Investigating different methods of vehicle detection from aerial images

Somayeh KHAJEHASANI [1,*]

[1] Department of Computer Engineering, University of Technology, Sirjan, Iran
email address: khajeh@sirjantech.ac.ir

Abstract

Superpixel segmentation method designed for aerial images for controlling segmentation with low failure rate is suggested for detecting vehicles from aerial maps with high resolution and accuracy. For greater efficiency of practice and recognition, significant areas are extracted based on segmented superpixel centers. After segmentation, iteration strategy of sample selection based on scattered provision is used to provide a small and complete training subset of the original set. Selected training subset is a method that has the ability to distinguish and differentiate to detect vehicles. Training and detection for the features network of histogram of oriented gradient (HOG) is used to extract features.

Keywords: vehicle detection, aerial images, high resolution, superpixel, segmentation, near-real-time

1. Introduction

Detecting vehicles or cars from aerial images provides very important information for traffic management and urban planning [1], [2]. To enhance efficiency of automatic detection, extraction and formation of typical and rational training set, a complete algorithm that uses scattered provision and superpixel segmentation for automatic vehicle detection with high resolution is provided. To move the detection window without side effects, superpixel-based segmentation was provided for segmenting high-resolution aerial images into a set of superpixels. The important parts are produced based on superpixel centers. Then, scattered provision is used to learn dictionaries and process segmentation. To form the optimal training subset, iteration of sample selection strategy based on scattered provision is used. During the training sample selection, completeness of positive and negative typical and rational samples is taken into account. Using the selected optimal training set, the dictionary of scattered provision with the ability to detect vehicles is achieved. In order to obtain high classification accuracy between vehicles and backgrounds, a training sample set including positive and negative samples is required. Therefore, for training a classifier element, selecting a small and complete subset of the training samples is essential. However, manual and non-automatic selection of all representative samples of a lot of negative samples is hard and time-intensive.

2. First algorithm for detecting vehicles Subtitle

2.1. Related works

Scattered provision and superpixel segmentation are of great significance in terms of computer vision. Scattered provision is successfully used in various fields, including face detection, classification of purposes and objects, image classification, elevating noise of the image, image reconstruction, objective prominence, and data compression [3], [4]. The development of superpixel segmentation and a new method for pre-processing of image, image segmentation, feature extraction, and goal and object tracking are provided. In recent years, most research has focused on image segmentation based on superpixel leading to different approaches. Typical and rational approaches include graph-based algorithms and gradient-based algorithms. The recent success of linear iterative clustering (SLIC) is based on edge-weighted centroidal voronoi tessellations (vcell) and enthalpy rate clustering [5], [6]. The use of scattered provision and superpixel segmentation is a new way for detecting vehicles in aerial images with high resolution. There are many approaches to detect vehicles in high-resolution aerial images. Most approaches can be divided into the two types of models of vehicles, including appearance-based implicit models and explicit models.

Appearance-based implicit models: an appearance-based implicit model usually involves image intensity or contextual characteristics that are calculated by mall core or window, where the given pixel surrounds a small cluster of pixels. Then, detection is carried out through testing the feature detectors of pixels in the middle of the image. Cheng et al. used dynamic Bayesian networks for detection of vehicles from aerial surveillance and found promising results in challenging data sets [8]. The color model, which is especially designed for separating machines from the background color, is still unable to prevent misdiagnosis because of overlap of vehicles and background color models. Another problem is that the approach must combine the recognized pixels to be single vehicles. When vehicles are parked in the vicinity of a small area, it is a hard task. In addition, testing and investigating detecting of all pixels not only increases computational complexity, but also increases the rate of misdiagnosis. Shao et al. investigated vehicle detection using multiple features (e.g. HOG, binary local plan and opposite histogram) and SVM core intersection [7].

The two defects of vehicle detection methods using explicit model are clear. First, the diagnosis is usually based on specified edges, which leads to be unconsolidated against noise and complicated background. Second, these methods are underperforming in cases of obstruction and deformations since the vehicle models are already defined as severe and consistent. Most up-to-date methods of vehicle detection (even commercial software) use implicit model to understand the distant images, which is because of its better generalization ability. But methods using implicit model still suffer from the following problems. Most methods are based on pixel or use a sliding window with sliders stage during detection. Pixel-based methods have heavy calculations. Moreover, these methods must recombine detected pixels into a single photo that is difficult when vehicles are parked close to each other. In the sliding window method, slider stage and step affect processing speed and detection recall rates. A great slider step may increase the speed of processing, but decrease recall rate. A small slider step may reduce recall rates, but will lead to high computational cost. Balancing detection recall rate and processing speed is a problem. The effective

strategy of scanning is desirable for improving scan performance. Second, the training samples are randomly or even manually selected. Manual selection of training samples is time-consuming. To detect vehicles in aerial images, complex background leads to many negative samples, which makes it difficult to manually select an optimal negative training subset. Considering the method of random selection may cause unstable performance and usually does not achieve optimal performance. A more efficient method of training samples selection must be created. As a result, a great need exists for extracting a method for solving the above two problems [15].

2.2 The general framework

As indicated in Figure 1, the general framework of the proposed method consists of two stages of comprehensive training and vehicle detection. In the training phase, the training images are first divided into superpixels using superpixel production method. According to superpixel centers, important parts are produced as a training set and then, a small training subset is selected for initial start of scattered dictionary. In this method, HOG describer network extracts parts as the input of the dictionary. Using the trained dictionary, similarities of remained training samples and vehicles are estimated. Negative samples with the highest similarity and positive samples with the least similarity are selected to be added to the training subset so as to train a new dictionary for the next iteration of sample selection. Iteration of training sample selection is continued to achieve convergence. Here, two situations are considered as convergence mode. First, the trained dictionary contains more than 2,000 samples and components. Second, segmentation accuracy of test images at a recall rate of .7 is more than 80%. When convergence happens, scattered provision dictionary is used to identify vehicles. At the detection stage, the test image is segmented to superpixels that the high-performance test image is scanned based on their centers. Based on scattered codes during scanning, candidates are classified into vehicles and background.
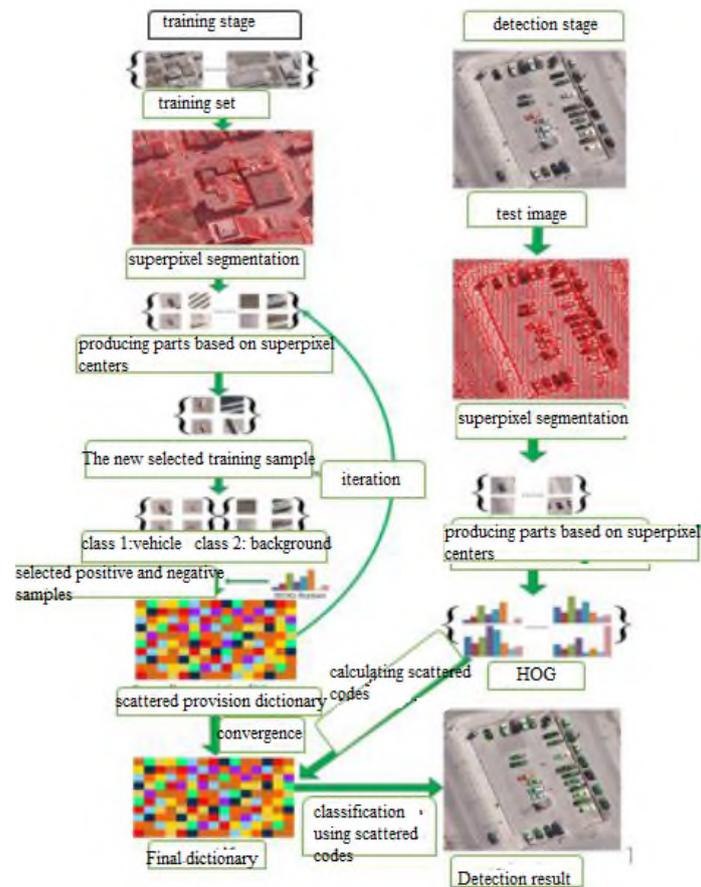
Figure 1: Framework of the proposed method including the two training and detection stages
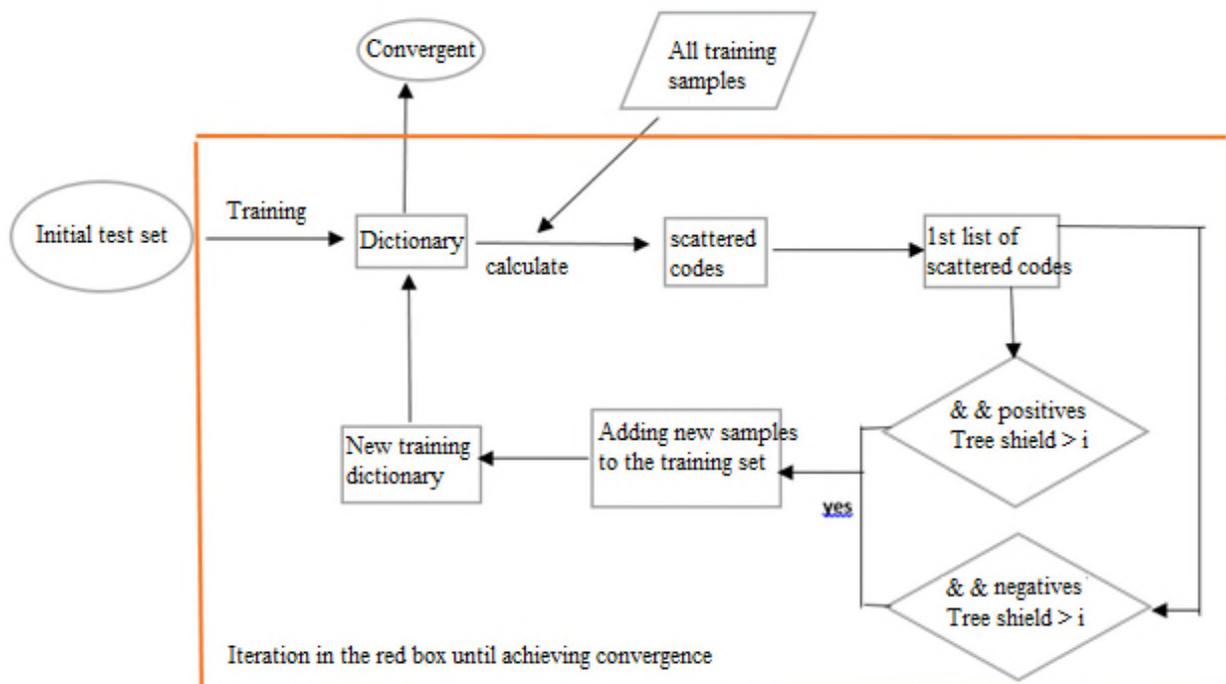

Figure 2: Framework of the proposed method for training sample selection

In this stage, the approach of automatic selection of training samples is introduced, which considers the difference of inner part and completeness of the inner part so as to form the intensive training set to improve the efficiency of segmentation and training.

Figure 2 shows the framework of selecting iterative training sample. First, dozens of samples of vehicles and background are selected manually for initializing the training set that are provided as a result of the new method of detecting vehicles from aerial images with high resolution aerial imagery using scattered provision and superpixel segmentation. Using superpixel segmentation, aerial images are first segmented to superpixels. Based on superpixel centers, key components are obtained for training and detection to benefit from training sample selection and efficiency of scanning. To form a complete and small subset, training sample selection method is proposed based on scattered provision so as to select more representative samples of the entire great training set. With the selected training subset, scattered provision dictionary with high capability of vehicle detection is achieved. Performance of algorithm is improved using the efficient direction estimation so that parts are forced to keep their original direction during training and detection. Three factors of the accuracy of detection in this method are supepixel segmentation size, sample selection iteration time, and completeness of the main entire training set.

3.   Second algorithm for detecting vehicles

Detecting vehicles in images is difficult due to the relatively small size of the target objects and complex backgrounds in the human-made areas. Especially if the goal is to identify the near-real-time, it is more difficult [9]. Detection in just few seconds from great pictures, without any additional information such as the database of roads and the exact size of the target, will be difficult. Here, vehicles in the main frame 21-megapixel image can be detected within seconds, without accurate information about the scale. In addition to the bounding box of vehicles, information on the direction and type (car / truck) is also extracted. First, a double or binary fast detector is employed using the integral channel characteristics in a soft cascade. In the next phase, multi-class classification is used on binary detector that gives the information on direction and vehicle type [16].

Collecting data on traffic and parking from a hovering platform provides a fast coverage over a larger area. Getting the same coverage by ground sensors requires the deployment of more sensors and more manual tasks. In the real-time system, aerial images are taken from the road, vehicles are detected, and multiple consecutive frameworks are traced. This system gives a more comprehensive and rapid information on the traffic situation by providing information about the number of vehicles, their speed and position. Detecting small sizes of vehicles (size of a car may be 12×30 pixels) and complex background of human-made objects that look like vehicles is a challenging problem. The same position and direction of objects detected by the track is caused by restrictions on the movement of vehicles. This is especially important in dense traffic scene in which the allocation and assignment becomes more challenging. The use of roads and parking lots depends more on the type of vehicle (e.g. a truck creates a traffic flow different from a private car). Having access to this rich information, a system can manage better infrastructure. The processing time in a real-time operating system has limited computing power. Therefore, processing method should also be possible and fast. Detection of vehicles is both strong and fast and provides information on the

direction and vehicle type fully automated only with respect to the input image. Bounding box of vehicles is detected using a sliding-window fast dual detector through integral channel features (ICFs) and an AdaBoost classifier in a soft-cascade structure [13]. Bounding boxes are classified for different directions and vehicle types, based on the histogram orientation gradient (HOG) profile [10]. This data set contains the original and non-orthorectified images frame that caused more difficulty because of the unknown GSD (Ground Sampling Distance).

Method provided for vehicle detection has the following features:

It uses features quickly calculated in soft-cascade structures. This has led to the detection process take place more quickly, i.e. the 21-megapixel images are shown in just seconds on the laptop.

The accurate GSD, street databases, and three-dimensional images are worked on single frame images without being georeferenced.

This method also assessed the direction and vehicle type in addition to the position and the location.

3.1. Steps for detecting vehicles

There are many articles and resources to identify vehicles in aerial images; e.g. Moranduzzo and Melagni use high-resolution UAV images for detecting vehicles [11]. For classification, support detector machine (SVM) and a detector with point features for the scale of constant feature describer are used. While other methods use a HOG catalog describer and then estimate and assess direction [12]. In another case, vehicles are detected by a deep neural network that is run on graphic processor unit in the sliding-window approach for fixed scale. In another method, vehicles are detected by online boosting of Haar characteristics such as, local binary patterns and directional histograms. They trained detector for vehicles in one direction and rotated images in 15-degree step during the test. The detector was trained for identified 35×70 pixel size of the objects and was tested on images of the same scale. Leitloff et al. used a two-step approach to identify vehicles; first, they used an AdaBoost classifier with Haar characteristics and then, applied a SVM on different geometry and radiometry profiles. Their road database was already used to detect roads in a specific direction. Although this method with a fast run on a CPU achieved good results, but was restricted with orthorectified images and areas covered by the road database. Tuermer et al. used a road and bounded it with areas under study and by eliminating buildings. HOG characteristics with AdaBoost classifier is used to detect vehicles in selected areas. This approach is bounded by the pair geo-referenced images and areas covered by the road database.

3.2. Multidirectional detection of vehicles

The problem of detecting vehicles is investigated in two steps.

The first step: a very fast binary object slider-window detector is used to provide vehicles' axis aligned bounding boxes without the knowledge of the type of vehicle or direction.

The second step: a multiclass classifier for bounding boxes is applied that detects the direction and type of vehicles. Processing steps are shown in Figure 3.
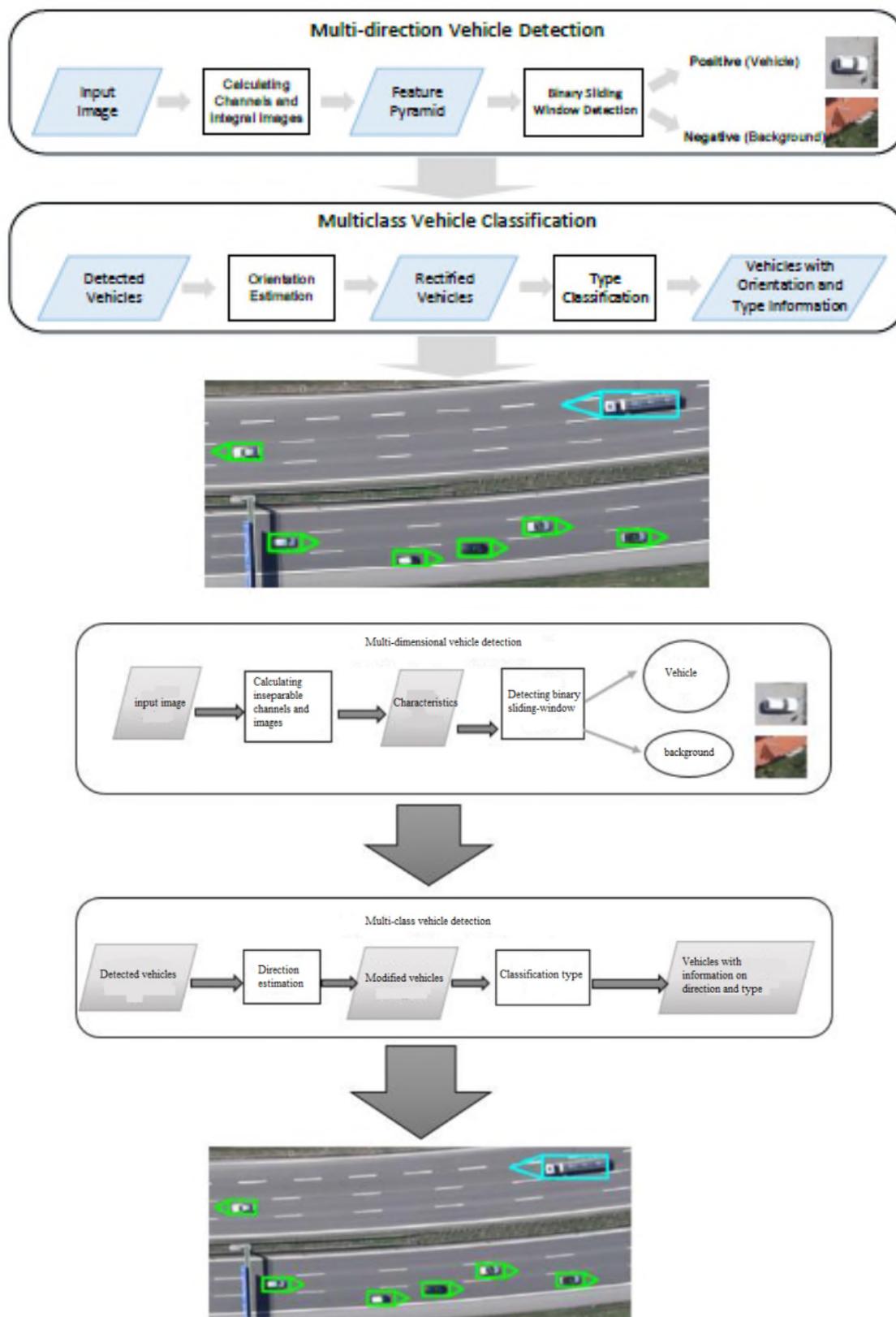
Figure 4: The proposed framework for vehicle detection

The input image is evaluated by a multidirectional vehicle detector. A sliding-window moves along the x-axis and y-axis. Characteristics are extracted from the detection window and are sent to binary trained classifier. The binary classifier identifies the current detection window which contains a positive object or not. Vehicles are detected and then processed to estimate their direction and class.

3.3 Multidirectional detection

Direction of vehicles in aerial images can be arbitrary. This increases intra-class changes of displays in the sliding-window of axis alignment. The solution used is a simple solution, but computationally expensive to train the detector for a specific direction and rotation of the input images and detection for each rotation. The calculations for each direction need a separate integral image and results will be processed slowly. To overcome this, two methods have been proposed: one of them is single trained classifier that is able to identify the vehicles' direction differently and the other is simple aggregated classifiers that each was sensitive to a particular direction.

Single classification method: a single binary classifier has trained samples that cover all directions. Training process is to deal with a variety of high inter-cluster and find common part of all positive samples. When the detector is used on input images, vehicles could be classified as positive samples in each direction.

Aggregated classification method: is another approach that reduces intra-cluster types by dividing training for different directions. Multiple binary classifications have been trained for each specific vehicle direction and the classifiers are used during detection step, respectively. The results of each classifier are summed up using banned non-maximum. Integral image does not need several calculations, it only needs to be classified.

3.4. Multiclass classification of vehicles

Bounding box detectors provide axis aligned of vehicles. In the next step, informations extracted by vehicle direction and type classification are modified and a two-step method including an estimator or direction estimator and a classifier is used. Sample is first selected to direction estimator and then, horizontally rotated based on direction estimation. Finally, it is processed by type classifier to identify the class type of the desired vehicle.

3.5. Direction estimation

Vehicle direction estimation is considered as a multiclass classification problem. Directions are clustered and each cluster is considered as a class. ICF characteristics can be quickly calculated, but there are lots of them; therefore, they are not suitable for multi-class classification of detectors with constant length. HOG strong features that have a detector with a constant length are used. A neural network with a hidden layer is used as a multi-class classifier [14].

3.6. Vehicle type classification

Type classifier needs to classify the input image into corresponding categories. Two types of class or category are defined, for example car and truck. But the proposed method can be extended to more categories. Object bounding box is horizontally rotated according to direction estimates. Unrelated context is cropped and HOG features are re-extracted and classified by type classifier.
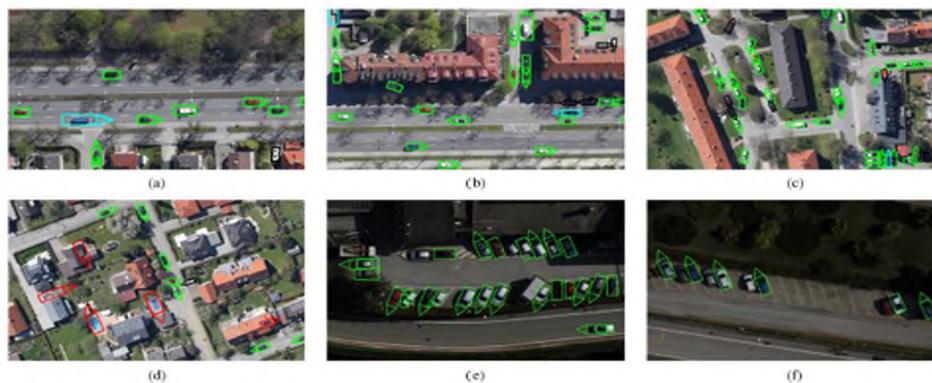

Figure 5: Detection results from the DLR test images

Green and cyan bounding boxes are the correct detected samples that represent cars and trucks, respectively. Black bounding boxes are the missed ones and red are the false positives. The results showed that this method works well in most scenarios of (a)-(c); however, the complicated rooftops or outdoor swimming pools may lead to false positive detections (d). The detection results are shown in (e) and (f).

3.7 Conclusion

In the first algorithm [15] of vehicle detection on high-resolution aerial images, scattered provision and superpixel segmentation is used.

Three factors of detection accuracy, i.e. supepixel segmentation size, sample selection iteration time, and completeness of the main entire training set, affect this method. Experimental analysis of iteration time of training sample selection and processing efficiency was also presented.

In the second algorithm, [16], vehicles can be detected with orientation and type information on aerial images in a few seconds on large images. The application of Integral Channel Features in a Soft Cascade structure results in both good detection performance and fast speed. The detector works on original images where no georeference and resolution information is available. As future work, the performance can be further improved using a deep neural network after using a binary detector. Since this has to be applied only to a fraction of the image, the speed of the detector would be still fast.

Accordingly, the second algorithm is better than the first one in terms of performance and detection speed. However, the first algorithm has greater accuracy than the second one.

References

[1]   B. Tian, Q. Yao, Y. Gu, K.Wang, and Y. Li, "Video processing techniques for traffic flow monitoring: A survey," in Proc. 14th Int. IEEE ITSC, 2011,pp.1103–1108

[2]   W. Liu, F. Yamazaki, and T. T. Vu, "Automated vehicle extraction and speed determination from QuickBird satellite images," IEEE J Sel. Topics Appl. Earth Observ. Remote Sens., vol. 4, no. 1, pp. 75–82, Mar2011.

[3]   J. Leitloff, S. Hinz, and U. Stilla, "Vehicle detection in very high resolution, satellite images of city areas," IEEE Trans. Geosci. Rmote Sens.,Vol. 48, no. 7, pp. 2795–2806, Jul. 2010.

[4]   J. Wright et al., "Sparse representation for computer vision and pattern recognition," Proc. IEEE, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[5]   S.Wang, H. Lu, F. Yang, and M.-H. Yang, "Superpixel tracking," in Proc IEEE ICCV, 2011, pp. 1323–1330.

[6]   X. Zhou, W. Jiang, Y. Tian, and Y. Shi, "Kernel subclass convex hull sample selection method for SVM on face recognition," Neurocomputing, vol. 73, no. 10/12, pp. 2234–2246, Jun. 2010

[7]   R. Achanta et al., "SLIC superpixels compared to state-of-the-art superpixel methods," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[8]   J. Wang and X. Wang, "VCells: Simple and efficient superpixels using edge-weighted centroidal Voronoi tessellations," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 6, pp. 1241–1247, Jun2012.

[9]   J. Leitloff, D. Rosenbaum, F. Kurz, O. Meynberg, and P. Reinartz, "An operational system for estimating road traffic information from aerial images," Remote Sens., vol. 6, no. 11, pp. 11 315–11 341, 2014

[10]   N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE CVPR, 2005, vol. 1, pp. 886–893.

[11]   T. Moranduzzo and F. Melgani, "Automatic car counting method for unmanned aerial vehicle images," IEEE Trans. Geosci. Remote Sens., vol. 52, no. 3, pp. 1635–1647, Mar. 2014.

[12]   T. Moranduzzo and F. Melgani, "Detecting cars in UAV images with a catalog-based approach," IEEE Trans. Geosci. Remote Sens., vol. 52, no. 10, pp. 6356–6367, Oct. 2014.

[13]   X. Chen, S. Xiang, C. Liu, and C. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," IEEE Geosci Remote Sens. Lett., vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[14]   S. Kluckner, G. Pacher, H. Grabner, H. Bischof, and J. Bauer, "A 3D teacher for car detection in aerial images," in Proc. IEEE 11th ICCV, 2007, pp. 1–8.

[15]   Ziyi Chen, ChengWang, Member, IEEE, ChengluWen, Member, IEEE, Xiuhua Teng, Yiping Chen, Haiyan Guan, Huan Luo, Liujuan Cao, and Jonathan Li, Senior Member, IEEE VOL. 54, NO. 1, JANUARY 2016

[16]   Kang Liu and Gellert Mattyus, Member, IEEE, VOL. 12, NO. 9, SEPTEMBER