ISPRS Journal of Photogrammetry and Remote Sensing 136 (2018) 85–105

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Object-based detection of vehicles using combined optical and elevation data

Hendrik Schilling *, Dimitri Bulatov, Wolfgang Middelmann

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Gutleuthausstr.1, 76275 Ettlingen, Germany

ARTICLE INFO

Article history:
Received 4 May 2017
Received in revised form 24 November 2017
Accepted 30 November 2017

Keywords:
Vehicle detection
Object-based classification
Data fusion
Elevation data
Random forest
High-resolution
Feature extraction
Cluster analysis

ABSTRACT

The detection of vehicles is an important and challenging topic that is relevant for many applications. In this work, we present a workflow that utilizes optical and elevation data to detect vehicles in remotely sensed urban data. This workflow consists of three consecutive stages: candidate identification, classification, and single vehicle extraction. Unlike in most previous approaches, fusion of both data sources is strongly pursued at all stages. While the first stage utilizes the fact that most man-made objects are rectangular in shape, the second and third stages employ machine learning techniques combined with specific features. The stages are designed to handle multiple sensor input, which results in a significant improvement. A detailed evaluation shows the benefits of our workflow, which includes hand-tailored features; even in comparison with classification approaches based on Convolutional Neural Networks, which are state of the art in computer vision, we could obtain a comparable or superior performance (F_1 score of 0.96–0.94).

1. Introduction

Automatic vehicle detection in airborne sensor data has numerous applications. Some of them, such as traffic analysis (road capacity assessment and congestion warning), ecological sciences (estimation of noise and air pollution), and city planning (e.g., availability of parking spaces), were mentioned in Holt et al. (2009) and represent an efficient and cheap alternative to a direct monitoring system. Furthermore, vehicle detection was employed by Leberl et al. (2007) to improve orthophotos and 3D city models by removing the frequently changing objects (persons, cars) and closing the gaps with inpainting methods. Finally, vehicle detection is also important for quick response applications, such as disaster management, where the presence of vehicles can be decisive for saving lives or identifying threats. In addition to the challenges that frequently arise, given that vehicles are rather small objects with strong signature variations (car types, occlusions, shadows, etc.), in the previous type of application vehicles can appear outside traffic routes, which renders their detection particularly difficult. Due to these numerous applications and challenges, vehicle detection in airborne data represents a captivating research area that has been studied for the last few decades.

The importance of vehicle detection is corroborated also by the variety of sensor data types for which detection pipelines have been developed. The two most important types are optical data, that is, image and image sequences, and 3D data, such as airborne laser data. While on the good side, a single aerial or even satellite image typically has a very high spatial resolution, the main disadvantages for vehicle detection in optical data are occlusions and variations in the appearance of vehicles. Thus, vehicles can easily be confused with other objects, such as roof dormers, air cooling systems, ping-pong tables, and garbage bins. The utilization of sets of images recorded with a time delay allows the detection of moving vehicles and therefore is widely used in traffic monitoring. However, this is possible only for very high depth-to-basis ratios that is, either almost planar scenes or nearly identical sensor position. More importantly, for the application mentioned above, identification of stationary vehicles is often essential. By taking the object height into account, the number of false alarms is expected to be reduced. As a consequence and also because of the consistent accuracy of the captured 3D points, the use of LiDAR point clouds for vehicle detection is gaining popularity. However, the disadvantages are the cost factor and frequently inadequate resolution. In addition, co-registration of LiDAR and optical data is a non-trivial issue. Fortunately, it is possible to compute a large and accurate

* Corresponding author.
E-mail address: hendrik.schilling@iosb.fraunhofer.de (H. Schilling).
URL: http://www.iosb.fraunhofer.de (H. Schilling).

https://doi.org/10.1016/j.isprsjprs.2017.11.023
0924-2716/

3D point set from aerial images captured from different positions by using state-of-the-art stereo reconstruction methods and use this set for detecting stationary vehicles. Although partly occluded regions, for instance, those between vehicles, are not always effectively covered, the advantages of photogrammetrically reconstructed point clouds are evident.

The goal of this study was to establish a workflow for the detection of stationary vehicles in combined optical and elevation data, which are co-registered with an accuracy below a few pixels, and to investigate the extent to which this workflow is superior to the existing workflows, in particular, those processing either single type of data. For this reason, a detailed literature survey is provided. Furthermore, to facilitate the aforementioned applications, we strive for a detection pipeline that is able to extract each individual vehicle separately, even in scenes where many vehicles are closely located. All stationary vehicles should be detected, even those partly occluded or not close to roads or parking spaces. Only moving vehicles were excluded from this study, since they are not represented correctly in photogrammetrically reconstructed data. One important goal was to reduce the number of parameters. However, where their use in the workflow cannot be avoided, the parameters, are automatically learned from available training data or derived directly from the input data, e.g., spatial resolution. This allows the workflow to be adapted to different datasets or new tasks. Finally, the computational effort should be considered within the requirements of remote sensing, that is, the ability to process large datasets.

Our proposed workflow consists of three consecutive stages: (1) candidate identification, (2) classification, and (3) single vehicle extraction (SVE). These steps are illustrated in Fig. 1. The task of candidate identification comprises reducing the search space from all possible image positions to a set of likely candidates. We utilize the fact that, at the considered resolution, most cars in the images contain two parallel lines. These lines are detected and connected to so-called stripes, which form the input for the next stage. As candidate identification frequently leads to a high number of false alarms, the second stage utilizes a supervised classification method to group the candidates into vehicle or background categories. We apply a customized discriminative feature set combined with an adapted fusion scheme to attain improved results. In the third stage, the exact position and orientation of each single vehicle are extracted from all the candidates classified as a vehicle. For this, an energy function depending on the orientation and location of the candidate is derived from training data using support vector machines (SVMs).

The rest of this article is structured as follows. Section 2 provides an overview of previous algorithms for vehicle detection. In Section 3, our method for candidate identification is described. In Section 4, an overview of the applied features, as well as the classification scheme, is provided. The method for SVE is introduced in Section 5. In Section 6, the considered evaluation metrics, datasets, and experimental setup are described and Section 7 presents a detailed evaluation of all the processing steps. We terminate our article with conclusions and outlook in Section 8.

2. Previous work

We decided to categorize the existing methods for vehicle detection according to the sensor data applied for accomplishing this task. The cheapest and most easily available type of data is (single) aerial, or, analogously, satellite images of very high resolution. We devote Section 2.1 to techniques for which the input is optical data. Then, in Section 2.2 we will review the existing approaches for which 3D data are available instead of or additionally to the optical data. Note that other sensor types that enable vehicle detection, such as multispectral and radar-based techniques, are not reviewed, since our experiments were performed with combined optical and 3D data. We conclude this section with a summary of the lessons learned and a specification of the contributions of this paper in Section 2.3.

2.1. Optical data

In his dissertation, Türmer (2014) proposed sub-categorizing the existing vehicle detection algorithms into gradient- and region-based. The first group of methods relies on changes in image gradients in the regions around vehicles. These changes may be analyzed using generic context-independent features, e.g., Haar-like features or histogram of oriented gradients (HoG) (Dalal and Triggs, 2005), local binary patterns (Ojala et al., 1994), or SIFT descriptors (Lowe, 1999), calculated and stored for every pixel, edge, etc. The texture patterns that statistically predominate in and around cars are determined using training examples. The test data are subjected to a classifier. These so-called implicit approaches, e.g., those presented in Leberl et al. (2007), Kembhavi et al. (2011), Liu and Mattyus (2015), Moranduzzo and Melgani (2014), Grabner et al. (2008), Shao et al. (2012), are characterized by very high-dimensional feature spaces and training databases. For example, Leberl et al. (2007) considered concatenation of heterogeneous features and performed classification with an online version of the well-known AdaBoost classifier. Detection was performed using an exhaustive search over all image data. As the final step, the exact locations were determined using a mean shift method. In the study of Liu and Mattyus (2015), integral channel features were computed and processed by a modification of the AdaBoost classifier. The authors additionally estimated orientation and car type by means of HoG features and a neural network. The orientation estimation was formulated as a multiclass classification problem. The widely used SIFT features were utilized in the study of Moranduzzo and Melgani (2014). After the search space was reduced to areas where vehicles are typically found, e.g., asphalt regions, SIFT key points were detected. These SIFT features were extended by red, green, and blue (RGB) and hue, saturation, and value (HSV) channels, and, together with morphological features, were subjected to an SVM. Finally, key points classified as belonging to the car class were grouped into single vehicles with an iterative procedure. These methods have in common that the achieved F_1 score is significantly below 0.90.

Recently, object detection using convolutional neural networks (CNNs) has become attractive because of the increased computational capacities and training datasets. An example was given by Chen et al. (2014), who introduced a "hybrid deep convolutional neural network," which was optimized to extract multi-scale features. By combining this method with a modified sliding window technique, it was possible to detect vehicles with high accuracy. However, the evaluation was limited to a few Google Earth images and single sensor data. Furthermore, dependency on sliding window techniques usually results in a time consuming process. The direction of the work of Ammour et al. (2017) was similar. They used a mean shift algorithm to create object proposals, a CNN to extract features, and a linear SVM for classification. Beside being used as classifiers only, CNNs can also be employed to create region proposals. In the context of vehicle detection, Deng et al. (2017) and Tang et al. (2017) presented two interesting examples of this strategy. They developed specialized region proposal networks to create candidates in larger images. While these methods are promising, their achieved F_1 score is only around 0.83. Unfortunately, it is not clear in what aspect improvement can be achieved. Two further approaches that use CNNs for semantic segmentation were proposed in Audebert et al. (2017a,b). Both methods are