Available online at www.sciencedirect.com

SCIENCE DIRECT

Robotics and Autonomous Systems 50 (2005) 41–49

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Visual correction for mobile robot homing

C. Sagüés*, J.J. Guerrero

Dpto. de Informática e Ingeniería de Sistemas/Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza, C/ María de Luna num. 1, E-50018 Zaragoza, Spain

Accepted 12 August 2004

Abstract

We present a method to send a mobile robot to locations specified by images previously taken from these positions, which sometimes has been referred as homing. Classically this has been carried out using the fundamental matrix, but the fundamental matrix is ill conditioned with planar scenes, which are quite usual in man made environments. Many times in robot homing, small baseline images with high disparity due to rotation are compared, where the fundamental matrix also gives bad results. We use a monocular vision system and we compute motion through an homography obtained from automatically matched lines. In this work we compare the use of the homography and the fundamental matrix and we propose the correction of motion directly from the parameters of the 2D homography, which only needs one calibration parameter. It is shown that it is robust, sufficiently accurate and simple.

Keywords: Mobile robots; Vision; Lines and homographies; Teaching by doing; Robot homing

1. Introduction

Robot navigation normally has involved the use of commands to move the robot to the desired position. These commands include the required position and orientation, which suppose good measurement of the motion made by the robot. However, odometry errors or slipping and mechanical drifts may make the desired position not to be reached. Therefore, the use of an additional perception system is mandatory. Vision is perhaps the most broadly researched perception system.

Using vision, some autonomous vehicles are able to execute tasks based on landmarks which give global localization (Map-Based Navigation). Others carry out specific tasks, building maps of the environment simultaneously (Map-Building-Based Navigation). Our system can be classified in a third group as Map-less navigation [2], because the robot can autonomously navigate without prepared landmarks or complex map-building systems. In our work the target positions are specified with images taken and memorized in the teaching phase. In the playback phase the mo-

* Corresponding author. Tel.: +34 976 761940, fax: +34 976 761914.

E-mail addresses: csagues@unizar.es (C. Sagüés); jguerrer@unizar.es (J.J. Guerrero).

0921-8890/$ – see front matter
doi:10.1016/j.robot.2004.08.005

tion correction to reach to the target position is computed from a projective transformation, which is obtained by processing the current image and the stored reference image. The geometric features extracted and matched are lines which have some advantages with respect to points [3], specially in man made environments.

This kind of navigation using a representation of the route with a sequence of images has been previously considered by a correlation based matching of the current and the reference images [7]. Extracting and matching geometric information from images is not currently time costly and geometric based approaches are less sensitive to noise or illumination changes than others. Of course, the data of the target images to memorize is lower in our proposal, since only extracted lines are stored. Other authors [9] also use vertical lines to correct robot motion, but using a calibrated trinocular vision system.

The recovering of motion from geometric features has been presented using the epipolar geometry [1]. Rotation and direction of translation are initially computed from the essential matrix. In addition, the steps to collision are also computed using a third image. However, there are situations where the fundamental matrix is not meaningful (small translations or planar scenes) and other models are needed to obtain motion [5,8]. We recover motion from homography solving matches of lines automatically. To match them, we use image information and the constraints imposed by the projective transformation. Besides that, robust statistical techniques are considered, which make the complete process useful in real applications. Some experiments show results in typical situations of visual robot homing (Section 6).

2. Motion from two images

In this work, motion information is obtained from the previously stored image and the current image, taken both with the same camera. We first discuss the ways to compute motion from two images of a man made environment, where straight lines and planar surfaces are plentiful.

Two perspective images can be geometrically linked by linear algebraic relations: the fundamental matrix and the homography. An homography relates points or lines in one image belonging to a plane of the scene with points or lines in the other image. On the other hand, fundamental matrix provides a general and compact representation of the geometric relations between two uncalibrated images of a general 3D scene [6]. It provides a mapping from an image point to its epipolar line, but neither a point to point nor a line to line mapping.

Let us suppose two images whose projection matrices in a common reference system are $\mathbf{P}_1 = \mathbf{K}[\mathbf{I}|\mathbf{0}]$ and $\mathbf{P}_2 = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ being $\mathbf{R}$ and $\mathbf{t}$ the camera rotation and translation, respectively, and $\mathbf{K}$ the internal camera calibration matrix.

The homography $\mathbf{H}$ can be related to camera motion [5] as,

$$\mathbf{H} = \mathbf{K}\left(\mathbf{R} - \frac{\mathbf{t}\,\mathbf{n}^{\mathrm{T}}}{d}\right)\mathbf{K}^{-1} \tag{1}$$

being $\mathbf{n}$ the normal to the scene plane and d the plane depth.

The camera motion and the planar structure can be computed from $\mathbf{H}$ when the camera is calibrated [13]. In this case two solutions for motion, with a scale factor for $\mathbf{t}$ and the depth of the plane, are computed.

Camera motion can also be related with the fundamental matrix, which is a 3×3 matrix of rank 2 which encapsulates the epipolar geometry. The fundamental matrix can thus be expressed as $\mathbf{F} = \mathbf{K}^{-T}\,([\mathbf{t}]_\times \mathbf{R})\,\mathbf{K}^{-1}$, being $[\mathbf{t}]_\times$ the antisymmetric matrix obtained from vector $\mathbf{t}$. Given the calibration matrix, the motion can be deduced from $\mathbf{F}$ as follows [5]:

- Compute the essential matrix $\mathbf{E} = \mathbf{K}^{\mathrm{T}}\,\mathbf{F}\,\mathbf{K}$
- Compute the singular value decomposition of matrix $\mathbf{E}$, in such a way that $\mathbf{E} = \mathbf{U}\,\mathrm{diag}(1, 1, 0)\,\mathbf{V}^{\mathrm{T}}$
- The camera translation, up to a scale factor is $\mathbf{t} = \mathbf{U}\,(0, 0, 1)^{\mathrm{T}}$
- The two solutions for the rotation matrix are $\mathbf{R} = \mathbf{U}\,\mathbf{W}\,\mathbf{V}^{\mathrm{T}}$ and $\mathbf{R} = \mathbf{U}\,\mathbf{W}^{\mathrm{T}}\,\mathbf{V}^{\mathrm{T}}$, being $\mathbf{W} = [(0, 1, 0)^{\mathrm{T}}, (-1, 0, 0)^{\mathrm{T}}, (0, 0, 1)^{\mathrm{T}}]$

The homography can be obtained from two images using point or line matches [4]. The fundamental matrix can be computed from corresponding points [14]. It can also be computed from homographies obtained through two or more planes, and in this case, corre-