Available online at www.sciencedirect.com
ScienceDirect
Information Sciences 178 (2008) 137–151
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

Acquisition of translational motion by the parallel trinocular

Chi-Cheng Cheng *, Gwo-Long Lin

Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-Sen University, 70 Lian-Hai Road, 80424, Taiwan, ROC

Received 5 November 2006; received in revised form 23 July 2007; accepted 24 July 2007

Abstract

This study presents a technique for recovering translational motion parameters using a parallel trinocular system and a least squares estimation scheme. The proposed approach overcomes the matrix singularity problem encountered when attempting to recover the motion parameters using a binocular scheme. To further reduce the computational complexity of the motion estimation process, the study also presents a compact closed-form scheme for estimating the translational motion parameters. The closed-form algorithm not only resolves the matrix singularity problem, but also avoids the requirement for matrix manipulation. As a result, it has a low computational complexity and is therefore an ideal solution for performing motion estimation in complex, real-world visual imaging applications. The performance of the closed-form algorithm is evaluated by performing a series of numerical simulations in which translational motions of various magnitudes and in various directions are recovered in both noise-free and perturbed environments. In general, the results demonstrate that the translational motion parameters can be accurately reconstructed provided that the motion in the depth direction is limited to small displacements only. Overall, the simulation results suggest that the parallel trinocular system and the motion parameter estimation scheme presented in this study represent a suitable basis for the development of artificial planar-array compound-like eyes for enhanced performance tracking and imaging applications.

Keywords: Binocular; Optical flow; Trinocular; Visual servo

1. Introduction

Estimating 3D motion parameters from a sequence of images is of crucial importance in many image analysis applications, ranging from robot vehicle guidance to obstacle avoidance, target tracking, the reconstruction of 3D models, and so forth. However, a number of major problems exist in the 3D motion parameter estimation field. Firstly, while monocular observers designed to visualize relative motion within a scene have the benefit of an extremely simple hardware structure, they are unable to recover translational motion parameters or the coordinates of 3D structures with any degree of reliability due to their inherent depth-speed ambiguity [6]. Since it is impossible to determine absolute values of the translation and depth using monocular schemes, 3D interpretation can only be achieved by applying an arbitrary scale factor to the relative

* Corresponding author. Tel.: +886 7 525 4236; fax: +886 7 525 4299.
E-mail addresses: chengcc@mail.nsysu.edu.tw (C.-C. Cheng), d8932808@student.nsysu.edu.tw (G.-L. Lin).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2007.07.021

translational motion and depth values [1]. Secondly, virtually none of the parameter reconstruction techniques presented in the literature provide reliable results when applied to the optical flow fields calculated from realistic scenes due to the difficulties involved in extracting accurate flow fields [6]. Thirdly, most parameter estimation algorithms designed to solve equations of motion are characterized by some form of nonlinearity. In such schemes, generating possible solutions typically involves executing an iteration process starting from some initial guess and has no guarantee of convergence. Furthermore, the presence of nonlinearity prompts the possibility of multiple results [5,41].

Motion estimation has long been a major concern in the image processing field. Seiffert and Michaelis [34] presented a scheme based upon an artificial neural network with 3D self-organizing maps, in which the motion parameters were estimated by making reference to *a priori* information regarding the image sequence history. The use of split-merge displacement estimation strategies [9] or Lie block-based transformation operators [31] has been proposed as a means of improving the motion estimation performance of data compression applications. However, such methods simply approximate the translation by considering the combined motions of individual objects within the video scene. This approach reduces the implementation complexity considerably compared to affine- or perspective-based reconstruction models, but inevitably degrades the estimation performance [37].

In a binocular system, given image pixels corresponding to the same physical point in space, the 3D location of the point can be accurately determined. This property of binocular systems has been exploited to recover 3D motion and depth from known optical flows [22] or to compute 3D information based upon image velocity differences in the left and right flow fields, respectively [32]. Barron and Eagleson [5] exploited the relationship between time-varying optical flows and the corresponding 3D physical structures to recover motion parameters by applying a Kalman filter to integrate the motion and structure parameters over time. Recently, Ogale and Aloimonos [28] applied horizontal and vertical slants to create a first-order approximation to piecewise continuity in order to estimate correspondence and occlusion in the left and right images of a binocular system. However, the approaches outlined in [5,28] involve abstruse mathematical manipulations. Consequently, resolving the correspondence problem in binocular stereo vision systems is a time-consuming process; particularly for complex real-time applications such as mobile robot navigation [8,30], for example. The matching-phase condition exploited in stereo vision techniques is obtained by exploiting a number of fundamental geometrical constraints, including the epipolar constraint, in which it is assumed that the match of a point in the reference image to lie along the epipolar line in the other image. Such constraints provide a rich source of information, which makes possible a faithful reconstruction of the scene. Unfortunately, the geometrical constraints of binocular stereo vision systems are insufficient to enable the construction of a unique solution [17,23]. However, trinocular vision systems enable the correspondence problem to be more easily resolved [15,36] and overcome many of the limitations inherent in binocular stereo matching schemes [2,11,30].

Due to rapid advances in the capabilities of integrated electronic circuits and ever-decreasing costs, CCD cameras have emerged as an attractive solution for image acquisition and processing applications. A typical application is that of multiple-camera visioning configurations such as trinocular vision systems. Such systems can be broadly classified as either right triangular [2,4,29,36], parallel [30], surrounding [23], divergent [33], orthogonal [40], or arbitrary [7,16,17]. The first four types are coplanar, while the remainder are non-coplanar. The right triangular configuration is coplanar in the XY plane, while the parallel, surrounding and divergent schemes are coplanar, as shown in Fig. 1, in the XZ plane. Each configuration is characterized by a unique set of geometrical constraints governed by the particular arrangement of the three CCD cameras.

Ohya et al. [30] investigated the stereo capture of targets using a triangulation technique in binocular systems. However, solving pattern-matching problems was found to be a highly time-consuming procedure. Accordingly, the authors added an additional camera to the vision system and demonstrated that the resulting parallel trinocular scheme enabled a significant reduction in the matching time. The developed visioning system was implemented on a robot, which was then trained using teaching and playback techniques. The results demonstrated that following the training process, the robot was capable of achieving autonomous navigation based on recognizable landmarks. However, the authors neither addressed the behavior of translational motion nor explored the potential sources of navigational error. Li and Duncan [20] used the image flow fields captured by two parallel stereo cameras to determine the 3D translational motion parameters with respect to various objects in the viewing area and to establish the correspondence between equivalent features in the left