Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
PHOTOGRAMMETRY & REMOTE SENSING
ISPRS Journal of Photogrammetry & Remote Sensing 61 (2007) 298–306
www.elsevier.com/locate/isprsjprs

Video camera registration using accumulated co-motion maps

Zoltán Szlávik [a,*], Tamás Szirányi [a,*], László Havasi [b]

[a] *Computer and Automation Research Institute of Hungarian Academy of Sciences, P.O. Box 63, H-1518 Budapest, Hungary*
[b] *Péter Pázmány Catholic University, Práter u. 50/a, H-1083 Budapest, Hungary*

Received 3 June 2005; received in revised form 3 August 2006; accepted 19 September 2006
Available online 15 November 2006

Abstract

The paper presents a method to register partially overlapping camera-views of scenes where the objects of interest are in motion even if unstructured environment and motion. In a typical outdoor multi-camera system the observed objects might be very different due to the changes in lighting conditions and different camera positions. Hence, static features such as color, shape, and contours cannot be used for camera registration in these cases. Calculation of co-motion statistics, which is followed by outlier rejection and a nonlinear optimization, does the matching. The described robust algorithm finds point correspondences in two camera views (images) without searching for any objects and without tracking any continuous motion. Real-life outdoor experiments demonstrate the feasibility of our approach.

Keywords: Image registration; Video sequences

1. Introduction

Image registration or matching is a basic task in many computer vision and digital photogrammetry applications. It is an important subtask of image based 3D reconstruction (Pollefeys et al., 2000), camera-view registration and calibration in multicamera systems (Szlávik et al., 2004a,b; Faugeras et al., 1992; Maas, 1999). View registration methods can be divided into two main classes, i.e. still-image based algorithms and image sequence based algorithms.

Still-image registration algorithms (Zhang et al., 1995; Barnard and Thompson, 1980; Cheng and Huang, 1984; Weng et al., 1992) search for static features in images such as: edges, corners, contours, color, shape etc. They are usable for image pairs with small differences; however, they may fail in case of occlusions, featureless regions or if the chosen features cannot be reliably detected. Due to the lighting changes, occlusions and view changes that occur in different views the images of the scene from various cameras may be very different, so we cannot base the decision solely on the still features.

In order to overcome the above-described problems image-sequence based methods were proposed. Video sequences, in fact, contain much more information than the static scene structure of any individual frame does; in particular, in the time-dimension information is also captured about the scene dynamics which is an inherent property of the scene. They are common to all video sequences recorded from the same scene, even if taken from different camera positions, or with different zoom-lens settings. References (Lee et al., 2000; Caspi et al.,

* Corresponding authors.
E-mail addresses: szlavik@sztaki.hu (Z. Szlávik), sziranyi@sztaki.hu (T. Szirányi).

0924-2716/$ - see front matter © 2006 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2006.09.014

2006) present approaches in which motion-tracks of the observed objects are aligned. However, in these cases a robust capability for object tracking is assumed; this is the weak point of the methods, since the tracking could fail in case of cluttered environments or unpredictable motions.

The main aim of the present paper is to propose an algorithm for the registration of different overlapping camera-views that is based only on the analysis of the detected changes, which assists to overcome the limitations of the above mentioned methods. It does not require any *a priori* knowledge about objects' motion, structure or appearance in the scene. The approach we propose in the paper is an extension of previously published sequence-based image matching methods for non-structured estimation. It aims to use statistics of concurrent motions – the so-called co-motion statistics (Szlávik et al., 2004a,b) – instead of trajectories of moving objects to find matching points in view pairs. Our previous conference articles (Szlávik et al., 2004a,b) introduced the basic idea of co-motion statistics with some preliminary results. The present paper contains a more detailed description and numerical analysis of the use of co-motion statistics for camera registration. Here, we will also show that it can be used in challenging circumstances, when the observed motions are hardly to be tracked.

2. Registration of views

An algorithm for the registration of views usually has the following steps:

1. Feature detection;
2. Extraction of candidate point-pairs;
3. Rejection of outliers and estimation of the model that does the matching.

If the observed changes are in the common view of both cameras then these changes are to be concurrently detected in cameras' views. Hence, the basic features of our algorithm are those pixels from cameras' views (images) where significant changes were detected and which are concurrently changing. They will serve as corresponding points for registration.

Here we also assume multi-camera systems containing a series of cameras with partial overlapping among the neighboring views, resulting in separately paired views. Basically, the input of our system is a pair of videos taken from the same scene and the algorithm registers those overlapping image regions where motion was observed. Thus, here we consider two cameras at arbitrary positions. We assume that we have standard cameras having no lens distortions, the cameras are fixed and their views are overlapping. However, possible lens distortions produced in cameras with wide-angle lenses can be corrected by using well-known methods described in Chapter 7.4 of Hartley and Zisserman (2003).

Generally, we do not assume object structure and we have no assumption about the motion of observed objects. In our algorithm we assume that objects are moving either on a plane or in 3D (this is the most general case). In most of the surveillance and traffic control applications multi camera systems observe motions on the groundplane. In this case the registration of views can be modeled by the homography between them (Hartley and Zisserman, 2003). In case of 3D motions the fundamental matrix and the epipolar geometry expresses the geometric relation between the images (Hartley and Zisserman, 2003). These geometric models provide additional constraints on point correspondences; hence more precise registration can be estimated.

Our algorithm is based on the analysis of the recorded scene dynamics and the features are the maxima of the so-called co-motion statistics, which represents the concurrent motion with respect to a given pixel in the image. These maxima form the set of candidate point-pairs. Due to the errors in feature detection, outliers – point-pairs that do not correspond to the same point in the 3D space – will be always present in the set of candidate point-pairs. It may come from the erroneous change detection and differences in projected shapes in different views. Therefore we have implemented a robust outlier rejection algorithm.

According to the above general schema for view registration the proposed algorithm has the following steps:

1. Feature detection and extraction of candidate point-pairs:
   a. Change detection
   b. Computation of co-motion statistics
   c. Extraction of maximums of co-motion statistics
2. Rejection of outliers and estimation of the model that does the matching:
   a. Rejection of outliers by introducing some neighborhood rules
   b. Rejection of outliers by estimating the geometric model that relates the input images

The major assumption is the time synchronization between the cameras, which can be done by using standard algorithms (e.g. Berkeley algorithm Tanenbaum and van Steen, 2001). When it exists, the motion information can be transformed into motion-statistics. Later we will show that this assumption by using further processing can be avoided.