

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/2595084>

Self-calibration of Stereo Cameras with a Probabilistic Camera Model including Lens Distortion

Article · February 1970

Source: CiteSeer

CITATIONS

0

READS

46

2 authors:



[André Redert](#)

Rodotti

46 PUBLICATIONS 827 CITATIONS

SEE PROFILE



[Emile Hendriks](#)

Delft University of Technology

146 PUBLICATIONS 2,137 CITATIONS

SEE PROFILE

Self-calibration of Stereo Cameras with a Probabilistic Camera Model including Lens Distortion

André Redert and Emile Hendriks

Information and Communication Theory Group, Department of Electrical Engineering
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
email: {P.A.Redert,E.A.Hendriks}@its.tudelft.nl, <http://www-ict.its.tudelft.nl>
phone: +31 15 278 6269, fax: +31 15 278 1843

Keywords: stereo camera model, lens distortion, correspondence estimation, epipolar constraint, self-calibration

Abstract

In this paper we present a new method for the self-calibration of a stereo camera pair including lens distortion. A single stereo image pair is used without calibration object or prior knowledge of the scene.

The method is based on the estimation of a stereo image correspondence field followed by extraction of the calibration parameters. The correspondence field is a Markov Random Field motion estimator with an epipolar constraint that does not need calibration in advance. A new probabilistic model for local lens distortion serves as basis for the constraint. A probabilistic stereo camera model is fitted to the correspondence field using a MAP criterion. The optimum is found by simulated annealing.

The results with the correspondence estimator show that it is capable of dealing with very large translations, rotations and scale differences, and thus, a large variety in stereo camera parameters such as position, orientation and focal distance differences. For the parameter extraction phase, in methods without lens distortion, only 7 parameters can be measured. Our results indicate that if lens distortion is included, at least 18 out of 19 parameters can be measured.

1. Introduction

In the area of 3-D measurements with stereo imaging, accurate calibration of the camera pair is crucial [1]. The reconstruction of scene points is done by estimation of corresponding pixels in left and right image and subsequent triangulation. Accurate triangulation can only be performed if all parameters of the cameras are calibrated (known). These are the external parameters which represent the position and orientation of the cameras, and the internal parameters such as focal length, pixel ratio, lens distortion and CCD mispositioning. As an important byproduct calibration reduces the correspondence problem from 2-D motion estimation to the more efficient and reliable 1-D disparity estimation [4].

There are two different techniques for calibration: fixed and self-calibration. In fixed calibration, all parameters are extracted off line by placing a special object with known geometry in front of the cameras and processing of the camera images. In the past 20 years, little has changed in fixed calibration techniques [2,8]. Although they provide the most accurate results, they suffer from a number of disadvantages. A special calibration object and user interaction is required. The parameters become useless after a change in the camera parameters due to e.g. zooming.

In self-calibration techniques the parameters are extracted using an image pair that contains the actual scene. This circumvents the disadvantages of fixed methods. However, without any reference to the standard meter, it is impossible to obtain 3-D reconstructions in meters. If absolute scale is not important for the application then self-calibration is a powerful alternative to fixed calibration.

For cameras without lens distortion, self-calibration provides at most 7 parameters in general [1]. Any camera model that has more parameters results in ambiguous scene reconstructions. In [3] self-calibration is performed in a structure from motion application, in which multiple images are taken by a *single* camera. This simplifies the stereo camera model significantly. Together with some additional constraints (such as a square pixels assumption) a model with a sufficiently low number of free parameters is obtained.

For general stereo cameras with lens distortion, it is not known how many parameters can be estimated. In [5] only the distortion parameters are found. In [9] all calibration parameters are found. However, lens distortion is assumed to be equal for both cameras and the images still contain some calibration pattern.

In this paper, we consider self-calibration of all camera parameters on the basis of a single image pair. The image pair contains the actual 3D scene of interest. A dense correspondence field is estimated, from which the calibration parameters are extracted, see Figure 1.

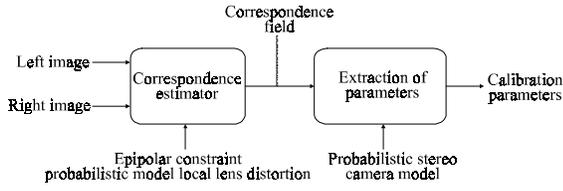


Figure 1: The self-calibration system

The correspondence estimator is a motion estimator with an epipolar constraint that does not need the internal and external camera parameters. It utilizes a probabilistic Markov Random Field model for lens distortion. The distortion MRF model is easily combined with an MRF model for motion estimation which are known for their high accuracy results [6].

Finally a probabilistic stereo camera model is fitted to the correspondence field obtained. We use the stochastic minimization procedure simulated annealing to find the parameters.

The paper is organized as follows. In the next section we describe our correspondence estimator based on motion estimation and a new probabilistic distortion model. Section 3 gives the probabilistic stereo camera model and section 4 describes the extraction of camera parameters from the correspondence field. Finally sections 5 and 6 give preliminary results and conclusions.

2. Correspondence estimation

In this section the goal is to find the corresponding pixels in left and right images, in the absence of calibration knowledge.

Correspondence estimation in a general pair of images requires a 2-D search, which can be performed by motion estimation [6]. However, in the specific case of stereo images the correspondences follow the so-called epipolar constraint [1] that reduces the search to 1-D along specific lines in both images, the so-called epipolar lines. Applying this constraint enhances the quality of the correspondence field obtained significantly.

The camera parameters determine the position and orientation of the epipolar lines. Without lens distortion, the epipolar lines are straight lines in the left and right images. With lens distortion, the epipolar lines become curved.

As correspondence estimator we use a Markov Random Field motion estimator [6] with an additional epipolar constraint that only uses the curvature of the epipolar lines. This is invariant to all camera parameters except lens distortion.

We estimate a correspondence vector field \vec{m} and a scalar field α_L simultaneously. The field \vec{m} defines the corresponding left and right pixels:

$$\begin{bmatrix} x_R \\ y_R \end{bmatrix} = \vec{m} \left(\begin{bmatrix} x_L \\ y_L \end{bmatrix} \right) \quad (1)$$

The scalar field $0 \leq \alpha_L < \pi$ models the direction of the epipolar lines in the left image, defined as the angle between the x_L axis and the line tangent to the epipolar curve. Figure 2 shows the situation in which there is no lens distortion and the epipolar curves are straight lines.

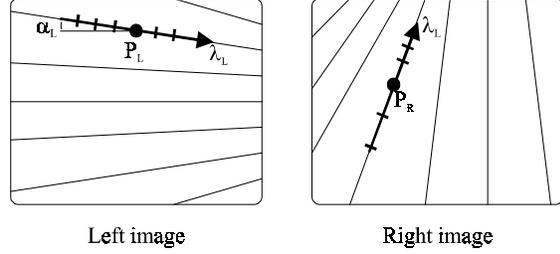


Figure 2: The epipolar direction field α_L

At each point P_L in the left image, we construct a line l_L tangent to the epipolar line and parametrize it by λ_L with $\lambda_L = 0$ at P_L .

For a distortionless left camera the following holds:

$$\frac{\partial}{\partial \lambda_L} \alpha_L = \cos \alpha_L \frac{\partial}{\partial x_L} \alpha_L + \sin \alpha_L \frac{\partial}{\partial y_L} \alpha_L = 0 \quad (2)$$

Now we project the line l_L onto the right image plane using the correspondence field m . As depicted in Figure 2, the uniform parametrization by λ_L is not preserved by the projection. It is affected by both calibration parameters and disparity, which is a function of the particular 3-D scene. When traveling on line l_L with constant velocity, the corresponding walk on line l_R in the right image will thus exhibit unknown accelerations.

If the right camera is also distortionless, the line l_R in the right image is a straight line. To remain on a straight line, the acceleration should always be parallel to the velocity. Formulated mathematically, we have:

$$\frac{\partial^2}{\partial \lambda_L^2} \vec{m} = K \cdot \frac{\partial}{\partial \lambda_L} \vec{m} \quad (3)$$

The value of K depends on the calibration parameters and the particular 3-D scene. It changes with image position.

For cameras with distortion, (2) and (3) do not hold. We now model any deviation from (2) and (3) by Gaussians with zero mean, which result in quadratic energy functions in an MRF model. After rewriting (2) and (3) into:

$$\alpha_L' = 0 \quad \vec{a} = K \vec{v} \quad (4)$$

we construct the curvature energy terms by:

$$E_L = |\alpha_L'|^2 \quad E_R = \frac{|\vec{a} \times \vec{v}|^2}{v^2 a^2} \quad (5)$$

The normalisation in E_R makes sure the energy is invariant to the absolute magnitude of both a and v , which depend on the particular 3-D scene.

We have now defined a correspondence estimator with 3 unknown scalar fields (motion m_x , m_y and epipolar angle α_L) and two field equations (2) and (3). The total MRF model is then defined including a luminance difference term [6]:

$$E_{TOT} = \sum_{\substack{\text{all pixels in} \\ \text{the left image}}} \gamma |I_L - I_R| + E_L + E_R \quad (6)$$

This term reflects the Constant Image Brightness (CIB) assumption, which states that corresponding pixels have similar luminance. The weight factor γ regulates the influence of the curvature energies relative to that of the luminance differences. It will be determined by experiment.

The a posteriori probability for \vec{m} and α_L after observation of the images is given by:

$$P_{\vec{m}, \alpha_L | I_L, I_R} \propto e^{-E_{TOT}} \quad (7)$$

The MAP solution of the motion and curvature field are found by minimization of E_{TOT} . A hierarchical simulated annealing algorithm is used to approximate the MAP solution.

3. Stereo camera model

In this section we describe our probabilistic stereo camera model. Figure 3 shows the complete model. Five reference frames are defined, the stereo frame SF with origin O_{SF} , the left/right lens frames LF_L and LF_R , and the left/right projection frames PF_L and PF_R . Without calibration object with a known length in meters, it is impossible to obtain measurements in meters. We therefore select the camera baseline as our unit of length. The frame SF is defined to be a right handed frame in which the two optical centres lie on the x-axis symmetrically around the origin.

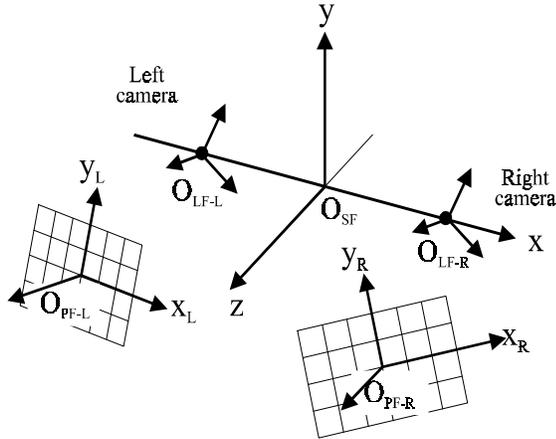


Figure 3: The stereo camera model

Gaussian probability densities are assigned to all parameters. We will now describe the parameters and their mean and variance.

In baseline units, the optical centres have fixed coordinates in SF , $(-1/2, 0, 0)$ for the left camera and $(+1/2, 0, 0)$ for the right camera. The orientations of the the left and right lens frames are defined by two sets of Euler angles (ϕ_x, ϕ_y, ϕ_z) . The lenses are present in the origins of the lens frames, oriented in the xy planes. We assume radial symmetry in the lenses and thus we can assume $\phi_z = 0$. The other two angles are modeled by $\mu = 0$ and $\sigma = 2$ rad. This introduces a small bias towards cameras that are aimed at the same object. The reference frame SF is defined up to a rotation around the x-axis. We can therefore introduce an arbitrary equation that eliminates either $\phi_{x:L}$ or $\phi_{x:R}$, such as $\phi_{x:L} + \phi_{x:R} = 0$.

We assume the CCD to be perfectly flat, have perfectly perpendicular image axes and perfectly rectangular pixels. The image formation is invariant for scaling of the triplet focal length, horizontal and vertical pixel size. Therefore we choose without loss of generality the horizontal size of the pixels equal to 1 (both cameras) and the vertical size equal to $R_{L/R}$, the pixel ratio. The ratio is modeled as $\mu = 1$, $\sigma = 0.3$.

The positions of the projection frames $PF_{L/R}$ (CCD chip) relative to the lens frames $LF_{L/R}$ are defined by two vectors $(O_{PF}^{X_{LF}}, O_{PF}^{Y_{LF}}, O_{PF}^{Z_{LF}})$. The first two numbers define the intersection of the lens optical axis with the CCD (mis-positioning) and are modeled by $\mu = 0$, $\sigma = 10$ (pixels). The third is the focal length modeled with $\sigma = \infty$.

The orientation of the projection frames $PF_{L/R}$ relative to the lens frames $LF_{L/R}$ is defined by two sets of Euler angles $(\theta_x, \theta_y, \theta_z)$. θ_z defines the rotation of the projection frame and is modeled with $\sigma = \infty$. The other two angles define non perpendicular CCD placement and are modeled with $\mu = 0$, $\sigma = 5^\circ$.

Since CCD misplacement is incorporated in several of the previous parameters, lens distortion can be modeled simpler than in [7]. We use only the radial distortion parameter K_3 with $\mu = 0$, $\sigma = 0.3$. This parameter relates the slope of the incoming and outgoing light rays by:

$$r_{out} = r_{in} + K_3 r_{in}^3 \quad (8)$$

The slopes are defined by the arctangent of the angles of the ray to the z axis of the lens frame.

We have now defined a stereo camera model that contains 19 parameters ϕ_i . It consists of 3 external parameters (three independent lens rotations), 14 internal parameters (six for CCD positions, six for CCD orientations and two pixel ratios) and 2 lens distortion parameters.

4. Extraction of calibration parameters

Once we have the correspondence field \vec{m} , we can extract the camera parameters. We use the following procedure.

If we have an estimate of the camera parameters, each corresponding pixel pair in the motion field \vec{m} can be triangulated to yield a 3-D scene point, see Figure 4.

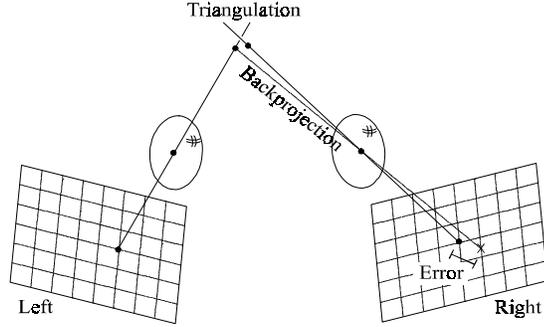


Figure 4: Extraction of scene points by triangulation of corresponding pixels

For the true motion field and parameters, each pair of constructed light rays intersect in a scene point. If there are errors in either motion field or parameters, the light rays do not intersect in general. By least squares estimation we can obtain the two closest points on the light rays. These points are projected back on the CCD's vice versa as is shown in Figure 4 for the right CCD.

The difference in forward and backward light rays for each CCD is assumed to be caused by errors in the estimated motion field. These errors are modeled to be independent for each correspondence. A 2-D Gaussian is used with $\mu=0$ and σ_m to be determined by experiment.

The total probability for the parameters given prior knowledge defined in section 3 and triangulation errors is defined as $P_{TOT} = e^{-E_{TOT}}$, with E_{TOT} equal to:

$$E_{TOT} = \sum_{\text{all parameters}} \left(\frac{\phi_i - \mu_i}{\sigma_i} \right)^2 + \sum_{\text{all pixel pairs}} \left(\frac{\Delta m}{\sigma_m} \right)^2 \quad (9)$$

We use a MAP criterion to define the best set of parameters. A simulated annealing procedure is used to minimize E_{TOT} (maximize P_{TOT}).

5. Experiments

We tested the system separately for the correspondence estimation and parameter extraction parts. In our experiments with the correspondence estimator defined in section 2, we did not observe convergence to a relevant approximation of the solution. For the moment, we downscaled the

complexity by discarding the α_L field. The energy functions in (6) were selected as follows:

$$E_L = 0 \quad E_R = |\vec{a}_x|^2 + |\vec{a}_y|^2 \quad (10)$$

with

$$\vec{a}_x = \frac{\partial^2}{\partial x_L^2} \vec{m} \quad \vec{a}_y = \frac{\partial^2}{\partial y_L^2} \vec{m} \quad (11)$$

This yields smoothing on the correspondence field which is invariant to scale, translation and rotation. Thus, it is insensitive to large differences in focal distance (zoom) and position/orientation of the camera pair. The drawback is that (5) applies the epipolar constraint and does not smooth, while (10) applies smoothing equally in all directions.

Figure 5 and 6 show results with two stereo pairs. In Fig. 5 the stereo pair consists of a natural image and a 90° rotated version. For the correspondence estimator, we found that $1 \leq \gamma \leq 10$ provides convergence to relevant solutions. Using the motion field, we interpolated the stereo pair. Clearly, the estimator can handle very large translations and rotations. Figure 6 shows a synthetic stereo pair and an interpolated image. Due to slight differences in focal distances, pixel ratios and viewing angles of the left and right cameras, the two scene objects have different sizes in the stereo image pair. The interpolated image shows that the estimator is capable of handling these differences in scale.

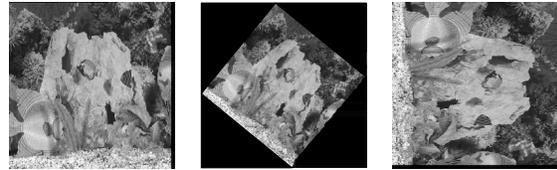


Figure 5: Left, interpolated and right image from a stereo pair with very large motion thru 90° rotation.



Figure 6: Left, interpolated and right image from a synthetic stereo pair with scale differences

Still, the motion fields of this downscaled estimator were not accurate enough for the parameter extraction algorithm. The results either did not converge, or converged to the wrong solution (for the images in Figs. 5 and 6 the true parameters are known).

To test the parameter extraction algorithm, we used a synthetic image pair with available true motion field ($\sigma_m = 0$) and true camera paraters, see Fig. 7.



Figure 7: Synthetic image pair and true motion field (m_x component)

With the algorithm of section 4, we were able to obtain all parameters, except for one focal length which had to be provided manually. Table 1 shows the true and measured parameters ϕ and ϕ^* for the left and right cameras. The $O_z = 1000$ has been entered explicitly for the left camera. The simulated annealing procedure is a costly computation: It took about 30 minutes on a SGI octane machine, while using a four times subsampled motion field (only 1/16 of all correspondences).

	ϕ_L	ϕ_L^*	ϕ_R	ϕ_R^*
ϕ_x	1°	0.998°	-1°	-0.998°
ϕ_y	15°	13.43°	-15°	-16.82
O_x	-4	-8.12	8	8.61
O_y	3.1415	3.11	9	8.99
O_z	1000	1000*	1000	1016.5
θ_x	2°	1.55°	2°	1.54°
θ_y	4°	5.01°	3°	4.74°
θ_z	10°	10.04°	10°	10.08°
R	1	1.00	1	1.01
K_3	1.5	1.50	1.5	1.55

Table 1: True ϕ and measured ϕ^* parameters.

6. Conclusions

We presented a self-calibration method for stereo cameras consisting of two parts. First correspondences are estimated in a stereo image pair that contains the actual scene of interest, after which the camera parameters are extracted from the correspondence field. As our results are very promising but can be improved, we will discuss them in somewhat large detail.

We designed a new correspondence estimator that is based on motion estimation with a local epipolar constraint. It is based on extraction and penalizing of the lens distortion of the cameras, without further knowledge of the camera parameters. Currently, we obtained only results with a downsampled version, that does not include the epipolar constraint. It does include invariance to translation, rotation and scale, and thus, invariance to most of the camera parameters such as camera position, orientation and focal distance. For image interpolation, subjectively very good results were obtained for stereo pairs involving very large translations, rotations and scale differences.

For the parameter extraction we used a simulated annealing method. This method is easy to design and

does not require analytical computation such as derivatives in e.g. downhill methods. The price is the very high computational load. Although the downsampled correspondence estimator yields very good interpolation results, we could not extract from it any relevant camera parameter set. Thus, the correspondences are accurate enough for photometric interpolation, but not accurate enough for geometric triangulation as shown in Fig. 4.

We therefore tested the parameter extraction algorithm with a synthetic stereo image pair with known true correspondence field and camera parameters, and found the following. In self-calibration methods for lens-distortion-free cameras, only 7 parameters can be obtained. The introduction of lens distortion enhances the quality of camera models by definition. Our results indicate that it also enables the measurement of more parameters, in addition to the distortion parameters themselves. We found that 18 out of 19 parameters in the camera model could be obtained. If the two lens distortion parameters are not counted, we are thus able to measure 16 internal and external parameters.

Our future research objective is twofold. First, we would like to incorporate the local epipolar constraint in the motion estimator. We expect that good results can be obtained with more computational power rather than conceptual changes. Secondly, we are investigating if the single remaining freedom in the calibration can be accounted for. Options that we consider are the use of a stereo camera model based on a moving single camera such as in [3], the before-hand calibration of camera-life-time constant parameters such as the pixel ratio, and the introduction of a third reference camera that is partly calibrated before-hand.

References

- [1] O. Faugeras, "Three-dimensional computer vision, a geometric viewpoint", MIT Press, 1993
- [2] F. Pedersini, D. Pele, A. Sarti and S. Tubaro, "Calibration and self-calibration of multi-ocular camera systems", in proceedings of the *International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging (IWSNHC3DI'97)*, Rhodos, Greece, pp. 81-84, 1997
- [3] M. Pollefeys, R. Koch, M. Vergauwen and L. van Gool, "Flexible acquisition of 3D structure from motion", in proceedings of the *IEEE Image and Multidimensional Digital Signal Processing (IMDSP) Workshop '98*, pp. 195-198, 1998
- [4] P.A. Redert, C.J. Tsai, E.A. Hendriks and A.K. Katsaggelos, "Disparity estimation with modeling of occlusion and object orientation",

Proceedings of SPIE VCIP98, Vol. 3309, pp. 798-808, 1998

- [5] G.P. Stein, "Lens distortion calibration using point correspondences", in *IEEE Conference on CVPR*, pp. 602-609, 1997
- [6] C. Stiller, "Object-based estimation of dense motion fields", *IEEE Transactions on Image Processing*, Vol. 6, No. 2, pp. 234-250, 1997
- [7] J. Weng, P. Cohen and M. Herniou, "Camera calibration with distortion models and accuracy evaluation", in *IEEE Transactions on PAMI*, Vol. 14, No. 10, pp. 965-980, 1992
- [8] H.J. Woltring, "Simultaneous multiframe analytical calibration (SMAC) by recourse to oblique observations of planar control distributions", *SPIE Vol. 166 Applications of Human Biostereometrics*, pp. 124-135, 1978
- [9] Z. Zhang, "On the epipolar geometry between two images with lens distortion", in proceedings of *Int. Conf. Pattern Recognition (ICPR)*, Vol. 1, pp. 407-411, 1996