

Online Calibration of a 4 DOF Stereo Head

N. A. Thacker* and P. Courtney.

AI Vision Research Unit, University of Sheffield
Sheffield, UK

(* now at the Department of Electrical and Electronic Engineering,
University of Sheffield, Sheffield, UK)

Abstract

This paper addresses the problem of recovering accurate 3D geometry from a 4 degree of freedom stereo robot head. We argue that successful implementation of stereo vision in a real world application will require a self tuning system. This paper describes a statistical framework for the combination of many sources of information for the calibration of a stereo camera system which would allow continual recalibration during normal use of the cameras. The calibration is maintained using modules at three levels: fixed verge, variable verge and pan/tilt/verge calibration. Together these modules provide the means to fuse data obtained at various head positions into a single coordinate frame.

1 Introduction

Computer vision systems which can deliver an accurate estimate of 3D geometry from stereo are now relatively commonplace in the computer vision literature. Our own stereo vision system TINA has been demonstrated to have useful 3D vision capabilities [5]. The vision system, which makes use of edge based representation and stereo matching, relies upon calibration in both the determination of epi-polars for the matching process and the calculation of 3D position from disparity. Implementing these algorithms on a moveable head rig poses a real problem of multiple parameter calibration.

Calibration has generally been achieved by a procedure whereby the calibration parameters are recovered once from a known stimulus with no concern for updating this calibration in future by any other means other than total replacement. In a practical vision system which is to be in continual use, such one-off calibration methods are inadequate. A moving camera system would have to 'look' at a calibration stimulus every time it was moved. For a practical stereo vision system, recalibration must be an integrated activity working with data available during normal use [9]. In addition, we cannot expect there to be sufficient information at any one time to obtain a completely accurate calibration of the system, so we require a method of combining data collected over a period of time into a consistent calibration. Ideally this method would also allow the integration of information from different image sources and can be described as "online calibration". The idea is not new, other authors have suggested mathematical frameworks for self calibrating systems [2]. Here we describe our own practical framework and a three stage calibration system which

maintains the full head calibration (figure 1) and demonstrate its use with hand eye coordination of our head and a robot arm.

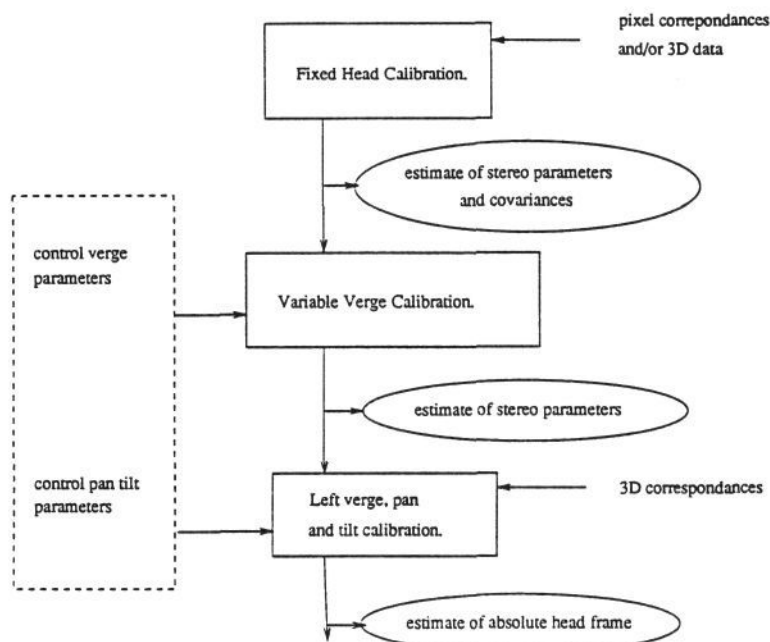


Figure 1. Head Calibration modules.

2 A Unified Mathematical Framework

A method is required for data combination, and this can be achieved via standard statistical methods by minimising a least-squares error measure (χ^2):

$$\chi^2 = (a - a_t)^T C_a^{-1} (a - a_t) + \sum_i (y_i - \phi_i(a_t))^T W_i^{-1} (y_i - \phi_i(a_t))$$

with respect to the parameters a_t , where χ_t^2 is a summed error criterion comprising a constraint term on the parameters a derived from previous data (which can be called a regularisation term) and a term for the current set of data y_i . C_a is the covariance matrix for the measurement vector a and the t subscripts denotes the iteration. W_i is the data measurement variance. This last term involves the data model ϕ , if this is linear then the model parameters can be estimated using a Kalman filter. If it is approximately linear then it can be linearised locally and solved using the Extended Kalman Filter (EKF). Both of these approaches are common in the computer vision literature [1].

The EKF uses the assumption that if ϕ is approximately linear then the χ^2 can be modeled as a quadratic around the current estimate at the minimum χ_0^2 :

$$\chi^2(\delta a) = \chi_0^2 + \delta a C^{-1} \delta a^T$$

where δa is the difference vector from the current estimate to the chosen point in calibration space. If this is true then the combined estimate of the total χ^2 from all combined data is also valid and should give the same result as if all the data had been minimised simultaneously. If however, the model is very non-linear the EKF may have to be iterated several times and depending on the degree of linearity this process may be unstable. Alternatively, the optimal combined estimate can be obtained directly by finding the parameters that minimise χ_t^2 . This is the method that we have adopted on the basis of increased robustness. We minimise the function iteratively using the downhill simplex method [4]. This gives the maximum amount of freedom for model parameter change and the inclusion of robust statistical measures. We limit the maximum contribution to the error score from each data point during minimisation, this effectively protects against outliers. Parameter tracking can be achieved by limiting the size of the covariance matrix so that new data takes preference over old [9].

To obtain a covariance matrix we must be minimising a χ^2 variable, this rules out a lot of calibration algorithms as candidates for optimal combination. Generally we cannot combine results unless the method takes correct account of the errors in the measurement system. In a stereo camera system we believe that the errors are mainly due to sampling noise and pixelation, and therefore best modelled in the image plane. The elements of the inverse covariance matrix are defined in terms of the *Hessian* by

$$C_{nm}^{-1} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial e_n \partial e_m}$$

which when close to the minimum of the function can be approximated using the *Jacobian*

$$C_{nm}^{-1} = \sum_i \frac{\partial \chi_i}{\partial e_n} \frac{\partial \chi_i}{\partial e_m}$$

We estimate the derivatives using numerical methods for purposes of model parameterisation flexibility. Any method that minimises an error metric in the image plane can be formulated as a χ^2 minimisation and combined within this statistical framework.

2.1 Fixed Stereo Camera Calibration

Obtaining a reliable stereo camera calibration is a particularly difficult task. A full camera model comprises both intrinsic parameters f (internal to the cameras) and extrinsic parameters e

(relative camera transformation specification). These two sets of parameters are often strongly correlated. For example the rotation of the camera and the translation of the centre of image co-ordinates will have virtually identical effects on our error criterion. The same is also the case for translation of the camera and changing the focal length. Correlations between parameters make it impossible to determine an isolated subset of the parameters properly without accurate prior knowledge of the remaining parameters. These correlations produce extended minima in the error surface so that many dissimilar sets of

calibration parameters may be equally valid, making statistical combination difficult.

The EKF and related methods (including ours), which take into account the covariance of the calibration parameters, can overcome these problems. New data is incorporated close to the current estimate where the covariance for the previous set of data is most reliable.

We can identify two sources of calibration data: known 3D measurements (for example the movement of a robot arm or known 3D objects) and epi-polar alignment of matched stereo correspondences. Both of these sources can be used to construct a χ^2 measured in the image plane. For calibration from 3D data it is assumed that data is provided on an accurately measured 3D object and that features on this object have been identified in the image planes of either camera. Such data can be obtained in a working system from the known motion of a robot arm or accurately known rigid 3D objects. A χ^2 is formed as the difference between the observed position of the image features and the predicted position, given the current estimates of the model. As the object is measured in an arbitrary co-ordinate frame the parameters describing the absolute transformation are redundant and only those relevant to the stereo camera system s are required. For calibrating from image correspondences the χ^2 is formulated following the numerical method of Trivedi [7]. The method can accommodate correspondence data either from matched epi-polar tangencies [3] or matched corners [8].

Thus for combination of calibration from these results with data from each source the stereo camera inverse covariance matrix C_s^{-1} is needed and the total combination cost function is given by :

$$\chi_t^2 = \delta s C_s^{-1} \delta s + \chi^2$$

Once this is minimised the stereo camera covariance matrix C_s^{-1} must be updated with the inclusion of all new data [10].

2.2 Variable Vergence Stereo Camera Calibration

A full parametric model of the vergence camera system would be capable of describing the whole space of possible configurations of the left right verge system. However, a global model is only applicable to a well engineered robot head. Moreover, this method of calibration would have to be developed almost independently of any other solutions for the fixed camera geometry so that we cannot build on previous methods.

Alternatively, we might adopt a look up table solution, this has the advantage that we can use the previous methods for calibrating fixed head configurations to fill the entries of the table. For a head like our own which has a movement resolution of 8 minutes of arc over a range of 60 degrees for two cameras there are 202500 possible configurations of which about 44300 may be regarded as viable stereo vision configurations (figure 2).

This degree of freedom we call an "asymmetric vergence" control paradigm. A simpler moveable head system is where the fixation of an object requires pan tilt and verge angles so that the left and right verge motors have equal and opposite motor position control parameters. This we call a "symmetric vergence" control paradigm. With this method however, the full number of possible configurations is 112. We may safely assume that each configuration

of the head would require an independent estimate of the extrinsic stereo camera parameters. Such independent entries for each configuration of the head would need extensive modification each time the system was disturbed.

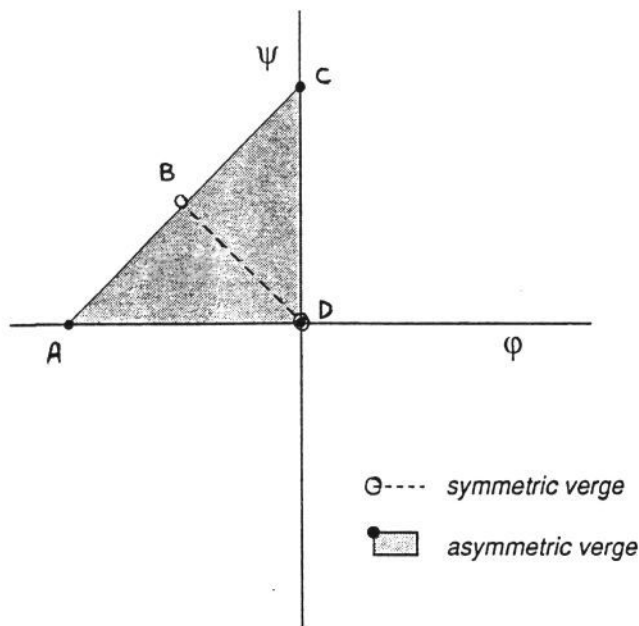


Figure 2. Defining interpolative lines and planes for symmetric and asymmetric calibration interpolation.

What we require is a compromise which combines the speed of training of global methods with the increased generality of the look-up table. The solution we have adopted is one which allows us to specify a local model that can be applied over a large number of possible verge configurations. We make the assumption that between two fixed verge configurations the stereo camera parameters can be linearly interpolated on the basis of the motor control parameters. In our camera model the relative stereo camera geometry e is stored as a quaternion and translation[9].

$$e = (q, t)^T$$

Linear interpolation would imply that the net effect of a change in verge rotation for both cameras about their verge axes can be approximated by a rotation about one axis and that small rotations about this axis are approximately linear in the left and right verge position control parameters Φ and Ψ . The range over which these assumptions are valid have been tested by simulation and it was found that the approximation will not measurably degrade 3D

geometry in our system over a rotation range of 0.2 radians (Figure 3).

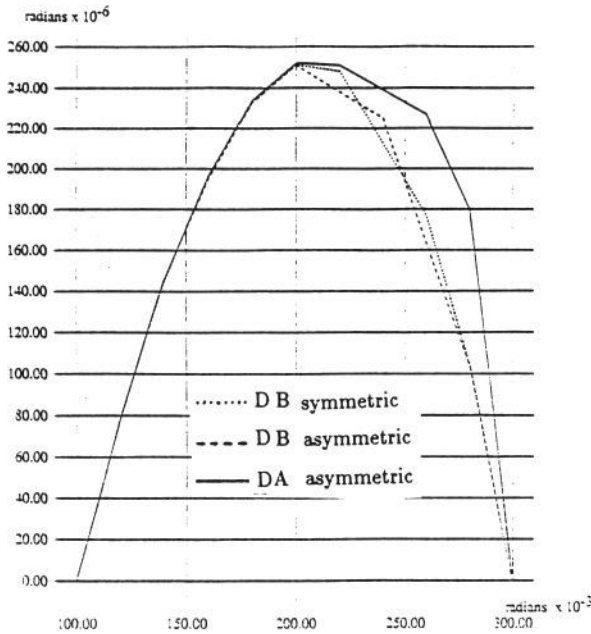


Figure 3. Simulated estimate of the error on interpolated vergence rotation for symmetric and asymmetric control paradigms.

Rotation errors are approximately quadratic and a maximum at the centre of the interpolated region. Thus, for the symmetric verge paradigm the whole useful range of the system can be defined with two calibration configurations defining the endpoints of the interpolation line e_1e_2 at Φ_1 and Φ_2 . The interpolated estimate of the relative camera geometry is given by

$$\hat{e} = \alpha_1 e_1 + \alpha_2 e_2$$

with

$$\alpha_1 = (\phi - \phi_1)/(\phi_2 - \phi_1)$$

$$\alpha_2 = (\phi_2 - \phi)/(\phi_2 - \phi_1)$$

Thus we have two stored calibrations as opposed to the 112 entries needed for a look-up table.

Similarly, for the asymmetric vergence paradigm, a large part of the useful range of the cameras can be defined using only three calibration configurations e_1 , e_2 and e_3 , which define an interpolative plane in configuration space.

$$\hat{e} = \beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3$$

with

$$\beta_2 = \frac{(\Psi_3 - \Psi_1)(\Phi - \Phi_1) - (\Psi - \Psi_1)(\Phi_3 - \Phi_1)}{(1 + (\Psi_2 - \Psi_1)(\Phi_2 - \Phi_1))(\Phi_2 - \Phi_1)(\Psi_3 - \Psi_1)}$$

$$\beta_3 = \frac{(\Phi_2 - \Phi_1)(\Psi - \Psi_1) - (\Phi - \Phi_1)(\Psi_2 - \Psi_1)}{(1 + (\Phi_2 - \Phi_1)(\Psi_2 - \Psi_1))(\Psi_3 - \Psi_1)(\Phi_2 - \Phi_1)}$$

$$\beta_1 = 1 - \beta_2 - \beta_3$$

By comparison, for verge rotations up to 0.2 radians this same region would require 3700 separate calibration entries in a standard look up table. More of the configuration space can be calibrated by defining similar three-point regions and calibrating these separately. We do not advocate extrapolation of the calibration outside the defined interpolation triangle.

The full symmetric and asymmetric models g including intrinsic parameters f_l and f_r can be written as

$$g = (f_l, e_1, e_2, f_r)^T$$

or

$$g = (f_l, e_1, e_2, e_3, f_r)^T$$

The current estimate of the calibration can be written as

$$\hat{s} = (f_l, \hat{e}, f_r)^T = g \nabla_g(s)$$

Our fixed camera calibration methods provide an estimate of the intrinsic and extrinsic camera parameters s and a full covariance matrix C_s^{-1} . These are combined into the model g using standard statistical methods as follows

$$g_t = g_{t-1} + C_{gt}(\nabla_g(s))^T C_s^{-1}(s - \hat{s}_{t-1})$$

$$C_{gt}^{-1} = C_{g(t-1)}^{-1} + (\nabla_g(s))^T C_s^{-1} \nabla_g(s)$$

It is vital for the stability of these methods that estimates of parameters s from new data is constrained with estimates of these parameters from previous data \hat{s} . This constraint must then be removed from the combined estimate of g to prevent double counting of data [10].

2.3 Calibration of the Pan/Tilt/Verge rotation axes

The above methods provide an interpolative estimate of the relative camera geometry and the intrinsic camera parameters for a restricted range of verge configurations. This can be used to provide estimates in 3D of the location of any observed stereo correspondance in the left camera coordinate system. We now need to be able to relate these coordinate frames for any configuration of the head. In practice this requires the determination of the translations between the pan tilt and left verge rotation axes and rotation scale factors. We call this set of parameters the head calibration parameters j . In practice only the parameters defining the rotation scale factors and transformation of the left camera into the left verge co-ordinate frame require calibration. The remaining transformation parameters can be determined by direct measurement.

We have implemented two methods for achieving this, the first again uses the robot and minimises error in the back projected position in the left image plane of the robot arm subject to j . The second uses the estimate of the verge

calibration and a stereo/temporal corner matcher [8] to provide estimates of temporally matched 3D image points in the left camera coordinate system. A χ^2 is then formed from the summed squared error between the first point and the second point transformed back into the first head configuration in scaled disparity space $(x/Az, y/z, I/\sqrt{2}.z)$ where A is the aspect ratio of the cameras and I the interocular separation. Disparity space errors provide a scaled approximation to the image plane error but are much easier to compute [6].

3 Results and Conclusions

The data combination method for fixed camera geometry was tested using data from robot motion, matched stereo corner correspondences and a calibration tile. The accuracy of estimated epipolar geometry was found to improve with the inclusion of new data as expected (figure 4). This system will allow the online recalibration of a fixed verge stereo camera system.

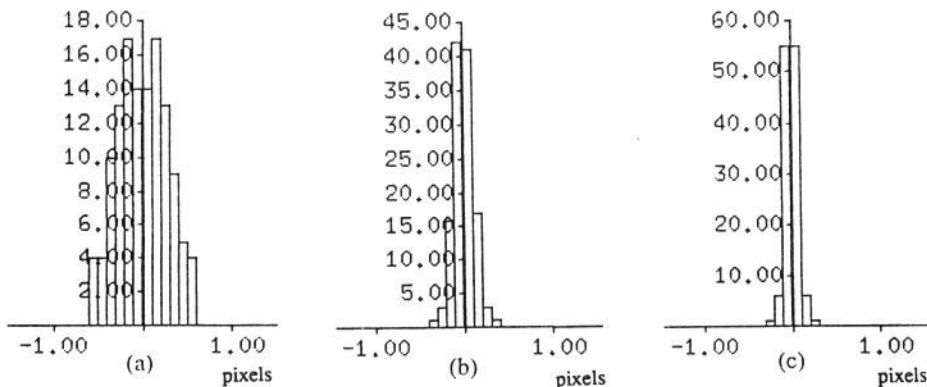


Figure 4. Epi-polar errors on test data after combination of data from
 (a) robot position,
 (b) matched stereo correspondences and
 (c) a calibration tiles.

The variable verge interpolation scheme was initialised with three fixed verge calibrations at the points $(0,-30)$, $(30,0)$ and $(30,-30)$ in verge motor configuration space (one motor count = 2.5 mRadian). This was tested by interpolating the camera geometry at the point $(25,-25)$, the epi-polar accuracy was found to be consistent with a fixed verge calibration at the same point (figure 5). This system supports online recalibration of a variable verge stereo camera system using data from the fixed verge calibration method over a relatively large range

of rotations.

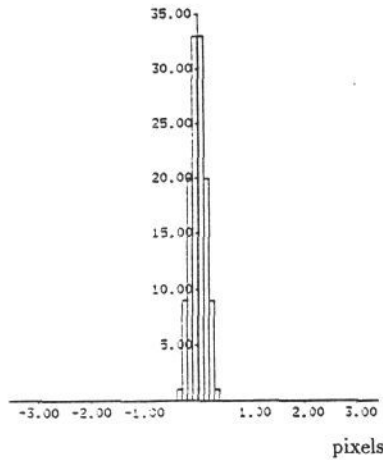


Figure 5. Epi-polar errors for interpolated camera calibration at (25,-25).

The pan/tilt and left verge parameters were obtained by calibrating on the back projected robot motion and matched static 3D points. The resulting full camera model was tested on unseen robot positions (figure 6). Outliers can be seen which are generated by several causes including stereo mis-matching and undershoot of the robot arm. This parameterisation of the system now permits 3D data from different head configurations to be combined into one coordinate frame and computation of head configurations for fixation of 3D world points.

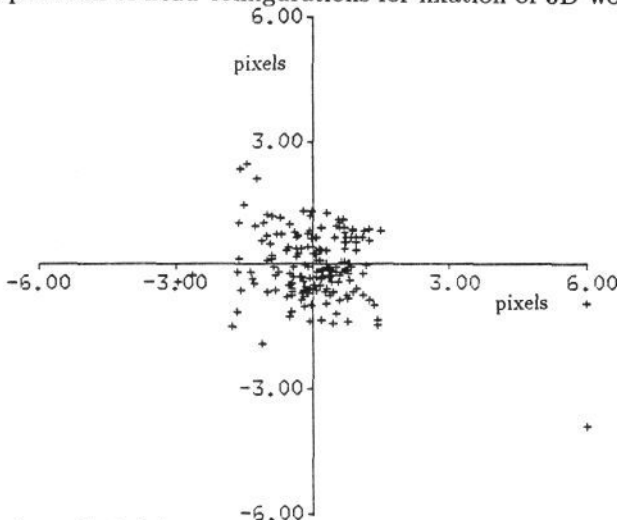


Figure 6. Back projected image plane errors after calibration of the full 4 DOF head system.

These results show that accurate calibration of a 4 DOF stereo head is feasible using a robust statistical framework which will permit online recalibration. Methods in stereo computer vision like those in our own TINA vision system requiring accurate stereo camera calibration can be supported with such a system.

Acknowledgements

We gratefully acknowledge the grant holders Prof. John E.W. Mayhew Dr. Paul Dean and Prof. John Frisby, the support of ESRC/MRC/SERC and the EEC, under the ESPRIT VOILA project, for the funding of this work. We also wish to thank our colleagues at AIVRU especially Li-Dong Cai, John Porrill and Stephen Pollard.

References

- [1] Ayache, N., **Artificial Vision for Mobile Robots**, MIT Press. 1991.
- [2] Faugeras, O.D. and G. Toscani, The Calibration Problem for Stereo, Proceedings of the CVPR, pp.15-20 1986.
- [3] Porrill, J. and S. Pollard, Curve Matching and Stereo Calibration, Image and Vision Computing, pp. 45-50. vol 9 no. 1 feb. 1991.
- [4] W.H.Press B.P.Flannery S.A.Teukolsky W.T.Vetterling, Numerical Recipes in C, Cambridge University Press 1988.
- [5] Rygol,M., S.B. Pollard, C.R. Brown, A Multiprocessor 3D Vision System, Concurrency vol 3, no 4 1991.
- [6] Sparks,E. and M Stephens, Integration of Stereo and Motion, Proceedings of BMVC 90.
- [7] Trivedi,H.P., Estimation of Stereo and Motion Parameters using a Variational Principle, Image and Vision Computing, v5, n2 ,pp.181-183 May 1987.
- [8] Thacker,N.A., Y.Zheng and R. Blackbourn, Using a Combined Stereo/Temporal Matcher to Determine Ego-motion, Proceedings of BMVC 90.
- [9] Thacker,N.A. and J.E.W.Mayhew, Optimal Combination of Stereo Camera Calibration from Arbitrary Stereo Images, Image and vision computing, pp.27-32 vol 9 no 1 Feb., 1991.
- [10] Thacker,N.A. and P. Courtney, Online Stereo Camera Calibration, AIVRU internal memo 62.