Proceedings of the IEEE ITSC 2006
2006 IEEE Intelligent Transportation Systems Conference
Toronto, Canada, September 17-20, 2006

Robust Classification and Tracking of Vehicles in Traffic Video Streams

Brendan Morris and Mohan Trivedi
University of California, San Diego
La Jolla, California 92093-0434
Email: {b1morris, mtrivedi}@ucsd.edu

Abstract— **The widespread use of cameras for traffic monitoring coupled with the availability of robust tracking algorithms has led to volumes of data. It is necessary to process this data for higher level tasks. One of these processing tasks is vehicle type classification, which can be used in a query based management system. This paper presents a tracking system with the ability to classify vehicles into three classes {Sedan, Semi, Truck+SUV+Van}. This system was developed after comparing classification schemes using both vehicle images and measurements. The most accurate of these learned classifiers was integrated into tracking software. This merging of of classification and tracking greatly improved the accuracy on low resolution traffic video.**

I. Introduction

Today there are an ever growing number of cameras being used for scene analysis. Many of these are applied to traffic monitoring because it is a low cost and passive method for data collection. Past research [1] has been devoted to accurately tracking vehicles. The relative strengths and limitations of the varied trackers are well researched. Background subtraction trackers perform quickly and can adapt to various lighting conditions but deal poorly with occlusion or adverse weather conditions. Interest feature detection can better model a vehicle with occlusion but still suffer from difficulties ensuring robust localization of salient features. Model based trackers are robust to illumination and occlusion but require models for all vehicles, limiting its scalability. The focus of more recent research is on higher level analysis of this tracking data. This is of great importance for dealing with the growing problem of urban congestion. The traffic congestion problem is costing Americans $63.1 billion a year. Any analysis that can help assess and direct planning and ameliorate this problem is needed.

One of the primary high level analyses one could envision is vehicle classification. Vehicle classification is particularly useful for re-identification [2] in multi-sensor networks [3] and anomalous event detection [4] as well as the more standard applications of traffic flow analysis and unobtrusive path tracking [5]. This paper demonstrates the effectiveness of combining tracking with classification for significantly improved classification results on low resolution traffic video. The technique is also general enough to be applied to a wide variety of surveillance scenes besides traffic.

II. System Overview

This paper presents a traffic monitoring system that is able to classify the type of vehicles in a highway scene. We classify vehicles into three main classes, Sedan, Semi, and an additional class, Truck + SUV + Van (TSV). Examples from each of the classes are shown in Fig. 1. The Van samples are composed of a variety of vehicles, e.g. minivans, that were either not explicitly separated or were vehicles that were difficult for a human viewer to classify. The Truck, SUV, and Van examples were grouped together because of the similar appearance and represent the most diverse class. The examples shown have been scaled for display purposes but demonstrate the difficulties for a classifier. The vehicles have differing scale, perspective view, and can be quite low resolution when in the lanes furthest from the camera, making the appearance of different classes similar. Fig. 4 shows a sample of the video output. The text accompanying each vehicle is a Track ID, class identifier number, followed by a estimate of speed and direction of travel. Our proposed system would be similar to [6] but without explicitly selecting regions to build classifiers. Instead the system will have one classifier trained for the entire scene and eventually should be invariant to the camera pose selected by a remote operator.

The vehicle classifier was built after doing a comparison of different classification schemes using either image based (IB) or image measurements based (IM) features. A feature transformation technique, principle component analysis (PCA) or linear discriminant analysis (LDA), was applied to manage the size of the data and classification was accomplished using a weighted K nearest neighbor classifier. The classifier scheme with the best performance on test sets from two days was integrated into tracking software to output track information that included a track vehicle type label of greatly improved accuracy.

III. Comparison Study

We are interested in comparing the performance between classifiers using the image of a vehicle as a feature and features derived from measurements taken from the image. The features are obtained automatically from traffic tracking software and used for classification.

A. Image Based (IB) Features

Here the image of the tracked object is used as a feature vector. To do a proper comparison each object was resized

1-4244-0094-5/06/$20.00 ©2006 IEEE

(a) Class 0 - Truck + SUV + Van (TSV)

(b) Class 1 - Sedan (c) Class 2 - Semi

Fig. 1. Sample Images from Each Class. Vehicles are Detected at Different Scales and Perspective.

to [64x32] pixels, generating a feature vector with 2048 components. Images as feature vectors have been used in human face recognition literature for many years with a high degree of success [7]. We borrow those techniques and attempt to apply them to the vehicle classification problem. The added complexity for vehicle classification is the large variation between classes. In face recognition, the structure of two different subjects is quite similar (eyes, nose, mouth) as opposed to vehicles which may not have the consistent structure necessary for alignment. In addition many of the face databases are constructed in very controlled environments to deal with lighting and background clutter.

B. Image Measurements Based (IM) Features

In addition to the image of tracked objects, a number of simple measurements are taken. We would like to work with measurements rather than the images directly because it is much cheaper computationally and storage-wise to maintain a database of features rather than images. The goal is to obtain as many simple measurements as possible and allow a classifier to decide which are best for classification. Currently, 10 different measurements make up the measurement feature vector. These measurements have been normalized by applying a homography transformation to the road ground plane allowing comparison of vehicles throughout the entire field of view. The feature vector consists of

- area
- bounding box [width, height]
- convex area
- ellipse [eccentricity, major axis, minor, axis]
- extent - proportion of pixels in bounding box to object
- solidity - proportion of pixels in convex hull to region
- perimeter

Using IM, objects are defined by features derived from an image blob. This generalizes well to many different types of scenes besides traffic, such as pedestrians, by defining all objects by the same set of simple measurements.

IV. Classification Methods

The image based feature vector is rather large and prohibits robust classification both computationally and because of the lack of adequate training data to properly fill a high dimensional feature space. Two different dimensionality reduction techniques, PCA and LDA, were used to ensure manageable classification. The same dimensionality techniques are used on the IM features, not for dimensionality reduction but to remove redundant information and to project the data into a space better suited for classification [8].

A. Principle Component Analysis

As a dimensionality reduction technique, PCA aligns data along the directions of greatest variance. Let $\chi = \{x_1, x_2, \ldots, x_N\}$ be a training set of N vectors each of dimension d. Construct the training matrix $\bar{X} = [x_1 - \mu, \ldots, x_N - \mu]$, where $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$. The PCA projection is found by solving the eigen problem

$$\bar{X}\bar{X}^T e = \lambda e. \tag{1}$$

Notice the rank of $\bar{X}\bar{X}^T$ is the minimum of d and N. When $N \ll d$ this decomposition can be computed more efficiently by solving

$$\bar{X}^T\bar{X} f = \gamma f. \tag{2}$$

The eigenvector e is then easily found by noting that multiplying both sides of (2) by $\bar{X}$ gives us (1),

$$\begin{aligned} \bar{X}^T\bar{X} f &= \gamma f \\ \bar{X}\bar{X}^T(\bar{X} f) &= \gamma(\bar{X} f) \\ \bar{X}\bar{X}^T e &= \gamma e, \end{aligned}$$

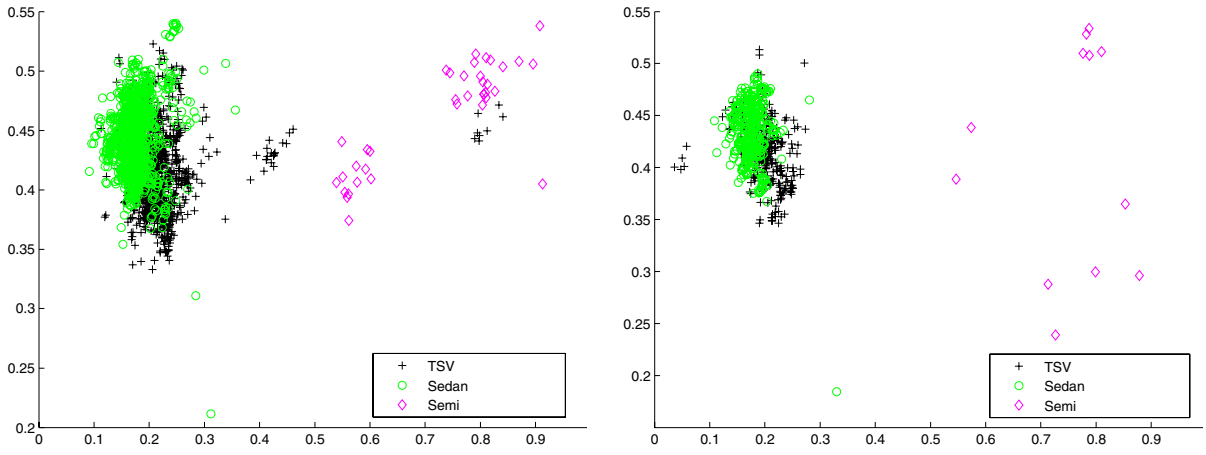
with $e = \bar{X} f$ (e must be normalized such that $\|e\| = 1$). The PCA projection, P_{PCA} is then formed from the M eigenvectors with largest corresponding eigenvalue. The new feature vector

$$x_{PCA} = P_{PCA} x = [e_1, \ldots, e_i, \ldots, e_M]x$$

is of dimension M.

B. Linear Discriminant Analysis

While PCA puts emphasis into retaining directions of large variance, it does not take into account the actual classes. Since large variance is not always best for classification, LDA tries to project onto a subspace of best discrimination by maximizing the separation between classes relative to separation within classes. Let $\chi_c = \{x_1, \ldots, x_{N_c}\}$ be a set



(a) Training Data (b) Test Data

Fig. 2. LDA projection of image measurement (IM) data. Sedan and semi classes are well separated but there is overlap between SUV and Truck classes.

of N_c training vectors for class c, each of dimension d, with mean $\mu_c = \frac{1}{N_c}\sum_{i=1}^{N_c} x_i$. The full training set, $\chi = \{\chi_1, \ldots, \chi_C\}$, is composed of the training samples from all classes and has mean $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$, where $N = \sum_c N_c$. The LDA projection is given by the maximization problem

$$P_{LDA} = \underset{x}{\operatorname{argmax}} \frac{|w^T S_B w|}{|w^T S_W w|} \tag{3}$$

where S_B is the between class scatter and S_W is the within class scatter matrices.

$$S_B = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^T$$
$$S_W = \sum_{i=1}^{C} \sum_{x_k \in \chi_C} (x_k - \mu_i)(x_k - \mu_i)^T$$

The solution to this maximization leads to the generalized eigen problem $S_B w = \lambda S_W w$. Again the top M eigenvectors are retained to obtain the LDA projection matrix,

$$x_{LDA} = P_{LDA} x = [w_1, ..., w_i, ..., w_M] x \tag{4}$$

C. NN Derivative Classification

In this experiment a weighted K nearest neighbor rule (wkNN) [6] was used to classify a transformed feature vector into a vehicle class. This a modification of the nearest neighbor (NN) classifier. The advantage of wkNN is that each sample is assigned to every class by a class weight (5) while NN only gives a binary indication of class membership. This class weight is a soft membership to each class, which builds robustness to noise and outliers. The L_2 norm was used as the distance metric to determine the similarity between vectors. The wkNN weight for each class indicates the strength of match and a label (6) is assigned corresponding to the class with the highest weight (7).

$$W_c = \sum_{\substack{i=1 \\ x_i \in \chi_C}}^{K} \frac{1}{\|x_i - x\|} \tag{5}$$

$$L(x) = \underset{c}{\operatorname{argmax}}\, W_c \tag{6}$$

$$W(x) = \max_{c} W_c \tag{7}$$

V. Track Based Refinement of Classification

After evaluating the performance of the different classifiers, the LDA-IM classifier was chosen for integration into tracking software. Using LDA-IM generates a simple classifier with low computational complexity and that generalizes well because of scene object independence. The system level block diagram follows in Fig. 3. Potential vehicles are detected by the *Object Detection* module which used an adaptive background subtraction scheme [9]. Taking the difference between the current video frame and the estimated background produces regions of moving objects. These regions are processed to produce vehicle blob detections. The detections are normalized by a homography transformation with respect to the freeway ground plane [10]. This transformation of the video makes the lanes of the freeway parallel and removes perspective distortion in the scene allowing comparison between vehicles in the closest lanes to those in the farthest lanes from the camera. Tracking is accomplished by using a Kalman filter on the center of mass of the detected object region [11]. The *Kalman Filter* module outputs a state vector, $[x, v]^T$, containing the position and velocity of the region. The Kalman filter is a state estimation tool that predicts the position of a vehicle in the next frame. The predicted state is used for data association by the *Track Builder* module to connect individual vehicles in consecutive frames into a track. Each region is also sent to the *Car Type Classifier* module to be classified. Each object at time t is assigned a soft class membership by normalizing

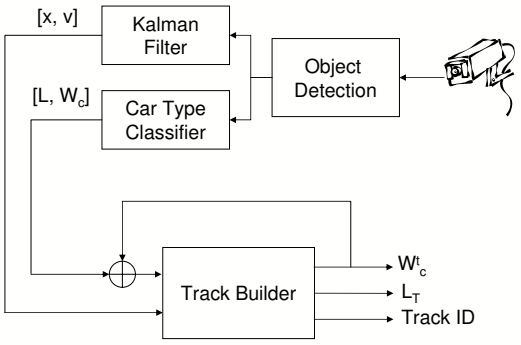


Fig. 3. Block Diagram for Tracker with Classification

TABLE I
IMAGE AND MEASUREMENT BASED CLASSIFICATION ACCURACY %

Classifier Type	TSV	Sedan	Semi	Total
Image Based Features				
PCA	69.45	88.89	100	80.36
LDA	74.91	52.78	25.00	62.19
Measurement Based Features				
PCA	74.65	83.02	86.11	79.50
LDA	65.09	85.49	91.67	76.43

the class weights (5),

$$W_c^t = \frac{W_c}{\sum_c W_c}. \tag{8}$$

This soft classifier deals with the uncertainty of producing one class label by allowing a test vehicle to be a member of every vehicle class with the class with largest weight, W_c^t, being the most likely true class. The *Track Builder* module builds vehicle tracks from the Kalman state, assigning each track a Track ID and a single class label to the track. The track vehicle label is determined by building a histogram [12] of class weights for each frame in a track T and selecting as label the class with the highest membership,

$$L_T = \underset{c}{\operatorname{argmax}} \sum_{t \in T} W_c^t. \tag{9}$$

By binning class the soft class membership into a track histogram the *Track Builder* is able to recover from misclassified examples by only assigning a final label as the most likely class along the entire track. By using an entire track, as much information as available in the video is being utilized for classification. The classification is no longer produced from a single image of a vehicle but a collection of complementary images.

VI. EXPERIMENTAL STUDY

A. Individual Image Classification

The classification results comparing image based and measurement based features are presented in Table I. The results are separated into image based and the measurement based techniques. All data was compiled from the output of the background subtraction based tracker. The tracker was run on a video sequence to obtain 1836 {TSV, Sedan, Semi} = {825, 974, 37} training examples and 611 {275,324,12} test examples. The classification accuracy rates are given for each vehicle class individually and a total for the entire dataset.

Looking at the image based results we see the best performance came using PCA in contrast to the face classification results from [7]. LDA performed worse than PCA even though it is designed for discriminability. Though the training set is well separated, applying the same transformation to the test set produced little separation. The top LDA basis cars do not resemble any type of vehicle, hinting at the failure of this procedure. When doing face classification, face images are resized to similar size and registered such that facial features align. This can not be done for vehicles because each class has a different appearance whereas a face is quite similar across all people.

Using the measurement features we achieve classification rates comparable to the image based PCA but with much less computation time, which is critical for a real time implementation. Although measurement based PCA performed better for this three class problem it only showed improvement in the TSV class. Both Sedan and Semi classes were more accurately classified using the LDA transformation. LDA is the preferred method because it will generalize better when adding new classes. In Figures 2(a) and 2(b) we see that even when using LDA there is overlap between classes, particularly between Sedans and TSV. Much of the error can be attributed to this overlap. Though the classification rate for IM LDA is 4% less than IB PCA, the increase in speed justifies its use.

The robustness of LDA IM based classifier was tested on another video sequence from a different day with 867 {360,483,24} test vehicles. These results from day 2 are presented in the last line of Table II. The classification confusion matrix for the two test days is also shown in Table II. The TSV and Sedan are most often confused because of their proximity the LDA feature space. This contributes to the lower classification accuracy of these classes. Accurate

(a) Vehicle 16: Correctly classified as TSV after short oscillation

(b) Vehicle 312: Tracking error combines bus with other vehicles

(c) Vehicle 287: Motorcycle classified as Van

(d) Vehicle 68,69: Track classification quickly recovers after tracking error

Fig. 4. Sample Track Based Classification Results

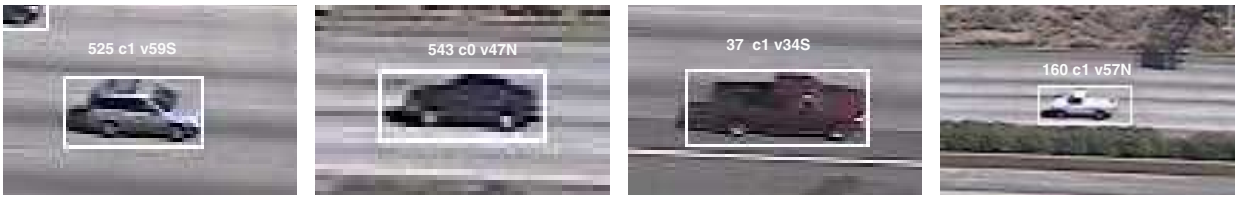
TABLE II
IMAGE MEASUREMENT DETECTION CLASSIFICATION CONFUSION MATRIX DAY 1+2

Day		True TSV	Sedan	Semi	Total
1+2	TSV	483	99	2	
	Sedan	151	708	0	
	Semi	1	0	34	
	Total	635	807	36	1478
	%Correct	76.06	87.73	94.44	82.88
1	%Correct	65.09	85.49	91.67	76.43
2	%Correct	84.44	89.23	95.8	87.43

classification of vehicles can not be accomplished using single images as traditionally done. More information is contained in video that needs to be utilized.

B. Tracking Based Classification

Tracking based classification uses the entire track of a vehicle for classification rather than just an individual frame image as shown previously. Each frame generates an individual example of a vehicle which can be classified more accurately when all occurrences in a track are combined. Table III gives the improved classification results obtained with tracking. Using the tracks we see an improvement of almost 10% in the Total classification rate. The accuracy of every type is improved, with significant improvements for Sedan and TSV. We see much lower confusion between TSV and Sedan as well as the complete removal of errors for the Semi type. The soft class membership 8 allows the *Track Builder* to absorb mis-classifications by only assigning a label as the most likely class along the entire track. In Fig. 4(a) object 16, a white Truck, enters into the camera view from the bottom left. Along the track it is mis-sclassified as Sedan (c1) going south with an estimated speed of 60 mph (v60N). The *Track Builder* is able to recover the correct label by the end of the track, where it is classfied as a TSV (c0) going 51 mph (v51S). The classification can also help improve tracking. In Fig. 4(b) we see a tracking error where a bus was grouped together with two other vehicles. The classification results can help track reasoning after tracking recovery as seen in Fig. 4(d). The breakup of track 68 into two tracks of differing classes indicates that there was either occlusion or grouping of vehicles in earlier frames. Most track based errors arose from TSVs classified as Sedans. Some examples of misclassifications using track based classification are shown in Fig. 5. Adding more classes might be able to better discriminate these missed examples by better characterizing Trucks and SUVs against Sedans. The track based classification results are promising and indicate the value of doing classification over spatio-temporal detections. Even though the individual detection classification may have ambiguities, as is the case for Sedan and TSV (Table II), using tracking greatly improved the



(a) Vehicle 525: SUV Classified as Sedan (b) Vehicle 543: Sedan Classified as TSV (c) Vehicle 37: Truck Classified as Sedan (d) Vehicle 160: Truck Classified as Sedan

Fig. 5. Sample Track Based Classification Errors

TABLE III

TRACK CLASSIFICATION CONFUSION MATRIX DAY 1+2

Day		True TSV	Sedan	Semi	Total
1+2	TSV	211	15	0	
	Sedan	29	321	0	
	Semi	0	0	7	
	Total	240	336	7	583
	%Correct	85.42	96.43	100	91.94
1	%Correct	89.19	98.09	100	94.51
2	%Correct	82.17	94.97	100	89.68

final classification accuracy by taking advantage of expanded information contained in video. Tracking will be crucial in further developing our system to recognize a wider range of vehicles.

VII. CONCLUDING REMARKS

The widespread deployment of cameras for traffic analysis has given researchers plenty of data to develop robust techniques for detection and tracking of vehicles. With tracking well solved, the next advancements will come from high level event recognition and prediction. One of the key tasks necessary for higher level analysis is vehicle classification. A simple classifier was built for speed and generality. The robustness of this simple classifier was improved almost 10% by integrating tracking information.

REFERENCES

[1] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," *Image and Vision Computing*, vol. 21, no. 4, pp. 359–381, Apr. 2003.

[2] G. T. Kogut and M. M. Trivedi, "Maintaining the identity of mulitiple vehicles as they travel through a video network," in *Proc. IEEE Conf. on Intell. Transport. Syst.*, Oakland, California, Aug. 2001, pp. 756–761.

[3] R. Chang, T. Gandhi, and M. M. Trivedi, "Vision modules for a multi-sensory bridge monitoring approach," in *Proc. IEEE Conf. on Intell. Transport. Syst.*, Oct. 2004, pp. 971–976.

[4] W. Hu, X. Xiao, D. Xie, T. Tan, and S. Maybank, "Traffic accident prediction using 3-d model-based vehicle tracking," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 677–694, May 2004.

[5] S. Bhonsle, M. Trivedi, and A. Gupta, "Database-centered architecture for traffic incident detection, management, and analysis," in *Proc. IEEE Conf. on Intell. Transport. Syst.*, Dearborn, Michigan, Oct. 2000, pp. 149–154.

[6] T. K. Osamu Hasegawa, "Type classification, color estimation, and specific target detection of moving targets on public streets," *Machine Vision and Applications*, vol. 16, pp. 116–121, Feb. 2005.

[7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 771–720, July 1997.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY: Wiley-Interscience, 2001.

[9] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Appoach*. Upper Saddle River, NJ: Prentice Hall, 2003.

[10] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer, 2005.

[11] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik, "A real-time computer vision system for measuring traffic parameters," in *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, San Juan, Puerto Rico, June 1997, pp. 495–501.

[12] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," in *Proc. Fourth IEEE Workshop on Applications of Computer Vision*, Princeton, New Jersey, Oct. 1998, pp. 8–14.