Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 2716–2729
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Self-calibration of a stereo rig using monocular epipolar geometries

Fadi Dornaika*

Institut Géographique National, Laboratoire MATIS, 94165 Saint Mande, France

Received 21 April 2006; received in revised form 30 October 2006; accepted 3 January 2007

Abstract

This paper addresses the problem of self-calibration from one unknown motion of an uncalibrated stereo rig. Unlike the existing methods for stereo rig self-calibration, which have been focused on applying the autocalibration paradigm using both motion and stereo correspondences, our method does not require the recovery of stereo correspondences. Our method combines purely algebraic constraints with implicit geometric constraints. Assuming that the rotational part of the stereo geometry has two unknown degrees of freedom (i.e., the third dof is roughly known), and that the principle point of each camera is known, we first show that the computation of the intrinsic and extrinsic parameters of the stereo rig can be recovered from the motion correspondences only, i.e., the monocular fundamental matrices. We then provide an initialization procedure for the proposed non-linear method. We provide an extensive performance study for the method in the presence of image noise. In addition, we study some of the aspects related to the 3D motion that govern the accuracy of the proposed self-calibration method. Experiments conducted on synthetic and real data/images demonstrate the effectiveness and efficiency of the proposed method.

Keywords: Self-calibration; Stereo rig; Intrinsic and extrinsic parameters; Cooperative stereo-motion; Epipolar geometry; Correspondence problem; Non-linear optimization

1. Introduction

In the last decade a number of researchers developed self-calibration methods for vision sensors that require no known reference object. Such methods can be used to determine the intrinsic camera parameters, the stereo geometry as well as the 3D shape of the observed scene (see Refs. [1–7], for a single moving camera and Refs. [8–14], for a moving stereo rig). The usefulness of the self-calibration techniques can be tangible in some cases where the sensor parameters are subject to variations and no known reference objects are available (active vision, space robots). In Ref. [10], authors use stereo correspondences across a sequence of stereo pairs. Using different projective reconstructions that are associated with each stereo pair, they propose an algorithm for the recovery of the internal parameters and the 3D Euclidean shape. In Ref. [3], a similar strategy dealing with planar scenes has been developed. In Ref. [13], authors use motion and stereo correspondences across two stereo pairs (one motion of the stereo rig). They propose a method that simultaneously recovers the two internal parameters and the motion of each camera as well as the stereo geometry. The proposed method relies on minimizing the discrepancies between the features and their epipolar lines.

All previous self-calibration techniques for stereo assume that the stereo correspondences are given. In general, solving the correspondence problem between left and right images (the two cameras are termed as the left and the right cameras here) has been proved to be a difficult problem. Several factors make the stereo correspondence problem difficult: occlusions, large disparities, little overlap, photometric and figural distortions. One can notice that significant differences in the cameras' characteristics like the gain and the resolution make the stereo correspondence problem even more challenging. On the other hand, in many cases it is much easier to find motion correspondences than solving for stereo correspondences.

In this paper, we address the following problem. Given two uncalibrated stereo pairs of unknown and arbitrary scenes obtained by a general and unknown motion, we like to recover the intrinsic and extrinsic parameters of the stereo rig without solving the stereo correspondence problem. Unlike the existing

* Correspondence address: IGN, Service de la recherche Laboratoire MATIS, 2 avenue Pasteur, 94165 Saint-Mandé Cedex, France.

E-mail address: fadi.dornaika@ign.fr.

0031-3203/$30.00 © 2007 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2007.01.008

methods for stereo rig self-calibration which consider motion and stereo correspondences as given, our approach uses motion correspondences only. Hence, all problems encountered by stereo correspondence can be avoided.

The organization of this paper is as follows. Section 2 presents some backgrounds as well as the problem we are focusing on. Section 3 presents the recovery of the intrinsic and extrinsic parameters of a stereo rig using monocular epipolar geometries. Section 4 provides an initialization method for the proposed non-linear method. Section 5 provides a performance study of the developed method in the presence of image noise. Section 6 studies some accuracy issues related to the 3D motion of the stereo rig. Section 7 describes experiments with real images. Finally, Section 8 provides some conclusions.

2. Backgrounds and problem formulation

2.1. Backgrounds

This study deals with the estimation of stereo rigs' intrinsic and extrinsic parameters from uncalibrated images. In order to make the paper self-contained, we briefly describe in this section the terminology of the main parameter entities, namely the camera matrix, the rigid displacement, and the fundamental matrix. Interested readers can find more details in Refs. [15,16]. In the sequel, vectors will be denoted by bold lower-case letters and matrices by bold capital letters. The transpose symbol will be denoted by T. For example, $\mathbf{K}^{\mathrm{T}}$ denotes the transpose of the matrix $\mathbf{K}$.

(1) *The camera matrix*: The camera matrix (also known as the calibration matrix) represents the perspective projection of 3D scenes onto the 2D image plane. By setting the world coordinate system to the camera coordinate system and by adopting the pinhole camera model, the camera matrix is represented by a 3×3 matrix, $\mathbf{K}$:

$$\mathbf{K} = \begin{pmatrix} \alpha & s & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{pmatrix},$$

where α is the focal length of the camera expressed in horizontal pixels, β is the focal length expressed in vertical pixels, s is the skew factor and (u_0, v_0) is the principal point—the intersection of the optical axis with the image plane. These five parameters are called the camera intrinsic parameters since they depend only on the electronic and optical characteristics of the camera. In general, the skew factor is set to zero—assuming that the image axes are orthogonal. Thus, the camera matrix is described by four parameters. When the camera is calibrated, i.e., the intrinsic parameters are known (using a calibration algorithm), then the 2D projection onto the image plane (perspective projection) of any given 3D point will be straightforward. Any given 3D point whose 3D coordinates are $(X, Y, Z)^{\mathrm{T}}$ will be projected on a point whose homogeneous image coordinates $(x, y, 1)^{\mathrm{T}}$ are given by $\mathbf{K}\ (X, Y, Z)^{\mathrm{T}}$, i.e.,

$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha & 0 & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix},$$

where λ is a non-zero scalar. In the above equation, the 3D coordinates are expressed in the camera coordinate system. Using non-homogeneous image coordinates, the above matrix equation reduces to the following two equations:

$$x = \alpha \frac{X}{Z} + u_0,$$

$$y = \beta \frac{Y}{Z} + v_0.$$

(2) *The rigid displacement*: The rigid displacement is a change of Euclidean coordinates in 3D space. This transform is composed of a 3D rotation and a 3D translation. The rotation is represented by a 3×3 orthogonal matrix $\mathbf{R}$ and the translation by a 3-vector $\mathbf{t}$. Therefore, a rigid displacement has six degrees of freedom (dof). Any rigid transform can be represented in a matrix form using the following:

$$\mathbf{D} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix}.$$

The use of a 4×4 matrix whose fourth row is (0, 0, 0, 1) is very useful for dealing with homogeneous coordinates as well as for composing several transforms. One can notice that the same 4×4 matrix can be used for representing rigid body motions. In this study, the rigid displacement will be used for representing three entities: (1) the relative pose between the two cameras composing the stereo rig—the stereo rig geometry, (2) the 3D motion of the right camera, and (3) the corresponding left camera 3D motion. In the sequel, these entities will be denoted by $\mathbf{D}_s$, $\mathbf{D}_r$, and $\mathbf{D}_l$, respectively.

(3) *The epipolar geometry and the fundamental matrix*: When the points in space and the two cameras (or two different views captured by the same camera) are in general position, it is not possible to predict the correspondence p' of a point p. Let $\mathbf{p}'$ and $\mathbf{p}$ be their respective homogeneous 2D coordinates—3-vectors. However, p' in the second image is not arbitrary: the corresponding 3D point P has to lie along the optical ray of p, and therefore p' is necessarily located on the projection of that optical ray in the second camera. This 2D line is called the *epipolar line* of the point p in the second image (see Fig. 1). The relationship between the point p and its epipolar line $\mathbf{l}'_p$ in the second image is projective linear. Therefore, there is a 3×3 matrix which describes this correspondence, called the fundamental matrix, giving the epipolar line of the point p: $\mathbf{l}'_p = \mathbf{F}\mathbf{p}$. If two points p and p' are in correspondence, then the point p' belongs to the epipolar line of p, therefore they satisfy the epipolar constraint:

$$\mathbf{p}'^{\mathrm{T}} \mathbf{F} \mathbf{p} = 0.$$