Efficient Vehicle Detection and Orientation Estimation by Confusing Subsets Categorization

Feimo Li, Xiaosong Lan, Shuxiao Li, Chengfei Zhu, Hongxing Chang
Integrate Information System Research Center
Institute of Automation, Chinese Academy of Science
Beijing, China
e-mail: lifeimo2012, lanxiaosong2012, shuxiao.li, chengfei.zhu, hongxing.chang@ia.ac.cn

***Abstract*—Aerial traffic surveillance requires algorithms that can accurately predict the locations and orientations of hundreds of vehicles in a large high resolution aerial image within seconds. Under this constraint, the classical cascaded detection framework based on boosting algorithms still remains an optimal choice. These methods, however, usually use many binary classifiers to enhance the localization performance resistant to orientation variances, which is not effective in distinguishing confusing orientations and subsets. This paper categorizes these confusing subsets automatically by analyzing the correlations between specific orientation angles and location deviations at local detection window regions, makes robust predictions on them by N-nary multi-class classifiers. This helps to reduce the required number of classifiers to less than half and improve both localization and orientation estimation accuracies, making it potential for additional speed optimization.**

Keywords-high resolution aerial image, vehicle detection, orientation estimation

I. INTRODUCTION

As the expense of aerial remote sensing has been greatly reduced by the recent advances in unmanned aerial system, it is now possible for related departments to make short-time-interval wide range traffic surveillance in urban regions. But this increase in acquisition bandwidth surpasses the capability of manual inspection, thus requires fast and efficient detection algorithms which can predict the locations and orientations of hundreds of vehicles. Detection vehicles in urban region from aerial image is a challenging task, since vehicle are small comparing to the size of the image, and there are prevalent man-made objects with similar texture appearance. To address these difficulties, a number of methods have been proposed for robust vehicle appearance modeling, which can mainly be divided into two categories: methods based on implicit models or explicit models.

Methods based on implicit models make no pre-assumption on the shape and size of the vehicles, and extract local features around a pixel point or a segment region for appearance modeling. For instance, in [1-3], features like Local Binary Pattern (LBP), Histogram of Gradients (HOG), Scale Invariant Feature Transformations (SIFT) etc. are firstly adopted to build a dense feature map, and then fed to the classifier to estimate the location of the vehicle or its components. These methods can make detection in sub-optimal situations when the vehicle is partially occluded, but the computation of a dense feature map can be very expensive, and the final decision on location requires unstable heuristic rules to cluster points to avoid overlapping. Segmentation based vehicle detection methods [4-6] can partially overcome this drawback with better spatial-contextual information utilization, and predict a better region fitted with the vehicle contour. But segmentation on large high resolution image can be very slow, and the appropriate segmentation scale is hard to choose. On the other hand, methods based on explicit model pre-estimate the vehicle geometry by constraining the local feature sampling region, which is often squared [7], rectangular [8,9] or elliptic [10]. Such sampling pattern is often termed as sliding window, and their derivations are mainly focusing on optimal feature selection [11,12] and ideal feature classifier tuning [8].

This paper tries to improve the detection efficiency of a fast sliding window based cascaded vehicle detection method proposed in [8], which originally uses a bundle of single-orientation sensitive binary classifiers to enhance the rotation-invariant localization and orientation angle discrimination. But the independent predictions from binary classifiers are not good at distinguishing adjacent orientation angles and confusing subsets, so this increase in computational cost with extra classifiers does not always promise a better performance. In this paper, a novel method based on Multiple Instance Learning (MIL) is proposed to categorize these confusing orientations and subsets by examining the local pattern correlations between samples with specific location deviations and orientations at local detection window regions. Finally, N-nary multi-class classifiers instead of binary classifiers are used to predict these categories for better discrimination on their subtle relationships. Experiments show the advantage of this method in better localization and orientation estimation.

II. IMPORTANCE OF LOCAL WINDOW REGION

Intuitively, the features from some local detection window regions are more discriminative for samples with specific location deviation or orientation angles. As will be shown in this section, such kind of discrimination can be quantitatively evaluated by using Multiple Instance Learning classifier.

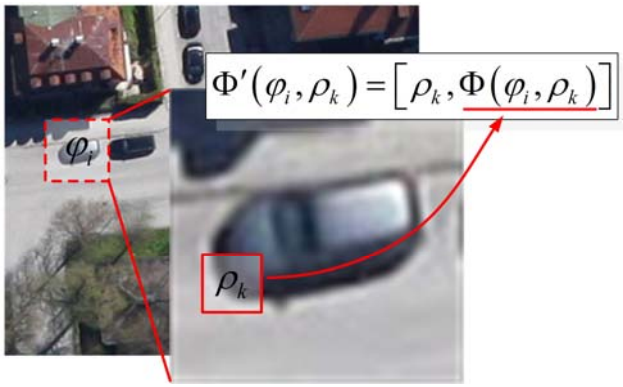

Figure 1. Definition of the bag and instance for local region based positional correlation estimation.

A. *Multiple Instance Learning on Local Window Region*

Multiple Instance Learning (MIL) is a machine learning concept derived from supervised learning. In MIL, labels are incompletely provided for unit called bag, which consisted of multiple data points called instances. A bag is recognized as positive if only one of its instances is positive. The goal of MIL is to train a classifier to predict the label or positive probability of a new bag and its instances, given the data of its instances. This attribute is useful for showing the importance of local region feature in predicting the positional positive likelihood of a given image sample.

As being shown in Fig. 1, the image patch from one of the N detection windows is designated as bag φ_i, and the local features from one of the K local region ρ_k as instance $\Phi(\varphi_i,\rho_k)$. In order to correlate local region with vehicle detection importance in MIL based estimation, the original feature vector is additionally indexed with its local coordination as $\Phi'(\varphi_i,\rho_k)=[\rho_k,\Phi(\varphi_i,\rho_k)]$. MILBoost [13] is choosen as the MIL classifier here, which is based on the boosting framework with its loss function defined as

$$L=-\log\prod_{i=1}^{N}p_i^{\bar{y}_i}\left(1-p_i\right)^{1-\bar{y}_i} \qquad (1)$$

In (1), p_i is the positive probability of image φ_i, and y_i is ground truth label. During training, this loss function is minimized by estimating the positive probabilities of instances related with the probability of the bag in

$$p_i=1-\prod_{k=1}^{K}\left(1-p_i^k\right) \qquad (2)$$

Where in (2), p_i^k is the positive probability of the k-th instance in bag φ_i. This probabilistic XOR definition ensures the positive probability of bag p_i increases as long as any of its instances' positive probability p_i^k is increased. So after the MILBoost classifier is trained with instance feature value $\Phi'(\varphi_i,\rho_k)$ and correlated bag label y_i, it is rerun on the training data and get the positive probabilities of the instances as $p_i^k=\Pr\left(\varphi_i\left|\Phi'(\varphi_i,\rho_k)\right.\right)$.

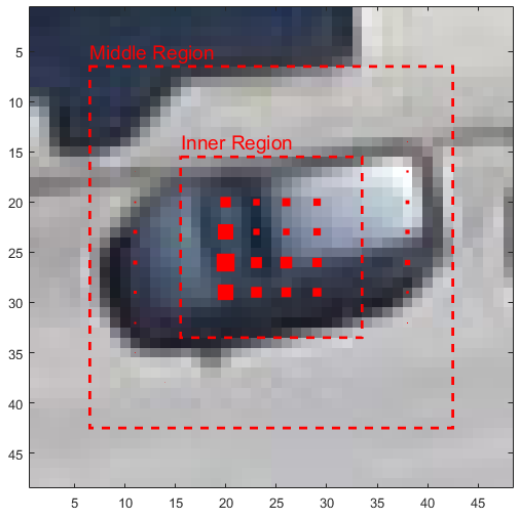

Figure 2. Importance of local regions with sample orientation in location estimation, greater importance has bigger square.

B. *Subset Correlated Importance for Local Regions*

The positive probabilities $\{p_i^k\}$ from previous subsection shows localization importance of local regional feature for each instance, which can be integrated by Bayesian rule to show the correlation between local feature region ρ_k and specific orientation θ_i or any sample subsets. For example, denoting the group of images with orientation θ_i as $G(\theta_i)$, and conditional probabilistic correlation between ρ_k and $G(\theta_i)$ as $\Pr(G(\theta_i)|\rho_k)$, then this probability can be calculated by marginalizing the probabilities of positive local features as in

$$\Pr\left(G(\theta_i)\left|\rho_k\right.\right)=\sum_{\varphi_i\in G(\theta_i)}\Pr\left(\Phi(\varphi_i,\rho_k)\right)\cdot p_i^k \qquad (3)$$

In (3), $\Pr\left(\Phi(\varphi_i,\rho_k)\right)$ is the distribution probability of one feature value at local region ρ_k from one image φ_i in group $G(\theta_i)$, which can be approximated by the reciprocal of set size $\left|G(\theta_i)\right|$ as $\Pr\left(\Phi(\varphi_i,\rho_k)\right)=1/\left|G(\theta_i)\right|$. The resulting local region importance with a certain orientation angle has been shown in Fig. 2, where two set of local regions are sampled from the squared inner region and middle region. Greater importance is illustrated with bigger red square. As can be seen, important regions lay along the main vehicle center line. Such calculation and visualization can also be made for any arbitrary sample subset.

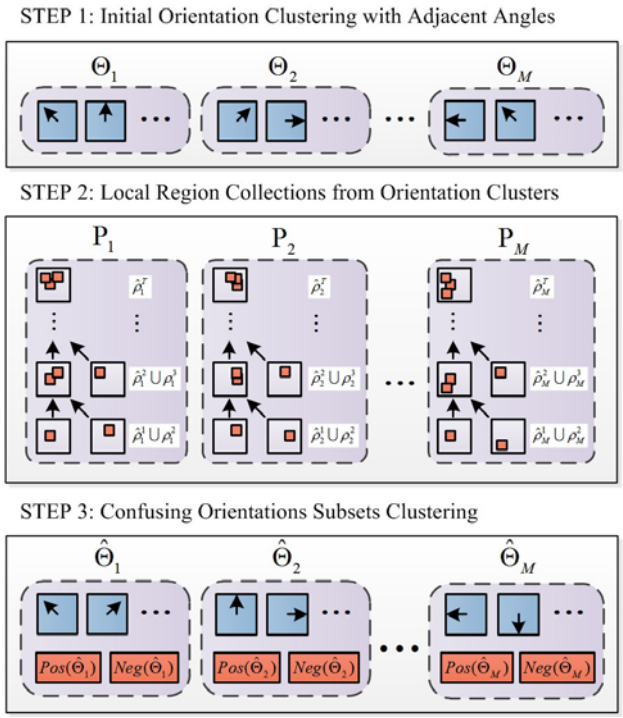

Figure 3. The 3 step confusing orientations and subsets clustering.

III. ORIENTATION AND SUBSET SENSITIVITY AT LOCAL DETECTION WINDOW REGIONS

So equipped with the conditional probability of local regional feature to image patches p_i^k and the calculated local region significance for sample groups to specific orientation $\Pr(G(\theta_i)|\rho_k)$, now it is possible to split out confusingly correlated orientations and subsets based on different local regions clusters. Once these orientations and subsets are found, they are assigned as the categorical outputs of several multi-class classifiers. Thus, ideally, each N-nary multi-class classifier is responsible for disentangling one set of inter-class relationships.

A. Three Step Orientation and Subset Clustering

Assuming there are M multi-class classifiers need to be built, then based on the methodology described in above, M varied set of local regions should be found. But finding the collection of distinctive local regions can easily fall into a chick and egg dilemma, since they can only be differentiated by the orientations and subsets they are more likely correlated with. So in compromise, an even splitting on orientation angles set is made as the prior information for local region clustering. This results in a 3-step procedures clustering method presented in below, and shown in Fig. 3 as three functional blocks.

STEP 1, even orientations splitting with adjacent angles: suppose the 360 degree orientation range has been evenly quantized into R categories as $\{\theta_i\}_R = \{\theta_i \mid i = 1...R\}$, then the splitting in this step generates M orientation subsets $\{\Theta_j \mid j = 1...M\}$, where the angles in each Θ_j are adjacent, and properties $\{\theta_i\}_R = \bigcup_{i=1}^{M} \Theta_i$ and $\Theta_i \bigcap \Theta_j = \varnothing$ hold for the clusters. Denote the group of positive samples with orientation fall within Θ_j as $G(\Theta_j)$, now the local regional positive probability of ρ_k with $G(\Theta_j)$ can be calculated similar to Equation (3) as in

$$\Pr\left(\Theta_j \middle| \rho_k\right) = \sum_{\varphi_i \in G(\Theta_j)} \Pr\left(\Phi\left(\varphi_i, \rho_k\right)\right) \cdot p_i^k \qquad (4)$$

STEP 2, local region collection based on hierarchical binary combination: for each orientation cluster Θ_j, T most correlated local regions are collected through a $T-1$ step hierarchical binary combination. At step t for Θ_j, one of the remaining $K-(t-1)$ local regions is tentatively combined with the $t-1$ selected regions $\hat{\rho}_j^{t-1} = \{\rho_k \mid k = 1...t-1\}$, then the region ρ_j^t with the highest probability $\Pr(\Theta_j \left| \hat{\rho}_j^{t-1} \bigcup \rho_j^t \right.)$ is selected to merge with existing ones. Since the conditional probability for region combination cannot be directly calculated from Bayesian rules, a "minimal-of-two" combination rule is devised as in

$$\hat{p}_{i,j}^t = \Pr\left(\varphi_i \middle| \Phi'\left(\varphi_i, \hat{\rho}_j^{t-1} \bigcup \rho_j^t\right)\right) = \min\{\hat{p}_{i,j}^{t-1}, p_{i,j}^t\} \qquad (5)$$

$$\Pr\left(\Theta_j \middle| \hat{\rho}_j^t\right) = \sum_{\varphi_i \in G(\Theta_j)} \Pr\left(\Phi'\left(\varphi_i, \hat{\rho}_j^{t-1} \bigcup \rho_j^t\right)\right) \cdot \hat{p}_{i,j}^t \qquad (6)$$

Where in (5), $\Phi'(\varphi_i, \hat{\rho}_j^{t-1} \bigcup \rho_j^t)$ is the combined feature vector of the $t-1$ selected regions with region t from image φ_j, $\hat{p}_{i,j}^{t-1}$ and $p_{i,j}^t$ are the abbreviations for the probabilities $\Pr(\varphi_i \mid \Phi'(\varphi_i, \hat{\rho}_j^{t-1}))$ and $\Pr(\varphi_i \mid \Phi'(\varphi_i, \rho_j^t))$. Equation (6) shows the method to calculate the conditional probability of the t local regions collection $\hat{\rho}_j^t$ for the cluster Θ_j, the completed T collection $\hat{\rho}_j^T$ is abbreviated as P_j in Fig. 3.

STEP 3, confusing orientations and subsets clustering based on local region collections: based on the M local region collections $\{\mathrm{P}_j \mid j = 1...M\}$ from STEP 2, the instance level probabilities $\Pr(\varphi_i \mid \Phi'(\varphi_i, \mathrm{P}_j))$ are caculated between each image patch and every region collection. After that, the probability $\Pr(\theta_i \mid \mathrm{P}_j)$ between every orientation angle and region collection is calculated in similar way as in Equation (3) and (6), and the orientation θ_i is assigned to the region collection P_j with the highest $\Pr(\theta_i \mid \mathrm{P}_j)$, where duplicated assignments are skipped. This forms M new orientation angle cluster $\{\hat{\Theta}_j \mid j = 1...M\}$, where size of new angle cluster equals to the old ones, and the splitting properties

$\{\theta_i\}_R = \bigcup_{j=1}^{M} \hat{\Theta}_j$ and $\hat{\Theta}_i \cap \hat{\Theta}_j = \varnothing$ still hold. Additionally, for each new cluster, the positive samples with other orientations and the negative samples are split into subsets based on manually set thresholds on the instance level probabilities $\Pr(\varphi_i \mid \Phi'(\varphi_i, \mathrm{P}_j))$, denoted as $Pos(\hat{\Theta}_j)$ and $Neg(\hat{\Theta}_j)$. Finally, the orientation cluster and subsets for the new cluster $\hat{\Theta}_j$ are made the categorical outputs for a multi-class classifier $f_m^j : x \mapsto \{\hat{\Theta}_j, Pos(\hat{\Theta}_j), Neg(\hat{\Theta}_j)\}$.

B. *Localization and Orientation Estimation based on Predictions from N-nary Classifiers*

The final localization prediction is made by a binary classifier using the classification confidence from the M base multi-class classifiers $\{f_m^j \mid j = 1...M\}$. And for those samples being predicted as positives, an extra N-nary classifier is applied to give out the orientation category labels. The binary classifier for localization is chosen to be the AdaBoost.M1 classifier. While for the N-nary orientation estimator, an Artificial Neural Network with one hidden layer is adopted in consistency with [8].

IV. Experimental Results

A. *The Munich Dataset and Data Augmentation*

The proposed classification scheme is compared with others on the same Munich dataset as in [8]. It contains 20 high resolution aerial images at size of 5616 x 3744 with ground spatial definition (GSD) up to 13 cm taken by the Canon Eos 1Ds Mark III camera system. The dataset is originally split in half for training and testing, and in order to achieve more accurate evaluation for the performance divergence between different classification schemes, 48 x 48 sized image patches are manually extracted at all possible sliding window locations to build training and testing image classification datasets. The gotten image patches with window to vehicle center deviation greater than 3 pixels are marked as negative, and the orientations for positives are evenly quantized into 16 categories with 22.5 degree spacing. Furthermore, every positive image patch is rotated to the other 15 directions for data augmentation, and the positive to negative quantity ratio are set to be 1:5 and 1:10 for the training and testing datasets.

Such configuration aims for enhancing the evaluation on orientation prediction performance, ensures every different classification scheme is trained and tested on data sets with same quantity of location deviations and orientation variances. The negative definition by fixed pixel variance eliminate the randomness post-processing procedures such as non-maximal suppression, thus making the comparison less prejudiced.

B. *Local Regional Features and Classifier Aggregations*

The feature type for local regions are chosen to be HOG with 9 bins, which can be approximated in a fast manner by ICF as being introduced in [8]. These local feature regions are sampled from the inner and middle regions shown in Fig. 2. Where in inner region, HOG features are of size 4 x 4, sampled at step length 2 in a 4 x 4 spatial grid. HOG features from middle ring are of size 6 x 6 along the edges, with 9 equally spaced along each side. This result in 52 local HOG features with total feature dimensions 52 x 9 = 468.

To facilitate description, abbreviations for conventional classification schemes by binary classifiers and proposed scheme by multi-class classifiers are listed in Table. 1.

TABLE I. Classifier Aggregation Naming Routine

Name	Meanings
N x Bin.	Classifier aggregation with N binary classifiers, each is sensitive to a specific range of orientation.
M x Mul. (A/R) + mP + nN	Classifier aggregation with M N-nary multi-class classifiers, each classifier predicts a cluster of orientations, along with m splitting on positives with other orientations and n splitting on negatives. The trailing (A/R) indicates whether the angles in the orientation cluster are adjacent or re-organized.

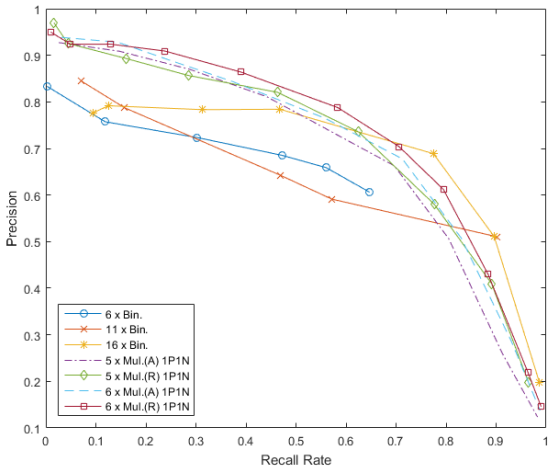

Figure 4. Precision-Recall curves between binary and multi-class classifier aggregations in localization.

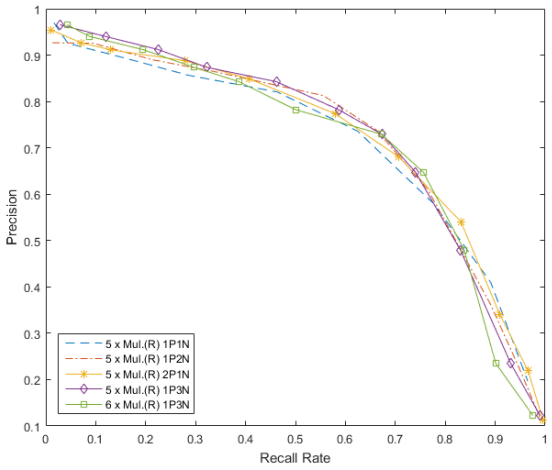

Figure 5. Precision-Recall curves of different multi-class classifier aggregations in localization.

C. *Localization Performance Evaluation*

With the augmented image classification dataset, localization evaluation is equivalent to positive negative binary classification assessment. So the classification schemes are compared by their Recall to Precision curves in Fig. 4 and Fig. 5. As in Fig. 4, for conventional classification scheme with binary classifiers, the increase in number of classifiers does not guarantee better localization performance. For instance, the classifier 11 x Bin. shows worse precision at high recall range compared with 6 x Bin. In contrary, the increased number of classifiers improves the accuracy for our proposed scheme with multi-class classifiers, since more N-nary classifiers provide richer information on more confusing subsets and orientations. Moreover, the reorganized orientations in "5 x Mul.(R) 1P1N" and "6 x Mul.(R) 1P1N" shows better performance than adjacent orientation clustering in "5 x Mul.(A) 1P1N" and "6 x Mul.(A) 1P1N". Benefits of extra splits on negatives and other positives are shown in Fig. 5. As can be seen, increased splits on negative samples will improve the precision at high recall range, while splits on the other positives will improve the prevision at lower recall range. While total classifier number increase brings overall performance improvements.

D. *Orientation Estimation Performance Evaluation*

Orientation estimation accuracies for methods based on pure HOG feature or multi-class confidence scores are compared in Fig. 6, in which X-axis shows the quantized orientation estimation errors ($0\times\Delta\theta$ means correct prediction with zero categorical deviation), and Y-axis shows the number of elements under each estimation error. As being shown, for all three estimation schemes, the most errors lie in $1\times\Delta\theta$ and $8\times\Delta\theta$ (with 180 degree flipped) category. While the "6 x Mul.(A) 1P1N" and "6 x Mul.(R) 1P1N" based on N-nary classification scores generally better than tat based on Raw HOG feature, and the orientation re-organization also improves the angle prediction accuracy.

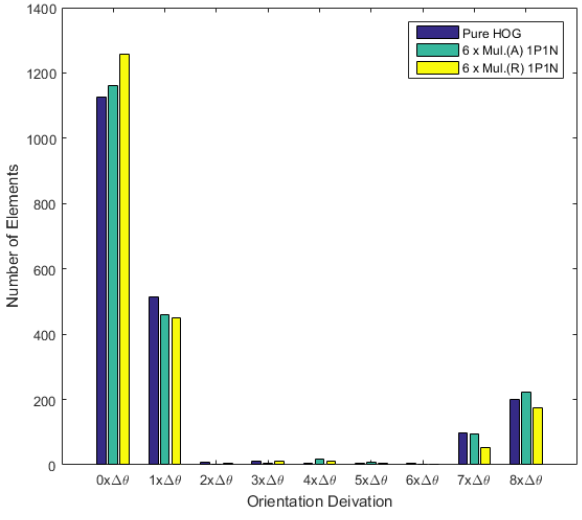

Figure 6. Orientation categorical deviation histogram between real and estimated orientations from different orientation estimation schemes

V. CONCLUSION

In this paper, a fast cascaded sliding window based vehicle detection algorithm is improved by using N-nary multi-class classifiers to model confusing relationships between subsets and vehicle orientation angles. These relationships are measured by Multiple Instance Learning based evaluation of positive confidence at local detection window regions. Correlated experiments show significant improvements in localization and orientation prediction performances for the resulting algorithm. Future efforts will focus on classification speed acceleration for real-time calculation with the reduced number of N-nary classifiers.

ACKNOWLEDGMENT

This work is supported by National Science Fundation of China (NSFC) under grantings No. 61302154 and No. 61573350.

REFERENCES

[1] Moranduzzo, T., Melgani, F.: A sift-svm method for detecting cars in uav images. In: Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, IEEE (2012) 6868 - 6871

[2] Moranduzzo, T., Melgani, F.: Detecting cars in uav images with a catalog-based approach. Geoscience and Remote Sensing, IEEE Transactions on 52 (2014) 6356 - 6367

[3] Moranduzzo, T., Melgani, F.: Automatic car counting method for unmanned aerial vehicle images. Geoscience and Remote Sensing, IEEE Transactions on 52 (2014) 1635 - 1647

[4] Tan, Q., Wang, J., Aldred, D.A.: Road vehicle detection and classifcation from very-high-resolution color digital orthoimagery based on object-oriented method.In: Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International.Volume 4., IEEE (2008) IV-475

[5] Ouyang, Y., Duval, P.L., Sheng, Y., Lavigne, D.A.: Robust component-based car detection in aerial imagery with new segmentation techniques. In: SPIE Defense, Security, and Sensing, International Society for Optics and Photonics (2011) 80200M-80200M

[6] Chen, Z., Wang, C., Wen, C., Teng, X., Chen, Y., Guan, H., Luo, H., Cao, L., Li, J.: Vehicle detection in high-resolution aerial images via sparse representation and superpixels. Geoscience and Remote Sensing, IEEE Transactions on 54 (2016) 103-116

[7] Razakarivony, S., Jurie, F.: Vehicle detection in aerial imagery: A small target detection benchmark. Journal of Visual Communication and Image Representation 34 (2016) 187-203

[8] Liu, K., Mattyus, G.: Fast multiclass vehicle detection on aerial images. Geoscience and Remote Sensing Letters, IEEE 12 (2015) 1938-1942

[9] Chen, X., Xiang, S., Liu, C.L., Pan, C.H.: Vehicle detection in satellite images by hybrid deep convolutional neural networks. Geoscience and Remote Sensing Letters, IEEE 11 (2014) 1797-1801

[10] Kembhavi, A., Harwood, D., Davis, L.S.: Vehicle detection using partial least squares. Pattern Analysis and Machine Intelligence, IEEE Transactions on 33 (2011) 1250-1265

[11] Elmikaty, M., Stathaki, T.: Car detection in high-resolution urban scenes using multiple image descriptors. In: ICPR. (2014) 4299-4304

[12] Tuermer, S., Kurz, F., Reinartz, P., Stilla, U.: Airborne vehicle detection in dense urban areas using hog features and disparity maps. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 6 (2013) 2327-2337

[13] Ali, K., Saenko, K.: Confidence-rated multiple instance boosting for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2014) 2433–2440.