THE CALIBRATION PROBLEM FOR STEREOSCOPIC VISION

O.D. Faugueras, G. Toscani
INRIA
78153 Le Chesnay (France)

Abstract.

The problem of calibrating a stereo system is extremely important in practical applications. We describe in this paper our approach for coming up with an efficient and accurate solution. We first review the pinhole camera model that is used and analyze its relationship with respect to the internal camera parameters and its position in space. We then study its behavior with respect to changes of coordinate systems.
This yields a constraint which is used in the meansquare solution of the calibration problem that we propose. Since an estimation of the uncertainty is also important, we suggest another solution based on Kalman filtering.
We show a number of experimental results and compare them with those obtained by Tsai [8]. We finish with two practical applications of our calibration technique: reconstructing 3D points and computing the epipolar geometry of a stereo system.

I Introduction.

The problem of calibrating a number of cameras is becoming extremely important in the field of Computer Vision to make 3D Perception possible. For Vision Systems used in industrial applications the ***accuracy*** of 3D reconstruction relies on it.
Furthermore it can prove important [2] for an intelligent vision system to have a measure of the ***uncertainty*** related to the reconstructed 3D points for the definition of a sensing strategy: where to look and how to increase the accuracy on the location of an object. Again this uncertainty depends, among other factors, on the calibration.
Even though this problem has been often considered as minor by many people as compared to the more noble problem of Stereo, or Motion, we think that this attitude is wrong for two reasons.
First, in Stereo, the epipolar constraint (detailed in the last Section of this paper) plays a very important role since it reduces the search for correspondences from a two-dimensional space to a one-dimensional one. Accurate knowledge of the epipolar geometry is therefore a prerequisite of all stereo algorithms and can be achieved only by optical calibration if we want to avoid complicated mechanical setups. Due to the lack of easily usable calibration procedures, this has lead most workers in the field to assume in their algorithms a very special camera geometry yielding a simple epipolar geometry with epipolar lines parallel to the scanlines, hoping either that this configuration could be achieved by careful mechanical alignment (not a very practical assumption), or that a more complicated geometry could be compensated for by performing the so-called epipolar transformation on both images in order to make the epipolar lines parallel to the scanlines. Unfortunately, the epipolar transform can only be derived through the very same calibration procedure which is discussed here, and applying it to the images implies interpolating them on a new grid and is therefore bound to be

i) computationally expensive

and/or

ii) inaccurate.

We show in the last Section that the computation of the epipolar geometry is extremely simple from the results of the calibration and that therefore rectifying the stereo pair should only be necessary when numerical correlation techniques are used in the stereo matching. Moreover, if one wants to use more than

NATO ASI Series, Vol. F52
Sensor Devices and Systems for Robotics
Edited by A. Casals

This work was partially supported by ESPRIT grant P940

two cameras the notion of epipolar transformation vanishes as fog in the sun.

The second reason is that after working for a while with calibration it becomes clear that the problem is fundamentally the same as that of estimating the motion (or more precisely the displacement) of a camera. Indeed, the knowledge of the displacement from the first camera to the second, i.e rotation and direction of translation, plus the knowledge of some parameters internal to each camera (to be defined later as the intrinsic parameters), completely defines the epipolar geometry. Thus, calibration is equivalent to motion analysis or more precisely, since usually the displacement between the two cameras is not small, can be attacked by token matching techniques [1,3]. The results obtained by the corresponding techniques which are not detailed in this paper but in the two previous references seem to indicate the possibility of self-calibration without the use of a any special pattern.

Our main concern is to use off-the-shelf cameras. Standard cameras and lenses have the advantage of being cheap and easily available, and the disadvantage of usually presenting high symmetric and asymmetric lens distortion, poor definition, together with an absence or poor reliability of information concerning intrinsic parameters.

Here, we analyse the pinhole model of camera from two points of view:

i) first, the relationship between the perspective transformation performed by such a model and the physical parameters attached to the camera.

ii) second, the properties of the transformation with respect to changes of coordinates.

In particular, we show that the position and orientation of the camera (extrinsic parameters), and some internal (intrinsic) parameters can be recovered from the perspective transformation matrix while guaranteeing that the computed displacement is indeed a displacement, i.e the rotation matrix is orthogonal which is usually not guaranteed by most of the other techniques.

In order to compute the perspective transformation matrix, it has to satisfy at least one constraint. We show that the most commonly assumed constraint (for its simplicity), yields to a solution which is absurd since the intrinsic parameters depend upon the choice of the world coordinate system. We then propose another constraint, based on the analysis of the properties of the perspective transformation matrix with respect to changes of coordinate systems, which does not suffer from this drawback.

We discuss how the matrix can be estimated both by meansquare and by linear filtering techniques (Kalman Filter) and present a number of ways of measuring the accuracy of the technique and comparing it with the predicted accuracy obtained by the Kalman Filter. One advantage of using the Kalman Filter is that it allows us to incorporate measurement noise at several levels : pixel noise, 3D reference points coordinates noise. We give experimental results and compare them to those obtained by a method reported recently by Tsai [8].

We finish by showing how the results of calibration can be used to reconstruct 3D points and compute the epipolar geometry.

Notations.

Let us define a few notations that will be heavily used in the remaining of this paper:
- bold characters are used for row or column vectors: **l**, **m**.
- outline characters are used for matrices: $\mathbb{M}$, $\mathbb{R}$.
- **lm** denotes the matrix product of row vector **l** by column vector **m**.
- **l** $\times$ **m** stands for the cross product of vectors **l** and **m**.
- $\mathbb{R}$**m** denotes the matrix $\mathbb{R}$ applied to the column vector **m**.

II The camera model.

In this section we describe the camera model which is used in the sequel.

II.1 The perspective transformation model.

We assume that the camera performs a perfect perspective transformation with center O_1 (the camera optic center) at a distance f (the focal length) of the retina plane (figure 1). Let us consider an ideal image plane parallel to the retina plane at a distance 1 from the optic center.