Image and Vision Computing 50 (2016) 27–41

Contents lists available at ScienceDirect

Image and Vision Computing

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

Degenerate motions in multicamera cluster SLAM with non-overlapping fields of view☆

Michael J. Tribou[a,*], David W.L. Wang[b], Steven L. Waslander[c]

[a] Aeryon Labs Inc., 575 Kumpf Drive, Waterloo, ON N2V 1K3, Canada
[b] Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada
[c] Department of Mechanical and Mechatronics Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada

ARTICLE INFO

Article history:
Received 29 May 2014
Received in revised form 21 September 2015
Accepted 15 January 2016
Available online 23 February 2016

Keywords:
SLAM
Computer vision
Multicamera cluster
Non-overlapping FOV
Degeneracy analysis
Critical motions

ABSTRACT

An analysis of the relative motion and point feature model configurations leading to solution degeneracy is presented, for the case of a Simultaneous Localization and Mapping system using multicamera clusters with non-overlapping fields-of-view. The SLAM optimization system seeks to minimize image space reprojection error and is formulated for a cluster containing any number of component cameras, observing any number of point features over two keyframes. The measurement Jacobian is transformed to expose a reduced-dimension representation such that the degeneracy of the system can be determined by the rank of a dense submatrix. A set of relative motions sufficient for degeneracy are identified for certain cluster configurations, independent of target model geometry. Furthermore, it is shown that increasing the number of cameras within the cluster and observing features across different cameras over the two keyframes reduces the size of the degenerate motion sets significantly.

1. Introduction

Precise robotic motion and manipulation tasks with respect to unknown target environments and objects require an accurate, real-time measurement of the relative position and orientation of the robot and target. Multicamera systems are often employed for robotic pose and target model estimation, as each camera is an inexpensive, light-weight, and passive device capable of collecting a large amount of environment information at high rates. Many researchers across different fields have investigated the use of cameras for the purpose of estimating motion and scene structure. As a result, many techniques using a variety of camera types and configurations have been detailed in the literature.

A camera cluster is composed of any number of simple perspective cameras mounted rigidly with respect to each other, as shown in Fig. 1, including configurations in which their fields-of-view (FOV) are spatially disjoint [1]. This arrangement makes effective use of the camera sensors to cover a large combined FOV with high resolution, and in general, is able to overcome the limitations of other camera configurations, such as scale and translation-rotation motion ambiguities [2]. Additionally, by arranging the cameras to look in many directions, the pose estimation is made more robust since when certain cameras do not see any point features suitable for tracking, the other cameras in the cluster can maintain the localization. In this scenario, camera arrangements with a smaller collective FOV may become lost causing the tracking operation to fail. Several multicamera cluster Simultaneous Localization and Mapping (SLAM) [3] algorithms have been proposed in the literature (e.g. [4,5,6]).

In order for any pose estimation system to operate successfully, the current state must be uniquely recoverable given the measurable outputs up to, and including the current time step. In the context of a multicamera cluster relative pose system, this means that the image measurements must contain sufficient information to recover the cluster motion and the target model parameters, including the proper global scale metric. Within the SLAM system, the quality of the motion and structure parameter estimates are fundamentally linked. Further the convergence properties and accuracy of the solution depends on the point feature constellation geometry and relative motion trajectory of the camera cluster and target during the estimation. Identifying and avoiding configurations where a solution is difficult to recover is vital to successful application to robotics operations.

Furthermore, the solution must be unique since convergence to a different configuration, which may also agree with the

☆ This paper has been recommended for acceptance by Enrique Dunn, Ph.D.
* Corresponding author. Tel.: +1 519 635 8971.
E-mail addresses: mjtribou@aeryon.com (M.J. Tribou), dwang@uwaterloo.ca (D.W.L. Wang), stevenw@uwaterloo.ca (S.L. Waslander).

http://dx.doi.org/10.1016/j.imavis.2016.01.005
0262-8856/

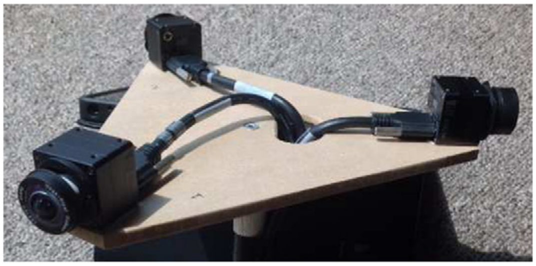

Fig. 1. An example camera cluster in which the three component cameras are rigidly-fixed with respect to each other.

measurements, would likely result in failure of perception and control operations. The visual SLAM system can be considered as a physical mechanism with the geometry and links represented in the parametrization. For the solution to converge to a unique configuration, it must be sufficiently constrained by the bearings-only image measurements or the mechanical system is not rigid and the solution can move without violating these constraints.

When the multicamera cluster is configured such that there is little or no spatial FOV overlap between the component cameras, the sensitivity of the image measurements to the global scale of the reconstructed model is low, particularly around specific motion profiles known as critical motions [7]. When the relative motion of the cluster is at or near critical, the global scale of the solution is extremely difficult, if not impossible, to recover accurately. In the presence of measurement noise, the solution will converge to an incorrect scale value. If the motion and structure estimates are used in a control loop, the incorrect scale could result in inefficient or unstable behaviour.

This work investigates the degenerate configurations when estimating the SLAM system states for a calibrated multicamera cluster over two keyframes while observing a set of point features in each camera and using an iterative optimization or recursive filter-based approach, minimizing the image space reprojection error of point feature measurements. This includes Bundle Adjustment (BA) [8] schemes as well as recursive filters such as an extended Kalman filter [9]. The main contribution is the identification of configurations of motion and target model structure leading to non-unique SLAM solutions.

Determining the system configurations leading to solution degeneracy is closely related to the concept of observability in control systems. In the study of observability for nonlinear systems, the local weak observability of the system can be determined by calculating the observability rank condition about any point in the state space [10]. This involves checking the column rank of a matrix containing the partial derivatives with respect to the system states, for increasing orders of Lie derivatives of the measurement model with respect to the system dynamics. When the matrix has full column rank, the system is locally weakly observable about that point.

For a SLAM system using only the visual measurements from the cluster cameras and a non-stationary target, the system does not have a model of the dynamics for the relative motion and therefore, only the zeroth-order Lie derivatives are non-zero. In this case, evaluating the observability rank condition is equivalent to checking the rank of the measurement Jacobian matrix, as will be done here in the degeneracy analysis in Section 4. If the system were to contain a model of the relative motion dynamics, and the extra information that comes with it, the higher-order Lie derivatives of the measurement model would contain non-zero terms and the added matrix rows would only increase the likelihood that the matrix has full column rank at any point in the state space. However, in this analysis, no such assumptions about the relative motion dynamics are made and the degenerate configurations arising from only using image measurements for a set of point features over two keyframes are identified.

The remainder of this paper is arranged as follows: Section 2 contains a review of the previous analyses for degenerate configurations of the multicamera cluster relative pose system; Section 3 presents the multicamera cluster SLAM system; the degenerate configurations of the pose estimation system are identified in Section 4; and finally, conclusions are drawn in Section 5.

2. Related work

Previous analyses identifying cluster motions leading to degenerate system solutions have assumed that the five degrees of freedom describing relative orientation and translation direction of the cluster are known using the well-studied single camera ego-motion estimation techniques (e.g. [8]). These include the work of Kim *et al.* [11], and Clipp *et al.* [7] for camera clusters with two component cameras, as well as that of the authors [12] for clusters with three component cameras. Of interest are the conditions when the image measurements from the camera cluster are able to allow for estimation of the final degree of freedom, corresponding to the translation magnitude and therefore, global system scale. The analyses show that when each point feature is seen by only one of the two cameras at both keyframes, the global scale of the solution is recoverable only when the relative translational and rotational motion are both non-zero, and does not result in the optical centres of each camera moving in concentric arcs on circles with a common centre at the intersection of the baselines at each keyframe [7]. When a third non-collinear camera is added to the cluster, the set of degenerate motions is reduced to those which result in all the three cameras moving in parallel [12].

Analyses of degeneracies of the full SLAM solution for multicamera clusters have focused on those associated with solving the generalized camera relative pose problem, either linearly using the Generalized Essential Matrix (GEM) [2], or aligning imaging rays in space for minimal cases of camera poses and points [13]. Sturm [14], Stewenius *et al.* [13], and Mouragnon *et al.* [15] discuss some degenerate cases, but Kim and Kanade [16] provide the most complete analysis. They identify the following degenerate configurations for generalized cameras using the seventeen point method [2]:

1. All of the observation rays pass through one common point before and after the camera motion.
2. The camera centres are on a line before and after the motion.
3. Each corresponding ray pair passes through the same local point in the general camera frame before and after the motion.

For a camera cluster with non-overlapping FOV, it is possible that each component camera observes its own mutually exclusive set of feature points over the two keyframes. In this case, the system satisfies condition 3 and the solution to the seventeen point algorithm is always degenerate. However, it is known from previous results that in certain configurations, other solution methods are able to recover an accurate estimate of the motion and structure. Consequently, the seventeen point algorithm does not always recover a solution when one exists. This problem was noticed by Li *et al.* [17], who have since modified the algorithm for use with non-overlapping clusters, but the subsequent degeneracy analysis has not been carried out. More importantly, the degenerate configurations are specific to the linear method of estimation. In this work, the minimization of image-space reprojection error is considered and the configurations for which an optimization of this type will fail are identified in the subsequent analysis.