ELSEVIER

Image and Vision Computing 20 (2002) 1–13

image AND vision COMPUTING

www.elsevier.com/locate/imavis

Performance analysis of a simple vehicle detection algorithm

H. Moon*, R. Chellappa, A. Rosenfeld

Center for Automation Research, University of Maryland, College Park, MD 20742-3275, USA

Received 3 December 2000; received in revised form 16 May 2001; accepted 6 June 2001

Abstract

We have performed an end-to-end analysis of a simple model-based vehicle detection algorithm for aerial parking lot images. We constructed a vehicle detection operator by combining four elongated edge operators designed to collect edge responses from the sides of a vehicle. We derived the detection and localization performance of this algorithm, and verified them by experiments. Performance degradation due to different camera angles and illuminations was also examined using simulated images. Another important aspect of performance characterization — whether and how much prior information about the scene improves performance — was also investigated. As a statistical diagnostic tool for the detection performance, a computational approach employing bootstrap was used. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Performance analysis; Vehicle detection; Aerial image; Bootstrap; Empirical evaluation

1. Introduction: performance characterization

The task of a computer vision algorithm can be specified in terms of two components: the range of images to be processed and the performance criterion that the algorithm should try to achieve. The algorithm can then be designed to handle every image in the given class, and at the same time, to optimize the specified criterion function.

The performance of the algorithm is evaluated in terms of these components, using two different methodologies: theoretical formulation of the relation between the imaging conditions and the criterion function, and empirical evaluation using real or simulated imagery.

If there is an underlying model for the image and for the possible perturbations of the image that we may expect in a real environment, and also if the algorithm is based on some mathematical formulation, it may be possible to predict the performance theoretically. Haralick [14] has asserted that performance characterization of a vision algorithm has to do with establishing the random variations in the output data that result from the random variations and imperfections in the input data. However, a scene model is not always available or is usually not complete; hence, the error propagation computed using a theoretical model is usually a crude approximation.

Numerous imaging conditions can affect the performance of an algorithm, including scene noise and camera and lighting angles. If we can quantify these scene factors and if we have a very large set of images annotated with these parameters, experiments on this set of images would give a very informative performance measure for the algorithm. In reality, gathering data for such a complete and systematic evaluation is not usually possible.

Most algorithms are based on some model or an underlying theoretical framework; however, analytical tools for predicting the behavior of algorithms are not usually available. On the other hand, it is usually easy to generate simulated images and investigate algorithm performance using these images. If a perturbation model for the scene is available, we can generate perturbed images and conduct a systematic evaluation. While this kind of experiment is commonplace in computer vision, there have been objections to this method: the scene model as well as the perturbation model are far from realistic. This is true; however, as stated in Ref. [11], simulation is inevitable to verify the correctness of implementations, and also to analyze the behavior of algorithms under varying conditions for which real images are not available. Some scene environments are not easy to model, and it is usually impossible to acquire annotated ground-truth data for all the relevant sets of parameters. In our analysis of the vehicle detection task, we followed the above guidelines.

The three most important characterizations of the task of vehicle detection are the detection, false alarm, and

* Corresponding author.

E-mail addresses: hankyu@cfar.umd.edu (H. Moon), rama@cfar.umd.edu (R. Chellappa), ar@cfar.umd.edu (A. Rosenfeld).

PII: S0262-8856(01)00059-2

localization performances. There are numerous relevant scene factors, including resolution, noise, camera and lighting conditions, parking lot layout, etc. It is important to identify which parameters affect the criterion function significantly and which do not. We have found that the camera and illumination angles are the most crucial factors. We tested extreme values of these factors to verify the degradation of performance. With low illumination angles, the vehicle detector produces many false alarms due to long shadows; also, very oblique camera angles contribute to a poor detection rate. Since these angles are the most crucial factors in algorithm performance, we can conjecture that some prior information about these parameters can reduce errors.

The noise level of the image how the vehicles are positioned in the parking lot, and the colors of the vehicles also affect performance. We have designed a vehicle detector based on a simple vehicle model and Canny's formulation of edge detection, and have investigated these issues in a controlled environment using a parking lot scene simulator, as well as in a real environment. The utility of site information was investigated using real images with ground-truthed vehicles and annotated parking lot orientations.

By using the local linearity of the vehicle operator response, statistical properties of its detection and localization performance were derived, and they were then verified by simulation. We can view this formulation as error propagation: via the geometry of the vehicle model, from the image domain to the detection probability and parameter space. This framework is one of the contributions of this work, which can be extended to more general problems.

As it turned out, the most serious problem in vehicle detection is the occurrence of many false alarms, due to spurious responses from adjacent vehicles, road structures, and buildings. We developed an empirical hypothesis test to remove false alarms and self-diagnose the detection and localization performance.

This paper is organized as follows: Section 2 introduces the task of vehicle detection (Section 2.1) along with a review of related work (Section 2.2). By extending Canny's formulation of step edge detection (Section 2.3), we provide a solution to the problem of vehicle detection (Section 2.4). The output of the algorithm for a small parking lot image is discussed in Section 2.5. A mathematical formulation is given in Sections 3.1 and 3.2. Performance evaluation using simulated images is discussed in Section 4. Experiments on real ground-truthed images are given in Section 5. Finally, a method of self-diagnosing the performance using bootstrap is introduced in Section 6.

2. The vehicle detection algorithm

2.1. The task of vehicle detection

Aerial parking lot images usually have quality limitations due to the low resolution of the camera optics and atmospheric turbulence. The vehicles in the images we have used to test our algorithm occupy about 7×17 pixels on the average, and have approximately rectangular shapes. The front and/or rear windshields are usually recognizable, depending on the color of the vehicle and lighting conditions.

Aerial vehicle images are generally simpler to model than other objects. While vehicles in a typical parking lot have different overall dimensions, shapes, relative positions of windshields, and colors, we can use a simple rectangular model for the vehicle boundaries. We also used some variations on this model to deal with camera and lighting conditions, as discussed in later sections.

2.2. Related work

Most of the work on vehicle detection or recognition [2,10,17] has been on ground images, mainly as preprocessing before tracking for surveillance or traffic applications. There have not been many papers on non-military vehicle detection in aerial imagery. The proposed method makes use of a similar vehicle model and essentially the same image features as in Ref. [8], where local edge detection, a generalized Hough transform (GHT), and 'rubber-band matching' are performed. The GHT is employed to narrow the search space for a vehicle centroid using edge information. A rectangular band of fixed width is examined in every position where the GHT response is high. The number of edge pixels inside the band is taken to be the likelihood of the presence of a vehicle at that position. Ref. [3] investigates parameter adjustment of the method used in Ref. [8] using a Bayesian and Neyman–Pearson framework. Ref. [18] uses a machine learning approach employing a hierarchical vehicle and road structure model and a multi-layer perceptron for classifying vehicles and non-vehicles. While there has been much work on the detection of military vehicles using SAR [1,9] or FLIR imagery taken either from the ground or from the air, these sensors do not usually provide the same level of geometric image signatures as visible light images do.

The proposed work is performed in the context of site monitoring and surveillance, such as vehicle activity monitoring [8], and vehicle detection on roads [7]. This work exploits a site model (the positions and structures of buildings and roads) and a context model (regions where certain activity is more probable) to detect changes and activities. Ref. [4] utilizes the spectral analysis of edge structure to learn regions of interest. Some of the performance issues investigated in this paper, such as the performace degradation due to illumination and acquisition angle changes, or the role of site and context models, were considered in this previous work.

2.3. Canny's formulation of edge detection

Our approach uses Canny's formulation [5,6] of optimal