The IEEE 5th International Conference on Intelligent Transportation Systems
3 – 6 September 2002, Singapore.

Mobile Vision-based Vehicle Tracking and Traffic Control

Tamer Rabie, Amer Shalaby, Baher Abdulhai
Dept. of Civil Eng.
Intelligent Transportation Systems Group
University of Toronto
ITS Center / R-SF3001
Toronto, Ontario, M5S 1A4
tamer@cs.utoronto.ca, {amer|baher}@ecf.utoronto.ca

Ahmed El-Rabbany
Dept. of Civil Eng.
Geomatics Engineering Group
Ryerson Polytechnic University
350 Victoria Street
Toronto, Ontario, M5B 2K3
rabbany@acs.ryerson.ca

Abstract

This paper discusses work-in-progress to develop a mobile, bus-mounted machine vision system for transit and traffic monitoring in urban corridors, as required by Intelligent Transportation Systems. In contrast to earlier machine vision technologies used for traffic management, which mainly rely on simple algorithms to detect certain traffic characteristics, the new proposed approach makes use of a recent trend in computer vision research; namely the active vision paradigm. Active vision systems have mechanisms that can actively control camera parameters such as orientation, focus, zoom, and vergence in response to the requirements of the task and external stimuli. Mounting active vision systems on buses will have the advantage of providing real-time feedback of the current traffic conditions while possessing the intelligence and visual skills which allow them to interact with a rapidly changing dynamic environment such as moving traffic.

Keywords: *Active Vision, Intelligent Transportation Systems, Bus-mounted Machine Vision, Vehicle Tracking, Stereo Analysis.*

1 Introduction

Machine vision and video technologies have found an increasing application in traffic management and transportation planning. Traffic surveillance and incident detection and management constitute the most widespread uses of video technology, while machine vision processing of license plate images for purposes of electronic toll enforcement make up most of the other common applications. Recently, these technologies have been used in adaptive traffic signal control systems, for monitoring speeds, and for collecting travel time data.

Intelligent Transportation Systems (ITS) is an emerging global industry that capitalizes on advanced technologies to better manage the dynamic, over-congested transportation networks of today. The current ITS boom has given rise to the need for a comprehensive real-time surveillance of traffic conditions over the transportation network to allow for dynamic control and management of traffic. Existing traffic detection technologies cover a wide spectrum of technologies as well as performance, ranging from modest pavement-buried inductive loop detectors to more advanced pole-mounted off-road detectors such as microwave, radar, and camera-based detectors. All existing detector types, however, share a common limitation of being point-detectors reflecting only traffic conditions at the locations of the detectors.

Many ITS applications require the collection of traffic data over an extended period of time at one or more locations. Wide area video traffic surveillance is one example. Such applications require the installation of numerous video cameras along with communications infrastructure to transmit video images to a Traffic or Transit Management Center (TMC) housing computer and video equipment for data processing and dissemination of information or control. In contrast, other applications require video images to be collected by means of tripod-mounted camcorders, with the images stored on videotape for subsequent viewing or machine vision processing [18].

In the past, vehicle-mounted machine vision technologies have been employed extensively in a variety of applications that mainly deal with Intelligent Vehicle (IV) and Automatic Vehicle Driving (AVD).Unlike existing technologies, the current research focuses on integrated development of a wide-area mobile-surveillance system where the tracker (the vision system on the bus) and the target (traffic in the camera view) and the processing tools (computer, systems and algorithms) are all in a mobile environment, posing new challenges to be addressed.

2 The Bus-Mounted Active Vision System

In this project, we develop a binocular machine vision system, mounted on buses in an urban corridor, to dynamically monitor traffic conditions along the corridor. Although the concept of using buses, aided with Automatic Vehicle Location (AVL) techniques, as probes of traffic conditions has recently gained considerable momentum, the use of mobile vision technology as proposed in this work has not been investigated to the best of our knowledge. Stationary real-time monocular machine vision systems have been recently developed for tracking vehicles under congested conditions for the purpose of traffic management [4].

From a technical perspective, the under development machine-vision-based approach is active and dynamic, as both the tracking agent and the targets are mobile in a dynamic environment. The level of complexity is significantly higher than the case of pole-mounted video technology where the camera and the background are both practically static, which mainly relies on passive-vision techniques to detect moving traffic objects and characteristics. The proposed approach will employ recent trends in computer vision research, namely the Active Vision Paradigm. Active vision systems have mechanisms that can actively control camera parameters such as orientation, focus, zoom, and vergence in response to the requirements of the task and external stimuli. Mounting active vision systems on buses will have the advantage of providing real-time network-wide feedback of the current traffic conditions while possessing the intelligence and visual skills that allow them to interact with a rapidly changing dynamic environment such as moving traffic. The main approach is to stabilize the visual field of view of our camera system by compensating for the vibrations due to the motion of the bus. This can be achieved by computing the displacement between successive video image frames and updating the gaze angles of the camera to cancel out this displacement which can be computed as a translational offset in the image coordinate system. Once the video image stream is stable, the next task is to detect and track vehicles that appear in the camera field of view. The relative speed of vehicles with respect to the speed of the bus can be estimated by tracking the motion of vehicles between frames. An accurate estimate of traffic speed can be obtained from the knowledge of the speed of the camera, which has the same speed as the bus, and the relative speeds estimated from the video images. Further traffic characteristics such as density and the presence of incidents can also be deduced from the image sequence.

2.1 Technical Rationale and Approach

Components similar to the bus mounted machine vision system are currently available in other fields. In particular, recent research carried out by Rabie and Terzopoulos [16, 19] on active vision in a simulated 3D virtual environment has enabled a new paradigm for computer vision research that is called "animat vision". The essence of the concept is to implement active vision systems allowing virtual animal robots (or animats) to understand perceptually the realistic virtual worlds in which they are situated so that they may effectively interact with other animats situated within these worlds. A set of active vision algorithms have been successfully implemented, within the animat vision framework, that integrate motion, stereo and color analysis. These algorithms support robust color object tracking, vision-guided navigation, visual perception, and obstacle recognition and avoidance abilities, enabling the animat to better understand and interact with its dynamic virtual environment.

We are combining the stereo, motion, and color algorithms together with a robust tracking algorithm, namely the KLT feature tracker, to increase the robustness and functionality of our overall vision system. The KLT feature tracker is based on the early work of Lucas and Kanade [14] and later developed fully by Tomasi and Kanade [20]. The only published, readily accessible description of this tracker is contained in a paper by Shi and Tomasi [17]. The final bus-mounted vision system will be able to detect moving targets of interest (moving vehicles) and segment their region of support using motion detection and optical flow estimation. It will then change its stereo camera gaze angles to fixate the detected target of interest. With the target inside the left and right camera's field of view, the feature tracker algorithm will take control to keep the moving target fixated and in view. Tracking will also facilitate the estimation of the relative speed of the tracked vehicle. The stereo algorithm will estimate the relative distance between the bus-mounted cameras and vehicles visible in front of the bus. These distances will be used by the tracker to filter out undesirable features (disconnected features and features on objects that are too close to the bus), and focus on tracking features that belong to the same vehicle.

The ongoing research builds on existing and tested vision techniques as a starting point from which we continue further development, augmenting them with enhancements suitable for the new application at hand. Finally, the developed vision components will be integrated into a sophisticated dynamic machine vision system that will be mounted on transit buses and calibrated to work at optimum performance.

2.2 The Prototype Bus-Mounted Vision System

The bus-mounted vision system consists of two main modules; the motion stabilization module and the stereo tracking module, as shown in Fig. 1. Together they implement a gaze control capability that enables the camera system to stabilize the visual world due to vibrations caused by the motion of the bus, as well as to detect visual targets in the camera field of view, change gaze to lock them in, and visually track them between image frames. Disparities between the stabilized left and right camera images will be estimated by the stereo tracking module, thus giving an estimate of the

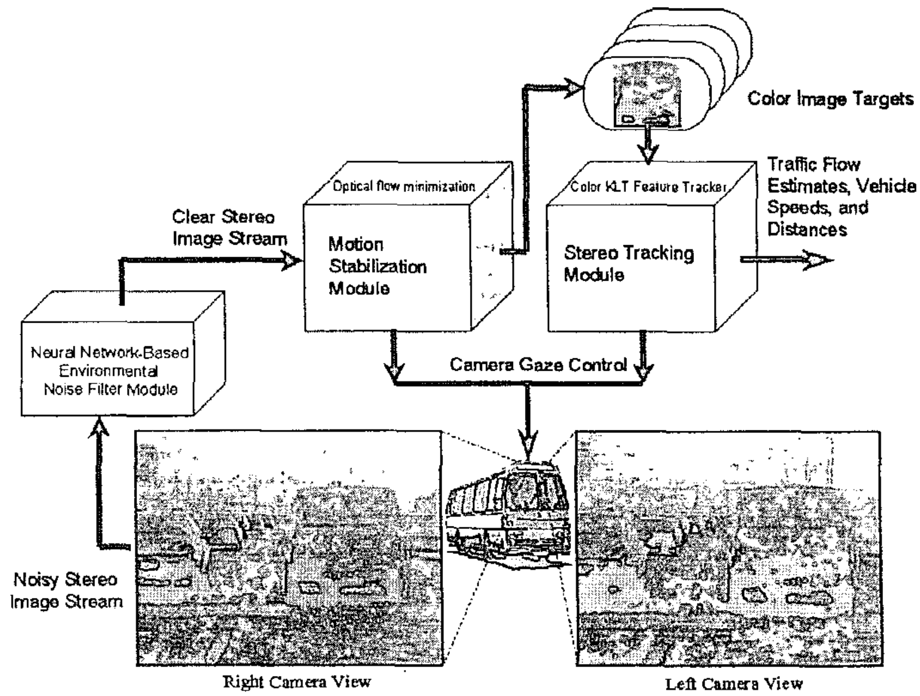

Figure 1: The Prototype Bus-mounted Active Vision System Diagram.

relative distance to objects in the image.

The vision system will have to contend with noisy images acquired under varying environmental conditions such as fog, rain, snow-fall, or extra darkness. An Artificial Neural Network (ANN) module, shown in Fig. 1, will be trained to receive the noisy stereo image stream, and recognize those conditions and their severity. To the neural network module, these conditions will appear as "noise frames" that are superimposed onto the ideal clear image stream that the system will initially be trained to deal with. The appropriate filtering algorithm within the neural network module would be, then, triggered to filter out that noise and restore a clear image stream.

2.3 Visual Field Stabilization using Optical Flow

It is necessary to stabilize the visual field of the stereo camera system because of the continuous motion of the bus. The optokinetic reflex in animals stabilizes vision by measuring image motion and producing compensatory eye movements. Stabilization is achieved by computing the displacement between successive image frames and updating the camera gaze angles by this displacement offset. The displacement is computed as a translational offset in the image frame coordinate system by a least squares minimization of the optical flow constraint equation between image frames at times t and $t-1$ [5].

Given a sequence of time-varying images, points imaged on the retina appear to move because of the relative motion between the eye and objects in the scene. The instantaneous velocity vector field of this apparent motion is usually called optical flow [3]. Optical flow can provide important information about the spatial arrangement of objects viewed and the rate of change of this arrangement [8]. Various techniques for determining optical flow from a sequence of two or more frames have been proposed in the literature [9, 1, 14]. Optical flow, once reliably estimated can be very useful in various computer vision applications. Discontinuities in the optical flow can be used in segmenting images into moving objects [5, 10]. Navigation using optical flow and estimation of time-to-collision maps have been discussed in [6] and [15].

Once the camera visual view is stabilized against any vibrations, the camera gaze is redirected at a moving target of interest. Redirecting gaze when a target of interest appears in the image frame can be a complex task. One solution would be to section the peripheral image into smaller patches or focal probes [5] and search all the probes for large motion fields. The strategy will work well for sufficiently small images, but for dynamic vision systems, that must process natural images, this approach is not effective. We propose a method based on motion cues to help narrow down the search for a suitable gaze direction. We propose to create a saliency image consisting of the optical flow field between two stabilized image frames. The saliency image S is then convolved with a circular disk of area equal to the expected area of the target object of interest as it appears in the image frame[1]. The blurring of the saliency image emphasizes moving objects in the image. The maximum in S is taken as the location of the fastest moving object and serves as the new gaze direction for the stereo camera system.

2.4 Target Tracking

Once the stereo camera system has been redirected to gaze at an appropriate target, the stereo tracking module assumes the task of selecting good features from the current image frame, consistently tracking these features over time. Tracking moving objects in video streams has been a popular topic in the field of computer vision in the last few years. The different tracking techniques for video data can be classified into four main approaches:

1. 3D Model based tracking: Three-dimensional model-based vehicle tracking systems have previously been investigated in the literature [13, 2]. The emphasis is on recovering trajectories and models with high accuracy for a small number of vehicles. The most serious weakness of this approach is the reliance on detailed geometric object models. It is unrealistic to expect to be able to have detailed models for all vehicles that could be found on the roadway.

2. Region based tracking: The idea here is to identify a connected region in the image associated with each vehicle and then track this region over time using a cross-correlation measure. Initialization of the process is most easily done by subtracting the background scene from the acquired image. A Kalman filter-based adaptive background model allows the background estimate to evolve as the weather and time of day affect lighting conditions. Foreground objects (vehicles) are detected by subtracting the incoming image from the current background estimate, looking for pixels where this difference image is above some threshold and then finding connected components [7]. This approach works fairly well in free flowing traffic. However, under congested traffic conditions, vehicles partially occlude one another instead of being spatially isolated, which makes the task of segmenting individual vehicles difficult. Such vehicles will become grouped together as one large blob in the foreground image.

3. Active contour based tracking: A dual to the region-based approach is tracking based on active contour models, or snakes [12]. The idea is to have a representation of the bounding contour of the object and keep dynamically updating it. The advantage of having a contour-based representation instead of a region based representation is reduced computational complexity. However, the inability to segment vehicles that are partially occluded remains. If one could initialize a separate contour for each vehicle, then one could track even in the presence of partial occlusion [4].

4. Feature based tracking: Finally, yet another approach to tracking abandons the idea of tracking objects as a whole but instead tracks sub-features such as distinguishable points or lines on the object. The advantage of this approach is that even in the presence of partial occlusion, some of the sub-features of the moving object remain visible. The technology of tracking points and line features is developed fully as the KLT feature tracker by Tomasi and Kanade [20] and a readily accessible description of this tracker is contained in the paper by Shi and Tomasi [17].

For our specific application, we require efficiency, robustness to occlusion, and real-time tracking at all times. The feature-based tracking approach, described above, satisfies our requirements. We, thus, propose to incorporate the KLT feature tracker into the stereo tracking module. This tracker locates good features by examining the minimum eigenvalue of each 2 by 2 image-gradient matrix, and features are tracked using a Newton-Raphson method of minimizing the difference between the two matrices in two consecutive frames. Multiresolution tracking allows for large displacements between images [17]. To further enhance the tracker algorithm, we propose to enrich the description of each pixel in the image by making use of the (R, G, B) color signals from our color images by including them in the feature description for each pixel. This is expected to improve the feature tracking process considerably by restricting the matching process to features of similar color composition, which can be considered as a sort of color-feature constraint.

2.5 Vehicle Speed Estimation

To help our tracking algorithm focus on the tracked vehicle and to reduce distraction due to background clutter in the image sequence, we feed our tracking algorithm difference images instead of the actual images. The difference images are created by subtracting the previously acquired im-

[1] Reasonably small areas suffice, since objects in the image frame are typically small in front of the bus-mounted camera. Methods for estimating appropriate areas for the object, such as Jagersand's information theoretic approach [11], may be applicable.

age of the road from the image acquired at the current time instant. This has the advantage of blocking out the details in the background of the tracked vehicle while only emphasizing the vehicle to be tracked. This is possible due to the fact that the closer the vehicle is to the camera, the larger its motion will be, and the farther away the vehicles are, the more insignificant their motions will be. Thus, when subtracting previous image from current image, only the large motions of the vehicle in front of the bus will be emphasized in the difference image, while the vehicles that are far away from the bus get cancelled out of the difference image.

To facilitate this, we make use of small camera fields of view, thus only capturing a small area of the image in front of the bus, where the possibility of capturing only a single vehicle is higher. Even if the camera captures more than one vehicle in front of the bus, the difference image will still emphasis the motion of the closer vehicle to the bus-mounted camera. Fig. 2-(a) shows three selected image frames of a longer sequence acquired by a camera mounted on a Toronto Transit Commission (TTC) bus on route in one of the streets of Toronto. Fig. 2-(b) shows the corresponding difference images with the three bright white points on the license plate of the 4x4 vehicle corresponding to the good features that the tracker algorithm was able to select and successfully track between image frames. The images clearly show that key features are correctly detected and tracked over time. The displacement of these tracked features between image frames will give the relative speed of the tracked vehicle, knowing that the video camera is recording at 30 frames per second. Given that we know the speed of the bus, through an integrated low cost GPS/Dead Reckoning system, we can estimate the speed of the tracked vehicle.

2.6 Stereo Analysis

Stereo analysis is a process of extracting scene depth information by measuring the disparity of corresponding points between left and right binocular images.

In future work we will make use of the stereo bus-mounted camera by employing a stereo algorithm to estimate the relative distance between the bus-mounted cameras (the observer) and the tracked vehicles visible in the image. This will allow the tracking module to reject objects that are too close or too distant, thus giving a more accurate estimate of the relative speed of the tracked vehicle. Tracking features on objects that are too close to the observer can be prone to errors due to the difficulty of accurately estimating large displacements typical of close objects. The estimated disparities will also be used to allow the tracker to focus on tracking feature points that have similar disparity estimates (low variance disparities) and where there are no disparity discontinuities (high variance disparities) in between them indicating that they belong to the same vehicle.

3 Conclusion

This paper has presented research-in-progress to develop a mobile, bus-mounted machine vision technology for transit and traffic monitoring in urban corridors, as required by Intelligent Transportation Systems. The ongoing research builds on existing and tested vision techniques as a starting point from which we continue further development, augmenting them with enhancements suitable for the new application at hand. The developed vision modules will be integrated into a sophisticated dynamic machine vision system that will be mounted on transit buses and calibrated to work at optimum performance.

Acknowledgements

The ongoing research described herein is supported by a Communications and Information Technology Ontario (CITO) grant.

References

[1] P. Anandan. A computational framework and algorithm for the measurement of visual motion. *Inter. J. Computer Vision*, 2:283–310, 1989.

[2] K.D. Baker and G.D. Sullivan. Performance assessment of model-based tracking. In *Proc. of the IEEE Workshop on Applications of Computer Vision*, pages 28–35, Palm Springs, CA, 1992.

[3] D.H. Ballard and C.M. Brown. *Computer Vision*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982.

[4] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik. A real-time computer vision system for measuring traffic parameters. In *IEEE Conference on Computer Vision and Pattern Analysis (CVPR'97)*, 1997.

[5] P.J. Burt, J.R. Bergen, R. Hingorani, R. Kolczynski, W.A. Lee, A. Leung, J. Lubin, and H. Shvaytser. Object tracking with a moving camera: An application of dynamic motion analysis. *Proc. IEEE Workshop Visual Motion*, pages 2 – 12, March 1989.

[6] M. Campani, A. Giachetti, and V. Torre. Optic flow and autonomous navigation. *Perception*, 24:253–267, 1995.

[7] B. Gloyer, H.K. Aghajan, K.Y. Siu, and T. Kailath. Vehicle detection and tracking for freeway traffic monitoring. In *Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 970–974, 1994.

[8] B.K.P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

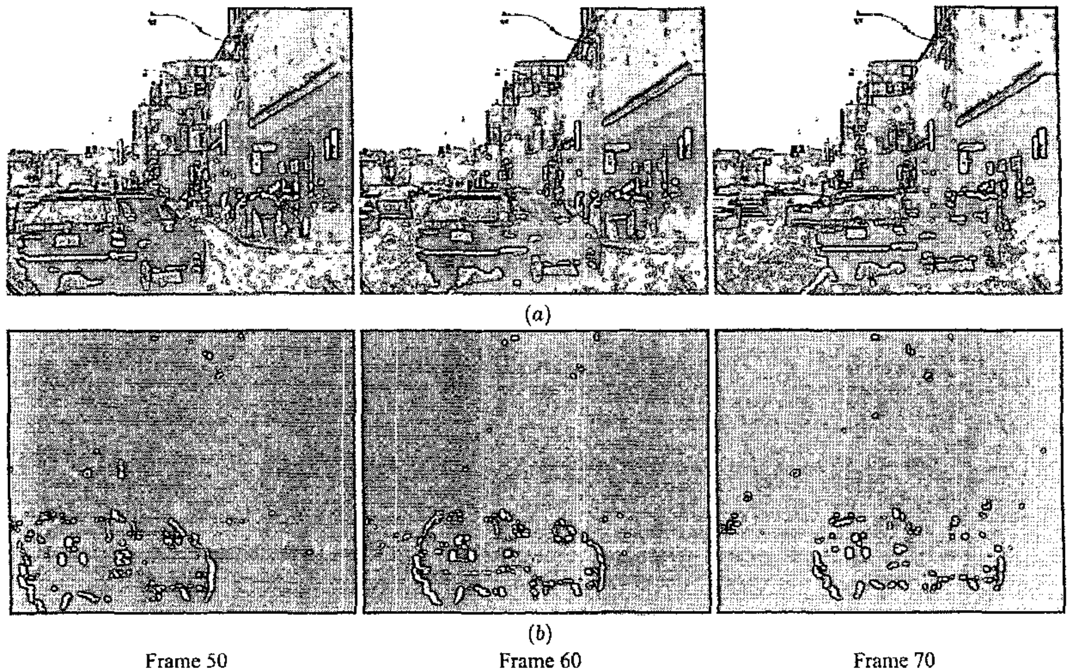

Figure 2: Good features represented as bright white feature points are tracked over the sequence of 20 difference image frames.

[9] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[10] M. Irani and S. Peleg. Image sequence enhancemenmt using multiple motion analysis. In *Proc. 11th IAPR, Int. Conf. on Pattern Recognition*, pages 216–221, 1992.

[11] M. Jagersand. Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach. In *Proc. Inter. Conf. Computer Vision (ICCV'95)*, pages 195–202, MIT, Cambridge, Massachusetts, June 20 - 23 1995.

[12] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. of Computer Vision*, 1(4):321–331, 1987.

[13] D. Koller, K. Daniilidis, and H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10:257–281, 1993.

[14] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. Image Understanding Workshop*, pages 121–130, 1981.

[15] F. Meyer and P. Bouthemy. Estimation of time-to-collision maps from first order motion models and normal flows. In *Proc. 11th IAPR, Int. Conf. on Pattern Recognition*, pages 78–82, 1992.

[16] T.F. Rabie and D. Terzopoulos. Motion and color analysis for animat perception. In *Proc. Thirteenth National Conf. on Artificial Intelligence (AAAI'96)*, pages 1090–1097, Portland, Oregon, August 4-8 1996.

[17] Jianbo Shi and Carlo Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[18] P.W. Shuldiner. Video technology in traffic engineering and transportation planning. *Journal of Transportation Engineering*, Sept./Oct. 1999.

[19] D. Terzopoulos and T.F. Rabie. Animat vision: Active vision in artificial animals. *Videre: Journal of Computer Vision Research*, 1(1):2–19, September 1997.

[20] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. Technical report cmu-cs-91-132, Carnegie Mellon University, April 1991.