J. Vis. Commun. Image R. 24 (2013) 1182–1190

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Image-based on-road vehicle detection using cost-effective Histograms of Oriented Gradients

Jon Arróspide [a,b,*], Luis Salgado [a,c], Massimo Camplani [a]

[a] Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, 28040 Madrid, Spain
[b] Altran Spain, Methods & Tools, 28022 Madrid, Spain
[c] Video Processing and Understanding Lab, Universidad Autónoma de Madrid, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 14 November 2012
Accepted 1 August 2013
Available online 9 August 2013

Keywords:
Image analysis
Intelligent transportation systems
Intelligent vehicles
Vehicle detection
Histograms of Oriented Gradients

ABSTRACT

Image-based vehicle detection has received increasing attention in recent years in the framework of advanced driver assistance systems. However, the variability of vehicles in size, color, shape, etc. poses an enormous challenge, especially for the vehicle verification task. Histograms of Oriented Gradients (HOGs) have successfully been applied to image-based verification of objects. However, these descriptors are computationally demanding and are not affordable for real-time on-road vehicle detection. In this paper, less-demanding HOG descriptors are proposed and evaluated that significantly lighten the computation by exploiting the a priori known vehicle appearance. The proposed descriptors are evaluated on a large, public database and the experiments disclose that the computation times are reduced in a factor of more than 5, thus rendering HOG-based real-time vehicle detection affordable, while achieving detection rates of over 96%.

1. Introduction

Intelligent Transportation Systems (ITS) are the focus of increasing interest in recent years. The rationale behind these systems is to unite the drivers, the vehicle and road infrastructure technologies in order to increase road safety. In particular, in-vehicle technologies allow to provide the driver with a better understanding of the situation in its neighborhood. In addition, emerging interest is devoted to inter-vehicle communication technologies that allow to transmit the information collected in the environment of the vehicle to other vehicles so that a network of interconnected nodes is created (e.g. see [1–3]). Among in-vehicle technologies, those aiming at preceding vehicle detection and tracking play a leading role due to the high proportion of accidents produced by other vehicles and their severity.

Solutions based on both active and passive sensors are reported in the literature to address on-board vehicle detection. Active sensors have some advantages over cameras such as the higher detection range and, particularly in the case of LIDAR, a higher field of view and angular resolution, enabling a very good 3D accuracy. However, cameras are more advantageous in terms of cost and flexibility, and also allow for a more complete understanding of the environment (apart from vehicle detection other functionalities as lane detection and vehicle recognition can be achieved), which has fostered the research in this direction in recent years (see for instance [4,5]). Video-based vehicle detection is usually broken down into two stages. In the former, the complete image is analyzed in search of regions potentially containing vehicles. To this end, some light features relating to vehicle appearance, such as edges [6], shadow [7] and symmetry [8], are typically used, so that hypotheses for vehicle presence can be swiftly made. Other authors exploit stereovision for hypothesis generation [9]. In the second stage, the hypothesized regions are verified so as to make a final decision on the vehicle presence. The focus of this article is the second stage, i.e., methods for vehicle verification.

State-of-the-art vehicle verification approaches resort to supervised classification techniques. Specifically, two classes are considered, vehicle and non-vehicle, whose characteristics are learnt through a training stage. In this context, the selection of good descriptors plays a very important role for the success of the classifier. Popular methods to generate descriptors include Principal Component Analysis (PCA) [10] and Gabor filters [11], among others. On the other hand, Histograms of Oriented Gradients (HOG) have proven to be very effective for verification of many objects. In particular, HOG were proposed by Dalal and Triggs [12] for people detection, where they have been widely and successfully applied since (e.g. [13,14]), and their use has also been extended to other applications such as eye localization [15].

Recent works have also been reported in the literature attempting to employ HOG in the field of vehicle detection, including

* Corresponding author at: Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, 28040 Madrid, Spain. Tel.: +34 913367353.

E-mail addresses: jal@gti.ssr.upm.es (J. Arróspide), L.Salgado@gti.ssr.upm.es (L. Salgado), mac@gti.ssr.upm.es (M. Camplani).

1047-3203/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jvcir.2013.08.001

on-road [16–18] and aerial images [19]. For instance, the authors in [16] use HOG in combination with Haar-like features for vehicle verification. In [17] HOG are utilized for coarse level discrimination between cars and other vehicles. Nevertheless, they involve heavy computation times and thus hinder real-time operation when it comes to vehicle verification. In the aforementioned works, in order to relieve the computation requirements, the orientation range is divided into only four bins and the sign of the gradient is discarded, thus jeopardizing the classification performance. An efficient implementation is proposed in [20] by adopting the integral image method [21]. However, they use the MIT CBCL database [22], which contains only 512 vehicle samples, only half of it corresponding to rear views. This is quite limited and thus does not ensure a good generalization, in fact only visual results are provided to support the experiments. The work in [18] has similar shortcomings: an AdaBoost classifier using HOG features is proposed, but only vehicles in the lane of the own vehicle are considered. In summary, the aforementioned works represent a very promising preliminary approach to the use of HOG for vehicle imaging, but there are still a number of issues that must be addressed. To begin with, existing approaches mirror the configurations used for other objects to vehicles, while no analysis of the particular properties of vehicle imaging is made. In addition, extensive quantitative experiments are lacking on the effectiveness of HOG descriptors for vehicle verification.

In this study an in-depth analysis of the applicability and effectiveness of HOG for vehicle imaging is performed. First, the performance of the standard HOG technique is evaluated and the main drawbacks of this approach are inferred. In contrast to existing approaches, the evaluation is realized in a large and heterogeneous database of vehicle and non-vehicle images in different areas of the road. Then, in order to relieve the unaffordable computational requirements of HOG, less demanding HOG descriptors are proposed by making use of the specific vehicle imaging properties, namely of the gradient-wise vehicle appearance. In particular, an in-depth analysis of the response of these descriptors according to different vehicle poses is performed. In contrast to existing approaches, the proposed cost-effective descriptors render fine orientation binning and consideration of gradient sign affordable. Experiments reveal that these descriptors achieve a significantly better trade-off between cost and performance in the vehicle verification task than standard HOG.

The paper is organized as follows. First, in Section 2 the fundamentals of HOG technique are reviewed in order to give readers new to this topic a general overview. Next, in Section 3 the performance of the classical HOG descriptor when applied to vehicle classification is evaluated. In Section 4 the proposed cost-effective HOG descriptors are presented and their performance is extensively evaluated under different vehicle poses and parameter combinations. Comparison between traditional HOG and the proposed HOG-based cost-effective descriptors is enclosed in Section 5. Finally, the main conclusions of the paper are gathered in Section 6.

2. HOG-based vehicle descriptor

Histograms of Oriented Gradients (HOG, [12]) evaluate local histograms of image gradient orientations in a dense grid. The underlying idea is that the local appearance and shape of the objects can often be well characterized by the distribution of the local edge directions, even if the corresponding edge positions are not accurately known. This idea is implemented by dividing the image into small regions called cells (see Fig. 1 (a)). Then, for each cell, a histogram of the gradient orientations over the pixels is extracted. The combination of the histograms of the different cells provides image representation.

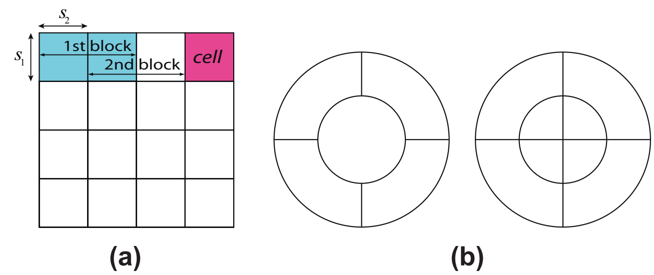

Fig. 1. Example of HOG grid. In (a) the standard rectangular configuration is shown, where each cell is of size $s_1 \times s_2$, blocks contains 1×2 cells, and there is a 2-fold block overlapping. In (b) two variants of a circular HOG configuration are shown.

The first step is thus to define the spatial and orientation binning, i.e., the size of the cells, and the number of evenly spaced orientation bins in the range [0°,180°) (the sign of the gradient is usually ignored). Gradient may be computed pixel-wise by means of any of the standard methods, e.g., Sobel or Prewitt operators. Then, for a given cell, each pixel in the cell votes for the bin containing its orientation. Votes are weighted according to the magnitude of the gradient. In addition, in order to tackle illumination and shadowing effects, the original proposal by Dalal and Triggs also comprises a final local contrast normalization step. This is performed by grouping cells in larger spatial structures, called blocks, and normalizing the response locally in each block in accordance with any standard measure, such as the L1 or the L2 norm. The final descriptor comprises the normalized responses from all the blocks. In fact, it is convenient to group the cells into overlapping blocks. For instance, if a block covers two cells in the horizontal direction, and the spacing between blocks equals the horizontal length of a cell, then there is 2-fold coverage of each image cell (see Fig. 1 (a)).

The original HOG technique presents two different kinds of configurations, called rectangular HOG (R-HOG) and circular HOG (C-HOG), depending on the geometry of the cells used. Specifically, the former involves a grid of rectangular spatial cells, as shown in Fig. 1 (a), and the latter uses cells partitioned in a log-polar fashion (see Fig. 1 (b)). The rear of vehicles typically contains rectangular structures, such as the license plate, the taillights, or the back window, and the contour of the vehicle itself is projected into the image with an almost rectangular shape. Therefore, rectangular grids adapt better the intrinsic structure of vehicles, and hence this geometry is used in this study.

3. Performance of HOG-based classifier

The performance of HOG as feature extractors is evaluated in the vehicle classification problem using the following experimental setup.

3.1. Experimental setup

A supervised classification approach is adopted, in which the feature vector extracted using the HOG descriptor is fed to a 2-class Support Vector Machine (SVM) classifier (see [23] for an introduction to SVM), which classifies samples as vehicles or non-vehicles. SVM classifiers have proven to yield excellent classification performance in the related literature and are thus also adopted in this study. In particular, a linear SVM is utilized as a baseline for comparison of the different descriptors.

Tests are carried out on the GTI vehicle database [24]. This is the most complete data set found on the literature for this purpose, as it comprises 4000 vehicle and 4000 non-vehicle images of size $R \times C = 64 \times 64$ with a wide variety of vehicle appearance in color,