Tina Memo No. 1991-002
Image and Vision Computing, 9(1), 27-32, 1990.

Optimal Combination of Stereo Camera Calibration from Arbitrary Stereo Images.

N.A.Thacker and J.E.W.Mayhew.

Last updated
6 / 9 / 2005

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

Optimal Combination of Stereo Camera Calibration from Arbitrary Stereo Images.

Neil A. Thacker, John E. W. Mayhew. while at AI Vision Research Unit, University of Sheffield

Abstract.

Many stereo correspondence algorithms require relative camera geometry, as the epipolar constraint is fundamental to their matching processes. We intend to build a eye/head camera rig to mount on the mobile platform COMODE to enhance the abilities of the TINA system to recover 3D geometry from its environment. Thus we will need to be able to associate camera geometry with particular head configurations. Generic calibration of such a system would require the ability to compute camera geometry from arbitrary stereo images. This paper describes a system which solves this problem using an established corner detector combined with a robust stereo matching algorithm and a variational solution for the camera geometry.

Keywords. Calibration, Corner detection, Stereo, Stereo matching.

Introduction.

We wish to develop a stereo eye/head camera rig which will support similar low level vision competences to primates, these are: foveation, vergence, saccades and tracking. This head configuration is currently under construction [Figure 1] and a simulation of the hardware has been used for the work presented here. We wish to be able to use this head with the TINA [1] vision system to recover stereo geometry and generate a 3D representation of the world. These low level vision competences will require stereo correspondence of well located image features. We show here that we can also use these correspondences to compute the relative camera geometry necessary to provide epipolar geometry for other stereo matching algorithms. Identification of such features can be achieved using an interest operator similar to that developed by Moravec [2]. The Plessey group [3] developed this idea further and the resulting edge and corner detector was used to obtain structure from motion [4]. Thus it seems natural to use the Moravec/Plessey corner detector as our starting point.

In order to use corners to generate the necessary camera translation and rotation parameters, we need to robustly match the sets of corners obtained. We cannot use the Plessey algorithm here as there may be substantial translations between views from two stereo cameras. Also, we cannot make much use of epipolar constraints as this would require the camera geometry which we are trying to obtain. This is not a difficult problem to solve provided we only require a subset of the total number of corners matched.

(Figure 1 about here)

Estimation of the camera geometry needs to be robust and unbiased, we would prefer to use the variational method proposed by Trivedi [5]. However, we would require in excess of 100 data points to provide sufficient calibration accuracy, which is large compared to the number found and matched in most scenes. For this reason we have applied standard statistical methods for data combination to the resulting calibration. We have extended this idea further to the calibration of a moving camera system which moves on a one dimensional trajectory in a space described by the calibration parameters.

Corner Detection and Matching.

The corner detector we use is that suggested by Harris and Stephens [2] which calculates an interest operator defined according to an auto-correlation of local patches of the image.

$$M_{uv} = \begin{bmatrix} (\partial I/\partial u)^2 * w & \partial I/\partial u \; \partial I/\partial v * w \\ \partial I/\partial u \; \partial I/\partial v * w & (\partial I/\partial v)^2 * w \end{bmatrix}$$

where u and v are image coordinates and $*w$ implies a convolution with a gaussian image mask. Any function of the eigenvalues α and β of the matrix M will have the property of rotation invariance. What is found is that the trace of the matrix $Tr(M) = \alpha + \beta$ is large where there is an edge in the image and the determinant $Det(M) = \alpha\beta$ is large where there is an edge or a corner. Thus edges are given when either α or β are large and corners can be identified where both are large. Corner strength is defined as

$$C_{uv} = Det(M) - kTr(M)^2$$

Corners are identified as local maxima in corner strength which are fitted to a two dimensional quadratic in order to improve positional accuracy which has been estimated as 0.3 pixels.

Given 5 or more correspondence points in the two images it is possible to compute the camera translation/rotation parameters for the left to right camera transformation. There are generally an order of magnitude more corners than this in even a relatively simple image. The corners are matched using a robust stereo matching algorithm which identifies reliable matches.

Image tokens can be matched in some cases using the following heuristics;

(a) restricted search strategies (eg epipolars in the case of stereo).

(b) local image properties (eg image correlation).

(c) uniqueness.

(d) disparity gradient (or smoothness) constraints.

For stereo matching potential matches are sought in a variable epipolar band, with a width determined by the accuracy of stereo calibration. As the corner detector finds local maxima in an auto-correlation measure it makes sense to compare possible matches between points on the basis of local image cross correlation. Lists of possible matches are generated, for corners in the left image to the right and right to left, and ordered in terms of the local image correlation measure;

$$M \;=\; \int_{-\infty}^{\infty} A^{-2}\; w_{uv}\; I_{uv} I'_{uv}\; du\; dv$$

with

$$A \;=\; \int_{-\infty}^{\infty} w_{uv} I_{uv}^2\; du\; dv \;\int_{-infty}^{\infty} w_{uv} I_{uv}^{'2}\; du\; dv$$

where w is a gaussian weighting function. This measure varies between 0 and 1 (close to 1 for good agreement), again the assumption has been made that there is little rotation about the viewing axis. This measure is invariant to the scale of the registered image intensity (assuming that no prior knowledge of the lighting conditions and individual camera aperture settings is available). Weak dependence on the absolute image intensity can be reintroduced using an asymmetry cut on the relative corner strength.

$$\frac{|C_1 \;-\; C_2|}{C_1 \;+\; C_2} \;>\; \eta$$

A value of 0.85 is generally chosen for η, this will allow a difference of 12 in relative corner strength or a factor of 1.8 in image intensity.

Only if the absolute value of the correlation measure is high ($M_{max} > \rho$) is the match accepted and added to the list of possible matches. *rho* can be set arbitrarily high to ensure that the underlying images are essentially identical and a value of 0.99 is generally used. We accept that this will inevitably result in some bias in matching ability for front-to-parallel surfaces. Candidate matches are only considered further if they involve the best correlation measure M_{max} found for that pair of points matched both ways between the left and right images. This algorithm implicitly enforces one to one matching and also eliminates incorrect matches resulting when a feature has only been detected in one image.

Due to the sparseness of corner data in many regions of an image it is difficult to impose a smoothness or disparity gradient constraint. However, it may be possible in future to constrain possible matching using the results from less sparse matching primatives such as edges.

On real images corner detection can be very noisy and setting a generic threshold for corner detection is problematic. Also high frequency textured regions generally give rise to many corners which, on the basis of the above heuristics, are unmatchable, as there are many similar candidate matches for each feature. Thus in real images it is difficult to automate the generation of a reliable set of correspondences, potentially preventing successful ego-motion calculation. What is required is a method of identifying those features which may be unreliably matched.

Unreliable features can be defined as those which have many candidate matches and consequently may be expected to be ambiguously matched. Ambiguous matches can be excluded by selecting matches where neither list of other candidate matches has an entry which is above a value of $M_{max} - \delta$. The required value of δ is defined by the expected variability of the cross-correlation value for correct matches and can be expected to be relatively constant for all images. δ can be defined so that only very unique matches are accepted as good, a value of 0.005 has been found generally to be sufficient. Such a reliability heuristic reduces the effects of feature detection thresholds on the matching of high frequency features.

If we have temporal match information, a more direct method of selecting reliable matches can be used. Temporal matches are sought using three dimensional positions of corner features combined with odometry information specifying the expected motion of COMODE. Match lists are generated between temporal pairs of images in exactly the same way as for the stereo matcher. The result is a set of possible matching lists for each point in each image to its stereo and temporal counterpart. A subset of correct matches is then selected by checking that the matching between all sets of stereo and temporal images is consistent.

After removal of non-unique matches there were generally between 20 and 100 matches fewer than 2 % of these were incorrect. This is enough to obtain an estimate of the camera rotation suitable for epi-polar matching, though generally too poor to obtain good geometrical accuracy. For this reason a method of combining the results from successive calibrations was required.

Camera Calibration.

It is possible to formulate the solution for an arbitrary camera rotation/translation (RT) from two sets of corresponding vector points in the images x_i and x_i' using a variational principle [5]. The small shifts δx_i and $deltax_i'$ needed to move these correspondences in each image, so that they satisfy an estimate of the transformation, can be approximated to linear order in an expansion about the current solution [Appendix 1] giving;

$$\delta x_i = - \frac{F_i S \nabla F_i^T}{\nabla F_i S \nabla F_i^T + \nabla' F_i S \nabla' F_i^T}$$

$$\delta x_i' = - \frac{F_i S \nabla' F_i^T}{\nabla F_i S \nabla F_i^T + \nabla' F_i S \nabla' F_i^T}$$

$$F_i = x_i^T (RT) x_i'$$

$$\nabla F_i = x_i^{'T} (RT)$$

$$\nabla' F_i^T = (RT) x^i$$

Where the rotation/translation constraint equation F^i uses the matrix formulation first suggested by Longuet-Higgins [6], which is a matrix alternative to writing the vector constraint equation;

$$F_i = x_i . t \times R x_i' \quad (= 0)$$

where t is the translation vector. This follows directly from the coordinate transformation equation which is valid for both points in the real world and image coordinates. The transformation matrix T and error matrix S are given by

$$T = \begin{bmatrix} 0 & e_6 & -e_5 \\ -e_6 & 0 & e_4 \\ e_5 & -e_4 & 0 \end{bmatrix} \qquad S = \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{bmatrix}$$

where $e_4 e_5 e_6$ are the direction cosines (xyz) of the translation between the optical centres of the cameras in the left camera frame.

Many of the constraints between elements of the rotation matrix can be imposed in a way that permits a unique reconstruction of the rotation matrix. This is done by parameterising the rotation matrix R in terms of Euler parameters (a quaternion representation [Appendix 2]). The error matrix allows proper account to be made of the asymetric nature of the x and y corner location accuracy introduced by the camera aspect ratio a with $\sigma_x = a\sigma_y$. The error in the z direction σ_z is set to zero as per the original implementation by Trivedi. This is a relatively simple model for the expected errors on the location of corners and a more principled one could be used if known. In our experience all corner locations are determined with the same accuracy within a factor of two.

An appropriately weighted sum of the minimum shifts required for each point to be independently consistent with the current transformation can be formed.

$$E = \sum_i E_i = \sum_i (\delta x_i S^{-1} \delta x_i + \delta x_i^{'T} S^{-1} \delta x_i')$$

note also that

$$E_i = \frac{F_i^2}{\nabla F_i S \nabla F_i^T + \nabla' F_i S \nabla' F_i^T} = F_i^2 / \sigma_i^2$$

The transformation matrix which is most consistent with the position of the observed correspondences can be obtained. This is done by minimising this sum with respect to the five free rotation and translation parameters e_1, e_2, e_3, e_5, e_6 while at the same time enforcing the following constraints.

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

Derivative information can be computed for each correspondence point [Appendix 3]. However, it was found that minimisation routines which could make use of this information were not very efficient or robust when used on this particular minimisation task. Minimisation is best done using a robust numerical minimisation routine as for example the simplex minimisation algorithm of Nelder and Mead (see for example [5]). This method lends itself to robust statistical methods should the fitted data be found to have a distribution which is non-normal.

The Trivedi algorithm has no adjustable parameters and yields errors in terms of image variables which can be used to judge the accuracy of the result. This information combined with knowledge of the corner detection accuracy allows rogue points to be iteratively removed from the fitting process.

The number of corners located in a pair of stereo images may not be sufficient to calibrate the camera geometry accurately. For this reason we need to be able to combine the estimates of the calibration variables e from several images. This can be done using the covariance matrix $[C]$ (as estimated as in Appendix 3) as follows

$$e_t \;=\; C_t(C_{t-1}^{-1}e_{t-1} + C^{-1}e)$$

and

$$C_t^{-1} \;=\; C_{t-1}^{-1} \;+\; C^{-1}$$

Flexibility can be obtained by limiting the size of C_t to that which provides the required calibration accuracy. This then allows the calibration to track any systematic changes in the camera system.

Calibrating a Moveable Head.

For a system which follows a one dimensional trajectory in a high dimensional calibration space we can approximate this trajectory locally using linear interpolation between data points. The calibration parameters must follow such a trajectory in the case of our simulated head when we restrict the control vergence rotation angles to be symetrical (Here symetrical is defined only in terms of control signals and places no restriction on the actual orientation of the cameras or their rotation axes). We can parameterize this curve using one free parameter ϕ the control vergence angle of both cameras obtained from accurate odometry. Using this parameter it is possible to interpolate calibration parameters across a range of camera angles

$$\hat{e} \;=\; (e'(\phi \;-\; \phi') \;+\; e''(\phi'' \;-\; \phi))/(\phi'' \;-\; \phi')$$

where e' and e'' the camera transformation parameters at ϕ' and ϕ''. These estimates can be concatenated into one calibration vector g which can be estimated from successive observations of e at known ϕ given the covariance C using a kalman filter.

$$g_t \;=\; g_{t-1} \;+\; C_{gt}(\nabla_g \hat{e})^T C^{-1}(e \;-\; \hat{e})$$

with

$$C_{gt}^{-1} \;=\; C_{gt-1}^{-1} \;+\; (\nabla_g \hat{e})^T C^{-1}(\nabla_g \hat{e})$$

The intrinsic parameters of the camera system, focal lengths and image centres are required as input parameters and could be assumed to be fixed for our camera rig. These can be determined independently using a combination of optical methods and alternative calibration algorithms [8]. We are currently working on several calibration systems which are to be unified within one statistical framework. The current implementation ignores radial distortions but these could easily be incorporated should it become necessary. The algorithm is independent of the magnitude of interocular separation and only the direction of translation between the cameras is determined. This is sufficient for obtaining epipolar geometry suitable for stereo matching but the interocular separation is needed for absolute depth measurements. This would imply that calibration for our moving head would be made simpler if the cameras were to rotate about their optical centre.

Results.

The Trivedi algorithm was first tested on simulated data. The head was simulated assuming that when fixating on objects it always adopted symetric vergence so that the transformation between cameras would be a function of the verge angle. The parameters used to monitor the resulting calibration accuracy of the method were the error on the obtained vergence angle and the sum of the squared minimum shifts required to make the simulated data consistent with the estimated transformation. The first of these gives a direct estimate of the limiting accuracy of depth measurement [Figure 2].

(Figure 2 about here)

Verge error can be estimated using the covariance matrix and improves as the results from several fits are combined. For a uniform distribution of n data points this was found to vary as approximately $(n-5)^{-1/2}$. The second parameter is directly related to the accuracy of the epipolar geometry which was generally found to less than 0.1 $pixels^2$ for $n > 20$. The Trivedi algorithm was found to deliver an unbiased estimate of the true transformation when the correct intrinsic camera parameters were supplied (see below). When calibrations were combined (as above) the resulting accuracy was consistent with that which would have been determined using the whole data set.

The correct calibration was also recovered following a shift in the simulated camera system [Figure 3] (corresponding to a knock on the real system). Recovery to a useable estimate was found to be an exponential function of the number of data points, as expected. The variation of transformation parameters e with camera vergence angle was found to be sufficiently linear to allow calibration over a 15 degree range using the method outlined above. The results indicate that only twice the number of correspondences required in the fixed camera method would be needed to obtain the same accuracy on reconstructed geometrical data.

(Figure 3 about here)

The performance of the algorithm was also investigated in the case where incorrect aspect ratios and image centres were provided. Errors on these parameters appear to provide the real limit on the accuracy of obtained stereo data, errors in the image centres of only 10 pixels can produce systematic depth errors of as much as 5 %. The effects of these errors are compounded when using the Tsai calibration algorithm with incorrect intrinsic parameters to determine the camera focal-lengths and interocular separation.

The algorithms were used to calibrate the camera geometry with several real scenes, using focal lengths and interocular separation obtained from the Tsai algorithm and corner matched correspondences. The estimate of the vergence measurement accuracy calculated from the covariance matrix can be seen in Table 1. The value of χ^2 was entirely dominated by the expected error in the y direction (by two orders of magnitude) corresponding to a reproduceability in position of 0.3 pixels. Points which were not consistent with the obtained camera geometry were excluded iteratively until the χ^2 was observed to be consistent with the corner location accuracy.

(Table 1 about here)

The overall accuracy of the rotation parameters was found to be in agreement with [5]. The new algorithm was found to be better than Tsai at determining the epipolar geometry on the same set of data points. There was agreement between both methods within the simulated errors for each process given the uncertainties on the intrinsic parameters.

Conclusion

It has been shown that a subset of robustly matched corner correspondences can be obtained from real images suitable for calibration purposes. A general purpose calibration algorithm has been demonstrated which enables optimal combination of calibration over a sequence of images. The method can be used to calibrate either fixed or moving head configurations (with symetric vergence). We believe that the method should be extendable to asymetric vergence configurations by interpolating on a plane defined between three calibration points.

Appendix 1.

To obtain the minimum shift *deltaxsubi* needed to make the observed data consistent with a constraint $Fsubi$ we can use the method of Lagrange. Here we minimise the expression;

$$E \;=\; \sum_i (\delta x_i^T S^{-1} \delta x_i \;+\; \delta x_i'^T S^{-1} \delta x_i') + \sum_i \lambda_i (F_i \;+\; \nabla F.\delta x_i \;+\; \nabla' F.\delta x_i')$$

This can be done analytically as follows;

$$\partial E/\partial \delta x_i \;=\; 2\delta x_i^T S^{-1} \;+\; \lambda_i \nabla F \;=\; 0$$

giving

$$\delta x_i \;=\; -\lambda_i S \nabla F^T/2$$

thus expanding the constraint equation about the point *xsubi*

$$2F_i \;+\; \nabla F \lambda_i S \nabla F^T \;+\; \nabla' F \lambda_i S \nabla' F^T \;=\; 0$$

giving

$$\lambda_i/2 \;=\; \frac{F_i}{\nabla F S \nabla F^T \;+\; \nabla' F S \nabla' F^T}$$

and hence δx_i and similarly for $\delta x_i'$.

Appendix 2.

The quaternion representation for the rotation of a coordinate frame can be written as follows

$$q \;=\; (e_0, e_1, e_2, e_3)$$

where

$$e_0 \;=\; cos(\theta/2) \qquad e_1 \;=\; r_0 sin(\theta/2)$$

and

$$e_2 \;=\; r_1 sin(\theta/2) \qquad e_3 \;=\; r_2 sin(\theta/2)$$

where r is a vector defining the axis of rotation and θ is the angle of rotation about that axis. The rotation matrix is then reparameterised as

$$R \;=\; \begin{bmatrix} e_0^2+e_1^2-e_2^2-e_3^2 & 2(e_1e_2+e_0e_3) & 2(e_1e_3-e_0e_2) \\ 2(e_1e_2-e_0e_3) & e_0^2-e_1^2+e_2^2-e_3^2 & 2(e_2e_3+e_0e_1) \\ 2(e_1e_3+e_0e_2) & 2(e_2e_3-e_0e_1) & e_0^2-e_1^2-e_2^2-e_3^2 \end{bmatrix}$$

Appendix 3.

The elements of the inverse covariance matrix are defined from the *chisup*2 variable by

$$\alpha_{nm} \;=\; \frac{1}{2}\partial^2\chi^2/\partial e_n \partial e_m$$

which can be constructed in our case from individual contributions from each data point i.

$$\alpha_{nm} \;=\; 1/(2\sigma_c^2) \sum_i \partial^2 E_i/\partial e_n \partial e_m$$

where σ_c is the estimated corner location accuracy and

$$E_i \;=\; \frac{F_i^2}{\nabla F_i S \nabla F_i^T + \nabla' F_i S \nabla' F_i^T} \;=\; F_i^2/\sigma_i^2$$

the first derivative is given by

$$\partial E_i/\partial e_n \;=\; \frac{2F_i \partial F_n/\partial e_n}{\sigma_i^2} \;-\; \frac{F_i^2 \partial \sigma_i^2/\partial e_n}{\sigma_i^4}$$

At the minimum the second term is found to be three orders of magnitude smaller than the first, allowing the second derivatives to be approximated to around the same accuracy using;

$$\partial^2 E^i/\partial e_n \partial e_m \;=\; \frac{2\partial F_n/\partial e_n \partial F_m/\partial e_m}{\sigma_i^2}$$

Acknowledgments.

We gratefully acknowledge the grant holders Dr. John E.W. Mayhew Dr. Paul Dean and Prof. John Friby and the support of ESRC/MRC/SERC for the funding of this project.

References.

[1] Porrill, J., S.B.Pollard, T.P.Pridmore, J.B.Bowen "TINA: The Sheffield AIVRU Vision System" Proc. 9th IJCAI. Vol.2 pp.1138-1144. 1987.

[2] Moravec, H.P. "Obstacle avoidance and navigation in the real world by a seeing robot rover" Ph.D Thesis, Stanford Univ., Sept., 1980.

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

[4] Charnley, D. and R.Blisset "Surface Reconstruction from Outdoor Image Sequences." Proceedings of the Fourth Alvey Vision Conference, pp.153-158 August 1988.

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

[6] Longuet-Higgins, H.C. "A Computer Algorithm for Reconstructing a Scene from Two Projections." Nature, Vol 293 pp.133-135 September 1981.

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

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

Figure Legends.

Figure 1. The Robot Head.

Figure 2. Percentage depth error on absolute depth measurement for specific verge angle accuracies. For relative depth errors simply multiply by two.

Figure 3. Variation of the optimal estimate of verge angle with time. 20 new data points were combined at each time step while the covariance matrix for the estimate was limited to a size which specified an error of 0.05 degress (generally requiring 400 data points). The figure shows how the estimate recovers after a shift in the camera system.

Table 1. Results from real scenes showing the improving calibration accuracy with increasing numbers of data points.