Pattern Recognition 43 (2010) 2265–2280

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Recursive estimation of motion and a scene model with a two-camera system of divergent view

Jae-Hean Kim [a,*], Myung Jin Chung [b], Byung Tae Choi [a]

[a] Digital Content Research Division, Electronics and Telecommunications Research Institute, 161, Gajeong-dong, Yuseon-gu, Daejeon 305-700, Republic of Korea
[b] Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, 373-1, Kusong-dong, Yusong-gu, Daejon 305-701, Republic of Korea

ARTICLE INFO

Article history:
Received 7 March 2008
Received in revised form
23 July 2009
Accepted 27 December 2009

Keywords:
Structure from motion
Field of view
Motion
Stereo
Stereo correspondence
Scene reconstruction
Extended Kalman filter

ABSTRACT

This paper deals with recursive reconstruction of a scene model from unknown motion of a two-camera system capturing the images of the scene. Single camera systems with a relatively small field of view have limited accuracy because of the inherent confusion between translation and rotation. Estimation results from the stereo camera systems are also compromised due to this confusion if the systems require the fields of view to intersect for stereo correspondence. The cameras constituting the two-camera system considered in this paper are arranged so that there is a small intersection of the fields of view. This configuration of divergent view improves the accuracy of the structure and motion estimation because the ambiguity mentioned above decreases due to a large field of view. In this paper, a recursive algorithm is proposed for fast scene model reconstruction using a two-camera system of divergent view. Using inversely inferred stereo correspondences in the intersection of the fields of view is also proposed to remove degeneracy of scale factor determination and to acquire more accurate results from the information redundancy. The results of the experiments with long term real image sequences are presented to demonstrate the feasibility of the proposed system.

1. Introduction

There have been many attempts to reconstruct a scene model from unknown motion of a camera. This research is known as *structure from motion* (SFM), which has been an important topic in computer vision and has been approached in a variety of ways. Many SFM algorithms based on image sequences can be categorized into two methods: a batch method and a recursive method. The time-consuming batch methods determine the estimates using all of the image observations simultaneously and produces the most optimal estimates [1–4]. The recursive methods using the *extended Kalman filter* (EKF) have been developed for the requirement of computational efficiency [5–9].

However, most of the work in the SFM has concentrated on single camera systems. The SFM algorithm using a single camera has limited accuracy due to a relatively small field of view because there is inherent confusion between translation and rotation [10]. For example, horizontal translation parallel with an image plane can be easily confounded with rotation around the vertical axis when a field of view is small. This effect is accumulated during long term camera movement and can be a significant factor of motion drift causing inaccuracies of a reconstructed scene model. This is also the case for the stereo camera systems moving in a scene if the systems require the fields of view to intersect largely for stereo correspondence.

To remove this ambiguity, several approaches have been suggested. These approaches can be categorized into three classes. First, a catadioptric vision sensor, which consists of a mirror and a camera, can widen the camera's field of view [11–15]. However, the sensor system has the disadvantage of non-uniform and low angular resolution, which can affect the estimation accuracy. Second, an active vision system that can change the viewing direction and see a wide angular range can be an alternative method [16–18]. The shortcomings of this system are a cumbersome mechanism and a finite time required to capture an image with a wide angular range. Finally, we can adopt a system having multiple cameras [19–25]. The system can acquire a large field of view and alleviate the ambiguity of confusion between translation and rotation, keeping uniform and high image resolution. Since, in these days, fully digital cameras are available inexpensively, which can acquire images with high image resolution and transfer them to PCs at video rates, it is more convenient and useful to install a system that has more than one camera.

In this paper, we propose an algorithm to reconstruct a scene model from unknown motion of a two-camera system capturing long term image sequences, which have point features appearing and disappearing. The two-camera system has divergent view so that there is a small intersection of the fields of view. We found in

* Corresponding author. Tel.: +82 42 860 1053; fax: +82 42 860 1050.
E-mail addresses: gokjh@etri.re.kr, kjh@cheonji.kaist.ac.kr (J.-H. Kim).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.12.015

experiments that using just two cameras can obtain much better accuracy than using a single camera.

A recursive framework using the EKF adapted for a two-camera system is proposed for fast scene model reconstruction and motion estimation. It has been shown that there must be a high degree of correlation between feature estimates to maintain filter consistency [26]. The EKF used in the proposed method satisfies this requirement with stochastic map representations in the covariance matrix. Although the EKF framework contains linearization errors, many works based on this have shown impressive and successful estimation results for motion and structure [7,27]. An alternative is the particle-filter based framework that can cope with non-linearity and multi-modality [28,29]. However, sufficient samples that increase computational complexity are needed to obtain the Bayesian optimal estimates.

The proposed approach can be viewed as an extension of our previous work [30] showing that an absolute scale factor can be acquired while a stereo camera moves in a scene without feature correspondences among the cameras. All of the degenerate motions in determining scale factors were reported in that work. Absolute scale factor is important for removing a scale factor drift in long term image sequences [7]. Although the degenerate motion can be avoided with additional rotational motion, it is not practical to have this type of motion continuously. To overcome this problem, it is also proposed in this paper to use inversely inferred stereo correspondences in the small intersection of the fields of view. These stereo correspondences give more accurate results as well due to the information redundancy. The initial stereo correspondences are robustly extracted by an absolute scale factor obtained from the first few frames of non-degenerate motion. These correspondences contribute to obtaining the next stereo correspondences continuously even under degenerate motion.

In this paper, the perspective camera model is used. This model corresponds to an ideal pinhole camera and is most commonly used model in the field of computer vision. The proposed method assumes that the intrinsic parameters of cameras are given so that the normalized image coordinates of measured feature points can be obtained. It is also assumed that the geometry between the two cameras are given and the scene viewed from the cameras is static.

2. Related works

The motion estimation algorithms for multi-camera systems in [19,22] were designed to obtain differential motion estimates. Although these methods also have their own applications, the results have limited accuracy for long term motion estimates because they only compute the differential values at certain camera position and do not integrate the information acquired from image sequences.

The two-view position framework for linear motion estimation of multi-camera systems was proposed in [23,25] using the generalized epipolar constraints. The linear solutions from these methods can be used as the initial estimates for the batch methods or the recursive methods to improve stability. However, these works did not address the problem of how to extend the framework to multi-view position framework using information from all image frames.

A multi-camera system of single-viewpoint structure was used in [20] to obtain planar motion and structure parameters by using observations in three different positions and non-linear optimization. The single-viewpoint structure means that all optical centers of the cameras coincide at one point and can be constructed by using a multi-mirror and multi-camera so that virtual optical centers coincide. Although this structure is somewhat more difficult to construct than a multi-viewpoint structure, it can make use of the vast array of techniques already developed for a single camera system. However, in the single-viewpoint structure, there is the scale factor ambiguity and the method in [20] also did not address the problem of the extension to the multi-view position framework.

An approach acquiring the camera motion parameters by using a multi-camera system for long term image sequences was suggested by Sato et al. [21]. This method requires some features of known 3D positions. These features are used as initial seed information to compute the camera motion parameters. These motion parameters can give the positions of the other natural features, which are then included in the feature group of known 3D positions. These steps are iterated and the parameters are finally refined with the batch method that is time-consuming.

Recently, for computational efficiency, the recursive framework using the EKF has also attracted attention in the field of vision based *simultaneous localization and mapping* (SLAM) [27,31,32]. The framework is similar to that of the proposed method. However, these works only considered a single camera system.

The methods using a stereo camera system have shown good results for motion and structure estimation, which have an absolute scale factor [33–35]. However, as long as the systems require many stereo correspondences, there is a limit for widening the field of view because features should be in the intersection of the fields of view.

3. Preliminaries

In this section, the models of the EKF are described for a two-camera system and the notation related to the models are introduced. This model was also introduced in [30].

3.1. State of the EKF

Fig. 1 shows a schematic diagram of the two-camera system in the process of observing features in each camera. A sensor coordinate system is attached to the two-camera system. From now on, we assume that one of the camera coordinate systems is selected as the sensor coordinate system and define this camera coordinate system as a *reference camera coordinate system*. A world coordinate system can be defined as a sensor coordinate system at the start position without a loss of generality. The six motion parameters of the sensor are represented with a matrix $\mathbf{R}_g(t)$ for rotation and a vector $\mathbf{t}(t)$ for translation, where $\mathbf{R}_g(t)$ is the rotation matrix of the world coordinate system relative to the current sensor coordinate system and $\mathbf{t}(t)$ is the position of the origin of the world coordinate system relative to the current sensor coordinate system.

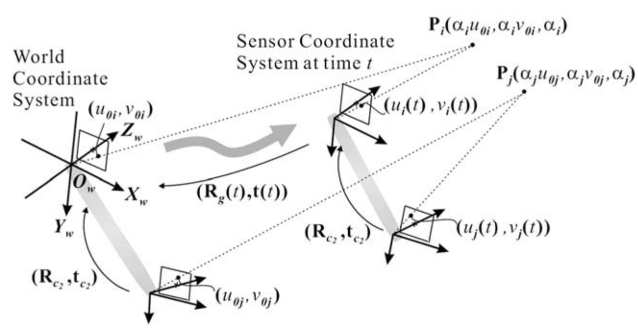

Fig. 1. A schematic diagram of the two-camera system in the process of observing features in each camera.