A Multi-camera Approach to Vehicle Tracking based on Features

Lin Yang, John Johnstone and Chengcui Zhang
Department of Computer and Information Sciences
University of Alabama at Birmingham
E-mail: {galabing, jj, zhang}@cis.uab.edu

Abstract

Automobile detection based on feature tracking has proven effective in handling vehicle occlusions. However, current systems adopt a single camera to compute the 3D coordinates of feature points, which leads to an under-conditioned problem. As a result, only a few stable features can be reconstructed successfully. In this paper, we extend the system by adding a second camera, and integrate structure from motion (SfM) techniques to reconstruct a large volume of 3D points without sacrificing accuracy.

1. Introduction

The last decade has seen a huge body of research work on automatic traffic monitoring. A good survey can be found in [5]. Among all the attempts, feature-based detection has emerged as a promising tool [1], because it can avoid the ambiguities caused by vehicle occlusions. Recently, Kanhere et al. [4] proposed to recover the 3D coordinates of feature points from the calibrated camera model. However, since 3D reconstruction from a single view is inherently an under-conditioned problem, the system has to estimate the height for each feature point to provide the extra constraint, and a majority of features that cannot be consistently measured are discarded.

In the proposed system, instead of height estimation, a second camera is adopted to provide the missing constraint in 3D reconstruction. Given two video sequences, the system starts by calibrating the two camera models off-line using the standard DLT algorithm [3]. Then, at each timestamp, the synchronized frames from both cameras are passed to the KLT module [7] to detect point correspondences, followed by a robust estimator to remove outliers. The 3D coordinates of corresponding feature points are computed by the linear triangulation method [3]. In order to handle vehicle occlusions, the reconstructed 3D points are grouped as in [6]. The output of our system is a set of 3D points, grouped as different vehicles, which can be found useful in a variety of applications such as traffic counts and speed estimation.

2. System architecture

The proposed system, called AutoDetect, consists of the following four sub-modules.

2.1. Camera calibration module

Since the cameras are stationary, the calibration is a one-time, off-line process. Our system adopts a similar interface introduced in [5] on the two cameras separately: after one video sequence is loaded, the system extracts and displays the first frame, and allows the user to select at least six points. Once the image points are selected and their corresponding world coordinates specified, the DLT algorithm is triggered to compute the camera model.

2.2. Inter-frame tracking module

Feature points are tracked using the KLT feature detector, which provides a rich set of features. KLT is run every k frame(s), and the value of k is determined by both the camera settings and the traffic situation. The system can work more efficiently as k grows, but fewer feature points will be tracked consistently due to the large motion. Our experiments are conducted at a highway intersection with the frame rate of 30fps, and setting $k = 5$ can produce satisfactory results.

As vehicles keep entering and leaving the scene, the system frequently needs to replace the missing features with newly detected ones. The system automatically runs the KLT detection module after every inter-frame tracking to maintain a constant number of feature points.

0-7695-3084-2/07 $25.00 © 2007 IEEE
DOI 10.1109/ISM.Workshops.2007.21

2.3. Two view reconstruction module

In order to recover the 3D coordinates of the tracked features, correspondences must be found in the other camera view. This is, again, done by the KLT tracking module. Note that KLT is an optical-flow based algorithm and cannot deal with large motion, so the two cameras should be reasonably close to each other. RANSAC [2], a robust estimator, is adopted to remove any false correspondences returned by KLT. Finally, the robust correspondences are passed to the triangulation module to recover their world coordinates, using structure from motion techniques [3].

2.4. Grouping module

The reconstructed 3D points may contain multiple vehicles and must be labeled accordingly. The grouping module is built following the algorithm in [6]. This module constructs a graph over time, with each vertex corresponding to a 3D feature point. Newly emerged vertices are connected to all the existing ones in the graph, and at each timestamp, the module updates the distance between each pair of points, and removes the edge between their corresponding vertices if the distance has varied too much. The connected components are returned as the grouping result at that timestamp.

3. Analysis and future work

This paper presents AutoDetect, an automatic traffic monitoring system based on two cameras. The use of the second camera improves the previous work [5] [4] in several ways. First of all, it avoids estimating the feature heights, and thus does not rely on the foreground classifier [4], which is subject to lighting changes. Secondly, it is able to reconstruct as many as 90% of the tracked points, which can provide more accurate traffic parameters. Finally, the reconstruction is based on two views, instead of any estimated constraints, so the results are more robust compared to [4].

There are still some possible extensions of this framework. A more robust feature tracker could replace KLT in the two view reconstruction module, so that the two cameras can be freely installed. Another extension would be using the reconstructed data for vehicle classification. Since AutoDetect is able to recover a large volume of points for a typical vehicle, it can gain more accurate vehicle parameters such as position and type by finding the best-fitting model to the 3D points.

4. Acknowledgement

The research of Dr. Chengcui Zhang is supported in part by the UAB ADVANCE program through the sponsorship of the National Science Foundation.

5. References

[1] D. Beymer, P. McLauchlan, B. Coifman and J. Malik, "A real time computer vision system for measuring traffic parameters", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 495 - 501.

[2] M. Fischler and R. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Communications of the ACM, 1981, vol. 24, no. 6, pp. 381 - 395.

[3] Hartley, R., and A. Zisserman, *Multiple View Geometry in Computer Vision, second edition*, Cambridge University Press, 2004.

[4] N. Kanhere, S. Pundlik and S. Birchfield, "Vehicle Segmentation and Tracking from a Low-Angle Off-Axis Camera", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 1152 -1157.

[5] N. Kanhere, S. Birchfield and W. Sarasua, "Vehicle Segmentation and Tracking in the Presence of Occlusions", *TRB Annual Meeting Compendium of Papers, Transportation Research Board Annual Meeting*, 2006.

[6] N. Saunier and T. Sayed, "A Feature-based Tracking Algorithm for Vehicles in Intersections", in *Proceedings of the 3rd Canadian Conference on Computer and Robot Vision*, 2006, vol. 00, pp. 59.

[7] J. Shi and C. Tomasi, "Good Features to Track", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593 - 600.