

An Improved Car Detection using Street Layer Extraction

Georg Pacher, Stefan Kluckner, and Horst Bischof

ICG,

University of Technology, Graz, Austria
pacher, kluckner, bischof@icg.tugraz.at

Abstract Automatic 3D modeling of urban environments has recently become a hot research topics since location aware applications (e.g. Virtual Earth, Google Earth) on the Internet aim for detailed models of the earth. In large scale 3D city models, moving objects, such as cars usually bear errors and distract the automatic reconstruction process. Therefore it is desirable to detect and remove these objects. This work introduces a strategy for extracting a street layer by using 2.5D height data and color information and demonstrates how this layer can be applied to improve the car detection process on high resolution aerial images. We detect street related objects, such as zebra crossings, to extract control points of a street layer, on which an occurrence of cars is accurately defined. These control points help to generate a Digital Terrain Model (DTM) and a color model for the street network. By introducing this information in the car detection process, the false positive rate can be decreased considerably. The extracted street layer is evaluated on available classification results and hand labeled ground truths. Furthermore, we compare the achieved performances of our approach to state-of-the-art car detection results on aerial images.

1 Introduction

Efficient and robust object detection in large scale aerial imagery has been drawing the attention of the vision community for the last years. In particular, detection of cars in aerial images has been an active research topic as a result of many civil and military applications e.g. [10, 21, 27]. In automatic 3D city modeling the removal of irrelevant objects directly from aerial images benefits a suitable visualization. For the detection process it is crucial to acquire high detection and low false positive rates, because the removal of false detected objects can also remove important data for the overall visualization process.

According to its remarkable detection performance, the boosting approach introduced by Viola and Jones [25] has been used to solve various object detection tasks (pedestrians, faces, vehicles, etc.). Nguyen et al. [16] applied an on-line version of AdaBoost [9] to the car detection problem in aerial images. A supervisor labels positive and negative samples for the on-line training of the appearance model based classifier. An evaluation of the detection results on hand labeled ground truth data gives a recall rate of approxi-



Figure 1: A car detection result on a small scene including some false positive detections.

mately 80% and a false positive rate of 10%. Figure 1 shows a car detection result on a sub-image including several false positive detections on rooftops.

The aim of our work is the reduction of false positive detections by integrating the available 2.5D height information. This additional information motivates the idea of extracting a street layer, and removing all detections on non street objects. Due to noisy range images and high reconstruction errors of moving cars, a direct estimation of the street layer by using the positions and height levels of detected cars does not give robust results.

A large variety of approaches for street extraction exists, using different data and algorithms. An (almost historic) overview is given in [8]. In [20] a Digital Elevation Model (DEM) is used to find and verify road boundaries in an image, assuming a regular grid form of urban roads. Several approaches use contours to track streets in aerial or satellite imagery [2, 18]. In [26] a pixel-wise classification and a reconstructed surface model are combined to perform a layer extraction for buildings, streets, vegetation and water surfaces on high resolution aerial images.

In our approach we introduce the idea of recognizing objects which are directly related to streets, have a distinctive pattern and are stationary. Se [22] and Uddin et al. [24] proposed methods to extract zebra crossings as road features in street side images. Here we propose a strategy to detect and combine zebra crossing-like structures in aerial images resulting in the generation of a street layer. Due to their size and unique appearance, zebra crossings are the most appropriate stationary street related objects. Detected zebra crossing can additionally be used to find further street related ob-

jects, or for verification of an extracted road network.

Gabor filtering techniques are widely used in unsupervised 2D texture analysis and object detection, e.g. [11] and [13]. The matched filter design gives suitable results in various fields of computer vision [17, 23]. These filters match the pattern exactly and therefore the signal-to-noise ratio gets maximized. In our approach we design a matched filter which is directly related to the pattern of a zebra crossing. After filtering with the matched filter, the resulting responses give the initial starting points for finding minimum cost linkages between them, considering the height values. Convenient graph-based segmentation strategies using a computation of shortest paths between a starting and an ending vertex are described in [7] and [15]. The *Dijkstra* algorithm [5] finds the single shortest path with lowest costs depending on the graph weights. In this work we assign the weights depending on color and height information.

For the extraction of the street layer including separated areas such as courtyards, we propose the use of a so-called Bald Earth model or Digital Terrain Model (DTM). In order to retain generality, hills and slopes should be modeled correctly. The DTM does not include the modeling of man made structures, such as houses. Therefore, a surface is interpolated using points from the range image which definitely are part of the street layer. A Thin-Plate Spline (TPS) [3] generates a globally smooth surface, which is fitted to the computed control points in a least-squares manner. Other state-of-the-art approaches for surface fitting to an non-regular positioned set of 3D points are B-splines and especially NURBS [19].

Given a DTM and the corresponding 2.5D range image, areas which lie on the ground level are easily identified. However, cars should only be located on streets and not on vegetation areas or even on water bodies. Hence, a further segmentation or terrain classification is needed. Starting from the extracted zebra crossings, a model of the color distribution for streets can be estimated. Then, this model can be used in a simplified segmentation approach similar to [6] using the Bhattacharyya distance [1].

This paper is organized as follows. First we review the generation of the required image data and explain briefly the applied method for car detection (Section 2). In Section 3 our proposed strategy for the street layer extraction is explained in detail. Section 4 demonstrates the results of our work, including the evaluation of the extracted street layer and the performance of the car detection on two different data sets. Finally, we conclude our work and give ideas for future work in Section 5.

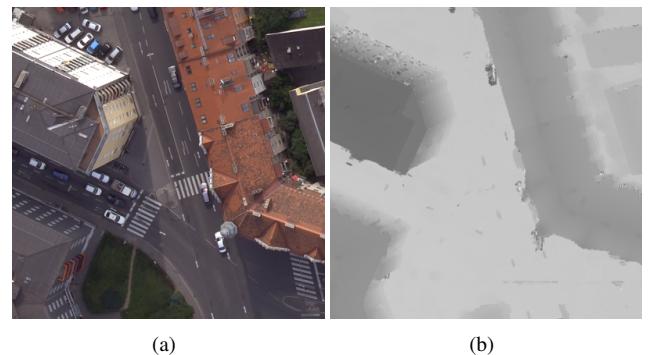
2 Preliminaries

Before introducing the details of our work, we give a short overview of the available data sources and the car detection framework, used to detect the cars in aerial images.

2.1 2.5D Height Model from Aerial Images

The 2.5D range data is derived from the high resolution aerial images taken by the *UltraCam_D* camera from *Microsoft Photogrammetry*. Each aerial image has a size of 11500×7500 pixels with a ground sampling distance of

approximately 8cm . In order to enable robust and fully automatic processing of the data, high image redundancy is ensured by capturing the images with high overlap. The exterior orientation of the images is retrieved by an automatic process as described in [26]. An area based matching algorithm produces a dense range image for each input image by using the camera orientations. The range images are computed from three neighboring input images using a plane sweeping approach [4]. The plane sweeping applies the normalized cross correlation as similarity measure and produces a 3D depth space which contains the depth hypotheses and their associated correlation values. The final range image is computed using a semi-global optimization approach proposed in [12]. Figure 2 depicts a typical sub-image of a scene and the corresponding range image.



(a)

(b)

Figure 2: (a) A typical sub-image of a scene in an aerial image and (b) the corresponding 2.5D range image providing the height information.

2.2 Car Detection

Similar to [16], we apply an on-line boosting for feature selection technique [9] to detect cars in high resolution aerial images. However, any other car detection approach can be used. In general, boosting combines several weak classifiers to form a strong one. The on-line selection of significant weak classifiers is performed on so-called selectors. Each selector consists of a set of weak classifiers and can be seen as an auxiliary classifier that switches between the weak learners. Each training sample trains all weak classifiers and the selector with the lowest error on the samples is then selected. Each weak learner directly corresponds to the response of a single extracted feature. We apply Haar-like features [25] and orientation histograms [14] to the detection process due to fast feature response computation by using integral images. After training the classifier is applied to a test image by exhaustively passing rotated sub-images according to an angle step of 15 degrees and a search window with 35×70 pixels.

In contrast to [16], where a mean-shift clustering is proposed to find the best detection results, we apply an orientation dependent non-maximum suppression right after the detection process. In our approach we train the classifier by using an on-line learning hand labeling procedure. After training, the classifier can be applied to various aerial images.

3 Street Layer Extraction

The main focus of this work is the extraction of a street layer and the use of this information to improve a car detection process. The street layer defines regions where cars can definitely appear. Therefore it consists of all flat areas on the ground which are not covered by vegetation or water. For a robust extraction of a street layer, reliable start points have to be found. Since most parts of the street layer are connected, a connectivity measure can be used as verification for the start points. Our proposed street layer extraction is divided into three subsequent parts.

First, initial response areas are computed using an oriented matched filter to detect zebra crossings as depicted in Figure 3. The detection of zebra crossings in aerial images is appropriate to estimate the street layer, since zebra crossings are directly related to the street level and have a unique discriminative pattern. Due to possible similar structures, the detected zebra crossings cannot be directly used for the street layer extraction. Therefore wrong detections have to be identified. This can be done by a grouping technique, as streets should connect most zebra crossings with each other. Therefore we perform a graph-based shortest path finding algorithm on the 2.5D range data and the color information. This step connects the extracted responses with minimum cost. The measurement of connectivity is then used to remove detections with low connectivity (e.g. on roofs). Besides the effect of grouping, this graph-based algorithm serves as a control point generator, where the color distribution for streets is estimated.

To improve the control point generation for closed streets of houses, we exploit the image redundancy and apply our strategy to three overlapping images. Since camera positions and orientations as well as the height for every pixel are known, the separately computed control points of each image can be merged in a common world coordinate system. These control points are then used to interpolate a DTM by a TPS surface for the whole scene. This height level and the additional color distributions estimate the street layer for each image by back projection. Furthermore the DTM can be used to remove reconstruction errors (e.g. from moving cars) and undesired objects on the street layer from the 2.5D height model. The following sections explain the successive extraction of the street layer.

3.1 Matched Filter Design and Response Areas

For our work, we assume that zebra crossings have a typically similar appearance with a regular striped structure and a frequent occurrence in urban environments. In Figure 3 samples of various extracted zebra crossings are shown.

Apart from the differing contrast we estimate a width of our artificial matched filter from these samples experimentally. The squared matched filter mask is based on the symmetric cos function and depends on the width w and the height h . Equation 1 denotes the computation of a horizontal oriented matched filter mask with $x, y = -k \dots k$.

$$f(x, y) = \begin{cases} \cos\left(\frac{x\pi}{w}\right), & |x| \leq 2w, |y| \leq h \\ 0, & \text{else} \end{cases} \quad (1)$$

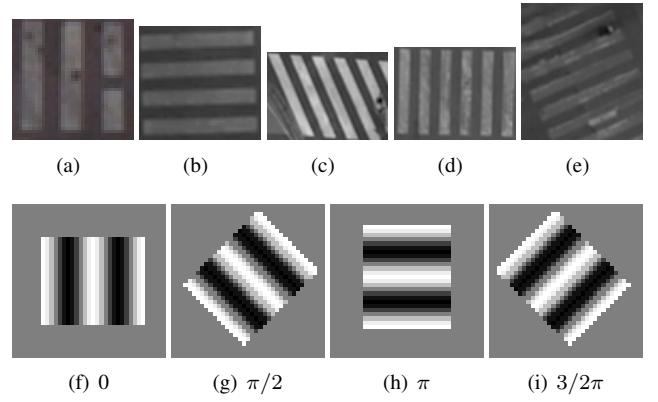


Figure 3: (a)-(e) Samples of several extracted zebra crossings and (f)-(i) matched filter masks for various given orientations.

In aerial images the problem of color fading decreases the contrast of the appearance of zebra crossings drastically. Here, an adaptive thresholding strategy performs a local contrast enhancement. In order to obtain responses for various orientations, the filter mask is rotated according to a given angle α (see Figure 3). Note, that these computations are done using integer values, due to computation time.

Then, the designed matched filter is applied to the pre-processed image using the convolution step $M(\alpha) = M * f(\alpha)$. Moreover, as defined in Equation 2, the achieved responses over all orientations get maximized.

$$M_{response} = \arg \max_{\alpha} M(\alpha) \quad (2)$$

A global segmentation with a threshold

$$t_g = \frac{2}{3} \arg \max_{x,y} M_{response}(x, y) \quad (3)$$

elaborates the maximum responses and results in a binary mask including many unlabeled areas. A fast contour based region labeling combines connected pixels to form summarized response regions R with assigned mean orientations, which represent the possible driving directions. Additionally, an investigation of area characteristics provides a selection of representative extracted zebra crossings for the next step of the street layer extraction. Figure 4 shows the results of the zebra crossing extraction for a fixed width of $w = 0.52m$ and $h = 1.2m$.

3.2 Graph-based Grouping using Color Models and Height Values

It is obvious that extracted response regions can also include irrelevant zebra crossing-like patterns. Therefore a graph-based grouping technique can perform outlier removal by including context information, such as height data and a color similarity measurement. We apply the idea of graph-based border detection to link the N extracted response regions R . We define two requirements for a valid linkage L between two extracted response regions R_i and R_j with $i, j = 1 \dots N, i \neq j$. First, L may not include discontinuities concerning their involved height values. Additionally, the color values along the connecting path must correspond to a specific street related color distribution. These

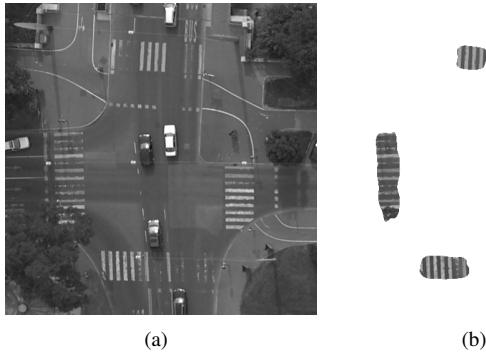


Figure 4: Extracted response regions using matched filters, (a) original gray value image, (b) found zebra crossing patterns after area investigation.

two requirements constrain the construction of a weighted bi-directional graph $G = (V, E)$. The set of vertices V is 4-neighborhood connected according to a regular grid covering the overall image as shown in Figure 5. In order to reduce computation time, we fix the grid spacing to 1.5 meters for all evaluated data sets.

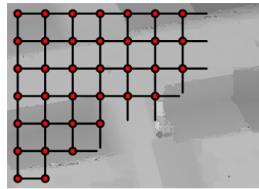


Figure 5: The constructed regular graph covering a sub-image of the range data.

The weights of the edges E get assigned as follows. To eliminate the height discontinuities, the weight of each edge is set to infinity if the height value difference $|H_n - H_{n+1}|$ between two neighboring vertices V_n and V_{n+1} is larger than $1.5m$ (slope of 45 degrees). If the slope of this edge is below that, we introduce color models to improve our connection finding strategy. The color distribution along streets is estimated using sample patches in both driving directions from each extracted region. This guarantees image patches which are part of the street, but not affected by the zebra crossings. In order to have an isotropic color feature space, the image is converted to the three dimensional CIE Luv color space. Assuming normal distributed color values, the distribution of extracted color values can be easily modeled by using the estimated mean and covariance matrix. For the mean color model estimation of streets $\bar{N}(\bar{\mu}, \bar{\Sigma})$ coarse color outliers (e.g. a colored car on the particularly extracted spot) are removed using an alpha-trimming technique.

The Bhattacharyya distance is applied as a similarity measure between Normal distributions $N_n = \{\mu_n, \Sigma_n\}$ and $N_m = \{\mu_m, \Sigma_m\}$ similar to [6]. The discretized Bhattacharyya distance β is given in Equation 4 and denotes a measure for statistical separability of probability distributions. Therefore, it gives a quantitative statement whether the color distributions are likely to be the same or not.

$$\beta(N_n, N_m) = \frac{1}{2} \ln \frac{|\frac{\Sigma_n + \Sigma_m}{2}|}{\sqrt{|\Sigma_n| |\Sigma_m|}} + \frac{1}{8} (\mu_m - \mu_n)^t \left[\frac{\Sigma_n + \Sigma_m}{2} \right]^{-1} (\mu_m - \mu_n). \quad (4)$$

By applying this distance measure patch-wise on each pixel of a test image, using the estimated mean color model \bar{N} , the image can be roughly segmented as shown in Figure 6. To avoid the extensive computation of distances to all pixels, we only estimate color models around the position of two neighboring vertices of the graph G .

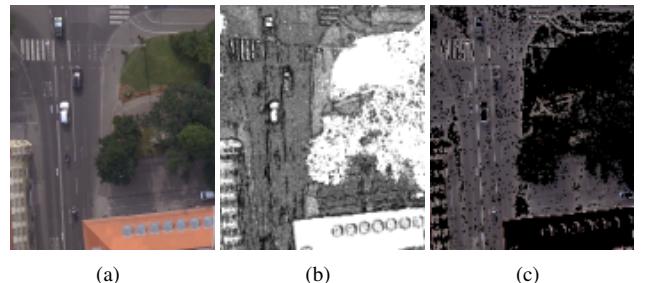


Figure 6: (a) Original image, (b) Bhattacharyya distances to the mean color model, obtained through knowledge about zebra crossings and (c) a rough segmentation with assigned original colors.

The overall weight assignment for neighboring vertices V_n and V_{n+1} including the color distributions N_n, N_{n+1} and the height difference is summarized in Equation 5.

$$\omega_{n,n+1} = \begin{cases} \infty, & \text{if } |(H_n - H_{n+1})| \geq 1.5m, \\ \max(\beta(\bar{N}, N_n), \beta(\bar{N}, N_{n+1})), & \text{otherwise} \end{cases} \quad (5)$$

Then, the *Dijkstra* algorithm [5] finds a single shortest path with lowest costs depending on the assigned weights. A linkage L_{ij} , connecting the response regions i and j , is determined with respect to the maximal costs C_{ij} for each computed path. By tracing all low-cost linkages between the extracted response regions, we establish a connectivity measurement P_i for each region R_i as defined in Equation 6.

$$P_i = \frac{1}{N-1} \sum_{k=1, i \neq j}^m L_{ik} \quad (6)$$

This connectivity measurement offers important information for a confident utilization in the following DTM interpolation step. A high connectivity P_i implies that R_i connects to many other extracted regions, while regions with low connectivity are isolated detections. Traversed vertices along linking paths between regions of high connectivity can then be used as additional points forming the overall set of control points Q . In Figure 7 the cropped sub-image includes the overlaid computed minimum cost paths.

To obtain an enlarged coverage of the street layer extraction we exploit the available image redundancy. Independently, the matched filter responses and the estimated connecting paths are extracted from three neighboring images.



Figure 7: Sample image of the data set with outlines of detected zebra crossings (blue) and the corresponding paths (green lines) connecting detections with minimum cost. The numbers depict the connectivity measure P_i for each found zebra crossing. The yellow dots represent the positions where the color samples for the color modeling are selected.

The collected paths provide an increased number of covering control points and facilitate the loop closing of streets. Since the camera orientations of each image are known, the control points are transformed into a common world coordinate system using the height data. As the next step, the DTM interpolation based street layer extraction step is performed on this world coordinate system.

3.3 DTM Interpolation and Street Layer Extraction

The extraction and grouping of the detected zebra crossings in overlapping images improves the description of streets characteristics using additional representative control points. Yet, separated areas like court yards, or streets without any zebra-crossing, might be left out. To find areas on similar ground level a DTM can be estimated and compared to the available 2.5D range data. A DTM specifies the height of the ground without buildings or trees. This information can further be used for estimation of building heights and is an useful byproduct of our approach. It is generated using the traversed vertices of the *Dijkstra* algorithm as explained in Section 3.2. The TPS interpolation [3] offers a closed-form solution for surface interpolation from a set of control points. The interpolated surface $\text{DTM}(x, y)$, which passes through the set of M control points Q has the form:

$$\text{DTM}(x, y) = a_0 + a_1 x + a_2 y + \sum_{i=1}^M w_i U(|Q_i - (x, y)|) \quad (7)$$

where U denotes the base function $U(r) = r^2 \log(r)$, (a_0, a_1, a_2) and w_i are the coefficients of the spline surface. Due to the solving of an $M \times M$ linear equation system, the interpolation using TPS is limited in the number of control points. Our graph-based grouping approach for three overlapping images produces a number of more than tens of thousands control points. Therefore, we apply a simple uniform random sampling to achieve a reduction of points, which can be efficiently solved using TPS interpolation. Figure 8 depicts the sampled control points and the interpolated DTM.

Using the spline surface, the street layer can be easily extracted by considering areas in the 2.5D range data,

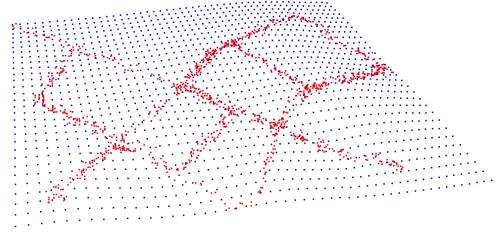


Figure 8: An illustration of the TPS interpolation result with inflated height values in z-direction. The computed grid points are indicated as blue and the involved control points as red dots, respectively.

which are at approximately the same height level, as the estimated DTM. These difference computations result in a rough ground layer segmentation. Non-street related regions, such as grassland or water bodies located at the same height level are still included. Therefore, in a post processing step, the estimated mean color model (see Section 3.2) is used to refine the segmentation result again by evaluation of the Bhattacharyya distances to the mean color model. Cars might still not be contained in this segmentation result, since they are slightly higher than streets and may be colored significantly different. Hence a morphological post processing is performed to add car-sized isolated areas to the street layer. This street layer mask is then used in the car detection process, to remove detections which do not lie on the street layer.

4 Experiments and Results

In this section we first evaluate the results of our proposed street layer extraction method. Then we show results on how the street layer improves a state-of-the-art car detection process. Our approach is evaluated on high resolution aerial images of the city of *Graz*. Two test regions have been chosen to evaluate the ability of our approach to adapt to a given scene. One region lies in the city center including the river *Mur*, while the other shows a more suburban area including green areas. The data set *Graz 1* has a size of 4000×4000 pixels, including 330 cars. In the test data set *Graz 2* there

	<i>Graz 1</i>		<i>Graz 2</i>	
	Our approach	[26]	Our approach	[26]
FPR	15.07 %	1.97 %	7.21%	2.60 %
TPR	96.47 %	81.52 %	99.24 %	76.21 %

Table 1: Comparison of our approach and [26] to the hand labeled ground truth for the two data sets *Graz 1* and *Graz 2*. Note that for our approach a high TPR is more important than a low FPR.

are 519 cars on a size of 6000×4500 pixels. The ground truth sets for cars include the location and the orientation of each car. We marked a car as positive for the ground truth, if more than 50% of its area is visible. With the known ground sampling distance of 8cm both data sets cover an area with a radius of approximately 200m . Additionally, we build a hand labeled ground truth of each test image including the street related areas.

4.1 Extracted Street Layer Evaluation

Our proposed street layer extraction method is compared to a hand labeled ground truth and to the classification results refined with height information as proposed in [26]. In Figure 9 the results of our approach are compared to the hand labeled ground truth qualitatively. It can be seen that parts of the river are classified as street, which is caused by the similarity of the color distribution of the river with dark parts of the streets. Furthermore, the dense matching has problems in reconstructing the river, resulting in scatter in the height data. To overcome these problems, either a different segmentation strategy would be needed, or the use of further information like near infrared imagery as in [26]. For a quantitative evaluation of this result, the false positive rate (FPR) and the true positive rate (TPR) are used. The FPR measures the areas which are falsely classified as street layer, while the TPR measures the right classified areas. In Table 1 our approach and the one given in [26] are compared with the hand labeled ground truth. It can be seen, that our approach performs better regarding the TPR, which states that we find more areas of the real street layer than [26]. On the other hand we perform worse with respect to the FPR. The reasons for that are the problems with the water surface of the river in data set *Graz 1*, and dark green sub urban courtyards in data set *Graz 2*. Nevertheless, our goal is a mask, which can be used to remove false positive car detections. Therefore, a high TPR is desirable such that detections on streets are not thrown away. It is of minor importance that all false detections are discarded, thus we are not so strict for the FPR. Note that the zebra crossing detection finds approximately 90% of the zebra crossings. It fails to find a zebra crossing if the markings on the street are worn out strongly. Wrong detections only occur on repetitive patterns on roofs, which are rejected by the grouping algorithm.

4.2 Improving the Car Detection

In this section we show quantitative results of our approach. The car detector is evaluated on two sub-regions of the *Graz* data set. It was trained with 260 positive car samples and 521 negative updates in a first step, obtained from a non-overlapping training data set. For detection, a grid with 90%

overlapping patches is used. The extracted street layer mask is then used as a post processing to remove detections which do not lie on the street layer. Thus, the false positive rate of the car detector is reduced.

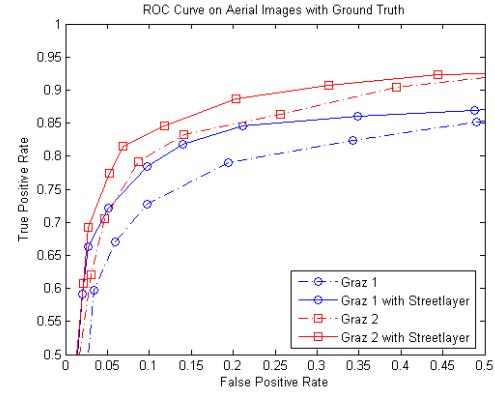


Figure 10: ROC curve for the data sets *Graz 1* and *Graz 2* of the car detection results with (solid lines) and without (dotted lines) using a street layer.

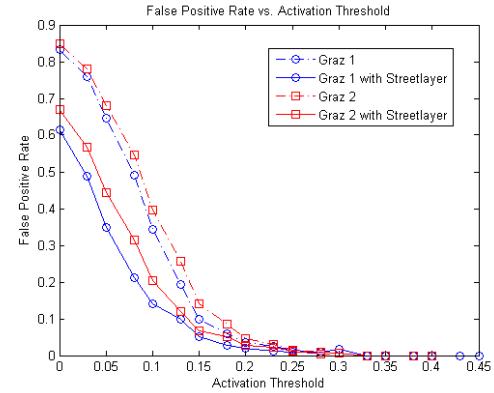


Figure 11: The false positive rate vs. the activation threshold of the car detector with (solid lines) and without (dotted lines) using a street layer.

Figure 10 demonstrates the detection results in terms of true and false positive rates. The results for both data sets with (solid lines) and without (dotted lines) using a street layer as context information are depicted. Figure 11 shows the improvement of the false positive rate for several activation thresholds of the car classifier. The activation threshold denotes the value of minimal confidence for the decision of a valid car detection. The evaluation shows that the false positive rate is decreased by a factor of 2 for both data sets. Figure 12 shows the visual results on a sub-image for the data set *Graz 2*. The yellow rectangles indicate the detected cars. The utilization of the extracted street layer eliminates the detected car like structures on rooftops or on the church roof successfully.

5 Conclusion

This paper proposed a strategy to reduce the number of false positive detections of a car detection process by introducing

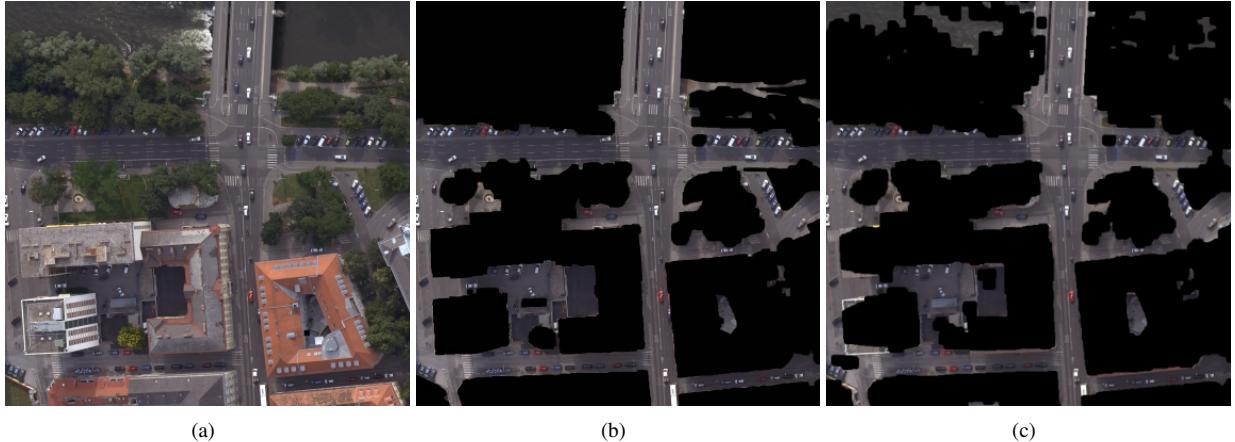


Figure 9: (a) shows the original image taken from data set *Graz 1*, (b) depicts the labeled ground truth, and (c) is the street layer mask obtained by our approach.

an extracted street layer as context information. We have shown how the detection of zebra crossing and available height information can be used to extract the layer, where cars can definitely appear. Additionally, a color model of the street facilitates the correct modeling of the ground. A TPS interpolation generates a smooth surface of the terrain. The comparison of the extracted street layer mask to classification results and hand labeled ground truth demonstrates the strength of our approach with more than 95% correct classified pixels. Applying the resulting street layer to the car detection process results in decrease by a factor of 2 of the false positive detection rate. Detections on roofs can successfully be removed.

Future work will address the problem of hilly terrain modeling. An adapted region growing algorithm on height and color information will result in better coverage of the control points for the surface interpolation. Furthermore, street color models using Gaussian Mixture Models instead of single mean and covariance estimation will give more robust results. Additionally, the global TPS interpolation can be replaced by a local surface interpolation method due to the computation of large data sets and the problem that streets might be sloped perpendicular to the driving direction. A local interpolation can be constraint to specific gradients between the extracted control points. Certainly, the evaluation of our approach has to be extended to data sets including hilly terrain.

Acknowledgments

This work has been supported by the APAFA Project No. 813397, financed by the Austrian Research Promotion Agency (<http://www.ffg.at>) and the Virtual Earth Academic Research Collaboration funded by Microsoft.

References

- [1] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematics Society* 35, pages 99–110, 1943.
- [2] M. Bicego, S. Dalfini, G. Vernazza, and V. Murino. Automatic road extraction from aerial images by probabilistic contour tracking. *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, 3:III-585–8 vol.2, 14–17 Sept. 2003.
- [3] F. L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(6):567–585, 1989.
- [4] R. T. Collins. A space-sweep approach to true multi-image matching. *Proceedings IEEE Conference Computer Vision and Pattern Recognition*, 00:358, 1996.
- [5] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.
- [6] M. Donoser and H. Bischof. Roi-seg: Unsupervised color segmentation by combining differently focused sub results. In *Proceedings IEEE Conference Computer Vision and Pattern Recognition*, 2007.
- [7] A. X. Falcão, J. K. Udupa, S. Samarasekera, S. Sharma, B. E. Hirsch, and R. A. Lotufo. User-steered image segmentation paradigms: Live-wire and live-lane. *Graphical Models and Image Processing*, 60(4):233–260, Jul 1998.
- [8] A. Fortier, D. Ziou, C. Armenaldis, and S. Wang. Survey of work on road extraction in aerial and satellite images, 1999.
- [9] H. Grabner and H. Bischof. On-line boosting and vision. In *Proceedings IEEE Conference Computer Vision and Pattern Recognition*, volume 1, pages 260–267, 2006.
- [10] S. Hinz. Detection and counting of cars in aerial images. In *International Conference on Image Processing*, volume 3, pages 997–1000, 2003.
- [11] A. Jain, N. Ratha, and S. Lakshmanan. Object detection using gabor filters. *Pattern Recognition*, 30:295–309, 1997.
- [12] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. *Proceedings of International Conference on Pattern Recognition*, pages (III):15–18, 2006.

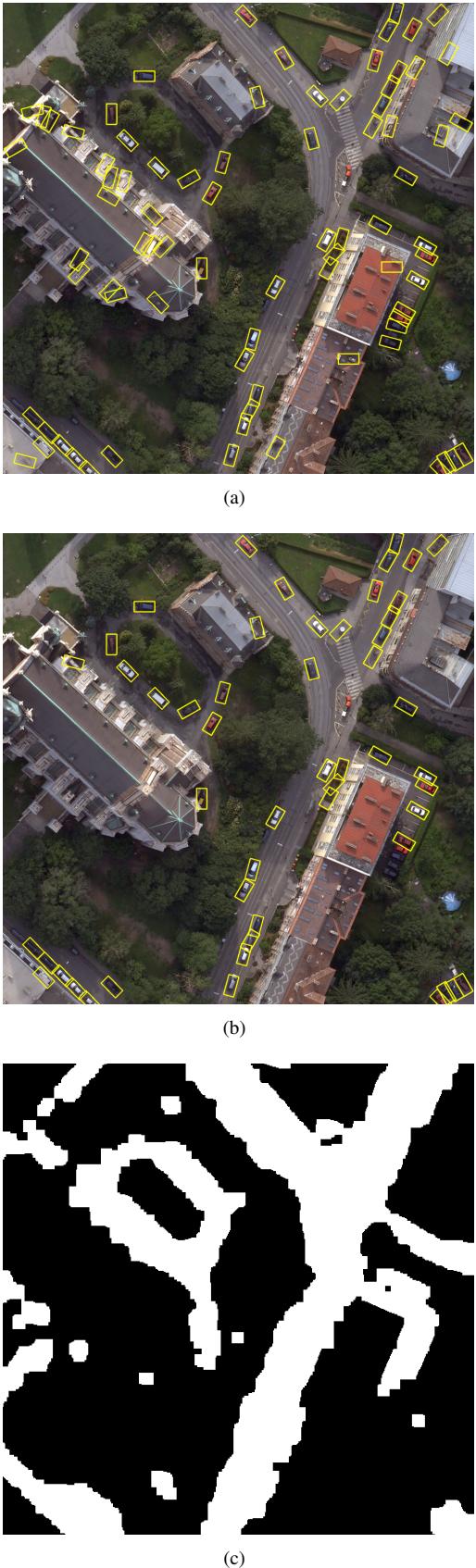


Figure 12: Visual detection results using (a) no street layer information, (b) with street layer and (c) the street layer mask.

- [13] V. Kyrki, J.-K. Kamarainen, and H. Kälviäinen. Simple gabor feature space for invariant object recognition. *Pattern Recognition Letters*, 25(3):311–318, 2004.
- [14] K. Levi and Y. Weiss. Learning object detection from a small number of examples: The importance of good features. In *Proceedings IEEE Conference Computer Vision and Pattern Recognition*, pages 53–60, 2004.
- [15] E. N. Mortensen and W. A. Barrett. Intelligent scissors for image composition. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 191–198, New York, NY, USA, 1995. ACM.
- [16] T. Nguyen, H. Grabner, B. Gruber, and H. Bischof. On-line boosting for car detection from aerial images. In *IEEE International Conference on Research, Innovation and Vision for the Future (RIVF07)*, pages 87–95, 2007.
- [17] L. O’Gorman and J.V. Nickerson. Matched filter design for fingerprint image enhancement. 2:916–919, 1988.
- [18] R. Peteri and T. Ranchin. Multiresolution snakes for urban road extraction from ikonos and quickbird images. In *23rd Annual Symposium Remote Sensing in Transition*, 2003.
- [19] L. Piegl and W. Tiller. *The NURBS book (2nd ed.)*. Springer-Verlag New York, Inc., New York, NY, USA, 1997.
- [20] K. Price. Road grid extraction and verification. *International Archives of Photogrammetry and Remote Sensing*, 3:101–106, 1999.
- [21] R. Ruskone, L. Guigues, S. Airault, and O. Jamet. Vehicle detection on aerial images: A structural approach. In *International Conference on Pattern Recognition*, volume 3, pages 900–904, 1996.
- [22] S. Se. Zebra-crossing detection for the partially sighted. In *Proceedings IEEE Conference Computer Vision and Pattern Recognition*, volume 2, pages 211–217, 2000.
- [23] M. Sofka and C. V. Stewart. Retinal vessel extraction using multiscale matched filters, confidence and edge measures. *IEEE Transactions on Medical Imaging*, 25(12):1531–1546, Dec. 2006.
- [24] M. S. Uddin and T. Shioyama. Bipolarity and projective invariant-based zebra-crossing detection for the visually impaired. *Proceedings IEEE Conference Computer Vision and Pattern Recognition*, 0:22, 2005.
- [25] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings IEEE Conference Computer Vision and Pattern Recognition*, volume I, pages 511–518, 2001.
- [26] L. Zebedin, A. Klaus, B. Gruber-Geymayer, and K. Karner. Towards 3d map generation from digital aerial images. *International Journal of Photogrammetry and Remote Sensing*, 60:413–427, Sept. 2006.
- [27] T. Zhao and R. Nevatia. Car detection in low resolution aerial image. In *International Conference on Computer Vision*, 2001.