

Towards Real Time Camera Self Calibration: Significance and Active Selection

Antonio Martos¹, Lars Krüger¹, Christian Wöhler²

¹Daimler AG, Group Research and Advanced Engineering, 89081 Ulm, Germany

²Image Analysis Group, Technische Universität Dortmund, 44221 Dortmund, Germany

martos.antonio@gmail.com, lars.krueger@daimler.com, christian.woehler@web.de

Abstract

In this study we use statistical methods for assessing the quality of camera self-calibration of a stereo vision system, considering accuracy, reliability, and significance of the solution. The results of the optimisation process are subject to model fitting analysis. In this context, we define figures such as significance, confidence levels, and error prediction curves across the complete field of view. Furthermore, we analyse the effects of the number and spatial distribution of image correspondences on the calibration quality and examine the efficiency of uniform random selection strategies to reach a predefined accuracy level with as few correspondences as possible. We introduce a strategy for active selection of correspondences based on the prediction error, which achieves a faster convergence of calibration quality than random selection without uniformity requirements. This active selection strategy is combined with random selection, relying on an adaptive selection probability determined based on the prediction band values, which leads to a faster and steadier convergence than uniform random selection. Our evaluation is performed using both synthetic data and real-world data from a vehicle-based stereo vision system. Our empirical analysis improves the understanding of camera calibration procedures and helps to find a good trade-off between accuracy and computational efficiency by specifically selecting the correspondences which are most relevant for the calibration process.

1. Introduction

Stereo vision systems require good knowledge of the geometric properties of the cameras used for image acquisition. For some methods (e.g. [5, 7]) a sufficiently accurate camera calibration is necessary even to operate properly, as they require rectified stereo images as an intermediate step. For other stereo approaches, the quality of the calibration will at least influence the quality of the output results.

Additionally, for real-time vision systems in mobile

robots or vehicle-based driver assistance systems, limited computing resources and time are available for self/autocalibration, while reliability and accuracy are still a fundamental requirement.

When using natural real-world scenes, such systems can sometimes provide a considerable amount of corresponding points for self-calibration, specially if several frames are accumulated. Still it may be difficult to predict the availability of correspondences in time and they may be not uniformly spread across the field of view or be redundant.

However, for the usual iterated non-linear optimisation methods [10], computation time and memory requirements of camera calibration algorithm largely depend on the number of correspondences used [9]. On the other hand, accuracy converges asymptotically with respect to the number of correspondences, thus the obvious choice to keep computation time within desirable limits is to select a limited number from the set of already available correspondences. This paper focuses on the analysis of different selection strategies aiming to achieve a desired quality level with as few correspondences as possible.

We concentrate on self-calibration of a stereo system for which ground truth information is not available and thus only such residual information is used. The most typical quality measures used in the computer vision literature on self-calibration methods are based on statistics of the reprojection error or epipolar constraints [9] measured in image space. However, these measures are not very consistent as they are highly dependent on the particular data set (number and spatial distribution of correspondences, noise levels). Furthermore, they provide information only on how well the model fits the particular data set used for calibration, but no information regarding how the model is expected to fit to new data in that range (interpolation), and especially outside the calibrated range (extrapolation).

In a general metrological context, successful calibration has to be consistent and systematic, and both the calibration and subsequent measurements should be traceable. In camera calibration it is also necessary to define when the system can be considered fully calibrated in a consistent

and meaningful way with respect to its prediction capabilities for new measurements. This property should be computed across the complete measurement range, i.e. the field of view (FOV), forming the so-called error prediction iso-curves. They represent the usual problems in calibration in an intuitive and easily interpretable way, regarding the possible occurrence of an inhomogeneous distribution of input data, over-fitting of the model parameters, and sensitivity to input noise.

2. State of the Art

2.1. Self-calibration and quality measures

Self/auto-calibration of stereo vision systems is a widely regarded research topic, see e.g. [4, 6, 17] for a general overview of the subject assuming geometric camera models. Mostly, a two-stage procedure is used: First, an approximate (linear) solution is computed that minimizes an objective function w.r.t. the unknowns; some of the camera parameters (either internal or external). Second, this solution is refined by a non-linear, iterative minimisation of a similar objective function, using the values from the first stage as starting values for the unknowns. For the remainder of this paper, we concentrate on the non-linear minimisation (Bundle Adjustment) as here the gain in computation time is more important when reducing the number of correspondences.

The ultimate quality measure in stereo vision should be the comparison of 3D scene reconstruction with ground truth data [13]. This can be obtained from semi-synthetic images, but by construction these will follow a particular camera model “perfectly”, therefore they are not complete for the purpose of this paper.

Ground truth data could be as well obtained by laser-scanning the scene [15] but this requires instrumentation which is not available in most real time systems. It is obvious that this kind of quality measure cannot be used for performing auto/self-calibration.

In [6] and many other publications, the objective function often used is (or is related to) the squared errors of the available residuals. However, related quantities such as the mean residual error (MRE) measured in pixels, or the related Residual Sum of Squares (RSS), are often provided as well as criteria for evaluation of the quality of different calibration methods.

From the regression analysis literature [2] it is known however, that the RSS will be artificially small when only a few measurements are used, i.e. when over-fitting of the model occurs, resulting in misleading conclusions about the quality of calibration.

This is relevant since we intend to use as few points for calibration as possible. Additionally we have to cope to some degree with possibly close to degenerate spatial distri-

bution of correspondences and extrapolation from the available measurements to the complete range of measurement, i.e. the whole field of view.

2.2. Optimisation and χ^2 test

Optimisation algorithms in calibration search for the minimum in parameter space of some objective function. In computer vision the most common choice is the residual sum of squares [6] $RSS = \sum_i r_i^2$, where r_i is the residual of correspondence i . The actual value in pixel² units is often neglected as meaningless since for optimisation it can be replaced by any other function having the same minimum. In fact this approach is equivalent to another well-known measure, the root mean square error $RMSE = \sqrt{RSS/N_m}$ with N_m as the number of measurements [2, 17]. However, this value may be measured in pixel units, although it still has no clear statistical meaning. The RMSE measure can be generalized by dividing each residual r_i by a weighting factor σ_i . If we assume that the residuals are independent of each other and that they are drawn from a known probability distribution with standard deviation σ_i , this function becomes a statistically interpretable quantity known as the (Pearson’s) χ^2 statistic:

$$\chi^2 = \sum_i \frac{r_i^2}{\sigma_i^2}. \quad (1)$$

The choice of this statistic as a goal function is justified by several reasons: It is dimensionless, system independent and statistically interpretable. It is also a standard measure of the goodness-of-fit [12]. The χ^2 distribution represents the probability of independent events (measurements) to occur in a purely random manner and is a maximum likelihood estimator under this assumption. For a simpler interpreta-

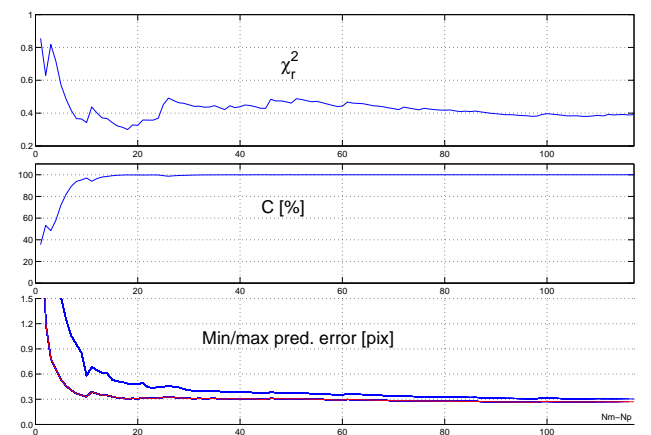


Figure 1. Example of the evolution of the χ_r^2 value (top), the p-value confidence C (centre), and the maximum and minimum predicted error values across the FOV w.r.t. the number of correspondences N_m randomly selected from real data (bottom).

tion, the reduced measure χ_r^2 is defined as

$$\chi_r^2 = \frac{\chi^2}{\nu}, \quad (2)$$

[12], where $\nu = N_m - N_p$ is the number of degrees of freedom with N_p as the number of free parameters in the model. This estimator provides a convenient rule of thumb. The conventional limit of significance corresponds to $\chi_r^2 = 1$. Lower values represent overfitting while higher values mean disagreement between the model and the data.

A more formal interpretation makes use of the χ^2 cumulative distribution function $\chi_{\text{cdf}}^2(x, \nu) = \int_0^x \chi^2(\alpha, \nu) d\alpha$. The values of χ^2 and the number of degrees of freedom can thus be translated into a p-value of confidence C , denoting the likelihood of the overall fit:

$$C = P_{\chi^2, \nu} = 1 - \chi_{\text{cdf}}^2(\chi^2, \nu) \quad (3)$$

The p-value will be near zero when it is unlikely that the model explains the data and near 1 when the model is consistent with the data. A common (but still arbitrary) reference threshold value is 95% (confidence level $\alpha = 0.05$). Fig. 1 shows an example of the evolution of χ_r^2 and C with an increasing number of randomly distributed correspondences.

3. Prediction surfaces

3.1. Definition

Although it is not new in computer vision to calculate confidence intervals for the fitted (camera) parameters [3], it is however surprisingly difficult to find references concerning confidence or prediction bands in the space of the dependent variables. In regression analysis, confidence bands are curves around the fitted model (function) which are expected to enclose the true model with a predefined probability or confidence level, e.g. 95%. This is caused by the uncertainty of the fit propagated to the model parameters. The extent of the bands gives an idea of how well the function fits the data at different positions in data space. On the other hand, the 95% prediction bands are curves enclosing the area that is expected to contain 95% of the (predicted) future data points when using that model. These curves include the uncertainty of the model fit (like the confidence bands) and additionally account for the scatter of the data around the fit. Therefore, the prediction bands are always broader than the confidence bands.

These bands can be calculated using Monte-Carlo or bootstrapping methods for repeatedly testing sample variations of model fit parameters and also of the measurements, taking into account their respective probability distribution if known. The bands can also be calculated analytically by linear error propagation, i.e. a first order approximation which is sufficient when the measurement noise has

a Gaussian-like distribution. Possible deviations from the Gaussian distribution in real data will be discussed later on.

3.2. Prediction surfaces in camera calibration

To compute prediction surfaces, the covariance matrix of the parameters is required. In our implementation we use the Jacobian \mathbf{J} computed in the last iteration of the optimisation step,

$$J_{ij} = \frac{\partial r_i}{\partial p_j} \quad (4)$$

$$K_{ij} = \frac{J_{ij}}{\sigma_i}, \quad (5)$$

where $\mathbf{p} = (\phi_1, \phi_2, \phi_3)$ denotes the parameters to be optimised (which in our example scenario correspond to the three relative orientation angles of the cameras) and \mathbf{K} the weighted Jacobian, which is used to compute the covariance matrix Σ according to

$$\Sigma = (\mathbf{K}^T \mathbf{K})^{-1}. \quad (6)$$

As stated in Sec. 3.1, prediction bands can be calculated for any position in the image. In this paper, these positions correspond to all integer pixel positions \mathbf{x}_1 in the left image. As the point measurements are independent of each other, we illustrate the computation for an arbitrary point (u, v) in the left rectified image.

To compute prediction surfaces, we use the Jacobian vector \mathbf{g} of the model function f with $g_j(u, v) = \partial f(u, v) / \partial p_j$ and the covariance matrix Σ to obtain a dimensionless value c according to

$$c(u, v) = \mathbf{g} \Sigma \mathbf{g}^T. \quad (7)$$

The half width σ_P of the prediction interval then corresponds to

$$\sigma_P(u, v) = \frac{\beta}{2} \sqrt{(c(u, v) + 1) \frac{\text{RSS}}{\nu}}, \quad (8)$$

where β is a constant derived from the inverse cumulative t-student distribution t_{cdf}^{-1} according to $\beta = t_{\text{cdf}}^{-1}(1 - \alpha/2, \nu)$ with α as the probability of the desired confidence level for the surfaces. At a confidence level of 95% we use $\alpha = 0.05$ throughout this paper.

The derivation of Eq. (8) is based on the derivation of the parameter confidence band as described in [8]. As both bands are symmetrical around the model function we can subtract the model to rectify the bands and perform a residual analysis. When calculated for a two-dimensional domain in image space, the term ‘‘band’’ is replaced by ‘‘surface’’ and conveniently represented by the error iso-curves overlaid on the images.

3.3. Usage and interpretation

In computer vision the evaluation of the overall quality of fit is commonly done by a single goodness-of-fit-like number [6], standard deviations of the model parameters [3, 11], and/or analysis of the correlation matrix of the parameters [18]. Error prediction curves, however, not only aim for justifying the fit of the model to the calibration data but also to predict the error of the system for future measurements across in the full data range. Their shape depends on the spread of the correspondences used in the fit. In areas close to several measurements consistent with the model, prediction bands are expected to be narrow, indicating a good local fit. Otherwise, when extrapolating to regions of the measurement space far away from the data used for calibration or when some measurements show a poor consistency with the model, the error prediction values increase.

This natural behavior is depicted in Fig. 2, where the same number of measurements are spread in a very different manner across the image. From the value of the bands one can immediately observe in which example the model works better. It should be noticed that “interpolation” among existing points is much more accurate than “extrapolation” to outside regions. Fig. 1 shows one possible evolution of the maximum and minimum values of the curves across the field of view for a random uniform selection sequence of the correspondences used for self-calibration, where the error converges asymptotically.

4. Epipolar Geometry

The objective function of the self-calibration problem is a constraint on the four-dimensional measurements of each correspondence: The 2D positions of corresponding points (u_l, v_l, u_r, v_r) in the left and the right image are subject to the epipolar constraint, thus only three of them are free and the fourth is a measure of the residual, i.e. the distance to the re-projected point orthogonal to the epipolar line.

For a two-view problem, the residual deviation from the epipolar constraint is the objective function to be minimized [6]:

$$e_i = \mathbf{x}_{l_i}^T ([\mathbf{T}]_{\times} \mathbf{R}(\phi_1, \phi_2, \phi_3)) \mathbf{x}_{r_i}, \quad (9)$$

where $[\mathbf{T}]_{\times}$ denotes the cross-product matrix of the translation vector \mathbf{T} between the two cameras, which is assumed to be constant. The term $\mathbf{R}(\phi_1, \phi_2, \phi_3)$ denotes a rotation matrix in terms of the XYZ -fixed rotation angles ϕ_1 , ϕ_2 , and ϕ_3 . An ideal solution would imply that each corresponding pair i of 2D feature points (x_l, x_r) from two undistorted images has a residual error $e_i = 0$. Since real point measurements are subject to noise (e.g. caused by image noise or the inherent accuracy of the correspondence analysis algorithm), calibration is usually stated as a least-mean-square

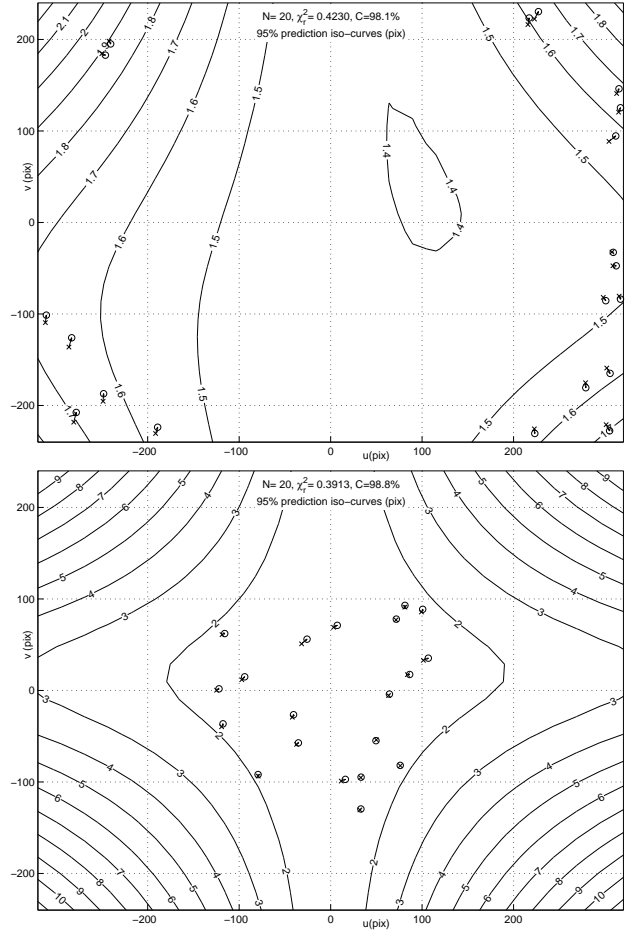


Figure 2. Two different distributions of 20 correspondences (circles: measurements; crosses: re-projections) provide half widths of the prediction interval that range from 1.4 to 2.2 pixels (top) and from 1.9 to 11 pixels (bottom).

solution determined by a nonlinear optimisation technique such as the Levenberg-Marquardt algorithm [10].

If an initial approximate calibration for the three rotation angles ϕ_1, ϕ_2, ϕ_3 is available, both image spaces may be rectified such that the epipolar constraint explicitly becomes $v_d = f(u, v, u_d)$. There is a bijective transformation for any point (u, v) in the first image with a horizontal disparity u_d to a point (x, y, z) in Euclidean space if the residual is $v_d = 0$. The residual has to be minimized such that the scene can later be reconstructed accurately. For this purpose, we restate Eq. (9) such it becomes explicit. The easiest way is to switch from orthogonal distance to vertical distance: The deviation of the measured value of v_r from the predicted value is computed as the orthogonal distance of v_r to the epipolar line of \mathbf{x}_l , where the normal begins at

the point u_r :

$$\begin{aligned} r_i(\mathbf{x}_{l_i}, \mathbf{x}_{r_i}) &\approx e_i(\mathbf{x}_{l_i}, \mathbf{x}_{r_i}) \\ &= v_{r_i} - f(\mathbf{x}_{l_i}, u_{r_i}). \end{aligned} \quad (10)$$

The error introduced by this approximation is related to the cosine of the angle between the epipolar line in the right image and the horizontal line. As we deal with a well constructed stereo vision system, this angle is always smaller than 1° and its influence is therefore negligible.

5. Evaluation

5.1. Experimental setup and synthetic test

The proposed procedure is demonstrated based on real images acquired with a vehicle-based stereo camera system attached to the windscreen. An embedded computer processes both images in real time, extracts common features and generates three-dimensional measurements. The camera system is pre-calibrated with approximate parameters, but the windscreen is expected to alter the model as thermal and mechanical stress cause deviations in time, specially in the relative pose or rotation of the second camera with respect to the first. To compensate this effect, the system needs to be auto-calibrated in the field using natural features. We assume a standard central projection camera model with known lens distortion correction.

We evaluated the described prediction surfaces using real-world and synthetic data. For the real-world dataset we calibrated a synchronized stereo-vision system with a baseline of 210 mm and a camera constant of 1026 pixels using [1]. Our evaluation relies on image pairs extracted from a longer test sequence, rectified such that epipolar lines correspond to pixel rows (standard epipolar geometry [6]). The features for self-calibration are extracted using a pixel accurate method based on the Census transform [14], where the sub-pixel accurate displacements are computed according to the KLT method [16].

With respect to our synthetic test, the data set consists of a variable number of correspondences that are uniformly distributed across the field of view. The noise of each point is Gaussian with $\sigma = 0.32$ pixels and a slight miscalibration is introduced to test a solution that has not yet reached its optimum.

5.2. Effects of noise in image features

For computing Eq. (2) we assumed before that the a-priori error of each correspondence comes from a known probability distribution whose individual uncertainties σ_i are also known. Formally, this assumption would require to extend the model by additional parameters to be integrated into the calibration procedure. However, many systems completely ignore such considerations. For example,

whenever RSS-like objective values are used, it is equivalent to a particular case of χ^2 where each residual in pixel units is assumed to be implicitly divided by a constant value $\sigma = 1$ pixel.

In some cases, image matching algorithms such as KLT or SIFT can provide values of the uncertainty or noise levels for the coordinates of each individual image feature. Still this value may not be reliable as it depends on several issues that are difficult to determine in advance, such as image resolution and quality, local image structure, geometry, size, shape and uniqueness of the feature, or sampling problems.

However, we assume in our scenario that noise can be determined empirically either as a global average single figure, or as a function with respect to image position. The latter will be shown as an alternate interpretation of the prediction curves. Although this may seem to lead to a circular definition, it is in fact a consistent posterior verification of the model, including the assumptions about noise as an additional model unknown as a consequence of the central limit theorem. This can be verified with a sufficiently large reference data set.

In our real-world data example, we start by provisionally taking an arbitrary accuracy goal assuming a uniform value of $\sigma_i = 0.5$ pixels for each correspondence. This is a conservative guess based on previous (subjective) observations of the accuracy of the KLT algorithm in our system and other considerations of the physical system specifications. Later we will show that this a priori estimation is not critical and can be calculated in a two-step procedure, as any other unknown parameter.

As a reference data set, we use 3215 correspondences extracted from a real stereo sequence. After filtering outliers, we perform a Levenberg-Marquardt optimisation [10] to find a near best-fit solution for camera orientation. Fig. 3 shows the rectified left image with overlaid correspondence pairs and computed prediction iso-curves.

In this case the calibration is nearly optimal. Notably, the shape of the curves confirms a residual error prediction which almost constantly amounts to 0.318 pixels across the complete image space. This observation supports the idea that a constant value may be sufficient to model image noise for this system, although $\sigma = 0.5$ pixels was an overestimation. The value $\chi_r^2 = 0.42$ indicates that the result is statistically significant with some overfitting, and the p-value being almost identical to 1.0 confirms the consistency of the fit. The gross a-priori estimation of the constant σ has quite a small effect in the calculation of the prediction curves. Table 1 shows the maximum of the prediction value σ_P and the value of χ_r^2 computed in the complete field of view for different values of σ .

We observe that the proposed method is quite tolerant w.r.t. reasonable guesses of σ , reaching a stable value for

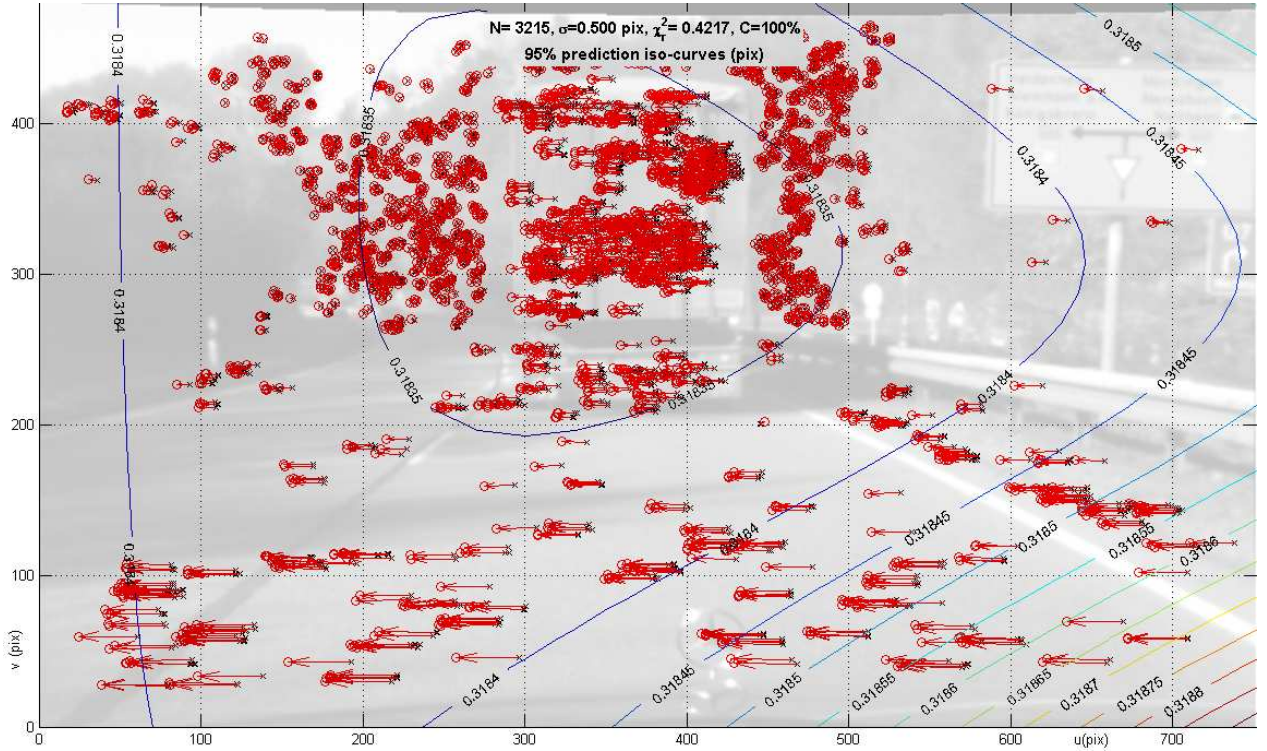


Figure 3. Left rectified image and prediction iso-curves using 3125 correspondences (circles: x_r , crosses: x_l), $\sigma = 0.5$ pixels. The resulting value of $\chi_r^2 = 0.42$ means that the fit is reliable and significant. The prediction curves are almost constant with $\sigma_P = 0.318$ pixels.

$\chi_r^2 \approx 1$, marking the limit of significance of the solution. The error prediction curves are stable when σ is underestimated and increase slowly when σ is overestimated. So far, this example illustrates an accurate but expensive calibration, using all available correspondences. It will be used as a reference for comparison with computationally much cheaper calibrations performed using a small number of correspondences.

The second order central moment (standard deviation) is computed from the real data, corresponding to 0.315 pixels, which is quite similar to the mean error of the prediction curves. For the following analyses both example values, $\sigma = 0.315$ pixels as the optimal error and $\sigma = 0.5$ pixels as an arbitrary conservative goal, serve as a reference in the graphs.

Table 1. Maximum predicted error (MPE) σ_P and χ_r^2 for different initial values of σ . The “true” mean residual is 0.315 pixels.

σ	0.10	0.25	0.32	0.50	1.00	2.00	5.00
MPE	0.318	0.318	0.318	0.318	0.320	0.326	0.365
χ_r^2	10.5	1.69	1.03	0.42	0.10	0.03	0.00

5.3. Correspondence selection

For an embedded real-time system, processing around 3000 correspondences is time-consuming. However, a solution which is significant and accurate enough can be reached with a much smaller number of correspondences. For our stereo system with $N_p = 3$ parameters (the camera rotation angles) we consider calibration experiments with an increasing number of correspondences, starting from $N_m = 4$ to $N_m = 120$, extracted from the complete available set following different selection strategies.

First, we consider based on synthetic data that each new selected correspondence is closest to a randomly uniform point generated in the field of view. One of these experiments is shown in Fig. 1 which indicates that high confidence and significance levels are reached soon.

It is intuitively clear that any selection strategy evaluated on smaller subsets of a large data set should not produce better error estimation than the full set. However, it can be seen in Fig. 4 that the RMSE tends to violate this principle, often providing error estimates below the limit of 0.315 pixels obtained when considering all available correspondences. Instead, we therefore propose to use the Maximum Predicted Error (MPE), which is the maximum value of the prediction curves calculated in a uniform grid of 25×25 points across the complete field of view. This

error estimate obeys the stated principle. Moreover, unlike the RMSE, it does not only describe the capability to fit the previous data but also predicts correct results for new measurements. The system can only be considered fully calibrated if the solution is significant at a desired p-value (e.g. 95%) and the MPE accuracy is in some desired range (e.g. 0.5 pixels) across the complete field of view. The density plot shows accumulation of different resulting RMSE and MPE values obtained from 1000 different Monte-Carlo experiments. Black areas show the most probable values. The density plot shows that using the strategy of random selection, at least 35 correspondences are required to achieve an accuracy of $MPE < 0.5$ pixels with a significance level of 95%.

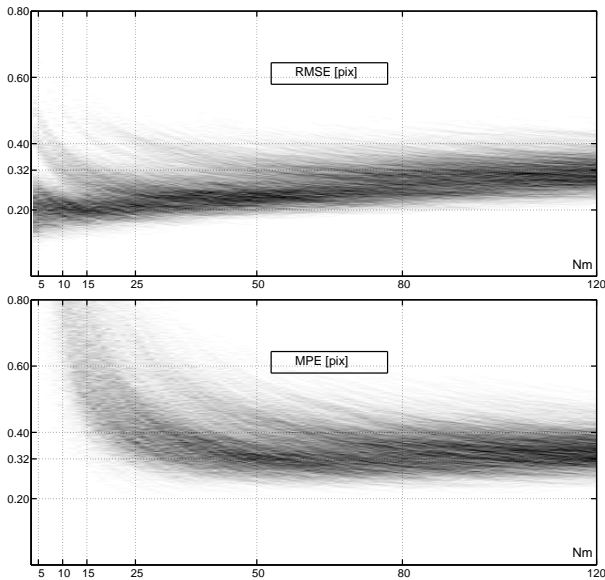


Figure 4. Comparison of the asymptotic convergence of RMSE (optimistic) and MPE (pessimistic) for randomly uniform selection on synthetic data. With few correspondences, the RMSE gives a premature impression of false accuracy (true $\sigma = 0.315$ pixels).

The strategy of uniformly random selection works quite well with synthetic data. It requires, though, either a uniform spread of the data or a large data set from which to extract more uniformly distributed subsets that cover the complete field of view. However, in vehicle-based real-time systems correspondences are often extracted sequentially concentrated in clusters, with very different densities depending on the scene. Even worse, statistics of residuals and errors vary among different areas, leading to systematic errors when the selection is not uniform.

An example is shown in Fig. 5 (top), where a real sequence was re-sampled by bootstrapping. Asymptotic convergence of MPE is slower here than in the uniform case and markedly multi-modal, causing a solution that is unstable with respect to the selection experiment. The latter is

due to the fact that several sequences start with all correspondences in the same area, thus providing limited coverage of the field of view. It requires more than 120 correspondences in this situation to reach an error threshold of 0.5 pixels.

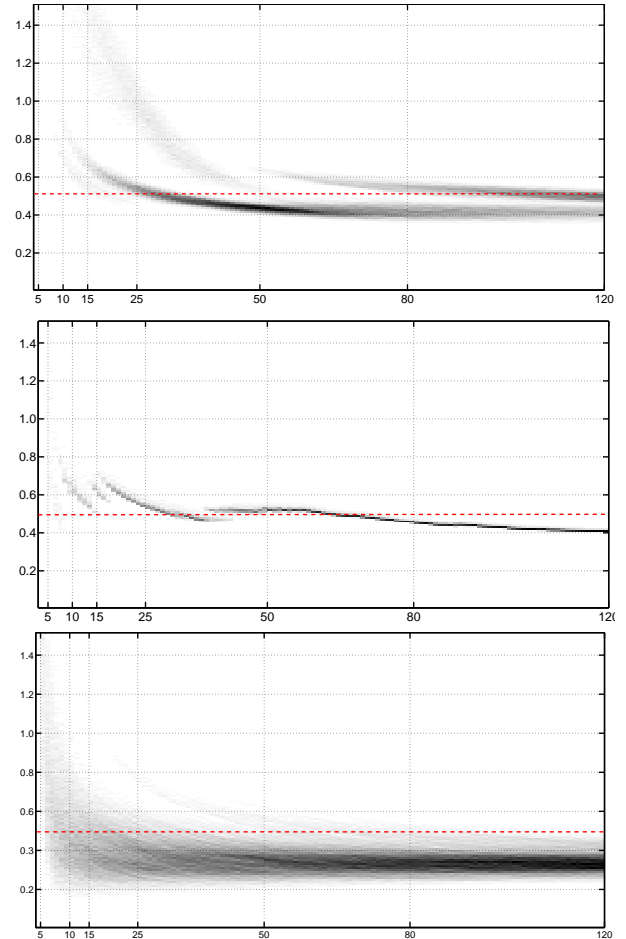


Figure 5. MPE density for real data clustered by temporal availability. Top: random selection; centre: active selection; bottom: adaptive selection.

A more sophisticated correspondence selection strategy is based on error prediction curves. An active selection strategy uses the correspondences processed in the last processing step to calculate prediction errors in the field of view and select correspondences from image regions where the predicted errors are large. Fig. 5 (centre) shows the density of the MPE values for this strategy on the same data. The convergence behavior is slower than for uniform random selection on synthetic data, but it is more stable and faster for real data, reaching the same threshold after about 70 correspondences. Notably, there was no attempt to detect or correct cases starting close to degeneracy, causing high variability in the left region. Nevertheless, the active

selection strategy remains stable.

Finally, we propose a random strategy with an adaptive probability of selection based on the prediction band values. Each new available correspondence is selected or discarded based on an individual random choice. The probability of selection is governed by the value of the prediction band in that area, corresponding to 1 for values at the maximum and 0 for those close to the minimum. Hence, correspondences in areas with larger prediction errors are more likely to be selected. When the prediction error is more or less homogeneous, this selection approach becomes purely random. Results are shown in Fig. 5 (bottom). This strategy converges faster and more steadily than pure random selection, it is still system independent and alleviates the problem of multi-modality.

6. Summary, Conclusions, and Future Work

In this study we have used statistical methods for assessing the quality of camera self-calibration of a stereo vision system, considering accuracy, reliability, and significance of the estimated parameters. The results of the optimisation process are subject to model fitting analysis. In this context, we have defined meaningful figures such as significance, confidence levels, and error prediction curves across the complete field of view. Furthermore, we have analysed the effects of the number and the distribution of correspondences on the calibration quality and examine the efficiency of random selection strategies to reach a predefined accuracy level with as few correspondences as possible. We have introduced a strategy for active selection of correspondences based on the prediction error, which achieves a similar level of calibration quality as random selection but without uniformity requirements. The active selection strategy has been combined with random selection, relying on an adaptive selection probability determined based on the prediction band values, which leads to a faster and steadier convergence than purely random selection. The evaluation has been performed using both synthetic data and real-world data from a vehicle-based stereo vision system. Our empirical analysis has improved our understanding of camera calibration procedures and helped to find a good trade-off between accuracy and computational efficiency by specifically selecting the correspondences which are most relevant for the calibration process.

Future work will address the application of the presented methods to multiple camera computer vision systems, to instrumented calibration, and to cross-calibration of cameras based on ground truth e.g. from RADAR or LIDAR sensors. A further extension will allow the computation of 3D error prediction volumes or their corresponding iso-surfaces for evaluation of the model fit over the complete observed three-dimensional volume.

References

- [1] J.-Y. Bouguet. Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc, 2008. 5
- [2] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley-Interscience, 1998. 2
- [3] C. Engels and D. Nister. Global uncertainty in epipolar geometry via fully and partially data-driven sampling. In *ISPRS BenCOS Workshop*, 2005. 3, 4
- [4] O. Faugeras and Q.-T. Luong. *The Geometry of Multiple Images*. MIT Press, 2001. 2
- [5] U. Franke, C. Rabe, H. Badino, and S. Gehrig. 6d-vision: Fusion of stereo and motion for robust environment perception. In *Proc. DAGM*, 2005. 1
- [6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004. 2, 4, 5
- [7] H. Hirschmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proc. CVPR*, volume 2, pages 807–814, 2005. 1
- [8] O. J. W. F. Kardaun. *Classical Methods of Statistics*. Springer-Verlag, Berlin, 2005. 3
- [9] T. Luhmann, S. Robson, C. Reeves, P. Wainwright, and S. Kyle. *Close Range Photogrammetry: Principles, Methods and Applications*. Whittles Publishing, 2006. 1
- [10] K. Madsen, H. B. Nielsen, and O. Tingleff. *Methods for Non-Linear Least Squares Problems*. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, 2004. 1, 4, 5
- [11] M. Pollefeys and L. V. Gool. Stratified self-calibration with the modulus constraint. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21:707–724, 1999. 4
- [12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992. 2, 3
- [13] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. of Computer Vision*, 47(1-3):7–42, 2002. 2
- [14] F. Stein. Efficient computation of optical flow using the census transform. In *Proc. DAGM*, pages 79–86, 2004. 5
- [15] C. Strecha, W. von Hansen, L. J. van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for highresolution imagery. In *Proc. CVPR*, 2008. 2
- [16] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical Report CMU-CS-TR-91-132, Carnegie Mellon University, 1991. 5
- [17] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment – a modern synthesis. In *Proc. ICCV, Intl. Workshop on Vision Algorithms*, volume 1883, pages 298–372, 2000. 2
- [18] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, volume 14, pages 965–980, 1992. 4