Available online at www.sciencedirect.com
SCIENCE DIRECT
Pattern Recognition 39 (2006) 1649–1661
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Absolute motion and structure from stereo image sequences without stereo correspondence and analysis of degenerate cases

Jae-Hean Kim*, Myung Jin Chung

Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, 373-1, Gusong-dong, Yusong-gu, Daejon, 305-701, Republic of Korea

Received 13 August 2005; accepted 5 December 2005

Abstract

This paper deals with the estimation of motion and structure with an absolute scale factor from stereo image sequences without stereo correspondence. We show that the absolute motion and structure can be determined using only motion correspondences. This property is very useful in two aspects: first, motion correspondence is easier to solve than stereo correspondence because sequences of images can be taken at short time intervals; second, it is not necessary that the rigid scene be included in the intersection of the field of view of the two cameras. It is also shown that the degenerate cases reported in this paper constitute all of the degenerate cases for the scheme and can be easily avoided.

Keywords: Structure from motion; Stereo; Motion; Motion correspondence; Stereo correspondence; Real-time vision; Extended Kalman filter

1. Introduction

To estimate camera motion and the structure of a scene simultaneously from an image sequence is currently one of the most exciting areas of computer vision and has been approached in many ways. This research is known as *structure from motion* (SFM). In particular, methods using stereo image sequences offer the advantages of having an absolute scale factor and more accurate results over methods using a monocular image sequence. If the stereo is calibrated, the scene reconstructed from the stereo correspondence has an absolute scale factor because the stereo geometry includes the absolute translational vector.

However, most works combining stereo and motion assume that both the stereo correspondence and the motion correspondence are given or cooperate to help each other [1–3]. This means that these algorithms can acquire more accurate estimates because they exploit information redundancy by using these two correspondences simultaneously. However, stereo-based algorithms suffer from the stereo correspondence problem, which is an ill-posed problem and seriously influences the depth information from the stereo [4]. In particular, there is a strong possibility of a false correspondence when there are inherent ambiguities in a scene, such as a repeated pattern over a large part of the scene, an occluded region that cannot be seen from one of the cameras, and features out of the field of view for one of the cameras.

This paper shows that the absolute scale factor can be determined using only motion correspondence. Moreover, it is shown that the degenerate cases reported in this paper constitute all of the degenerate cases and can be easily avoided. From these results, we can construct an algorithm having two useful properties. First, motion correspondence is easier to solve than stereo correspondence because sequences of images can be taken at short time intervals. There has been a great deal of research on motion correspondence for motion and structure estimation from image sequence [5–9]. Furthermore, because the disparity between two consecutive frames is usually very small, it is possible to monitor the changes of appearance of features during motion correspondence for discriminating and abandoning features

* Corresponding author. Tel.:+82 42 869 8029; fax: +82 42 869 3410.
E-mail address: kjh@cheonji.kaist.ac.kr (J.-H. Kim).

0031-3203/$30.00 © 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2005.12.002

mistracked due to tracking problems such as occlusion and glossy reflection [5,6]. Second, it is not necessary that the rigid scene be included in the intersection of the fields of view of the two cameras. We can reconstruct the scene in the union of the fields of view and acquire a wider range of the field of view.

There are two potential methods for implementing the proposed scheme by sequences of images. One is a two-stage approach. In the first stage, the motions of each camera are estimated up to scale by any SFM algorithm. In the second stage, the scale factors for each camera are determined by using the two motion estimation results. However, this two-stage approach does not give optimal estimates because the motion constraint derived from the stereo geometry is not enforced in the first stage, but enforced separately in the second stage. This is a critical issue when the motion is near the degenerate motion cases or the level of image measurement noise is high. The other is a one-stage approach. This approach minimizes the criterion obtained from the measurement equation that includes the constraint of the stereo geometry. This paper shows that this approach can give more optimal estimates than the two-stage approach. We adopt the *extended Kalman filter* (EKF) to implement the one-stage approach. The EKF has been used in various SFM research for real-time application and shows good performance [10–12].

The work most closely related with that of the present paper was presented by Weng et al. who reported the first attempt to combine stereo and motion without using stereo correspondence [13]. Their study verifies the feasibility of scale factor determinacy without stereo correspondence and explains how to determine the absolute scale factor. However, the optimality of their method can be disputed because it uses a two-stage approach. Moreover, the method was only applied to a two-view framework and has not been extended to the use of sequences of images. Also, they did not extract all degenerate cases and did not consider the degenerate cases in a multi-view framework.

An approach acquiring stereo correspondences from motion correspondences was suggested by Ho et al. [14]. It was demonstrated that the stereo correspondence problem due to the ambiguities could be greatly alleviated by this approach. The approach also has the strengths of using relatively short image sequences and having low time complexity. However, the candidate features must be extracted in advance at both stereo images simultaneously in order to be matched, and a few initial stereo correspondences are needed. Also, an affine camera model is used. Dornika et al. [15] extended this approach to a perspective camera model, which is the most commonly used model. Unfortunately, the algorithm also requires a few initial stereo correspondences. Moreover, given that the stereo correspondences are inferred from the re-projection, Ref. [15] may not be completely free from the ambiguities of the stereo correspondence problem. An algorithm to calibrate the stereo geometry using only motion correspondence was described in Ref. [16]. That method also uses a two-stage approach like [13]. The calibration results are up to scale since only motion correspondence is used. It is worthwhile noting that the degenerate motions proved in this paper are included in the degenerate motions reported in Ref. [16]. The degenerate motions lead to singularities in their solution constraints. However, verification of whether the degenerate cases reported in their work constitute all of the cases was not shown.

2. The scale factor determinacy

In this section, we show that the scale factor can be determined by using only motion correspondence from stereo image sequences if, and only if, the motion is not included in a degenerate motion set. Consequently, we can see that the degenerate cases reported in this paper are all of degenerate cases and can be easily avoided.

Consider the coordinate systems shown in Fig. 1, which shows a schematic diagram of a moving stereo camera. Let $\mathbf{X}^0$, $\mathbf{X}^i$, $\mathbf{M}_1^i$, and $\mathbf{M}_2^i$, for $i = 1, \ldots, N$, be the homogeneous

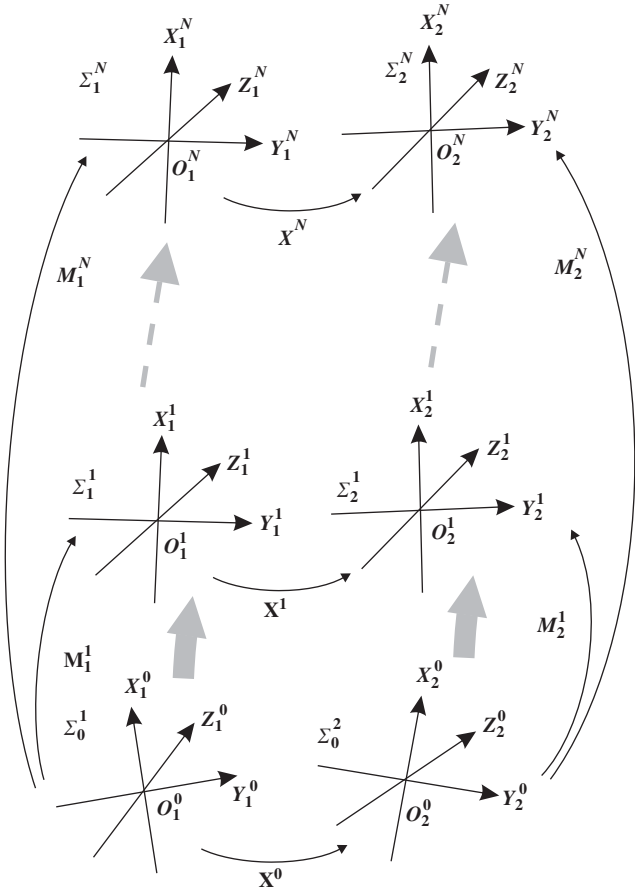

Fig. 1. The schematic diagram of a moving stereo camera. There are homogeneous relationships $\mathbf{M}_1^i\mathbf{X}^i = \mathbf{X}^0\mathbf{M}_2^i$, for $i = 1, \ldots, N$.