VEHICLE DETECTION FROM HIGH-RESOLUTION AERIAL IMAGES BASED ON SUPERPIXEL AND COLOR NAME FEATURES

Ziyi Chen[a], Liujuan Cao[a], Zang Yu[a], Yiping Chen[a], Cheng Wang[a],*, Jonathan Li[a]
[a]Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Information Science and Engineering, Xiamen University, 422 Siming Road South, Xiamen, Fujian 361005, China

ABSTRACT

Automatic vehicle detection from aerial images is emerging due to the strong demand of large-area traffic monitoring. In this paper, we present a novel framework for automatic vehicle detection from the aerial images. Through superpixel segmentation, we first segment the aerial images into homo-geneous patches, which consist of the basic units during the detection to improve efficiency. By introducing the sparse representation into our method, powerful classification ability is achieved after the dictionary training. To effectively describe a patch, the Histogram of Oriented Gradient (HOG) is used. We further propose to integrate color information to enrich the feature representation by using the color name feature. The final feature consists of both HOG and color name based histogram, by which we get a strong descriptor of a patch. Experimental results demonstrate the effectiveness and robust performance of the proposed algorithm for vehicle detection from aerial images.

Keywords: vehicle detection; aerial image; superpixel; sparse representation; color name

1. INTRODUCTION

Coming with the rapid development of vehicle based traffic systems, it have become emerging to monitor such massive scale traffic information via various sensors [1, 2]. The fundamental task here is to monitor the traffic flow, vehicle density, and parking situation are partially acquired. A large number of fixed ground sensors, such as induction loops, bridge sensors, stationary cameras, radar sensors, etc. are required to efficiently monitor vehicles and gather traffic information [3, 4]. However, these methods fail to provide a complete overview of the traffic situation, which is a vital information source for studying road networks planning, modeling, optimization, and traffic-related statistics.

The demand for gathering an overview of traffic situations leads to the monitoring of vehicles via alternate methods. Among them, remote sensing is widely considered as one of the most promising and convenient platform to that end., such as remote sensing images captured by satellites or airplanes. Compared to satellite images, aerial images are usually preferred because of the higher spatial resolution of ranging 0.1 m to 0.5 m [5, 6] and easier data acquisition [7].

In the literature, there are many works focusing on the task of automatic vehicle detection from high-resolution aerial images [3, 5, 6, 8-14]. Most of the approaches can be separated into two types of models: appearance-based implicit and explicit models.
The appearance-based implicit model typically adopts image intensity or texture features computed using a small window or kernel surrounding a given pixel or a small cluster of pixels. The detection is conducted by examining feature vectors of the image's immediate surrounding pixels [13]. Cheng et al., using dynamic Bayesian networks for vehicle detection from aerial surveillance, which has achieved promising results on a challenging data set [12]. However, the color model, specially designed for separating cars from the background, still can't avoid false and missing detection due to the overlap of cars and backgrounds. Another problem is that the proposed approach is powerless to separate cars that are parked in close proximity. Also, the detection has to check over all pixels, which not only increases the computational complexity, but also increases the false detection rate. As a subsequent work, Moranduzzo et al. combined Scale Invariant Feature Transform (SIFT) and Support Vector Machine (SVM) for detecting cars from unmanned aerial vehicles (UAV) images [15].

* cwang@xmu.edu.cn;

2nd ISPRS International Conference on Computer Vision in Remote Sensing (CVRS 2015), edited by Cheng Wang, Rongrong Ji, Chenglu Wen, Proc. of SPIE Vol. 9901, 99010M · © 2016 SPIE · CCC code: 0277-786X/16/$18 · doi: 10.1117/12.2234762

Regarding the explicit model, a vehicle is usually described by a box, wire-frame representation, or morphological model. Subsequently, car detection is carried out by matching a car model to the image with either "top-down" or "bottom-up" strategies [13]. Zhong et al. utilized grayscale opening transformation and grayscale top-hat transformation to identify potential vehicles in the light or white background, and then used grayscale closing transformation and grayscale bot-hat transformation to identify potential vehicles in the black or dark background. Then, size information is employed to eliminate false alarms [13]. Their approach exhibits a good performance on highway aerial images. However, the gray value estimates of the background and a geographic information system (GIS) data are required. As a result, this method is not suitable for general scenes. In most cases, car detection methods using implicit model usually can achieve a better performance than the methods using explicit model due to the better generalization ability of implicit model.

However, existing methods employing the implicit model still suffer from two problems. First, an efficient scanning strategy is desired, which can replace the time-consuming pixel-based and normal slide window scanning. Second, due to the variations of objects' colors and the interference of noise, illumination variation, etc., the color information which is meaningful for object recognition [16] is usually neglected.
In this paper, we introduce a superpixel segmentation scheme to segment the images into patches to improve the scanning efficiency. Thus, we can scan the test images through sliding by superpixels to replace the pixel-based or slide window strategy. To utilize the color information and keep robust to noise and illumination variation, we proposed to use the color name feature, which can effectively represent the color information. We further combine color name feature with the HOG feature to describe each superpixel. To further improve the detection performance, the sparse representation which has been successfully applied in many fields [17-20] is applied for the model learning and classification procedure.

The rest of the paper is organized as follows. In section 2, superpixel segmentation used in the experiments is described. In section 3, the feature extraction method is presented. In section 4, the sparse representation is described. In section 5, experimental results are presented and discussed. Conclusions and future works are given in section 6.

2. SUPERPIXEL SEGMENTATION

The superpixel segmentation used in this paper is the Simple Linear Iterative Clustering (SLIC) scheme proposed in [21]. In the following, we give a brief introduction about SLIC.

Given the desired number of approximately equally sized superpixels k, the color images are first represented in CIELAB color space. The clustering procedure begins with an initialization step where k initial cluster centers $C_i = [l_i, a_i, b_i, x_i, y_i]^T$ are sampled on a regular grid spaced S pixels apart. To produce roughly equally sized superpixels, the grid interval is $S = \sqrt{N/k}$ where N is the total pixel number. Next, in the assignment step, each pixel i is associated with the nearest cluster center whose search region overlaps its location. Through search region restrict, fast processing speed is fulfilled. During the pixel assignment, the spatial information and color information are combined to measure the distance between a pixel and its nearby centers. Since each pixel can be represented as a 5 D vector similar as centers, the distance measurement between pixel i and cluster center j is defined as follows:

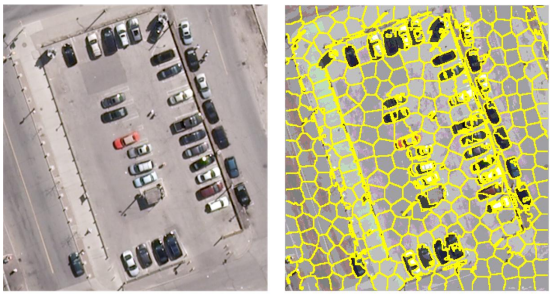
$$\begin{aligned} D &= \sqrt{(\frac{d_c}{m})^2 + (\frac{d_s}{S})^2}, \\ d_c &= \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2}, \\ d_s &= \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}. \end{aligned} \tag{1}$$

where m is the space distance factor. The Eq. (1) also can be simplified as:

$$D = \sqrt{d_c^{\,2} + d_s^{\,2}(\frac{m^2}{S})} \tag{2}$$

Fig. 1 shows a segmentation result of SLIC on an aerial image. It can be seen that the aerial image was effectively segmented into meaningful superpixels with high boundary recall.

3. FEATURE EXTRACTION



(a) (b)

Fig. 1. The segmentation result of SLIC on an aerial image. (a) is the original image, (b) is the superpixel segmentation result.

3.1 Color Name

In our method, a feature extraction which combines HOG and color name information [16] is used for effectively describing a patch. Here, we give a brief introduction about the color name learning. The color name descriptor is defined as a vector containing the probability of a color name given an image region R:

$$CN = \{p(cn_1 \mid R), p(cn_2 \mid R), ..., p(cn_{11} \mid R)\} \tag{3}$$

with

$$p(cn_i \mid R) = \frac{1}{P}\sum_{x \in R} p(cn_i \mid f(x)) \tag{4}$$

where cn_i is the *i-th* color name, x are the spatial coordinates of the P pixels in region R, $f = \{L^*, a^*, b^*\}$, and $p(cn_i \mid R)$ is the probability of a color name given a pixel value. The probabilities $p(cn_i \mid R)$ are computed from a set of images collected from Google. To learn color names, 100 images per color name are used. Fig.2 shows a color naming result of an aerial image.

We get the color mapping between RGB and color names, which has been learned through Google images in [22] and contains 11 color names (i.e. black, blue, brown, grey, green, orange, pink, purple, red, white, and yellow).

3.2 Feature Extraction

For HOG feature extraction, an image is divided into $N \times N$ non-overlapping pixel regions, which are known as cells. For each cell, the histogram of the pixels is calculated. Through combining the histograms of individual cells, the local appearance of a patch is effectively represented. In our method, we use 31-dimensional histogram vector to describe each cell.

To embed the color name information into the HOG feature, we calculate the color name histogram of each divided cell. Then, we extend the 31-dimensional HOG vector with the 11-dimensional color names histogram vector to construct a new feature vector. For cell C_i, the representation is obtained as follows:

$$C_i = [HOG_i, CN_i] \tag{5}$$

where HOG_i and CN_i represent the *i-th* cell's HOG vector and color name histogram vector. Thus, we get 42-dimension feature vector for each cell. Through computing all cells' concatenated feature vector, a patch is represented.

4. SPARSE REPRESENTATION

Let $Y = [y_1 \cdots y_N] \in R^{n \times N}$ denote N n-dimensional input signals. Then, learning a reconstructive dictionary with K items for sparse representation of Y can be accomplished by solving the following problem:

$$< D, X >= \arg\min_{D,X} \| Y - DX \|_2^2 \; s.t. \forall i, \| x_i \|_0 \le T \tag{6}$$

where T is a sparsity threshold. Eq. (6) can be replaced by a $l_1 - norm$ problem:

$$< D, X >= \arg\min_{D,X} \| Y - DX \|_2^2 + \gamma \| X \|_1 \tag{7}$$

where γ is a parameter to balance the reconstruction error and the sparsity of representation codes. The equality of Eq. (6) and Eq. (7) was proved in [23]. The K-SVD algorithm [24] is an iterative approach to minimize the energy in Eq. (7) and learns a reconstructive dictionary for the sparse representation of signals. Reversely, given a dictionary D, the sparse representation x_i of an input signal y_i is computed as:

$$x_i = \arg\min_x \| y_i - Dx \|_2^2 + \gamma \| x \|_1 \tag{8}$$

The orthogonal matching pursuit (OMP) algorithm [25] is used to solve Eq. (8). Due to the discrimination of the sparse codes among different classes, the distribution scores of sparse codes can be directly used for classification. The distribution scores of sparse code x_i can be computed as:

$$\mathrm{S}(c_i) = \frac{\sum_{x_i \in c_i} \| x_i \|}{\sum \| x_i \|} \tag{9}$$

where c_i represents the class i.

5. EXPERIMENTAL RESULTS

We tested the performance of our algorithm on an aerial image, covering the city of Toronto, with a size of pixels and a color depth of 24 bits per pixel (RGB). In our experiment, we cut the image into subareas and selected several subareas for training and testing. In the experiment, 13 sub-images for training and 8 images for testing were selected. The total number of cars in the testing set is 1589, and each car contains about 38×16 pixels.

For training, we manually selected 180 cars and randomly selected 1080 background patches as the training set. In our method, the slide window size is 61×31. We first extracted the features introduced by our paper from the training set to train a sparse representation dictionary. After that, the dictionary was used for the vehicle detection in the test images.
For comparison, we also tested other three popular methods (including HOG + Linear SVM, HOG + kernel SVM, and SIFT + SVM) on the test images using a slide window method with 5 pixel slide step in both vertical and horizontal directions. Fig. 3 shows the comparison result of the four methods in test images. In Fig. 3, although the HOG + Linear SVM and HOG + Kernel SVM performed a bit better than our method when the recall rate is lower than 0.6, our method performed best when the recall rate is higher than 0.6. The curves show that due to the introducing of color name information into our method and the meaningful segmentation of superpixels for effective scanning the test images, our method is most robust among the four methods. To show the visual experimental results by our method, we exhibit a detection result of an area in test images in Fig. 4. The green and red rectangles represent the correct and wrong detection, respectively. Fig. 4 shows that our method can achieve a high recall rate and maintain a high detection precision at the same time. Even under a complex background as in Fig. 4 (a), our method still worked well and robust.

6. CONCLUDING REMARKS

This paper presents a vehicle detection method from aerial images based on superpixel segmentation and sparse representation. The major work relies on two points. First, we scan the image according to the superpixel segmentation result to significantly accelerate the speed of sliding windows. Second, we introduce the color name feature to be combined with the traditional HOG based descriptor to fully describe the patch information. Through such an effective representation of a patch, we make the detection more robust by using sparse representation. The experimental result and comparison illustrate our method's superior performance and robustness.

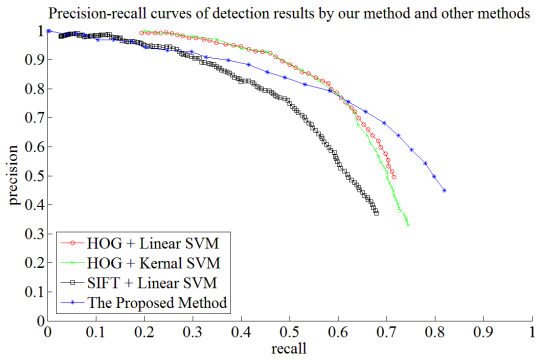


Fig. 2. The precision-recall curves of detection results by our method and other three methods in test images.

(a)

(b)

Fig. 3. Detection results of two subareas in test images. Green and red rectangles are correct detections and wrong detections, respectively.

REFERENCES

[1] Du, R., Chen, C., Yang, B. *et al.*, "Effective Urban Traffic Monitoring by Vehicular Sensor Networks," Vehicular Technology, IEEE Transactions on PP, 1 (2014).

[2] Liu, W., Yamazaki, F., and Vu, T. T., "Automated vehicle extraction and speed determination from Quickbird satellite images," Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 4(1), 75-82 (2011).

[3] Zheng, Z., Wang, X., Zhou, G. *et al.*, "Vehicle detection based on morphology from highway aerial images." 5997-6000.

[4] Leitloff, J., Hinz, S., and Stilla, U., "Vehicle detection in very high resolution satellite images of city areas," Geoscience and Remote Sensing, IEEE Transactions on, 48(7), 2795-2806 (2010).

[5] Kembhavi, A., Harwood, D., and Davis, L. S., “Vehicle detection using partial least squares,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, 33(6), 1250-1265 (2011).
[6] Grabner, H., Nguyen, T. T., Gruber, B. *et al.*, “On-line boosting-based car detection from aerial images,” ISPRS Journal of Photogrammetry and Remote Sensing, 63(3), 382-396 (2008).
[7] Moranduzzo, T., and Melgani, F., “Detecting Cars in UAV Images With a Catalog-Based Approach,” Geoscience and Remote Sensing, IEEE Transactions on 52, 6356 - 6367 (2013).
[8] Hinz, S., "Detection and counting of cars in aerial images." 3, III-997-1000 vol. 2.
[9] Reinartz, P., Lachaise, M., Schmeer, E. *et al.*, “Traffic monitoring with serial images from airborne cameras,” ISPRS Journal of Photogrammetry and Remote Sensing, 61(3), 149-158 (2006).
[10] Choi, J.-Y., and Yang, Y.-K., [Vehicle detection from aerial images using local shape information] Springer, (2009).
[11] Khan, S. M., Cheng, H., Matthies, D. *et al.*, "3D model based vehicle classification in aerial imagery." 1681-1687.
[12] Cheng, H.-Y., Weng, C.-C., and Chen, Y.-Y., “Vehicle detection in aerial surveillance using dynamic bayesian networks,” Image Processing, IEEE Transactions on, 21(4), 2152-2159 (2012).
[13] Zheng, Z., Zhou, G., Wang, Y. *et al.*, “A novel vehicle detection method with high resolution highway aerial image,” Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 6, 2338 - 2343 (2013).
[14] Moranduzzo, T., and Melgani, F., “Automatic Car Counting Method for Unmanned Aerial Vehicle Images,” Geoscience and Remote Sensing, IEEE Transactions on, 52(3), 1635-1647 (2014).
[15] Moranduzzo, T., and Melgani, F., "A SIFT-SVM method for detecting cars in UAV images." 6868-6871.
[16] Shahbaz Khan, F., Anwer, R. M., van de Weijer, J. *et al.*, "Color attributes for object detection." 3306-3313.
[17] Chen, F., Yu, H., and Hu, R., “Shape sparse representation for joint object classification and segmentation,” Image Processing, IEEE Transactions on, 22(3), 992-1004 (2013).
[18] Zepeda, J., Guillemot, C., and Kijak, E., “Image compression using sparse representations and the iteration-tuned and aligned dictionary,” Selected Topics in Signal Processing, IEEE Journal of, 5(5), 1061-1073 (2011).
[19] Yang, J., and Yang, M.-H., "Top-down visual saliency via joint crf and dictionary learning." 2296-2303.
[20] Cheng, M., Wang, C., and Li, J., “Sparse Representation Based Pansharpening Using Trained Dictionary,” Geoscience and Remote Sensing Letters, IEEE 11, 293 - 297 (2013).
[21] Achanta, R., Shaji, A., Smith, K. *et al.*, “SLIC superpixels compared to state-of-the-art superpixel methods,” Pattern Analysis and Machine Intelligence, IEEE Transactions on 34, 2274 - 2282 (2012).
[22] Van De Weijer, J., Schmid, C., Verbeek, J. *et al.*, “Learning color names for real-world applications,” Image Processing, IEEE Transactions on, 18(7), 1512-1523 (2009).
[23] Donoho, D. L., “For most large underdetermined systems of linear equations the minimal ” Communications on pure and applied mathematics, 59(6), 797-829 (2006).
[24] Aharon, M., Elad, M., and Bruckstein, A., “-svd: An algorithm for designing overcomplete dictionaries for sparse representation,” Signal Processing, IEEE Transactions on, 54(11), 4311-4322 (2006).
[25] Tropp, J. A., and Gilbert, A. C., “Signal recovery from random measurements via orthogonal matching pursuit,” Information Theory, IEEE Transactions on, 53(12), 4655-4666 (2007).