Motion of an uncalibrated stereo rig : self-calibration and metric reconstruction

Zhengyou Zhang, Quang-Tuan Luong, Olivier Faugeras

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Motion of an Uncalibrated Stereo Rig: Self-Calibration and Metric Reconstruction

Zhengyou Zhang

Quang-Tuan Luong

Olivier Faugeras

N° 2079

Octobre 1993

PROGRAMME 4

Robotique,
image
et vision

ISSN 0249-6399

Motion of an Uncalibrated Stereo Rig: Self-Calibration and Metric Reconstruction

Zhengyou Zhang
Quang-Tuan Luong
Olivier Faugeras

Programme 4 — Robotique, image et vision
Projet Robotvis

Rapport de recherche n ° 2079 — Octobre 1993 — 18 pages

Abstract: We address in this paper the problem of self-calibration and metric reconstruction (up to a scale) from one unknown motion of an uncalibrated stereo rig, assuming the coordinates of the principal point of each camera are known (This assumption is not necessary if one more motion is available). The epipolar constraint is first formulated for two uncalibrated images. The problem then becomes one of estimating unknowns such that the discrepancy from the epipolar constraint, in terms of distances between points and their corresponding epipolar lines, is minimized. The initialization of the unknowns is based on the work of Maybank, Luong and Faugeras on self-calibration of a single moving camera, which requires to solve a set of so-called Kruppa equations. Redundancy of the information contained in a sequence of stereo images makes this method more robust than using a sequence of monocular images. Real data have been used to test the proposed method, and the results obtained are quite good.

Key-words: Camera Calibration, Stereovision, Reconstruction, Self-calibration

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

(Résumé : tsvp)

Mouvement d'un système stéréoscopique non-calibré : calibration automatique et reconstruction métrique

Résumé : Cet article décrit une méthode pour la calibration automatique des caméras d'un système stéréoscopique, et la reconstruction métrique (à un facteur d'échelle près) à partir d'un mouvement inconnu de ce système, en supposant que les coordonnées du point principal de chaque caméra sont connues (cette hypothèse n'est pas nécessaire si l'on fait un mouvement supplémentaire). La contrainte épipolaire est d'abord formulée pour deux images non-calibrées. Le problème devient celui de l'estimation des inconnues pour lesquelles l'écart de la contrainte épipolaire, quantifié par les distances des points à leurs lignes épipolaires correspondantes, est minimal. L'initialisation des inconnues est fondée sur les travaux de Maybank, Luong et Faugeras sur l'auto-calibration d'une caméra en mouvement, qui passent par la résolution d'un ensemble d'équations dites de Kruppa. A cause de la redondance de l'information dans une séquence d'images stéréoscopique, cette méthode donne des résultats plus robustes que l'utilisation d'une séquence monoculaire. Des données réelles ont été utilisées pour tester l'algorithme proposé, et de bons résultats ont été obtenus.

Mots-clé : Calibration de caméras, Vision stéréoscopique, Reconstruction, Calibration automatique

Contents

List of Figures

List of Tables

1 Introduction

It is well recognized [15, 17, 18] that stereoscopic cues play an important role in understanding the 3-D environment surrounded us and provide a robust way for 3-D reconstruction. In order for the 3-D reconstruction to be possible, one need to know the relationship between the 3-D world coordinates and their corresponding 2-D image coordinates for each camera, and the relative geometry between the two cameras. This is the purpose of camera calibration. A wealth of work on camera calibration has been carried out by researchers either in Photogrammetry [1, 3] or in Computer Vision and Robotics [25, 7, 12, 26, 23, 28, 29] (see [27] for a recent review). The usual method of calibration is to compute cameras parameters from one or more images of an object of *known size and shape*, for example, a flat plate with a regular pattern marked on it. One problem is that it is impossible to calibrate online, while the cameras are involved in a visual task [16]. Any change of camera calibration occurring during the performance of the task cannot be corrected without interrupting the task. The change may be deliberate, for example the focal length of a camera may be adjusted, or it may be accidental, for example the camera may undergo small mechanical or thermal changes. In many situations such as vision-based planetary exploration, it is not very practical to calibrate cameras with a calibration apparatus. We can either send a calibration apparatus together with the planetary rover and show it each time we need to calibrate the cameras, which is not very realistic, or we can pre-calibrate the stereo system on the ground, which is not reliable.

Recently, a number of researchers in Computer Vision and Robotics are trying to develop online camera calibration techniques, known as self-calibration. The idea is to calibrate a camera by just moving it in the surrounding environment. The motion rigidity provides several constraints on the camera intrinsic parameters. They are more commonly known as the

epipolar constraint, and can be expressed as a 3×3, the so-called *fundamental matrix*. Hartley [8] proposes a singular-value-decomposition method to compute the focal lengths from a pair of images if all other cameras parameters are known. Trivedi [24] tries to determine only the coordinates of the principal point of a camera. The most noticeable work is the theory of self-calibration proposed by Maybank and Faugeras [16]. They show that a camera can be in general completely calibrated from three different displacements. At the same time, they propose an algorithm using tools from algebraic geometry. However, the algorithm is very sensitive to noise, and is of no practical use. Luong, in cooperation with them, has developed a real practical system as long as the points of interest can be located with sub-pixel precision, say 0.2 pixels, in image planes [13, 5].

In this paper, we describe a self-calibration method for a binocular stereo rig from one displacement using a simplified camera model (i.e., the principal points are known), although it can be easily extended to include all parameters if more displacements are available. Because of the exploitation of information redundancy in the stereo system, our approach yields more robust calibration result than only considering a single camera, as to be shown by experiments with real images. Sect. 2 describes the calibration problem to be addressed in this paper. Sect. 3 summarizes the epipolar constraint, which is the fundamental constraint underlying all self-calibration techniques. Sect. 4 deals with the details of the problem solving, including additional constraints present in a stereo rig. As the problem is solved by a nonlinear optimization, an initial estimation of the camera parameters must be supplied, which is described in Sect. 5. Experimental results with real data are provided in Sect. 6.

2 Problem Statement and Notations

2.1 Camera Model

A camera is described by the widely used pinhole model. The coordinates of a 3-D point $M = [x, y, z]^T$ and its retinal image coordinates $\mathbf{m} = [u, v]^T$ are related by

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbb{P} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} ,$$

where s is an arbitrary scale, and $\mathbb{P}$ is a 3×4 matrix, called the perspective projection matrix. Denoting the homogeneous coordinates of a vector $\mathbf{x} = [x, y, \cdots]^T$ by $\tilde{\mathbf{x}}$, i.e., $\tilde{\mathbf{x}} = [x, y, \cdots, 1]^T$, we have $s\tilde{\mathbf{m}} = \mathbb{P}\tilde{M}$.

The basic assumption behind this model is that *the relationship between the world coordinates and the pixel coordinates is linear projective*. This allows us to use the powerful tools of projective geometry, which is emerging as an attractive framework for computer vision [19]. With the state of the art of the technology, camera distortion is reasonably small, and the pinhole model is thus a good approximation.

The matrix $\mathbb{P}$ can be decomposed as

$$\mathbb{P} = \mathbf{A} [\mathbf{R}\ \mathbf{t}] ,$$

where $\mathbf{A}$ is a 3×3 matrix, mapping the normalized image coordinates to the retinal image coordinates, $(\mathbf{R}, \mathbf{t})$ is the displacement (rotation and translation) from the world coordinate system to the camera coordinate system. The most general matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \begin{bmatrix} -fk_u & fk_u \cot\theta & u_0 \\ 0 & -\frac{fk_v}{\sin\theta} & v_0 \\ 0 & 0 & 1 \end{bmatrix} , \tag{1}$$

where

- f is the focal length of the camera,
- k_u and k_v are the horizontal and vertical scale factors, whose inverses characterize the size of the pixel in the world coordinate unit,
- u_0 and v_0 are the coordinates of the principal point of the camera, i.e., the intersection between the optical axis and the image plane, and
- θ is the angle between the retinal axes. This parameter is introduced to account for the fact that the pixel grid may not be exactly orthogonal. In practice it is very close to $\pi/2$.

As is clear, we cannot separate f from k_u and k_v. In the following, we use the following notations: $\alpha_u = -fk_u$ and $\alpha_v = -fk_v$. We thus have five intrinsic parameters for each camera: α_u, α_v, u_0, v_0 and θ.

2.2 Problem Statement

The problem is illustrated in Fig. 1. The left and right images at time t_1 are respectively denoted by I_1 and I_2, and those at time t_2 are denoted by I_3 and I_4. A point $\mathbf{m}$ in the image plane I_i is noted as $\mathbf{m}_i$, and a point M in 3-space expressed in the coordinate system attached to the i-th camera is noted as M_i. The second subscript, if any, will indicate the index of the point in consideration. Thus $\mathbf{m}_{ij}$ is the image point in I_i of the j-th 3-D point, and M_{ij} is the j-th 3-D point expressed in the coordinate system attached to the i-th camera.

Without loss of generality, we choose as the world coordinate system the coordinate system attached to the left camera at t_1. Let $(\mathbf{R}_\mathrm{s}, \mathbf{t}_\mathrm{s})$ be the displacement between the left and right cameras of the stereo rig. Let $(\mathbf{R}_\mathrm{l}, \mathbf{t}_\mathrm{l})$ be the displacement of the stereo rig between t_1 and t_2 with respect to the left camera. Let $(\mathbf{R}_\mathrm{r}, \mathbf{t}_\mathrm{r})$ be the displacement of the stereo rig between t_1 and t_2 with respect to the right camera. Let $\mathbf{A}_\mathrm{l}$ and $\mathbf{A}_\mathrm{r}$ be the intrinsic matrices of the left and right cameras, respectively. The problem can now be stated as: