Tina Memo No. 1992-002
Internal.

Online Stereo Camera Calibration.

Neil A. Thacker and Patrick Courtney.

Last updated
6 / 9 / 2005

This document forms part of the **Features and Measurement** Series
available from www.tina-vision.net.

WWW.TINA-VISION.NET

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Online Stereo Camera Calibration.

Neil A. Thacker and Patrick Courtney

while at Vision Research Unit, University of Sheffield

Preface

The work presented in this document was not published in a journal. However the approach was extended to accommodate a 4 DOF stereo head and this was published a year later at the BMVC. The stereo head rig has now been dismantled, but, this document describes the basis for the calibration software which still survives in the **Calib Tool** *in the Tina vision system, and the motivation for taking this approach.*

Abstract.

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 results in a calibration system that would allow continual re-calibration during normal use of the cameras. We demonstrate this with the use of measured 3D objects, accurate robot motion and matched stereo features in a working vision system.

Introduction.

Camera calibration is necessary to relate camera data to the physical world and it is also important in making some vision problems tractable. Our own stereo vision system TINA has been demonstrated to have useful 3D vision capabilities (Rygol et al 1991) that may soon find real applications in an industrial environment. 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.

In the past, approaches to camera calibration have concentrated on computational efficiency and the choice of camera model (eg: Tsai 1987). Often the search for computational efficiency has led to solutions which compromise the requirements of the original problem and result in inaccuracy and bias. With the continued development of computer technology, computational efficiency becomes less and less important in the solution of such problems. Without the requirement of algorithmic speed many of the criteria for judgement of calibration algorithms simply disappear. We are left with the question: what is it we require in order to use stereo computer vision algorithms in a real environment?

Calibration has generally been done by a procedure whereby a known object (calibration stimulus) is placed in front of the cameras and the imaged positions of features on this object have been used to recover camera model parameters. This calibration stimulus is often a specially constructed object which would generally not be visible to the system in normal use. The calibration parameters are recovered once 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 and perhaps have some freedom of movement, such one-off calibration methods are next to useless. For fixed cameras the normal flow of visual input would need to be interrupted periodically for re-calibration. A moving camera system would have to 'look' at a calibration stimulus every time it was moved. In order to have a practical stereo vision system, re-calibration must be an integrated activity working with data available during normal use (Thacker and Mayhew 1991). In addition we cannot expect there to be enough information at any one time to obtain a complete accurate calibration of the system so we require a method of combining data over a period of time into a consistent calibration. Ideally this method would also allow the integration of information from different sources and can be described as "online calibration". The idea is not new, other authors have suggested mathematical frameworks for self calibrating systems (Faugeras and Toscani 1986). Here we demonstrate the practical realisation of such a scheme.

Stochastic vs Deterministic Design Principles.

Science proceeds by building models and testing them with data. In quantitative science models fall into two categories, *deterministic* models and *stochastic* (probabilistic) models. The most familiar examples of determistic models are the laws of classical physics, where simple observable variables are related by parameters. Unfortunately, the phenomena studied by scientists are rarely as predictable as the simple deterministic theories would suggest. In the presence of measurement errors and variability of experiments, deterministic formulae are unable to give definitive estimates of parameters. It then becomes necessary to extend these deterministic models so that we can predict the range of probable future observations and indicate degrees of uncertainty. Problems such as this prompted the mathematician Gauss to develop his *theory of errors*, and later led others such as Neyman and Fisher, to lay the groundwork for what we now refer to as the science of statistics.

Many approaches to camera calibration are based upon a deterministic approach to estimation of parameters. Such work can be recognised as formulating sets of equations based upon classical geometrical problems and then attempting to solve the resulting expressions. Though this is often done in the guise of minimisation framework, this approach should not be confused with a truly quantitative estimation.

Generally, a quantitative estimation, with appropriate estimates of uncertainty or precision, will require the explicit use of a solid methodology, such as *Likelihood.* Without such an approach it becomes impossible to perform the tasks required of estimated parameters, in particular we cannot perform error estimation, data fusion and parameter updates. Although calibration provides only one example of this issue, we could argue that without the use of a quantitative methodology, many of the basic questions in science are just not approachable. Algorithms for calibration and estimation based upon solution of deterministic models may therefore be adequate for graphics applications (where the data is all accessible to whatever accuracy is required), but should be deemed strictly unsuitable for image interpretation.

We should therefore prefer to see any algorithms intended for use in a vision system designed explicity using statistical principles. Importantly, these approaches should identify any assumptions made regarding not only the deterministic models on which they are based, but also any assumed statistical distributions, which lead to specific forms of algorithm. Only by doing this can the various alternatives between algorithms and implementations be theoretically compared and later experimentally tested [1], there is no alternative [2].

A Unified Statistical Framework.

A method is required for data combination, and this can be achieved via standard statistical methods by minimising the quantity:

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

with respect to the parameters a_t , where χ^2_t 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 can be derived from probability theory on the assumption of Gaussian distributions for both the fitted data and the parameter estimates. The 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). These approaches are both common in the computer vision literature (Ayache 1991).

The elements of the inverse covariance matrix are defined in terms of the *Hessian* by

$$C^{-1}_{nm} \ = \ 1/2\partial^2\chi^2/\partial e_n \partial e_m$$

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

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

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^2_0 :

$$\chi^2(\delta a) \ = \ \chi^2_0 \ + \ \delta a C^{-1} \delta a^T$$

[1] Research which ignore these issues may well result in algorithms which cannot be objectively tested and may therefore be regarded as strictly unscientific.

[2] Well there is, it's called chaos.

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. Parameter tracking can be achieved by limiting the size of the covariance matrix so that new data takes preference over old (Thacker and Mayhew 1991).

Sources of Data.

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. 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. For the purposes of this paper we will identify two constraints on visual data which can be used for online camera calibration.

The first constraint is derived with known 3D information which can be associated directly with image measurement. One such source of calibration data is the visual appearance of simple calibration stimuli, but models of accurately machined work objects could also be used. A common use for a vision system is in the visual control of a robot manipulator. The movement of these devices is generally quite accurate Cartesian co-ordinates (1 mm. for our RTX robot) and the movement of a few points on a manipulator over a period of time would provide useful information about the robot to camera transformation as well as the relative stereo camera transformation.

The second constraint is that stereo matched features from an unknown 3D environment must lie on epi-polar lines, which are calculatable from relative camera geometry. We have two methods for extracting such data, the first is corner matching (Thacker 1991) and the second is curve matching which delivers accurately measured matched epi-polar tangencies (Porril 1990).

Both of these sources of data can be formulated to give χ^2 error functions in the image plane.

Stereo Camera Calibration.

Obtaining a reliable stereo camera calibration is a particularly difficult task. A full camera model comprises both intrinsic parameters (internal to the cameras) and extrinsic parameters (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 often have virtually identical effects on our error criterion. The same is also the case for translation of the camera and changing the focal length. These correlations produce extended minima in the error criterion so that many dissimilar sets of calibration parameters may be equally valid. Correlations between parameters make it impossible to determine an isolated subset of the parameters properly without accurate prior knowledge of the remaining parameters. Incorrect specification of the remaining parameters can lead to large systematic errors when the calibration results are used to relate observed image data back to the world.

A particular set of data need not be sufficient to determine or even provide constraints on all of the parameters. Some forms of 3D information are insufficient to determine all of the camera parameters uniquely. For example, planar data can only provide enough information to determine 8 free parameters in the absence of radial distortion and this is generally fewer than the number required for a complete camera model. Calibration with such data leads to problems (such as depth compression or elongation) when attempting to make visual 3D measurement away from the area of the planar data. Epi-polar information obtained from stereo matching does not constrain the inter-ocular separation and only provides weak constraints on the intrinsic camera parameters.

Attempts to combine isolated calibrations on several sets of data using their covariances are unstable because the separation of the parameters in parameter space is large compared to the range over which the covariance matrix can provide a reliable model of the error surface. The EKF and related methods (including ours) overcome these problems [Ayache 1991]. New data is incorporated close to the current estimate where the covariance for the previous set of data is most reliable.

The Camera Model.

The robust method of data combination that we have adopted is relatively less sensitive to specific forms of the camera model than standard calibration procedures operating on a subset of calibration parameters to give a closed form solution. Obviously the camera model should describe the stereo camera system with a minimum number of degrees of freedom. However, provided the model adequately describes the data many representations of the same system will probably be just as good.

First we must choose a parameterisation of the stereo camera system which is suitable for calibration from stereo matches. Our starting point is the Trivedi algorithm [3], the method minimises the distance that two known correspondences must be moved on the image plane in order to bring them into epi-polar agreement. The Trivedi formulation of this solution uses a quaternion representation of the between camera transformation (e_0, e_1, e_2, e_3) and a unit vector (e_4, e_5, e_6) for the direction of translation which combined with two constraint equations

$$e_0^2 \ = \ 1 \ - \ e_1^2 \ - \ e_2^2 \ - \ e_3^2 \qquad e_4^2 \ = \ 1 \ - \ e_5^2 \ - \ e_6^2$$

leaves only five free parameters e_1, e_2, e_3, e_5, e_6. This is a particularly useful method of taking into account all of the constraints on the transformation between two coordinate frames. The final parameter that is needed to completely describe the relative camera transformation is the inter-occular separation (e_7). The remaining parameters required to describe the full stereo camera system are the intrinsic camera parameters for the left and right camera. The fitted camera variables are;

a) image centres ($f_1 and f_2$).

b) focal length (f_3).

c) aspect ratio (f_4).

d) radial distortion (f_5 *to* f_{5+n}).

The number of radial distortion terms and the specific model adopted is dependent on the optical qualities of the lenses, for our purposes we find that one term, first order (quadratic) distortion, is often sufficient (but see Weng et al. 1990). The full stereo camera model thus includes two sets of intrinsic camera parameters f_l , f_r and one set of transformation parameters e.

Calibration from 3D data requires the determination of a further transformation and we choose that from the 3D model co-ordinate system to the left camera frame comprising

a) quaternion representation of rotation matrix ($g_1 - g_3$).

b) translation vector to the object origin ($g_4 - g_6$).

Calibrating from known 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. 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, on the basis that the source of error is in the image measurement. As the object is measured in an arbitrary co-ordinate frame the g parameters are redundant and only e , f_l and f_r are relevant to the stereo camera system, we shall call this subset of parameters s. Thus for combination of calibration from these results with data from a different coordinate frame only the stereo camera inverse covariance matrix C_s^{-1} is needed and the total combination cost function is given by :

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

The first term may be derived from physical estimates of camera model parameters and their expected error and thereby used for regularisation. χ_t^2 is minimised using the parameters for the left camera to object transformation g and the stereo camera s . This is done using a robust iterative numerical minimisation algorithm (we use the simplex minimisation of Nelder and Meade [2] but any robust method may be adopted).

The stereo camera covariance matrix C_s^{-1} must be updated with the inclusion of all new data. To do this we first calculate the full covariance matrix C using singular value decomposition to reduce numerical instability (Cai 1990).

$$C \ = \ (J^T J)^{-1} \ = \ (V W^2 V^T)^{-1}$$

where J is the Jacobian matrix ($\nabla_t \chi$). This matrix can be computed via derivative methods using the camera model or by numerical estimation. Numerical estimation gives greatest flexibility for subsequent modification of the camera model. This gives

$$C = VW^{-2}V^T$$

The columns and rows corresponding to the g parameters are eliminated from this matrix to determine the contribution to the camera covariance matrix from the current set of data. Finally this matrix is inverted by a robust numerical inversion scheme (we use singular value decomposition again) and added to C_s^{-1} to obtain the total stereo camera inverse covariance.

Calibrating from Image Correspondences.

The χ^2 is formulated following the numerical method of Trivedi (Trivedi 1987) with the inclusion again of the regularisation term obtained from previous calibrations. Minimisation of this χ^2 subject to the stereo camera calibration parameters e , f_l and f_r determines the optimal set of parameters which are consistent with previous calibrations and the current set of data. Although there are no constraints from this new data on the inter-occular separation and only weak constraints on the intrinsic parameters they may change during minimisation, due to the correlations stored in the covariance matrix between these parameters and those which are constrained by the new data. Failure to take correct account of these correlations would lead to a camera model which may produce reliable epi-polars but would no longer be suitable for calculating 3 dimensional position from disparity. The contribution to the stereo camera covariance matrix can be computed directly from the χ^2.

$$C^{-1} = J^T J$$

The total inverse covariance matrix is updated as for the previous method.

Results.

An RTX robot arm was positioned before a set of stereo cameras at a distance of 5 - 10 inter-occular separations. A feature on an object held in the robot gripper was identified and tracked though 80 random positions visible from both cameras (Figure 1(a) and (b)). The known 3D position of the robot arm and the corresponding positions in the left and right images provide suitable data for camera calibration (Figure 1(b)). We wish to show that the combination method can be used to update and refine an ongoing estimate of the calibration parameters. For this the data was processed twice, once with the full data set ((calibration (A)), and again as four groups of 20 data points with three combinations (calibration (B)). At each combination stage the covariance on the absolute transformation parameters is lost but those on the stereo camera transformation and the intrinsic parameters are maintained. Thus we would expect calibration (A) to give the highest accuracy. Figure (4) shows the radial error distribution of the total set of 80 data points for each calibration result. In experiment (B) it was found that the accuracy of the calibration as a description of the complete data set improved with each additional set of data and the final result provides as good a model of the camera parameters as the results from experiment (A). For these cameras there was no observed trend on the the radial error plots to indicate radial distortion in these particular lenses so the radial distortion term was set to zero. Protection of the algorithm against rogue data points (due to hard limiting of the robot motion for example) was crucial. For this we employed a statistical test on the back projected position of each new data point to check that it was consistent with at least one of the current camera models.

To demonstrate the successive combination of stereo matches and known 3D data the stereo corner matching algorithm (Thacker 1991) was run on the scene shown in figure 2(a). A further set of 3D data was then obtained from a conventional calibration grid (Figure 3(a) and (b)). All three sets of data were then combined using a-priori estimates of nominal image location accuracy for each data set. These differences arise because the tile data is obtained from the intersection of fitted lines (0.1 pixels), the corners are located using a standard corner detector (Harris 1988) (0.3 pixels) and the RTX data contains both corner detection errors and RTX location accuracy (0.5 pixels).The epi-polar errors of a set of test tile data were used to monitor the accuracy of the calibration. Again accuracy improved with addition of each set of data (Figure 5).

Conclusions.

We have shown that by adopting a unified statistical framework for camera calibration based on error measurement in the image plane it is possible to use standard statistical method to combine two sources of data into a single calibration estimate. We suggest that such methods of calibration will be necessary if stereo vision systems are ever to be used in real world applications where periodic re-calibration using standard methods may be impractical. We have demonstrated the utility of this method on three sources of vision data including robot motion and stereo correspondences, both of which could be obtained during normal use of the vision system. Once this framework of data combination has been adopted any algorithm which minimises errors in the image plane can be formulated for incorporation. Other methods for camera calibration which enforce particular physical properties of the world such as rigidity, perpendicularity, planarity could be investigated.

References.

Ayache, N. "Artificial Vision for Mobile Robots." MIT Press. 1991.

Faugeras, O.D. and G. Toscani, "The Calibration Problem for Stereo" Proceedings of the CVPR, pp.15-20 1986.

Harris, C. and M.Stephens, "A Combined Corner and Edge Detector.", Proceedings of the Fourth Alvey Vision Conference. pp.147-151 August 1988.

Porrill, J. and S. Pollard. ,"Curve Matching and Stereo Calibration.", Image and Vision Computing., pp. 45-50. vol 9 no. 1 feb. 1991.

W.H.Press B.P.Flannery S.A.Teukolsky W.T.Vetterling, Numerical Recipes in C. Cambridge University Press 1988.

Rygol,M., S.B. Pollard, C.R. Brown., "A Multiprocessor 3D Vision System.", Concurrency, 3,4 1991.

Trivedi,H.P., "Estimation of Stereo and Motion Parameters using a Variational Principle.", Image and Vision Computing, 5,2,pp.181-183 May 1987.

Tsai,R.Y., "An efficient and Accurate Camera Calibration Technique for 3D Machine Vision.", IEEE Computer Vision and Pattern Recognition, pp.364-374 1987.

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.

Thacker,N.A., "Corner Detection and Matching.", AIVRU Internal memo, (submitted to ECCV 1992) 1991.

Li-Dong Cai. and N.A. Thacker., "Some notes on Solution of Data Modelling of Camera Calibration.", AIVRU internal memo.

Weng J., P Cohen and M. Herinou., "Calibration of Stereo Cameras using a non-linear Distortion Model.", IEEE Pattern Recognition, pp.246-25 2 1990. .bp

Figure Legends.

Figure 1 (a). The RTX robot arm and matched corner features.

Figure 1 (b). Three dimensional reprojection of robot positions used for calibration (RTX data).

Figure 2 (a). Left image of a scene of a box with matched corner features shown as crosses.

Figure 2 (b). Three dimensional reprojection of matched corner features used for calibration (BOX data).

Figure 3 (a). Calibration grid.

Figure 3 (b). Three dimensional reprojection of data used for calibration (GRID data).

Figure 4. Results from experiment (A).

(a). Radial error versus radius (radial distortion) of RTX data from the calibration for the whole data set (80 points).

(b). Radial error versus radius (radial distortion) of RTX data with final calibration from the combination of 4 x 20 data points.

Figure 5. Results rom experiment B.

(a). Off epi-polar error for test data after calibrating with the RTX data set.

(b). Off epi-polar error for test data after combination of the RTX and BOX data set.

(c). Off epi-polar error for test data after combination of the RTX , BOX and GRID data sets.