VIDERE

Videre: Journal of Computer Vision Research

Quarterly Journal

Fall 1997, Volume 1, Number 1

The MIT Press

Article 3

An Effective Technique for Calibrating a Binocular Stereo Through Projective Reconstruction Using Both a Calibration Object and the Environment

Zhengyou Zhang, Olivier Faugeras, Rachid Deriche

Videre: Journal of Computer Vision Research (ISSN 1089-2788) is a quarterly journal published electronically on the Internet by The MIT Press, Cambridge, Massachusetts, 02142. Subscriptions and address changes should be addressed to MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; phone: (617) 253-2889; fax: (617) 577-1545; e-mail: journals-orders@mit.edu. Subscription rates are: Individuals $30.00, Institutions $125.00. Canadians add additional 7% GST. Prices subject to change without notice.

An Effective Technique for Calibrating a Binocular Stereo Through Projective Reconstruction Using Both a Calibration Object and the Environment

Zhengyou Zhang,[1] Olivier Faugeras,[1] Rachid Deriche[1]

We present a novel technique for effectively calibrating a binocular stereo rig using the information from both scenes and classical calibration objects. The calibration provided by the classical methods is only valid for the space near the position of the calibration object. Our technique tries to make the best use of the rigidity of the geometry between two cameras. The idea is to first estimate precisely the epipolar geometry which is valid for a wide range in space from all available matches, extracted from both the environment and the calibration objects. This allows us to conduct an accurate projective reconstruction. Using the *a priori* knowledge of the calibration object, we are finally able to calibrate the stereo rig in a Euclidean space. The proposed technique has been tested with a number of real images, and significant improvement has been observed. A WWW demo with more examples and experimental data is available at http://www.inria.fr/robotvis/personnel/zzhang/CalibEnv/CalibEnv.html.

Keywords: camera calibration, stereo vision, on-line calibration

1. INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, F-06902 Sophia-Antipolis Cedex, France, Email: {zzhang,faugeras,der}@sophia.inria.fr

Introduction

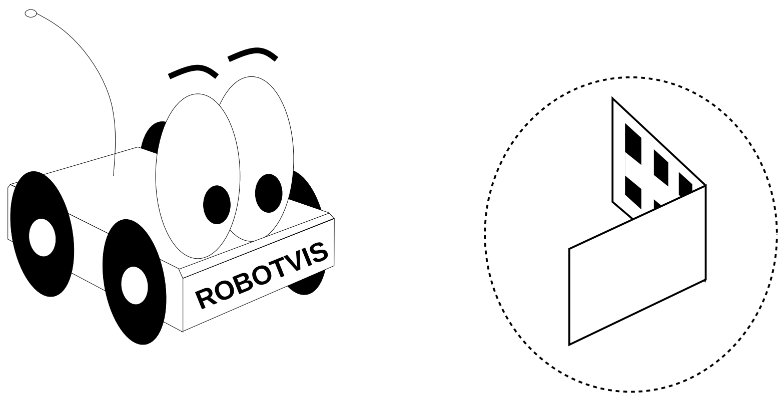
The technique described in this article has been developed as part of a perception system for applications of planetary intervention, although the technique itself can be used in other applications. A correlation-based binocular stereovision system has been chosen because of its robustness. Since a metric reconstruction of the environment is necessary for the path planning and navigation of the planetary rover, the stereovision system must be calibrated *strongly*, in the sense that a 3-D *Euclidean* reconstruction can be performed. The calibration should be performed on site, because if we do it before launching it will soon not be valid due to, for example, vibration and temperature variation during the flight.

Classical calibration techniques first estimate the perspective projection matrix for each camera, and then deduce the epipolar geometry from the projection matrices. The calibration of each camera is performed by observing a calibration object whose geometry is known with a very good precision in 3-D space, and can be done very efficiently. This is illustrated in Figure 1, where the calibration apparatus is a two-plane model, on which a checker pattern is painted. The coordinates of the corners on the checker pattern are known precisely with respect to a coordinate system attached to this apparatus. However, this approach is difficult to bring into operation for space applications, because human interaction is usually required to place the calibration apparatus such that it is visible from both cameras. Since, up to now, there is no robust and fully automatic self-calibration technique, the calibration object cannot yet be thrown away, and we mount it on the vehicle. Furthermore, the classical approach suffers from two problems:

- The fact that the set of image points in one camera corresponds to the set of image points in the other camera is not used in the classical approach, which degrades the precision of the estimation of the epipolar geometry.
- The calibration is only valid for the volume around the position of the calibration object. The performance of the calibration degrades when we go away from that position.

The second problem is more critical in our situation. Because the calibration object is mounted on the vehicle, its distance to the cameras is only about 1.5 meters, and the range of interest is up to 10 meters or more. Refer again to Figure 1. The circle around the calibration object

Figure 1. Illustration of classical camera calibration.



illustrates the valid work volume within which the calibration is valid. On the outside of it, the calibration will become worse and worse. The reason is that calibration is a process of fitting a camera model to a set of data. The model is only an approximation of the real camera, and furthermore the data is noisy. The fitted model can describe well the observed data, but not so for the space where no data is available. This is similar to the problem of extrapolation in numerical analysis. Extrapolation is much more hazardous than interpolation, because much less constraint is available for a point to be extrapolated. We can also consider a dual example: if the calibration object occupies only a small part of the image, the calibration will be poor for the other part of the image.

The calibration technique proposed in this paper overcomes the above two problems:

- The first stage of our calibration technique is to estimate the epipolar geometry between two cameras by minimizing the sum of the squared distances between points and their epipolar lines.
- Matches from the environment, which do not belong to the calibration object, can and should be used in the estimation of the epipolar geometry in order to increase its validity range.

Furthermore, our algorithm can take advantage of the rigidity of the geometry between two cameras by incorporating matches established at different instants, thus yielding an estimate of the epipolar geometry asymptotically valid for the whole range.

The idea underlying our approach is based on the now well-known fact that a projective structure can be reconstructed from two views given that the epipolar geometry is known [3, 6]. If the positions of a set of points are known in 3-D Euclidean space, then the projective distortion matrix (collineation) can be computed to bring the projective structure to a Euclidean one, which eventually allows us to recover the camera projection matrices. Our method thus consists of the following steps:

1. Estimate the epipolar geometry using all available information.
2. Compute the camera perspective projection matrices with respect to a projective reference frame.
3. Reconstruct projectively the points belonging to the calibration object in 3-D space.
4. Estimate the collineation matrix between the projective structure and the known Euclidean structure.

5. Compute the camera perspective projection matrices with respect to the Euclidean reference frame attached to the calibration object.

Determining the Epipolar Geometry of the Stereo

The most crucial step is determining the epipolar geometry, because everything else depends on the precision of the estimated epipolar geometry. The epipolar geometry can be described by a 3×3 matrix $\mathbf{F}$, which is known as the *fundamental matrix*. This matrix is defined up to a scale factor, and its determinant is zero. Thus, the fundamental matrix is only defined by seven parameters.

In order to estimate the fundamental matrix, we must establish correspondences between two images. In order to reduce the amount of information to be processed, we first extract a set of points of interest in each image, which correspond to points of high curvature. We then try to match these points between images using a classical correlation technique, followed by a relaxation procedure which uses contextual information (neighborhood) to disambiguate matching candidates. However, as the two matching techniques use heuristics and the only geometric constraint, i.e., the epipolar constraint, is not available, there exist inevitably a number of false matches. The false matches will severely affect the precision of the fundamental matrix, or even make the estimation useless, if we directly use the whole set of matches. For this reason, we apply a robust technique called *least-median-squares* to detect these false matches, in order to estimate accurately the fundamental matrix. The idea underlying this technique is to find a fundamental matrix which is consistent with a majority of matches by searching in the parameterization space. This is a very robust technique, and can detect false matches as numerous as 50% of whole data. This stage has been fully described in [11]. An automatic and robust algorithm has been developed to estimate the unknown epipolar geometry, in terms of the fundamental matrix, between two images.

All matched points are then used together to effectively estimate the fundamental matrix $\mathbf{F}$:

$$\min \sum_i w_i \left(d^2(\mathbf{m}_i, \mathbf{F}\mathbf{m}'_i) + d^2(\mathbf{m}'_i, \mathbf{F}^T \mathbf{m}_i) \right)$$

where w_i stands for the uncertainty measure associated to each measured point match $(\mathbf{m}_i, \mathbf{m}'_i)$, and $d(\mathbf{m}, \mathbf{l})$ denotes the Euclidean distance from point $\mathbf{m}$ to the line $\mathbf{l}$. Linear and nonlinear criteria, as well as different parameterizations for $\mathbf{F}$, have been considered in [8, 11] to accurately determine $\mathbf{F}$.

Computing the Camera Projection Matrices

Once we have computed the epipolar geometry (in terms of the fundamental matrix) between two images, we are able to compute a projective reconstruction of the scene. In this section, we show how to compute the camera projection matrices with respect to a projective basis.

Factorization Method

Let $\mathbf{F}$ be the fundamental matrix for the two cameras. There are an infinite number of projective bases which all satisfy the epipolar geometry. One possibility is to factor $\mathbf{F}$ as a product of a skew matrix $[\mathbf{e}']_\times$ ($\mathbf{e}'$ is in fact the epipole in the second image) and a matrix $\mathbf{M}$, i.e., $\mathbf{F} = [\mathbf{e}']_\times \mathbf{M}$.

Here $[\mathbf{e}]_\times$ is the skew matrix such that $[\mathbf{e}]_\times \mathbf{x} = \mathbf{e} \times \mathbf{x}$ (and $\times$ denotes the cross product). A canonical representation can then be used: $\mathbb{P} = [\mathbf{I}\ \ \mathbf{0}]$ and $\mathbb{P}' = [\mathbf{M}\ \ \mathbf{e}']$.

The factorization of $\mathbf{F}$ into $[\mathbf{e}']_\times \mathbf{M}$ is generally not unique, because if $\mathbf{M}$ is a solution, then $\mathbf{M} + \mathbf{e}'\mathbf{v}^T$ is also a solution for any vector $\mathbf{v}$. One way to do the factorization is as follows [7]. Since $\mathbf{F}^T\mathbf{e}' = \mathbf{0}$, the epipole in the second image is given by the eigenvector of matrix $\mathbf{F}\mathbf{F}^T$ that has the smallest eigenvalue. Using the relation

$$\|\mathbf{v}\|^2 \mathbf{I}_3 = \mathbf{v}\mathbf{v}^T - [\mathbf{v}]_\times^2 \quad \text{for all vector } \mathbf{v} ,$$

we have

$$\mathbf{F} = \frac{1}{\|\mathbf{e}'\|^2}(\mathbf{e}'\mathbf{e}'^T - [\mathbf{e}']_\times^2)\mathbf{F} = \frac{1}{\|\mathbf{e}'\|^2} \underbrace{\mathbf{e}'\mathbf{e}'^T\mathbf{F}}_{\mathbf{0}} + [\mathbf{e}']_\times \underbrace{\left(-\frac{[\mathbf{e}']_\times}{\|\mathbf{e}'\|^2}\mathbf{F}\right)}_{\mathbf{M}}$$

The first term on the right hand is equal to 0 because $\mathbf{F}^T\mathbf{e}' = \mathbf{0}$. We can thus define the $\mathbf{M}$ matrix as

$$\mathbf{M} = -\frac{[\mathbf{e}']_\times}{\|\mathbf{e}'\|^2}\mathbf{F}$$

This decomposition is used in [1].

Choosing a Projective Basis

Another possibility is to choose as a projective basis five pairs of points, any four points of which not being coplanar, between the two cameras. We can, of course, choose five corresponding points we have identified. However, the precision of the final projective reconstruction will depend heavily upon the precision of the pairs of points. In order to overcome this problem, we have chosen the following solution. We first choose five arbitrary points in the first image, noted by $\mathbf{m}_i$ ($i = 1, \ldots, 5$). For each point $\mathbf{m}_i$, its corresponding epipolar line is given by $\mathbf{l}'_i = \mathbf{F}\mathbf{m}_i$. We can now choose an arbitrary point on $\mathbf{l}'_i$ as $\mathbf{m}'_i$, the corresponding point of $\mathbf{m}_i$. Finally, we should verify that none of the four points are coplanar, which can be easily done using the fundamental matrix ([3], credited to Roger Mohr). The advantage of this method is that the five pairs of points satisfy exactly the epipolar constraint.

Once we have five pairs of points $(\mathbf{m}_i, \mathbf{m}'_i)$, ($i = 1, \ldots, 5$), we can compute the camera projection matrices as described in [3].

Which Method?

For the moment, the second method is used. Theoretically, the two methods should produce identical results. It is, however, possible that the results differ because of numerical stability due to the configuration of the projective basis.

Projective Reconstruction

Now that the camera projection matrices of the stereo with respect to a projective basis are available, we can reconstruct 3-D structures *with respect to that projective basis* from point matches. This reconstruction can be done for all matches, but for calibration purposes, we only need to reconstruct those points corresponding to the calibration objects.

Given a pair of points in correspondence: $\mathbf{m} = [u, v]^T$ and $\mathbf{m}' = [u', v']^T$. Let $\tilde{\mathbf{x}} = [x, y, z, t]^T$ be the corresponding 3-D point in space with respect to the projective basis chosen before. Following the pinhole model, we have

$$s\,[u, v, 1]^T = \mathbb{P}\,[x, y, z, t]^T \tag{1}$$

$$s'\left[u', v', 1\right] = \mathbb{P}'\,[x, y, z, t]^T \tag{2}$$

where s and s' are two arbitrary scalars. Denote $\mathbf{p}_i$ be the vector corresponding to the i-th row of $\mathbb{P}$, and $\mathbf{p}'_i$ be the vector corresponding to the i-th row of $\mathbb{P}'$. The two scalars can then be computed as

$$s = \mathbf{p}_3^T\tilde{\mathbf{x}} \qquad s' = \mathbf{p}_3'^{\,T}\tilde{\mathbf{x}}$$

Eliminating s and s' from Equation (1) and Equation (2) yields the following equation:

$$\mathbf{A}\tilde{\mathbf{x}} = \mathbf{0} \tag{3}$$

where $\mathbf{A}$ is a 4×4 matrix given by

$$\mathbf{A} = \left[\mathbf{p}_1 - u\mathbf{p}_3, \mathbf{p}_2 - v\mathbf{p}_3, \mathbf{p}'_1 - u'\mathbf{p}'_3, \mathbf{p}'_2 - v'\mathbf{p}'_3\right]^T$$

As the projective coordinates $\tilde{\mathbf{x}}$ are defined up to a scale factor, we can impose $\|\tilde{\mathbf{x}}\| = 1$, then the solution to Equation (3) is well known to be the eigenvector of matrix $\mathbf{A}^T\mathbf{A}$ that has the smallest eigenvalue.

If we assume that no point is at infinity, then we can impose $t = 1$, and the projective reconstruction can be done exactly in the same way as for the Euclidean reconstruction.

The previous method has the advantage of providing a closed-form solution, but it has the disadvantage that the criterion that is minimized does not have a good physical interpretation. Instead, we can carry out the minimization in the image plane, that is, we can minimize the following criterion:

$$(u - \frac{\mathbf{p}_1^T\tilde{\mathbf{x}}}{\mathbf{p}_3^T\tilde{\mathbf{x}}})^2 + (v - \frac{\mathbf{p}_2^T\tilde{\mathbf{x}}}{\mathbf{p}_3^T\tilde{\mathbf{x}}})^2 + (u' - \frac{\mathbf{p}_1'^{\,T}\tilde{\mathbf{x}}}{\mathbf{p}_3'^{\,T}\tilde{\mathbf{x}}})^2 + (v' - \frac{\mathbf{p}_2'^{\,T}\tilde{\mathbf{x}}}{\mathbf{p}_3'^{\,T}\tilde{\mathbf{x}}})^2$$

Hartley and Sturm [5] show that it can be formulated as solving a sixth order polynomial. We use any standard iterative minimization technique, where the initial estimate of $\tilde{\mathbf{x}}$ can be obtained by using the above closed-form technique.

We have implemented both methods. Our experiences show that the improvement through the nonlinear minimization is rather small, mainly because the points on the calibration pattern are very precisely localized. Other projective reconstruction techniques can be found in [5, 9].

Estimating the Projective Distortion Matrix

Now we have reconstructed a set of 3-D points $\tilde{\mathbf{x}}_i = [x_i, y_i, z_i, t_i]^T$ $(i = 1, \ldots, n)$ with respect to a projective basis. For each of these points, we know precisely its 3-D coordinates in a Euclidean reference frame, because it belongs to the calibration object. Let the set of 3-D Euclidean points be $\tilde{\mathbf{X}}_i = [X_i, Y_i, Z_i, 1]^T$ $(i = 1, \ldots, n)$. The projective points $\tilde{\mathbf{x}}_i$ are related to the Euclidean points $\tilde{\mathbf{X}}_i$ by a collineation (a 4×4 matrix), which we would like to call the *projective distortion matrix* $\mathbb{D}$, i.e.,

$$\lambda_i\,[x_i, y_i, z_i, t_i]^T = \mathbb{D}\,[X_i, Y_i, Z_i, 1]^T \tag{4}$$

or

$$\lambda_i \tilde{\mathbf{x}}_i = \mathbb{D}\tilde{\mathbf{X}}_i \tag{5}$$

where

$$\mathbb{D} = \begin{bmatrix} D_{11} & D_{12} & D_{13} & D_{14} \\ D_{21} & D_{22} & D_{23} & D_{24} \\ D_{31} & D_{32} & D_{33} & D_{34} \\ D_{41} & D_{42} & D_{43} & D_{44} \end{bmatrix}$$

and λ_i is an arbitrary scalar, because $\mathbb{D}$ is only defined up to a scale factor.

From Equation (5), the two vectors $\tilde{\mathbf{x}}_i$ and $\mathbb{D}\tilde{\mathbf{X}}_i$ are related by a scalar. Let $\mathbf{v} = \mathbb{D}\tilde{\mathbf{X}}_i = [v_1, v_2, v_3, v_4]^T$, then we have the following three independent equations:

$$\begin{aligned} y_i v_1 - x_i v_2 &= 0 \\ z_i v_1 - x_i v_3 &= 0 \\ t_i v_1 - x_i v_4 &= 0 \end{aligned} \tag{6}$$

Let $\mathbf{x} = [D_{11}, D_{12}, \ldots, D_{44}]^T$ be the vector of the 16 parameters of the distortion matrix to be computed. It is easy to show that Equation (6) is equivalent to the following equation:

$$\mathbf{B}_i \mathbf{x} = 0 \tag{7}$$

where

$$\mathbf{B}_i = \begin{bmatrix} y_i \tilde{\mathbf{X}}_i^T & -x_i \tilde{\mathbf{X}}_i^T & \mathbf{0} & \mathbf{0} \\ z_i \tilde{\mathbf{X}}_i^T & \mathbf{0} & -x_i \tilde{\mathbf{X}}_i^T & \mathbf{0} \\ t_i \tilde{\mathbf{X}}_i^T & \mathbf{0} & \mathbf{0} & -x_i \tilde{\mathbf{X}}_i^T \end{bmatrix} \tag{8}$$

Given n correspondences $(\tilde{\mathbf{X}}_i, \tilde{\mathbf{x}}_i)$, we have n equations of the type of Equation (7). The problem is then to estimate $\mathbf{x}$ by minimizing the following error function:

$$\mathcal{F} = \sum_{i=1}^{n} (\mathbf{B}_i \mathbf{x})^2 = \mathbf{x}^T \left(\sum_{i=1}^{n} \mathbf{B}_i^T \mathbf{B}_i \right) \mathbf{x} \tag{9}$$

Let $\mathbf{B} = \sum_{i=1}^{n} \mathbf{B}_i^T \mathbf{B}_i$, which is a symmetric matrix. As $\mathbb{D}$ is only defined up to a scale factor, we can normalize $\mathbf{x}$ with $\|\mathbf{x}\| = 1$. It is well known that the solution to Equation (9) is the eigenvector of $\mathbf{B}$ corresponding to the smallest eigenvalue of $\mathbf{B}$.

Many other methods exist for estimating the projective distortion matrix [10]. The above method is simple, but it is not clear what is being minimized. We are currently evaluating different methods. One meaningful cost function is the sum of squared distances between the reconstructed points and the known 3-D Euclidean points:

$$\mathcal{F} = \sum_{i=1}^{n} \|\mathbf{X}_i - \widehat{\mathbf{X}}_i\|^2$$

where $\widehat{\mathbf{X}}_i$ is the vector consisting of the first 3 elements of $\mathbb{D}^{-1}\tilde{\mathbf{x}}_i$ divided by its fourth element.

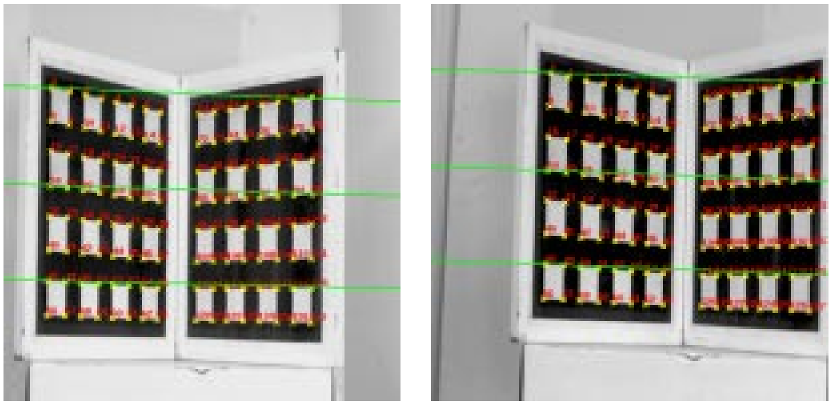


Figure 2. A two-plane model used in the classical calibration technique.

Table 1. Comparison of average distance with different fundamental matrices. The underlined numbers compare the calibration result with the proposed technique and that with the classical one for the "head" image.

	Distance (pixels)		
Fundamental Matrix	**Two-Plane**	**Head**	**All**
two-plane	0.10	<u>1.90</u>	1.19
head	5.22	0.90	3.63
all	0.18	<u>0.97</u>	0.72

Recovering the Euclidean Camera Projection Matrices

Let the Euclidean camera projection matrices be $\mathbb{M}$ and $\mathbb{M}'$. Following the pinhole model, we have

$$
\begin{aligned}
s\,[u, v, 1]^T &= \mathbb{M}\,[X, Y, Z, 1]^T \\
s'\left[u', v', 1\right]^T &= \mathbb{M}'\,[X, Y, Z, 1]^T
\end{aligned}
$$

Compare the above equations with Equation (1) and Equation (2), and we have at once the Euclidean camera projection matrices given by

$$
\begin{aligned}
\mathbb{M} &= \mathbb{P}\mathbb{D} \\
\mathbb{M}' &= \mathbb{P}'\mathbb{D}
\end{aligned}
$$

Now the calibration of the binocular stereo is achieved. The calibration matrices are estimated with respect to the world coordinate system associated to the calibration object, but taking into account the information from the environment.

Experimental Results

Here is an example. Figure 2 shows a stereo pair of a two-plane model which is used by the classical calibration technique [4]. The two images are taken simultaneously. The corners (indicated by crosses) are detected with very high precision by intersecting the lines of the checkered pattern. Since the coordinates of their corresponding points are known in space, we can compute the perspective projection matrix for each camera, from which we can deduce the epipolar geometry (denoted by $\mathbf{F}_{\text{two}-\text{plane}}$). The average distance between the detected corners and their epipolar lines is 0.1 pixels (see Table 1 for a comparison of the results). This suggests that the calibration result is very good, as far as the two-plane model is concerned, at this specific position.

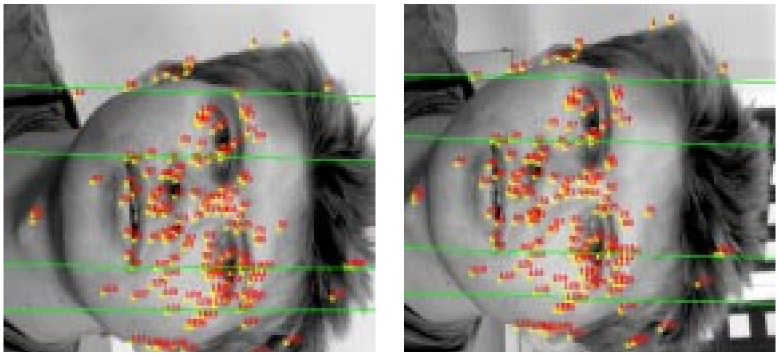


Figure 3. The stereo pair of Hervé's head, together with several epipolar lines given by the classical calibration technique.

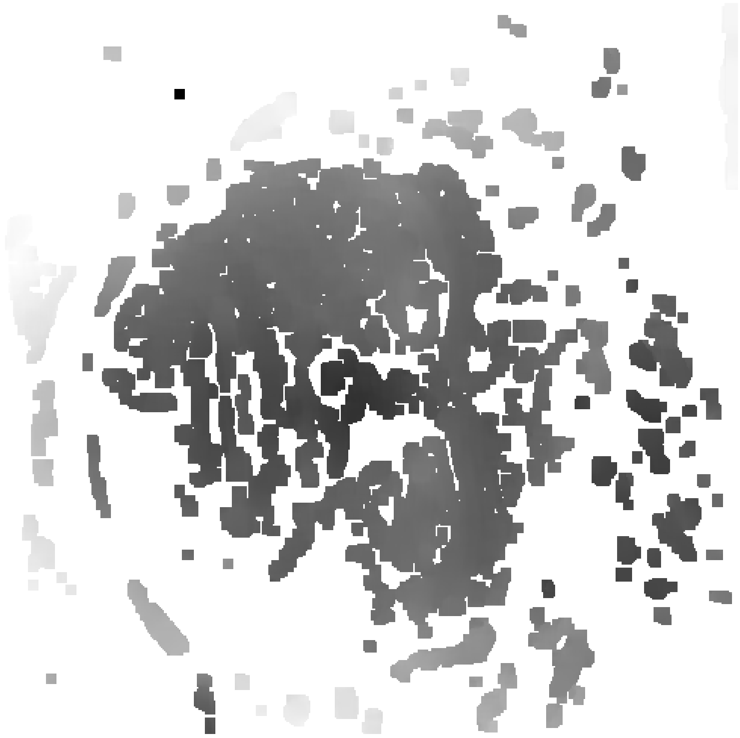


Figure 4. Disparity map obtained with the correlation technique using $\mathbf{F}_{\text{two-plane}}$. The darker a point is, the closer it is to the cameras. Completely white points are those not correlated.

However, the calibration is not as valid for other parts of space. Figure 3 shows a pair of images taken by the same stereo rig. The matches have been established automatically by the program `image-matching` as described in [11] without using $\mathbf{F}_{\text{two-plane}}$. The corresponding epipolar geometry is denoted by $\mathbf{F}_{\text{head}}$, and the average distance between points and epipolar lines is 0.9 pixels. If we consider $\mathbf{F}_{\text{two-plane}}$ (obtained with the classical calibration technique), the average distance becomes 1.9 pixels, which is quite high. In Figure 3, the epipolar lines corresponding to match 11, 30, 39, and 131 are displayed, respectively. An obvious deviation can be observed. As a matter of fact, our correlation-based stereo system does not work correctly if $\mathbf{F}_{\text{two-plane}}$ is used. The result is shown in Figure 4, where we see that only a small part of the face has been correlated. On the other hand, the epipolar geometry $\mathbf{F}_{\text{head}}$ is not valid for the points on the two-plane model (the average distance is 5.22 pixels). The epipolar lines corresponding to match 0, 24, and 48 are shown in Figure 2.

Now, we combine both sets of matches found on the two-plane model and on the head, and compute another fundamental matrix which is

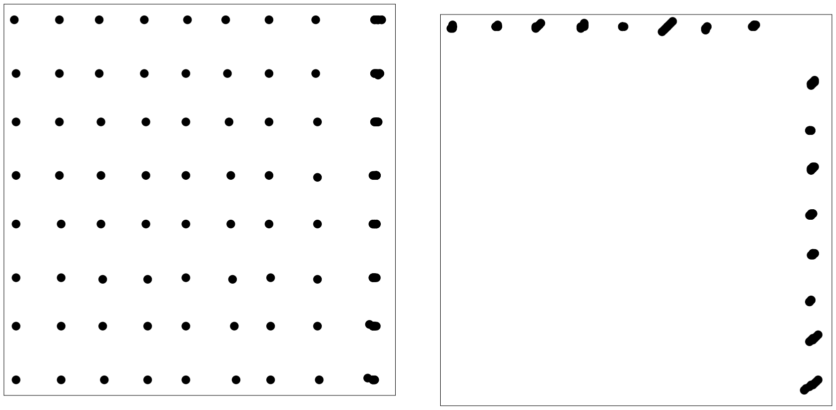


Figure 5. 3-D reconstruction of the two-plane model: orthographic projections on one of the planes of the two-plane model and on a plane orthogonal to both planes.

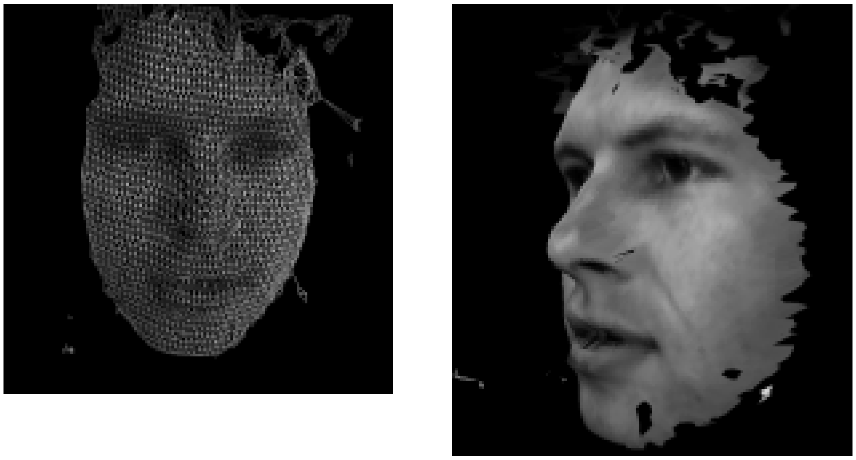


Figure 6. 3-D reconstruction of Hervé's head: two projections.

denoted by $\mathbf{F}_{\text{all}}$. The average distance for the whole set of matches is 0.72 pixels. If we only consider the set of matches on the head, the average distance is 0.97 pixels, which should be compared with 1.9 pixels when the epipolar geometry estimated by the classical method is used (see Table 1 for a complete comparison of the results).

From $\mathbf{F}_{\text{all}}$ and the knowledge of the two-plane model, we calibrate the stereo rig using the technique described in this paper. The calibration is now valid both for the two-plane model and for the head. The 3-D reconstruction of the two-plane model is shown in Figure 5. This calibration also makes the correlation-based stereo system work decently. The technique described in [2] is then used to reconstruct the head in 3-D space, which is shown in Figure 6. This should be compared to that shown in Figure 4.

Conclusion

We have described a novel method for calibrating a binocular stereo by using the information from the scene (although the 3-D positions are not known) as well as that from a classical calibration object. The method is divided into five steps:

1. Estimate the epipolar geometry using all available information.
2. Compute the camera perspective projection matrices with respect to a projective reference frame.
3. Reconstruct projectively the points belonging to the calibration object in 3-D space.

Videre rough pages 1997/9/10 14:26

4. Estimate the collineation matrix between the projective structure and the known Euclidean structure.
5. Compute the camera perspective projection matrices with respect to the Euclidean reference frame attached to the calibration object.

The idea is to first estimate precisely the epipolar geometry which is valid for a wide range in space from all available matches. This allows us to conduct a projective reconstruction. Using the *a priori* knowledge of the calibration object, we are eventually able to calibrate the stereo rig in a Euclidean space. Experiments show that we improve the validity of the calibration parameters by integrating additional visual information other than that from the calibration object.

In order to quantify how much accuracy in the epipolar geometry is lost due to the change in distance, and how much improvement in the accuracy of 3-D reconstruction is gained with the proposed technique, one should, as the reviewer suggests, perform additional experiments by taking a sequence of images of the same calibration object, but at different distances. This is a part of our future work.

Acknowledgments

The contribution of Fréderic Devernay is gratefully acknowledged. The comments from the reviewer have allowed the authors to improve the paper. This work was partly carried out under the contract Eureka IARES.

References

[1] Paul Beardsley, Andrew Zisserman, & David Murray. Navigation using affine structure from motion. In J-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 2 of *Lecture Notes in Computer Science*, pages 85–96, Stockholm, Sweden, May 1994. Springer-Verlag.

[2] Frédéric Devernay & Olivier Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 208–213, Seattle, WA, June 1994. IEEE.

[3] Olivier Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, volume 588 of *Lecture Notes in Computer Science*, pages 563–578, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.

[4] Olivier Faugeras & Giorgio Toscani. The calibration problem for stereo. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 15–20, Miami Beach, FL, June 1986. IEEE.

[5] R. I. Hartley & P. Sturm. Triangulation. In *Proceedings of the ARPA Image Understanding Workshop*, pages 957–966. Defense Advanced Research Projects Agency, Morgan Kaufmann Publishers, Inc., 1994.

[6] Richard Hartley, Rajiv Gupta, & Tom Chang. Stereo from uncalibrated cameras. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 761–764, Urbana Champaign, IL, June 1992. IEEE.

[7] Q.-T. Luong & T. Viéville. Canonic representations for the geometries of multiple projective views. In J-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 1 of *Lecture*

Notes in Computer Science, pages 589–599, Stockholm, Sweden, May 1994. Springer-Verlag.

[8] Quang-Tuan Luong. *Matrice Fondamentale et Calibration Visuelle sur l'Environnement-Vers une plus grande autonomie des systèmes robotiques*. PhD thesis, Université de Paris-Sud, Centre d'Orsay, December 1992.

[9] C. Rothwell, G. Csurka, & O. Faugeras. A comparison of projective reconstruction methods for pairs of views. In *Proceedings of the 5th International Conference on Computer Vision*, pages 932–937, Boston, MA, June 1995. IEEE Computer Society Press.

[10] Z. Zhang. Computing projective distortion matrix (collineation). Technical report, INRIA Sophia, Robotvis, November 1993.

[11] Z. Zhang, R. Deriche, O. Faugeras, & Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78:87–119, October 1995.