ROBUST CAMERA CALIBRATION
A Generic, Optimization-based Approach

Stephan Rupp and Matthias Elter
Fraunhofer Institute for Integrated Circuits (IIS), Am Wolfsmantel 33, 91058 Erlangen, Germany

Keywords: Robustness, Camera Calibration, Optimization, Genetic Algorithm, Heuristics.

Abstract: The estimation of camera parameters is a fundamental step for many image guided applications in the industrial and medical field, especially when the extraction of 3d information from 2d intensity images is in the focus of a particular application. Usually, the estimation process is called camera calibration and it is performed by taking images of a special calibration object. From these shots the image coordinates of the projected calibration marks are extracted and the mapping from the 3d world coordinates to the 2d image coordinates is calculated. To attain a well-suited mapping, the calibration images must suffice certain constraints in order to ensure that the underlying mathmatical algorithms are well-posed. Thus, the quality of the estimation severly depends on the choice of the input images. In this paper we propose a generic calibration framework that is robust against ill-posed images as it determines the subset of images yielding the optimal model fit error with respect to a certain quality measure.

1 INTRODUCTION

Camera calibration is an indispensable step for augmented reality or image guided applications where quantitative information should be derived from images. Usually, a camera calibration is obtained by taking images of a special calibration object and extracting the image coordinates of projected calibration marks enabling the calculation of the projection from the 3d world coordinates to the 2d image coordinates. To attain this, the calibration images must suffice certain constraints in order to ensure that the underlying mathmatical algorithms are well-posed. In the literature, ill-posed setups are often referred to as *singularities* or *degenerated configurations* (Sturm and Maybank, 1999; Zhang, 2000). Unfortunately, in everyday calibration work, some of the acquired images yield significant calibration errors or even originate from such ill-posed configurations and their determination is rarely obvious. Hence, a mechanism that automatically identifies such images is desirable or at least a calibration method that is robust with respect to these configurations.

In our contribution, we address this problem and propose a generic calibration framework, that is robust against ill-posed configurations because it automatically chooses images that result in low calibration errors. The framework is generic in the sense that it is independent of a certain calibration technique since it is parameterized by the applied calibration algorithm.

2 RELATED WORK AND CONTRIBUTION

Camera calibration has been studied intensively in the past years, starting in the photogrammetry community (McGlone and Mikhail, 1940) and more recently in computer vision (Tsai, 1987; Sturm and Maybank, 1999; Zhang, 2000; Heikkilä and Silvén, 2000). According to Heikkilä and Silvén (Heikkilä and Silvén, 1997), there are four main problems when designing a calibration procedure: control point localization in the images, camera model fitting, image correction for radial and tangential distortion and estimating the errors originated in these stages. Most of the research has been devoted to model fitting and only few works can be found in literature about the other stages of the pro-

Rupp S. and Elter M. (2007).
ROBUST CAMERA CALIBRATION - A Generic, Optimization-based Approach.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IFP/IA*, pages 61-68

cess such as feature point localization, cf. (Mateos, 2000). In addition to this, Ouellet et. al. (Ouellet and Hébert, 2004) propose an interactive approach in order to predict improper calibration images that feature blurred, circular calibration marks. They analyze the images' marks with an acutance-based quality measure (Rangayyan et al., 1997) in order to quickly decline images that suffer from static or motion blur. This in combination with an interactive assistant tool for geometric camera calibration eliminates the need to carefully examine each of the images and thus facilitates the calibration process.

Concerning the identification of degenerated configurations, literature neglects the problem of an automatic image selection, that determines the images that are likely to result in small model fit errors. However, this is an important topic since ill-posed configurations can negatively influence the over-all calibration procedure and thus lead to significant errors as much as poor image quality does.

Hence, we propose a generic and extensible calibration framework that is robust against singularities. The framework is based on discrete optimization technique and makes use of local search methods in order to reject image sets that are likely to contain images from degenerated configurations. Due to the fact that local search methods require an initial start solution, we analyze the use of Genetic Algorithms in comparison with the a our random sampling method suggested in (Rupp et al., 2006). In addition, the framework is extensible, so that i.e. the accutance-based blur detection of Ouellet et. al. can be easily integrated in order to automatically exclude images from calibration that feature blurred calibration marks and thus improve the robustness of camera calibration against poor image quality and singularities.

3 METHODS

In general, calibration is the problem of estimating values for the unknown parameters in a sensor model in order to determine the exact mapping between sensor input and output.

The calibration of a imaging device is usually performed by observing a special calibration object, which is in most cases a flat plate with a regular pattern marked on it using colors causing a high contrast between the marks and the background. The pattern is chosen such that the image coordinates of the projected reference points can be measured with high accuracy. Once the relationship between the 2d image coordinates and 3d world coordinates is known, the transformation $\mathcal{C}$ of the visual system can be estimated. In practice, a set of n observations (input images) $\mathsf{I} = \{\iota_1, \iota_2, \ldots, \iota_n\}$ is considered whereas some of the acquired images may originate from ill-posed configurations. Typically, singularities are seldomly known beforehand, so that neither considering all the n images nor a human-made subset selection will in general yield the *optimal* calibration result - particularly for non-expert users.

We present a robust calibration framework that applies optimization techniques in order to automatically determine the optimal subset out of the pool of aquired images yielding the best calibration result with respect to a quality measure.

3.1 Optimization Terminology

The term *optimization* refers to the study of problems in which one seeks to minimize or maximize a real function $\phi : \mathsf{X} \to \mathsf{R}$ by systematically choosing the values of the variables from within an allowed set X. Typically, the function ϕ is called *objective function*, its domain X is the *solution space* and an element of X is referred to as *solution* or *state*. The optimization persues minimization or maximization of ϕ that means to identify the *global optimal solution* $\mathbf{x}_{\text{opt}} \in \mathsf{X}$ such that $\phi(\mathbf{x}_{\text{opt}}) < \phi(\mathbf{x})$ (minimization) or $\phi(\mathbf{x}) < \phi(\mathbf{x}_{\text{opt}})$ (maximization), for all $\mathbf{x} \in \mathsf{X}$. Frequently, there are some auxiliary conditions defined on X that reduce the solution space. These *constraints* are usually described by a predicate P defined on X or a set of (in)equalities. The solutions sufficing these constraints are called *feasible solutions* and define the *set of feasible solutions* $\mathsf{Y} \subseteq \mathsf{X}$.

If X is countable and finit, the optimization problem is named *discrete optimization problem*, if additionally $\mathsf{X} \subseteq 2^{\mathsf{G}}$ holds, with G being a certain basic set, the optimization problem is called *combinatorial optimization problem* (Lee, 2004).

3.2 Modelling

The framework makes use of optimization and thus requires a formulation of the image selection task being suitable for the application of optimization techniques. Hence, we assume the n elements of the calibration image set I being (partially) ordered by an arbitrary relation. We identify an element at position i (the i-th image ι_i) with the i-th unit vector

$$\{\iota_i\} \in \mathsf{I} \quad \mapsto \quad \mathbf{e}_i = (\underbrace{0\,0 \ldots \overset{i}{1} \ldots 0\,0}_{n})^T \quad i = 1, \ldots, n,$$

and model a certain subset by the coordinate vector $\mathbf{x} = (x_1 \ldots x_n)^T, x_j \in [0,1]$, for example

$$\mathbf{x} = (0\,1 \ldots 0 \ldots 1)^T = 0\,\mathbf{e}_0 + 1\,\mathbf{e}_1 + \ldots + 0\,\mathbf{e}_k \ldots + 1\,\mathbf{e}_n.$$

Here, $s_j = 1$ denotes the containedness of the j-th image in the corresponding subset. With this modelling, the image selection is equivalent to the combinatorial optimization problem (cf. Sec. 3.1):

$$\mathbf{x}_{\text{opt}} = \text{argmin}\, \phi(\mathbf{x}), \quad (1)$$

the solution space $\mathsf{X} = \{0,1\}^n$ is due to the coordinate vector representation.

Combinatorial optimization are sometimes easy to solve, i.e. they can be solved in polynomial time, but more often - such as in this case - polynomial-time algorithms are not known (Garey and Johnson, 1979) and one usually resorts to heuristics that are not guaranteed to find an optimal solution (Pitsoulis and Resende, 2002), but a so-called sub-optimal solution $\mathbf{x}_{\max} \in \mathsf{Y}$ being close to the optimal solution $\mathbf{x}_{\text{opt}} \in \mathsf{Y} \subseteq \mathsf{X}$ instead.

3.3 Heuristics

Heuristic algorithms are based on searching the local neighbourhood of a current solution for an improvement. Given a current solution $\mathbf{x} \in \mathsf{X}$, the elements of the neighbourhood $\mathcal{N}(\mathbf{x})$ of $\mathbf{x}$ are those solutions that can be obtained by applying an elementary operation to $\mathbf{x}$. Local search methods start from an initial solution $\mathbf{x}_0 \in \mathsf{X}$ and iteratively generate a serie of improving solutions $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$.

At the k-th iteration, $\mathcal{N}(\mathbf{x}_k)$ is searched for an improving solution $\mathbf{x}_{k+1}$ such that $\phi(\mathbf{x}_k) < \phi(\mathbf{x}_{k+1})$ or $\phi(\mathbf{x}_k) > \phi(\mathbf{x}_{k+1})$ respectively. If such a solution is found, it is made the current solution. Otherwise, the search ends with $\mathbf{x}_k$ as a local optimum.

Since heuristic optimization strategies are based on a neighbourhood relation $\mathcal{N}$, two image selections are defined to be neighbours, if they differ by excatly one image:

$$\mathcal{N}(\mathbf{x}) := \{\mathbf{y} \in \mathsf{X} : d(\mathbf{x},\mathbf{y}) = 1\}$$

Thus, the neighbourhood of a solution $\mathbf{x}$ is given by the Hamming distance defined on the search space $\mathsf{X} = \{0,1\}^n$.

3.4 Framework

A configuration of the framework is defined by the tuple $\mathcal{F} = (\mathsf{I}, \alpha, \omega, \phi, \tau)$ comprising the set of input images I, the calibration algorithm α, the quality measure ϕ, the optimization strategy ω and finally a criterium τ that terminates the optimization. In general, the optimization strategy may cover arbitrary techniques, however, due to the the special structure of Eq. (1) the choice of a particular strategy is restricted to the class of local search method O_{LS} described above.

Algorithm 1 Generic robust camera calibration

Require: $\mathsf{I}, \alpha, \phi, \omega \in \mathsf{O}_{\text{LS}}, \omega_{\text{init}}, \tau$
  $\mathbf{x} = \omega_{\text{init}}()$
  $\mathbf{x}_{\max} = \mathbf{x}$
  while $!\tau$ **do**
    $\mathbf{y} = \omega(\mathbf{x}), \quad \mathbf{y} : \mathbf{y} \in \mathcal{N}(\mathbf{x}) \wedge P$
    if neighbour $\mathbf{y}$ exists **then**
      $\mathbf{x} = \mathbf{y}$
      if $\phi(\alpha(\mathbf{x}))$ is better than $\phi(\alpha(\mathbf{x}_{\max}))$ **then**
        $\mathbf{x}_{\max} = \mathbf{x}$
      end if
    end if
  end while
  return $\mathbf{x}_{\max}$

A generic algorithm in pseudo code is depicted in Alg. 1. Due to the limitation on heuristic algortihms the framework features an additional parameter, that is responsible for finding a suitable start solution ω_{init} (initialization). Typically, the termination criterion τ is given by the maximum number of iterations $k_{\max}$, a convergence term measuring the relative improvement or a combination of both.

Before entering the optimization loop, the initialization strategy ω_{init} is applied in order to find a feasible solutions that acts as the start solution and that is made the current (sub)optimal solution $\mathbf{x}_{\max}$. Then, the loop is entered and repeated until the termination criterion is met. For the current solution $\mathbf{x}$, an improving, feasible solution $\mathbf{y}$ from within the current solution's neighbourhood $\mathcal{N}(\mathbf{x})$ is identified by the local search method ω. If such a solution exists, the visual system's parameters are estimated with the calibration algorithm α and the images represented by $\mathbf{y}$. Subsequently, the calibration result is compared with the current optimal solution using the quality measure ϕ.

4 APPLICATION

In the following we demonstrate the use of the framework with common choices for the camera calibration algorithm α, the objective function ϕ and exemplarily a variant of the standard downhill search method ω_{SD}. The described configurations serve simultaneously as setups for the subsequent experiment and result section.

4.1 Initialization

As mentioned above (Sec. 3), heuristic optimization starts from a given intial solution $\mathbf{x}_0$. Due to the

bration with the images $\mathbf{x}$:

$$\phi_{\text{BPE2d}}(\mathbf{P}) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \underbrace{\left\| \begin{pmatrix} u_{ij} \\ v_{ij} \\ 1 \end{pmatrix} - \mathbf{P} \cdot \begin{pmatrix} X_{ij} \\ Y_{ij} \\ Z_{ij} \\ 1 \end{pmatrix} \right\|_2}_{\varepsilon_{ij}}$$

The projection error of a single calibration feature ε_{ij} is given by the Euclidean distance between its initially extracted image coordinates $(u_{ij}, v_{ij})'$ and the corresponding 3d world coordinates $(X_{ij}, Y_{ij}, Z_{ij})'$ being projected to the image plane with the projection matrix $\mathbf{P}$ acquired by the calibration procedure.

In order to improve the calibration process, we propose to combine the previous error measure ϕ_{BPE2d} with a term that assesses the spatial error when reconstructing 3d points from 2d image point correspondences by means of the calibrated projection matrix $\mathbf{P}$. For this, we incorporate the regression error $\varepsilon_{\text{PlaneFitError}}$ with respect to a plane that has been fitted into the intersection points of back projected rays by means of the projection matrix $\mathbf{P}$:

$$\phi_{\text{BPE}} := \phi_{\text{BPE2d}} + \varepsilon_{\text{PlaneFitError}} \qquad (2)$$

Due to the fact that this error function ϕ_{BPE} calculates the projection errors as well as the out-of-plane errors for all the images of the initial calibration image set I, it can be used as an indicator for singularities. The smaller the value of ϕ_{BPE} is, the better the calibrated parameters fit the model. Thus, the optimization persues minimization of $\phi_{\text{BPE}} : \{0,1\}^n \to \mathsf{R}_0^+$ in order to identify the best image subset, whereas a huge value of ϕ_{BPE} indicates the containdeness of a singularities within the image subset.

4.4 Optimization Strategy

As a representant of the vast number of local search algorithms, we exemplarily consider a variant of the simple and common downhill heuristic. Both algorithms make use of a conceptual skier that constantly moves downhill in the value landscape. For this, the basic version just seeks for a neighbour with an equal or better solution. Thus, it chooses a deterministically or stochastically determined neighbour $\mathbf{x}_{k+1} \in \mathcal{N}(\mathbf{x}_k)$ that yields a smaller back-projection error than the current solution $\mathbf{x}_k$:

$$\phi_{\text{BPE}}(\mathbf{x}_k) < \phi_{\text{BPE}}(\mathbf{x}_{k+1})$$

The *steepest descent* local search method ω_{SD} acts as a stronger formulation as it always consideres the best solution within the neighbourhood. Thus, the algorithm replaces the current feasible solution $\mathbf{x}_k$ with a new solution $\mathbf{x}_{k+1}$ according to:

$$\mathbf{x}_{k+1} = \underset{\mathbf{x}_{k+1} \in \mathcal{N}(\mathbf{x}_k)}{\text{argmax}} \ \phi_{\text{BPE}}(\mathbf{x}_k)$$

with $\phi_{\text{BPE}}(\mathbf{x}_k) < \phi_{\text{BPE}}(\mathbf{x}_{k+1})$.

Table 1: Comparison of human-made selections with the global optimum $\mathbf{x}_{\text{opt}}$, the selection of all images and the proposed methods. The mean projection error is given in pixel and calculated with respect to the whole input image set. The bold values determine the best result within a group.

Method	**Average**	**# Img.**	**Std.Dev.**
Expert 1	0.179088	8	./.
Expert 2	0.180657	6	./.
Expert 3	**0.178398**	10	./.
Expert 4	0.178818	18	./.
Expert 5	0.182151	4	./.
Expert 6	0.178776	11	./.
Expert 7	0.178678	7	./.
Expert 8	0.178643	9	./.
All	**28.2622**	20	./.
$\omega_{\text{init,MCM}}$	0.178386	./.	1.10e-5
$\omega_{\text{init,MCM}} + \omega_{\text{SD}}$	**0.178376**	./.	1.64e-5
$\omega_{\text{init,GA}}$	0.178410	./.	2.72e-5
$\omega_{\text{init,GA}} + \omega_{\text{SD}}$	**0.178379**	./.	2.44e-5
$\mathbf{x}_{\text{opt}}$	**0.178320**	11	./.

5 EXPERIMENTS AND RESULTS

For an evaluation of our approach, we asked different persons with a background in computer vision to calibrate cameras and compared their calibration results with those that have been obtained with the proposed methods. The experts calibrated several cameras of different resolution and manufacturers, each from $n = 20$ images of a 14-by-10 checkerboard (with $m = 117$ calibration marks) whereas some of them originated from ill-posed configurations. In order to compare the expert's over-all performance, the globally optimal solution for all image subsets was determined too. For this, an exhaustive search of the search space has been performed and the minimum projection errors for the configurations that comprise of two images, three images and so on up to 20 images have been identified. Starting with configurations of only two images is due to the Zhang method that requires at least two different views of the calibration pattern (see Sec. 4.2). In contrast to taking the minimum number of images, considering all images corresponds to the procedure typically persued in everday calibration work. The exhaustive search procedure has found the global optimum of 0.178320

the global optimium as well as with human-made decisions exhibit that the calibration can be significantly improved by automated image selection. Even though one of the image sets manually selected by one of the experts performed almost equally well as the proposed automatic methods, this is not the case in general.

Considering the selection algorithms, the Monte Carlo initialization method $\omega_{\text{init,MCM}}$ performs slightly better than the genetic algorithm $\omega_{\text{init,GA}}$. Similarly, the heuristic optimization results in lower backprojection errors with the stochastic initialization but at the expense of longer responses. Regardless of which initialization method is used and no matter if it is followed by a heuristic optimization, the calibration result is robust with respect to outlier images, i.e. images that were taken from ill-posed views. Hence, selecting good image sets for camera calibration no longer requires long lasting experience or time-consuming trial and error.

REFERENCES

Fischler, M. A. and Bolles, R. C. (1981). Random sampling consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381–395.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability - A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company.

Goldberg, D. E. (1989). *Genetic Algorithms in Search and Optimization*. Addison-Wesley.

Heikkilä, J. and Silvén, O. (1997). A four-step camera calibration procedure with implicit image correction. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1106–1112.

Heikkilä, J. and Silvén, O. (2000). Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1066–1077.

Lee, J. (2004). *A First Course In Combinatorial Optimization*. Cambridge Press.

Mateos, G. G. (2000). A camera calibration technique using targets of circular features. *5th Ibero-America Symposium On Pattern Recognition (SIARP)*.

McGlone, C. and Mikhail, E. (1940). *Manual of Photogrammetry*. ASPRS, 1 edition.

Ouellet, J.-N. and Hébert, P. (2004). Developing assistant tools for geometric camera calibration: Assessing the quality of input images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 80–83.

Pitsoulis, L. S. and Resende, M. G. C. (2002). Greedy Randomized Adaptive Search Procedures. In Pardalos, P. and Resende, M., editors, *Handbook of Applied Optimization*, pages 178–183. Oxford University Press.

Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. E. L., and Alim, O. A. (1997). Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on Medical Imaging*, 16(6):799–810.

Rupp, S., Elter, M., Breitung, M., Zink, W., and Küblbeck, C. (2006). Robust Camera Calibration using Discrete Optimization. *Enformatika Transactions on Engineering, Computing and Science*, 13:250 – 254.

Sturm, P. and Maybank, S. (1999). On plane-based camera calibration: A general algorithm, singularities, applications. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 432–437.

Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuaricy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Transactions on Robotics and Automation*, 4:323–344.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334.