Segmentation methods for detection of stationary vehicles in combined elevation and optical data

Dimitri Bulatov and Hendrik Schilling
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation
Gutleuthausstr. 1, 76275 Ettlingen, Germany, dimitri.bulatov@iosb.fraunhofer.de

***Abstract*—Detection of vehicles in remote sensing data represents a captivating and challenging task that has been studied during many years. The state-of-the-art detection tools can be subdivided into implicit and explicit methods; the latter ones provide detection results by means of some explicitly characterizing features. Mostly, these methods rely on optical aerial images in which vehicles appear distorted. However, 3D elevation data and orthophotos are increasingly available and typically used to perform a full context-based scene analysis of which vehicles are an indispensable part. In this paper, we propose to combine elevation and optical data for segmentation of vehicle-like objects. To do this, several strategies, their advantages and disadvantages, will be discussed. Since any segmentation method also produces numerous false alarms, we will briefly describe the complete vehicle detection pipeline. The results indicate that sensor data fusion is crucial for obtaining the most accurate results in a reasonable time. For example, using trapezoids or stripes formed in optical and elevation data allows one to detect almost all targets with a very high accuracy exceeding the results obtained from single sensor data. We perform an extensive evaluation of all presented methods and outline the main ideas for correction of the existing shortcomings and for a closer embedding of vehicle detection into the process of urban terrain reconstruction from sensor data.**

Fig. 1. View of a urban terrain reconstruction result for the complete data set Vaihingen from Sec. III, where detection of buildings and trees, as well as reconstruction and texturing of building roofs was performed as in [2]. Areas of the orthophoto occupied by vehicles are inpainted by means of the method [3] and three 3D models of vehicles were inserted as an example.

I. Introduction and related work

Detection and recognition of small objects, especially vehicles, is a very important and interesting yet a complicated problem of Computer Vision and its applications. In remote sensing, the importance of vehicle detection is stated by numerous applications such as road verification, traffic monitoring, and civil security. Using airborne data, large areas can be surveyed. But there are drawbacks like low resolution, high variation of background objects and, in general, low availability of labeled training data as the appearance of objects strongly depends on the sensor and recording parameters. In our applications, the vehicle detection module is embedded into the highly complex pipeline for urban terrain reconstruction and modeling from sensor data and its representation in simulation environments for training and rehearsal purposes. With respect to a photo-realistic simulation, we represent the ground texture by an orthophoto in which appearance of parking (or moving) vehicles is not desirable. They should be better identified and overpainted, see, for example, [1] and Fig. 1. Detection of vehicles is not only interesting because it is a theoretically challenging problem of finding a needle in a haystack, but also because of numerous challenges due to variations in their appearances, occlusions, reflections, shadows, and of course, varying resolution.

It is out of scope of this paper to handle all possible kinds of vehicle configurations, sensor data, and detection methods. We concentrate on detection of stationary vehicles, because moving objects should be better detected by considering changes between subsequent images additionally to the extracted features [4]. Principally, related work suggests that detection of vehicles from aerial images alone is possible, see [1], [5], [6], [7], [8], [9], [10], [11]. The first three methods can be considered as implicit, which means that rather generic, context-independent features (e.g. Haar-like or Histogram of oriented Gradients) are calculated and stored for every pixel of the image. These features are processed by a sequence of weak classifiers (boosting) which eventually creates a strong classifier. On the contrary, the explicit approaches use typical features of the car shape. An example of such an approach, [10], has found out that especially the rectangular shape, presence of front windshields, position of the car preferably in a street-like regions (streets, parking lots, etc.) and shadow casts are important – in the mentioned order – for vehicle detection in aerial images. The 3D wire-frame model for several vehicle types is used for comparison of simulated projections into the image with the actual image content. The weak point of the algorithm is that only cars parallel to the street course can be detected.

Though excellent results were obtained using aerial images only, one should expect things to become more interesting and hopefully efficient if elevation information is available for feature extraction instead of or additionally to color images. However, it seems that – despite the continuously improving quality of sensors and state-of-the-art methods for 3D reconstruction from aerial images – many authors do not rely on features based on 3D results but rather use 3D information for projection and narrowing the search space. From other references concerning detection of vehicles using elevation information, the principal novelties of [12] are given firstly by a simplified computation of the Normalized Digital Surface Model (NDSM, that is, the difference between the elevation map DSM and the digital terrain model DTM) and secondly by increasing the number of training examples by considering small rotations and scaling of the already existing ones. In [13], the disparity calculation is applied to improve segmentation of roads for which a coarse information is given from external sources. This improved classification of street pixels yields a smaller region for searching for hypotheses; consequently, vehicles positioned further away from the streets cannot be detected. Both methods use boosting classifier applied on simple image features.

In this work, we wish to explore to what extent segmentation of vehicle-like objects can be performed in *combined elevation and optical data*. The typical input for any semantic urban terrain reconstruction procedure is a DSM and an orthophoto; both may stem from airborne laser scanning or from the 3D reconstruction given aerial images [14]. At each case, as [15] pointed out, elevation data is an extremely useful source of information, which can and should be extensively exploited. Our second difference with respect to [10] is that we are interested in improving the quality of segmentation of cars without constraints on orientation, vicinity to a-priori estimated street course, etc. A possibly clean segmentation, performed either after [5], [6] or before [13], [16] actual detection, always leads to better classification results and reduces computation time. State-of-the-art segmentation algorithms as well as customized ones will be presented in Sec. II with the aim at obtaining as many as possible correct and complete hypotheses. The elimination of false alarms is then carried out at a later stage (classification), as will be explained in Sec. III together with the description of data sets. Also, since there is no approach which equally suits all applications, a reasonable evaluation strategy will be discussed in order to get a well-grounded quality assessment. Finally, Sec. IV summarizes the contents of the paper and outlines ideas for future work.

II. Tools for vehicle detection

The task of this section is to investigate to what extent segmentation of vehicle-like regions can be performed in combined elevation and optical data. As already mentioned above, the important criteria for the detection quality are *per-object completeness* (as many targets as possible) as well as *per pixel accuracy* (clean segmentation in order to make features more discriminative). The per-object false alarm rate is less important at this stage: though it negatively affects the computation time, most of the false alarms can be discarded by means of computationally cheap features together with a state-of-the-art classifier.

D1: [Top-hat operator] The initial segmentation of the procedure proposed by [15] implied application of the *top-hat* filter on the elevation map. The top-hat transformation is actually a thresholded difference between the image and its morphological opening with a given filter size. All elevated objects that have a larger area than defined in the filter are deleted by this procedure. Hence, the size of the filter coarsely corresponds to the typical size of a vehicle. Additionally, in order to perform the separation of elevated objects along gradient jumps (that is, vehicles near the building wall), pixels with a gradient norm value exceeding a fixed threshold are set to background level. This is done by filling heights of these points by a morphological reconstruction process [17]. Strictly speaking, computation of DTM is not necessary for this module, which saves computing time and is therefore an advantage. However, computation of DTM may be necessary for other steps of the building reconstruction process.

D2-D3: [Elevation, NDVI, and planarity thresholding] In case DTM is computed, e. g. , by means of the standard procedure [2], the intuitive way for segmentation is to subject the NDSM to a threshold describing the typical vehicle size and to suppress the pixels with a high value of NDVI (Normalized Differenced Vegetation Index) as well. In absence of the near infrared channel of the orthophoto, the green channel can be used. The method D2 means that connected components of the resulting binary image are labeled and considered as vehicle hypotheses. The main disadvantage of this method is short-comings of the NDSM: cars parked over parking levels or small hills cannot be detected by height-thresholding. The same problem arises also for groups of cars parking close to each other. The planarity map, computed by means of the approach based on Eigenvalue analysis, is then applied in order to separate outstanding objects. Now, the detection result [D3] is given by the intersection of the mask D2 with the thresholded planarity image. Vehicles should be more clearly distinguished because the area they occupy has mostly low planarity.

D4-D6: [Stripes] Many authors [9], [10] rely on quasi-rectangular shapes of vehicles and search for such structures in images. Because the probability that at least one side of the vehicle is occluded is not negligible, we applied a stripe detector which has as inputs straight line segments [18] detected in the orthophoto [D4], in the DSM [D5], and in the union of both [D6] while the union of D4 and D5 is denoted by D6a. A computationally efficient way to compute stripes is given by the work of [19]. Besides, the domain for computation of lines can be restricted to the region specified by D2: thus, lines lying in regions with a too high value of NDVI or implausible value of NDSM are excluded from consideration. By assessing the distances between the centroids of the remaining lines and their orientation modulo π, a preselection of candidates for stripes can be obtained with a range search method. After this, two more computationally expensive tests are applied. First it must be checked whether their distances lie within a predefined range and second, whether their displacement lies below a fixed threshold. Given the set of stripes, it is an analogous, straight-forward job to compute the rectangles (denoted as R-structures) and the so-called U-structures, which comprise only three rectangle edges. For the sake of completeness – since the number of detected targets will not increase – the R and U-structures were computed in the orthophoto where the threshold deviation to the right angle was set to 10°. In the approach denoted by D4a, we evaluate both R- and U-structures while in the approach D4b, merely R-structures are taken into consideration.

A different situation was described in [9], where rectangles were computed in optical images and pixels lying in several rectangles were given high likelihoods to be a car. Hence, this method has a disadvantage that one vehicle may belong to more than one stripe. In the case of vehicles parked densely, numerous stripes are formed from parking marks, neighboring vehicles and their shadows. To accelerate the computation in the future, it will be necessary to cluster these stripes and compute a mean stripe from each cluster. For now, these segments are given by pixels lists.

D7: [Stability of planarity], presupposes a search in planarity maps for small stable components which differ from background by a value corresponding to the typical car-size. The algorithm used for detection of hotspots is [20]. Similar to D3, this method relies on the fact that regions of constantly high planarity are surrounded by those where elevation differences are contributing to low planarity values. The disadvantages of this method are: a large set of parameters which are sometimes difficult to interpret and the necessity to deal with multiple components.

D8-D9: [Marker-based Watershed Algorithm] is a well known tool for image segmentation. Considering a grayscale image as a topological relief, the watershed algorithm finds catchment basins, where edges separate adjacent regions. We apply the watershed algorithm to the gradient norm of the elevation map, because similarly to D7, we expect the car centers to correspond to a local minimum of the gradient and the vehicle contours to correspond to the local maxima. Since in its original version, the watershed algorithm is sensitive to noise (as one segment is built for each local minimum), we propose to use markers, where search for minima must be performed. We differentiate between markers found in the orthophoto [D8] and in the [D9]. We search for salient components (either by their elevation or intensity) by means of techniques similar to top-hat filtering. The procedure based on watershed is rather fast and can be even accelerated by setting to infinity the planarity values for pixels in the complement of D2 (where no vehicles are expected), thus avoiding building the watershed components within this mask.

D10: [Mean-shift] was proposed by [16] for segmentation of 3D objects in point clouds. The basic idea is to find the local maximum of a density function in a non-parametric feature space. In our case, X,Y,Z coordinates are directly used as features. Using a Gaussian kernel function, the weighted mean of the density is computed for every 3D point within an iterative procedure. Hence, in a local 3D neighborhood, each point is "drawn" to the corresponding cluster center. The output is a dense segmentation of the input point cloud which again can be filtered according to NDVI and relative elevation.

III. Computational results

A. Evaluation strategy

In [21], three important goals were proposed for evaluation of segmentation for its suitability for vehicle detection: to maximize the similarity of areas between the binarized target image and the output segmentation as well as to minimize under-segmentation and over-segmentation, which is actually the same as to maximize the per-pixel values for precision and recall, respectively. Given results of detection, it is a straightforward to obtain both per-object and per-pixel *precision* (also *correctness*), *recall* (also *completeness*), and *overlap*, which are defined as: $TP/(TP+FP)$, $TP/(TP+FN)$ and respectively, $TP/(TP+FN+FP)$, where TP, FN and FP denote the number of true positive, false negative, and false positive, respectively, targets or pixels. The term overlap, which in Photogrammetry is sometimes referred to as *quality*, is closely related to $accuracy = (TP+TN)/(TP+FN+FP+TN)$. However, we prefer to use the overlap measure since is not diluted by true negatives, which are quite large numbers in our case.

In our efficient implementation of estimating per-pixel precision and recall for a target t and a detected segment a, pixels lists of t and a are sampled into a 2D histogram in such a way that the number of true positives $TP(t,a)$ is found in the corresponding bin. Then,

$$\text{precision}(t,a) = \frac{TP(t,a)}{A(a)}, \text{recall}(t,a) = \frac{TP(t,a)}{A(t)},$$
$$\text{overlap}(t,a) = \frac{TP(t,a)}{A(a)+A(t)-TP(t,a)}, \tag{1}$$

where $A(t), A(a)$ are the cardinalities of the target and of the region, respectively, and can be easily computed in advance. Note that $A(t) = TP(t,a) + FN(t,a)$ and $A(a) = TP(t,a) + FP(t,a)$. We decided to use overlap as our first measure:

$$\text{overlap}(t) = \arg\max_a (\text{overlap}(t,a)), \tag{2}$$

while the overall overlap is the average value of (2) over all targets. The remaining, per-object accuracy measure of [21] is dependent on the application. For some applications, the symmetric measure is rather suitable, while for example for many rescue missions, a detector with one missed target should be penalized much more than a detector with one false alarm. Because of this and because a bulk of false alarms can be discarded rapidly, we opt for assessing the per-object recall (completeness) only. We denote by

$$\text{per-obj. recall} = A(\{t|\text{overlap}(t) > \delta\}) / A(\{t\}), \tag{3}$$

where $\delta \approx 0.25$ is a constant. Finally, sometimes the target was detected within more than one region. It can happen if a car cabin lies in a different region as the motor hood or if the whole car lies in two or more stripes each of which is given by pixel list. Hence, the number of clusters (alarms per target) and the total number of alarms are reported as well.

B. Datasets

The dataset Vaihingen, which is a small town in Southern Germany, is the well-known ISPRS benchmark test site on urban object detection. It includes a DSM and a digital orthophoto of a rather high resolution (10pix/m), computed by the method [14] from several 16 bit color infrared images. The whole data set covers an area of around 1km^2, but for testing purposes, a image fragment of 2500×2500 pixels was chosen. We show in Fig. 2, left, this fragment; in Fig. 1, we show a 3D view of the urban terrain reconstruction. It contains boundary representations of building models with textured roofs, obtained by [2], generic tree models, and the ground surface textured with the orthophoto, where vehicles are inpainted using the method [3]. The 188 masks around the vehicles were interactively selected in the orthophoto; however, 17 of them were not considered because these moving vehicles are not represented well in the DSM. Among the remaining 171 targets, 11 are partly occluded and therefore especially challenging.

The second dataset represents an urban area in Munich city, also reconstructed from several images by [14]. Here, many vehicles lie in shadowy areas along the streets, see Fig. 2, right. In the area $270 \times 350\text{m}^2$, 269 cars are used for evaluation.

C. Evaluation and discussion

We performed evaluation of all methods on the datasets Munich and Vaihingen, for both of which, an example area is shown in Figs. 3 and 4, respectively. Table I records the values for the three important evaluation criteria discussed in Sec. III-A. Furthermore, the distributions for the overlap measure for the dataset Vaihingen, see (2), are visualized in Fig. 5 to track how the number of found targets (equivalently, per-object recall) may change if δ in (3) is increased.

Fig. 2. Our datasets: Vaihingen (on the left) and Munich (on the right)

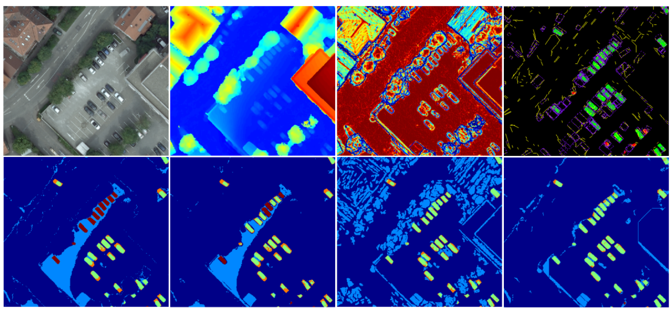

Fig. 3. Detailed view of evaluation of the data set Vaihingen. Top row, left to right: Orthophoto, NDSM, planarity map, and the ground truth mask with fully visible and partly occluded vehicles specified by red and green color, respectively. Additionally, line detected in the orthophoto (yellow) as well as the stripes formed from these lines (light blue, dashed) are shown. Bottom row: results of the detection by method D2, D3, D7, and D8 respectively. We show in light-blue, dark-red, orange and green the (per-object and per-pixel) false positives, the targets not detected – because of non-sufficient accuracy in (3) – the per-pixel false negatives and true positives, respectively.

One can see that the method based on tophat filtering D1 and the threshold-based methods D2 and D3 achieve a moderate performance for both data sets. The main advantage of these approaches is a low computational burden. Morphological operators are well understood and there are many efficient implementations allowing a calculation even in real-time applications. Additionally, it must be mentioned that the results for the methods D1-D3 are much better for high quality (for example airborne laser scanning) data. However, in DSMs obtained from images, separation of closely neighbored objects is hardly possible because the elevation values between them are often interpolated. Also, especially in the areas where computation of DTM contained systematic errors (a slightly elevated parking deck in Fig. 3), the method D2 merged all vehicles in the elevated areas. This situation can be partly corrected with the method D3, and even better with D7-D9, where regions of a high gradient value are used to separate the vehicles from each other. Watershed with markers and ISOL-based method could detect all, even partly occluded vehicles in (and around) the parking deck of Fig. 3. As a disadvantage, several cars become subdivided into more than one region, like the engine hood and the roof separated by the windshield. This leads to a moderate detection accuracy, though the number of not detected targets is decreased. The problem of interpolated elevation values is more severe in the dataset Munich. Hence, many planarity values are not reliable and lead to merging vehicles into groups, with a consequence of a lower performance of the methods D3, D7 and D8.

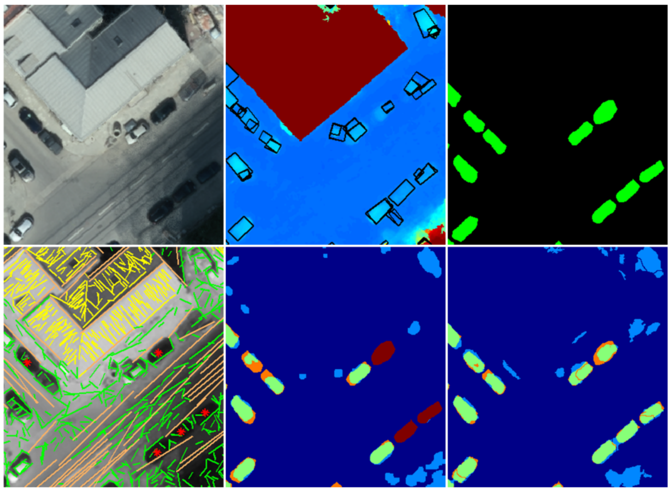

Fig. 4. Detailed view of evaluation of the data set Munich. Top row, left to right: Orthophoto, (truncated) NDSM with stripes formed from lines detected there, and the ground truth mask. In the bottom row, left: gray image of orthophoto with detected line segments. Only green segments are considered for stripes because orange segments are too long and yellow segments have a z-coordinate or NDVI value not plausible for vehicles. By red stars, we denote vehicles not detected by D4. Bottom, middle and right: results of detection by tophat filter (D1) and watershed with markups (D8).

The approach based on stripes in the orthophoto, D4, bears much potential since for both considered datasets, 92% and 82% of vehicles were contained in at least one stripe. Mostly, non-detected vehicles have an explanation in aliasing effects in orthophoto that itself is a result of dense 3D reconstruction. Due to these effects and also shadowy areas, line detection may fail in certain image regions. Another source of errors emerges often in case of cars forming a queue along the

TABLE I. VALUES FOR THE PER-OBJECT RECALL, FOR THE AVERAGE PER-PIXEL OVERLAP, FOR THE AVERAGE NUMBER OF CLUSTERS PER TARGET, AND THE TOTAL NUMBER OF SEGMENTS FOR DETECTION METHODS D1-D10. THE BEST VALUE IS SPECIFIED BY THE GREEN CELL COLOR WHILE THE WORST RESULT IS EMPHASIZED BY RED COLOR.

Vaihingen data	D1	D2	D3	D4	D4a	D4b	D5	D6	D6a	D7	D8	D9	D10
per-obj. recall, %	88	68	80	92	61	30	81	96	95	92	97	95	99
av. overlap, %	60	42	51	55	32	14	53	71	66	56	59	63	55
clusters per target	1.06	1.08	1.21	10.5	5.28	3.37	3.96	29.3	18	1.47	1.19	1.23	1.93
num. segments $\cdot 10^3$	1.8	7.1	3.6	8.7	0.33	1.3	3.1	21.4	11.8	14.1	1.6	1.6	9.6
Munich data	D1	D2	D3	D4	D4a	D4b	D5	D6	D6a	D7	D8	D9	D10
per-obj. recall, %	83	65	70	82	37	14	69	98	94	55	88	90	99
av. overlap, %	52	42	44	46	20	06	44	66	59	31	51	55	60
clusters per target	1.03	1.26	1.24	7.77	3.15	1.19	3.00	21.1	9.73	1.27	1.43	1.37	1.86
num. segments $\cdot 10^3$	2.4	5.6	3.1	7.8	0.14	0.98	2.2	18.7	10.1	4.1.0	2.0	2.2	18.0

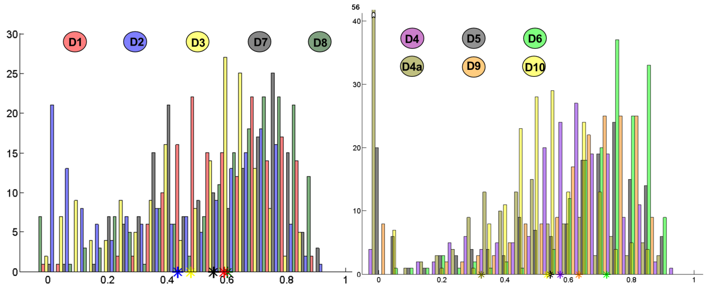

Fig. 5. Two histograms with overlap measures for the dataset Vaihingen. The different colors correspond to the methods specified above and described in Sec. II. Additionally, the mean values of overlap (second row of Table I) are specified by stars. The distribution for method D4b is not shown since its performance is even worse than that of D4a.

streets because long lines are then built and the computation of stripes fails. This can be seen in the shadowy areas of the data set Munich, see Fig. 4, bottom left, where all undetected targets are accompanied by a long orange line. Subdividing such long lines may solve these problems, which in the DSM-based approach (D5) are by far less acute.

Due to aliasing effects in the orthophoto, the approaches based on U+R- respectively on R-structures have turned out to be a less suitable detection tool for the considered data. Even after we replaced straight line segments of [18] with the Canny edges proposed by [9], there was no significant improvement in the performance. The final observation is that actually the best methods for vehicle detection D6 and D6a indeed strongly rely on the sensor data fusion, that is lines detected both in the orthophoto and in the DSM. The disadvantages of these methods are the computation time and the very high number of clusters.

Mean-shift is the only algorithm that can be directly calculated in airborne 3D point clouds and needs no initial DSM calculation. We obtained the highest recall in comparison to the other approaches, however, at the cost of creating a full segmentation result. Similar to functionality of D3, D8 and D9, most cars are divided into two clusters. The main disadvantage of this method is the extremely high computational effort.

D. Classification

Segmentation by one of the tools described in Sec. II and evaluated in previous section are the first stage in a standard three-step process chain. What follows – e. g. in the workflow proposed by [15] – is the feature extraction and two-class classification needed to reduce the false alarm rate and provide reliable results. For each segment (vehicle hypotheses), features are extracted. These can comprise 1) straight-forward region properties (such as area and eccentricity) which are easy to compute and are often very discriminative, 2) orientation histograms, Haar-like features, local binary patterns, etc. 3) features arising from those used for hypothesis identification (e. g. intersection of segments with lines and stripes stemming from the orthophoto and DSM), and finally 4) other features exploiting properties of the car in all available data (assessment of the region by elevations of its pixels, their planarity values, NDVI values, distribution of colors, number of empty bins in an equally sampled histogram, and many others).

To create labels for training data, the segments are compared with the ground truth mask using the per-pixel overlap measure described above. Segments with overlap larger than 0.25 are used as training data for the class car, segments with overlap 0 (no pixel in a car) are used for background training data. The testing is performed on a data set spatially separated from the training data. All segments in this area are classified with a state-of-the-art tool into background or car. We applied a random forest classifier [22] to the normalized (zero-mean, standard deviation one) features. For the numeric evaluation, precision and recall per object are used. To keep the section short, only the methods ”stripes” (D6) and watershed (D8) are tested for the dataset Vaihingen since these methods obtained the best recall value with a relatively low – however different – number of segments. After classification, precision and recall are 0.94 resp. 0.88, for stripes and 0.89 resp. 0.87 for watershed. This shows that – despite the somewhat smaller percentage of found targets and the high number of overall segments – the final detection result is more successful for stripes, due to a far better mean accuracy.

IV. CONCLUSIONS AND OUTLOOK

The goal of this paper was to identify useful algorithms for segmentation of vehicles for a following object-based classification and to analyze their performance on two datasets of different quality. Several algorithms applied to elevation and image data (orthophotos) as well as to the 3D point clouds were tested, starting with a simple thresholding over identifying stripes, regions of stable planarity (watershed transformation) as well as point cloud segmentation by the mean-shift method. All methods are fully-automatic, they are – with an exception of the procedure based on mean shift – rather fast, and require no training data. One could argue that since in the later stage of the algorithm proposed in Sec. III-D, we *do* need these training data. Hence, in future work, it will be indispensable to compare the detection results produced in this contribution with those induced by implicit methods.

For the most non-trivial and promising methods considered, plenty of work remains to be done: Subdividing long lines and reducing the number of clusters for the stripes-based methods as well as meaningful merging of hypotheses based on watershed and mean-shift. For the latter method, the computation time needed to create segmentation represents the main drawback. With the exception of this method, we could see a dramatic increase of accuracy results by combining different kinds of sensor data. For example, straight lines detected in images and in DSMs often complement each other as one can see in violet and in green histograms in Fig. 5, right.

In this work, detection of vehicle is performed with the same data which is typically used for semantic reconstruction of urban terrain. This means that this task and the modules for detection and reconstruction of other instances (buildings, roads, etc.) can strongly benefit from each other. The application of inpainting (see Fig. 1), whose potential has to be investigated yet, was one example of how to improve the authenticity of the orthophoto. In the future, integration of car detection module into an urban terrain reconstruction pipeline should be considered as well.

ACKNOWLEDGEMENT

The Vaihingen dataset was generated by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF), [23], see http://www.ifp.unistuttgart.de/dgpf/DKEP-Allg.html. In particular, DSM of the datasets Vaihingen and Munich were obtained using the SURE software [14] and provided by Norbert Haala and Mathias Rothermel, to whom the authors express their deep gratitude.

REFERENCES

[1] F. Leberl, H. Bischof, H. Grabner, and S. Kluckner, "Recognizing cars in aerial imagery to improve orthophotos," in *Proc. of the 15th annual ACM international symposium on Advances in geographic information systems*. ACM, 2007, p. 2.

[2] D. Bulatov, G. Häufel, J. Meidow, M. Pohl, P. Solbrig, and P. Wernerus, "Context-based automatic reconstruction and texturing of 3D urban terrain for quick-response tasks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 93, pp. 157 – 170, 2014.

[3] C.-B. Schönlieb and A. Bertozzi, "Unconditionally stable schemes for higher order inpainting," *Communications in Mathematical Sciences*, vol. 9, no. 2, pp. 413–457, 2011.

[4] X. Shi, H. Ling, E. Blasch, and W. Hu, "Context-driven moving vehicle detection in wide area motion imagery," in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 2512–2515.

[5] H. Grabner, T. T. Nguyen, B. Gruber, and H. Bischof, "On-line boosting-based car detection from aerial images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 63, no. 3, pp. 382–396, 2008.

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

[7] 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.

[8] S. Hinz and U. Stilla, "Car detection in aerial thermal images by local and global evidence accumulation," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 308–315, 2006.

[9] 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.

[10] T. Zhao and R. Nevatia, "Car detection in low resolution aerial images," *Image and Vision Computing*, vol. 21, no. 8, pp. 693–703, 2003.

[11] Z. Zheng, G. Zhou, Y. Wang, Y. Liu, X. Li, X. Wang, and L. Jiang, "A novel vehicle detection method with high resolution highway aerial image," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 6, pp. 2338–2343, 2013.

[12] B. Yang, P. Sharma, and R. Nevatia, "Vehicle detection from low quality aerial lidar data," in *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*. IEEE, 2011, pp. 541–548.

[13] 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. 6, no. 6, pp. 2327–2337, 2013.

[14] M. Rothermel, K. Wenzel, D. Fritsch, and N. Haala, "Sure: Photogrammetric surface reconstruction from imagery," in *Proceedings LC3D Workshop, Berlin*, 2012, pp. 1–9.

[15] H. Schilling, D. Bulatov, and W. Middelmann, "Object-based detection of vehicles in airborne data," in *SPIE Remote Sensing*. Intern. Society for Optics and Photonics, 2015, pp. 9643–420U.

[16] W. Yao, S. Hinz, and U. Stilla, "Extraction and motion estimation of vehicles in single-pass airborne LiDAR data towards urban traffic analysis," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 260 – 271, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271610000961

[17] P. Soille, *Morphologische Bildverarbeitung*. Springer, 1998.

[18] J. Burns, A. Hanson, and E. Riseman, "Extracting straight lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 4, pp. 425–455, 1986.

[19] U. Soergel, E. Cadario, A. Thiele, and U. Thoennessen, "Feature extraction and visualization of bridges over water from high-resolution insar data and one orthophoto," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 1, no. 2, pp. 147–153, 2008.

[20] C. Anderer, U. Thönnessen, M. F. Carlsohn, and A. Klonz, "Ein Bildsegmentierer für die echzeitnahe Verarbeitung," *DAGM-Symposium*, pp. 380–384, 1989.

[21] A. C. Holt, E. Y. Seto, T. Rivard, and P. Gong, "Object-based detection and classification of vehicles from high-resolution aerial photography," *Photogrammetric Engineering & Remote Sensing*, vol. 75, no. 7, pp. 871–880, 2009.

[22] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[23] M. Cramer, "The DGPF test on digital aerial camera evaluation – overview and test design," *Photogrammetrie, Fernerkundung, Geoinformation*, vol. 2, pp. 73–82, 2010.