under 1% in the computation of the focal length, and an almost imperceptible error in the vergence angles. When both cameras had identical elevation, but one of the optical axes was tilted up to 3 deg, the relative rms error in the estimate of f was 1.2%, while the error in β_1 and β_2 was 0.1 deg. The simultaneous effect of both discrepancies was found to be more relevant causing relative errors in the focal length of up to 3% and an error of up to 0.3 deg in the vergence angles. These tests were carried out in the presence of noise in the image coordinates governed by a Gaussian distribution with zero mean and standard deviation, σ, of 0.5 pixels. Note that care was taken in these tests to avoid degenerate configurations of the synthetic stereo head.

8 Experimental results with real images

8.1 Self-calibration and reconstruction via a single moving camera

The self-calibration method was first tested on a single moving camera with its optical axis confined to the ground plane. The intrinsic parameters of the camera remained fixed. Note that this corresponds to case 2 in section 3 where direct formulae are obtained for the focal length of the camera and for the virtual vergence angles. Figure (7) shows the initial and final images taken by a CCD Panasonic camera with a focal length of 8.5 mm. The dimensions of the images were 768(cols) $\times$ 576(rows) and they were not corrected for radial distortions. The virtual left and right vergence angles were set at approximately 2 and 8 deg respectively. The principal point of the images was assumed to be located at the image centre for this experiment. The length of the baseline was estimated to be 300mm. Note that the scene contains a calibration grid. This was not used for calibration purposes but to provide ground truth for the 3D reconstruction as the positions of the points in the calibration grid pattern could be accurately measured.

After estimating the fundamental matrix that relates both views, the direct formulae for self-calibration given in (10) were used to compute the focal length and the virtual vergence angles. Metric reconstruction of the scene was then performed using the estimated values of the calibration parameters and knowledge of the baseline length. Various views of the reconstructed calibration grid have been displayed in Figure 8. The quality of the reconstruction is patent in view of the preservation of metric invariants such as the orthogonal relationship between the two boards of the calibration grid, the alignment of the squares and the equal lengths of the sides of different squares. The quantitative performance was measured by computing the length of the sides of the squares in the grid and the angle formed by the two pages of the calibration book. Encouragingly, the relative rms error in the measurement of the length of the sides of the squares was found to be 3%, while the rms error in the measurement of the angle was 2.2 deg.

Fig. 7. Pair of images obtained by a moving camera with fixed intrinsic parameters undergoing ground plane motion.

(a) (b) (c) (d)

Fig. 8. Metric reconstruction of the calibration grid: (a) Front view (b) Top view (c) Side view (d) Lateral view

8.2 Self-calibration and reconstruction via a dynamic stereo head

Many real-world experiments were carried out, one of which is detailed here. Figure 9 shows the initial and final stereo pairs used to capture an office scene. Two Pulnix TM-6CN cameras were employed, with left and right cameras having lenses with focal lengths of 12mm and 8mm, respectively. Note therefore that the right images exhibit a much wider field of view than their

left counterparts. The dimensions of all images were $768(\text{cols}) \times 576(\text{rows})$. No corrections were made for such as radial image distortion. After capturing the first stereo pair, the stereo head was moved away from the scene in a direction parallel to the ground plane, before being rotated a little about the vertical axis. The vergence angles of the two cameras were varied after the motion to maintain significant overlap of the four images, and the baseline of the stereo rig (the distance between the two optical centres) was kept approximately constant. Effort was made to ensure that the virtual vergence angles describing the orientation of the left camera's initial and final images were of sufficiently dissimilar absolute magnitude so as to avoid a sensitive configuration. A similar effort was made with respect to the right camera's images. A poor quality head apparatus was used, this acting to further challenge the robustness of the approach.

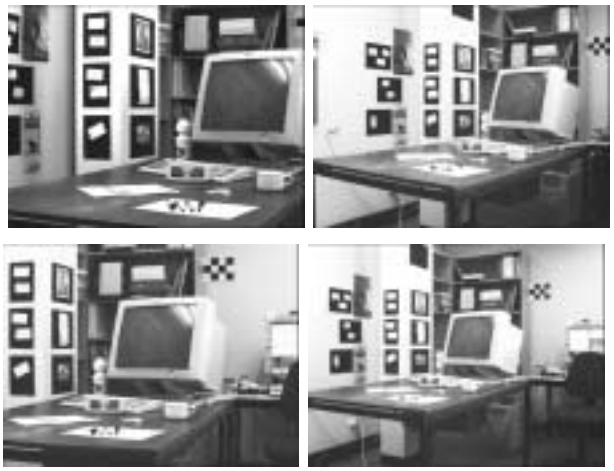


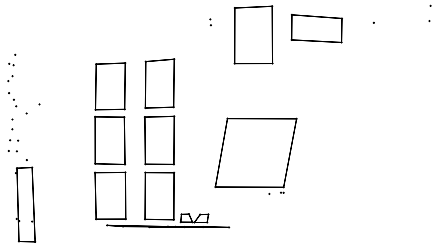
Fig. 9. Initial (above) and final (below) image pairs of a dynamic stereo head.

Feature points were determined and matched using the software of Zhang (the algorithm is described in [24]). The corresponding points were then expressed with respect to an origin at the predetermined principal points.

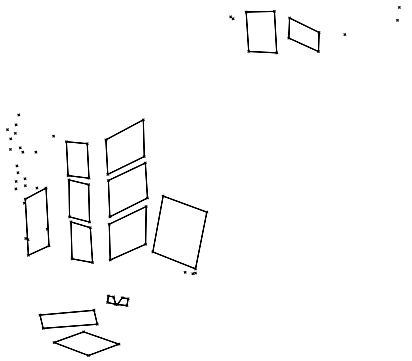
8.2.1 Self-calibration

The four fundamental matrices were computed using the technique described in section 5. The closed-form self-calibration method described in section 4 was then applied to obtain values for the vergence angles, focal lengths and motion of the head, from which reconstructions were computed via the standard method of triangulation. Applying the self-calibration formulae to these fundamental matrices, the effective focal lengths of the left and right cameras were estimated to be 1639.4 pixels and 1050.2 pixels, respectively. These distances have a ratio of 1.561, and relate to the nominal focal lengths of 12mm and 8mm. The left and right vergence angles (in deg) for the initial pair were estimated at 0.23 and 10.19, whilst those for the final pair were estimated at 3.97 and 7.11.

(a) Overhead view



(b) Table-level view



(c) Intermediate view

Fig. 10. Reconstruction from self-calibration

As noted previously, ground truth values for self-calibration parameters were not obtained (and for this reason head-motion estimates are not given). However, the quality of the estimates of the various parameters may be assessed indirectly by analysing the results of most importance, these being the reconstructions.

8.2.2 Reconstruction

The cluttered office scene contains a number of orthogonal surfaces of regular size that afford visual checking of the reconstructions. The vertical, rectangular pillar exhibits pieces of paper of sizes A5 and A4 on the left and right faces, respectively. Foreshortened A4 sheets are visible on the surface of the desk. The computer screen is tilted a little off the vertical.

The reconstructed scene was displayed using various visualisation tools. Three snapshots are given in Figure 10. To improve the visualisation of the reconstruction, some of the reconstructed 3D points have been connected with line segments.

The reconstruction proves to be of excellent quality, despite the fact that it is obtained without knowledge of focal lengths, relative orientations and motion of the stereo head. The overhead view shows that the reconstruction preserves very well the orthogonal and parallel relationships between various surfaces. The pieces of paper on the desk have good shape in the overhead view, while desktop points are seen to be coplanar in the table-level view.

Finally, we note that reconstruction is made more difficult when lenses with significantly different focal lengths are used, as was the case here. It is harder to obtain good estimates of the fundamental matrices in this situation since the spatial distribution of corresponding points is less extensive. Our reconstructions obtained with cameras having similar focal lengths prove to be more accurate.

9 Conclusion

A main aim of this work has been to show that there are some special sensitivities that must be taken into account when attempting to self-calibrate a stereo head. A further aim has been to explore and develop some of the computational techniques that need to be employed if a fully robust self-calibration method is one day to be attained.

New analytical forms were given for the fundamental matrix associated with the stereo head. Elegant and manipulable forms were obtained by adopting a novel parameterisation whereby relative orientation is expressed in terms of vergence angles. Inspection of the explicit forms of fundamental matrices associated with various configurations revealed that the stereo head is a degenerate set-up for self-calibration. This is meant in the sense that the simpler situation of a stereo head reduces the capacity to carry out self-calibration in comparison with the capacity to self-calibrate a pair of cameras in general po-

sition. The extent to which a reduced form of self-calibration may be achieved was then explored, and closed-form solutions were given. It was noted that when the cameras of the stereo head verge inwards by the same angle, the configuration becomes acutely degenerate.

In order to explore how a stereo head might more fully be self-calibrated, a single step motion of the head was considered. Here, a pair of images is captured at two separate times, the head having moved in the meantime. Clearly, more information is obtained in this way, but at the expense of introducing several new unknowns, since the motion of the head is assumed unspecified and the vergence angles may change. It emerged that full self-calibration could be obtained by considering not just the fundamental matrices associated with the left and right cameras in initial and final positions but also the fundamental matrix associated with each camera in its initial and final positions. Closed-form solutions in the elements of the four fundamental matrices were given for the parameters of the head and its motion. An extension to the analysis was given that permitted a measure of upward tilting of the head.

Robustness issues were considered next. It was examined how the fundamental matrix might be calculated in the case of a static stereo head with identical cameras, or, equivalently, a single, fixed camera undergoing ground plane motion. Computing an estimate of the fundamental matrix by assuming 5 zero entries (as per the analytical form) proved to be less effective than assuming there were only 2 zero entries. A procedure was given for calculating the fundamental matrix; this involved translation and scaling of data points, the use of an M-estimator technique and the epipolar distance for detecting outliers. This amounted to adapting standard methods to our particular problem.

The effect on the quality of self-calibration was then considered in relation to either error in the location of the principal point or proximity to an isosceles configuration. It was shown that cognisance of these sensitivities is necessary if self-calibration of a stereo head is to be successful.

Various practical experiments were conducted. In synthetic tests on images of a cloud of points, an assessment was made of the sensitivity of the self-calibration method to noise in the location of image points. Care was taken to avoid near-degenerate congurations. Empirical results suggest that errors in the estimates of various parameters vary approximately linearly with change in the standard deviation of noise in the locating of corresponding points. Our method was then applied to a pair of real images obtained by a single moving camera viewing a calibration grid, and the resulting reconstruction was found to be very accurate. Finally, two pairs of images of an office scene were captured by a moving stereo head. Self-calibration of the head and its movement was carried out, and the reconstructed scene again possessed excellent form, suggesting that the approach holds promise.

Further work would more fully explore robustness techniques for achieving improved reconstruction, including use of the cross fundamental matrices, and the rotational constraint. In addition, techniques might be developed for detecting the extent to which a stereo head departs from the idealised model, along with correction measures. However, it is clear that this endeavour will require the use of more than just corresponding points. Also of interest would be to more comprehensively model noise in the system and to develop statistical techniques for optimal parameter estimation (eg. see [13]).

Acknowledgements

The authors are grateful for the valuable remarks of the two referees which served to strengthen the paper. The authors are also grateful for the comments of Zhengyou Zhang, Heping Pan and Charlie Wiles. Michael Schneider provided helpful advice. This work was in part supported by Grant SAB95-0264, issued by the Dirección General de Investigación Científica y Técnica of Spain.

References

[1] L. de Agapito, *A Stereo Vision System for a Binocular Head: Hierarchical Feature Matching and Direct Methods for Self-Calibration*, PhD thesis, Departamento de Informática y Automática, Universidad Complutense de Madrid, Spain (1996).

[2] K. J. Bradshaw, P. F. McLauchlan, I. D. Reid, and D. W. Murray, Saccade and pursuit on an active head/eye platform, in: *British Machine Vision Conference* (1993), vol. 1, pp. 35–44.

[3] M. J. Brooks, L. de Agapito, D. Q. Huynh, and L. Baumela, Direct methods for self-calibration of a moving stereo head, in: B. Buxton and R. Cipolla, editors, *Proc. European Conference on Computer Vision* (Cambridge, UK, 1996), vol. 2, pp. 415–426.

[4] M. J. Brooks, W. Chojnacki and L. Baumela, Determining the egomotion of an uncalibrated camera from instantaneous optical flow, in: *Journal Optical Society of America A, 14, 10* (October 1997), pp. 2670-2677.

[5] J. L. Crowley, P. Bobet, and C. Schmid, Dynamic calibration of an active stereo head, in: *Proc. International Conference on Computer Vision* (1993), pp. 734–739.

[6] O. D. Faugeras, N. Ayache, and Z. Zhang, A preliminary investigation of the problem of determining ego and object motions from stereo, in: *Proc. International Conference on Pattern Recognition* (1988) vol. 1, pp. 242–246.

[7] O. D. Faugeras, Q. T. Luong, and S. J. Maybank, Camera Self-calibration: theory and experiments, in: G. Sandini, editor, *Proc. European Conference on Computer Vision* (Santa Margherita Ligure, Italy, 1992), Lecture Notes in Computer Science 588, pp. 321–334.

[8] M. A. Fischler and R. C. Bolles, Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the Association for Computing Machinery* 24(6):381–395 (1981).

[9] G. H. Golub and C. F. V. Loan, *Matrix Computations*, Johns Hopkins series in the mathematical sciences. Johns Hopkins University Press, Baltimore, 2nd edition (1989).

[10] R. I. Hartley, Estimation of relative camera positions for uncalibrated cameras, in: G. Sandini, editor, *Proc. European Conference on Computer Vision* (Santa Margherita, Ligure, Italy, 1992), pp. 579–587.

[11] R. I. Hartley, In defence of the 8-point algorithm, in: *Proc. International Conference on Computer Vision* (Cambridge, MA, USA, 1995), pp. 1064–1070.

[12] P. J. Huber, *Robust Statistics*, John Willey and Sons (1981).

[13] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, The Netherlands (1996).

[14] R. K. Lenz and R. Y. Tsai, Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 10(5):713–720 (1988).

[15] F. Li, M. Brady and C. Wiles, Fast computation of the fundamental matrix for an active stereo vision system, in: B. Buxton and R. Cipolla, editors, *Proc. European Conference on Computer Vision* (Cambridge, UK, 1996) vol. 1, pp. 157–166.

[16] Q. T. Luong and O. D. Faugeras, Active stereo with head movements, in: *Proc. Singapore International Conference on Image Processing* (1992), pp. 507–510.

[17] Q.-T. Luong, R. Deriche, O. Faugeras and T. Papadopoulo, On determining the fundamental matrix: analysis of different methods and experimental results, Rep. No. 1894, INRIA (1993).

[18] S. J. Maybank and O. D. Faugeras, A theory of self-calibration of a moving camera, *International Journal of Computer Vision*, 8, 2 (1992), pp. 123–151.

[19] H.-P. Pan, M. J. Brooks and G. N. Newsam, Image resituation: initial theory, in: *SPIE Videometrics* (1995), vol. 2598, pp. 162–173.

[20] P. Stelmaszyk, H. Ishiguro, and S. Tsuji, Mobile robot navigation by an active control of the vision system, in: *Int. Joint Conf. on Artificial Intelligence* (1991), vol. 2, pp. 1241–1246.

[21] P. Torr, *Motion segmentation and outlier detection*, PhD thesis, Department of Engineering Science, University of Oxford (1995).

[22] T. Viéville, Q. Luong and O. Faugeras, Motion of points and lines in the uncalibrated case, *International Journal of Computer Vision* 17, 1 (1994).

[23] T. Viéville and O. Faugeras, Motion analysis with a camera with unknown, and possibly varying intrinsic parameters, in: *Proc. International Conference on Computer Vision* (Cambridge, MA, USA, 1995) IEEE, pp. 750–756.

[24] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial Intelligence*, 75(1-2):87–120 (1995).

[25] Z. Zhang, Q.-T. Luong, and O. Faugeras, Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction, *IEEE Trans. on Robotics and Automation* 12(1):103–113 (1996).

[26] Z. Zhang, Determining the epipolar geometry and its uncertainty: a review, Report No. 2927, INRIA (1996).

[27] A. Zisserman, P. A. Beardsley, and I. Reid, Metric calibration of a stereo rig, *Proc. IEEE Workshop on Representations of Visual Scenes* (Cambridge, MA, USA, 1995), pp. 93–100.

A Notation semantics

Our notation differs from the standard notation of Faugeras et al. [7] (henceforth termed the Faugeras notation). Symbols $\mathbf{F}$, $\mathbf{T}$, $\mathbf{R}$ and $\mathbf{A}$ denote in this work the fundamental, translation, rotation and intrinsic-parameter matrices, respectively. Let the corresponding matrices in Faugeras notation be denoted F, T, R and A. Herein, the epipolar equation has the form $\mathbf{m}^T \mathbf{F} \mathbf{m}' = 0$, where $\mathbf{F} = \mathbf{A}^T \mathbf{T} \mathbf{R} \mathbf{A}'$. This contrasts with Faugeras notation, where $\mathbf{m}'^T F \mathbf{m} = 0$, and $F = A'^{-T} T R A^{-1}$. The full list of notational relationships is now given:

$$\mathbf{F} = \sqrt{\det(A)\,\det(A')}\,F^T, \quad \mathbf{A} = -\sqrt{\det(A)}\,A^{-1}, \quad \mathbf{R} = R^T, \quad \mathbf{T} = -R^T T R.$$