A PATTERN RECOGNITION AND FEATURE FUSION FORMULATION FOR VEHICLE REIDENTIFICATION IN INTELLIGENT TRANSPORTATION SYSTEMS

Ravi P. Ramachandran | *Glenn Arr* | *Carlos Sun* | *Stephen G. Ritchie*

Rowan University
Glassboro, NJ 08028
ravi@rowan.edu

Rowan University
Glassboro, NJ 08028
arr3344@rowan.edu

University of Missouri
Columbia, MO 65211
csun@missouri.edu

University of California
Irvine, CA 92697
sritchie@uci.edu

ABSTRACT

Vehicle reidentification is the process of reidentifying or tracking vehicles from one point on the roadway to the next. By performing vehicle reidentification, important traffic parameters including travel time, section density and partial dynamic origin/destination demands can be obtained. This provides for anonymous tracking of vehicles from site-to-site and has the potential for improving Intelligent Transportation Systems (ITS) by providing more accurate data. This paper presents a new vehicle reidentification algorithm that uses four different features, namely, (1) the inductive signature vector acquired from loop detectors, (2) vehicle velocity, (3) traversal time and (4) color information (based on images acquired from video cameras) to achieve high accuracy. A nearest neighbor approach classifies the features and linear feature fusion is shown to improve performance. With the fusion of four features, more than a 91 percent accuracy is obtained on real data collected from a parkway in California.

1. INTRODUCTION

The Bureau of Transportation Statistics (BTS), in April of 2001, presents the following figures quantifying the transportation problems in the United States [1]. The average annual person-hours of delay per eligible driver for the twenty most congested cities in the U.S. is 57 hours or 2.4 days in 1997. In terms of fuel consumption, the aforementioned delays translate into over 5 billion gallons of wasted fuel in 1997. BTS also estimates that annual congestion cost for the twenty most congested cities was over 57 billion dollars in 1997. The statistics for the years leading up to 1997 were also similar according to the BTS report. Over the previous decade, the average number of transportation fatalities was over 40,000 per year with highway modes claiming the most fatalities.

Intelligent Transportation Systems (ITS) is hailed by many to be a major contributor in the improvement of our transportation system in conjunction with other traditional methods [2]. One way of classifying the benefits of ITS is through the use of seven "E"s: efficiency, energy, environment, economics, education, enhanced safety, and enforcement. One of the most critical components in the success of ITS is the eyes of the system or intelligent surveillance. Without thorough and accurate knowledge of existing conditions on our transportation networks, it is impossible to adequately develop strategies that will optimize our transportation system.

The vehicle reidentification problem is the task of matching a vehicle image detected at one location (upstream) with the image generated by the same vehicle detected at a downstream location at some later time. In other words, it is the tracking of vehicles from point to point along the transportation network. In accomplishing this reidentification process, we first acquire the different features (like inductive vehicle signatures and attributes from color information). Second, identification or classification is performed using a nearest neighbor classifier and linear feature fusion (pattern recognition framework) [3]. Inductive vehicle signatures are unique deviations in the inductance of a loop detector caused by the passage of a vehicle. Color information from video cameras is also used since it is uncorrelated with signature information and can be extracted from imperfect video images.

Our algorithm is the basis for development of a multi-detector fusion system for intelligent surveillance. Automatic vehicle reidentification for intelligent surveillance has tremendous practical traffic applications. The derivation of section travel times and densities are useful to transportation engineers for the purpose of traffic operations, planning, and control. The travel time is the time taken by a vehicle to go from one point to another. The density is the number of vehicles passing through a section of roadway over a fixed period of time. Accurate travel times and densities can be instrumental in feedback control, vehicle routing, traffic assignment, dynamic origin/destination demand estimation, and traveler information systems.

Previous investigations using inductive loop detectors for vehicle reidentification seek to correlate vehicle signature patterns, lengths or aggregate traffic parameters [4][5]. The Karhunen-Loeve transform on the vehicle signatures has been attempted in [6]. A freeway control system using a dynamic traffic flow model and vehicle reidentification technique is the subject of [7]. A lexicographical optimization for vehicle reidentification on freeways is discussed in [8] for which a 78% accuracy is obtained. The reidentification accuracy is the number of vehicles identified correctly divided by the total number of vehicles assessed and is expressed as a percent.

The approach in [8] performs reidentification by matching individual vehicles. However, there is valuable traffic flow information that can be incorporated into our new reidentification algorithm to improve the accuracy. The new algorithm uses the fact that vehicles tend to travel in groups or platoons. Platoon in this context refers to a group of vehicles in chronological sequence in close proximity to each other. The new algorithm is a more general formulation since the limits of vehicle platoons are individual vehicles (size of platoon is one) in the least congested case and all vehicles (size of platoon is very large) in the extremely congested case. By comparing multiple vehicles instead of individual vehicles the reidentification accuracy is improved while individual vehicle matches are still maintained. We have experimentally determined that the best size of a platoon is three vehicles.

0-7803-7402-9/02/$17.00 ©2002 IEEE

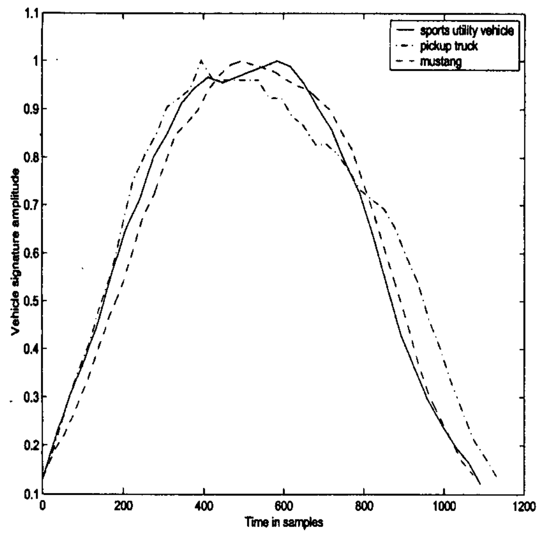

Fig. 1. Examples of vehicle signatures of a sports utility vehicle, a pickup truck and a mustang (a car)

2. FEATURE EXTRACTION

An intrinsic step in the pattern recognition algorithm formulation of the vehicle reidentification problem is the process of feature extraction. The next step is classification which is described later. Feature extraction seeks not only to use vehicle signatures but also color information and other salient data that would sufficiently differentiate vehicles. The first feature vector we use is the vehicle signature itself (denoted as $\mathbf{s}$). For both the upstream and downstream locations, there are two inductive loops each recording a signature. Since the two signatures are almost identical, only one of them is used for vehicle reidentification. The chosen vehicle signature vector is transformed to be speed invariant and is re-interpolated as equally spaced samples of the original acquired signature. The second feature we use is the vehicle velocity v (a scalar feature) and is computed as the distance between the two inductive loops divided by the turn on times of the two loops. The purpose of using two loops is in getting the feature v as opposed to $\mathbf{s}$. The third feature is the platoon traversal time p (a scalar). The quantity p is the difference between the time the last vehicle in the platoon crosses an inductive loop and the time the first vehicle in the platoon crosses the inductive loop. Figure 1 shows examples of the vehicle signatures ($\mathbf{s}$) of three different types of vehicles, namely, a sports utility vehicle, a pickup truck and a mustang (a car). The differences in the signatures allows for reidentification.

The color information is a feature vector $\mathbf{c}$ formed as follows. The video image of the vehicle is transformed into JPEG format with each pixel having a red-green-blue (RGB) value ranging from 0 to 255. There are a total of 256^3 possible RGB triplets each indicating a particular color and shade. For a particular image, each component of $\mathbf{c}$ corresponds to the percentage of pixels having a particular RGB value. This implies that the dimension of $\mathbf{c}$ is prohibitively high at 256^3. Instead of using every possible RGB triplet, the colors are quantized or grouped into subsets. Pixels with colors that are in the neighborhood of each other are grouped into a single triplet. This process helps to improve reidentification accuracy since the aggregated space is more tolerant to noise. Quantization of the RGB values to a level of 5 in order that each pixel has an RGB value from 0 to 4 gave the best reidentification accuracy. Quantization levels up to 30 were tested. Now, there are $5^3 = 125$ RGB triplets and the dimension of $\mathbf{c}$ is 125. Each component of $\mathbf{c}$ corresponds to the percentage of pixels having a particular quantized RGB value.

3. FUSION AND CLASSIFICATION APPROACH

The vehicle reidentification problem is the task of matching a vehicle detected at an upstream location with the same vehicle detected at a downstream location at some later time. In doing the matching, we use the L_1 distance measure between the upstream feature $\mathbf{f}_u$ and the downstream feature $\mathbf{f}_d$ as given by

$$d(\mathbf{f}_u, \mathbf{f}_d) = \sum_{i=1}^{m} |\mathbf{f}_u(i) - \mathbf{f}_d(i)| \tag{1}$$

where i denotes the ith component of the feature vector and m is the vector dimension. From Fig. 1, it is observed that the number of components of the signatures may be different for different vehicles. In this case, the vector with fewer components is padded with values of 0.12 before taking the L_1 distance. If the size of the platoon is denoted as N_p, the L_1 distance for the overall platoon, D_p, is

$$D_p = \sum_{j=1}^{N_p} d(\mathbf{f}_u^j, \mathbf{f}_d^j) \tag{2}$$

where $\mathbf{f}_u^j$ and $\mathbf{f}_d^j$ are the upstream and downstream features for vehicle j. Fusion of the four features is performed to get an overall fusion distance D given by

$$\begin{aligned} D &= w_s \sum_{j=1}^{N_p} d(\mathbf{s}_u^j, \mathbf{s}_d^j) + w_c \sum_{j=1}^{N_p} d(\mathbf{c}_u^j, \mathbf{c}_d^j) \\ &+ w_v \sum_{j=1}^{N_p} d(v_u^j, v_d^j) + w_p d(p_u, p_d) \end{aligned} \tag{3}$$

where w_s is the fusion weight applied to the vehicle signature distance, w_c is the fusion weight applied to the color information feature, w_v is the fusion weight applied to the velocity feature and w_p is the fusion weight applied to the platoon traversal time feature. As before, the subscripts u and d refer to upstream and downstream, respectively. Also, the superscript j refers to the jth vehicle in the platoon. Note that the platoon traversal time feature applies to the entire platoon and not for any individual vehicle. The fusion weights add up to one and are determined during training as described in the next section. This fusion strategy is known as linear fusion and has been shown to be effective in speaker recognition (recognizing a speaker from his or her voice samples) [9]. The distance D between each candidate upstream platoon and a detected downstream platoon is computed. The upstream platoon that achieves the smallest D is matched to the downstream platoon. A nearest neighbor classification approach is used [3]. The final step is to match individual vehicles within the already matched upstream and downstream platoons. This is again done by a linear fusion of the four features.

4. METHODOLOGY AND RESULTS

The traffic data used for the study was collected on June 30, 1998 in Irvine, California. The data site consists of an upstream and downstream detector station bounding a two-lane section of Alton Parkway within the intersections of Telemetry and Jenner streets. Each detector station has double inductive loops in a speed trap configuration and 3M Canoga detector cards. The distance between the two detector stations is 130 m (425 ft). The inductive loops are standard 1.83 m by 1.83 m (6 ft x 6 ft) rectangular loops that are commonly used by many transportation agencies. The data are collected during the morning peak between approximately 8:00 AM and 9:30 AM. This dataset contains 581 vehicles. The first 100 vehicles are used for training. The remaining 481 are used for testing and performance evaluation.

The video collection setup consists of four video cameras recording two lanes of traffic in each of the upstream and downstream locations. From this continuous video footage, one can visually identify many of the v ehicles by type and color. The first step in the data reduction process is to capture the video data into the computer. A video capturing board is used to digitize the video footage into still images stored in JPEG format. The processing algorithm reads each of the still image files and stores the image as a variable of the "C++" image class. This image class contains the RGB (red-green-blue) values of the vehicle image and other information such as the vehicle record number, lane, and time of arrival. The vehicle record number is a unique identification number used to match the video image to the inductive signatures. The RGB color space is used because of its simplicity in representing images. The image class is created with the ability to manipulate the RGB values of each pixel in an image. Each pixel has RGB values ranging from 0 to 255 (8 bit).

Processing the vehicle images involves four main steps; namely, contrast stretching, background subtraction, quantization, and establishment of the feature vector $\mathbf{c}$. The distribution of light and dark pixels is the contrast of an image. Ideally, an image containing a wide distribution of intensities utilizes the full dynamic range. However, if an image is either too light or too dark, some intensities are not utilized and are wasted. During the acquisition of vehicle images, contrast stretching was applied to enhance images.

Subtraction is the process of determining the differences between two images, one that contains a vehicle and a roadway and the other having just the background of the roadway without any vehicle. Background subtraction will produce the image of the vehicle without the surrounding roadway. The details of the subtraction process are as follows. After some investigations, it is verified that the processing algorithm creates all bitmaps of the same pixel height and width. Therefore, the same coordinates in each image should represent the same physical location. A comparison is performed between the two images. First, each pixel in the image containing the vehicle is examined to see if it has the same RGB values as the pixel in the identical location in the background image. If so, that pixel in the image is then set to null. In theory, all of the pixels representing the background should be identical in each of the two images. Therefore, one would expect the resulting background subtracted image to have a vehicle surrounded by null pixels. In actuality, only about 30-40 % of the subtracted image are blacked out, leaving non-vehicle pixels in the image. This problem is due to variability in both the data collection and processing. Such variability can be due to the quantization of a continuous signal into 256^3 or 16.8 millions colors, image changes between the time when the background image and the vehicle image were captured, and other randomness associated with the video hardware setup. Instead of examining a pixel to see if it has the same RGB values, the comparison is relaxed to examine if it has similar RGB values. Now, if the L_1 distance between the RGB vector of the pixel in the image containing the vehicle and the RGB vector of the pixel in the identical location in the background image is below a threshold, that pixel is set to null. The threshold is determined from the training data and is chosen to maximize reidentification accuracy using the color information only. It is found that threshold values between 70 and 85 gave the best performance, thereby indicating that is not sensitive to small changes in the threshold. We continue by choosing a threshold value of 80.

As mentioned earlier, the RGB values are quantized to a level of 5. This gave the best reidentification accuracy using the training data and color information only. Quantization levels up to 30 were tested. There are $5^3 = 125$ RGB triplets and the dimension of $\mathbf{c}$ is 125. Each component of $\mathbf{c}$ corresponds to the percentage of pixels having a particular quantized RGB value.

At a detector station, two vehicle signatures and the velocity feature are acquired. As mentioned earlier, only one of the vehicle signatures is used. Also, the velocity is the distance between the two inductive loops at a station divided by the turn on times of the two loops. Once a platoon is identified, the platoon traversal time, p, is computed. We have also described how the color feature $\mathbf{c}$ is found. Two important issues in implementing the algorithm is to determine the fusion weights and the size of the platoon N_p. All possible fusion weight combinations under the conditions that (1) each weight is between zero and one (inclusive), (2) each weight is incremented in steps of 0.005 and (3) the sum of the weights is one, are applied to the training data of 100 vehicles. The combination that achieves the best reidentification accuracy for the training data is used for testing and performance evaluation. There is no bias in the fusion weights towards the data used for testing. In fact, these weights are found independently of the test data.

A platoon is detected at the downstream site. A list of upstream candidate platoons are generated subject to a time window constraint that eliminates platoons that are not within a reasonable time window. Each upstream platoon is then compared with the downstream platoon by the computation of the distance D (see Eq. (3)). The upstream platoon candidate that most closely resembles the downstream platoon or equivalently, which minimizes D, is selected. To be selected as an upstream candidate platoon, each vehicle in the upstream platoon must have a travel time greater than L_t and less than U_t. The travel time is the time taken by the vehicle to go from the upstream to the downstream detector stations. The quantity L_t is the difference between the travel time of the first vehicle in the downstream platoon and the maximum travel time for all vehicles in the training set. The quantity U_t is the difference between the travel time of the last vehicle in the downstream platoon and the minimum travel time for all vehicles in the training set. In detecting platoons, the two lanes of traffic are treated separately in that lane changes are not considered. This is due to (1) our objective of establishing the feasibility of using pattern recognition and data fusion on a simpler case upon which more complicated scenarios such as overtaking can be later added, (2) our avoiding computational overhead in having a much greater number of feasible upstream platoons again for concept feasibility purposes and (3) the fact that on this relatively short stretches of roadway, lane changes are infrequent. Moreover, we found that lane changes for our data collection site accounted for only 2 percent of the total traffic.

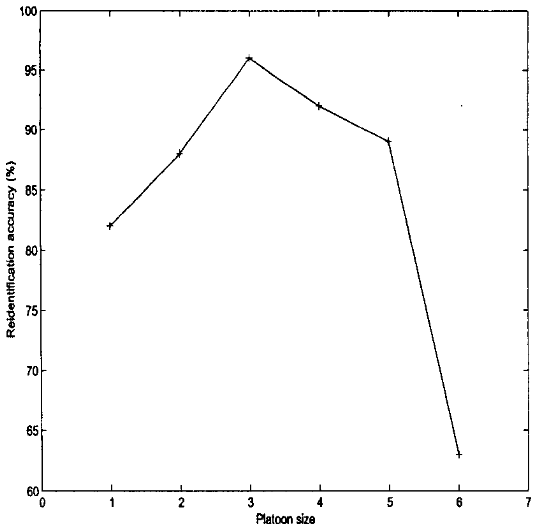

Fig. 2. Vehicle reidentification accuracy versus platoon size for the training data

Vehicle Signature Fusion Weight w_s	Velocity Fusion Weight w_v	Color Fusion Weight w_c	Platoon Traversal Time Fusion Weight w_p	Accuracy (percent) (percent)
1	0	0	0	86.87
0	1	0	0	54.92
0	0	1	0	75.82
0	0	0	1	74.78
0.6	0.4	0	0	87.39
0.005	0	0.995	0	89.29
0.022	0	0	0.978	88.08
0	0.013	0.987	0	76.51
0	0.37	0	0.63	72.71
0	0	0.985	0.015	82.21
0.005	0.025	0.97	0	90.5
0.015	0.425	0	0.56	89.12
0.01	0	0.71	0.28	90.5
0	0.015	0.98	0.005	79.45
0.005	0.015	0.955	0.025	91.36

Table 1. Vehicle reidentification accuracy for the features and all possible fusion combinations using the test data

The platoon size was varied from 1 to 6 and the reidentification accuracy was computed for the training data only. Three of the four features were used in that color information was neglected. Figure 2 shows the results and $N_p = 3$ is the chosen platoon size.

The training data is needed to determine various algorithm parameters as outlined above. We do a performance evaluation by considering the features one at a time and all possible fusion combinations. The test data (481 vehicles not corresponding to the training data) is used. Table 1 gives the results. Fusion of the four features leads to more than a 91 percent accuracy.

5. SUMMARY AND CONCLUSIONS

The feasibility of using a pattern recognition approach and linear fusion for vehicle reidentification is clearly established. With the fusion of four features, more than a 91 percent accuracy is obtained on real data collected from a parkway in California. Fusion of 3 features can result in accuracies of 90.5 percent. Once the vehicle reidentification problem is solved, the derivation of important traffic parameters can be accomplished. Travel times are derived straightforwardly using the difference in the vehicle arrivals times between two points. The derivation of section density is also straightforward by utilizing a counter for the vehicles that have entered the road section and have not reappeared.

6. ACKNOWLEDGEMENT

The research reported in this paper was supported by the California Department of Transportation (Caltrans) and PATH (Partners for the Advanced Transit and Highways). The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of any state agencies. This paper does not constitute a standard, specification, or regulation.

7. REFERENCES

1. Bureau of Transportation Statistics, "National Transportation Statistics", United States Department of Transportation, Report BTS01-01, Washington DC, April 2001.
2. ITE, "Intelligent Transportation Primer", Institute of Transportation Engineers, Washington DC, 2000.
3. R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, John Wiley and Sons, 2000.
4. D. Dailey, "Travel Time Estimation Using Cross Correlation Techniques", Transportation Research Part B, Vol. 27B, No. 2, pp. 97-107, 1993.
5. B. Coifman, "Vehicle Reidentification and Travel Time Measurement in Real-Time on Freeways Using Existing Loop Detector Infrastructure", 77th Annual Transportation Research Board Meeting, Washington DC, January 1998.
6. P. Bohnke and E. Pfannerstill, "A System for the Automatic Surveillance of Traffic Situations", Institute of Transportation Engineers Journal, pp. 41-45, January 1986.
7. R. Kuhne, "Freeway Control Using a Dynamic Traffic Flow Model and Vehicle Reidentification Techniques", Transportation Research Record 1320, pp. 251-259, 1991.
8. C. Sun, S. G. Ritchie, K. Tsai and R. Jayakrishnan, "Use of Vehicle Signature Analysis and Lexicographic Optimization for Vehicle Reidentification on Freeways", Transportation Research Part C, Vol. 7, pp. 167-185, 1999.
9. K. R. Farrell, R. P. Ramachandran and R. J. Mammone "An Analysis of Data Fusion Methods for Speaker Verification", IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Seattle, Washington, pp. 1129-1132, May 1998.