Available online at www.sciencedirect.com
ScienceDirect
Image and Vision Computing 25 (2007) 1422–1431
www.elsevier.com/locate/imavis

Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks

Xiaoying Jin, Curt H. Davis *

Department of Electrical and Computer Engineering, University of Missouri – Columbia, Columbia, MO 65211, USA

Received 8 December 2005; received in revised form 19 November 2006; accepted 14 December 2006

Abstract

High-resolution satellite imagery has recently become a new data source for extraction of small-scale objects such as vehicles. Very little vehicle detection research has been done using high-resolution satellite imagery where panchromatic band resolutions are presently in the range of 0.6–1.0 m. Given the limited spatial resolution, reliable vehicle detection can only be achieved by incorporating contextual information. Here, a GIS road vector map is used to constrain a vehicle detection system to road networks. We used a morphological shared-weight neural network (MSNN) to learn an implicit vehicle model and classify pixels into vehicles and non-vehicles. A vehicle image base library was built by collecting more than 300 cars manually from test images. Strategies to reduce the false alarms and select target centroids were designed. Experimental results indicate that the MSNN performed very well. The detection rate on both training and validation sites exceeded 85% with very few false alarms. By learning the implicit vehicle model through a MSNN, our method outperforms a baseline blob detection method.

Keywords: Vehicle detection; High-resolution satellite imagery; Neural networks; Feature extraction; IKONOS

1. Introduction

With the availability of a new generation of high-resolution commercial satellite images having spatial resolutions less than 1 m, small-scale objects such as roads [1], buildings [2], and vehicles can be readily seen. This provides an important new data source for urban remote sensing applications. In this paper, we develop a method for automatic vehicle detection and counting in high-resolution satellite imagery. Vehicle detection has applications in military and homeland surveillance, transportation planning and management, and intelligent traffic guidance systems. Most vehicle detection research has been done using aerial imagery with a spatial resolution of 0.35 m or less [3–10]. Vehicle detection research has seldom been done using high-resolution satellite imagery where spatial resolutions of the panchromatic band are presently in the range of 0.6–1.0 m [11,12]. Burlina et al. [6] postulated that vehicles are context-sensitive objects whose detection requires information about the surrounding environment. Accordingly, much of the previous research has relied on external information like digital maps or site models. The human eye is able to identify vehicles from high-resolution satellite imagery primarily because of context. Because of the limited spatial resolution of IKONOS data (1 m), we believe that reliable vehicle detection can only be achieved by incorporating contextual information. By reliable detection, we mean a high detection rate along with a small false alarm rate.

Existing approaches for vehicle detection can be categorized based on the underlying type of vehicle modeling. There are two basic kinds of vehicle models that have been used: (1) an appearance-based implicit model, and (2) an explicit model. The implicit model typically consists of intensity or texture features computed using a window that surrounds a given pixel. Detection is performed by

* Corresponding author. Tel.: +1 573 884 3789; fax: +1 573 882 0397.
E-mail addresses: jinxy65201@yahoo.com (X. Jin), DavisCH@missouri.edu (C.H. Davis).

0262-8856/$ - see front matter
doi:10.1016/j.imavis.2006.12.011

checking feature vectors of the surrounding image pixels. Ruskone et al. [5] utilized aerial imagery with 0.3–0.4 m resolutions. They exploited a two-step analysis strategy that was composed of vehicle detection followed by the validation through line clustering. A multilayer perceptron (MLP) analyzed the intensity values of a pixel's neighborhood for vehicle detection and classification. Then perceptual grouping theory was used to group the vehicles into lines for validation. Papageorgiou and Poggio [13] presented a trainable system for vehicle detection from aerial imagery views taken from a stationary camera. A Harr wavelet transform was used to describe the object classes in terms of local, oriented, and multiscale intensity differences around adjacent regions. The vehicle model was derived by training a support vector machine classifier using a large set of positive and negative examples. Michaelsen and Stilla [14] performed vehicle detection using infrared (IR) image sequences with a spatial resolution of 1 m. A spot detector was developed by checking the patterns of the neighborhood gray values around each pixel. They used large-scale vector maps of roads to exclude most non-vehicle locations. Evidence from overlapping areas of frames from the IR image sequence was accumulated in the scene. Sharma [11] presented three algorithms for vehicle detection from high-resolution satellite imagery. These were based on principal component analysis (PCA), a Bayesian Background Transformation (BBT), and gradient detection. The BBT method achieved the best results in all the test cases. In the BBT method, the probability of a pixel being a vehicle was estimated based on intensity changes from the background. From this summary of related work, we see that the implicit vehicle model was either derived through direct spatial or statistical analysis on the neighborhood of each pixel, or learned by training the features on a set of positive and negative samples.

A few authors have proposed the use of explicit models for vehicle detection. Here, a vehicle is usually described by a box or wire-frame representation. Detection is carried out by either matching the model "top-down" to the image or by grouping low-level features (e.g., edges) "bottom-up" to construct structures similar to the model. Burlina et al. [4] and Moon and Chellappa [9] represented a vehicle as a 3D box with width, length, and height. Site models were used to constrain vehicle detection to parking lots or roads. Edge points were first detected from the aerial imagery, and then a modified generalized Hough transform (GHT) was used to locate possible vehicles by extracting the centers of candidate rectangles. Vehicle hypothesis was generated by comparing the rectangles centered at the detected centers with the local edge map. Liu and Haralick [7] presented a technique that utilized ground-truth databases for vehicles in a vertical-view image dataset with spatial resolution of 0.3 m. The vehicles were modeled as rectangles and the centers were detected by edge detection and GHT. The performance was improved by a Bayesian classifier algorithm using eight input features. Zhao and Nevatia [8] presented a system to detect cars in aerial images with 0.1 m resolution. The generic model included a wire-frame geometrical model and a surface reflectance model. After psychological testing, they determined important features for human detection of vehicles. A Bayesian minimum risk classifier was developed to integrate multiple cues for a decision. Schlosser et al. [10] characterized vehicles in aerial imagery with 0.15-m resolution using an adaptive 3D wire-frame representation. They concluded that explicit models are more robust for vehicle detection in complex urban areas from high-resolution aerial images. During extraction, the model automatically adapted the expected saliency of radiometric and edge features depending on the measured vehicle color and actual scene illumination direction.

In this paper, we present a vehicle detection strategy for high-resolution IKONOS satellite imagery. At 1-m resolution, the image detail is too poor to use a vehicle detection approach based on an explicit model. Moreover, the space between vehicles in crowded parking areas is typically less than a pixel and this makes reliable detection in these areas extremely difficult for 1-m imagery. Because of the limited spatial resolution of IKONOS data, we conclude that reliable vehicle detection can only be achieved by incorporating contextual information. Hence, we use a GIS road vector map to constrain vehicle detection to road networks and this is consistent with a large body of previous work using context-supported approaches [4,6,9–12,14]. Road networks can also be obtained by automated road extraction systems [1].

Automated vehicle detection can be considered a specific application of automatic target recognition (ATR) research. The goal of ATR research is to achieve a high probability of target detection while simultaneously yielding a low probability of false alarm. Feature extraction and decision-making are both important components of ATR systems. Unfortunately, designing effective feature extraction strategies is a very difficult task. Since knowledge of vehicles in 1-m imagery is vague, it is not easy to build an explicit feature set for vehicles. Mathematical morphology is a useful technique for image processing, pattern recognition, shape analysis, and feature extraction [15]. Morphological shared-weight neural networks (MSNNs) have been successfully used for ATR research from multiple sensors including infrared, optical camera, SAR, and LADAR [16–19]. The morphological structure elements are learned as feature extractors simultaneously with the classification parameters. Here, we use a MSNN in our vehicle detection approach to learn an implicit vehicle model that incorporates both spatial and spectral characteristics.

This paper is outlined as follows. We give a brief summary of the MSNN in Section 2. The dataset used to construct the vehicle example base and used in our experiments is described in Section 3. Our vehicle detection approach is proposed in Section 4. In Section 5, detection results and evaluation are presented. Conclusions are summarized in Section 6.