2014 22nd International Conference on Pattern Recognition

Car Detection in High-Resolution Urban Scenes using Multiple Image Descriptors

Mohamed Elmikaty and Tania Stathaki
Department of Electrical and Electronic Engineering
Imperial College London, United Kingdom
{mohamed.elmikaty11, t.stathaki}@imperial.ac.uk

***Abstract*—Robust and efficient detection of cars in urban scenes has many useful applications. This paper introduces a framework for car detection from high-resolution satellite images, wherein a novel extended image descriptor is used to depict the geometric, spectral and colour distribution properties of cars. The proposed framework is based on a sliding-window detection approach and it begins with a pre-prepossessing stage, which discards detection windows that are very unlikely to contain cars, e.g., plain areas and vegetation, followed by the computation of a concatenated feature vector of Histogram of Oriented Gradients, Fourier and truncated Pyramid Colour Self-Similarity image descriptors that is then fed to a pre-trained linear Support Vector Machine classifier to discriminate between the feature and non-feature subspaces. For post-processing, a non-maximum suppression technique is used to eliminate multiple detections. The performance of the proposed framework has been assessed on the Vaihingen dataset and results show that it exceeds the performance of the current state-of-the-art car detection algorithms.**

I. Introduction

Car detection is an important research area in computer vision that has drawn the attention of many researchers as it can be exploited in various applications, such as security, military, traffic-monitoring and urban planning applications. However, it is a very challenging task in computer vision. The reason for this is that there are great differences in appearance, size, and external boundaries among cars in addition to the strong dependence of their appearance on the viewing angle. Nevertheless, differences among the external boundaries of cars are less when cars are captured from a top view by airborne vehicles or satellites than the cases of front, side or rear views. Based on this and on the current availability of several commercial earth-observation satellites [1], and unmanned aerial vehicles that provide high-resolution aerial images with a ground sample distance (GSD) of the order of few centimetres, our goal is to develop a car-detection framework that can operate reliably in real urban scenes captured by satellites.

II. Related Work

Several methodologies for car detection are found in the open literature [2]. Early works have employed 2-D and 3-D explicit models alongside image descriptors as a result of the unavailability of high-resolution images, wherein fine details of objects are obvious, before the wide existence of earth-observation satellites. For example, in [3], [4] and [5], authors addressed the problem of car detection as a 3-D-object detection problem. More specifically, they created a 3-D model, which depicted the geometric characteristics of cars, and incorporated radiometric characteristics. Detection was carried out by matching the model to the image. Choi and Yang [6] adopted a similar bottom-up approach. They used a mean-shift clustering algorithm to form blobs of pixels exploiting geometric and radiometric features. They corroborated the similarity between the distribution of pixels inside these blobs and the expected distribution of pixels of a car using a log-polar-shape descriptor. Problems with using explicit models usually stem from the high computational cost of the matching process which incurs unreasonable processing time.

Later, several car detection techniques have used combinations of multiple features in order to capture different characteristics of cars. For instance, Nguyen *et al.* [7] employed an on-line boosting algorithm. They exploited Haar wavelets [8], Histogram of Oriented Gradients (HOG) [9], Local Binary Patterns (LBP) [10] and used integral images for the fast computation of these features. Then, they used a mean-shift clustering algorithm as a post-processing stage. A similar multiple-feature approach was devised by Kembhavi *et al.* [11], who introduced Colour Probability Maps (CPM) and Pairs of Pixels (PoP) and integrated them with HOG features to form a set of image descriptors of 70,000 dimensions. They reduced this high-dimensional image descriptor using a Partial-Least-Square (PLS) algorithm. Shao *et al.* [12] incorporated HOG, LBP and opponent histograms into a single descriptor of 6760 dimensions. They utilised a greedy non-maximum suppression algorithm and a clustering-on-appearance-based approach to group multiple detections. Although it is widely known that using more than a single feature improves the overall performance of the detection algorithm, the performance is highly dependent on the choice of features.

Other car detection techniques have made use of road maps to extract roads and car parks as these regions are likely to contain cars. As an example, Moon *et al.* [13], [14] exploited the locations of roads and buildings and assumed that cars in parking areas exhibit a 2-D rectangular shape in aerial images. Subsequently, they applied the Derivative of the Double-Exponential (DODE) filter and the problem of car detection was converted into a problem of detecting parallelograms. Similarly, Zheng *et al.* [15] extracted roads and car parks and in order to identify regions that contain cars, they used a morphological pre-processing stage and a classifier that discriminated between car and non-car pixels. In a similar fashion, Eikvil *et al.* [16] utilised a segmentation stage that isolated regions with high probabilities to contain cars, followed by two stages of object classification, in addi-

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.737

tion to multi-spectral images, geometric properties and road networks. Tuermer *et al.* [17], as well, filtered areas that were unlikely to contain cars using road databases, disparity maps and a pre-processing stage before the application of a HOG object detector. Despite the fact that this class of car detection techniques provide very acceptable results, they require the knowledge of precise road maps *a priori* and a map-projection method or a global positioning system (GPS).

In this paper, we focus on the detection of cars in high-resolution urban scenes using the second class of detection techniques which is based on the use of a combination of image descriptors. Our main contributions are: i) the development of a multiple-descriptor vector that robustly detects cars in high-resolution urban scenes, ii) the introduction of rotation-invariant HOG features for objects that possess a dominant edge orientation, iii) the introduction of truncated Pyramid Colour Self-Similarity (tPCSS) and iv) the development of a pre-processing stage based on the physical characteristics of cars and tailored in order to discard detection windows of low probabilities to contain a car.

The paper is organised as follows: in §III, an overview of the proposed multi-stage type of methodology is presented followed by the illustration of the implementation of each stage. Information about the training and testing data is given in §IV. Results are discussed in §V and conclusions are drawn in §VI.

III. Proposed Framework

Cars are expected to appear as parallelogram-like shapes that correspond to their external boundaries. When edges are extracted from a high-resolution satellite image using one of the common edge detectors, e.g., Canny edge detector [18], an image patch placed around a car might reveal two small parallelograms inside that correspond to the front and rear wind-screens and several additional almost parallel and/or almost perpendicular lines that resemble the external boundaries. This set of parallelograms and lines exhibits a significant number of high-curvature points, e.g., corners. Nonetheless, it cannot be utilised to robustly identify cars, because other objects with parallelogram-like shapes, such as building roofs which have similar edge maps, can be easily misidentified as cars. This was our motive to employ more than one image descriptor in order to capture different image cues. The choice of which additional cue, we should combine with the well-established gradient type of descriptors, is a challenging problem since the overall performance of the final detection is highly dependent on the type of object class under investigation. In addition, theoretical foundations about this choice are still immature and only some guidelines are found in the open literature. Consequently, in most cases, decisions are made using empirical results.

The proposed framework consists of a pre-processing stage followed by the computation of three image descriptors, namely HOG, Fourier and tPCSS descriptors over a 64 x 64 detection window and the associated stride length is eight pixels. These descriptors characterise the spatial distribution of pixels, spectral information and local colour distribution as it will be shown later in this section. A fixed-size window has been used as cars in the dataset, we have worked on, are of almost same size. These feature descriptors are concatenated to form a single hybrid descriptor that is fed to a linear SVM classifier, which discriminates between the feature subspace and the non-feature subspace. Finally, a non-maximum suppression technique is applied to eliminate multiple detections per single true object.

A. Pre-processing Stage

During the pre-processing stage, detection windows that are very unlikely to contain cars (have low level of objectness) are eliminated based on several cues to shorten the processing time and help reduce the false-alarm rate [19]. First, a matrix of the magnitudes of gradients is computed over the detection window using a non-smoothing filter of [-1, 0, 1] vertically and horizontally for each colour channel separately and only the maximum gradient is kept. If the number of gradients that have magnitude near zero is greater than 1365 (one third of the detection window size), the detection window is ignored because this indicates that the detection window does not contain sufficient amount of gradients and most probably it belongs to a flat area, such as empty roads. It is worth mentioning that this matrix of gradients is used later in computing HOG features. Consequently, the formation of this matrix does not require extra processing time.

Second, for a grey-scale version of the RGB image under examination, a typical detection window that contains a car is expected to have a bi-modal intensity histogram. One mode represents the grey shade of the car's body and the other mode represents the dominant grey shade of the surrounding background. In addition, the score of the bins of very dark shades, which correspond to the grey levels of front and rear wind-shields, must exceed a pre-defined threshold. Therefore, we convert the entire RGB image to a grey-scale image for which we create a 16-bin histogram. Detection windows that have a single mode with a very high score (greater than 2500) are discarded as these are of high probabilities to be parts of roads, vegetation areas or roofs of buildings. Detection windows that do not have a sufficient number (less than 500) of dark pixels of intensity less than 25 are discarded as well. Despite the simplicity of the pre-processing stage, it helps reduce regions that will be checked by the classifier and consequently reduces the processing time.

B. Computation of the Extended Image Descriptor

As previously mentioned, three types of features are computed. Implementation details of these features are given below.

1) Histogram of Oriented Gradients: Histogram of Oriented Gradients was devised by Dalal and Triggs [9] mainly for pedestrian detection. However, it has been recently exploited in many detection algorithms for objects other than pedestrians. It is a powerful single-feature descriptor that captures the spatial distribution of edge pixels [20]. In order to compute the HOG descriptor for an image $I(x, y)$, a fixed-size detection window of 64 x 64 pixels is subdivided into evenly spaced cells of 8 x 8. For each cell, the first discrete derivatives $\frac{\partial I}{\partial x}$ and $\frac{\partial I}{\partial y}$ of I are computed using a non-smoothing filter of [-1 0 1] to compute the intensity gradient at each pixel in the vertical and horizontal directions for each colour channel separately and only the maximum gradient magnitude $M(i,j) = \sqrt{(\frac{\partial I(i,j)}{\partial x})^2 + (\frac{\partial I(i,j)}{\partial y})^2}$

is kept. Then, orientations $O(i,j) = arctan(\frac{\partial I(i,j)}{\partial y} / \frac{\partial I(i,j)}{\partial x})$ are accumulated in a 1-D histogram of nine-orientation bins with a weighted vote equals to the magnitude of the gradient where i and j represent the pixel indices. Votes are interpolated tri-linearly in both position and orientation. A 16 x 16 pixel square block, which consists of four overlapping cells, is formed. The contrast of each block of cells is locally normalised using $L2{-}norm, h \rightarrow h/\sqrt{\| h \|_2^2 + \epsilon^2}$, where h is the unnormalised descriptor and ϵ is a small constant equals to 0.01 in our implementation, in order to have a better invariance to illumination conditions.

Rotation-Invariant HOG: The original implementation of the HOG descriptor does not accommodate for rotation invariance. However, as mentioned before, an edge map of a car consists of a set of almost parallel and/or almost perpendicular lines. Therefore, if the dominant edge orientation within the patch is computed, it will be perpendicular to the orientation of the car in the original image. A nine-bin global histogram of orientations is formed to compute the dominant edge orientation and the bins of the local histogram of oriented gradients of each cell are circularly shifted so that rotation invariance is achieved.

2) Fourier Descriptor: The second image descriptor, we employ, is the Fourier descriptor, which captures the spectral information of a given image. The definition of the 2-D discrete Fourier Transform can be found in [21]. In our implementation, the detection window is rotated so that detection windows that have cars will have a vertical orientation, followed by cropping a centred square of 48 x 48 pixels to remove extra parts that are added as a result of the rotation of the original detection window. Then, the Fast Fourier Transform of a grey-scale version of the image is computed. Furthermore, we use a shifted spectrum, wherein the two-dimensional frequency pair (0,0) is transferred from the top left corner of the given image patch to the middle (centre) of it. The Fourier feature vector is formed by reshaping the matrix that contains the $L2$ normalised magnitudes of Fourier Transform coefficients into a one-dimensional vector.

An important feature of the Fourier transform is that if an edge line exists in the spatial domain, it will appear as a line in the spectral domain but with an orientation perpendicular to its original orientation. This feature is exploited in our algorithm as cars mainly consist of dominant line edges with almost parallel/perpendicular orientations, whereas other objects are expected to have edges with different orientations as shown in figure 1. As a result, the Fourier descriptor of a window containing a car has a very distinctive shape, wherein most of the energy is concentrated along the vertical and horizontal axes. Hence, the feature vector of Fourier coefficients for a car patch will be different from that of a non-car patch, wherein energy is distributed on most of the coefficients.

3) Truncated Pyramid Colour Self-Similarity Descriptor: Self-similarity descriptors are regularly used to describe the local or global similarity between pixels in a given image. Regarding a detection window that contains a car, it is expected that adjacent pixels that belong to a car body should exhibit a great similarity. Furthermore, there is always a local variation between pixels that belong to windscreens and pixels that belong to car bodies. Consequently, using a self-similarity descriptor will depict these distinctive features of local similarity

Fig. 1. Application of the Fourier transform to a car and non-car rotated patches. From left to right: original patches, corresponding edge maps, rotated patches and corresponding spectral domains.

Fig. 2. Samples of the training dataset. Left: Positive training data; right: negative tarining data

and variation across intensity levels. We modified the Colour Self-Similarity descriptors proposed by [22] and [23] to form the tPCSS descriptor. This self-similarity descriptor depends on the calculation of the intersection of local histograms to measure the similarity between two intensity histograms, has a low computational cost and a low dimensionality and can be computed as follows:

$$d_{intersection}(H_C, H_N) = \left\lceil \sum_i min(H_C(i), H_N(i)) \right\rceil \quad (1)$$

where H_C and H_N are the two grey-level histograms of the central cell and a neighbouring cell respectively , $H_C(i)$ and $H_N(i)$ are scores of the i^{th} bins of the two histograms and $\lceil \rceil$ indicates that the result is truncated at the size of an unsigned integer of eight bits. We have found that this truncation is necessary in order to have a high recall rate as it tolerates some variation among positive patches.

In our implementation of the self-similarity descriptor, the extracted rotated grey-scale detection window in the previous stage is divided into a number of square cells. The descriptor is then computed on two levels separately. In level one, the cell size is 4 x 4 pixels, whereas in level two, the cell size is 8 x 8 and an overlapped neighbourhood of eight cells in both levels. The self-similarity feature vector is formed by the concatenation of the $L2$ normalised responses from the intersections of cell histograms.

All values of thresholds and parameters used in the pre-processing and feature computation stages were found by testing excessively the properties of a large number of image patches from the training dataset.

Fig. 3. The three test areas in Vaihingen city - Germany. They contain typical urban backgrounds, wherein there are different builings, roads and vegetation. Left image: area 1: a dense development consisting of historic buildings having rather complex shapes, but also has some trees; middle image: area 2: a few high-rising residential buildings that are surrounded by trees; right image: area 3: a pure residential area with small detached houses

C. Classification

The proposed overall extended image descriptor is formed by the concatenation of 1764 HOG features, 2304 Fourier features and 928 tPCSS features. The length of the extended feature vector is 4996 features. A linear SVM classifier with a regularisation parameter "c" equals 15, which was chosen using cross validation on the training dataset, is fed by the formed descriptor vector to distinguish between the feature subspace and the non-feature subspace.

D. Post-processing

Since a sliding detection window has been adopted in our framework, multiple detections per one single true detection are expected. Bounding boxes that overlap by more than 50% are eliminated except the one that has the highest output confidence score. Figure 4 shows that multiple detections are successfully eliminated.

IV. Datasets

The Vaihingen dataset has been used for training and testing purposes. This is a set of high-resolution (8cm GSD) satellite images that were captured over the city of Vaihingen - Germany by the German Society for Photogrammetry.

A. Training Dataset

We cropped 754 positive patches, which contained cars placed at different orientations and on various backgrounds, from areas other than the areas specified by the German association for testing, shown in figure 2. The square patches were cropped such that there were at least six pixels around the body of each vehicle so that gradients could be efficiently computed and then patches were rescaled to a size of 64 x 64 pixels. In order to increase the diversity among car orientations in the training dataset, mirrored and horizontally flipped versions of these 754 rescaled patches were included. For the negative training data, a total of 3601 patches, including patches from different backgrounds and bootstrapped hard negatives that were fired as false alarms during validation, were used.

B. Testing Dataset

The proposed car-detection framework was tested on the three areas specified by the German association and they are shown in figure 3.

V. Experimental Results

In this section, the performance of the proposed framework is assessed based on a number of quantitative measures: the detection rate, the false-alarm rate and the miss-detection rate by plotting precision-recall curves. Precision (Pr) and recall (Rc) can be defined by equations 2 and 3:

$$Pr = \frac{\#Actual\ true\ detections}{\#Total\ detections} \tag{2}$$

$$Rc = \frac{\#Actual\ true\ detections}{\#Total\ true\ detections} \tag{3}$$

A detected bounding box (B_d) is considered true if it matches a corresponding ground truth bounding box (B_g) using equation 4 and there is no other bounding boxes matching the same ground truth [24]. In case of more than one bounding box matching the same ground truth, only one of them is considered a true detection and others are considered false alarms. Ambiguous objects, such as objects which are partially occluded, are not taken into account when measuring the performance of the proposed framework.

$$\frac{Area(B_d \cap B_g)}{Area(B_d \cup B_g)} \geq 0.5 \tag{4}$$

Figure 4 shows examples of the test areas and the experimental results of the proposed algorithm. It can be observed that few false alarms are fired when patches under examination contain objects with very similar structure to cars or parts of different objects form together a very similar structure.

Figure 5 shows the precision-recall curve for the proposed framework in different test areas. It performs best in areas that contain small detached houses, whereas in areas that contain

Fig. 4. Experimental results for the proposed car-detection framework in urban scenes (sample subimages of the test areas are shown). The number at the top of each bounding box represents its confidence score.

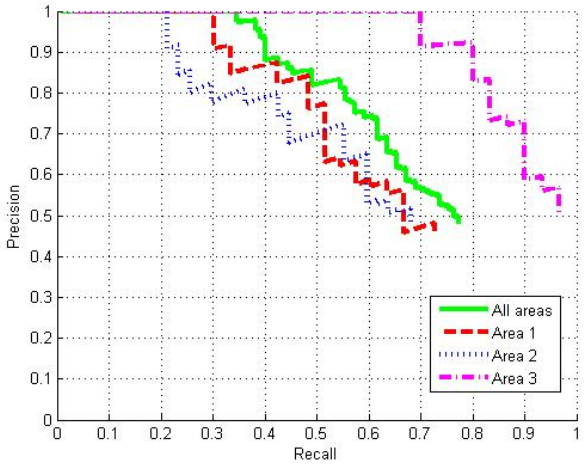

Fig. 5. Precision-recall curves for the proposed framework in different test areas.

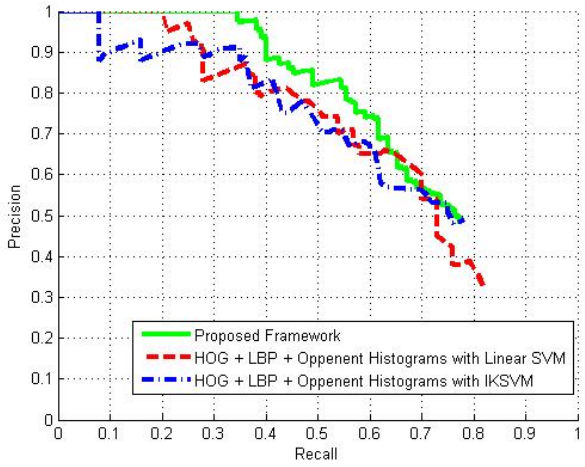

Fig. 6. Precision-recall curves for proposed framework and the work in [12]

high-rising residential buildings and historic buildings with complex shapes, the performance degrades. Furthermore, we compare our results with the latest proposed algorithm [12] for car detection using the same dataset as it follows a similar multiple-feature approach (6760 dimensions), which are HOG, LBP and opponent histograms. Figure 6 shows that for the same recall rate, the proposed framework achieves higher precision rate than that of the work in [12]. Moreover, we are using a lower-dimensional image descriptor (4996 dimensions)

combined with a linear SVM classifier (faster classification).

Running-Time Analysis: Our system has been implemented in MATLAB and all of our experiments were run on an Intel-Core-i5 CPU with speed of 2.5 GHz and a Random Access Memory of 6.00 GB size. Whilst the machine that was used has multiple processing cores, our current implementation of the algorithm uses only a single core. The algorithm is able to process $\approx$ 76 detection windows per second, which indicates that the current implementation has a near-real-time performance.

Effect of Using the Pre-processing Stage: The pre-processing stage helps reduce the average computational time per frame by 41.74% and the rate of false alarms by 26.26%, whilst keeping the same recall rate of the SVM classifier.

VI. Conclusion

We have developed a novel framework for car detection from high-resolution satellite images that can be employed for detection in urban scenes based on the use of multiple image descriptors, namely the HOG, Fourier and truncated Pyramid Colour Self-Similarity, combined with a linear SVM classifier. Upon testing the algorithm on the Vaihingen dataset, results have shown that the proposed framework outperforms the state-of-the-art. Furthermore, it can be concluded that the optimal choice of image cues to be used to depict a particular class of object can lead to the achievement of an improved performance using low-dimensional feature vectors and a linear classification.

Acknowledgement

This work is funded by EPSRC and BAE SYSTEMS Military Air and Information (UK). The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [25] http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html.

References

[1] "The earth observation handbook," 2012.

[2] V. Ramakrishnan, A. K. Prabhavathy, and J. Devishree, "A survey on vehicle detection techniques in aerial surveillance," *International Journal of Computer Applications*, vol. 55, no. 18, pp. 43–47, October 2012.

[3] T. Zhao and R. Nevatia, "Car detection in low resolution aerial image," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001, vol. 1, pp. 710–717 vol.1.

[4] S. Hinz, "Detection and counting of cars in aerial images," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, 2003, vol. 3, pp. III–997–1000 vol.2.

[5] S. Hinz, C. Schlosser, and J. Reitberger, "Automatic car detection in high resolution urban scenes based on an adaptive 3D-model," in *Remote Sensing and Data Fusion over Urban Areas, 2003. 2nd GRSS/ISPRS Joint Workshop on*, 2003, pp. 167–171.

[6] J. Choi and Y. Yang, "Vehicle detection from aerial images using local shape information," in *Advances in Image and Video Technology*, T. Wada, F. Huang, and S. Lin, Eds., vol. 5414 of *Lecture Notes in Computer Science*, pp. 227–236. Springer Berlin Heidelberg, 2009.

[7] T. Nguyen, H. Grabner, H. Bischof, and B. Gruber, "On-line boosting for car detection from aerial images," in *Research, Innovation and Vision for the Future, 2007 IEEE International Conference on*, 2007, pp. 87–95.

[8] P. Viola and M. Jones, "Robust real-time face detection," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001, vol. 2, pp. 747–747.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 1, pp. 886–893 vol. 1.

[10] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[11] A. Kembhavi, D. Harwood, and L.S. Davis, "Vehicle detection using partial least squares," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 6, pp. 1250–1265, 2011.

[12] W. Shao, W. Yang, G. Liu, and J. Liu, "Car detection from high-resolution aerial imagery using multiple features," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, 2012, pp. 4379–4382.

[13] H. Moon, R. Chellappa, and A. Rosenfeld, "Optimal edge-based shape detection," *Image Processing, IEEE Transactions on*, vol. 11, no. 11, pp. 1209–1227, 2002.

[14] H. Moon, R. Chellappa, and A. Rosenfeld, "Performance analysis of a simple vehicle detection algorithm," *Image and Vision Computing*, vol. 20, no. 1, pp. 1 – 13, 2002.

[15] H. Zheng, L. Pan, and L. Li, "A morphological neural network approach for vehicle detection from high resolution satellite imagery," in *Neural Information Processing*, J. King, I.and Wang, L. Chan, and D. Wang, Eds., vol. 4233 of *Lecture Notes in Computer Science*, pp. 99–106. Springer Berlin Heidelberg, 2006.

[16] L. Eikvil, L. Aurdal, and H. Koren, "Classification-based vehicle detection in high-resolution satellite images," {*ISPRS*} *Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 1, pp. 65 – 72, 2009.

[17] S. Tuermer, F. Kurz, P. Reinartz, and U. Stilla, "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*, vol. PP, no. 99, pp. 1–11, 2013.

[18] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[19] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 11, pp. 2189–2202, 2012.

[20] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 4, pp. 743–761, 2012.

[21] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, Thomson Learning, 3rd edition, April 2007.

[22] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 1030–1037.

[23] N. Wang, X. Gong, and J. Liu, "A new depth descriptor for pedestrian detection in rgb-d images," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 3688–3691.

[24] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2008 (VOC2008) Results," .

[25] M. Cramer, "The DGPF-test on digital airborne camera evaluation-overview and test design," *Photogrammetrie-Fernerkundung-Geoinformation*, vol. 2, no. 2, pp. 73–82, 2010.