Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

Video Based Surround Vehicle Detection, Classification and Logging from Moving Platforms: Issues and Approaches

Tarak Gandhi and Mohan M. Trivedi

Abstract—**This paper discusses the issues and approaches involved in developing a mobile vehicle-mounted system to detect, classify, and log the surrounding vehicles in a database for efficient query-based retrieval. This system consists of three components (1) Vehicle sensing, detection, and tracking (2) Feature extraction and classification (3) Database storage and retrieval. Relevant research in each of these components is described in order to guide the development of such a system. Such a system has applications including traffic analysis from mobile probes, analyzing driver behavior based on surrounding vehicles, as well as surveillance from mobile platform.**

I. Introduction and Motivation

INTELLIGENT vehicle systems have been mainly focused on detecting lanes, vehicles, pedestrians, and other obstacles on the road in order to warn the driver or take appropriate action for avoiding accidents. However, another potential application of mobile camera systems is to record the view of surroundings and detect, classify, and store the objects of interest that passed around the host vehicle.

Fixed cameras mounted in the infrastructure have been used for traffic monitoring and analysis for quite some time. However, a static camera has a limited coverage and a large number of cameras would be required to monitor the traffic patterns. An alternative solution is to use probe vehicles that move around a large area in order to obtain traffic statistics [1]. Such a vehicle would have one or more cameras that would continuously monitor the surroundings of the vehicle. The vehicles would be automatically detected and classified. Using the location information from GPS system, the traffic flow along the route of the vehicle can be determined. Such a system can keep record of detected and classified vehicles which can be useful for surveillance applications as well as for studying the effect of various types of objects on the driver's behavior. Such a system could also identify particular types of objects such as police cars, postal vans, etc. By logging the recorded vehicles, it would be possible to retrieve vehicles based on user queries.

This paper describes the issues involved in the development of a system that records the scene taken from vehicle mounted camera, automatically detects the vehicles, extracts their shape and appearance properties to classify the vehicles and store them in database for efficient query-based retrieval. The architecture of this surround vehicle detection, classification, and logging system is shown in Figure 1. Vehicles are detected using sensors and tracked over the entire field of view of the sensor. The CAN bus information as well as the position information from GPS are used to find the location of the host vehicle. The position of the detected vehicles relative to the host vehicle, which is estimated by the detection algorithm, can also be used to find the locations of those vehicles on the road. Shape and appearance based properties are extracted from the vehicles and using these properties, the vehicles are classified into different categories. The properties along with classification are used to create vehicle record that is sent to the database. A query engine accepts queries from user interface and retrieves the relevant records from the database. Issues in each of these stages are discussed in the subsequent sections.

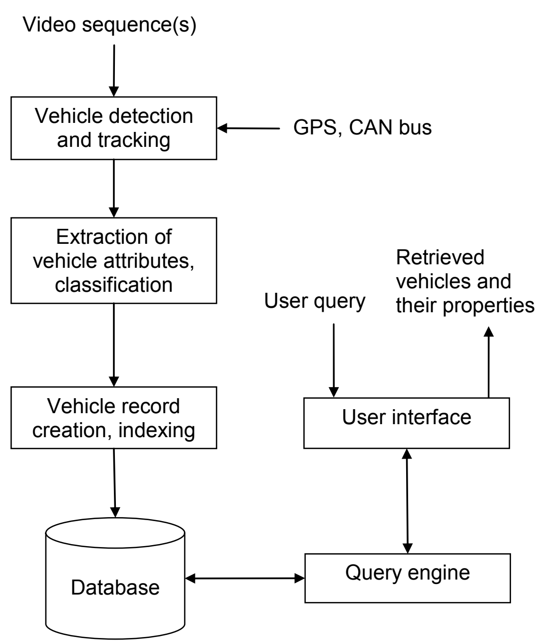


Fig. 1. Surround vehicle classification, logging, and retrieval system.

This research has been sponsored by the Technical Support Working Group (TSWG) of the US Dept. of Defense, and the UC Discovery Grant.

Tarak Gandhi is with the Computer Vision and Robotics Research Laboratory, University of California San Diego, USA (phone: 858-822-0053; e-mail: tgandhi@ucsd.edu).

Mohan M. Trivedi is with the Computer Vision and Robotics Research Laboratory, University of California San Diego, USA (e-mail: mtrivedi@ucsd.edu).

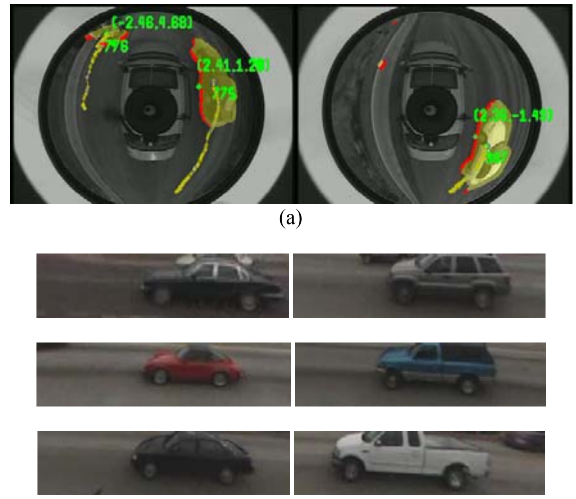
II. Vehicle Sensing, Detection, and Tracking

Commonly used sensors for detecting vehicles are imaging sensors and the 'time-of-flight' sensors. Imaging sensors can capture a high-resolution perspective view of the scene. Omnidirectional and other wide field of view cameras can monitor the entire scene around the vehicle at a comparatively lower resolution. On the other hand, active PTZ camera systems can zoom in to focus their attention on specific object of interest. An active vision system can detect the objects in omni camera and automatically direct the PTZ camera to capture high-resolution image of objects. Time-of-flight sensors such as RADARs and LASER scanners directly give accurate information about object distance. They can complement the imaging sensors in order to resolve distance ambiguities to get more robust detection. However, in the case of RADARs, the angular resolution is often limited. On the other hand, though the LASER scanners give a good angular resolution (0.25 degrees is currently available) in horizontal direction, they scan a single plane or a limited number of planes limiting the angular resolution in vertical direction.

Various approaches have been developed for detecting vehicles from moving platforms as described in the survey paper [2]. The preprocessing or attention focusing stage processes raw data using simple cues and fast algorithms to identify potential vehicle candidates. This stage needs to have high detection rate even at the expense of allowing false alarms. The classification and verification stage then applies more complex algorithms to the candidates from the attention focusing stage in order to separate genuine vehicles from false alarms. However, the boundary between these stages is often blurred and some approaches combine the detection and recognition stages into one. The detected vehicles are tracked over time to get trajectories.

For example, image motion cues can be used to detect vehicles or other objects based on their independent motion or height above the road which gives parallax when the host vehicle is in motion. Binocular stereo can also be used to obtain distances to the scene points based on disparity between images obtained from two cameras. The depth information offers valuable cues for separating vehicles from background, and provides accurate 3-D location of the vehicle.

Figure 2 shows examples of vehicle detection using video from a single omnidirectional camera using motion based approach described in detail in [3]. The road is assumed to be planar and its approximate image motion is first estimated and compensated from vehicle speed from the CAN bus, and the nominal calibration of the camera with respect to the road plane. The objects with independent motion and height would have large residual motion making it possible to separate them from road features. However, the features on the road may also have some residual motion due to errors in the vehicle speed and calibration parameters. Spatio-temporal gradients of the motion compensated images are obtained and used with optical flow constraint in statistical estimation framework to improve the estimates of the vehicle and camera parameters. Presence of residual motion in conjunction with constraints on vehicle length and separation is used to hypothesize the presence of vehicles. The vehicles are tracked over time using Kalman filter.



(a)

(b)

Fig. 2. Motion-based vehicle detection using an omni camera on top of a vehicle [3] (b) Snapshots of different types of vehicles automatically detected from the omni camera.

However, it should be noted that motion and stereo cues do not directly discriminate between different kinds of objects such as vehicles, pedestrians, and animals. Size, shape and appearance cues can be used to separate different types of moving objects. Size constraints can be used to perform coarse level separation between objects. On the other hand, shape and appearance cues would be useful for filtering out clutter that is not removed by previous stages, and to perform fine level classification of objects. For this purpose, characteristic features are extracted from images and a trained classifier is used to separate objects into different categories. It should be noted that this requires extensive modeling or training to detect the wide variety of vehicles that are seen on the road.

In order to obtain the trajectories of vehicles, the detected candidates are tracked over time. Kalman filter with data association techniques are often used for performing vehicle tracking. Tracking also helps to eliminate false alarms intermittently reported by the detection algorithms. Furthermore, instead of obtaining a single view, one now has multiple views of the same target from different viewpoints which can help in reconstructing the 3-D shape of the vehicle.

III. Feature Extraction and Classification

Considerable research has been performed on classifying vehicles from stationary platforms. Some of these ideas can be extended to perform classification from moving vehicle if a reasonably accurate bounding box of the vehicle is obtained. Classification of vehicles can be performed in hierarchical manner. At the coarse level, vehicles can be classified using image based measurements such as bounding box size, blob area, and moments. Since the number of such features is limited, fewer amounts of training data are required.

For example, Gupte et al. [4] use dimensions of bounding box to categorize detected vehicles into cars and non-cars. Huang and Liao [5] use a number of measurements such as size, aspect ratio, compactness in a hierarchical manner to discriminate between 7 classes of vehicles such as cars, SUVs, buses, and different types of trucks. Morris and Trivedi [6] use a feature vector of 17 image-based measurements. However, some of these features may be dependent on each other or may not contain sufficient discriminatory information. Hence, the feature space is projected to a lower dimensional space using Fisher's Linear Discriminant Analysis (LDA). Weighted k-Nearest Neighbor approach is used to classify the vehicles into 7 classes including cars, SUVs, vans, and trucks. The classification is improved by integration over the entire track using maximum likelihood approach. Figure 3 shows an example of classification using this method.

It should be noted that the performance of image measurement based classification depends on good estimation of the location and the bounding box of the vehicle, which may not always be accurately available especially from moving camera detection. On the other hand, shape and appearance based approaches can combine the detection of vehicles with classification by scanning over the entire image for vehicles. However, if the approximate location and size are obtained before, they could be used to reduce the search space and eliminate many false alarms.

Texture descriptors such as Histograms of oriented gradients (HOG) have been proposed in [7] to classify objects. This approach divides the image into rectangular cells and computes the histogram of the gradient orientations in each cell as shown in Figure 4. These histograms are used as feature vectors for the Support Vector Machine to distinguish between pedestrians and other objects. The approach has also been extended to finding other objects. Koch and Malone [8] use the HOG at multiple scales to distinguish between vehicles and other objects such as animals and people in thermal infrared images. The results from individual frames are fused over the entire vehicle track using sequential probability ratio test.

We applied the extraction of HOG features followed by SVM classifier to the omni images obtained from a moving vehicle. The images were used to obtain virtual perspective views looking towards the side of the vehicle. The gradient of the full image was obtained using scale space approach described in [8]. To check the effectiveness of the classifier itself, we manually labeled the positions of the bottom center of the vehicles. Assuming that the vehicle lies on

Fig. 3. Image and track based classification (a) cars (b) van (c) SUV. Note that (b) and (c) were misclassified using single image but correctly classified using full track [6].

Fig. 4. Histogram of oriented gradients (HOG) representation of vehicle snapshots. The red arrows denote the gradient orientations with the length corresponding to the frequency of that orientation in the local histogram around that point.

road plane, the Y-coordinate of the position gives the estimate of the image size of the object. Based on this, a bounding box of appropriate dimensions was selected that would include the vehicle in most cases. The HOG features were computed as follows:

(1) Subdivide the bounding box into an MxN grid.
(2) For each grid element, quantize the gradient directions in range of 0 to 180 degrees into K bins and histogram weighted by image intensity was obtained.
(3) Apply smoothing in spatial and orientation directions to the histogram array to reduce sensitivity to discretization.
(4) Stack the resulting array into a B=MxNxK dimensional vector.

For this experiment, the values of M=16, N=6, and K=4 were used giving number of bins B=384. Figure 4 shows examples snapshots showing some of the vehicles along with the HOG representation shown by red arrows. These vectors were fed to the SVM classifier [9] with 4 classes (1) car (2) van (3) pickup truck (4) No vehicle.

In order to test the efficiency of the classifier, we used a set consisting of 390 samples, with 157 vehicles and 233 without vehicles with the bounding boxes specified as described above. The set was equally divided into training and testing sets. The SVM was trained using the training set. In order to optimize the classification accuracy, cross-validation was performed by re-splitting the training set into partitions, training on one and testing on another. This was repeated multiple times in order to obtain optimal selection of parameters. The trained SVM classifier was applied to the test set. Table 1 gives the confusion matrix for the test data. The false alarm rate (non-vehicle classified as vehicle) is 0 % (0/108) and the detection rate considering all vehicles as same class is 98.8 % (85/87). For classification of vehicles, the accuracy is 64.3 % (56/87) showing that quite a few of the vehicles are misclassified. However, if the minivans and pickups are combined into one class, the accuracy of discriminating cars vs. other vehicles is 82.8 % (72/87). Note that the classification results are for the idealized case where the locations of the vehicles are manually selected. The overall accuracy of the system would depend on how accurately the detection stage finds the location of the vehicle.

It can be seen that the HOG approach is good for detecting vehicles, but discrimination between different types of vehicles is less reliable. In fact, HOG approach has been mostly used for coarse level distinction between entirely different categories of objects rather than fine level distinction between different types of objects in the same category where the similarities between objects are more significant than the differences. In order to perform fine classification, detailed shape features of the object may be needed. Ma and Grimson [10] use edge based descriptors as features. The edge points are extracted and grouped, and the features are formed from the edge segments. The SIFT features introduced by Lowe [11] are used since they have been shown to give good classification performance. They show good performance in discriminating between sedans, taxis, and minivans. Negri et al. [12] use oriented contour point based voting algorithm to identify the make and model of vehicles. The features used for this approach are based on the location and orientation of edge points in the image containing the vehicle.

TABLE 1
CONFUSION MATRIX FOR HOG BASED CLASSIFICATION

Ground truth	Sedan	Minivan	Pickup	None	Total
Sedan	**34**	2	0	0	**36**
Minivan	8	**17**	9	0	**34**
Pickup	4	7	**5**	1	**17**
None	0	0	0	**108**	**108**

IV. DATABASE STORAGE AND RETRIEVAL

Image databases efficiently store the properties of detected object and retrieve specific vehicles based on user queries. In addition to the shape, size, and appearance based features described above for classifying the vehicles, color information is also stored in order to address specific queries such as:

(1) Get all red cars that overtook the host vehicle.
(2) Find all police cars that were seen by the host vehicle.
(3) What is the percentage of SUVs that have a particular color?

The color distribution is represented by taking a histogram of the object pixels. However, the spatial information would be lost in this process. Hence, the histograms for color and shape features can be integrated to obtain more reliable discrimination [13]. Alternatively, appearance maps [14] can be generalized for representing the spatial distribution of color. Since the side view of a vehicle is predominantly rectangular in shape, the vehicle can be divided into 2-D rectangular grid. The color statistics of each grid element can be represented in various ways such as average value, histogram, or Gaussian mixture model.

Based on the above discussion, the typical examples of features stored can be:

(1) Blob measurements such as bounding box size, area, moments, etc.
(2) Vehicle trajectory over time to infer location and speed.
(3) Color features such as histogram or appearance map.
(4) Texture features such as histogram of oriented gradients.
(5) Edge based descriptors for fine distinction between vehicle shapes.

A query engine would match the features extracted from the query to the entries in the database using various similarity measures [15]. However, if the database is large, linear search for matching all entries is inefficient. Indexing schemes are needed for organizing efficient search. Indexing of features with high dimensionality is a difficult problem that many researchers have tried to address.

V. Concluding Remarks

This paper discussed the issues involved in the development of a mobile platform-based vehicle classification and logging system. Such a system could be useful for many applications such as intelligent traffic probes, driver behavior analysis, and mobile surveillance. Relevant research in the individual components of the system was described so that educated decision can be made for the overall design of the system.

Experimental results obtained from motion based detection and histogram-of-gradients (HOG) based classification were discussed. It was seen that the HOG approach was more appropriate for confirming the presence of vehicle and discrimination at coarse level such as between cars and other vehicles. However, discrimination at finer level into larger number of classes would require more selective approaches such as [10].

Acknowledgment

We thank the Technical Support Working Group (TSWG) of the US Dept. of Defense and the UC Discovery Grant for their sponsorship of the research. We also thank the members of CVRR laboratory for their valuable suggestions.

References

[1] Tamer Rabie, Gasser Auda, Ahmed El-Rabbany, Amer Shalaby, and Baher Abdulhai, "Active-Vision-based Traffic Surveillance and Control," *Proc. of the Fourteenth Canadian Vision Interface Conference (VI 2001),* pp. 87-93, Ottawa, Canada, June 7-9, 2001.

[2] T. Gandhi and M. M. Trivedi, "Vehicle Surround Capture: Survey of Techniques and a Novel Omni Video Based Approach for Dynamic Panoramic Surround Maps," *IEEE Transactions on Intelligent Transportation Systems,* Sept 2006.

[3] T. Gandhi and M. M. Trivedi. "Parametric Ego-Motion Estimation for Vehicle Surround Analysis Using an Omnidirectional Camera." *Machine Vision and Applications*, 16 (2), 85-95, February 2005.

[4] S. Gupte, O. Masoud, R. F. K. Martin, and N. P. Papanikolopoulos, "Detection and Classification of Vehicles," *IEEE Transactions On Intelligent Transportation Systems*, Vol. 3, No. 1, pp. 37-47, March 2002.

[5] Chung-Lin Huang and Wen-Chieh Liao, "A Vision-Based Vehicle Identification System," Proc. International Conference on Pattern Recognition, 2004.

[6] B. Morris and M. M. Trivedi, "Improved Vehicle Classification in Long Traffic Video by Cooperating Tracker and Classifier Modules", *Proc. IEEE International Conference on Advanced Video and Signal based Surveillence,* Nov 2006.

[7] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, p. 886-893, June 2005.

[8] Mark W. Koch and Kevin T. Malone, "A Sequential Vehicle Classifier for Infrared Video using Multinomial Pattern Matching," Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop, June 2006.

[9] Chin-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for Support Vector Machines," Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001.

[10] Xiaoxu Ma and W. Eric L. Grimson, "Edge-based rich representation for vehicle classification," Proc. IEEE International Conference on Computer Vision, 2005.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, November 2004.

[12] Pablo Negri, Xavier Clady, Maurice Milgram, and Raphael Poulenard, "An Oriented-Contour Point Based Voting Algorithm for Vehicle Type Classification," *Proc. International Conference on Pattern Recognition*, 2006.

[13] Ediz Saykol, Ugur Gudukbay, Ozgur Ulusoy, "A histogram-based approach for object-based query-by-shape-and-color in image and video databases," *Image and Vision Computing* Vol. 23, pp. 1170–1180, 2005.

[14] T. Gandhi, M. M. Trivedi, "Panoramic Appearance Map (PAM) for Multi-Camera Based Person Re-identification," *Proc. IEEE International Conference on Advanced Video and Signal based Surveillence*, Nov 2006.

[15] Image Databases: Search and Retrieval of Digital Imagery, Ed. Vittorio Castelli and Lawrence D. Bergman, John Wiley and Sons Inc., New York, 2002.