Available online at www.sciencedirect.com

ScienceDirect

Pattern Recognition 40 (2007) 2837–2845

PATTERN RECOGNITION

THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/pr

Camera calibration based on receptive fields

Jianbo Su*

Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, PR China

Received 28 November 2005; received in revised form 17 January 2007; accepted 18 January 2007

Abstract

Camera calibration is to identify a model that infers 3-D space measurements from 2-D image observations. In this paper, the nonlinear mapping model of the camera is approximated by a series of linear input–output models defined on a set of local regions called receptive fields. Camera calibration is thus a learning procedure to evolve the size and shape of every receptive field as well as parameters of the associated linear model. Since the learning procedure can also provide an approximation extent measurement for the linear model on each of the receptive fields, calibration model is consequently obtained from a fusion framework integrated with all linear models weighted by their corresponding approximation measurements. Since each camera model is composed of several receptive fields, it is feasible to unitedly calibrate multiple cameras simultaneously. The 3-D measurements of a multi-camera vision system are produced from a weighted regression fusion on all receptive fields of cameras. Thanks to the fusion strategy, the resultant calibration model of a multi-camera system is expected to have higher accuracy than either of them. Moreover, the calibration model is very efficient to be updated whenever one or more cameras in the multi-camera vision system need to be activated or deactivated to adapt to variable sensing requirements at different stages of task fulfillment. Simulation and experiment results illustrate effectiveness and properties of the proposed method. Comparisons with neural network-based calibration method and Tsai's method are also provided to exhibit advantages of the method.

Keywords: Camera calibration; Receptive field; Sensor fusion; Uncertainty; Weighted regression

1. Introduction

Camera calibration is an extensively studied topic in photogrammetry, computer vision and robotics communities. The purpose of it is to establish a mapping between the camera's 2-D image plane and a 3-D world coordinate system so that a measurement of a 3-D point position can be inferred from its projections in cameras' image frames. Camera calibration approaches developed so far can be classified into two categories. The explicit calibration methods develop solutions by analyzing physical model of camera imaging so that calibration is to identify a set of modelling parameters of physical meanings, whereas the implicit calibration methods resort to realizing a nonlinear mapping function that can well describe the input–output relation.

The explicit calibration methods can provide camera's physical parameters, which are important in some applications, such as computer graphics, virtual reality, 3-D reconstruction, etc. A camera model should be adopted before calibration, which consists of an internal model and an external model. They can be calibrated independently or combinatively. Calibration of internal model focuses on whether the model considers image plane distortions [1,2] or not [3–5], and in what way the distortions are modelled [6–8]. On the other hand, calibration is basically performed by observing a reference object whose geometry in 3-D space is precisely known [9]. Therefore, the reference object used in calibration determines calibration algorithms and procedures, which could be a 3-D reference cubic, a 2-D plane [10], a 1-D line [11], or any reference object without specifications [12–14]. Zhang gave a very good summary of calibration methods along this line [11]. Wong took advantage of a revolution object as a reference which is very easy to find [5], while Zhang proposed a flexible calibration technique by observing a

* Tel./fax: +86 21 3420 4276.
E-mail address: jbsu@sjtu.edu.cn.

0031-3203/$30.00 © 2007 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2007.01.024

calibration plane for at least twice [8], which might be the most convenient calibration method with enough accuracy developed so far. In addition, due to an incorporated zoom lens or an unstable working environment, camera parameters may be varying during the working process. Many methods have been developed to address this problem [15,16].

Calibration of the external model is often investigated by being evolved to be a PnP ($n \geqslant 3$) problem [12], i.e., to determine the camera's position and orientation with respect to a scene, a fixed world coordinate system, or another camera system from n correspondence points. Abundant results have been obtained in this category [17–19]. In robotics, external calibration of a camera is also termed as hand-eye calibration and generally formulated to solve the equation $AX = XB$, where X contains parameters of the external model [20–22]. This often occurs when a camera is fixed at a robotic manipulator and the relation of the camera with respect to the host robot is to be recognized.

Generally, the external model of a camera is calibrated after its internal model is obtained, although there was an effort to invert this procedure [14]. However, this two-step method may easily cause error propagation from internal model to external one, and the final performance is limited. Accordingly, more researches focus on integrated methods that can calibrate the internal and external models simultaneously. This is often accomplished by minimizing different kinds of cost functions, emphasizing on back-projection or reconstruction errors of reference points in 3-D space. Solutions have been available from different optimization algorithms, such as gradient descent [23], maximum likelihood [8], internal analysis [3], Kalman filter [24], neural networks [25,26], genetic algorithms [1,4,27], etc. Unfortunately, these methods generally suffer from poor convergence, susceptibility to getting trapped in local extrema, or slow convergence speed that are critical for successful applications.

The explicit calibration methods described above need to adopt models to describe projection imaging process as well as nonlinear distortions of image plane. Despite of extensive efforts, the models are surely not enough to describe nature of a camera, thus inevitably erroneous in formulations. On the other hand, it requires deep understanding of computer vision theory and principles behind those calibration algorithms, which is apparently intricate to average users. As a matter of fact, a large variety of applications actually do not care about details of camera imaging, nor the internal and external parameters of camera(s) involved, e.g., stereovision systems, surveillance, robot navigation, or sensor fusion. In these applications, multiple cameras are often deployed to construct a stereovision [28] or a multi-camera system [23,29] in order to provide measurements for 3-D surroundings. For these vision systems, details of a single camera model are not important and explicit calibration methods appear to be too fragile and expensive. These factors motivate to explore implicit calibration methods, which directly recognize nonlinear relation between 2-D positions of a feature in image planes and its 3-D position in space. There was an effort on implicitly calibrating camera by adopting a set of intermediate parameters of no physical meanings [6]. But most of works along this line are based on neural networks [30–32], which take advantage of capability of neural networks to approximate a continuous function with arbitrary accuracy. The neural network is trained offline, with the image positions of a feature in cameras as inputs and the 3-D coordinate of the feature as output. This method is simple in methodology, thus has presumably been accepted by those who have resort to a stereovision system but are not expertised in computer vision theory.

However, it is often confusing to use neural network for camera calibration in practice since it is not clear what kind of structure of the neural network should be deployed and which learning algorithm is acknowledged to converge to a reasonable performance. It is known that neural networks of improper structures may have over-fitting problem, and its training procedure may be unfavorably time-consuming or even divergent. In addition, since neural network has negative inference problem that makes it weak in generalization ability, it is often difficult to obtain satisfactory accuracy in the areas that are far from image center. In this case, different neural networks might be imposed to map 2-D to 3-D relations for different regions in image planes, which of course increase system complexity. Moreover, if the configuration of a vision system is changed, its neural network-based calibration model should be trained again, even if most of the cameras in the system are the same. This will increase the cost of the system, especially when reconfiguration of the vision system is frequent.

In this paper, we propose a new implicit method for multi-camera calibration based on receptive fields and data fusion strategy [33]. The receptive field weighted regression (RFWR) algorithm was first proposed by Schaal [34] as an incremental learning method that can overcome negative inference and bias-variance dilemma problems, which exist evidently in many learning frameworks including neural networks. A nonlinear function is approximated by a set of linear functions, each of which associates with a weighting function describing its approximation accuracy. The definition domain of a linear function is called a receptive field. The final estimation to the nonlinear function is the regression result of all linear functions' estimations weighted by their respective weighting functions. Accordingly, the nonlinear model of a camera can be learned and realized by RFWR models. The nonlinear mapping between a 3-D position and its 2-D image projections defined by a stereovision system or a multi-camera vision system can be implemented based on the RFWR models of each camera with the help of a weighted average fusion algorithm. The number of receptive fields is evolved automatically according to predefined approximation accuracy. So it is not necessary to determine structure of the calibration model or its initial parameters a prior, which facilitates its utility to a great extent.

The paper is organized as follows: Section 2 presents preliminaries of RFWR algorithm. Section 3 shows how the algorithm is applied to camera calibration problem. Simulation and experiment results of the proposed method are reported in Sections 4 and 5, respectively, followed by conclusions in Section 6.