

FUSION OF WAVELET TRANSFORM AND COLOR INFORMATION FEATURES FOR AUTOMATIC VEHICLE REIDENTIFICATION IN INTELLIGENT TRANSPORTATION SYSTEMS

Glenn Arr

Rowan University
Glassboro, NJ 08028
arr3344@rowan.edu

Carlos Sun

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

Ravi P. Ramachandran

Rowan University
Glassboro, NJ 08028
ravi@rowan.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 fusion based vehicle reidentification algorithm that uses four different features, namely, (1) the wavelet transform of 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 92 percent accuracy is obtained on real data collected from a parkway in California. Also, it is found that the wavelet transform improves performance and reduces the dimension of the feature vector when compared to the raw vehicle signatures.

1. INTRODUCTION

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 [1]. One of the most critical components in the success of ITS is the eyes of the system or intelligent surveillance. The vehicle reidentification problem comprises an important aspect of intelligent surveillance. Vehicle reidentification 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. 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.

In accomplishing this reidentification process, we first acquire the different features (like inductive vehicle signatures, the signature wavelet transform and color information). Second, identification or classification is performed using a nearest neighbor classifier and linear feature fusion (pattern recognition framework) [2]. Inductive vehicle signatures are unique deviations in the inductance of a loop detector caused by the passage of a vehicle. The wavelet transform [3] of the signature gives time/frequency information that can discriminate among vehicles. Color information from video cameras is also used since it is uncorrelated with signature information and can be extracted from imperfect video images.

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. A recent algorithm [9] uses the fact that vehicles tend to travel in groups or platoons to improve accuracy. Platoon in this context refers to a group of vehicles in chronological sequence in close proximity to each other. Also, the algorithm in [9] accomplishes multi-detector fusion of inductive signatures and color information. In this paper, the investigation is on the use of the wavelet transform of the inductive signature in an effort to improve reidentification accuracy and ascertain whether certain frequency bands are more important in achieving this accuracy. Fusion of only a subset of the wavelet transform coefficients (corresponding to certain frequency bands) with the color information leads to a better accuracy than using the vehicle signature and simultaneously lowers the feature vector dimension.

2. FEATURE FUSION AND CLASSIFICATION APPROACH

The inductive vehicle signature is a feature vector (denoted as 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 sam-

ples of the original acquired signature. The second feature is the wavelet transform coefficients of \mathbf{s} and is denoted as \mathbf{t} . The vector \mathbf{t} is formed by concatenating the coefficients of each band of the wavelet decomposition. Experiments show that the Haar wavelet [3] is very successful and is hence, used. The third feature 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. Using two loops is necessary in getting the feature v . The fourth 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.

The color information is the fifth 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 [9]. 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.

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

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

where i denotes the i th component of the feature vector and m is the vector dimension. 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. In the case of the wavelet transform, the feature vectors \mathbf{t} are of the same dimension because the signatures are padded to get the same dimension. If the size of the platoon is denoted as N_p , the L_1 distance for the overall platoon, D_p , is

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

where \mathbf{f}_u^j and \mathbf{f}_d^j are the upstream and downstream features for vehicle j . In [9], a platoon size of 3 was found to be the best and hence, we continue to use $N_p = 3$. Also, in [9], a performance of 91.36% reidentification accuracy was obtained by linear fusion of four features to get an overall fusion distance D given by

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

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 j th 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. 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 [2]. 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.

In this paper, a linear fusion strategy of \mathbf{t} (instead of \mathbf{s}) is attempted to improve the performance. The fusion distance D is now given by

$$\begin{aligned} D = & w_t \sum_{j=1}^{N_p} d(\mathbf{t}_u^j, \mathbf{t}_d^j) + w_c \sum_{j=1}^{N_p} d(\mathbf{c}_u^j, \mathbf{c}_d^j) \\ & + w_v \sum_{j=1}^{N_p} d(v_u^j, v_d^j) + w_p d(p_u, p_d) \end{aligned} \quad (4)$$

where w_t is the fusion weight applied to the wavelet transform distance and the other weights are as defined before. The nearest neighbor classification approach is again used.

3. EXPERIMENTAL METHODOLOGY

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 579 vehicles. The first 200 vehicles are used for training. The entire set of 579 vehicles 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 vehicles 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} . 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 described in [9]. As mentioned earlier, the RGB values are quantized to a level of 5. 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.

The training of the system using the first 200 vehicles is performed to determine the (1) threshold for background subtraction of the vehicle image, (2) number of quantization levels for the RGB values of the vehicle image, (3) set of fusion weights and (4) platoon size N_p . These were individually determined to maximize vehicle reidentification accuracy over only the 200 vehicles.

For the testing and performance evaluation, the parameters found during training are simply used without any reoptimization to avoid performance bias. A platoon of size $N_p = 3$ 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) and Eq. (4)). 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.

4. RESULTS

The benchmark for comparing our results is that obtained in [9] in which the fusion of the signature \mathbf{s} , velocity v , platoon traversal time p and the color feature \mathbf{c} was performed. The best fusion weights are $w_s = 0.004$, $w_v = 0.012$, $w_p = 0.092$ and $w_c = 0.892$. The reidentification accuracy is 91.36 percent.

The first experiment is to try different wavelets with varying number of vanishing moments and varying levels of decomposition to see which gives the best accuracy in the absence of fusion. The Daubechies wavelet, the Coiflet and the Symlets were investigated with the number of vanishing moments ranging from 1 to 8 and the number of levels of decomposition ranging from 2 to 20. It was found that the Daubechies wavelet with 1 vanishing moment and 8 levels of decomposition was the best. This is the Haar wavelet [3].

The second experiment is the fusion of the Haar wavelet transform with the other features (except for \mathbf{s}). Table 1 shows the results. The best accuracy is obtained by fusing all four features (92.06 percent) and is an improvement over the benchmark.

The third experiment is the fusion of all five features, namely, \mathbf{s} ,

Wavelet Transform Fusion Weight w_t	Velocity Fusion Weight w_v	Color Fusion Weight w_c	Platoon Traversal Time Fusion Weight w_p	Accuracy (percent)
1	0	0	0	90.67
0	1	0	0	66.32
0	0	1	0	78.76
0	0	0	1	78.41
0.3	0.7	0	0	91.36
0.03	0	0.97	0	91.36
0.28	0	0	0.72	91.36
0	0.02	0.98	0	80.14
0	0.03	0	0.97	77.89
0	0	0.01	0.99	83.07
0.02	0.02	0.96	0	91.36
0.25	0.4	0	0.35	91.88
0.03	0	0.87	0.1	91.54
0	0.024	0.887	0.089	86.18
0.03	0.02	0.85	0.1	92.06

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

\mathbf{t} , v , p and \mathbf{c} . This diminished the performance suggesting that there is much redundancy or correlation in the information represented by \mathbf{s} and \mathbf{t} . The fusion of \mathbf{s} , \mathbf{t} and subsets of the three other features also led to a performance below the obtained 92.06 percent.

The 8 level Haar wavelet decomposition first splits the signature into a lowpass band depicting the frequency range $[0, f_s/4]$ and a highpass band depicting the frequency range $[f_s/4, f_s/2]$. The sampling frequency is f_s . The split is done by filtering and subsampling by 2. The highpass band is referred to as Band 1 and the signal resulting from filtering and subsampling are the wavelet transform coefficients for Band 1. The second level of decomposition splits the lowpass band signal into another lowpass band corresponding to the frequency range $[0, f_s/8]$ and a highpass band corresponding to the frequency range $[f_s/8, f_s/4]$. The highpass band is referred to as Band 2 and the leads to a set of wavelet transform coefficients. The lowpass band is again split and this continues until the 8th level of decomposition is performed. There will be a total of 9 bands with Band i corresponding to the frequency range $[f_s/2^{i+1}, f_s/2^i]$ for $1 \leq i \leq 8$. Band 9 corresponds to the frequency range $[0, f_s/512]$. The feature \mathbf{t} are the transform coefficients of each band, concatenated as one vector. With appropriate padding over the entire database, the vector dimensions of \mathbf{s} and \mathbf{t} are 4272 and 4282, respectively. The dimension of \mathbf{t} is slightly more due to filtering. The fourth experiment is to evaluate the reidentification accuracy of the wavelet transform coefficients for each band without fusion. Table 2 gives the results.

The fifth experiment is to examine if the transform coefficients of only a particular band can be fused with v , p and \mathbf{c} to get a better performance. We perform this experiment with Band 7 only since it gives the best individual performance, along with bands 5 and 6, but contributes a lower vector dimension than bands 5 and 6. It is found that the performance of 92.92 percent is obtained when the transform coefficients for Band 7 is fused with the other features. This is important in that the vector dimension is substantially re-

Band	Vector Dimension	Accuracy Percent
1	2136	91.02
2	1069	91.19
3	535	90.67
4	268	90.67
5	135	91.19
6	68	91.19
7	35	91.19
8	18	90.67
9	18	88.08

Table 2. Vehicle reidentification accuracy for the wavelet features for each band without fusion

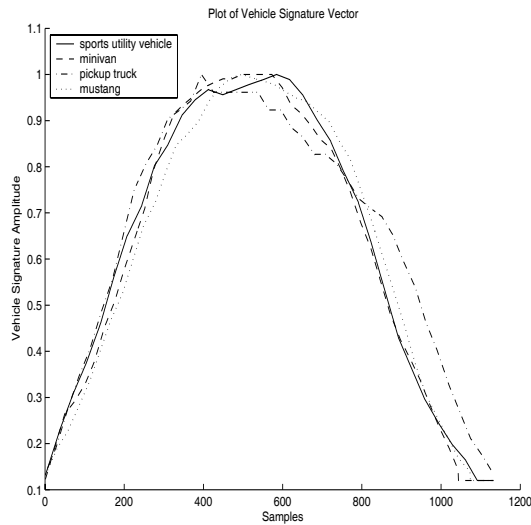


Fig. 1. Examples of vehicle signatures of a sports utility vehicle, a minivan, a pickup truck and a mustang (a car)

duced with a slight increase in performance.

Figure 1 shows examples of the vehicle signatures (*s*) of four different types of vehicles, namely, a sports utility vehicle, a minivan, a pickup truck and a mustang (a car). Figure 2 shows the Harr wavelet transforms (*t*) of these vehicle signatures for Band 7 only. The differences in the feature vectors allow for successful reidentification.

5. SUMMARY AND CONCLUSIONS

In this paper, it is shown that the use of the wavelet transform of the inductive signature improves vehicle reidentification accuracy. Also, using the transform coefficients of only one frequency band further improves accuracy and simultaneously lowers the feature vector dimension.

6. ACKNOWLEDGEMENT

The research reported in this paper was supported by the California Department of Transportation (Caltrans) and PATH (Partners

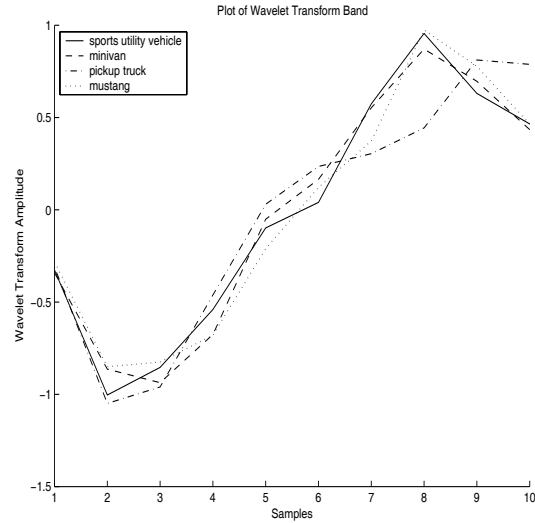


Fig. 2. Examples of the Harr wavelet transform of the vehicle signatures shown in Fig. 1 for Band 7

for the Advanced Transit and Highways).

7. REFERENCES

1. ITE, "Intelligent Transportation Primer", Institute of Transportation Engineers, Washington DC, 2000.
2. R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, John Wiley and Sons, 2000.
3. P. S. Addison, *The Illustrated Wavelet Transform Handbook*, Institute of Physics Publishing, 2002.
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. R. P. Ramachandran, G. Arr, C. Sun, and S. G. Ritchie, "A Pattern Recognition and Feature Fusion Formulation for Vehicle Reidentification in Intelligent Transportation Systems", *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Orlando, Florida, pp. IV-3840-IV-3843, May 13-17, 2002.