Object-based Detection and Classification of Vehicles from High-resolution Aerial Photography

Ashley C. Holt, Edmund Y.W. Seto, Tom Rivard, and Peng Gong

Abstract

Vehicle counts and truck percentages are important input variables in both noise pollution and air quality models, but the acquisition of these variables through fixed-point methods can be expensive, labor-intensive, and provide incomplete spatial sampling. The increasing availability and decreasing cost of high spatial resolution imagery provides an opportunity to improve the descriptive ability of traffic volume analysis. This study describes an object-based classification technique to extract vehicle volumes and vehicle type distributions from aerial photos sampled throughout a large metropolitan area. We developed rules for optimizing segmentation parameters, and used feature space optimization to choose classification attributes and develop fuzzy-set memberships for classification. Vehicles were extracted from street areas with 91.8 percent accuracy. Furthermore, separation of vehicles into classes based on car, medium-sized truck, and buses/heavy truck definitions was achieved with 87.5 percent accuracy. We discuss implications of these results for traffic volume analysis and parameterization of existing noise and air pollution models, and suggest future work for traffic assessment using high-resolution remotely-sensed imagery.

Introduction

Accurate vehicle counts and truck percentages are important for traffic volume analysis, and serve as input variables to both noise pollution and air quality models. For example, vehicle traffic counts are directly linked to noise-related health impacts in urban environments (Seto *et al.*, 2007). In order to calculate noise volumes for a given highway or street segment, noise pollution models such as the Federal Highway Administration Traffic Noise Model (Menge *et al.*, 1998) require traffic volume and vehicle-type percentages along each street segment.

Likewise, traffic volumes, vehicle type distributions and vehicle densities are important inputs to models of urban vehicular emissions (Lyons *et al.*, 2003); these models, in conjunction with pollutant dispersion models, aid in the assessment of human exposure and overall air quality (Kunzli *et al.*, 2000). In addition to serving as important input variables to both noise pollution and air quality models, accurate vehicle counts also have application to transportation planning, traffic flow studies, pavement maintenance programs, environmental justice studies (Forkenbrock and Schweitzer, 1999; Forkenbrock and Weisbrod, 2001) and the promotion of pedestrian and bicycle safety (Wachtel and Lewiston, 1996).

There are several different methods, involving both direct and remote sensing, which can be used to detect vehicle traffic volumes along a given street segment. Technologies such as pneumatic tube systems are portable and can provide constant monitoring along a street or highway segment; on the other hand, this method is labor-intensive, involving the installation and subsequent removal of equipment for each monitoring location and time period, and in most cases does not provide the capability of detecting different vehicle types or sizes (Bellemans *et al.*, 2000). An alternate option is the use of remotely sensed counts derived from fixed-point counters, such as loop counters or video monitoring. For example, video and infrared monitoring can provide constant-stream information about traffic flows along selected streets (Graettinger *et al.*, 2005; Pless and Jurgens, 2004). One drawback with fixed-point counters, however, whether they involve direct or remote sensing technologies, is that sampling may be limited and is often conducted reactively, that is to say, on streets with known high traffic volumes or with pre-existing traffic or pedestrian-safety issues.

Over large metropolitan areas with complex traffic patterns, detection technologies which provide more complete sampling capabilities can add an additional layer of traffic information to data acquired from direct or fixed-point sensors. The increasing availability of affordable aerial photography and high-spatial resolution satellite imagery will provide new opportunities for deriving Annual Average Daily Traffic (AADT) (McCord *et al.*, 2002). In addition, these data could contribute to spatially-comprehensive assessments of noise and air pollution model parameters, such as traffic counts, as well as car, truck, and urban bus percentages.

Ashley C. Holt and Peng Gong are with the Department of Environmental Science, Policy, and Management, College of Natural Resources, University of California, Berkeley, CA (gong@nature.berkeley.edu).

Edmund Y.W. Seto is with Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA.

Tom Rivard is with the Department of Public Health, San Francisco, CA.

Photogrammetric Engineering & Remote Sensing
Vol. 75, No. 7, July 2009, pp. 871–880.

0099-1112/09/7507–0871/$3.00/0

This study presents an object-based approach to detect and classify vehicles from high-resolution aerial photography. While previous studies have focused on the detection of vehicles in remotely sensed imagery along highways (McCord *et al.*, 2002), this study extends an object-based approach to detect vehicles from four different street types (freeways/highways, arterials, bus routes, and residential streets) in a major metropolitan area. We conclude that vehicles can be accurately detected using an object-based target detection approach. More specifically, this study demonstrates that vehicles can be classified (e.g., as cars, medium trucks, and heavy trucks) using an object-based approach. We discuss implications of these results for traffic volume analysis and parameterization of existing noise and air pollution models, and suggest future work for traffic assessment using high-resolution remotely sensed imagery.

Study Area and Data Description

Aerial photographs of San Francisco, California (Figure 1) were acquired by HJW Geospatial using a Zeiss RMK TOP15 aerial survey camera, which captures red, green, and blue (RGB) portions of the spectrum at a resolution of around 0.17 meters. Images were acquired on 26–27 August 2004, in the afternoon between the hours of 1200 and 1900. Before analysis, the images were projected using Electronic Field Study 2.6 (Pictometry EFS). From a set of over 2,000 images, six images were selected for analysis. Two images (Figure 2a and 2b) were selected for training and cross-validation of the segmentation and classification processes, while the remaining four images (Figure 2c through 2f) were reserved for independent validation of the classification hierarchy and process. These images were selected from different neighborhoods and land-use types (residential, commercial, and industrial) in San Francisco, and were chosen because they captured a high density of vehicular traffic.

In addition to the aerial photographs, we also used GIS ancillary data layers for this analysis. These data layers included polygon outlines of the city blocks, sidewalks, and curbs in San Francisco, and street center lines for four different types of streets: freeways/highways, arterials (defined as major roads with large traffic volumes), bus routes (as outlined by current San Francisco Municipal Transportation Agency (MUNI) bus and light-rail routes maps), and residential streets.

Methods

An overview of the main steps in the segmentation and classification process is described in Figure 3.

Creation of Training and Validation Data Sets

In order to create training and validation data sets for classification, polygons (rectangles) were digitized around vehicle shapes along streets in each of the six study images. These images were loaded into an ArcGIS® 9.1 project (ESRI) along with neighborhood boundaries and street-centerline shapefiles for four types of streets: freeways/highways, arterials (major roads with large traffic volumes), MUNI bus routes, and residential streets. If a street segment fell into multiple categories, it was assigned to the category with the highest potential traffic volume (for example, a street segment that was both a bus route and a freeway/highway would be assigned to the "freeway/highway" category).

To evaluate vehicle type distributions and shape properties of different vehicle classes, we panned across the images, visually identifying and digitizing polygons

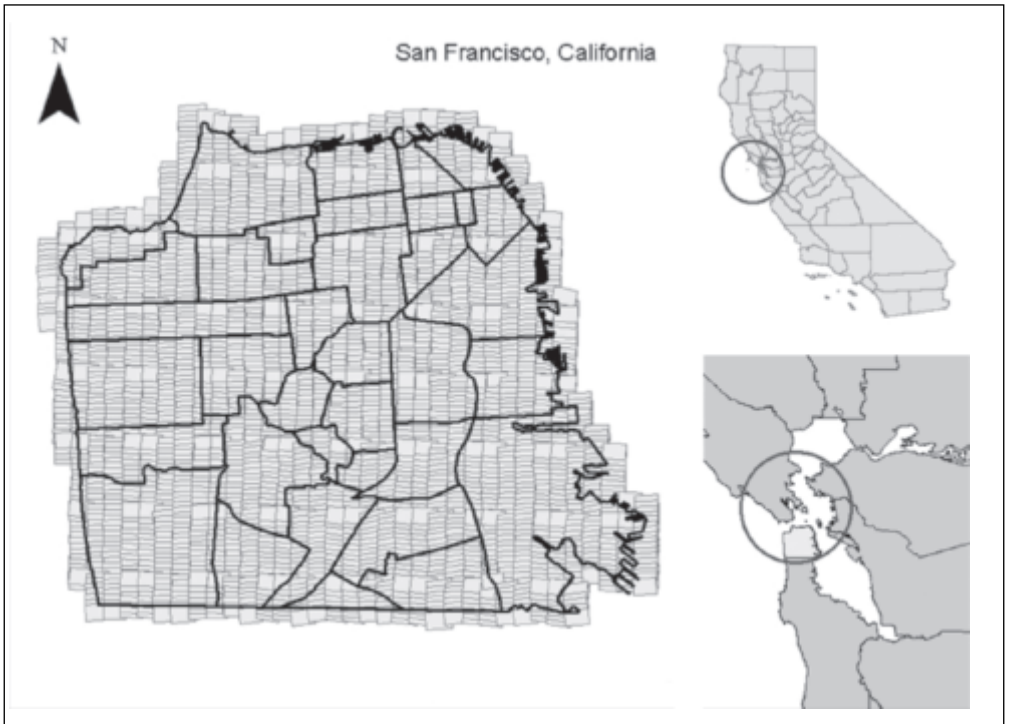

Figure 1. Study area, neighborhoods, and outlines of images. Over 2,000 images of San Francisco were acquired on 26–27 August 2004, in the afternoon between the hours of 1200 and 1900. Of these, six images were selected for analysis in the current study.

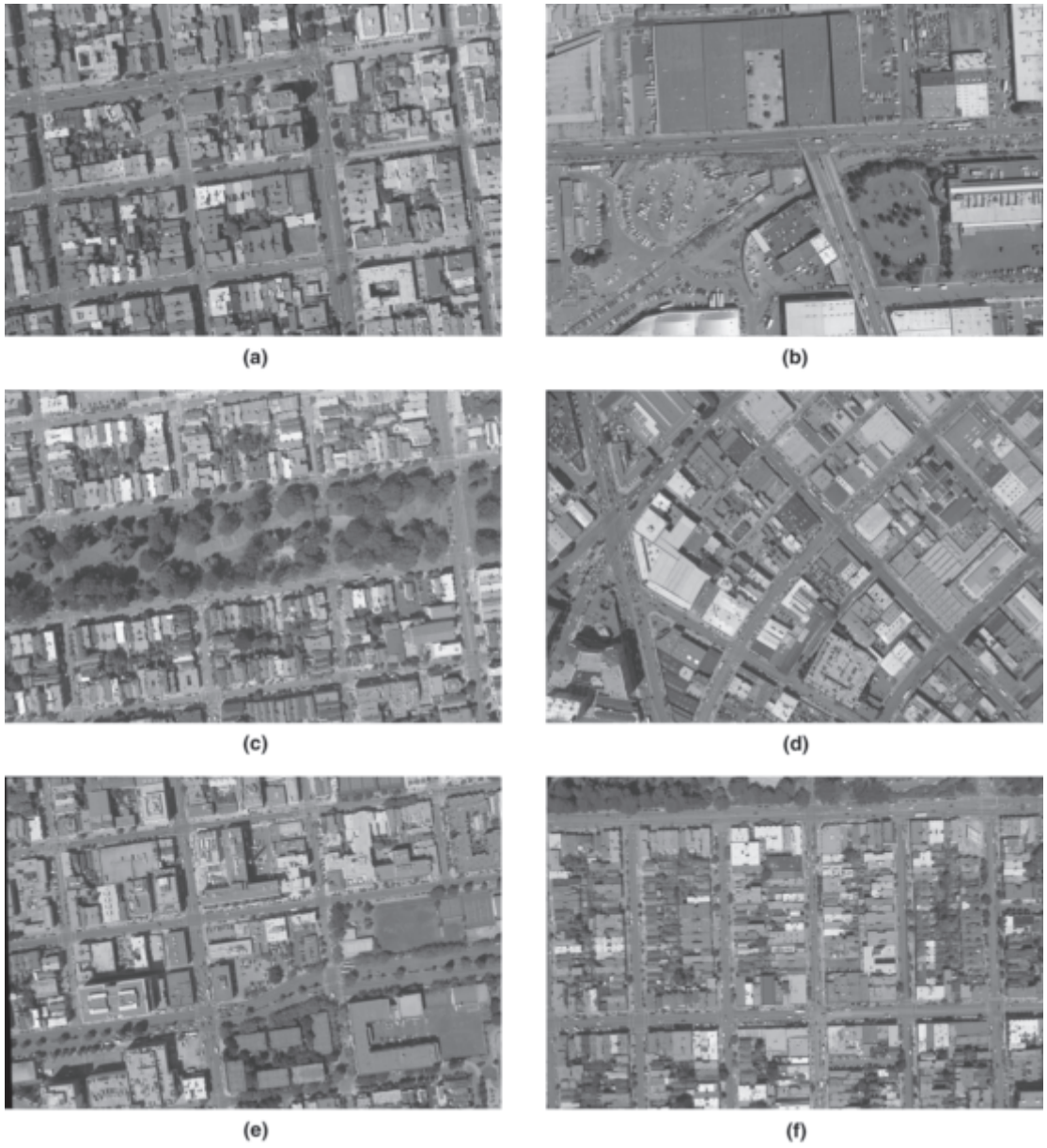

Figure 2. Training images (a) and (b), and validation images (c) through (f), shown in grayscale.

(rectangles) around vehicles. Each digitized vehicle shape was classified by visual interpretation into one of three vehicle classes (cars, medium trucks, and heavy trucks or buses), as defined by the Federal Highway Administration Traffic Noise Model (TNM) model parameters (Table 1).

We visually identified and digitized rectangular polygons around 553 vehicles from the six images. Of these 553 vehicle objects, 512 objects (92.6 percent) were identified as automobiles, 14 objects (2.5 percent) were identified as medium trucks, and 14 objects (2.5 percent) were identified as heavy trucks or buses. Histograms of vehicle sizes for each class were generated for later use in creating membership functions for fuzzy classification (Figure 4). This set of 553 digitized objects was split into two groups for training and validation. 130 objects were associated with the two training images used for determining segmentation parameters and developing the classification rules; the remaining 423 reference objects, which were digitized from the four images used for validation, were used to independently evaluate the accuracy of these processes. In addition to the reference objects, 500 non-vehicle points were digitized to evaluate commission error; 100 of these points were used for training of the object-based process and 400 were reserved for validation.

Multi-resolution Segmentation

The first step in object-oriented image classification is image segmentation, which involves the grouping of pixels into more meaningful object primitives. Correct image segmentation is a prerequisite to successful image classification. At the same time, this task requires explicit knowledge representation (Gong, 2003), as the definition of a "correct" or

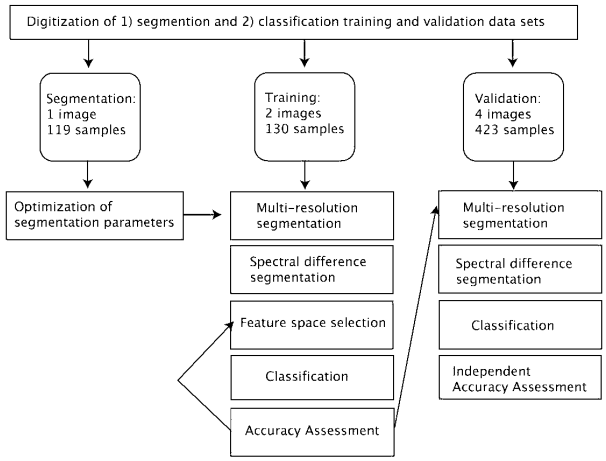

Figure 3. Schematic overview of segmentation and classification steps.

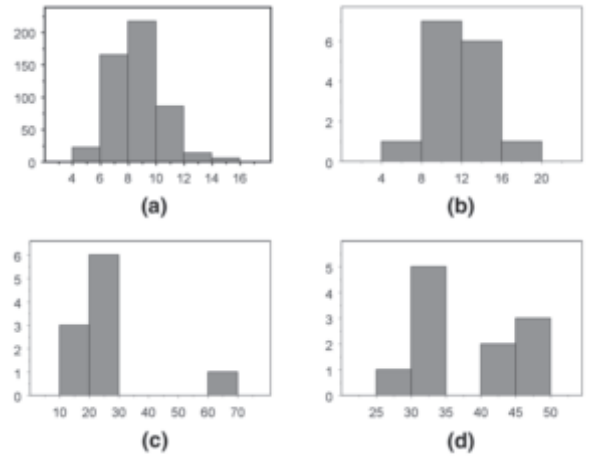

Figure 4. Histograms of vehicles sizes (in square meters) from the training samples created for each vehicle class: (a) cars, (b) medium trucks, (c) heavy trucks, and (d) buses.

"best" segmentation result or set of segmentation results will vary based on image type, classification scheme, and classification algorithm. Furthermore, optimal segmentation results are dependant on not only the choice of segmentation algorithm or procedure, but are also often influenced by the choice of user-defined parameter combinations which are required inputs for many segmentation programs (Radoux and Defourny, 2006; Carleer *et al.*, 2005).

While several studies (for example, Meinel and Neubert, 2004; Pal and Pal, 1993) have examined and compared various published, open-source, and commercially available segmentation algorithms, other studies (for example, Cardoso and Corte-Real, 2005; Zhang *et al.*, 2005) have focused on the question of how to conduct the evaluation of segmentation results, and on developing general methods for determining which technique or set of parameters produce the most meaningful object primitives. Zhang (1996) categorized the latter types of studies based on their focus on three methodological categories: analytical methods, empirical goodness methods, and empirical discrepancy methods. Analytical methods are those which analyze the algorithms themselves, and do not require application to a set of test images in order to evaluate the segmentation results. Empirical goodness methods are those which employ an unsupervised evaluation approach to test images, and rely on intrinsic properties of the resulting objects (such as region shape or contrast) to evaluate the segmentation results.

Evaluation methods which fall into the last category, often called empirical discrepancy methods, are generally considered more effective than the first two (Zhang, 1996). Empirical discrepancy methods are also called supervised evaluation methods, because they employ *a priori* knowledge about reference segmentation objects to determine the quality of image segmentation. Thus, segmentation can be accomplished in a number of different ways; while the spatial attributes of the best image segmentation or set of image segmentation layers will vary between applications, optimal segmentation results can be achieved for a given application by treating image segmentation as a supervised classification task.

In this study, we developed an empirical discrepancy method for choosing segmentation parameters such that the segmentation results are optimal given a set of training objects. We developed this method in the context of applicability to high spatial resolution remotely sensed data, and to address the need for a quantitative, user-supervised procedure for choosing best segmentation parameters. The goal was to develop an objective metric which maximizes the area matched and number of training objects (targets) matched, and which minimizes under- and over-segmentation for desired image object primitives.

For the proposed empirical discrepancy method of choosing segmentation parameter values, we defined optimal segmentation parameter values as those which

TABLE 1. VEHICLE CLASS DEFINITIONS (SOURCE: TNM MODEL PARAMETERS)

Vehicle Type	Class Definition
Cars/Automobiles (A)	All vehicles having two axles and four tires and designated primarily for transportation of nine or fewer passengers, i.e., automobiles, or for transportation of cargo, i.e., light trucks. Generally, the gross vehicle weight is less than 4,500 kg (9,900 lb).
Medium Trucks (MT)	All cargo vehicles having two axles and six tires. Generally, the gross vehicle weight is greater than 4,500 kg (9,900 lb) but less than 12,000 kg (26,400 lb).
Heavy Trucks (HT)	All cargo vehicles having three or more axles. Generally, the gross weight is greater than 12,000 kg (26,400 lb).
Buses (B)	All vehicles designed to carry more than nine passengers.

would (a) minimize the total area outside of training objects, for segmentation objects which overlapped training objects, and (b) in the case where different segmentation parameter combinations produced similar (optimal) outcomes for step (a), we defined optimal segmentation parameter values as those which would produce the least number of sub-objects which overlapped the boundaries of each training object. We defined the optimal results of a supervised segmentation approach as those that would optimize the ratio of training objects area to segmentation objects area, minimize under-segmentation, and minimize over-segmentation.

Goal 1: Maximize the Match of Training Object Area to Segmentation Object Area

Each training object a_i in the set of training objects A will have j segmentation objects (b_{ij}) whose centers are located within the boundaries of a_i. Ideally, for each training object a_i the total area of these j segmentation objects should equal the area of a_i. For each image segmentation result, we define the ***Area Matched*** criterion as:

$$\textbf{\textit{Area matched}} = \sum_{i=1}^{m} \frac{\sum_{j=1}^{n} Area(b_{ij})}{Area(a_i)}. \quad (1)$$

Goal 2: Minimize Under-segmentation

Each training object a_i in the set of training objects A will have j segmentation objects (b_{ij}) whose centers are located within the boundaries of a_i. If $j = 0$ for training object a_i, then we say that a_i has not been identified given that particular set of segmentation parameters, and the image is under-segmented for object a_i. Thus, to minimize the number of objects that are under-segmented, we maximize the proportion of objects which *are identified*. For each image segmentation result, we define the set A' as objects which have been identified

$$\boldsymbol{A}' = All\ a_i\ where\ (Center\ within(\ b_{ij}, a_i)) \quad (2)$$

and define the ***Proportion identified*** criterion as:

$$\textbf{\textit{Proportion identified}} = \frac{Cardinality(A')}{Cardinality(A)} \quad (3)$$

Goal 3: Minimize Over-segmentation

Each training object a_i in the set of training objects A will have j segmentation objects (b_{ij}) whose centers are located within the boundaries of a_i. In addition to meeting the first two criteria, an optimal segmentation result will produce one segmentation object that matches the geometry of each training object (ideally, a 1:1 ratio). For each image segmentation result, we define the set A' as objects which have been identified

$$\boldsymbol{A}' = All\ a_i\ where\ (Center\ within(b_{ij}, a_i)) \quad (4)$$

and the set B as segmentation objects whose centers are located within the boundaries of a member of set A'. For each image segmentation result, we define the ***Over-segmentation*** criterion as:

$$\textbf{\textit{Over-segmentation}} = \frac{Cardinality(B)}{Cardinality(A')} \quad (5)$$

Combining these criteria, we can define an objective metric which can be maximized to indicate the target segmentation parameters for a given application. Since the goal is to maximize the areas matched and the proportion identified, while minimizing over-segmentation, the objective metric is defined as:

$$\textbf{\textit{Optimal segmentation}} = \frac{AreaMatched \times ProportionIdentified}{Oversegmentation} \quad (6)$$

eCognition® 5.0 (Definiens, 2005) was used to conduct multi-resolution image segmentation, which involves knowledge-free extraction of image objects. Baatz and Schape (2000) describe the segmentation algorithm used by the eCognition® software, which begins with single-pixel objects and employs a region-growing algorithm to merge pixels into larger objects; pixels are merged based on whether they meet user-defined homogeneity criteria. Each multi-resolution segmentation task must be parameterized by the user, and involves settings of three parameters: Scale, Color-versus-Shape, and Compactness-versus-Smoothness. As defined by Baatz and Schape (2000), the Scale parameter controls the amount of heterogeneity allowed in the segmentation image objects. The Color-versus-Shape parameter defines the extent to which overall homogeneity is defined by the spectral homogeneity (as opposed to shape). The Smoothness-versus-Compactness parameter controls whether segmentation results are optimized for image objects in regard to smooth borders, or those which have more compact shapes. For segmentation training objects, we used 119 vehicle objects which we digitized from one of the training images (as previously described). We ran 150 independent segmentation trials using the parameter combinations described in Table 2. For each segmentation result, a value for the objective metric was calculated. Because the problem of optimizing segmentation parameters is a multi-response optimization problem, we also fit models to test the importance and optimal values of each of the segmentation parameters (scale, color/shape, and compactness/smoothness). We used JMP statistical software (SAS Institute) to fit a second-order polynomial model to describe the sensitivity of each defined segmentation criterion (***Area matched, Proportion identified, and Over-segmentation***) to the segmentation parameters. A nested model was fit to each response, and then terms that were not significant were removed to produce a reduced model. Prediction formulas were saved for each model, and displayed as overlapping contour profiles to identify a range for each of the Scale, Color, and Smoothness variables that would optimize each of the defined segmentation goals.

TABLE 2. 150 COMBINATIONS OF SEGMENTATION PARAMETERS

Parameter	Number of Settings	Values
Scale	6	10 20 30 40 50 60
Color/Shape	5	0.1/0.9 0.3/0.7 0.5/0.5 0.7/0.3 0.9/0.1
Compactness/Smoothness	5	0.1/0.9 0.3/0.7 0.5/0.5 0.7/0.3 0.9/0.1

TABLE 3. SIGNIFICANT PARAMETERS FOR REDUCED MODELS FOR EACH RESPONSE VARIABLE

Term	Proportion Identified	Over-segmentation	Area Matched
Scale	×	×	×
Color	×	×	×
Smoothness			
Scale * Color	×	×	×
Scale * Smoothness			
Color * Smoothness			
Scale^2		×	×
Color^2	×		×
Smoothness^2			

TABLE 4. eCOGNITION® SEGMENTATION PARAMETERS

	Parameter			Objects
Level	Scale	Color	Compactness	Average number
1	10	0.8	0.8	28474
2	36	0.6	0.5	14610

The reduced model provided insight into which segmentation parameters were most influential (Table 3). For this experiment, the Scale, Color, and (Scale * Color) interaction terms were important variables in predicting all three responses. The Smoothness term was not important. Based on the results of this analysis, segmentation parameters were chosen which optimized the objective metric. The selected parameters for the classification level, as well as a smaller-scale segmentation used for texture analysis, are described in Table 4.

Once the street areas were segmented based on the methods described above, we used a series of spectral difference segmentations to merge pavement objects in the "Street" portion of the image. Spectral difference segmentation merges objects which are below a user-defined threshold of spectral similarity. This allowed for background subtraction to eliminate objects which met spectral and texture criteria for classification as pavement, so that only objects which met spectral and geometric criteria for vehicle objects would be considered for classification. This process yielded a set of objects in the street areas of the image which were then considered as candidates for classification as vehicles.

Feature Space Optimization and Feature Selection

The eCognition® 5.0 software provides the user with many different feature choices which can be combined to develop classification rules. These features include spectral, texture, object shape, and scene descriptions. In order to choose a subset of these attributes which would be the most effective for this application, we used the eCognition® Feature Space Optimization tool (Definiens™) to choose the object features which would be most effective at separating the vehicle classes from each other and from surrounding non-vehicle objects, such as vegetation and pavement. We chose 10 to 15 sample objects for each category of vehicle and non-vehicle objects. We chose 25 potential classification features, based on *a priori* knowledge of likely classification attributes. These candidate features included geometric properties of image objects, as well as spectral and texture information derived from the images. The Feature Space Optimization tool was used to select the best features for separating the classes. Based on results from this step, three shape features (Main Direction, Density, and Rectangular Fit) and two texture features (Density of sub-objects and Mean of sub-objects) were identified as being important for separating the classes; these features are described in detail in Table 5. In addition to these features, we used the layer values as well as knowledge about the sizes of different vehicles (developed during the training object digitization step, and described in Figure 4) to create fuzzy set memberships for each of the vehicle classes.

TABLE 5. DESCRIPTION OF VARIABLES SELECTED FOR CLASSIFICATION BASED ON FEATURE OPTIMIZATION (FEATURE DESCRIPTIONS FROM eCOGNITION® 5 REFERENCE MANUAL)

Feature	Description
Layer Values	
Color: Layer values for R, G, B	Layer mean value calculated from the layer values of all pixels forming an image object
Standard Deviation: Layer values for R, G, B	Standard deviation calculated from the layer values of all pixels forming an image object
Shape	
Main direction	The direction of the eigenvector belonging to the larger of the two eigenvalues derived from the covariance matrix of the spatial distribution of the image object
Density	The area covered by the image object divided by its radius
Rectangular Fit	Percentage of rectangle that fits inside a rectangular approximation of the image object
Texture	
Density of sub-objects: stddev	Standard deviation calculated from the densities of the subobjects
Mean of sub-objects: stddev	Layer mean values calculated from the subobjects

Classification and Extraction of Vehicle Shapes

An object-based classification was conducted, using the TNM vehicle class definitions and fuzzy rulesets that were created based on the results of feature space optimization. The eCognition® classification process is object-oriented, which means that child objects inherit properties from parent classes. Following this model, the images were first classified into *City Blocks*, *Sidewalks*, and *Street* areas based on GIS thematic layers. Objects in the *Street* areas were then classified as vehicles based on the fuzzy rulesets that we developed. Finally, these *Vehicles* were classified as cars, medium trucks, and heavy trucks or buses. The classification hierarchy is illustrated in Figure 5. The class definitions utilized information about object texture, shape, and proximity to curb areas. Vehicles were further classified as either heavy trucks or automobiles/cars based on texture and vehicle dimension estimates. An example of the classification result is shown in Figure 6.

Validation

The segmentation and classification rules were first tested on the two validation images. After satisfactory cross-validation results were achieved, the classification hierarchy was saved and applied to the four independent validation images (the images which had not been used to create the classification hierarchy). The results were assessed to validate the accuracy of the classification scheme, as well as its performance across several images from different city neighborhoods and street types. The classified objects were exported from eCognition® as shapefiles, and evaluated in ArcGIS® to create error matrices for both the training image and objects, as well as for the validation images and objects.

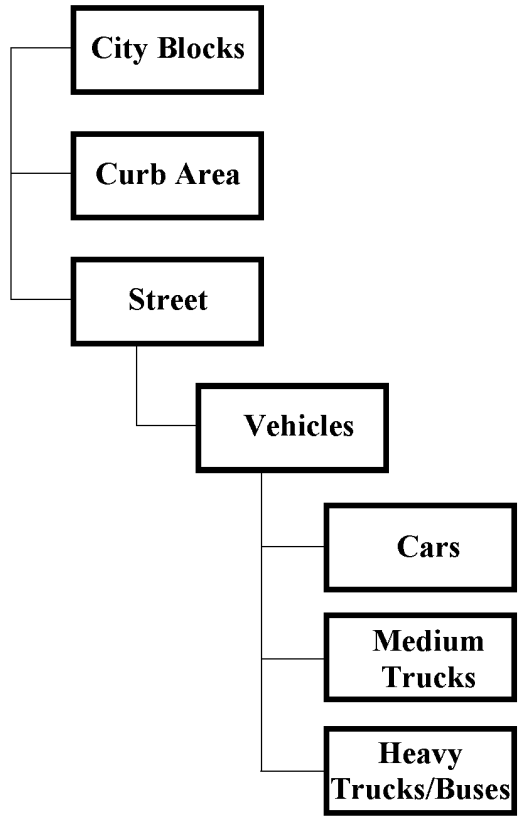

Figure 5. Object-oriented classification hierarchy. At the top level, the images are classified into City Blocks, Curb Areas, and Street areas. Within the Street areas, objects are further identified as Vehicle candidates based on background subtraction of pavement areas. These Vehicle candidate objects are then classified as Cars, Medium Trucks, and Heavy Trucks or Buses based on attributes derived from feature space optimization.

Results

For these training images, vehicles were extracted from street areas with 90 percent accuracy (Table 6). Classification of vehicles as cars, medium-sized trucks, and buses/heavy trucks was achieved with 88.7 percent accuracy (Table 7). For the four images used for independent validation, vehicles were extracted from street areas with 91.8 percent accuracy (Table 8). Classification of vehicles as cars, medium-sized trucks, and buses/heavy trucks was achieved with an overall accuracy of 87.5 percent. Cars were classified with 78.7 percent accuracy, medium trucks were classified with 64.3 percent accuracy, and heavy trucks and buses were classified with 57.1 percent accuracy (Table 9).

Discussion

In this study, vehicles were extracted from street areas with 91.8 percent accuracy, and further separation of vehicles into classes based on car, medium-sized truck, and buses/heavy truck definitions was achieved with 87.5 percent accuracy. The remaining confusion between cars, medium trucks, and heavy trucks/buses can be attributed to two primary factors. First, the error may be associated with the relatively small sample size of training objects for the latter two classes (medium trucks and heavy trucks/buses), which reflects the average distribution of different vehicle types in this city (i.e., cars are the prevalent vehicle class, and it is more difficult to find numerous samples of the truck and buses). The classification accuracy would likely be improved by additional imagery, additional sampling, and rule-set modification for these vehicle classes. The development of a "library" of vehicle samples from different metropolitan areas would also be useful future work in extending these results.

Secondly, classification error may be related to the potential for user error in creating the digitized training and testing sets. In this study, the creation of training and testing data involved the translation of vehicle class definitions, which are based on weight and axle size, into photo-interpretation decisions that are made from an aerial perspective. Particularly for the medium trucks class, it was sometimes difficult to determine from an aerial view the number of axels and weight of the vehicle. Other possible vehicle definitions that could be used include the Federal Highway Administration's commercial vehicle length and width standards (FHWA, 1996).

Vehicle classification accuracy could be improved by including additional geospatial, spectral, and temporal information in the analysis. For example, the time at which each aerial photograph was taken could be compared with the time designations for known "no parking" or "tow away" zones expand or contract the area of analysis for moving vehicles, as opposed to including only street areas beyond a fixed distance from the street curb. Secondly, the identification of shadowed areas could be used to both adjust classification of vehicles within shadowed areas, as well as to boost classification of street objects by using shadows to determine the height of the object (Niu *et al.*, 2002). Alternately, lidar data can be used to support

Figure 6. Example of classification results. City blocks and curbs are shown in blue; street areas are shown in white. Classified vehicles are shown in red (cars), orange (medium trucks) and yellow (heavy trucks and buses). A color version of this figure is available at the ASPRS website: **www.asprs.org**.

TABLE 6. VEHICLE TARGET DETECTION RESULTS AND ACCURACY FOR TRAINING IMAGES

Classes		Reference: Vehicle	Reference: Other	Sum	User's Accuracy
Classification results	Vehicle	**108**	1	109	0.99
	Other	22	**99**	120	0.82
Sum		130	100	230	
Producer's Accuracy		0.831	0.99	Overall accuracy:	**0.90**
				Kappa value:	**0.801**

vehicle-height detection (Toth and Grejner-Brzezinska, 2006). Finally, overlapping images could be used for change detection, to improve the accuracy of classification for vehicles versus objects that are actually pavement or non-moving objects.

We used GIS ancillary data to mask out blocks and curb areas, which allowed for a focused analysis of the street areas. However, this ancillary data may not be available in all situations (e.g., in cities which do not maintain geodatabases with block and curb spatial data). In the absence of this type of pre-existing information, the extraction of roads networks from aerial photographs or satellite imagery is possible, as demonstrated by several studies (Gao and Wu, 2004; Gong and Wang, 1997; Haverkamp, 2002; Unsalan and Boyer, 2004; Xiao, Tan and Tay, 2005). These types of techniques could be applied as a preliminary step to mask out non-street areas of the imagery.

Video and infrared monitoring can provide constant-stream information about traffic flows along selected streets (Graettinger *et al.*, 2005), but may provide limited spatial sampling. The methods outlined in this study provide a useful guideline for conducting traffic distribution analysis over large study areas. The independent validation demonstrates that this object-based method can be applied accurately across various street types and neighborhoods. While the application of aerial photography for traffic analysis may be limited by its low temporal resolution, the increasing availability of affordable aerial photography and high-spatial resolution satellite imagery (such as Ikonos and QuickBird) will provide new opportunities for spatially-comprehensive assessments of noise and air pollution model parameters, such as traffic counts, as well as car, truck, and urban bus percentages (Sharma *et al.*, 2006). Furthermore, the increased availability of high resolution imagery through technologies like Google™ Earth, combined with detection methods such as those outlined in this study, can provide a standardized method for comparisons of traffic across communities around the world.

Conclusions

The results of this study have application in transportation planning, traffic flow studies, pavement maintenance programs, environmental justice studies, and the promotion of pedestrian and bicycle safety. Traffic impacts health through

TABLE 7. CLASSIFICATION RESULTS AND ACCURACY FOR TRAINING IMAGES

		Reference					
Classes		Car	Medium Truck	Heavy Truck / Bus	Other	Sum	User's Accuracy
Classification results	Car	**95**	2	0	1	98	0.97
	Medium Truck	1	**5**	0	0	6	0.83
	Heavy Truck / Bus	0	0	**5**	0	5	1
	Other	21	1	0	**99**	121	0.82
Sum		117	8	5	100	**230**	
Producer's Accuracy		0.812	0.625	1	0.99	Overall Accuracy:	**0.887**
						Kappa value:	**0.796**

TABLE 8. VEHICLE TARGET DETECTION RESULTS AND ACCURACY FOR VALIDATION IMAGES

		Reference			
Classes		Vehicle	Other	Sum	User's Accuracy
Classification results	Vehicle	**364**	8	373	0.978
	Other	59	**392**	451	0.869
Sum		424	400	**824**	
Producer's Accuracy		0.860	0.98	Overall accuracy:	**0.918**
				Kappa value:	**0.838**

TABLE 9. CLASSIFICATION RESULTS AND ACCURACY FOR VALIDATION IMAGES

		Reference					
Classes		Car	Medium Truck	Heavy Truck / Bus	Other	Sum	User's Accuracy
Classification results	Car	**311**	4	0	2	317	0.981
	Medium Truck	20	**9**	6	4	39	0.231
	Heavy Truck / Bus	6	0	**8**	2	16	0.5
	Other	58	1	0	**392**	451	0.869
Sum		395	14	14	400	**823**	
Producer's Accuracy		0.787	0.642	0.571	0.98	Overall Accuracy:	**0.875**
						Kappa value:	**0.771**

multiple pathways, such as air quality, noise pollution, pedestrian injuries, and neighborhood walkability. In addition, the extent of these impacts not only depends upon traffic volumes but traffic type. Thus, the ability to understand these impacts at a small scale using detailed vehicle identification is an important input for identifying community-level health impacts. This study presents an object-based approach for detecting and classifying vehicles from high-resolution aerial photography. The results of this study demonstrate that vehicles can be accurately detected using an object-oriented target detection approach. The collection of traffic counts is a valuable input for planning and environmental analysis; however, it also represents a costly burden for local and state government. The use of remote sensing techniques can reduce cost associated with traffic analysis, and expand traffic analysis to both highways and other street types. The increased availability of high spatial resolution remotely-sensed data can also enable more complete analysis of traffic-related urban management issues, even in cities or areas where traditional ground-based methods are not available.

Acknowledgments

The authors thank Lynette Yang for contributing to the development of the training and validation data sets for this study, Qian Yu, and the three anonymous reviewers for their constructive feedback on this manuscript. This research was funded in part by a grant from the University of California Toxic Substances Research & Teaching Program, and the City of San Francisco Department of Building Inspection. The Geospatial Information and Imaging Facility at UC Berkeley provided use of computer facilities and Definiens eCognition® 5.0 software.

References

Baatz, M., and A. Schape, 2000. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation, *Angewandte Geographische Informationsverarbeitung XII*, Karlsruhe, pp. 12–23.

Bellemans, T., B.D. Schutter, and B.D. Moor, 2000. On data acquisition, modeling and simulation of highway traffic, *Proceedings of the 9th IFAC Symposium on Control in Transportation Systems 2000*, 13–15 June, Braunschweig, Germany, pp. 22–27.

Cardoso, J.S., and L. Corte-Real, 2005. Toward a generic evaluation of image segmentation, *IEEE Transactions on Image Processing*, 14(11):1773–1782.

Carleer, A.P., O. Debeir, and E. Wolff, 2005. Assessment of very high resolution satellite image segmentations, *Photogrammetric Engineering & Remote Sensing*, 71(11):1285–1294.

FHWA, 1996. *Federal Size Regulations for Commercial Motor Vehicles*, U.S. Department of Transportation, Federal Highway Administration.

Forkenbrock, D.J., and L.A. Schweitzer, 1999. Environmental justice in transportation planning, *Journal of the American Planning Association*, 65(1):96–111.

Forkenbrock, D.J., and G.E. Weisbrod, 2001. *NCHRP Report 456, Guidebook for Assessing the Social and Economic Effects of Transportation Projects*, Transportation Research Board.

Gao, J., and L. Wu, 2004. Automatic extraction of road networks in urban areas from IKONOS imagery based on spatial reasoning, *Proceedings of the XXth ISPRS Congress*, 12–23 July, Istanbul, Turkey, unpaginated CD-ROM.

Gong, P., 2003. Information extraction for human settlements, *Manual of Remote Sensing* (A. Renz, editor), American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland.

Gong, P., and J. Wang, 1997. Road network extraction from airborne digital camera data, *Geographic Information Sciences*, 3(1–2):51–59.

Graettinger, A.J., R.R. Kilim, M.R. Govindu, P.W. Johnson, and S.R. Durrans, 2005. Federal Highway Administration vehicle classification from video data and a disaggregation model, *Journal of Transportation Engineering*, 131(9):689–698.

Haverkamp, D., 2002. Extracting straight road structure in urban environments using IKONOS satellite imagery, *Optical Engineering*, 41(9):2107–2110.

Kunzli, N., R. Kaiser, S. Medina, M. Studnicka, O. Chanel, P. Filliger, M. Herry, J.F. Horak, V. Puybonnieux-Texier, P. Quenel, J. Schneider, R. Seethaler, J.C. Vergnaud, and H. Sommer, 2000. Public-health impact of outdoor and traffic-related air pollution: A European assessment, *The Lancet*, 356(9232):795–801.

Lyons, T.J., J.R. Kenworthy, C. Moy, and F.D. Santos, 2003. An international urban air pollution model for the transportation sector, *Transportation Research Part D*, 8:159–167.

McCord, M., P. Goel, Z. Jiang, B. Coifman, Y. Yang, and C. Merry, 2002. Improving AADT and VDT estimation with high-resolution satellite imagery, *Proceedings of the Pecora 15/Land Satellite Information IV Symposium 2002*, 10–15 November, unpaginated CD-ROM.

Meinel, G., and M. Neubert, 2004. A comparison of segmentation programs for high resolution remote sensing data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXV(B4):1097–1102.

Menge, C., C. Rossano, G. Anderson, and C. Bajdek. 1998. *FHWA Traffic Noise Model (TNM), Final Report, FHWA-PD-95-010*, U.S. Department of Transportation, Federal Highway Administration.

Niu, X., F. Xu, R. Li, C. Merry, M. O'Kelly, and T. Matisziw, 2002. Detection of trucks from georeferenced aerial photographs, *Proceedings of the ACSM-ASPRS Conference and Technology Exhibition*, 19–26 April, Washington, D.C., (XXII FIG International Congress), unpaginated CD-ROM.

Pal, N.R., and S.K. Pal, 1993. A review on image segmentation techniques, *Pattern Recognition*, 26:1277–1294.

Pless, R., and D. Jurgens, 2004. Road extraction from motion cues in aerial video, *Proceedings of the ACM Conference on Geographic Information Systems*, pp. 31–38.

Radoux, J., and P. Defourny, 2006. Influence of image segmentation parameters on positional and spectral quality of the derived objects, *Proceedings of the 1st International Conference on Object-based Image Analysis (OBIA 2006)*, unpaginated CD-ROM.

Seto, E.Y.W., A. Holt, T. Rivard, and R. Bhatia, 2007. Spatial distribution of traffic induced noise exposures in a US city: An analytic tool for assessing the health impacts of urban planning decisions, *International Journal of Health Geographics*, 6(24), URL: *http://www.ij-healthgeographics.com/content/6/1/24* (last date accessed: 10 April 2009).

Sharma, G., C.J. Merry, P. Goel, and M. McCord, 2006. Vehicle detection in 1-m resolution satellite and airborne imagery, *International Journal of Remote Sensing*, 27(4):779–797.

Toth, C.K., and D. Grejner-Brzezinska, 2006. Extracting dynamic spatial data from airborne imaging sensors to support traffic flow estimation, *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(3–4):137–148.

Unsalan, C., and K.L. Boyer, 2004. A system to detect houses and residential street networks in multispectral satellite images, *Proceedings of the 17th International Conference on Pattern Recognition*, 23–26 August, Cambridge, UK, (IEEE), unpaginated CD-ROM.

Wachtel, A., and D. Lewiston, 1996. Risk factors for bicycle-motor vehicle collisions at intersections, *Journal of Safety Research*, 27:195–195.

Xiao, Y., T. Tan, and S. Tay, 2005. Utilizing edge to extract roads in high-resolution satellite imagery, *Proceedings of the IEEE International Conference on Image Processing*, 11–14 September, pp. 637–40.

Zhang, H., A. Fritts, and S.A. Goldman, 2005. A co-evaluation framework for improving segmentation evaluation, *Proceedings of SPIE Signal Processing and Target Recognition XIV*, pp. 420–430.

Zhang, Y.J., 1996. A survey on evaluation methods for image segmentation, *Pattern Recognition*, 29(8):1335–1346.