An empirical evaluation of factors influencing camera calibration accuracy using three publicly available techniques

Wei Sun · Jeremy R. Cooperstock

Abstract This paper presents an empirical study investigating the effects of training data quantity, measurement error, pixel coordinate noise, and the choice of camera model, on camera calibration accuracy, based on three publicly available techniques, developed by Tsai, Heikkilä and Zhang. Results are first provided for a simulated camera system and then verified through carefully controlled experiments using real-world measurements. Our aims are to provide suggestions to researchers who require an immediate solution for camera calibration and a warning of the practical difficulties one may encounter in a real environment. In addition, we offer some insight into the factors to keep in mind when selecting a calibration technique. Finally, we hope that this paper can serve as an introduction to the field for those newly embarking upon calibration-related research.

Keywords Camera calibration · Camera parameters · Coordinate transformations · Distortion models · Accuracy evaluation

1 Introduction

The decreasing cost of computers and cameras has brought rapid growth in stereo-based applications. In consequence, an ever-increasing population of researchers are coming to depend on camera calibration for their projects.

Although camera calibration is a mathematically well-defined problem and numerous methods have been developed to provide solutions, these methods make different simplifying assumptions as to how they should be used in practice and which variables need to be measured. It is also not clear how certain factors such as training data quantity, measurement error, and the choice of camera model influence calibration accuracy. The effect of noise has been studied by Lavest et al. [1] and Zhang [2]; however, their experiments involved a small volume of 3D space and did not use separate test data to verify the scalability of calibrated results. Moreover, a major practical concern is the degree of effort involved in providing a given camera calibration algorithm with the training data it needs to achieve some required accuracy.

Using three publicly available techniques, through extensive experimentation with separate training and test data, we conducted an empirical study of the impact of noise, either in world or pixel coordinates, and training data quantity, on calibration accuracy. We also included a detailed comparison of various models to determine the relative importance of different distortion components, whether the addition of higher-order radial terms or decentering distortions actually improves calibration accuracy, and if so, to what degree. Our contributions are first, to provide a quick answer to researchers who need an immediate camera calibration solution; and second, to give some insight into practical difficulties one may encounter in a real environment, and which factors to keep in mind when looking for a suitable calibration technique. Finally, we hope to offer an introduction to the field for those newly embarking upon calibration related research.

The calibration methods chosen for experimentation were developed by Tsai [3], Heikkilä [4] and Zhang [5]. An important reason for this choice is that the source code for the three methods is publicly available, well developed, well tested, and therefore provides a fair comparison. In fact, only open source algorithms were serious candidates for a study of this purpose, because they are open to study and readily integrated into a larger system.

Tsai’s method represents a conventional approach that is based on the radial alignment constraint (RAC) and requires accurate 3D coordinate measurement with respect to a fixed reference. Among the conventional calibration methods surveyed by Salvi et al. [6], including those developed by Hall et al. [7], Faugeras–Toscani [8], and Weng et al. [9], Tsai’s was reported to exhibit the best performance. His method

W. Sun (✉) · J. R. Cooperstock
Department of Electrical and Computer Engineering, Centre for Intelligent Machines, McGill University, 3480 University Street, Montreal, Quebec H3A 2A7, Canada
E-mail: {wsun, jer}@cim.mcgill.ca
Fax: +1-514-398-7348
has been widely used in multi-camera applications [10, 11]. Heikkilä’s method, also world-reference based, although not included in that survey, employs the more general direct linear transformation (DLT) technique by making use of the prior knowledge of intrinsic parameters. It also involves a more complete camera model for lens distortions. Zhang’s method is a different special case of Heikkilä’s formulation. It combines the benefits of world-reference based auto-calibration approaches, which enables the linear estimation of all supposedly constant intrinsic parameters. This method is flexible in that either the camera or the planar pattern can be moved freely and the calibration procedure is easily repeatable without redoing any measurements. These three methods of course form part of a large space of potential algorithms and can each be generalized in a variety of ways. We have applied the researchers’ names to particular forms of their algorithms as embodied in published code.

2 Calibration methods

This section first provides a brief review of the existing calibration literature, which focuses mainly on camera modeling and calibration algorithm development, and then explains the mathematical details of the methods we investigate, followed by our evaluation criteria.

2.1 A brief literature review

Camera calibration has received increased attention in the computer vision community during the past 2 decades [6, 12]. According to the nature of training data, existing calibration methods can be classified as coplanar or non-coplanar. The coplanar approaches perform calibration on data points limited to a planar surface of a single depth. These methods are either computationally complex or fail to provide solutions for certain camera parameters, such as the image center, the scale factor, or lens distortion coefficients [13]. In contrast, the non-coplanar approaches, to which our study is confined, use training points scattered in 3D space to cover multiple depths, and do not exhibit such problems.

Non-coplanar approaches fall into a number of categories. World-reference based calibration is a conventional approach requiring the 3D world coordinates and corresponding 2D image coordinates of a set of feature points [3, 4, 6, 8, 9]. The disadvantage of this approach is that either a well-engineered calibration object is required, or the environment has to be carefully set up to achieve accurate 3D measurements. Geometric-invariant-based methods use parallel lines and vanishing points as calibration features. Although world coordinate measurement is not required, special equipment may be necessary for measuring certain variables, for example, the ratio of focal lengths [14]. Implicit calibration methods [15] have no explicit camera model and may achieve high accuracy, but are computationally expensive because of the large number of unknown variables involved and they do not reveal the physical parameters of a camera. Auto-calibration or self-calibration approaches determine camera parameters directly from multiple uncalibrated views of a scene by matching corresponding features between views, despite unknown camera motion and even changes in some of the intrinsic parameters [16, 17]. Unfortunately, due to the difficulty of initialization [12], auto-calibration results tend to be unstable [18]. Planar auto-calibration [19] addresses this initialization difficulty by using multiple views of a planar scene, taking advantage of the fact that planes are simple to process and allow reliable and precise feature or intensity-based matching. Sturm–Maybank [20] and Zhang [5] both extend this idea by taking into account the relative geometric information between planar feature points, with Zhang’s method estimating lens distortion coefficients, a factor not included in the former work. According to Sturm and Maybank’s singularity analysis [20], degenerate situations can be easily avoided. Another extension of the multiplanar approach suggests the use of angles and length ratios on a plane but provides no experimental results [21]. It is worth noting that these multiplanar calibration methods differ from the coplanar methods reviewed by Chatterjee and Roychowdhury [13]. The methods described here rely on a planar calibration pattern positioned at various orientations. They exploit the coplanarity constraint on the points in each sample set to reduce or eliminate the need for 3D measurement, but still sample a 3D region. Similarly, a one-dimensional object can be positioned at various spots and in various orientations in a 3D space to provide non-coplanar points for calibration [22].

The choice of camera model [23], varying mainly in its characterization of lens distortion, may also influence calibration results. Tsai used the second order radial distortion model [3] while Zhang employed both the second- and fourth-order terms [5]. Heikkilä included two further decentering distortion components [4], while Lavest et al. even added the sixth-order radial term [1]. Weng et al. introduced a thin prism distortion whose coefficients could be merged with those of the decentering distortion in actual calibration [9]. Most camera models assume zero skewness, i.e., the angle between x and y image axes is 90° [3, 4, 9], but Lavest et al. [1] and Zhang [5] estimate skewness as a variable.

2.2 Calibration methods for experimentation

After reviewing these calibration methods, we chose Tsai [3], Heikkilä [4] and Zhang’s [5] methods for experimentation. The mathematical details are now provided.

2.2.1 Camera model

ten projective model to map 3D scenes to the 2D camera image plane. Despite different formulations for lens distortion, the mapping between world and image points proceeds