21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Structural Signatures for Passenger Vehicle Classification in Video

Ninad Thakoor and Bir Bhanu
Center for Research in Intelligent Systems, University of California, Riverside, USA
$\{$*nthakoor,bhanu*$\}$*@cris.ucr.edu*

Abstract

In this paper, we focus on a challenging pattern recognition problem of significant industrial impact: classifying vehicles from their rear videos as observed by a camera mounted on top of a highway with vehicles traveling at high speed. To solve this problem, we present a novel feature called structural signatures. From a rear view video, a structural signature recovers the vehicle side profile information which is crucial in its classification. As a vehicle moves away from a camera, its surfaces deform differently based on their relative orientation to the camera. This information is used to extract the structure of the vehicle which captures the relative orientation of vehicle surfaces and the road surface. We present a complete system which computes the structural signatures and uses them for classification of passenger vehicles into sedans, pickups and Minivans/SUVs in highway videos. We analyze performance of the system on a large dataset.

1. Introduction

Vehicle classification is an important industrial application of the pattern recognition technology. Vehicle class information can be useful in traffic analysis, security applications, surveillance tasks, and law enforcement. For a vehicle classification system to be useful in a real world application, it must be robust to illumination changes, shadows, partial detections, occlusion, tracking failure, imaging system changes, camera viewpoint changes etc. Current vehicle classification methods which rely on blob features or appearance features cannot meet these requirements.

Gupte *et.al.*[1] use vehicle dimensions to classify their side views in real-time, however the classification is only limited to sedans or non-sedans. Ma and Grimson [3] classify sedans vs. taxis and sedans vs. minvans in their edge based approach with constellation model. They use oblique side views of vehicles in their work. A side view of a vehicle can be easily occluded on multi-lane roads. Additionally, most of the cameras deployed along the road capture rear or front views of the vehicle, reducing the applicability of side view based techniques. Morris and Trivedi [4] also use side views of vehicles and blob features to classify vehicles and suffer from the shortcomings mentioned above. In [2], Kafai and Bhanu use a hybrid dynamic Bayesian network to classify rear views of vehicles. They use features such as locations and dimensions of landmarks (e.g. license plates, tail lights) as well as their spatial relationships in the network. Detection of these high-level landmarks is challenging under varying environmental conditions.

Other class of vehicle recognition focuses on recognizing make-and-model of the vehicles. Petrovic and Cootes [7] use square mapped gradients of frontal views of vehicles to identify their make-and-model. Clady *et.al.* [5] use oriented contour features of frontal views to classify vehicles. Pearce and Pears [6] use a recursive partitioning scheme with Harris corner features to identify the class of vehicles. All these approaches use appearance information which can change widely under varying environmental conditions. Applicability of these approaches in real world scenarios is thus limited.

All these current methods either capture the appearance or blob structure of the vehicle. None of these methods use structure information that can be inferred from multiple views in the classification. In view of the state of the art, the contributions of this paper are:

1. Development of a vehicle classification system from the rear view video of the vehicle which separates vehicles in 3 classes: sedan, pickup and Minivan/SUV.
2. Introduction of a novel feature called structural signatures to capture side profiles of vehicles from rear.
3. Integration of information from multiple video frames in the signature computation as well as classifier decision making.
4. Validation with 879 real-world videos of vehicles.

2. Motivation

Figure 1 shows vehicle rear views from various categories. The canonical vehicle surfaces visible from its rear view are either almost parallel or perpendicular to the road. As seen in the top row of Figure 1, these surface orientations alone are discriminative enough to separate sedans, pickups and minivans/SUVs. The goal of structural signatures is to capture these surface orientations reliably with minimal computational effort.

From the rear view, the structure of the vehicle can be characterized along any vertical axes. The structure is almost the same towards the center of the images. We

978-4-9906441-1-6

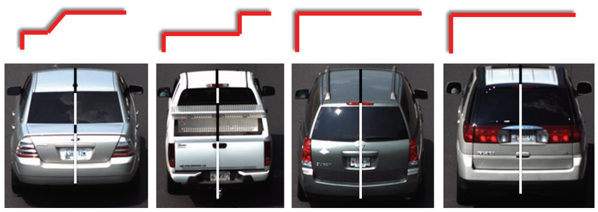

Figure 1: Top: Canonical structure, Bottom: Rear views (Black line: parallel to the road; White line: almost perpendicular to the road)

choose to encode the structure along the axis of bilateral symmetry of the vehicle as: (a) All types of passenger vehicles exhibit strong bilateral symmetry from the rear. (b) The axis of symmetry is robust to partial detection of the vehicle as well as to spurious regions such as shadows being detected as a part of the vehicle. (c) It is robust to illumination variation, body color change, image resolution change, etc. Thus, the axis of symmetry can be detected consistently and reliably.

3. Technical Approach

By analyzing motion of an object with time, its structure can be recovered. While this is the general principle of structure-from-motion approaches, we would like to reduce the complexity of the solution by imposing additional constraints of our problem.

3.1. Principle of the Technique

As the vehicle surfaces hold a constant relationship with the road independent of the camera, we choose to analyze the surfaces with respect to their road projection instead of their image projections.

Let an object on a plane be imaged by a camera. The object is moving away from the camera. The motion of the object is negligible in the image x direction. This adequately describes a vehicle moving on a road being captured from the rear by a camera mounted above the road. For simplicity, the object is assumed to be a cuboid. The side view of the object at a time instance t is shown in Figure 2. We analyze the horizontal and vertical faces of the cuboid represented by P_1P_2 and P_1P_3 respectively.

Theorem 1. *The height of the projection of the surface parallel to the road does not change with time.*

The projection of the surface P_1P_2 at time t be $\alpha(t)d$. Since $\triangle ODP_1(t) \sim \triangle OP_0P_1'(t)$, $\triangle ODP_2(t) \sim \triangle OP_0P_2'(t)$. Using properties of similar triangles:

$$\frac{h-\delta}{h} = \frac{DP_1(t)}{P_0P_1'(t)} = \frac{DP_1(t)+d}{P_0P_1'(t)+\alpha(t)d} = \frac{1}{\alpha(t)}. \quad (1)$$

As $\alpha(t)$ is a constant, the height of the projection of the surface parallel to the road does not change with time. $\square$

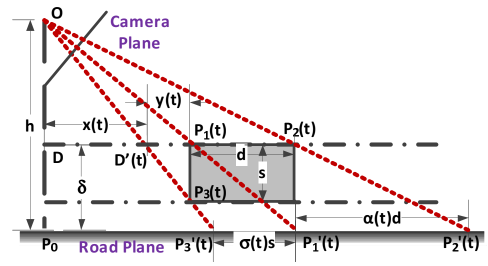

Figure 2: Projection on the road

Theorem 2. *The height of the projection of the surface perpendicular to the road changes with time.*

The projection of the vertical line P_1P_3 at time t be $\sigma(t)s$. Since, $\triangle ODP_1(t) \sim \triangle OP_0P_1'(t)$, $\triangle ODD'(t) \sim \triangle OP_0P_3'(t)$, we get,

$$\frac{P_0P_3'(t)}{\rho s} = \frac{x(t)}{y(t)}, \quad (2)$$

As, $\triangle ODD'(t) \sim \triangle P_3(t)P_1(t)D'(t)$.

$$\frac{x(t)}{y(t)} = \frac{h-\delta}{s}. \quad (3)$$

From (2) and (3), $\sigma(t) = P_0P_3'(t)/(h-\delta)$. As $\sigma(t)$ changes with time, the height of the projection of the surface perpendicular to the road changes with time. $\square$

Based on the properties of vertical and horizontal surfaces, we develop an approach to compute the structural signatures which is used in the vehicle classification system shown in Figure 3.

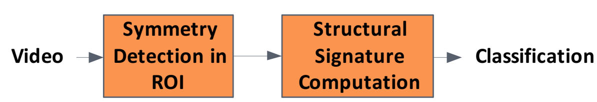

Figure 3: System overview

3.2. Symmetry Detection

Before a vehicle can be classified, it has to be detected from a video. For each region of interest (ROI) selected by the moving object detection, a bilateral axis of symmetry is established. The axis is established through a voting scheme. Given the orientation of the ROI, the axis of symmetry is assumed to be vertical, i.e., it corresponds to one of the ROI columns.

For a candidate axis location corresponding to the jth column of the ROI R with the edge magnitude E and their quantized orientations O, the votes are counted as,

$$V(j) = \sum_{\forall i, j^-, j^+ : (i,j^+) \in R, (i,j^-) \in R} v(i, j^-, j^+) \quad (4)$$

where

$$v(i, j^-, j^+) = \begin{cases} \min(E_{i,j^-}, E_{i,j^+}), & O_{i,j^-} = O'_{i,j^+}; \\ 0, & \text{Otherwise.} \end{cases}$$

Additionally, $j^+ = j + \Delta$, $j^- = j - \Delta$ and $O'_{i,j} = \pi - O_{i,j}$. For the candidate axis location j, Δ takes values from 1 to $\min(j, \text{width}(\text{ROI}) - j)$. The axis of symmetry is assigned to the column with the highest number of votes. A small rectangular *template* around the axis of symmetry is selected and it is tracked in subsequent frames.

3.3. Structural Signature Computation

To establish the structural signature between two frames i and j, only the templates from i to j are needed to be analyzed. First a row-to-row correspondence between the templates is established. This can be achieved in two ways: one, by matching the template pairs of adjacent frames and then propagating the matches; two, by performing multi-frame matching on all the templates in a single operation.

Before structural signatures for the vehicle can be generated, candidates for vehicle surfaces are generated in the ith frame. This is achieved by splitting the template in N *surface elements* of equal heights. The rows corresponding to the edges separating the surface elements be $E_1, E_2, \ldots, E_N, E_{N+1}$. Each adjacent pair of edges (E_n, E_{n+1}) forms a surface element $S_{n,n+1}$. The projection of this surface element on the road is $P_{n,n+1}$ which is computed using the camera to road homography. For a pair of frames (i, j), the structural signature can be computed as,

$$S_{i,j} = \left(\frac{P^i_{1,2} - P^j_{1,2}}{P^i_{1,2}}, \ldots, \frac{P^i_{N-1,N} - P^j_{N-1,N}}{P^i_{N-1,N}} \right) \quad (5)$$

which represents the normalized changes in the height of surface projections.

4. Experimental Results

• **Video Data:** The proposed system was validated with videos recorded at two freeway locations over several days. The camera was setup on top of a freeway lane at about 22 feet with depression angles of 8 to 10 degrees capturing more than 200 feet of the lane. Videos were captured at the 1600×1200 resolution at 12fps allowing for more than 15 frames with complete view of vehicles traveling at freeway speeds ($\sim$60mph). Figure 4 shows an example frame from these videos.

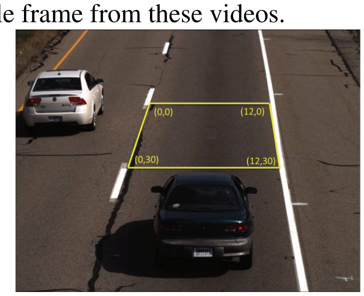

Figure 4: Example frame

For each camera view used in the video, a camera to road homography was established by detecting lane separation markings. These markings are typically a 10' white strip followed by a 30' gap. Using the lane width of 12', two additional points were located on the solid white lane (See Figure 4) to estimate the homography.

• **Symmetry Detection:** Vehicles were detected at the 400×300 resolution with a moving object detection technique using combination of frame difference and optical flow. Detected vehicles which did not lie in the right side lane or were not entirely in the frame were discarded. The full resolution ROI was then processed to compute the structural signature. In the very first frame where the vehicle was completely visible, the axis of symmetry was established. The edge orientations and magnitudes to vote for the axis of symmetry were found with Gabor filters with orientations $0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$.

• **Structural Signatures:** Centered around the detected axis of the symmetry, a template of width 7 pixels and height same as the ROI was extracted. The vehicle was tracked in the next frame using this template. The template was then updated to the matched region in the next frame. The tracking was continued till the vehicle exited the view. From the tracked templates, a row-to-row correspondence was established by either using frame-to-frame or multi-frame matching. After comparing the performance for various number of surface elements, their number N was fixed at 10.

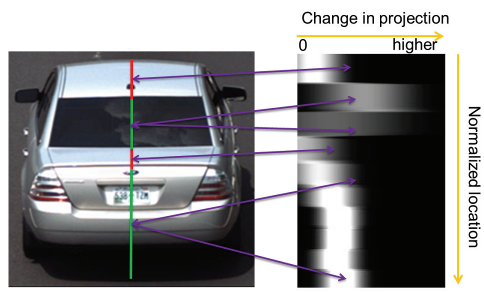

Figure 5: Structural signature occurrences for sedans (Right: Brighter means more frequent; darker means less frequent)

For each of the vehicles, the structural signature was then established using (5). Figure 5 shows a histogram like representation of structural signatures for sedans. As expected, values close to zero are observed for vehicle top and trunk top which are parallel to the road. For other parts of the vehicles, nonzero values are more frequent in the structural signatures. Thus, structural signatures capture the canonical structure of the vehicles.

• **Classification:** For classification, initially SVM and edit distance based classifiers were compared. As the SVM classifier outperformed the accuracy of the edit distance classifier (65.4%) significantly, an SVM classifier was chosen for subsequent experiments. The signatures were used to train an SVM classifier with a radial basis function kernel. We used two fold cross validation to obtain the classification accuracies.

Figure 6 shows classification accuracies when vary-

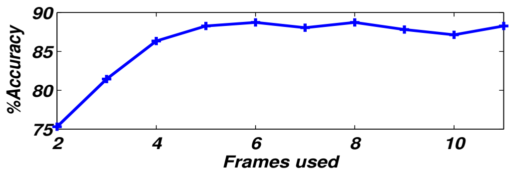

Figure 6: Number of frames used vs. Accuracy

ing number of frames are used to compute the structural signatures. When the number of frames used is low, the amount of evidence collected is low leading to lower quality of signatures and lower classification accuracy. However, with only 5 frames the classifier performance levels out at about 88%. This is possibly due to the tracking errors introduced which negate the evidence being added.

• **Comparison with Baseline and Multi-frame classification:** We also carried out experiments with variations of our approach. We compared these variation versus the single best classifier (88.7% accuracy) which carries out multi-frame matching on 6 frames and uses (5) to compute the signatures. The classification accuracies for the variations are shown in Table 1.

Table 1: Variation vs. Accuracy: MFn indicates multi-frame matching with n frames, FFn indicates frame-to-frame matching with n frames.

Variation	%Accuracy
Baseline: No road projection (MF6)	84.3
No normalization (MF6)	86.4
Frame to Frame tracking (FF6)	87.5
Best single (MF6)	88.7
Voting (MF2-MF11)	90.1
Weighted (MF1-MF11)	90.3

We establish the baseline accuracy by using the structural signature computed directly using the image coordinates. Since the camera settings are similar for most of the videos, the baseline structural signatures perform reasonably well giving accuracy of about 84%. When (5) is modified to remove the normalization factor in the denominator, the performance of the classifier drops to 86.4%. When frame-to-frame matching is used, the matcher cannot recover from matching failure in intermediate frames. This leads to reduced performance of 87.5%. Finally, we fuse the individual classification results shown in Figure 6. A simple voting based classifier which assigns the most frequent class as the true class, results in accuracy of 90.1%. A weighted voting, where each classifier vote is weighted according to its training accuracy, results in slightly improved performance of 90.3% as it avoids voting ties.

• **Discussion:** Table 2 gives the confusion matrix for the weighted classifier. Pickups have the lowest classification accuracy as their beds can carry items which can deform their structural signature leading to incorrect classification. On the other hand, minivans and SUVs have the simplest structures and this results in the highest accuracy. Additionally, some sedans such as hatchbacks and some pickup trucks which carry camper shells have structural signature similar to Minivans/SUVs which result in misclassifications.

Table 2: Confusion matrix

Class→ Decision↓	Sedan	Pickup	Minivan/ SUV
Sedan	270	12	8
Pickup	3	159	6
Minivan/SUV	38	18	365
%Accuracy	86.8	84.3	96.3

5. Conclusions

We presented structural signature features for classification of rear view videos of vehicles. It used information from multiple video frames to infer the vehicle structure unlike current state of the art approaches which either use blob features or appearance features from frame-to-frame. The structural signatures are independent of the appearance which makes them less susceptible to illumination changes and imaging system variations. Use of the road projection allows significant variations in camera angles. Incorporating symmetry makes our system robust against shadows, partial detections and occlusions. The proposed system uses computationally inexpensive techniques such as change detection, edge voting based symmetry, template tracking to realize the structural signatures. While our OpenCV based preliminary C++ implementation runs at 20fps, further optimizations will make real-time implementation viable on general purpose computation platforms such as computers, GPUs and DSPs.

6. Acknowledgment

This work was supported in part by a grant from Federal Signal Corporation.

References

[1] S. Gupte, O. Masoud, R. Martin, and N. Papanikolopoulos. Detection and classification of vehicles. *IEEE Trans. Intel. Trans. Sys.*, 3(1):37 –47, Mar. 2002.

[2] M. Kafai and B. Bhanu. Dynamic bayesian networks for vehicle classification in video. *IEEE Trans. Ind. Inform.*, 8(1):100 –109, Feb. 2012.

[3] X. Ma and W. Grimson. Edge-based rich representation for vehicle classification. In *Proc. IEEE ICCV*, 2005.

[4] B. Morris and M. M. Trivedi. Learning, modeling, and classification of vehicle track patterns from live video. *IEEE Trans. Intel. Trans. Sys.*, 9(3):425 –437, Sept. 2008.

[5] P. Negri, X. Clady, M. Milgram, and R. Poulenard. An oriented-contour point based voting algorithm for vehicle type classification. In *Proc. ICPR*, 2006.

[6] G. Pearce and N. Pears. Automatic make and model recognition from frontal images of cars. In *Proc. IEEE AVSS*, 2011.

[7] V. S. Petrovic and T. F. Cootes. Analysis of features for rigid structure vehicle type recognition. In *Proc. BMVC*, 2004.