Pakistan Journal of Information and Technology 1(2): 75-79, 2002


# Stratified Approach to 3D Reconstruction

Muhammad Arif, Huang Xinhan and Wang Min
Department of Control Science and Engineering
Huazhong University of Science and Technology, Wuhan 430074, Hubei, P. R. China

**Abstract:** 3D reconstruction problem from images can be classified into three strata each of which is equivalent to the estimation of a specific geometry group. The simplest being projective, then affine, next metric and finally Euclidean structure. The advantage of stratification is that the images do not need to be from calibrated cameras in order to obtain reconstruction. In this paper results for both camera calibration and reconstruction are presented to verify that it is possible to obtain a 3D model directly from features in the images for man-made world.

**Key Words:** Computer Vision, Projective Reconstruction, Affine Reconstruction, Euclidean Reconstruction Absolute Conic, Plane at Infinity, Vanishing Points

## Introduction

Obtaining three-dimensional models of scenes from images is a fairly new research field in computer vision. It has a wide range of applications in such diverse fields as robotics, architecture, archeology, art history, and forensic sciences. Nowadays however more and more interest comes from the multimedia and computer graphics communities. The use of 3D models and environments on the Internet is becoming common practice. Virtual reality environments and three-dimensional product catalogues have become feasible and commercially attractive knowing the internal and external camera parameters (Trucco and Verri, 1998) associated with a sequence of two or more images of a scene, it is possible to reconstruct geometry by back-projecting matched points in the images to 3D points in the world coordinate system.

Traditionally, the camera calibration was obtained off-line and used images of special calibration objects (Tsai, 1989 and Tsai, 1987). A decade ago, Faugeras *et al.* (1992) and Maybank and Faugeras, 1992) introduced the idea of self-calibration, where the camera calibration can be obtained from the image sequences themselves, without requiring knowledge of the scene. The method developed by Hartley (1994) computes the parameters of interest in steps using non-iterative and iterative estimation techniques.

Many other methods have been proposed that constrain the motion or the scene to simplify the problem (Dorn, 1993 and Polleyfeys *et al.*, 1996). Recent general methods (Heyden and Astrom, 1996); Triggs, 1997 and Pollefeys *et al.*, 1998) utilize the relationship between certain projective geometry entities, the absolute quadric, the absolute conic, and their projections on the image plane. Pollefeys *et al.* (1998) allow changing focal length of the camera assuming fixed aspect ratio and zero skew.

Most of the autocalibration techniques require nonlinear solution methods. An alternative strategy is to first recover affine structure, then solve for camera intrinsic parameters using relations generated by affine calibration. The advantage of this two-stage approach is that the equations on the internal parameters are linear. This stratified (Luong and Vieville, 1994) approach is suitable for constructing man-made world with regular scene geometry such as parallelism and orthogonality of lines and known relative lengths of line segments to provide calibration constraints.

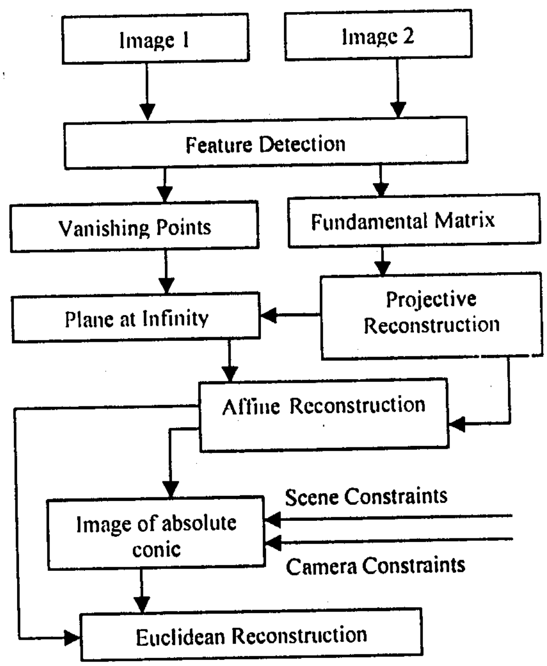



Fig. 1: Overview of the System

In this paper we investigate the stratified approach to reconstruct objects with a single moving uncalibrated camera. The overview of the implementation is given in Fig. 1.

**Hierarchy of Scene Reconstruction:** Suppose a set of image correspondences $x_i \leftrightarrow x_i^{\prime}$ are given. It is also assumed that correspondences come from a set of 3D unknown points $\overline{X}_i$ with some real existing cameras with unknown matrices $\overline{P}_1$ and $\overline{P}_2$. The reconstruction task is to find the camera matrices $\overline{P}_1$ and $\overline{P}_2$ as well as the 3D points $\overline{X}_i$ such that

$$x_i = \overline{P}_1\overline{X}_i \qquad x_i^{\prime} = \overline{P}_2\overline{X}_i \qquad \text{for all } i. \qquad (1)$$

However without further knowledge we can reconstruct the scene and cameras only up to projectivity. That is, we obtain $P_1$ and $P_2$ and $X_i$ from the measured image points $x_i$ which differ from camera matrices $\overline{P}_1$ and $\overline{P}_2$ and scene points $\overline{X}_i$ by an unknown projective transformation:

$$\sigma^k\overline{P}_k = P_kH^{-1} \quad for \quad k=1,\ 2$$
$$\lambda\overline{X}_i = HX_i \quad for \quad i=1,..,n \qquad (2)$$

Where $H$ is a $4x4$ matrix with rank 4 describing the 3D-to-3D projective transformation.

In order to upgrade projective structure $X_i$ to Euclidean structure $\overline{X}_i$ we have to recover $H$ with following decomposition:

$$H = H_eH_aH_p = \begin{bmatrix} sR & T \\ 0 & 1 \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} I & 0 \\ V^T & 1 \end{bmatrix} \qquad (3)$$

The first part $H_e$ depends on the choice of coordinate system and scale factor, the second part $H_a$ corresponds to affine upgrade, and is directly related to the knowledge of intrinsic parameters of the camera, and the third transformation is projective upgrade $H_p$ which transform points lying on a particular plane $[V^T \quad 1]X_i = 0$ to points at infinity.

This basic observation sets the stage for the so-called stratified approach to Euclidean reconstruction and self-calibration. Once the projective reconstruction is computed from image correspondences, further knowledge about scene structure or camera can be used to stratify the projective reconstruction to affine or Euclidean reconstruction. The term Euclidean Transformation is used in this paper to mean a similarity transformation, namely the composition of rotation, a translation and a uniform scaling.

**Projective Reconstruction:** It has been shown that a projective reconstruction requires only point (Berdsley *et al.*, 1994; Faugeras, 1992 and Hartley *et al.*, 1992) or line (Hartley, 1994) matches between two (three) views from unclaibrated cameras.

For two images, this step requires the determination of epipolar geometry (Faugeras, 1992). The epipolar geometry is entirely represented by fundamental matrix $F_{3x3}$ of rank 2. The fundamental matrix enables the determination of the projective matrices of two cameras, up to an unknown, but common, projective transformation (Hartley, 1992). Linear as well as nonlinear methods have been proposed to estimate fundamental matrix (Longuet-Higgins, 1981; Hartley, 1995; Zhang, 1998; Xu and Zhang, 1996; Torr and Murray, 1997) and Luong and Faugeras, 1996). To estimate fundamental matrix, Interest points in images can be found by a corner detector such as given in (Harris and Stephens, 1988).

Let the first camera coincides with the origin of the world coordinate system. The projective camera matrix for the first camera is then defined as

$$P_1 = [I_{3x3} \quad 0_3] \qquad (4)$$

The second projective camera matrix is chosen such that the epipolar geometry corresponds to the retrieved fundamental matrix (Zhang, 1998). Usually it is defined as

$$P_2 = [M \quad \beta e_2] \qquad (5)$$

Where $e_2$ is the epipole in the second image and the variable $\beta$ represents the global scale of the reconstruction. As the scale is unknown, it is arbitrarily chosen and set to 1. Matrix $M$ is defined as

$$M = -\frac{1}{\|e_2\|^2}[e_2]_x F$$

Where $[e_2]_x$ is the antisymmetric matrix of $e_2$.

The matrix $M$ is not necessarily unique, because if $M$ is solution, then $M + e_2v^T$ is also a solution for any vector $v$ (Zhang, 1998).

Given two camera matrices, 3D projective coordinates of scene points can be computed from image correspondences by triangulation (Hartley and Strum, 1995).

**Affine Reconstruction:** To upgrade the projective structure to affine structure we need to find the transformation matrix $H_p$. This step involves finding plane at infinity $\Pi_\infty = [V^T \quad 1]^T$ in the coordinate frame of projective reconstruction. In the true reconstruction, plane at infinity has coordinates $(0\ 0\ 0\ 1)^T$. $H_p$ thus maps plane at infinity to its canonical position. The transformation $H_p$ is now applied to all points and the two cameras to upgrade projective reconstruction to affine reconstruction. The plane at infinity in projective reconstruction cannot be identified unless some extra information is given. In practice this can be done (Horaud and Csurka, 1998) using (i) special camera motions (ii) exploiting special scene structure such as parallel lines, or (iii) using fixed entities under rigid motion.

The most obvious method is the knowledge that 3D lines are in reality parallel. The intersection of two parallel lines in space gives a point on plane at infinity and image of this point is vanishing point. Fig. 2 illustrates 3 vanishing points in an image of a building. Suppose that three sets of parallel lines can be identified in the scene. Each set intersects in a point on the plane at infinity. Provided each set has a different direction, the three points will be distinct. Since three points determine a plane, this information is sufficient to identify the plane at infinity.

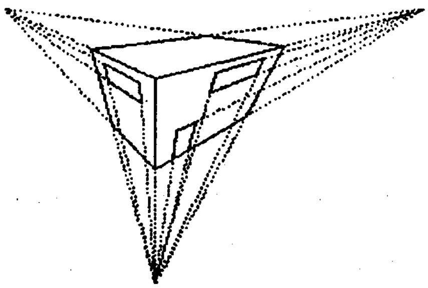


Fig. 2: Three Vanishing Points in an Image

Line segments in the images corresponding to parallel lines in three orthogonal directions can be computed using canny edge detector (Canny, 1986) and in the subsequent processing: edge linking; segmentation of the edgel chain at high curvature points; and finally, straight line fitting by orthogonal regression to the resulting chain segments. Vanishing points are then estimated using a maximum likelihood estimator described in (David, 2001).

Projecting the three vanishing points into space with the two projection camera matrices of (4) and (5), the points at the infinity are estimated. Triangulation (Hartley and Strum, 1995) is employed to find the best estimate of each point in space. From the three points at infinity, the plane at infinity is calculated as follows:

$$V_i^T \Pi_\infty = 0 \qquad for \quad i = 1,2,3. \tag{6}$$

Where $V_i$ are the points at infinity and $\Pi_\infty$ is the non-zero solution of the above linear homogeneous system.

The situation is not completely lost, in case above structure constraints are not available. Pollefeys *et al.* (1996) use modulus constraint. This approach requires a solution to a set of fourth order polynomial equations and multiple views (four) are necessary. Hartley (1993) uses chiral inequalities to give a range of possible values for. $\Pi_\infty$ where the optimum is achieved using linear programming approach.

**Euclidean Reconstruction:** The key to Euclidean reconstruction is the identification of absolute conic $\Omega_\infty$, which is a planar conic lying on plane at infinity (Hartley and Andrew, 2000). It has been shown that known image of dual absolute conic in the left ands right image is equivalent to knowing the internal parameters of the camera.

Suppose that in the affine reconstruction, the image of the absolute conic as seen by the camera with matrix $P = [M \mid m]$ is a conic $\omega$. It has been shown in (Hartley and Andrew, 2000) that the affine reconstruction may be transformed to a metric reconstruction by applying a 3D transformation of the form $\mathbf{H_a}$ given in (3). *A* in $H_a$ is obtained by Cholesky Factorization (Golub and Van Loan, 1989) from the equation

$$AA^T = \left(M^T \omega M\right)^{-1} \tag{7}$$

provided that that the right hand side of (7) is positive definite.

This approach of metric reconstruction relies on the identification of the image of absolute conic $\omega$ and it can be computed by combining constraints arising from scene, camera, and motion.

**Scene Constraints:** Commonly encountered constraints in man-made environments are orthogonality constraints between sets of lines in 3D. A pair of vanishing points $v_i$ and $v_j$ in the image plane corresponding to a set of orthogonal lines in 3D have to satisfy the following constraint:

$$v_i^T \omega v_j = 0 \tag{8}$$

**Constraints on Camera:** Computation of $\omega$ can be simplified when some of the intrinsic parameters are camera are known. The most commonly used assumptions are zero skew and known aspect ratio. For zero skew we have

$$\omega_{12} = \omega_{21} = 0 \tag{9}$$

And if pixels are square then

$$\omega_{11} = \omega_{22} \tag{10}$$

**Motion Constraints:** The image of absolute conic only depends on the calibration parameters of the camera not on the position and orientation of the camera. If both images are taken with the same camera then we have $\omega = \omega'$ i.e., the image of absolute conic is same in both images.

Between two views of a scene there is a planar homography induced by plane at infinity known as the infinite homography represented by a 3x3 homogeneous matrix $H_\infty$ (Liebowitz and Zisserman, 1999). It maps image of absolute conic between two views. This implies that

$$\omega' = H_\infty^{-T} \omega H_\infty^{-1} \tag{11}$$

Where $\omega'$ is the image of the absolute conic in the second view.

For $\omega' = \omega$, (11) gives a set of linear equations in the entries of $\omega$. This set of linear equations places four constraints on $\omega$, and it has 5 degrees of freedom, so it is not completely determined. However by combing constraints arising from scene and camera $\omega$ can be determined uniquely.

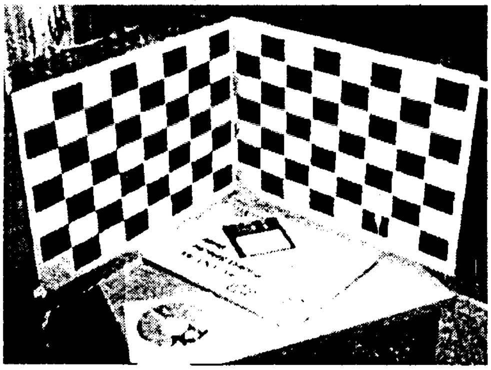
## Results and Discussion

Fig. 3 (a and b) shows two images of the calibration grid captured from two vantage points by Panasonic's V-CP450/WV-CP454 color CCTV camera. Fig. 3. (c) shows some of the reconstructed points in the scene. Fig. 4 shows 3D models of the calibration grid, up to a scale from three different vintage points after texture mapping by VRML (Virtual Reality Modeling Language). The reconstruction appears to be visually correct. Indeed the angle between the two planes computed from the reconstruction was 91.3°. The actual angle is 90°. To enhance the realism texture was mapped from the first image. However this has created a bias towards the selected image and imaging artifacts like sensor noise, unwanted specular reflections or shading of the first image have directly transformed onto the reconstructed model.

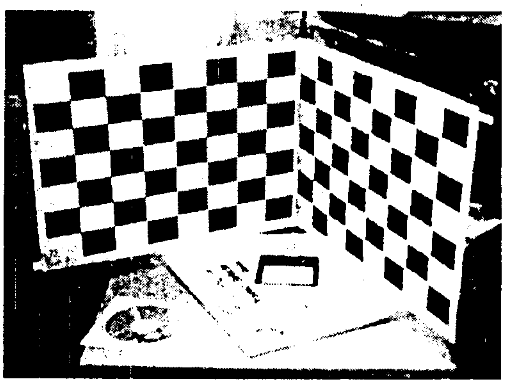
Table 1 compares results of camera parameters obtained by the method outlined above and to the results obtained by the method proposed by Zhang (Luong and Vieville, 1994).

Table 1: Comparison with Zhang's Method

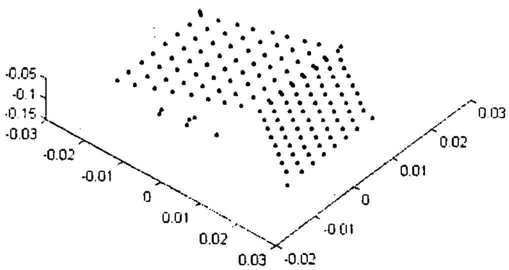
| | Above Method | Zhang's Method |
|---|---|---|
| $u_0$ | 197.0437 | 203.2790 |
| $v_0$ | 150.9275 | 157.7441 |
| $f_u$ | 683.2961 | 676.15192 |
| $f_v$ | 683.2961 | 674.75288 |
| s | 0 | 0 |



(a)



(b)




(c)

Fig. 3: Images Taken by a Moving Camera (a and b) and Reconstructed Points in The Scene (c)

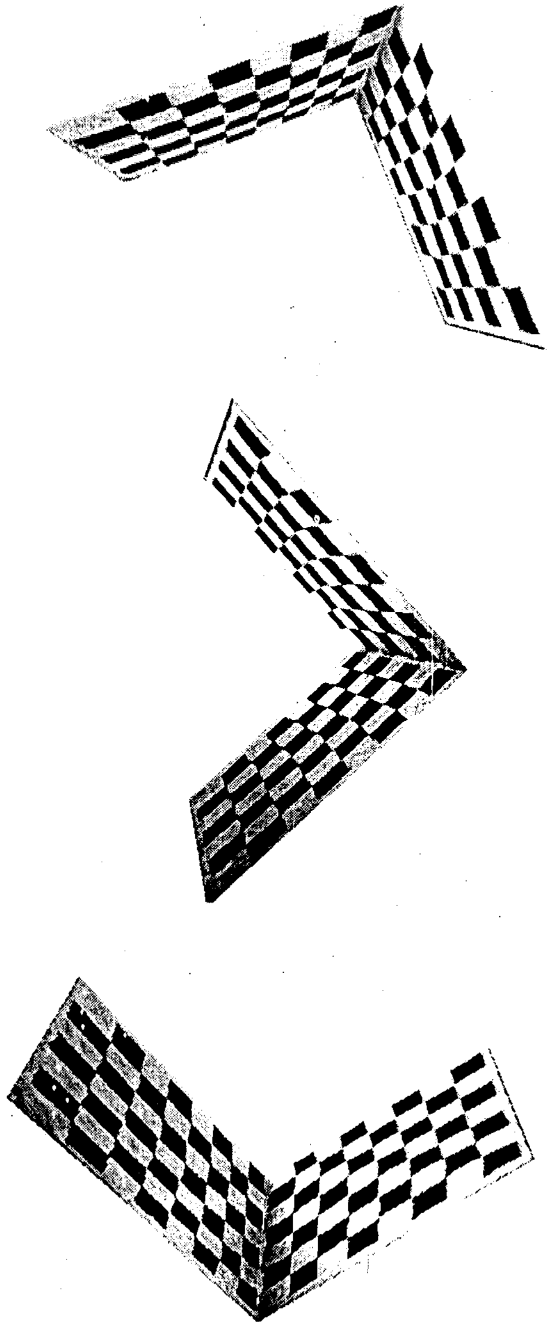


Fig. 4: Texture Mapped Reconstructed Calibration Grid Seen from Different Viewpoints

## Acknowledgement

We are grateful to CSC, P. R. China for financial support

## Conclusion

This paper represents camera calibration and 3D reconstruction from images based on stratified approach. This approach is suitable for man-made objects with rich geometry. We are currently working on developing algorithms to automate the reconstruction process.

It must also be noted that for certain configurations, however, calibration parameters cannot be determined uniquely. Sequences of camera motions for which such ambiguities arise are termed " critical motion sequences" and have been systematically classified by Strum (Horaud and Csurka, 1998) in case of constant internal parameters. Additional scene or motion constraints may help to resolve the ambiguity, but clearly the best way to avoid degeneracies is to use that are far from critical.

## References

A. Heyden and K. Astrom, 1996. Euclidean Reconstruction From Constant Intrinsic Parameters. In Proc. Int. Conf. on Pattern Recognition, Pages 339-343.

B. Triggs, 1997. Autocalibration and the Absolute Quadric. In Proc. IEEE Conf. Computer Vision and Pattern Recognition, Pages 609-614.

C. Harris and M. Stephens, 1998. A Combined Corner and Edge Detector. Proceedings Alvey Vision Conference, pages 189-192.

David Liebowitz, 2001. Camera Calibration and Reconstruction of Geometry from Images. PhD thesis. Dept. of Engineering Science Uni. of Oxford.

D. Liebowitz and A. Zisserman, 1999. Combining Scene And Auto-Calibration Constraints. In Proc. 7th International Conference On Cmputer Vision, Kerkyra, Greece, Pages 293-300.

E. Trucco, A. Verri, 1998. Introductory Techniques for 3-D Computer Vision. Prentice Hall, New Jersey.

G. Xu and Z. Zhang, 1996. The Epipolar Geometry in Stereo: Motion and Object Recognition. Kluwer Academic Publishers.

G. H. Golub and C. F. Van Loan, 1989. Matrix Computation. The Johns Hopkins Uni. Press, Baltimore.

H.C. Longuet-Higgins. 1981. A Computer Algorithm for Reconstructing a Scene from Two Projections. Nature, 239:133-135.

J. F. Canny, 1986. A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8:679-698.

L. Dorn, 1993. Dynamic Camera Self-Calibration From Controlled Motion Sequences. in Proc. IEEE Conf. Computer Vision and pattern Recognition, 501-506.

M. Polleyfeys, L. Van Gool, and Oosterlinck, 1996. The Modulus Constraint: A New Constraint for Self-Calibration. in Proc. Int. Conf. on Pattern Recognition, 349-353.

M. Pollefeys, R. Koch, And L. Van Gool, 1998. Self-Calibration And Metric Reconstruction in Spite Of Varying and Unknown Internal Camera Parameters. in Proc. Int. Conf. Computer Vision, 90-95.

M. Pollefeys, L. Van Gool and Andre Oosterlinck, 1996. The Modulus Constraint: A New Constraint for Self-Calibration. in Proc. I. C. P. R.

O. D. Faugeras, 1992. What Can be Seen in Three Dimensions with an Uncalibrated Stereo Rig? in Computer Vision - ECCV '92, LNCS-Series Springer-Verlag, 563-578.

O. D. Faugeras, Q. T. Luong, S. and Maybank, 1992. Camera Self-Calibration: Theory and Experiments. in Proc. European Conference on Computer Vision, LNCS 588, 321-334, Springer-Verlag,

Q. T. Luong and O. D. Faugeras, 1996. The Fundamental Matrix: Theory, Algorithms, and Stability Analysis. International J. of Computer Vision 17: 43-75, 1996.

Q. T. Luong and T. Vieville, 1994. Canonical Representations of the Geometries of Multiple Projective Views. Report: UCB/CSD 93-772, Uni. Of Berkeley, California.

P. Berdsley, A. Zisserman, and D. Murray, 1994. Navigation Using Affine Structure and Motion. in Proc. European Conference on Computer Vision, LNCS 800/801, 85-96. Springer-Verlag,

R. Hartley, 1992. Invariants of Points Seen in Multiple Images. Technical Report, G. E. CRD, Schenectady.

R. S. Torr and D. W. Murray. The Development and Comparison of Robust Methods For Estimating the Fundamental Matrix. International J. of Computer Vision 24:3 271-300, 1997.

R. Tsai, 1989. Synopsis Of Recent Progress on Camera Calibration for 3d Machine Vision. in Oussama Khatib, John J. Craig, and Tomas Lozano-Perez, Editors, The Robotics Review, 147-159. MIT Press.

R. Tsai, 1987. A Versatile Camera Calibration Technique for High-Accuracy 3d Machine Vision Metrology Using off-The-Shelf TV Cameras and Lenses. IEEE J.L of Robotics and Automation, 3: 323-344.

R. I. Hartely, 1994. An Algorithm For Self-Calibration From Several Views. in Proc. IEEE. Conf. Computer Vision and Pattern Recognition, 908-912.

R. Hartley, R. Gupta, and T. Chang, 1992. Stereo From Uncalibrated Cameras. in Proc. Computer Vision and Pattern Recongnition.

R. I. Hartley, 1995. In Defense of Eight Point Algorithm. In 5th International Conference on Computer Vision, 1064-1070, MIT, Cambridge, Massachusetts, IEEE Comp. Soc. Press.

R. Hartley, 1994. Projective Reconstruction From Line Correspondence. in Proc. Computer Vision and Pattern Recognition.

R. Hartley and P. Strum, 1995. Triangulation. In Proc. Conference Computer Analysis of Images and Patterns, Prague, Czech Republic.

R. Hartely, 1993. Cheirality invariants. in Proc. D. A. R. P. A. Image Understanding Workshop, 745-753.

R. Hartley, and Andrew Zisserman, 2000. Multiple View Geometry in Computer Vision. Cambridge Uni. Press.

R. Horaud and G. Csurka, 1998. Self-calibration and Euclidean Reconstruction using Motions of a Stereo rig. in Proc. Int. Conf. Computer Vision, 96-104.

S. Maybank and O. D. Faugeras, 1992. A Theory of Self-calibration of a Moving Camera. International J. of Computer Vision, 8:123-151.

Z. Zhang, 1998. Determining the Epipolar Geometry and Its Uncertainty: A Review. International J. of Computer Vision 27, 2: 161-195.

Z. zhang, 1998. A Flexible New Technique For Camera Calibration. Technical Report MSR-TR-98-71, Microsoft Research.