Automatic Car Detection in High Resolution Urban Scenes Based on an Adaptive 3D – Model

Christian Schlosser, Josef Reitberger, Stefan Hinz*

***Abstract* – This paper introduces a new approach to automatic car detection in monocular high resolution aerial images. The extraction is based on a 3D-model that describes the prominent geometric features of cars by a wireframe representation. Furthermore, vehicle color, windshield color, and intensity of a car's shadow area are included as radiometric features. During extraction, the model automatically adapts the expected saliency of these features depending on vehicle color measured from the image and the actual illumination direction given a priori. Car extraction is carried out by matching the model „top-down" to the image and evaluating the support found in the image. In contrast to most of the related work, our approach neither relies on external information like digital maps or site models, nor it is limited to one single vehicle model. Various examples illustrate the applicability of this approach. However, they also show the deficiencies which clearly define the next steps of our future work.**

***Index terms* – Car Detection, Urban Areas, Aerial Imagery, Image Analysis**

I. Introduction

This paper deals with automatic detection and counting of cars in high resolution aerial imagery. Research on this topic is motivated from different fields of application: Traffic-related data play an important role in urban and spatial planning, e.g., for road planning, for estimation or simulation of air and noise pollution, etc. In recent years, attempts have been made to derive traffic data also from aerial images, because such images are used in many fields of urban planning. Therefore, an algorithm that automatically detects and counts vehicles in aerial images would effectively support traffic-related analyses in urban planning. Furthermore, because of the growing amount of traffic, research on car detection is also motivated from the strong need to automate the management of traffic flow by intelligent traffic control and traffic guidance systems. 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 up-date.

The remainder of this paper is organized as follows. Section 2 discusses related work on automatic car detection and counting in aerial imagery. The vehicle model underlying our detection algorithm will be developed in Sect. 3. Then, Sect. 4 outlines details of the algorithm and Sect. 5 discusses results achievable with our approach.

II. Related Work

Related approaches to car detection can be distiguished based on the underlying type of modelling they use:

Several authors propose the use of an appearance-based, *implicit* model [1,2,3,4]. The model is created by example images of cars and typically consists of grayvalue or texture features and their statistics assembled in vectors. Detection is then performed by computing the feature vectors from image regions and testing them against the statistics of the model features. The other group of approaches incorporates an *explicit* model that describes a vehicle in 2D or 3D, e.g., by a box- or wire-frame representation [5,6,7,8,9,10,11]. In this case, detection relies on either matching the model „top-down" to the image or grouping extracted image features „bottom-up" to construct structures similar to the model. If there is sufficient support of the model in the image, a vehicle is assumed to be detected.

Of course, both approaches have their advantages and disadvantages. Regarding *complex urban areas*, however, we found the approach of explicit modelling superior for following reasons (similar comments are given in [12]):

Due to the high "density" and closeness of different objects in urban areas, objects impose strong influence on each other, e.g., trees may occlude cars partially, buildings cast shadows, materials like glasses or varnish may cause reflections or specularities on cars, etc. Since such influences mostly appear in form of local radiometric disturbances, a model emphasizing a structural description of a vehicle (as an explicit model) seems much more robust than one relying mainly on radiometry as the implicit model does. Another disadvantage of the implicit approach is, that the results are completely dependent on the training data, while it cannot be assured that the training data capture changes in illumination, viewpoint, and possible influences caused by neighboring objects correctly. In contrast, explicit modelling better allows to focus on the fundamental and robust features of cars and, furthermore, it better allows to employ a hierarchy of levels of detail. However, because of the small size of vehicles, it is clear that a very detailed model is necessary in order to avoid misdetections of objects that are fairly similar to vehicles.

* corresponding author:
Stefan Hinz, Chair for Photogrammetry and Remote Sensing,
Technische Universität München, Arcisstr. 21, 80333 München, Germany
URL: http://www.RemoteSensing-TUM.de

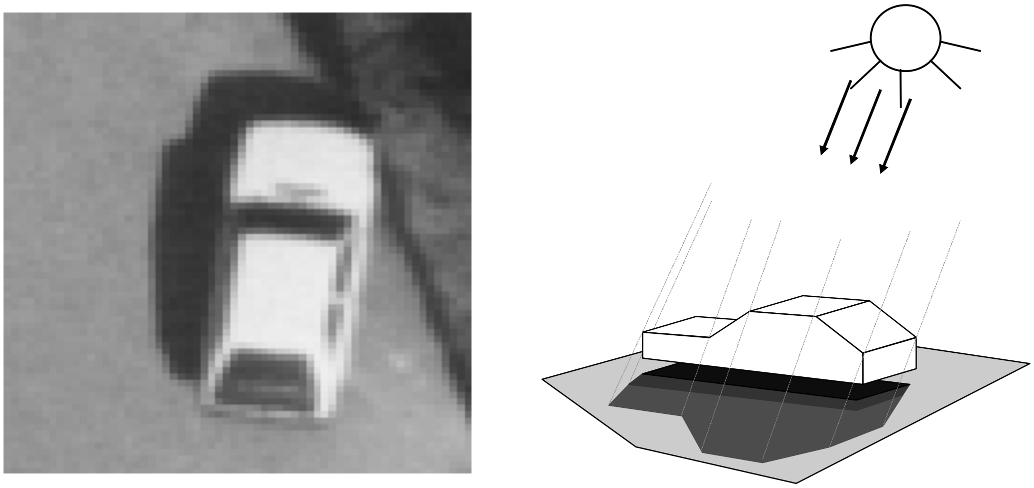


Figure 1: Image of car (left) and example of model (right)

III. Vehicle Model

In our approach, we use an explicit model that consists mainly of geometric features and also some radiometric properties. Geometrically, a car is modelled as 3D object by a wire-frame representation. Hence, an accurate computation of the car's shadow projection derived from date, daytime and image orientation parameters is possible and added to the model. The model further contains substructures like windshield, roof, and hood (see Fig. 1). As radiometric feature, color constancy between hood color and roof color is included. Please note, that color constancy is a relative measure and therefore independent of uniform illumination changes. The only absolute radiometric feature included is the darkness of the shadow region.

The main difference of our approach compared to related work, however, is that our vehicle model is *adaptive* regarding the expected saliency of certain features in the image. Consider, for example, the edge between a car's hood and windshield. In case of a bright car we expect a strong grayvalue edge since a windshield is usually very dark, while in case of a dark car the grayvalue edge may disappear completely. Hence, we model the expected saliency of a particular feature depending on vehicle color, vehicle orientation, view point (position in the image), and sun direction. View point and sun direction are given by the image orientation parameters and vehicle orientation and color are measured from the image (see also Fig. 2).

A disadvantage of the detailed description is, that a large number of models is needed to cover all types of vehicles. To overcome this problem a tree-like model hierarchy may be helpful having a simple 3D-box model at its root from which all models of higher level of detail can be derived subsequently.

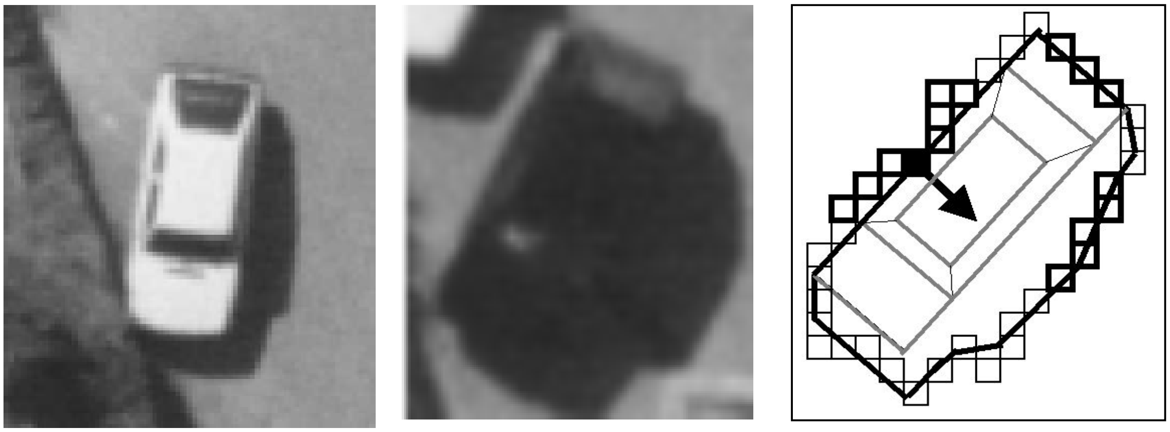
IV. Vehicle Detection

Detection is carried out by a top-down matching algorithm. A comparison with a grouping scheme that groups image features (edges and homogeneous regions) into car-like structures has shown that matching the complete geometric model top-down to the image is more robust. A reason for this is that, in general, bottom-up grouping needs reliable features as seed hypotheses which are hardly given in the case of such small objects like cars. Another disadvantage of grouping refers to the fact that we must constrain our detection algorithm to monocular images, since vehicles may move within the time of two exposures. Reconstructing a 3D-object from monocular images by grouping involves much more ambiguities than matching a model of the object to the image.

The steps of detection can be summarized as follows:

- Find edge pixels and compute edge direction
- Project the geometric model incl. shadow region to edge pixel and align its reference point and direction with the edge direction
- Measure reference color / intensity at roof region
- Adapt the expected saliency of each feature depending on position, orientation, color, and sun direction
- Measure features from the image: edge amplitude support of each model edge, edge direction support of each model edge, color constancy, darkness of shadow
- Compute a matching score (a likelihood) by comparing measured values with expected values
- Based on the likelihood, decide whether the car hypothesis is accepted or not

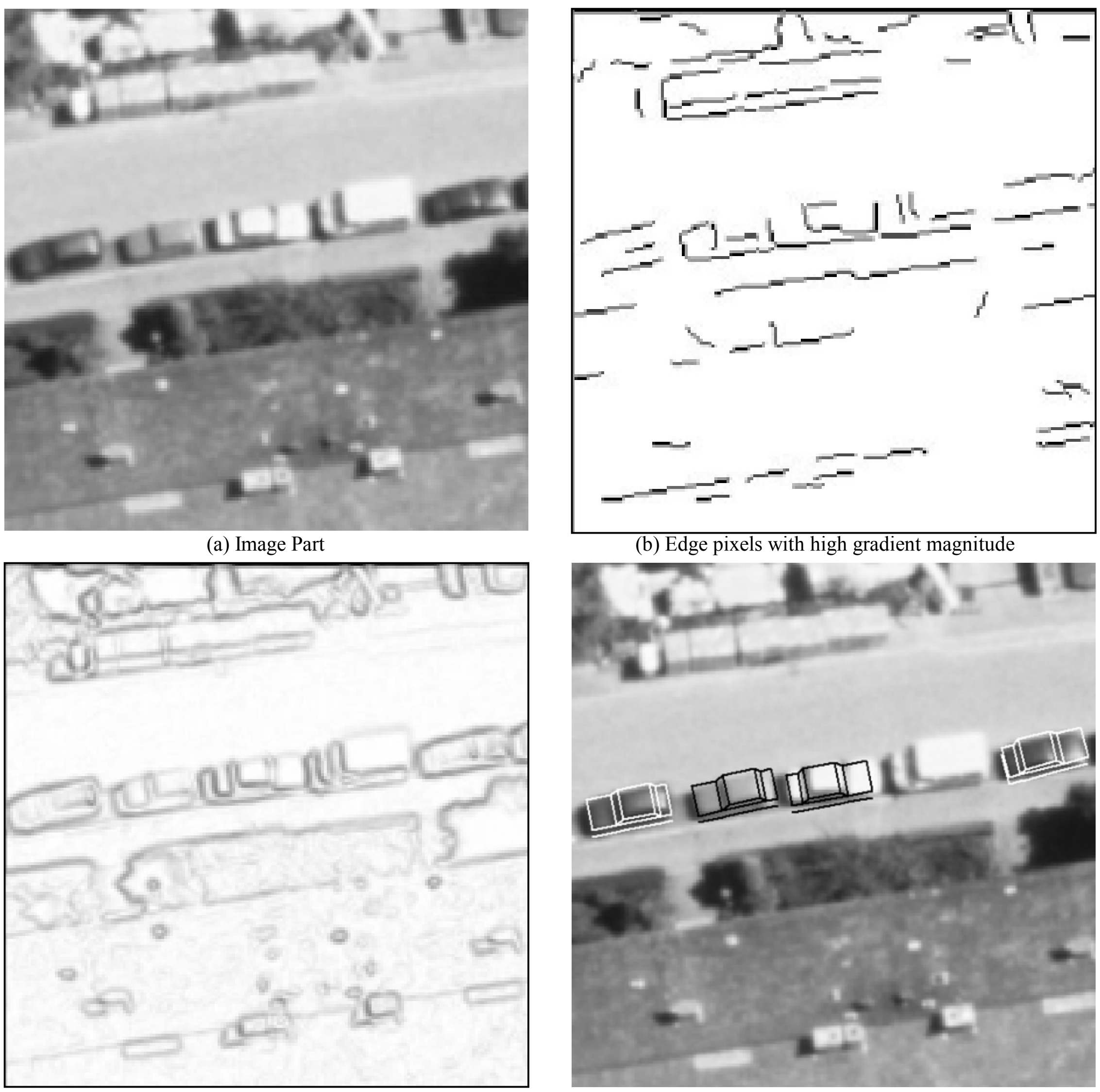
In the following, the evaluation measures involved are explained in more detail, while Figs. 2 and 3 illustrate the individual steps of matching.



(a) Bright car. (b) Dark car. (c) Saliency of model edges for dark car

Figure 2: Intensity-adaptive matching principle.

The car model is projected to an image position with high gradient magnitude (bold pixels in c) and aligned with the gradient direction (arrow). The matching score is calculated using *all* edge pixels (bold and thin pixels in c). Accoring to the intensity measured within the roof region, the expected saliency of each model edges is adapted. In case of a dark car (b), model edges belonging to the boundary of car- and shadow-region are expected to be salient (black lines in c) whereas all remaining model edges are assumed to be less salient (gray lines in c).



(a) Image Part (b) Edge pixels with high gradient magnitude

(c) Gradient magnitude image (d) Detected cars. White wireframe indicates car declared as "dark vehicle" and vice versa.

Figure 3: Steps of car detection.

The match of a model edge with the underlying image is calculated by comparing directional and positional features. Let $\Delta\alpha$ be the orientation difference between the gradient ∇I_i at a certain pixel i and the normal vector of the model edge and, furthermore, let d_i be the distance between this pixel and the model edge, then the match M_e [0 ; 1] of a model edge e with n pixels involved is computed by

$$M_e = \frac{1}{n}\sum_{i=1}^{n}\left|E_e - 0.5\cdot\left(A_i + D_i\right)\right|$$

with

$$A_i = \frac{\Delta\alpha_i}{\pi}\cdot\frac{\|\nabla I_i\|}{c_1} \quad , \quad D_i = \sqrt{\frac{d_i}{r}\cdot\frac{\|\nabla I_i\|}{c_1}}$$

and with E_e being the expected saliency of the model edge, r being the maximum buffer radius around the model edge, and c_1 being a constant to normalize the gradient magnitude $\|\nabla I_i\|$ into the range [0 ; 1]. Finally, the quality of the geometric match of the complete model is calculated as the length-weighted mean of all matches M_e.

Furthermore, darkness and homogeneity M_s of a shadow region s are evaluated by

$$M_s = \sqrt{\left(1-\frac{\mu_s}{c_2}\right)\cdot\left(1-\frac{\sigma_s}{c_3}\right)}$$

with μ_s and σ_s being mean and standard deviation of an image region and c_2 , c_3 being normalization constants.

To speed up runtime, a number of enhancements and pruning steps have been employed. The most important ones are:

- To avoid redundant computations for projecting models into image space, a database containing all possible (projected) 2D models is created beforehand which is accessed via indices during detection. Since image scale and sun direction are approximately constant for a given scene, the only free parameters are model orientation and x,y position in the image. A reasonable discretization for these variables is derived automatically from image scale and average vehicle size.
- The model is projected only to those positions where edge amplitude has passed a local non-maximum and noise suppression. Though, for calculating the matching score, all pixels are taken into account.
- The calculation of features is ordered in such a way, that implausible hypotheses appear yet after a few computations, thus allowing to abort matching immediately.

V. RESULTS AND DISCUSSION:

We tested our detection algorithm on a series of high resolution aerial images (ca. 15 cm ground resolution) of complex downtown areas. No preclassification of regions of interest has been carried out. The results show that nearly all passenger cars have been detected and that the false alarm rate is quite low (see Fig. 4). However, larger vehicles like vans or small trucks whose geometry deviates from the model too much have been missed. Problems occur also in regions where the complete road is darkened by building shadows. To overcome such kind of problems, a preclassification of large shadow regions would be necessary, so that the vehicle model can be adapted accordingly. The most encouraging outcome of the results, however, is the low rate of misdetections. This is a clear consequence of the detailed model employed.

VI. REFERENCES

[1] R. Ruskone, L. Guiges, S. Airault, and O. Jamet, 1996. „Vehicle detection on aerial images: A structural approach“, 13th International Conference on Pattern Recognition.

[2] A. Rajagopalan, P. Burlina, and R. Chellappa, 1999. „Higher-order statistical learning for vehicle detection in images“. International Conference on Computer Vision, 1999

[3] H. Schneiderman and T. Kanade, 2000. „A Statistical Method for 3D Object Detection Applied to Faces and Cars“. Computer Vision and Pattern Recognition, 2000.

[4] C. Papageorgiou and T. Poggio, 2000. „A trainable system for object detection“. International Journal of Computer Vision 38(1), pp. 15–33.

[5] P. Burlina, R. Chellappa, C. Lin, and X. Zhang, 1995. „Context-Based Exploitation of Aerial Imagery“. IEEE Workshop on Context-based Vision.

[6] P. Burlina, V. Parameswaran, and R. Chellappa, 1997. „Sensitivity Analysis and Learning Strategies for Context-Based Vehicle Detection“. In: DARPA Image Understanding Workshop, 1997.

[7] G. Liu, L. Gong, and R. Haralick, 1999. „Vehicle Detection in Aerial Imagery and Performance Evaluation“. ISL-Technical Report, Intelligent Systems Laboratory, University of Washington, Seattle, WA, 1999.

[8] G. Liu, 2000. „Automatic Target Recognition Using Location Uncertainty“. PhD thesis, University of Washington, Seattle, WA, 2000.

[9] T. Zhao and R. Nevatia, 2001. „Car detection in low resolution aerial image“. International Conference on Computer Vision, 2001.

[10] S. Hinz and A. Baumgartner, 2001. „Vehicle Detection in Aerial Images Using Generic Features, Grouping, and Context”. Pattern Recognition 2001 (DAGM Symposium 2001), Lecture Notes on Computer Science 2191, Springer Verlag, 2001.

[11] H. Moon, R. Chellappa, and A. Rosenfeld, 2002. „Performance Analysis of a Simple Vehicle Detection Algorithm“. Image and Vision Computing 20(1), pp. 1–13.

[12] P. Suetens, P. Fua, and A. Hanson, 1992. „Computational strategies for object recognit ion“. AC M Computing Survey s 24(1), pp. 5–60.

Figure 4: Results. White wireframe indicates car declared as "dark vehicle" and vice versa.