Detection of Stationary Vehicles in Airborne Decimeter Resolution SAR Intensity Images using Morphological Attribute Filters

OLIVER MAKSYMIUK [1], ANDREAS BRENNER [2] & UWE STILLA [1]

Abstract: The increasing resolution of modern SAR sensors leads to new approaches for object detection in SAR images. Objects appear not just as blobs but rather show a certain structure which can be exploited by advanced approaches already established in image processing. But an increase in detail comes at the cost of an increased complexity. In this work we use Mathematical Morphology (MM) to detect stationary vehicles in airborne decimeter resolution SAR intensity images. Mathematical Morphology has become popular because of its proven utility and sound mathematical foundations. Here we use the concept of connected components to simplify the image and to describe parts of the vehicles. The detection is a subsequent process. It is shown that a robust detection based on the explicit description of the vehicle appearance is possible.

Zusammenfassung: Bedingt durch die hohe Auflösung moderner SAR Sensoren ergeben sich viele neue Ansätze für die Detektion von Objekten. Es ist nunmehr möglich abseits von statistischen Methoden auch leistungsfähige Verfahren aus der Bildverarbeitung zu verwenden. In dieser Arbeit werden Methoden aus der mathematischen Morphologie verwendet um in SAR Intensitätsbildern stationäre Fahrzeuge zu detektieren. Dabei wird vor allem von Attributfiltern Gebrauch gemacht, die basierend auf den verbundenen Grauwertkomponenten eine Bildtransformation durchführen. Das Ergebnis ist eine Objektbeschreibung in der eine Komposition relevanter Fahrzeugmerkmale zur finalen Detektion genutzt wird. Es wird gezeigt, dass sich mittels dieser Zerlegung und anschließender Korrelation ein robustes Detektionssystem ergibt.

1 Introduction

The increase of the overall traffic volume imposes new challenges on traffic research and planning. Therefore, airborne and spaceborne imaging systems that provide synoptic views of complex traffic situations and the associated context have become interesting (STILLA et al., 2004) and can be considered as complementary systems to road sensors. In recent years, many approaches for detection of moving stationary and moving vehicles using data from different sensors have been developed, e.g. from optical satellite imagery (LEITLOFF et al., 2010), aerial image sequences (LENHART et al., 2008), aerial thermal videos (KIRCHHOF et al., 2006), and LiDAR point clouds (YAO et al, 2011).

1) Photogrammetrie & Fernerkundung, Technische Universität München (TUM), Arcisstraße 21, 80333 München, http://www.pf.bv.tum.de
2) Fraunhofer Institut für Hochfrequenzphysik und Radartechnik FHR, Fraunhoferstraße 20, 53343 Wachtberg, http://www.fhr.fraunhofer.de

Furthermore, there are furthermore many investigations in the context of Moving Target Detection (MTI) based on along track interferometry (CERUTTI-MAORI et al., 2008) but the detection of stationary vehicles has received only little attention so far. In (MAKSYMIUK et al., 2012) it was shown that robust detection of stationary vehicles in SAR intensity images is possible. The approach makes use of supervised learning and employs a set of different features ranging from statistical to spectral features as well as clustering mechanisms. It is a property of the former approach that there is no explicit formulation of the characteristic parts of the object and that most of the features capture additionally some of the surroundings which do not belong to the object itself. In supervised learning a set of features is generated and based on a training set the algorithms determine the importance of the features. Nevertheless it is not enforced that the features capture only the characteristic or essential parts of the objects, although the machine learning procedures ensure that the features are suitable to distinguish between the classes. In this work we want to perform detection only on the characteristic parts which were examined explicitly a priori for the different classes. A multiscale approach is used to represent an object by connected gray level components and therefore as a set of bright and dark components. Since the body of vehicles generally consists of metal it is expected that to cause strong backscattering which in turn is represented as bright connected component in the image. Since SAR images suffer from speckling pixel based methods are likely to fail. A better approach is to work on a higher level than on single pixel values. For this reason it seems promising to work on connected gray level components. They implicitly contain additional information about the objects' shape.

2 Mathematical Morphology and Image Representation

In this section, we present some basics of Mathematical Morphology (MM), the hierarchical image representation by Min- or Max-trees and attribute filters. These are essential for our proposed detection approach, which is introduced afterwards. We will show how to extract relevant features from the images and use them to detect stationary vehicles.

2.1 Morphological Operators and Connected Components

Image segmentation is a low-level processing task and of critical importance for most high-level applications such as object detection. It aims to identify and separate different regions of the images by their visual properties. Ideally the processing should not introduce artificial structures or other kind of artifacts. In this work we use a morphological decomposition related to granulometry and morphological reconstruction to divide the image into connected components with respect to gray levels. Generalized to a gray level image $f(x, y)$ where x, y are the pixel coordinates, a connected component at level t is defined as a set of pair wise connected pixels from the cross section of f at level t, which is the set of all pixels whose values are greater than or equal to t. Fig. 1 shows an example of connected components for a gray level image showing three vehicles oriented across track. At level 0 the whole image is a connected component. In general connected components as image segmentation result are of interest because of the assumption that pixels which are in a connected component belong to the same object.

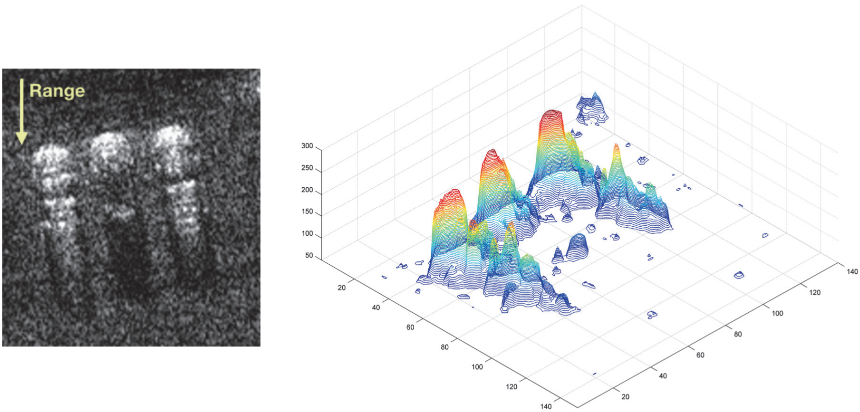

Fig. 1: Image with three vehicles oriented across track (left) and their corresponding connected components at different grayscale levels (right). For better visual representation it is obtained from the median filtered image and only the upper gray levels are presented. The property being a convex set is obvious in this representation.

2.2 Attributes and Attribute Filters

The first morphological reconstruction methods were based on the size of the objects. The well known operator *opening by reconstruction* removes all objects which do not meet the following criterion T, which uses the size as attribute: *the given structure element must fit inside the region*. This criterion is *increasing* since if a connected set meets this criterion then all its supersets also satisfy it. Other properties of the openings are that they are *anti-extensive* and *idempotent*. But since the majority of descriptive criterions like e.g. eccentricity are *nonincreasing,* the attribute opening must be generalized to attribute thinning (BREEN, 1996). However, a nonincreasing criterion T can introduce some artificial structures, which is undesired. In this case the filtering has to take care of this by imposing a kind of increasing property (see section 2.3.1).

2.3 Hierarchical Image Representation

The operators introduced in the last section belong to the class of *connected operators*. These operators describe and filter the image in terms of the *flat zones* rather than single pixels. As already mentioned they do not introduce any new contours and thus are very attractive as information preserving filter. They implicitly work on a data structure called *Max-Tree* (SALEMBIER et al., 1998) for anti-extensive (opening, thinning) or *Min-Tree* for extensive (closing, thickening) operators. This data structure is a hierarchical representation of the image in terms of the connected components, where the pixels assigned to node N_h^k are the smallest superset of all pixels attached to its successors. Here, h is the gray level and k an index variable. This property also explains the multi-scale characteristic since small components are located near to the leafs whereas the root represents the whole image and is the node with the lowest intensity (Max-tree). It is possible to reconstruct an image Q perfectly from the tree.

2.3.1 Attribute Filtering

Within the Max-tree representation an attribute filtering contains three different stages: construction, filtering, and image restitution. There are several strategies of filtering a Min/Max-Tree with respect to the attached attributes. If the criterion used for filtering is increasing for Max-Trees, all the approaches will deliver the same result. But if the criterion does not meet the condition the attribute filtering becomes more complicated and leads to different results depending on the strategy (DALLA MURA, 2010). In this work we use the *subtractive* rule as *nonpruning* strategy removing only nodes which do not satisfy the criterion and update its successors (URBACH, 2007). If a node is removed its pixels are assigned to the highest ancestor meeting the criterion and are therefore lowered in their gray level. The same amount is also used to lower the gray level of the successors. This means that a leaf node having many predecessors which do not meet the criterion has a low gray level at the end of the filtering. This can be interpreted as a kind of *contrast* with respect to the attribute and is an indicator for the exposedness of the component. An example for tree creation, filtering and image restitution is shown in Fig. 2. It should be noted that the tree creation has to be done only once. Attributes can

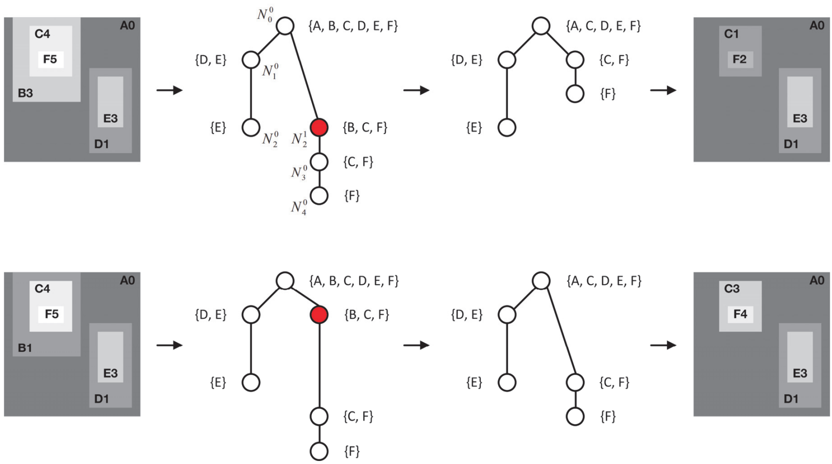

Fig. 2: Exemplary tree creation, subtractive attribute filtering and image restitution. The number given in the identifier of the gray level patches indicates their gray level. The red node (component B) is assumed to fail a criterion and is therefore removed. Note the difference between the upper and lower example with respect to the gray level of components C and F.

be computed at the time of tree creation. After this process the tree can be used to filter for different attributes. Due to this the approach is computational effective.

3 Part-based Representation by Attribute Filters

The set of methods presented in the last section, especially the attribute filters, allow the description of an image as a hierarchy of connected components with additional attributes attached. However the detection of objects which are not a single connected component is not straightforward. Our proposed method is to decompose the class of objects into multiple common parts P_j as can be seen in Fig. 3, where j is some index variable. Since the connected components work on gray levels the parts represent bright (convex) or dark (concave) parts and every part is considered as a single connected component with a set of attributes. Attribute filters (thinning, thickening) allow extracting them easily from the image neglecting the other components and without introducing artificial artifacts but it remains to model their spatial distribution or arrangement. One approach to tackle this problem is to use tree matching procedures to find node structures which best represent the hierarchy and match the attributes (KUMAR et al., 2011). Unfortunately this is a nontrivial problem especially if not all parts of the object are enforced to exist for every single object. Although there are some procedures using heuristics the problem remains difficult. However the problem statement can be reduced in complexity if only the spatial arrangement has to be analyzed. Assuming that the parts match the criterios, a correlation can be used to analyze and assess the position of the parts within the object. For this process we use a mask M_j with a predetermined size (depending on the size of the vehicles) containing a distribution centered at the expected position of the specific part. The assessment of every part is therefore given by $C_j = Q_j * M'_j$, where M'_j is the normalized mask and Q_j the reconstructed image as the result of an attribute filtering. The final detection result is then a linear combination $Y = \sum_j \lambda_j C_j$ with $\sum_j \lambda_j = 1$.

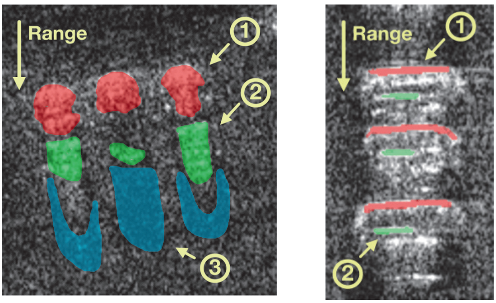

Fig. 3: Decomposition of vehicles of different orientation. Cross track (left) and along track (right). The decomposition is based on bright (convex) and dark (concave) gray level sets. For both classes the components 1 and 2 are convex components residing in the Max-tree whereas component 3 (shadow area) is based on the Min-tree as it is a concave component.

With the linear combination it is possible to put more emphasis on reliable parts e.g. the bright line for vehicles oriented along track is less probable caused by random events than the smaller one below. In order to use the exposedness of the component as a measure in the detection process, Q_j is not normalized. To scale C_j it is simply divided afterwards by $t_{\max}$. It should be noted that there is no investigation in determining a model for an understanding which parts of the vehicle are causing the backscattering visible in the image. One subtle difficulty is the choice of the common parts and their corresponding descriptive attributes. On the one hand they should represent a whole class of objects but should on the other hand allow discriminating between them and other objects in the scene.

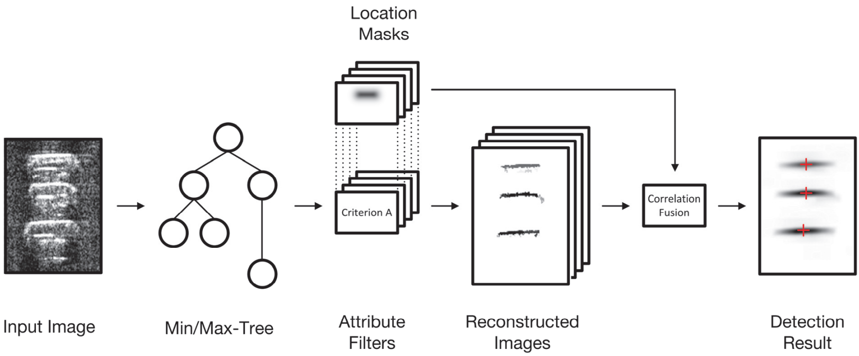

Fig. 4: Flow chart of the vehicle detection based on the decomposition profile. The process of detecting and evaluating the upper bright line is shown. Every attribute filter has its own location mask assessing the position of that part within the object. The final detection result is a linear combination of all part detections. Looking at the reconstructed image it can be clearly seen that only specific parts pass the filter.

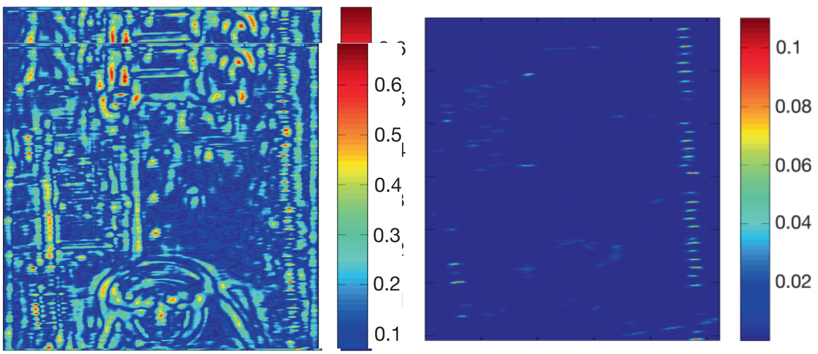

Fig. 5: Comparison between normalized correlation with the mean image of vehicles oriented along track (left) and our proposed approach (right) working with decompositions. Note the difference in scale.

One solution is to apply supervised learning procedures to determine the best representation. If there are not enough data for training a classifier the parameters need to be defined just by user knowledge or inspection of the visual appearance. The advantage of the decomposition is the reduction of artifacts, thus the correlation is performed only on data which is essential and characteristic for the object with high probability. Fig. 5 shows a comparison between a standard correlation and our proposed approach. Obviously the responses from the vehicles in the image are much more distinctive and compared to their local neighborhood making it remarkably easier to detect them.

4 Experiments

The dataset for the evaluation consists of three SAR images obtained from the PAMIR sensor (BRENNER, 2012) working in X-Band with a resolution of 15 cm in range and 7 cm in azimuth. The images contain 77 vehicles oriented along track and 29 vehicles oriented cross track.

4.1 Attributes for Part Based Detection

In this work the attributes and their values are determined statistically by using mean images of vehicles oriented across track and along track. The following attributes are used for components description: Area, direction obtained from main axes, moment of inertia (measure for the elongation of a shape), standard deviation in gray levels, length of the diagonal of the bounding box, and the volume. The latter is just the product of all pixel values residing in the component. These attributes are sufficient to obtain a filtering which removes many components not belonging to vehicles.

4.2 Detection Results

After receiving the result from the fusion of all decomposition parts the local maxima are detected and marked after thresholding with a centered rectangle. The size of the rectangle is choosen according to the size of the location masks. We used a mask size of 35 times 70 pixels for vehicles oriented along track and 80 times 40 pixels for vehicles oriented cross track. The threshold was chosen such that the objects have to exhibit a lower attribute contrast.

5 Conclusions

This paper presented an approach to object detection by decomposing the object into multiple connected components in a morphological sense. Every component has a set of descriptive attributes. The hierarchical representation as Min/Max-Tree allows using powerful attribute filters as information preserving image transformation. The spatial composition of the objects

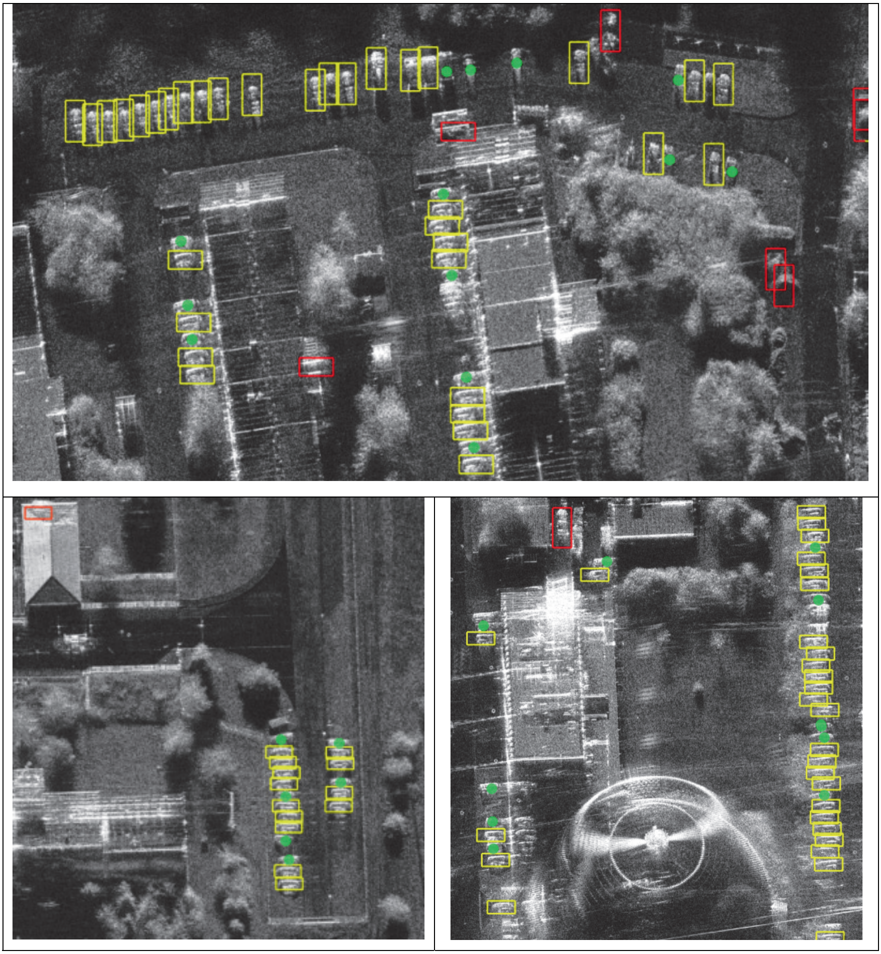

Fig. 6: Detection results of the proposed approach. Although there are some missing detection the method works remarkably well. False detections are marked with a red rectangle. Missing detection are indicated by green points. However, there are no viable ground truth data.

parts is handled by using simple correlation with location masks. The detection performance is only slightly lower than the more sophisticated approach using Support Vector Machines or Random Forests in a supervised learning setup (MAKSYMIUK et al., 2012). It can thus serve as a approach to guide more sophisticated by computationally expensive methods by ruling out areas which contain no vehicle. In future both approaches can also be combined using the connected components and their attributes as input for supervised learning classifiers. As a consequence one may conclude that the use of higher level features in SAR images has advantages to the use of single pixel values. Further investigations are necessary to determine how to model the objects' parts in the best manner.

6 References

BREEN, E. J. & JONES, R., 1996: Attribute Openings, Thinnings, and Granulometries, Computer Vision and Image Understanding, **64** (3), pp. 377-389.

BRENNER, A. R., 2010: Proof of concept for airborne SAR imaging with 5 cm resolution in X-band, 8th European Conference on Synthetic Aperture Radar (EUSAR), pp. 1-4.

BRENNER, A. R.; ESSEN, H. & STILLA, U., 2012: Representation of stationary vehicles in ultra-high resolution SAR and turntable ISAR images, 9th European Conference on Synthetic Aperture Radar (EUSAR), pp. 1-4.

CERUTTI-MAORI, D.; KLARE, J.; BRENNER, A. R. & ENDER, J. H. G., 2008: Wide-Area Traffic Monitoring with the SAR/GMTI System PAMIR, IEEE Trans. on Geoscience and Remote Sensing, **46** (10), pp. 3019-3030.

DALLA MURA, M.; BENEDIKTSSON, J. A.; WASKE, B. & BRUZZONE, L., 2010: Morphological Attribute Profiles for the Analysis of Very High Resolution Images, IEEE Trans. on Geoscience and Remote Sensing, **48** (10), pp. 3747-3762.

KIRCHHOF, M. & STILLA, U., 2006: Detection of moving objects in airborne thermal videos, ISPRS Journal of Photogrammetry and Remote Sensing, **61** (34), pp. 187-196

KUMAR, R.; TALTON, J. O.; AHMAD, S.; ROUGHGARDEN, T. & KLEMMER, S. R.,2011: Flexible Tree Matching, International Joint Conference on Artificial Intelligence

LEITLOFF, J.; HINZ, S. & STILLA, U., 2010: Vehicle Detection in Very High Resolution Satellite Images of City Areas, IEEE Trans. on Geoscience and Remote Sensing, **48** (7), pp. 2795-2806

LENHART, D.; HINZ, S.; LEITLOFF, J. & STILLA, U., 2008: Automatic traffic monitoring based on aerial image sequences, Pattern Recognition and Image Analysis, **18** (3), pp. 400-405

MAKSYMIUK, O.; SCHMITT, M.; BRENNER, A. R. & STILLA, U., 2012: First Investigations on Detection of Stationary Vehicles in Airborne Decimeter Resolution SAR Data by Supervised Learning, Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS), pp. 3584-3587.

PESARESI, M. & BENEDIKTSSON, J. A., 2001: A New Approach for the Morphological Segmentation of High-Resolution Satellite Imagery, IEEE Trans. on Geoscience and Remote Sensing, **39** (2), pp. 309-320.

SALEMBIER, P.; OLIVERAS, A. & GARRIDO, L., 1998: Antiextensive Connected Operators for Image and Sequence Processing, IEEE Trans. on Image Processing, **7** (4), pp. 555-570.

SALEMBIER, P. & GARRIDO, L., 2000: Binary Partition Tree as an Efficient Representation for Image Processing, Segmentation, and Information Retrieval, IEEE Trans. on Image Processing, **9** (4), pp. 561-576.

SOILLE, P. & PESARESI, M., 2002: Advances in Mathematical Morphology Applied to Geoscience and Remote Sensing, IEEE Trans. on Geoscience and Remote Sensing, **40** (9), pp. 2042-2055.

STILLA, U.; MICHAELSEN, E.; SOERGEL, U.; HINZ, S. & ENDER,J. H. G., 2004: Airborne monitoring of vehicle activity in urban areas, International Archives of Photogrammetry and Remote Sensing, **35**, pp. 973-979.

URBACH, E. R; ROERDINK, J. B. T. M. & WILKINSON, M. H. F, 2007: Connected Shape-Size Pattern Spectra for Rotation and Scale-Invariant Classification of Gray-Scale Images, IEEE Trans. on Pattern Analysis and Machine Intelligence, **29** (2), pp. 272-285.

YAO, W. & STILLA, U. 2011: Comparison of Two Methods for Vehicle Extraction From Airborne LiDAR Data Toward Motion Analysis, IEEE Geoscience and Remote Sensing Letters, **8** (4), pp. 607-611.

Publikationen der Deutschen
Gesellschaft für Photogrammetrie,
Fernerkundung und Geoinformation e.V.

Band 22 **2013**

Vorträge

**33. Wissenschaftlich-Technische
Jahrestagung der DGPF**

27. Februar – 1. März 2013
in Freiburg i. B.

Dreiländertagung D - A - CH

ISSN 0942-2870

Eckhardt Seyfert, Hrsg.