Estimation of traffic parameters in urban areas from satellite images

Jens Leitloff*, Stefan Hinz[†] and Uwe Stilla*

*Photogrammetry and Remote Sensing, [†]Remote Sensing Technology
Technische Universität München, Arcisstrasse 21, D-80333 München, Germany
{Jens.Leitloff, Uwe.Stilla, Stefan.Hinz}@bv.tum.de

Abstract— In this paper an approach for automatic vehicle detection from satellite imagery is presented. In space-borne images single vehicles have a very limited size and a detection scheme which only focuses on single vehicles would deliver ambiguous results. Therefore we extract vehicle queues because their repetitive pattern improves the distinction from similar objects. Additionally information about the location and approximately size of roads are used to limit the search space and thus decrease the number of false detection. Single vehicles can be determined by analysis of the width and contrast behavior of the queues. The results show that the analysis of width and contrast information extracts vehicles with high reliability, but since only a certain number of vehicles appear grouped, the overall completeness is not sufficient yet. The used data sets are QuickBird images covering the inner city of Munich (Germany) with a ground sampling distance of approximately 0.6 m for the panchromatic channel.

I. INTRODUCTION

A. Motivation

Recently, vehicle detection from remotely sensed data has achieved considerable attention on international conferences (e.g. [1], [2]) and journals [3]). Besides the scientific challenge of identifying, classifying and possibly tracking objects that are relatively small compared to the Ground Sampling Distance (GSD) of the sensors used, vehicle detection is motivated by different fields of application: For active traffic flow management, e.g. near real time information, which traditionally comes from permanent installed sensors like inductions loops or stationary cameras, is needed. Among others, approaches for acquisition of near real-time data from aerial images and LiDAR [4] or video cameras [5] show great potential of the usage of airborne sensors for this task. Furthermore, traffic-related data plays an important role in urban and spatial planning, i.e., for road planning, for estimation or simulation of air and noise pollution, etc. An algorithm that automatically detects and counts vehicles in airborne or spaceborne images would effectively support traffic-related analyses in urban planning. Other fields of application are found in the context of military reconnaissance and extraction of geographical data for Geo-Information Systems (GIS), e.g., for site model generation and update.

While airborne imagery usually has a GSD of approx. 10-25cm, which is very advantageous for detection and classification of vehicles, the spatial coverage of these data is quite limited. On the other hand new optical sensor systems on satellites provide images with a resolution of one meter or better and more extended ground coverage than airborne systems. The GSD of satellite imagery seems good enough - at least visually - to derive traffic relevant data. Although the coarser resolution is certainly a limiting factor in terms of detection performance of single vehicles, the potential of acquiring wide-area traffic data allows the derivation of very valuable traffic parameters on a global scale. Such information comprises, for instance, the traveling time at the time of image acquisition which helps to indicate the driver the currently best route to his destination. Due to long revisit intervals of current high resolution satellite systems this application is quite limited at the moment. On the other hand, also the existing satellite systems allow for the estimation of statistical, long-time parameters like traffic density per road segment, per route, or per (sub-) network. Such parameters can be derived from each (cloud-free) satellite's pass and help to validate and improve traffic models since they complement other traffic parameters estimated from sources like induction loops.

B. Related Work - Overview

Approaches for vehicle detection can be analyzed and grouped from different points of view. We have chosen a categorization based on the underlying type of model used for the extraction:

- Implicit models (for single vehicles)
- Explicit models (for single vehicles)
- Global models (for vehicle queues)

Approaches using an implicit model consider the vehicle detection as a classification problem. During a training phase the classification system learns the model parameters through sample images, i.e. in an appearance-based way. Afterward the image is evaluated pixel-by-pixel or tile-by-tile with respect to the learned model and classified into "vehicle" and "non-vehicle" regions. Main advantages of these approaches are their good generalization property and flexibility since specific object features do not need to be modeled. On the other hand, the detection performance is strongly correlated with the choice of representative training data. Examples for this type of modeling are the approaches presented in [6] and [7]. The latter reach promising results for images taken with terrestrial cameras. However, our own experiments showed that, in aerial

and space images, vehicles have a much greater radiometric variety because of reflections, specularities, shadows, or other influences of adjacent objects. Hence, it seems, in particular, hard to set up representative training data.

Nevertheless, due to the coarse resolution of satellite images, implicit modeling seems to be the method of choice for single vehicles. The few approaches for extraction of vehicles from satellite images also use implicit modeling. A comparison of different methods for vehicle detection from simulated satellite imagery of simple highway scenes can be found in [8]. Since most of the presented algorithms need a reliable background images, they are limited to certain scenes. A more advanced approach based an neural- networks can be found in [9]. They achieve good result for isolated vehicles in relatively simple scenes.

Approaches which are based on explicit modeling use specialized filters or two- or three-dimensional wire frame representations for describing a vehicle. Extraction is done by matching the model to the image "top-down" and evaluating the match according to the support found in the image (see [10] & [11]). Alternatively extracted image primitives can be grouped "bottom-up" to construct a description similar to the predefined model (e.g. [12]). An important prerequisite for successful applications is the abibility to detect sub-structures like hood or windshield, since they are the key features to make the model distinctive compared to similar objects. As it can be seen from Fig. 1 , in satellite imagery vehicles appear only as blobs with barely visible sub-structures. Therefore, an explicit model seems not applicable.

Vehicles show quite regular pattern if they appear grouped (e.g. in traffic jams or parking lots), the use of these repeating occurrences during the extraction we refer to as global modeling. The approach presented in [13] directly focuses on vehicle queues. They use the spectral signature of small image patches derived from a Fourier transform to extract regular object configurations (i.e. military convoys). Even if high resolution imagery is available, the use of global models seems useful to increases the detection quality as show in [11]. There the initial extraction is realized by a comprehensive explicit model. Afterward missed vehicles can be found by applying a less constrained vehicle model to gaps of the queues. In doing so, not only queues but also isolated cars can be extracted as long as they belong to the set of reliable hypotheses.

Summarizing the discussion and keeping the appearance of vehicles in (sub-) meter resolution satellite images in mind, following conclusions for a vehicle detection scheme can be drawn:

In general, the resolution is too coarse to contain detailed information like sub-structures of cars. Reliable hypotheses are not to expect if only local features are used. Furthermore, explicit modeling is only helpful for larger features which relate to the complete vehicle such as length, width, mean contrast, etc. An implicit model is applicable to isolated cars as long as their appearance is not much influenced by adjacent objects, shadows, or specularities. Hence, it seems not much superior to an explicit blob model. Utilizing a global model
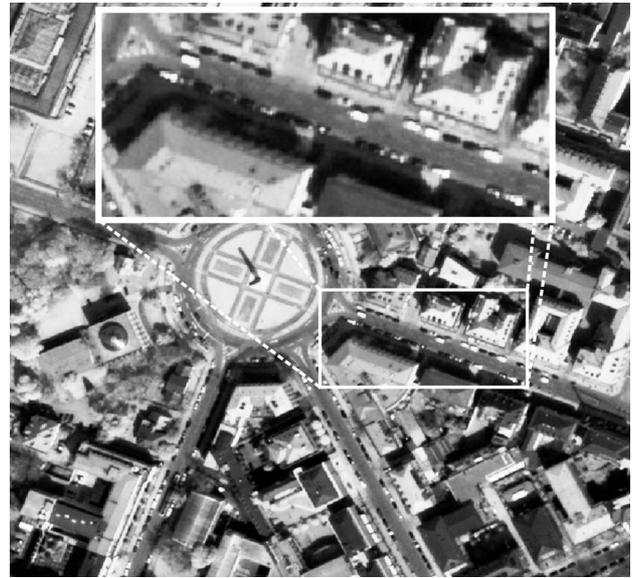

Fig. 1. Vehicles in QuickBird imagery, GSD = 0.6 m

for vehicle queues is supposed to deliver the most promising initial hypotheses, since its repetitive pattern makes a vehicle queue distinctive as a whole compared to other objects.

The use of context information is necessary to compensate, at least partially, for the missing detailed image information. However, context information - may it stem from a digital map or from extraction of related objects like roads, buildings, or vegetation - should be regarded as uncertain, although being very helpful in terms of reducing the search space and providing Regions of Interest.

Consequently, our vehicle detection scheme focuses primarily on the extraction of vehicle queues. After validation and selection, the remaining hypotheses may serve as starting point to further complete the overall detection by searching for isolated vehicles in their neighborhood.

II. Vehicle Detection

In this section the underlying queue model is introduced, before the queue and single vehicle extraction is described in detail.

A. Vehicle Queue Model

As summarized in the previously section it is helpful to use a global model for the initial extraction. Therefore, we define a vehicle queue as a ribbon with distinct symmetries along and across its local orientation. This model is essentially identical to that introduced in [11] and has also been used in previous publications (e.g. [14]). Since this model is originally designed for aerial images, a number of modifications regarding the significance of different features must be applied:

Therefore we define, that a vehicle queue

- needs sufficient length, limited width and low curvature;
- shows a repetitive pattern along the medial axis, both in contrast and width (Fig. 2(a))

- has a length and width that correspond to the vehicle dimensions;
- collapses to a line in Gaussian scale space, i.e. when smoothing the image accordingly (Fig. 2(b)).

We would like to emphasize, that this queue model differs from the previously mentioned approaches in a way that - in particular, through the scale-space description - the queue is modeled as a unique structure and not just as a composite of its underlying, smaller elements. At first glance, this seems to be of little importance. Still, it provides the basis for detecting a queue hypothesis as a whole (even though at a coarser scale) rather than constructing it from smaller elements. Thereby global knowledge can be incorporated from the very beginning of the extraction.

B. Vehicle Queue Extraction

As mentioned in the introduction the use of additional information, i.e. road location, is necessary to compensate the low resolution. Therefore GIS data is used to determine Region of Interest.

Then the line extraction is carried out by applying a differential geometric approach [15]. This algorithm is primarily based on the computation of the second image derivatives, i.e. the local curvatures of the image function. Parameters for the line extraction are chosen corresponding to vehicle geometry (vehicle width: w) and radiometry (expected contrast to road: c). Thus, the necessary input parameters for line extraction t_L, t_H - which will be explained in detail afterward - can be calculated as:

$$\sigma = \frac{w}{2\sqrt{3}} \qquad (1)$$

$$t = c\frac{-w}{\sqrt{2\pi}\cdot\sigma^3}e^{-\frac{1}{2}\left(\frac{w}{2\sigma}\right)^2} = c\frac{-\sqrt{6}\cdot12}{\sqrt{\pi}\cdot w^2}e^{-\frac{3}{2}} = c\cdot a$$

$$t_L = c_L \cdot a, \qquad t_H = c_H \cdot a$$

where σ defines the preliminary smoothing factor, calculated from the maximum expected width (e.g. 2.5 meter). t_L and t_H define the hysteresis thresholds for the second partial derivate of the image at each point. If the value is higher than t_H a point is immediately accepted as a line point. All points with a smaller second derivate than t_L are rejected. Points between t_H and t_L are accepted if they can be connected to
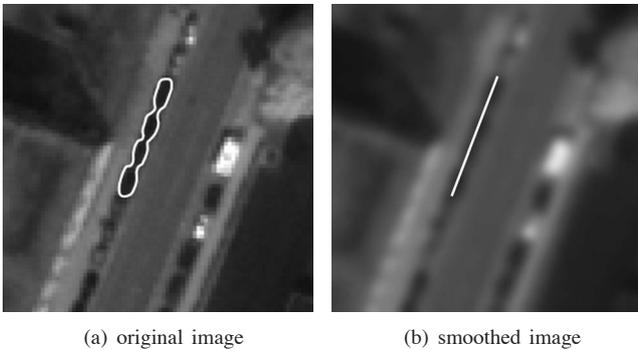

(a) original image   (b) smoothed image

Fig. 2.   Queue model
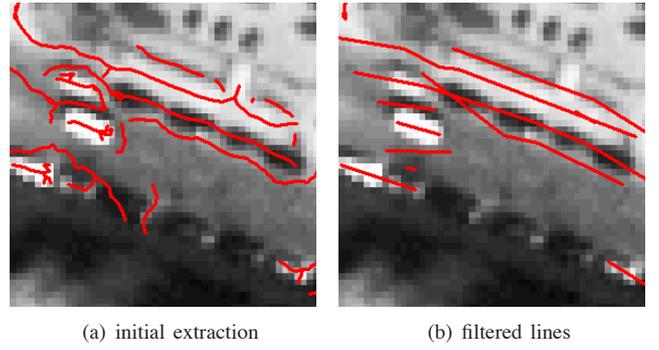

(a) initial extraction   (b) filtered lines

Fig. 3.   Line extraction and filtering

already accepted points. In order to achieve initial hypotheses, the parameters for c_L (minimum acceptable contrast) and c_H (contrast for queues definitely to be accepted) are chosen quite relaxed.

Additionally, the line extraction algorithm is supported by morphologically filtering of the image with a directional rectangular structuring element oriented along the particular road segment. In doing so, the queues are enhanced and substructures in bright cars are almost completely removed. The relaxed parameter settings lead to a huge number of false hypotheses but also return nearly all promising hypotheses for vehicle queues. Fig. 3(a) shows some results for the extraction of bright and dark lines. However, since the line extraction requires a minimum amount of contrast between vehicles and the road surface, gray vehicles cannot be extracted reliably because they hardly emerge from their surroundings.

Bright and dark lines are extracted separately. They are connected if they fulfill some distance and collinearity criteria. In our case, a maximum distance of one vehicle length must not be exceeded. From Fig. 3(a) it also becomes clear that the merging of parallel lines would lead to significant positional errors and is therefore prevented. The final processing steps consist of geometrical smoothing by polygonal approximation and resampling [16] as well as testing all resulting lines against a minimum length threshold and a maximum direction difference to the road. Results of the merging and filtering process are illustrated in Fig. 3(b).

After extracting lines as medial axes of a ribbon, the width and the contrast functions are determined. The algorithm to find the ribbon width in each line point is based on profiles spanned perpendicular to the local line direction (see Fig. 4) and determining each profile's gray values by bilinear interpolation. Fig. 5 shows a 3D visualization of extracted gray value profiles for a single vehicle queue. Then, for each profile, local maxima are determined in sub-pixel precision by fitting a second-order polynomial to the first derivative in each profile point. The first maximum value found on either side of medial axis is supposed to correspond to the vehicle boundary, i.e., the distance between the two maxima yields the queue width (see Fig. 4). If no maximum is found, gaps in the width function are closed afterwards by linear inter- or
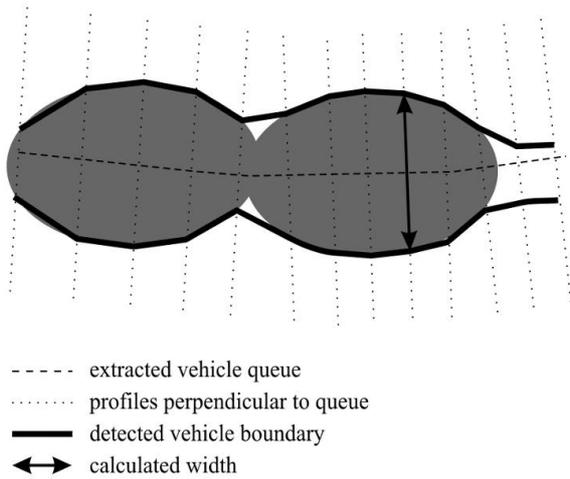
Fig. 4. Definition of width function

----- extracted vehicle queue
.... profiles perpendicular to queue
——— detected vehicle boundary
←——→ calculated width

extrapolation.

Results of width determination are illustrated in Fig. 6. It can be seen that most edges correspond to vehicle sides. Because of weak contrast between vehicles and road surface, several outliers are present, which are to be removed by median filtering the width function.

To determine the contrast function of a ribbon, a reference gray value outside the vehicle regions must be defined. The actual gray values in the direct neighborhood of a vehicle, however, are often influenced by adjacent objects or shadows and deliver barely reliable estimates of the reference gray value. A better way to determine the contrast function is to estimate the road surface brightness in the neighborhood of a vehicle queue and use this estimate as reference gray value. Assuming that - despite of the presence of some vehicles - the most frequent gray values in the RoI correspond to the road surface and that there occur less disturbances trough vehicles
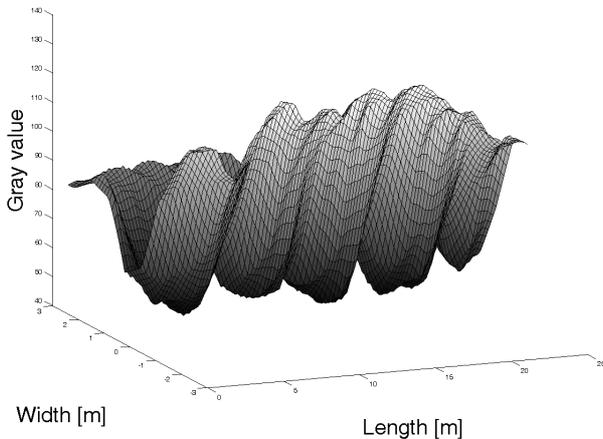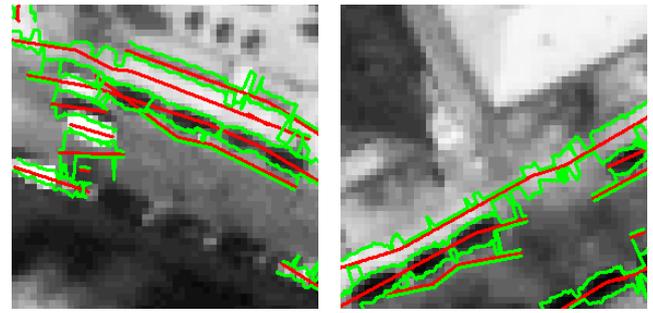

Fig. 5. Gray value profile of a ribbon

(a) scene 1　　　　　　　(b) scene 2

Fig. 6. Extracted queues: medial axis (red) and corresponding edges (green)

and shadows in the center of a road than at the road sides, the following simple procedure has been implemented to compute the road surface brightness:

- project the start and end point of each extracted medial axis onto the GIS road axis, thereby defining the relevant road section
- dilate the section by approximately the width of one lane
- calculate the median gray value of this image region yielding the road surface brightness

Since the gray values along the medial axes have already been extracted, the contrast function simply results from the absolute difference of these values and the reference gray value. In Fig. 7, examples of the width and the contrast function of a ribbon are shown. Furthermore, it illustrates that both functions show mutually correlated repetitive patterns which will be used to detect single vehicles.

C. Single vehicle determination

For extraction of single vehicles from a ribbon, Gaussian kernels are fitted to the width and the contrast function (see Fig. 7). Of course, different kernels like a second-order polynomial could be used instead. However, the estimated parameters of a fitted Gaussian kernel do not only relate to the desired vehicle dimensions but also allow establishing a link to the particular scale used for line extraction - especially the Gaussian kernels fitted to the contrast function. The rationale of the procedure outlined in the following should thus be understood as an attempt to embed the vehicle detection into the same scale space framework as the line extraction approach.

The calculation of the unknown parameters of each Gaussian kernel is done by least squares fit. The notation corresponds to [17].

The functional model of a Gaussian function to be fitted to a predefined interval of the width functions has the following form:

$$w\left(a_w, \sigma_w, \mu\right) = \frac{a_w}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma_w}\right)^2} \tag{2}$$

with:

$w(\ldots)$　width as function of a_w, σ_w and μ
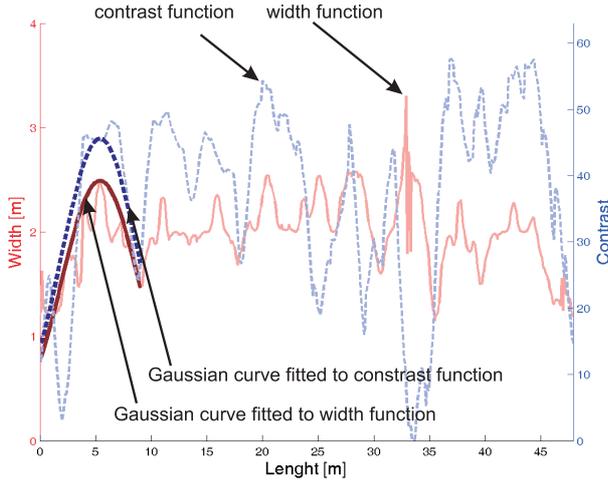
a_w　　　the amplitude of the fitted Gaussian kernel

Fig. 7. Width and contrast function of a ribbon

(a) scene 1   (b) scene 2

Fig. 8. Extraction results: correct (green), missed (yellow) and false (red) detections

σ_w	second-order moment of the Gaussian kernel
μ	first-order moment of Gaussian kernel
x	position of w along the interval under investigation

Similarly, the contrast c can be defined as a Gaussian function of the contrast amplitude a_c and the second-order moment σ_c. Since a vehicle should yield its maximum at the same position of both width and contrast function, is a shared parameter in both functions. Fig. 7 illustrates an example of the contrast and the width signal. The fitted Gaussian curves for the first interval are also included. These intervals are defined by two consecutive minima in a smoothed version of the function. It is also apparent from Fig. 7 that introducing σ as a shared parameter would not lead to satisfactory results. The two functions differ too much in shape so that the accuracy of the estimated unknown would decrease significantly. The unknown parameters of both Gaussian functions can be summarized in the vector $\mathbf{x}$:

$$\mathbf{x^T} = \left(\begin{array}{ccccc} a_w & \sigma_w & \mu & a_c & \sigma_c \end{array} \right)$$

It is clear that (2) and the Gaussian function for the contrast are nonlinear. Therefore, the determination of the unknown parameters is an iterative process.

The initial values for $\mathbf{x^0}$ are chosen as:

- μ^0 to the position of the maximum of the current interval
- a_w^0 can be calculated as $a_w^0 = w_\mu \cdot \sqrt{2\pi} \cdot \sigma_w$ where w_μ is the width at maximum μ (calculation of a_c^0 is done similarly)
- $\sigma_w^0 = \sigma_c^0$ is calculated according to the supposed vehicle length by (1)

If the width and the contrast functions exhibit the expected repetitive pattern, the Gaussian function can be fitted to the observations within a few iterations. As the final result of the least squares adjustment, we obtain the parameters describing a fitted Gaussian kernel for a given interval including the accuracies of the 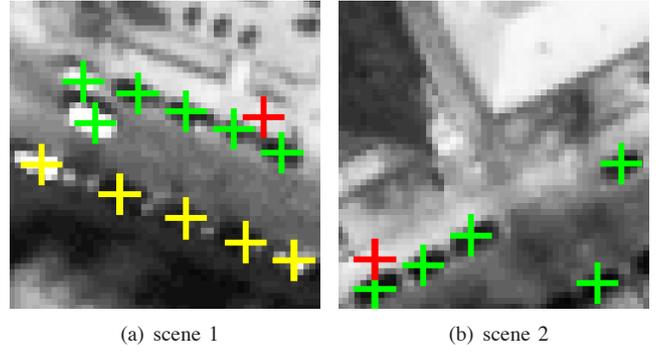estimation. Thresholds are applied to these parameters in order to discern false and correct hypotheses. The required thresholds were obtained by offline tuning using other scenes.

In some cases, multiple detections of the same vehicle occur due to neighboring ribbons. Therefore, an overlap analysis is carried out in which all overlapping (or close-by) hypotheses are mutually tested for consistency. In case of conflicts the worse hypothesis is rejected.

III. RESULTS AND DISCUSSION

In Fig. 8 some results of our extraction approach are shown. Green crosses correspond to correct extractions, red crosses represent misdetections. The results are encouraging, especially when recalling Fig. 7, which give an impression of the "noisiness" of the contrast and the width function.

However, there are also a number of misdetections. Such failures could be overcome, for instance, when analyzing neighborhood relations of extracted vehicles more in-depth. Incorporating such additional reasoning into the approach would allow to further reduce the misdetection rate.

The figures also show that a number of cars are not extracted, i.e. the completeness of this approach is quite low. However, one has to keep in mind that not all vehicles are contained in queues and, furthermore, that the line extraction does not extract all existing queues. In fact, tests have shown that approximately 60% of all vehicles are contained within the ribbons that serve as initial hypotheses. Also the edge detection procedure for determining a ribbon's width could be improved to prevent the least squares adjustment for aborting.

For numerical evaluation, manually created reference data have been utilized and the well-known criteria "correctness" and "completeness" are introduced as evaluation measures:

$$completeness = \frac{TP}{TP+FN}$$

$$correctness = \frac{TP}{TP+FP}$$

with: TP true positives
        FP false positves
        FN false negatives

The measures refer to single vehicles, i.e. true positives are correctly extracted vehicles, false positives are misdetections, and false negatives are missed vehicles with respect to the reference data. Table I summarizes the evaluation depending on the five types of reference data included:

   a) all verhicles

   b) only bright and dark vehicles, i.e. without gray vehicles

   c) only bright vehicles

   d) only dark vehicles

   e) only vehicles contained in queues (without gray vehicles)

Column e) is included, because our approach is designed to extract single vehicles from queues. Furthermore gray vehicles have been excluded from the reference in b) and e) since they almost show no contrast to their surroundings. In addition, we like to mention that the acquisition of reference data for some vehicles is certainly not free of errors. Even a human observer is sometimes not able to identify all vehicles in an image scene with high confidence. Therefore, our reference data can only be treated as a very good approximation of the real "ground-truth".

TABLE I

EVALUATION RESULTS

	Reference data				
	(a)	(b)	(c)	(d)	(e)
Completeness [%]	31.1	36.1	30.1	38.6	43.9
Correctness [%]	73.5	70.5	68.2	71.1	63.2

The numerical assessment of the results obtained when applying the approach to a large, complex urban scene confirms the discussion above. Despite the weak completeness, the good correctness of the eventually extracted vehicles allows to serve as starting point for searching additional vehicles. Therefore the next steps of implementation will include the search for isolated vehicles using the information from the previously queue detection. Preliminary investigations using a differential blob detector [18] for accomplishing this task have already been undertaken.

REFERENCES

[1] R. Bamler and S. Chiu, "Spaceborne traffic monitoring from sar and optical data," *Session at IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2005)*, 25 - 29 July 2005.

[2] "Joint Workshop of ISPRS and DAGM - Object extraction for 3D city models, road databases and traffic monitoring (CMRT'05)," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, U. Stilla, F. Rottensteiner, and S. Hinz, Eds., vol. 36, Part 3/W24, Vienna, Austria, 29-30 August 2005.

[3] S. Hinz, R. Bamler, and U. Stilla, Eds., *Theme Issue: Airborne and Spaceborne Traffic Monitoring*, ser. ISPRS Journal of Photogrammetry and Remote Sensing. Elsevier, 2006, vol. 61, no. 3-4.

[4] C. K. Toth and D. Grejner-Brzezinska, "Traffic flow estimation from airborne imaging sensors: A performance analysis," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. Part 1/W3, p. 7 p. (on CDRom), 2005.

[5] I. Ernst, M. Hetscher, K. Thiessenhusen, M. Ruhé, A. Börner, and S. Zuev, "New approaches for real time traffic data acquisition with airborne systems," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. Part 3/W24, pp. 69–83, 2005.

[6] A. Rajagopalan, P. Burlina, and R. Chellappa, "Higher-order statistical learning for vehicle detection in images," in *Proc. 7th IEEE International Confernce on Computer Vision (ICCV'96)*, vol. 2, Kerkyra, Corfu, Greece, 20-27 September 1996, pp. 1204–1209.

[7] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, no. 1, pp. 15–33, 2000.

[8] G. Sharma, "Vehicle detection and classification in 1-m resolution imagery," MSc Thesis, Ohio State University, Columbus, USA-OH, 2002, 125 p.

[9] X. Jin and C. H. Davis, "Vector-guided vehicle detection from high-resolution satellite imagery," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARRS'04)*, vol. 2, Anchorage, USA-AK, 20-24 September 2004, pp. 1095–1098.

[10] T. Tan, G. Sullivan, and K. Baker, "Model-based localisation," *International Journal of Computer Vision*, vol. 27, no. 1, pp. 5–25, 1998.

[11] S. Hinz, "Integrating local and global features for vehicle detection in high resolution aerial imagery," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, no. Part 3/W28, pp. 119–124, 2003.

[12] P. Burlina, R. Chellappa, C. Lin, and X. Zhang, "Context-based exploitation of aerial imagery," in *Proc. IEEE Workshop on Context-Based Vision (ICCV'95)*, Cambride, USA-MA, 19 June 1995, pp. 38–49.

[13] P. Burlina and R. Chellappa, "A spectral attentional mechanism tuned to object configurations," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1117–1128, August 1997.

[14] J. Leitloff, S. Hinz, and U. Stilla, "Detection of vehicle queues in QuickBird imagery of city areas," *Photogrammetrie - Fernerkung - Geoinformation (PFG)*, vol. 4, pp. 315–325, 2006.

[15] C. Steger, "An unbiased detector of curvilinear structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 2, pp. 113–125, February 1998.

[16] U. Ramer, "An iterative procedure for the polygonal approximation of plane curves," *Compter Graphics and Image Processing*, vol. 1, pp. 244–256, 1972.

[17] E. M. Mikhail, *Oberservation and Least Squares*. Copyright: Thomas Y. Crowell Company Inc., New York: IEP - A Dun-Donnelly Publisher, 1976.

[18] S. Hinz, "Fast and subpixel precise blob detection and attribution," in *Proc. IEEE International Conference on Image Processing (ICIP'05)*, Genua, 2005, pp. 11–14.