Single- and Multi-Sensor Techniques to Improve Accuracy of Urban Vehicle Classification

ISI-TR-2006-614

Unkyu Park[1,3] Fabio Silva[1] Mengzhao Hou[2] John Heidemann[1,3]
Genevive Guiliano[2] Xi Wang[3] Nikhil Prashar[2]
[1]Information Sciences Institute, [2] School for Policy Planning and Development, [3]Department of Computer Science
University of Southern California

Abstract

Vehicle traffic sensors are an essential part of urban traffic management systems. However, humans are still widely used when short-term traffic classification is required, because current traffic data collection systems do not meet requirements for high accuracy, rapid deployment, and low cost. This paper explores the use of a sensor network for vehicle traffic classification. We accomplish low cost and easy deployment by using simple embedded computers and inductive loop sensors that can be taped directly to the roadway. We explore several techniques to improve accuracy both at individual sensors and by combining readings from multiple sensors. Traditional approaches typically assume high-speed traffic; we instead focus on errors that are induced at low and varying speeds, since we expect sensors to often be deployed at locations where vehicles will slow or stop. The contributions of this paper are first, a comprehensive evaluation of approaches to reduce error: from feature definition, combining readings from a single sensor, and combining results from multiple independent sensors. Second, we evaluate these approaches against both on-line and off-line human observations, demonstrating sensor accuracy better than or slightly worse than on-line human classification depending on the similarity of the categories, and nearly optimal for length-based classification.

General Terms

Experimentation, Measurement

Keywords

Sensor Networks, Multi-sensor Fusion, Vehicle Classification

1 Introduction

Vehicle tracking has been a central application for sensor networks since 2000. Much of this work has focused on military surveillance applications, where individual vehicles move in unconstrained environments [2, 6, 9, 15]. However, there has been relatively little work exploring sensor networks applied to the much more common case of urban vehicle traffic, where vehicles are constrained to roadways, but vehicle density is much greater (the work of Cheung *et al.* is a notable exception [3, 4].). In this paper we explore vehicle monitoring and data collection for *transient, urban situations*. Specific users of this system include traffic management around construction zones or during emergencies, and for transportation planning and modeling.

Urban roadways carry thousands of vehicles each day, and elaborate vehicle traffic monitoring systems have been developed to manage traffic flows. Currently deployed vehicle traffic monitoring systems consist of either *emplaced* and relatively accurate sensors such as in-ground induction loops or elevated video cameras, or of *deployable* but less accurate sensors, such as pneumatic tubes. Both have strengths and limitations: sophisticated, emplaced traffic control systems today can be accurate and are essential to managing traffic flow, but such systems cover only major roadways and cannot be quickly deployed to new areas; they require substantial amounts of investment and planning to extend. Deployable systems, on the other hand, are more flexible. They can be used for short-term data collection, but current systems provide less accurate estimates of vehicle class and speed, particularly in dense or low-speed traffic.

Sensor networks provide a potential solution to this need for observing vehicles in urban environments. Ideally, small, battery-powered sensor nodes, attached to deployable sensors such as tape-down inductive loops, can detect and classify vehicles. More importantly, collections of individual sensor nodes can band together, both to improve overall classification accuracy, and eventually hopefully to do short-term tracking of vehicles in constrained areas, such as port facilities or distribution centers.

Although there has been a great deal of research in new sensor technologies to improve classification accuracy (see Section 3 for a review of this work), finding a good combination of accuracy, deployability, and cost has remained problematic. In fact, in spite of the research to date, manual counting ("people in lawn chairs") is still widely used when accurate vehicle classification is required. Video recording is also common; however line-of-sight and power requirements can limit deployability, and post-processing solutions to translate video into vehicle classes are still a subject of study.

This paper describes an easily deployable sensor network designed to classify vehicle traffic. We exploit three characteristics of sensor networks to improve classification accuracy: first, we *carefully select features* to minimize error caused by interpreting raw sensor data (Section 4.2). Second, we *exploit multiple estimates from a single sensor* to reduce the variance and improve estimation precision (described in Section 5.2, benefits quantified in Section 7.3). Finally, some classes of errors are uncorrelated across multiple sensors, so we *combine readings from different sensors* to improve classification accuracy (see Sections 5.3 and 7.6).

The main contribution of this paper is a comprehensive evaluation of approaches to reduce error: from feature definition, combining readings from a single sensor, and combining results from multiple independent sensors. While others have previously looked at the general areas of single-sensor signal analysis and multi-sensor fusion, and one group explored single-sensor approaches to vehicle classification [3, 4]), we are the first to consider this combination of techniques in the context of a sensornet for vehicle classification. In addition, we expect the lessons from our study of this specific domain to carry over to similar sensor networks with smart sensors.

The second contribution of this work is a comparison of three approaches to vehicle classification: on-line manual observation, off-line manual classification from video, and automatic, sensornet-based classification. This comparison is in Section 7.4. While all approaches strive for low error rates, none is perfect, each has different classes of errors. While some classification (car vs. semi-trailer) can be perfect, we show that distinguish between cars and trucks are inherently difficult because of ambiguity in the categories. Both humans and sensors face challenges consistently distinguishing these categories, thus it is appropriate to compare accuracy of automated system to human systems, in addition to striving to do better. With two categories (truck vs non-truck), we show that our system can exceed human accuracy with multi-sensor fusion (97% vs. 87%), while individual sensors roughly equal human accuracy (81–85% vs. 87%). In the more difficult case of distinguishing between three categories, multi-sensor techniques are needed to bring sensor-based classification near human levels (raising accuracy from single-sensor 57–70% to multi-sensor 74%, compared to 83% human). In addition, we demonstrate that inherent overlap of vehicle categories limits the accuracy of classification by length as we describe in Section 4.3; our dynamic observations have accuracy similar to what would be obtained by perfect-sensors.

We review related work in more detail in Section 3, considering work from both the sensor network and traffic management communities. Briefly, our work differs from most sensor networks in that we focus on classification of traffic in urban environments. In this case vehicles travel in predictable patterns (on roads), but traffic may be dense (sometimes bumper-to-bumper) over a wide range of speeds, making separations between vehicles difficult to identify. Unlike most other classification systems from traffic engineering, we focus on deployable systems that exploit multiple inexpensive sensors, rather than attempting to make a very accurate single sensor.

Figure 1. Deployment of a narrow inductive loop for a blade sensor.

To explore our approach we analyze data from a 1500-vehicle sample taken from a half-day experiment at USC, as described in Section 7.1

2 Goals of Vehicle Classification

The goal of vehicle classification is to detect vehicles on a roadway and identify their type (car, truck, semi-trailer, etc.). This information is needed by several groups, sometimes to use directly, and sometimes indirectly (for example, to predict vehicle length or weight). Transportation system managers are interested in the types of vehicles using the transportation network to help judge vehicle performance characteristics, impacts on road wear, and roadway safety. In addition, researchers are interested in understanding vehicle distributions to provide input to traffic models. Property developers often are required to perform traffic studies that evaluate vehicle flow rates and types.

The Federal Highway Administration (FHWA) has established a 13-category classification system to describe vehicles in terms of length, weight and number of axles [7]. The general scheme is passenger cars (FHWA class 2), light trucks (minivans, SUVs, pickup trucks, class 3), and various categories of larger trucks (class 4–13), identified by number of axles, length and average weight. It is straightforward to distinguish vehicles by number of axles, but more difficult to distinguish similar two-axle vehicles. The blending of the passenger car and SUV segments of the automobile market in recent years makes it particularly difficult to distinguish "passenger cars" from SUVs.

For our research we explore two different classification schemes: a *two-category* classification, where we distinguish between trucks and non-trucks (FHWA classes 2–3 and 4–13); and passenger cars, light trucks (SUV, pickups, etc.), and a *three-category classification* of passenger cars, light trucks, large trucks (FHWA classes 2, 3, and 4–13, respectively). As we describe in Section 4.2, we map vehicle length to category. We measure *wheelbase*, the distance from front-to back-wheels as the vehicle length. We describe this classification scheme in Section 4.3.

3 Related Work

There is a range of prior work in vehicle tracking, classification, and sensor fusion. We briefly review and compare this work to ours below.

Sensing for Vehicle Tracking in Constrained Environments: In-road traffic sensors are ubiquitous in most urban environments. The need for efficient traffic flow has sparked significant investment in novel uses of both existing and new sensors. A number of sensor technologies have been considered. Pneumatic tubes and piezoelectric sensors detect wheel crossings; inductive loops and magnetic sensors detect vehicle mass; and infrared, ultrasound, radar or laser ranging, and video, employ different levels of imaging. We survey these elsewhere [10]. Important differentiators here are ease of deployment, robustness, and cost. In-roadway inductive loops are widely used and quite robust, but require construction to deploy. Video approaches remain relatively expensive, both in use and deployment (which may require elevation).

In spite of the large amount of research done on traffic sensors, systems for temporary deployment often fall back on simple pneumatic tubes, or manual, human observers, either physically present or interpreting videotape. For our work we use portable inductive loops because they retain the deployability of tubes but can provide much greater information.

Closest to our work is that of Oh *et al.* [12], Sun *et al.* [13], and Cheung *et al.* [3, 4]. Oh *et al.* use the same IST Blade sensor we do, but targeted at arterial speeds (20–45mph) rather than slow speed. In addition, they select a very different set of features, use a neural network for vehicle reidentification rather than classification, and so do not explicitly estimate vehicle speed or length. Sun *et al.* add a neural network to loop sensor output, explicitly trying to differentiate a custom set of categories including cars separate from SUVs, as well as larger vehicles. They report good accuracy, 82–87%, although they do not indicate if their errors are in the difficult-to-distinguish categories or not. Cheung *et al.* instead use custom sensor nodes with magnetometers, measuring the change in the Earth's magnetic field. They use both length-based classification and a novel "Hill Pattern Classification". While they obtain highly accurate vehicle counts (98%), their classification accuracy (82% into 5 FHWA classes corresponding to different axle-count large trucks) [4] is slightly better than ours (74% into 3 FHWA classes corresponding to passenger cards, small trucks, and large trucks). However, it is important to note that their classes are quite distinct and can be distinguished by axles counts and large differences in the length. We instead focus on FHWA classes 2 and 3, difficult to distinguish cars vs. small trucks We also explore the use of multiple features in a single signal, as well as multiple independent sensors to improve classification accuracy.

Sensor Networks for Vehicle Tracking in Unconstrained Environments: Vehicle tracking was one of the first problems for distributed sensor networks [2, 9, 15]. In general, these approaches focus on tracking relatively sparse (clearly distinct) targets without assumptions about target motion. To cope with these challenges they exploit multiple sensors for data from different viewpoints and use information theoretic techniques to estimate the vehicle path [2, 15]. More recent work has focused on less powerful nodes [6] and approaches to accommodate individual sensor noise [9], but still addresses relatively sparse targets. By contrast, our work focuses on densely packed vehicles on a busy roadway, and we exploit the capabilities of powerful cross-road sensors to make the problem tractable.

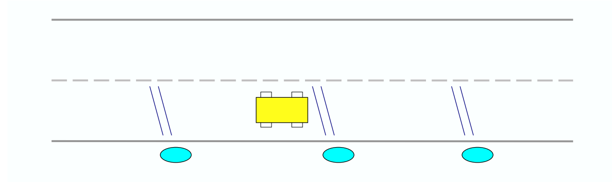

Figure 2. Deployment of three sensor nodes and six sensors along a roadway.

Sensor Fusion for Improved Accuracy: Sensor fusion is an important approach to exploit multiple sensors. Zhao *et al.* use information theoretic techniques to coordinate cooperation in a multi-sensor environment, selecting the sensor based on maximum information gain [15]. Brooks *et al.* explore collaborative signal processing and identify the level of sensor independence (how correlated or uncorrelated each is with others) as an important issue [2]. Gu *et al.* exploit cluster-based processing to correlate readings from multiple sensors [9].

4 A Sensor Network For Traffic Classification

Our long-term goal is to develop sensor networks that allow automatic accurate vehicle classification and reidentification. In this paper we explore the first steps: identifying a state-of-the-art stand-alone sensor that can be integrated into a deployable sensor network; investigating its accuracy, both stand-alone and in collaborative sensor network; and verifying this accuracy through a field experiment.

4.1 Sensor Hardware and Deployment

Our focus is a sensor network that is rapidly deployable, low cost, and sufficiently accurate. Our basic sensor network consists of several individual sensor nodes, each connected to a IST-222 high-speed detector card and a pair of Blade inductive loop sensors [11]. We selected the IST detector card because of its potential for sampling at high resolutions (up to 1.2 kHz), and the Blade inductive loop because of its sensitivity to vehicle features. The IST detector can be integrated with sensor network platforms such as the Intel Stargate [5] via USB, and we expect to use networks such as 802.15.4 for low-power, short-range communication. The systems as a whole should be relatively inexpensive (currently less than US$4000). The complete system is easily deployable since both the sensor nodes and the Blade sensor can operate on battery power. Pairs of inductive loops must be taped down across traffic lanes using asphalt tape, thus requiring a brief interruption of traffic.

Figure 1 shows deployment of a pair of loops as part of our trial. Figure 2 shows the logical configuration of our array of traffic sensors: each individual sensor node connects two closely separated blade sensors (a car is shown approaching the middle pair of sensors).

The use and configuration of two inductive loops is important to our design. We place the loops of the IST blade sensor about 4 inches wide at approximately a 20° angle to traffic. Both of these details are required to generating a signature that captures characteristics of the vehicle: because the inductive loop is narrow, it generates a signature sensitive to specific car features (axles and engine), rather than a simple presence or absence. Because it is angled, each wheel crosses the sensor separately and usually shows up distinctly. We place an array of sensor pairs along the roadway to facilitate sensor fusion; by separating the sensors we improve resilience to independent errors as discussed later in Section 5.3.

We report here on analysis conducted with data from two locations. First, we discuss our single sensor algorithms using data from one location. We then address multi-sensor fusion using data from two locations. We have not yet integrated our multi-sensor algorithms with communications protocols; currently we manually process multi-sensor data at a central site. We believe a distributed implementation is not difficult, and plan to exploit sensor deployment information to make the system largely self-configuring, as in prior work [14].

4.2 Single-Sensor Classification

Signal processing at an individual sensor is fairly traditional: we begin with noise removal, segment the data into individual vehicles, extract features pertinent to our analysis, then classify vehicles; we examine each of these steps next.

Noise elimination: In some cases we have observed significant crosstalk, environmental noise, and signal drift in our measurements. In addition, communications from the IST card to the host computer is not perfect. Crosstalk arises when inductive loops driven by different sensor cards are close to each other, and also due to slightly different sensor clock rates in the case that multiple sensor cards are attached to the same sensor node.

Drift and noise occur due to temperature change and other environmental effects. We filter each of these out using standard techniques. We observe around a 14% data loss on the USB bus between the sensor and the host (this behavior is a known problem of the specific model of detector cards we were using); we correct for this by interpolating the missing data.

Segmentation: When noise is eliminated, we are left with a continuous signal. We next isolate individual vehicles with a three-step process. We first detect active segments by observing large signal deviations from a running mean. We then merge temporarily close active segments to allow for vehicles that have "flat" areas between wheel wells. Finally, we grow segments by half their length at front and back to ensure we capture a complete segment, including leading and trailing features. As a special case, when growing a segment would cause overlap with a neighboring segment, we grow to the midpoint between segments. Ideally, after segmentation, each segment corresponds to exactly one vehicle. In practice we find that occasionally (about 5% of the time, see

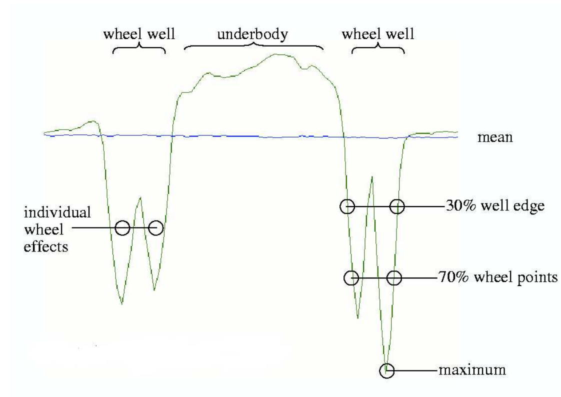

Figure 3. Sample signature indicating vehicle parts (signature #280, site BN)

segmentation errors in Table 10) vehicles that are very near to each other appear in a single segment.

Feature extraction: We experimented with several possible features for vehicle classification, including axle count, body width, and wheelbase (axle-to-axle distance). We converged on a two-level set of features. We directly extract the edges of wheels (70% wheels points, described below), then figure these and estimate speed and wheelbase. Our first goal is to determine wheel edges. Figure 3 shows a sample signature of a two-axle car crossing a sensor. Wheels show up as large dips in the signature, the underbody as bumps between the wheels. Because the vehicle crosses the inductive loop at an angle, each wheel of the same axle produces a distinct dip located near the other. We experimented with different algorithms to reliably extract each wheel. Our two main approaches were to identify peaks and to identify large changes in direction. Although peak identification is attractive, consistent results are difficult because peaks tend to be rounded, particularly at higher sampling frequencies (because wheels are round). In addition, depending on the angle the car crosses the sensor, we may get two clearly distinct peaks or a single merged peak. Because of these difficulties we adopted the "steep slope" algorithm shown in Algorithm 1.

It is important that our algorithm adapt to a wider range of vehicle speeds. To do so, we adjust the parameter v based on vehicle speeds. First, we start with a default v value 240 ($0.2R$, where R is the sampling rate) to estimate the vehicle speed, since it is not necessary to separate each wheel well to get speed estimates. Then, we adjust v according to the estimated speed as shown Algorithm 1. If v is too large, we cannot separate the wheel wells; on the other hand, if it is too small, we cannot capture the all wheels in the wheel well. In case of extremely low speed (less than 5mph), we use 680 ($0.4R$) for v to prevent from collapsing every wheel well into one. The other parameters are somewhat arbitrary. We initially chose the S value as 70%, with the goal of identifying a point on the steep part of the signature. We studied values from 30–80% and found they did not affect classification accuracy (a graph of these results is omitted due to space constraints).

Algorithm 1 The steep-slope algorithm for wheel edge extraction.

Normalize the signature values from 0.0 (lowest underbody) to 1.0 (highest wheel peak)
Compute m, the mean of all sensor readings over entire signature
Identify wheel wells by finding the first value greater than $m+(L*m)$ through the last value greater than $m+(L*m)$, allowing up to v consecutive values below $m+(L*m)$
For the first and last wheel wells:
    Find the maximum value M in the well
    Define the start-wheel-point as the first value in the well greater than $S*(M-m)$
    Define the end-wheel-point as the last value in the well greater than $S*(M-m)$
Parameters:
$v : 2/(VehicleSpeed)*SamplingRate$, number of samples allowed below $m+(L*m)$ in a wheel well (between wheel peaks)
L : 0.3, wheel well start threshold (fraction of mean)
S : 0.7, target for steep slope (fraction between mean and peak)

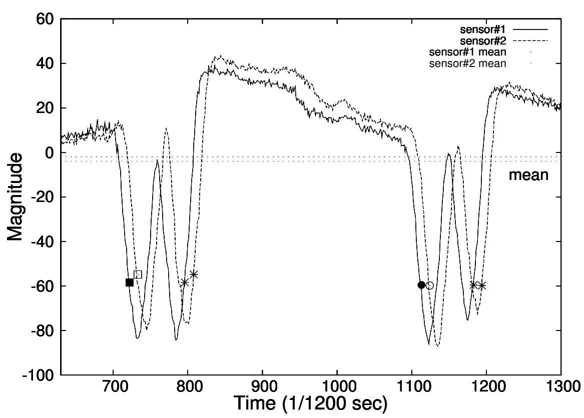

Figure 4. A signature with two-paired sensors (signature #130, site BN)

Speed Estimation: To estimate vehicle speed we compare the time difference when a wheel-point crosses each of the two closely placed loops. Figure 4 shows a time-correlated sample from both of the paired loops for one signature. We compare pairs of shapes (squares or circles), and similar observation at the front and back of the first and last wheel wells. Since the loops are a known distance apart, a speed estimate is simply the distances divided by the time between the same feature at each adjacent loop. To reduce error, we match the wheel-points for the start and end of the first and last wheel wells, giving us four estimates of speed. We then average these values. We discuss how this approach addresses different errors in Section 5.2, and quantify these benefits in Section 7.3.

Wheelbase: To estimate wheelbase, we observe the front of the first and last wheel wells: solid shapes (squares or circles) in Figure 4. With two paired sensors, we can get four estimates front and back of sensors 1 and 2.

Classification: Given a speed estimate, we classify vehicles by wheelbase length, the distance from the first to the last wheel axle. Given our wheel-points on the first and last wheel wells, we have four wheelbase estimates on each of our two paired sensors. As with speed estimates, we average these four readings.

Finally, we map length to vehicle classification as described in the next section.

Table 1. Length-based vehicle classification

FHWA classes	meaning	symbol	wheelbase length
Two-category system			
2 and 3	non-trucks	non-T	< 170”
4 to 13	commercial trucks	T	≥ 170”
Three-category system			
2	passenger cars	P	< 118”
3	small trucks/SUVs	S*	118 to 170”
4 to 13	commercial trucks	T	≥ 170”

Clearly our classification algorithm is quite simple. A much more sophisticated system would, for example, match the entire signature against a database of known vehicle types. However, even this simple classification system is sufficient to explore the use of multiple sensors to reduce error. In Section 8 we describe future directions defining more appropriate classification schemes.

4.3 Classification by Wheelbase

Table 1 shows our mapping from length to our two- and three-category classification systems and the relationship to the FHWA classes. We selected the two-category system because automatic classification of truck vs. non-trucks is relatively easy, while we will show that distinguishing cars from SUVs (P from S*) is much more difficult and thus represents a “worst case” for automated classification systems.

A critical problem with the FHWA system is that the boundaries between classes 2 and 3 are indistinct. In fact, the FHWA website says “because automatic vehicle classifiers have difficulty distinguishing class 3 from class 2, these two classes may be combined into class 2” [7]. This problem applies also to human observations (we quantify human accuracy in Section 7.2); we next consider how length relates to classes.

To evaluate the wheelbase-based classification with perfect sensors we surveyed wheelbase lengths of 47 vehicles from Ford [8] and government sources [1]. Tables 2 and 3 show classification accuracies based on wheelbase assuming perfect length determination. Even though we classify vehicles with their exact lengths, not all surveyed vehicles are correctly classified: 96% of vehicle models are correctly

Table 2. Two-category classification with perfect sensors

Class	Total	Correctly classified	Incorrectly classified
non-T	42	41	1
T	5	4	1
Total	47 (100%)	45 (96%)	2 (4%)

Table 3. Three-category classification with perfect sensors

Class	Total	Correctly classified	Incorrectly classified
P	24	24	0
S*	18	12	6
T	5	4	1
Total	47 (100%)	40 (85%)	7 (15%)

classified in two-category classification. In three-category classification only 85% of models are correctly classified, since there are many SUVs on either side of our threshold, regardless of where it is placed.

These results assume static analysis, in that the percentages are based on numbers of vehicle types. In practice, the population of vehicles of each type varies, as does the particular set of vehicles observed at any site. Therefore, dynamic measurements may differ depending on the mix of observations.

4.4 Using Multiple Sensors

A defining characteristic of sensor networks is the use of many relatively inexpensive sensors. We therefore wish to explore if multiple sensors can improve the best-possible classification results of a single sensor. Our hypothesis is that classes of errors are independent, so combining values from moderately separated sensors can eliminate these errors.

As shown in Figure 2, we place several pairs of sensors at several places on a roadway. We expect sensors to communicate through a local, low-power, wireless network such as provided by 802.15.4 or similar networks. Sensor nodes will not share raw sensor readings, but instead individual evaluations of vehicle type, coupled with data about their confidence in the classification. Such a system must have several components: a configuration system to automate initial deployment, communications protocols to share information between sensors, a signature matching algorithm to identify which signatures at one sensor correspond to signatures at another sensor, and a classification preference algorithm to select which classification is best. We plan to exploit the simple, constrained topology of the road, so configuration and communications are straightforward as each sensor interacts with its immediate neighbors. Signature matching and classification preference are the keys to improving accuracy with multiple sensors. We discuss these algorithms in Section 5.3 after reviewing potential types of error.

5 Types of Errors and Error Recovery

To consider how multiple sensors might improve accuracy, we first evaluate the types of error that arise in this application, then consider how to make a single sensor as effective as possible, and finally how multiple sensors can further improve accuracy.

5.1 Types of Error

We review the types of errors we expect in Table 4. There we evaluate each error for its generality, if it is specific to this application or applies to all sensors; dependence, if we expect multiple sensors to exhibit this error consistently or independently; how we address it, in-sensor with multiple estimates or multiple sensors; and if we observed it in our examples.

We observed significant amounts of *environmental noise* in our data, both due to temperature drift and sensor cross talk. A later revision of the IST Blade sensor handles noise elimination in the sensor itself, but for our data collection experiment we filtered noise manually post-facto. Inductive loops respond to vehicle mass relative to its distance from the sensor, thus they are less sensitive to vehicles that are higher off the ground. Loop sensitivity can be controlled by adjusting width, so potentially multiple loops of different widths could detect a wide range of vehicles. For our main experiments we used a loop width of about 4 inches.

We did not observe any *sensor failure* in our system, but it would be an issue for larger deployments.

An *insufficient sampling rate* or too close placement of sensors can result in imprecise speed and length estimates, since a change of a single sample interval corresponds to a noticeable change in estimate. Sampling rate or sensor distance must be adjusted to expected speeds, as we explore in Section 6.

Vehicle type errors, refer to different distances of vehicles from the ground. We did observe distance affecting the quality of sensor signatures, however it was not a major cause of misclassification.

Mis-channelization is when only part of the vehicle crosses the loop because it is changing lanes. *Changing speeds* occur when a vehicle alters its speed over the sensor, making estimation difficult. *Mis-segmentation* occurs when two cars travel so close that they appear to the sensor to be a single vehicle. All of these errors are specific to vehicle classification, but each occurs independently at different sensors and so should be correctable in the sensor network. We observe and correct several mis-segmentation errors as described in Section 7.3.

5.2 Single-Sensor Error Recovery

We used three general techniques to improve individual sensor readings: sharp feature detection, internal consistency checking, and cross-checking with multiple features .

Our early approach identified speed and wheelbase by using peaks of the signature to estimate the exact wheel locations. This approach proved inaccurate at high sampling rates because wheels provide rounded features to the inductive loop, since the loop has an effective low-pass filter, and wheels are round. To address these problems, we shifted to identifying the sharp slopes that correspond to the edges of the wheel peaks, as determined by the S threshold, a fraction of the distance from mean to the peak value (currently 70%). Since the slopes on the edges change rapidly, slopes tolerate much more error than peaks. In fact, we experimented with thresholds corresponding to 30–80% of the distance between the mean and maximum sensor values and found classification accuracy unchanged.

Table 4. Types of errors in vehicle classification

Type of error	Generality	Dependence	Addressed	Observed
Environmental noise	General	Either	(In-sensor) or post-facto	Yes
Sensor failure	General	Independent	(Multi-sensor)	No
Insufficient sampling	General	Independent	(Design)	Yes
Vehicle type	Application	Dependent	(Multi-sensor)	No
Mis-channelization	Application	Independent	Multi-sensor	Yes
Imprecise speed	Application	Dependent	Single sensor	Yes
Changing speeds	Application	Independent	Single sensor	Yes
Mis-segmentation	Application	Independent	Multi-sensor	Yes

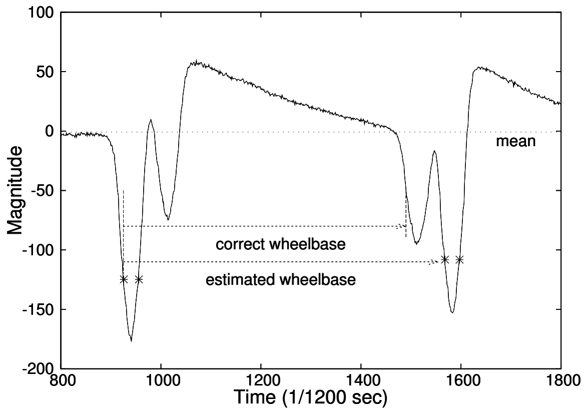

Figure 5. A signature with missing wheels (signature #96, site BN)

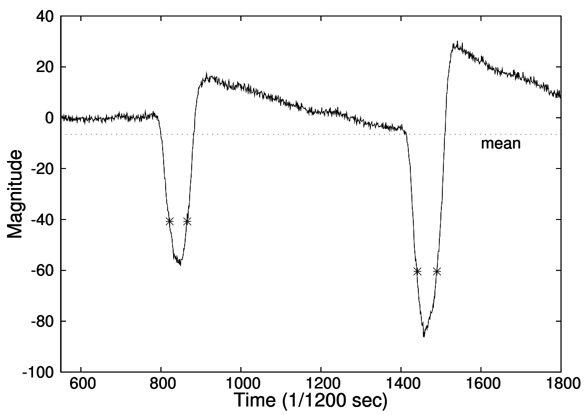

Figure 6. A signature with channelization error (signature #251, site BN)

Second, we check the features for internal consistency. Our primary approach is to evaluate where detected wheels are placed inside the wheel well. We expect to get two peaks in each wheel well, corresponding to the left and right wheels. However, if one of the peaks is much smaller than the other, sometimes it is missed by the steep-slope algorithm. One wheel may be omitted from a wheel well when channelization errors occur (Section 5.1). Ignoring these errors can result in significant inaccuracy in wheelbase estimation. Figure 6 shows a signature in channelization error. For example, in Figure 5, if we miss a right wheel in the first and the left wheel in the last, the estimate will be longer than the proper estimate.

To solve this problem, we take several steps to examine wheel placement for internal consistency. (We explored adjusting the steep slope target *S* to catch low peaks, but in general one cannot catch all low peaks and be robust to noise.) We count the number of wheels in each wheel well. If believe a wheel is missing, we identify which wheel is present, or perhaps determine that we cannot tell which wheel we observe. There are four possible wheel placement states: both wheels present, left wheel or right wheel missing, and unable to determine which wheel is missing. There are 16 possible combinations of these cases when we consider detection at the first and last wheel wells.

When both wheels are present we get two wheelbase estimates per loop. If a wheel is missing but we can identify which wheel is present we calculate a single wheelbase estimate with the present wheel. However, when we cannot determine which wheel is present, or if different wheels are present in the front and back wells, we know that our wheelbase estimate will be incorrect. In these cases we report our best estimate along with a lower confidence value in this estimate, potentially allowing multi-sensor error recovery to select a better estimate at another sensor as explained in the next Section 5.3.

Finally, we obtain multiple estimates of each feature in a single signature. We can find four wheel-points for each signature in the beginning and ending of the front and back wheel wells. (These are indicated by stars in Figure 4, in addition to the squares and circles.) This gives us four estimates of speed and length. Although these estimates are not completely independent, averaging them provides much less variance than any individual reading. This approach also partially corrects for vehicles that change speed over the sensor. We quantify the benefit of this in Section 7.3. Although reduction in variance does not necessarily translate into improved accuracy, it does imply less susceptibility to noise. We quantify the reduction in variance in Section 7.4.

5.3 Multi-Sensor Error Recovery

To use multiple sensors to reduce errors we must identify when multiple signatures correspond to the same vehicle and then which classification is most accurate: signature matching and classification preference.

Signature matching: We plan to exploit the constrained topology of a roadway to simplify signature matching. If sensors are placed on a roadway without intersections or exits, then the order of vehicles and signatures is fixed and so signatures can be matched based on timing and ordering. Missing signatures can be inferred by gaps, and mis-segmentations by a very long signature at one sensor followed by a short signature at the next.

Classification preference: Given two matched signatures with different classifications one must choose which classification is more likely to be correct. We are experimenting with two algorithms: quality-best and shortest-best.

The *quality-best* algorithm favors sensors that are able to consider multiple estimates that report consistent values. This approach addresses speed variability, since a larger number of speed and length estimates reduce the impact of variability and allow the sensor to estimate its confidence in the estimated speed and length (more variance indicates less confidence in the estimate). We also adjust the confidence according to the internal consistency of the signature by considering where we believe wheels are placed in a signature. If there are missing wheels and their placements are inconsistent, we reduce the confidence value to half of what we get from the variance.

The *shortest-best* algorithm is much simpler. In our experiment we found some errors were due to mis-segmentation. We can detect mis-segmentation as a long signature at one sensor with two short signatures at the other. Our system is more likely to merge two adjacent signatures than split a long signature, therefore in these cases the shortest-best algorithm selects the two-signature interpretation over the single long signature.

Finally, for analytic purposes, we consider an *oracle* algorithm. The oracle algorithm assumes a perfect classification preference algorithm that always chooses the correct classification if it is present at either sensor. Such an algorithm is impossible to realize in a real system—we can implement it only because we already have ground truth. We present it to provide an upper bound on how well sensor fusion can do.

Current Status: We have implemented classification preference based on sensor quality, and multi-sensor comparison. We have not implemented a complete matching algorithm, since in our data collection experiment sensors were separated by a large distance and an intersection, so for our preliminary analysis we manually associated signatures at two sensors.

6 Single-Sensor Calibration

Before evaluating our approaches with the field experiment, we examine the estimation accuracy of using a single sensor. The goal in this section is to determine the accuracy of a single sensor to estimate speed and wheelbase.

To evaluate the accuracy of a single-sensor, we conducted a short experiment on the roof of ISI parking structure. We tested two sensor separation distances: 18" and 36" at 300Hz sampling rate. We also experimented with two target speeds: 10 mph and 20 mph to demonstrate that different speeds affect estimation accuracy. Ground truth vehicle speeds are established by using a radar gun (Model: Bushnell 10-1911, accuracy: $\pm$1mph, precision: 1mph).

Table 5 and 6 shows detailed results. Speed estimation (Table 5) indicates that sensor readings are within 5–17% error. Linear regression of the correlation of estimated speed and radar gun speed shows an r-square of 0.85–0.90, indicating a very strong, but not perfect correlation between speed and ground truth speed. Although this accuracy is good, for speed we have a very poor reference "ground truth": the precision of the radar gun is only 1mph, so we believe some of this error corresponds to the poor precision of those measurements. In fact, with the wide-spaced sensors (36"), error is within this precision. With narrow-spaced sensors, errors are around 2.3mph, confirming that accuracy falls off with narrow-spaced sensors at these sampling rates. Finally, we observe that absolute error in speed is roughly constant for these speeds, and the relative error actually decreases.

When we turn to length (Table 6), we see much higher accuracies. The length estimation error is less than 6% which indicates that we can accurately estimate length using a single sensor. We also observe that the wide-spaced sensor is again more accurate. Note that length estimates are based on speed estimates, so we expect these results to be dependent. The main difference is that we have accurate ground truth of length.

We didn't investigate exhaustively the effect of distance and sampling rate on the accuracy, but we at least confirmed the configuration we used is appropriate for speed and length estimation. In Section 7, we experimented with 4.5" separation distance and 1200Hz sampling rate. We didn't explicitly examine this configuration, but the accuracy would be roughly equal to the configuration of 18" and 300Hz, because accuracy is proportional to the product of distance and sample rate.

7 Evaluation

To evaluate and develop our algorithms we collected a 3-hour traffic dataset with three approaches to sensing: Blade sensors using a prototype of our system, human observers, and videotape. Originally we expected to use human observations as ground truth, but, as we describe below, we found significant errors when we compared their observations against video data. This section describes the details of that experiment, compares accuracy of off-line video analysis to on-line human observation, considers two- and three-category automated classification, and finally the ability of multiple sensors to correct errors.

7.1 Data Collection Experiment

From 7 am to noon, August 6, 2004, we collected traffic data at the USC campus. Working with Steven Hilliard of IST, we collected 1500 detections of vehicles at three locations on the USC campus. Sensor data was supplemented with human observers and videotape to provide ground truth

Table 5. Single sensor accuracies for speed from the 18" and 36" sensor separation distance with 300Hz sampling. Standard deviations are given in brackets.

Sensor separation	Target speed (mph)	# samples	Achieved speed (mph)	Estimated speed (mph)	Absolute error (mph)	Relative error
	10	5	13.5 [1.97]	15.6 [1.36]	2.1	17.1%
18"	20	6	19.0 [1.87]	21.6 [2.94]	2.6	13.3%
	Overall	11	16.0 [3.40]	18.3 [3.73]	2.3	15.4%
	10	10	12.6 [0.84]	13.2 [1.54]	1.3	10.4%
36"	20	10	19.3 [0.94]	20.2 [0.69]	0.9	5.0%
	Overall	20	16.0 [3.55]	16.7 [3.77]	1.1	7.7%

Table 6. Single sensor accuracies for wheelbase from the 18" and 36" sensor separation distance with 300Hz sampling. Standard deviations are given in brackets.

Sensor separation	Actual length (inch)	Estimated length (inch)	Absolute error (inch)	Relative error
18"	107.5	113.4 [2.80]	6.0	5.6%
36"	107.5	107.7 [1.98]	2.6	1.6%

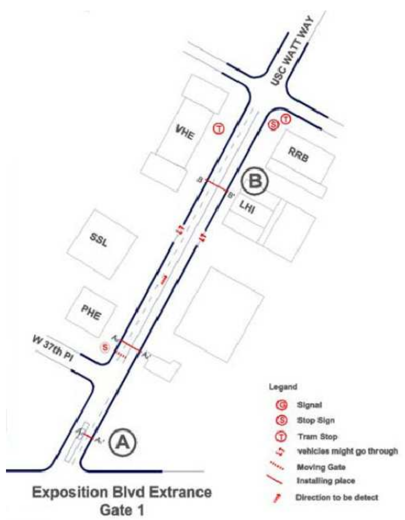

Figure 7. Placement of sensors for data collection.

data. We selected three locations on internal campus streets to get a mix of low- and moderate-speed traffic. We selected a data collection day when construction was underway on campus, allowing us to capture a mix of periodic traffic, including the USC shuttle bus, construction traffic, including cement mixers and 18-wheel trucks, and general automobile traffic. In addition to general traffic, we selected two passenger cars and ran them over each sensor 10 or more times to provide a baseline known vehicle to evaluate re-identification and sensor consistency.

Details about exact deployment locations are available in a technical report [10]; here we report on two locations: site A, near the campus entrance, and site B northbound (or BN), about 100m north of site A, past an intersection. Figure 7 shows this layout. Site A is next to a parking kiosk and has two lanes. Typical vehicle speeds are quite slow (mean speed 8mph); stops in lane #1 (immediately next to the kiosk) were frequent. Here we report only data from lane #2, further from the kiosk. Site B consisted of north (BN) and southbound (BS) lanes, mid-block. Mean speeds were 16mph. For most of our analysis we consider site BN. We use site A to confirm our BN results and see how our approach works on slower traffic, and to examine multi-sensor fusion for vehicles that pass through both sites A and BN. We use the data from 9:15 am to 10:50 am, discarding earlier data because of incomplete video ground truth, and later data because of numerous records from our two test cars.

7.2 Manual Classification and Video to Establish Ground Truth

To validate our sensor-based results we collected both manual counts by graduate students and video records of traffic. Although intended to provide "ground truth", both manual and video counts had errors that we describe below. Thus, in addition to providing an alternate count to compare sensor-based counts against, it also unexpectedly provided an alternate estimate of error rates to help put sensor-based accuracy into context.

First we consider on-line, manual observation. By "on-line" we mean classification done by human observes while vehicles arrive. This kind of manual observation is widely used today.

We did not expect any errors in our on-line manual counts. However, in practice we found that errors crept in for several reasons. At times, vehicles came too quickly to write down type, plate, and other information before the next vehicle arrived. Manual logs therefore had missing or incomplete entries. Interruptions in counting (when counters or videotapes were changed) also caused missing records. Finally, we recorded both vehicle license plates and types. Sometimes plates were not visible, and different observers sometimes used different indication for the same vehicle types, particularly with unusual types, or with trucks, SUVs, station wagons.

Next we consider off-line, manual analysis of videotape.

Table 7. Accuracy of on-line manual two-category classification (Site BN).

	Total considered	Correct	Total incorrect	Incorrectly classified: Undercounts	Overcounts	Wrong as non-T	Wrong as T
Total	248 (100%)	216 (87%)	32 (13%)	28 (11%)	0 (0%)	......4 (2%)......	
non-T	211	183	28	26	0	–	24
T	37	33	4	2	0	2	–

Table 8. Accuracy of on-line manual three-category classification (Site BN).

	Total considered	Correct	Total incorrect	Incorrectly classified: Undercounts	Overcounts	Wrong as P	Wrong as S*	Wrong as T
Total	248 (100%)	207 (83%)	41 (17%)	28 (11%)	0 (0%)	.........13 (5%).........		
P	127	109	18	14	0	–	4	0
S*	84	65	19	12	0	5	–	2
T	37	33	4	2	0	0	2	–

While still human driven, the fact that data was on video rather than live allows pausing and review of busy periods or inconclusive vehicle types. We used both these techniques to get what we believe as the most accurate possible classification.

Video collection losses were more predictable. We were unable to deploy a camera at one site, and there were interruptions when videotapes where changed. Finally, video resolution was insufficient to read license plates. However, with careful post-facto analysis (and re-analysis) of the video we produced what we believe is complete ground truth. While essential for careful evaluation of our results, the cost of repeated manual reviews of video is probably not warranted for typical counting experiments.

To determine accuracy of manual counts we analyzed data from one site (site BN, from 9:15 am to 10:50 am). We discarded 35 records of motorcycles, electric carts, and bicycles, since our sensor analysis focuses only on cars and larger vehicles. These changes left us with 248 manual observations.

Tables 7 and 8 summarize our comparison of on-line manual classification to multi-pass, off-line analysis of the video. First, these tables show that for the vehicles that are counted, accuracy is quite good: only 5% of vehicles that have a recorded classification type are mis-classified. Second, the data shows the main problem with manual counting: inability to keep up when many vehicles are present. This shows up as a large number of undercounts (28 cases, 11% of the records). This was particularly a problem in the period around 10:40 am. In addition, in one case a human observer misclassified a passenger car as a cart, excluding it from this survey. Finally, we didn't observe any overcount, vehicles recorded as present that were not there. We discuss how these results compare to computer sensors in a later section.

In comparing two- and three-category classifications (Tables 7 and 8), the results are quite similar. This similarity is because humans are generally good at identifying vehicle type, while sensors must do so indirectly, perhaps by wheelbase.

7.3 Error Recovery in a Single Sensor

Before quantifying classification accuracy and multi-sensors techniques to reduce error, we first evaluate single-sensor techniques to improve estimation consistency.

To evaluate the benefits of multiple features we compare the amount of variance we see across multiple readings. In Section 4.2, we show that we can determine up to four estimates of speed and length from each sensor pair. We expect that two classes of error, imprecise speeds (due to sensor inaccuracy), and changing speed (if cars alter their speed over the sensor), can be addressed by exploiting multiple estimates at a single sensor.

Figure 8 shows how variance changes when we consider 1, 2, 3, or all 4 estimates of each speed and length for data from both sites A and BN. Unfortunately we do not have ground truth about speed and length for our main experiment, so we computed our best-estimate of speed and length by taking the mean of all four measurements. For each signature, we compute the difference of an n-sensor estimate against this best estimate. We consider all possible combinations of n sensors for each signature, so the 1-estimate values consider 4 measurements per signature, the 2-estimate considers 6 combinations per signature (combinations of estimate 1-2, 1-3, 1-4, 2-3, 2-4, and 3-4), and the 3-estimate considers 4 per signature (estimates 1-2-3, 1-2-4, 1-3-4, 2-3-4).

For the left and center graph in this figure, we first estimated speed and length for each signature using the mean of all four estimates. As can be seen, use of multiple estimates greatly reduces measurement variance. Variance is larger when speeds are faster at the BN site. Finally, variation in length estimates is considerably less than variation in speed. Three-category classifications are affected by even small variations, however, since there are many vehicles near the dividing line between P and S*.

To test against ground truth, we took an additional, custom experiment when we ran a vehicle with known length across sensors spaced at 18 and 36 inches (details of the experiment are in Section 6). The right side of Figure 8 shows these results. The trend in the right graph against known ground truth matches the trend in the middle graph: more estimates reduce variability. However, this experiment also confirms that less variability corresponds with more accuracy.

7.4 Sensor-Based Classification

We next consider sensor-based classification using data collected from both A and BN sites as described in Sec-

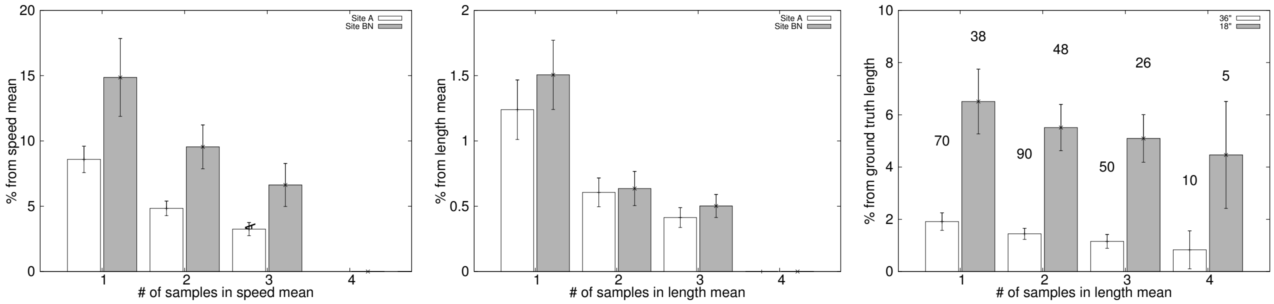

Figure 8. Comparison of the use of multiple estimates from a single signature. Speed (left) and length (middle) compared to four estimates, and length compared to to ground truth (right). Left and middle are sites A (white bars) and BN (gray bars), right from custom experiment.

tion 7.1.

As discussed in Section 7.2, we manually deleted records of motorcycles, carts, and bicycles from the dataset. It should not be hard to eliminate these automatically, but our current focus is on the harder problem of distinguishing cars from trucks, not the easier problem of cars from bicycles. As future work we plan to automate this filtering. After filtering we were left with 164 records on site A and 248 records on site BN. In addition, we found 22 records (9 on site A and 13 on site BN) were missing from sensor data set, due to segmentation errors. Although we do not yet automatically detect these, we believe we can do so relatively easily. We therefore report these values as "segmentation error (SE)" on Table 9 and 10.

We started by classifying vehicles in two groups: trucks (including 18-wheelers, cement mixers, and panel-trucks) and non-trucks (includes everything else other than trucks). Results are shown in Table 9. With just two categories (trucks, type T, vs. non-trucks), our classification rates are comparable to current, state-of-the-art published results from Sun *et al.* [13]. These results were particularly encouraging given that our system was not tuned to specifically deal with the high variability in vehicle speed that we found.

In addition to the two categories of vehicles mentioned above, we expanded our initial classification algorithm to include a third category, S*, comprised of SUVs, pick-up trucks, vans and minivans. Results are shown in Table 10. Our accuracy with three categories is not as high as with only two categories. Our algorithms depend primarily on vehicle length, but the vehicles in class S* and P overlap in length. No classification algorithm based only on length will be able to separate these categories accurately. This can be seen in the data in Table 10 where many "medium-size" vehicles are incorrectly classified as passenger cars (51 out of 139). By comparison, classification of trucks is quite good, with very few trucks being misclassified as S*, and very few S* or P types being classified as T.

This experiment suggests, first, that our approach is very appropriate if the goal is to classify trucks from non-trucks. For studies about road damage, this level of classification may be sufficient. It also suggests that very fine-grain classification of passenger cars, vans, pickups, SUVs will be quite difficult, given the blending of these vehicle types.

It is useful to compare the accuracy of our sensor network-based classification system to manual (human) classification described previously. First, considering just counting, manual classification was 83–87% accurate. By comparison, our classification system had accuracy rates of 81–85% for two-category classification and 57–70% for three-category classification. From this we conclude that our single-sensor system is comparable to humans for two-category counts, because our system can handle vehicles as rapidly as they occur, while humans can become overloaded and make errors when too many vehicles appear quickly.

We also must state that for three-category classification, humans remain more accurate when compared to our current system. This result is because, even though manual counts miss many vehicles, humans are much better at distinguishing "SUV-like" vehicles (type S*) than our system which uses simple length-based measures.

Table 11. Effect of careful feature extraction (single sensor)

Site	**simple** (peaks)	**careful feature** (edges) without consistency	with consistency
Two-category classification			
A	110 (67%)	133 (81%)	133 (81%)
BN	202 (81%)	209 (84%)	212 (85%)
Three-category classification			
A	55 (34%)	93 (57%)	94 (57%)
BN	157 (63%)	164 (66%)	174 (70%)

7.5 Effect of Careful Feature Extraction

We earlier discussed the importance of using distinct features, and of detecting common problems such as channelization error (Section 5.2).

To quantify the benefits of our improved feature detection; Table 11 compares overall classification accuracy of two and three-category analysis with and without these improvements. The three columns compare detection of simple peaks only, with detection of careful features (edges), with and without consistency checking.

As can be seen, the shift from peaks to edges is helpful at both sites. It is particularly helpful at site A (improving accuracy 14% for two-category and 23% for three-category) because there speeds are generally lower (8mph compared to 16mph at site BN), making wheel peaks much less distinct

Table 9. Classification Results, Trucks vs Non-Trucks

Site		Total considered	Correctly classified	Unable to classify [segmentation errors]	Incorrectly classified
A	Total	164 (100%)	133 (81%)	16 (10%) [9 SE]	15 (9%)
	T	8	6	2	0
	Non-T	156	127	14	15
BN	Total	248 (100%)	212 (85%)	18 (7%) [13 SE]	18 (7%)
	T	37	30	3	4
	Non-T	211	182	15	14

Table 10. Three-category Classification Results

Site		Total considered	Correctly classified	Unable to classify [segmentation errors]	Incorrectly classified			
					Total	Wrong as P	Wrong as S*	Wrong as T
A	Total	164 (100%)	94 (57%)	16 (10%)[9 SE]	54 (33%)			
	P	101	72	9	20	–	9	11
	S*	55	16	5	34	30	–	4
	T	8	6	2	0	0	0	–
BN	Total	248 (100%)	174 (70%)	18 (7%)[13 SE]	56 (23%)			
	P	127	90	13	24	–	17	7
	S*	84	54	2	28	21	–	7
	T	37	30	3	4	0	4	–

because of longer time spent over the sensor.

Consistency checking is most helpful at site BN (improving accuracy 1% and 4%) because that sensor suffered from a number of missing wheel cases. Site A suffers from a number of channelization errors which we detect but cannot correct. This detection does alter the quality estimate, however, allowing multi-sensor techniques to select the estimate without channelization error as described below.

7.6 Use of Sensor Fusion

Finally, we wish to investigate our hypothesis that multiple sensors can help resolve independent errors. In our experiment, 39 vehicles passed through both sites A and BN. We compare our two sensor fusion algorithms from Section 5.3, shortest-best and quality-best, with an oracle algorithm. Recall that the oracle algorithm takes the correct classification if either individual sensor is correct, thus providing a theoretical upper bound on performance.

Table 12 summarizes the results of this comparison, showing that our sensor fusion algorithms can correct several kinds of independent errors such as mis-segmentation and mis-channelization (described in Table 4). For two-category classification, both shortest- and quality-best always select the correct classification, matching the oracle. We cannot achieve more than the oracle unless we enhance individual sensor accuracy.

In three-category classification, the quality-best algorithm improves the accuracy by 5–14%. This improvement suggests that the confidence value we used for quality-best algorithm captures accuracy of the individual sensor's estimation and helps us to select the better classification. The shortest-best algorithm does not do as well as quality-best because it is designed to correct only segmentation errors and does not attempt to consider other kinds of errors.

Reviewing the errors from Table 4, we addressed mis-channelization error by detecting it at single sensors (Section 7.5), then selecting the best quality single-sensor with quality-best (Section 7.6). We handled imprecise speeds by selecting an appropriate sensor spacing and sampling frequency (Section 6). Changing speeds are best handled in single sensors. Segmentation errors are addressed by the shortest-best and quality-best sensor fusion algorithms (Section 7.6).

Table 12. Multi-sensor classification accuracy.

Total vehicles at both A and BN:	39 (100%)
Two-category classification	
single sensor:	
A alone:	36 (92%)
BN alone:	36 (92%)
multi-sensor combining A and BN:	
oracle:	38 (97%)
shortest-best:	38 (97%)
quality-best:	38 (97%)
Three-category classification	
single sensor:	
A alone:	24 (61%)
BN alone:	27 (69%)
multi-sensor combining A and BN:	
oracle:	32 (82%)
shortest-best:	25 (64%)
quality-best:	29 (74%)

8 Future Work

There are several areas of immediate future work, including revising our algorithms and completing system integration, classification systems, and developing and fielding more sophisticated sensor networks for vehicular applications. We have only begun to explore the use of Blade sensor data for vehicle classification and identification. A more careful analysis of both our single and multi-sensor approaches is needed. We also are in the process of integrating hardware (Blade sensors, Stargates, and supporting hardware) and software (drivers, data collection, and classification) to provide a complete, fieldable system.

One result of our work is to demonstrate that current classification schemes, designed for human observers, are poor matches for information available to computer sensors. We hope to evaluate the reasoning behind classification schemes and suggest approaches that are more sensor-friendly. The Federal Highway Administration (FHWA) has established thirteen vehicle classification standards for the purpose of collecting and analyzing traffic data [7]. Motivations for classification may include vehicle correlation with pavement damage, safety, or efficiency of road usage.

Although thus far we have explored vehicle classification, we plan to explore vehicle re-identification in the future.

9 Conclusions

This work has evaluated single- and multi-sensor techniques to improve accuracy of vehicle classification systems for urban roadways. Through extensive analysis of a field experiment we demonstrated that stand-alone sensors provide accuracy similar to human observers for two-category classification, and that accuracy in a three-category system is limited by the inherent ambiguity in mapping vehicle length to class. We showed the importance of extracting the most data possible from each individual sensor, and how quality estimates at individual sensors allow multiple sensors to improve accuracy further.

10 Acknowledgments

This work was supported in part by USC/CSULB METRANS 2003-04 grant #65A0047, SURE-SE. We would like to thank Steven Hilliard (IST) for his guidance on the IST sensors; and him and Paul Dell'aquila, Sara Hayden, and Dominic Adamietz (all of USC) for their assistance in conducting our August 2004 experiments. The authors will make our dataset available to other researchers upon request.

11 References

[1] National Highway Traffic Safety Administration. NHTSA Wheelbase and track width information. `http://www.nhtsa.dot.gov/`.

[2] Richard R. Brooks, Parmesh Ramanathan, and A. M. Sayeed. Distributed target tracking and classification in sensor networks. *Proceedings of the IEEE*, 91(8):1163–1171, August 2003. Invited Paper.

[3] Sing Yiu Cheung, Sinem Coleri, Baris Dundar, Sumitra Ganesh, Chin-Woo Tan, and Pravin Varaiya. Traffic measurement and vehicle classification with a single magnetic sensor. In *Proceedings of the Annual Meeting of the Transportation Research Board*. Transportation Research Board, January 2005.

[4] Sing Yiu Cheung, Sinem C. Ergen, and Pravin Varaiya. Traffic surveillance with wireless magnetic sensors. In *Proceedings of the 12th World Congress on Intellgient Transport Systems*, San Francisco, California, USA, November 2005.

[5] Intel Stargate Computer. Crossbow Technologies SPB400.

[6] Prabal Dutta, Mike Grimmer, Anish Arora, Steve Bibyk, and David Culler. Design of a wireless sensor network platform for detecting rare, random, and ephemeral events. In *Proceedings of the FourthIEEE International Workshop on Information Processing in Sensor Networks*, Los Angeles, California, USA, April 2005. IEEE.

[7] Federal Highway Administration and Office of Highway Policy Information. FHWA vehicle types. web page `http://www.fhwa.dot.gov/policy/ohpi/vehclass.htm` and `http://www.dot.state.oh.us/techservsite/availpro/Traffic_Survey/SchemeF/FHWA_Scheme_F_Report.PDF`, October 2003.

[8] Ford. Ford Vehicle. `http://www.fordvehicles.com/`.

[9] Lin Gu, Dong Jia, Pascal Vicaire, Ting Yan, Liqian Luo, Ajay Tirumala, Qing Cao, Tian He, John A. Stankovic, Tarek Abdelzaher, and Bruce H. Krogh. Lightweight detectionand classification for wireless sensornetworks in realistic environments. In *Proceedings of the Third ACM SenSys Conference*, pages 205–217, San Diego, California, USA, November 2005. ACM.

[10] John Heidemann, Fabio Silva, Xi Wang, Genevieve Giuliano, and Mengzhao Hu. *SURE-SE: Sensors for Unplanned Roadway Events—Simulation and Evaluation: Final Report*. USC/Information Sciences Institute, METRANS Transportation Center, May 2005.

[11] Inductive Signature Technologies Inc. IST sensor model 222.

[12] Cheol Oh, Stephen G. Ritchie, and Shin-Ting Jeng. Vehicle reidentification using heterogeneous detection systems. In *Proceedings of the 83rd Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2004. Transportation Research Board.

[13] Carlos Sun, Stephen G. Ritchie, and Seri Oh. Inductive classifying artificial network for vehicle type categorization. In *Proceedings of the Computer-Aided Civil and Infrastructure Engineering*, pages 18(3):161–172. Computer-Aided Civil and Infrastructure Engineering, 2003.

[14] Xi Wang, Fabio Silva, and John Heidemann. Infrastructureless location aware configuration for sensor networks. Technical Report ISI-TR-2004-589, USC/Information Sciences Institute, June 2004.

[15] Feng Zhao, Jaewon Shin, and James Reich. Information-driven dynamic sensor collaboration for tracking applications. *IEEE Signal Processing Magazine*, 19(2):61–72, March 2002.