

A CASCADED ENSEMBLE CLASSIFIER FOR OBJECT SEGMENTATION IN HIGH RESOLUTION POLARIMETRIC SAR DATA

Marc Jäger, Andreas Reigber

Microwaves and Radar Institute
Department of SAR Technology
German Aerospace Center (DLR)

Olaf Hellwich

Computer Vision and Remote Sensing Group
Berlin University of Technology

ABSTRACT

The paper proposes a novel approach to object classification and segmentation in multi-channel (e.g. polarimetric) SAR data. The classifier is intended for particularly difficult problems, where objects of interest exhibit a high degree of radiometric, polarimetric and geometric heterogeneity, both within individual object instances and across the object category as a whole. Classification is based on a non-parametric characterization of scene contents that avoids model assumptions liable to fail in this scenario. The classifier structure is based on a combination of techniques developed for related problems in computer vision: the cascade architecture helps breaking down the problem into manageable stages while random forests provide a powerful framework for learning and combining discriminative classification rules. In addition, scale space techniques explicitly introduce non-local, contextual and geometric information into the classification process. Preliminary results illustrate the potential of the proposed approach with respect to the task of building segmentation in dual-polarized TerraSAR-X data.

Index Terms— Synthetic Aperture Radar, SAR Polarimetry, Classification, Texture

1. INTRODUCTION

The accurate segmentation of objects in high-resolution synthetic aperture radar data is treated as a classification problem in which a non-parametric description of local signal statistics plays an important role. The proposed classifier is trained from examples and is not specific to a particular object category or acquisition mode; in particular it extends naturally to multi-channel SAR data such as (partially-) polarimetric acquisitions. The principal focus is on difficult problems where the objects of interest may have a complex internal structure with a large radiometric and polarimetric diversity within individual objects as well as between different instances of the same object type. This type of problem will continue to become more relevant as increasing sensor resolutions continue to reveal more of the spatial structure and texture of objects in SAR acquisitions.

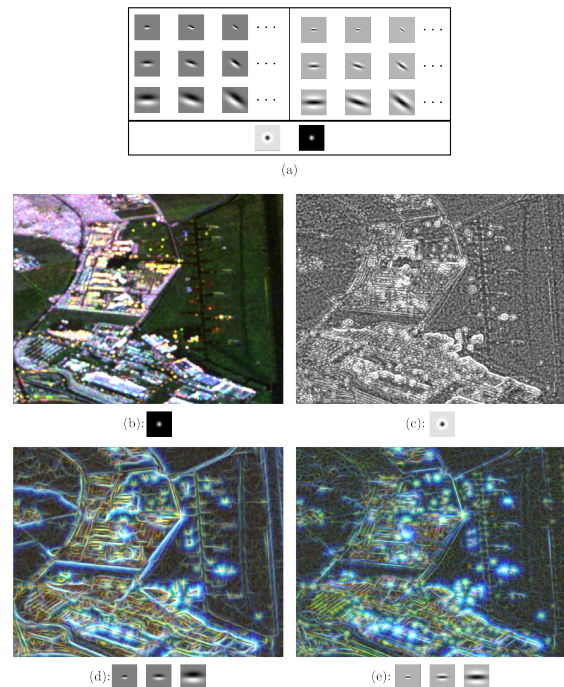


Fig. 1. The filter bank applied to characterize local image structure. a) The eight families of filter applied. Clockwise from the top left: oriented 1st order, oriented 2nd order, Gaussian and Laplacian filters. b) Sample input data, convolved with the Gaussian filter. c) The Laplacian response. d) 1st order responses for the three filter scales. e) 2nd order responses for the three filter scales.

The approach is largely inspired by recent advances in the segmentation and categorization of objects in optical imagery. Here, Shotton *et al.* [1] have proposed an approach based on randomized forest[2] classifiers using texture-related information. A somewhat similar approach, combining boosting and conditional random fields, was previously proposed in [3]. Both techniques achieved impressive results and are, implicitly or explicitly, based on the so-called *texton* representation of optical imagery. This representation has gone through several re-formulations before reaching the current state of the art in [4].

The current approach differs from previous techniques by the introduction of a new cascade-based classifier architecture. Cascade classifiers were first introduced in [5] and extended in [6] to solve object detection problems in real time. Although computational cost is not a primary concern here, cascades are attractive in the context of object segmentation since they provide mechanisms to deal with the enormous diversity of the background in object segmentation tasks. In addition, scale space analysis techniques are introduced to allow the classifier to discriminate on the basis of non-local image structure geometry.

Section 2 briefly reviews the non-parametric descriptor used as a basis for classification. The classifier architecture and the training process are summarized in section 3. Section 4 introduces the scale space techniques adopted to account for geometric object properties, while section 5 presents and briefly discusses preliminary results for building segmentation in partially polarimetric TerraSAR-X data. Section 6 summarizes the contributions and points to future work.

2. TEXTON DESCRIPTORS

Texton descriptors, introduced in [4], are extracted in a two step process which begins with applying a multi-scale filter bank to obtain a local signal characterization as illustrated in figure 1. Each sample of a given image is then compared to a so-called dictionary on the basis of the filter responses associated with it. The dictionary contains entries that describe local image structure that is recurrent in terms of geometric structure and radiometric/polarimetric scattering characteristics.

Finally, an image region is described as the histogram of dictionary entries that have been matched within it. In practice, this histogram is a local description of the observed backscatter characteristics as well as the relative abundance and constellation of line, edge and point like geometric image constituents. Importantly, the process of histogram accumulation is inherently non-parametric, i.e. it involves no model assumptions concerning the distribution obtained, and is therefore suitable for complex, heterogeneous scene contents that are otherwise difficult to model satisfactorily.

For a full description of an extension of this approach to fully polarimetric SAR, which is beyond the scope of this ar-

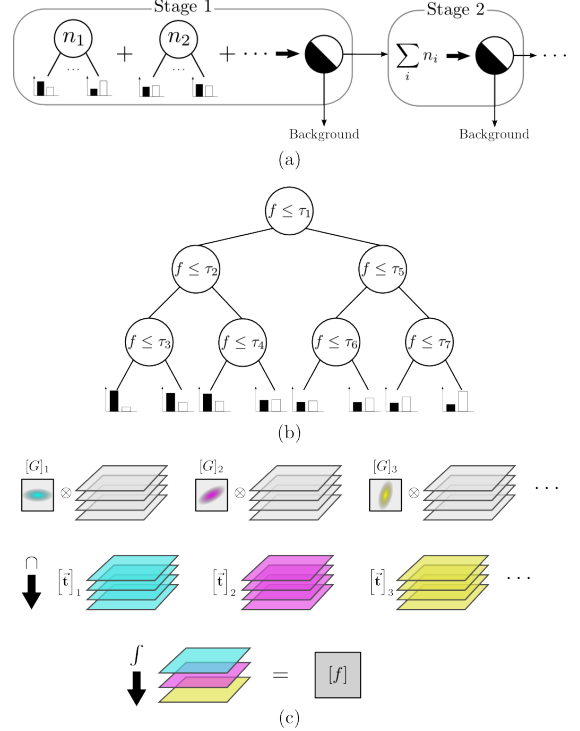


Fig. 2. Classifier and decision tree structure. a) Structure of the cascade classifier. b) Extraction of features from texton vectors. c) The binary decision tree (with leaf posterior distributions) in a cascade node.

ticle, the reader is referred to [7].

3. CASCADE CLASSIFIER

The classifier produces a bi-level segmentation of a multi-channel (e.g. polarimetric) SAR data set by assigning each resolution cell to one of the classes $c = 0$ (background) or $c = 1$ (foreground/object). As illustrated in figure 2a, the proposed classifier consists of several *stages* in accordance with the cascade architecture introduced by Viola and Jones in [5]. A given input sample traverses the first stage of the cascade, at the end of which it is either irrevocably assigned to the background class $c = 0$ or propagates to the next stage where the process repeats. Any sample that traverses all stages in the cascade without being assigned to the background is then deemed to be part of the foreground and is assigned $c = 1$.

For the present purposes, this architecture helps to overcome difficulties stemming from the fact that the background class is, in practice, extremely diverse. Instead of attempting to characterize the entire background in a single, monolithic classifier, stages break down the classification process and allow the learning process to focus on one (manageable) part of the background at a time. Latter stages in the classifier can assume that most background has been eliminated and focus on

finding specific rules to isolate the few remaining but difficult background samples.

In the proposed architecture, each stage of the classifier is itself a random forest [2] in which a number of weak classifiers combine to produce a single strong classifier. The individual weak classifiers are represented by the nodes n_1, n_2, \dots in figure 2a. Internally, each node takes the form of a binary decision tree in which each leaf l contains a posterior density $p(c|l)$ over classes $c \in \{0, 1\}$, as illustrated in figure 2b. Each binary decision is based on comparing a different, scalar *feature* derived from the texton descriptors to a threshold.

When the classifier is applied to SAR data after training is complete, each image sample traverses each decision tree and the posteriors $p(c|l)$ encountered are averaged. At the end of each stage, the accumulated posterior determines whether a given sample traverses to the next stage or not.

The pool of scalar features used in binary decisions is generated by locally averaging the texton vectors (local histograms) with a bank of oriented low-pass filters of a particular size and anisotropy, as shown in the first row of figure 2c. The filtered results are compared to an arbitrary template using one of a number of distance measures (e.g. Euclidian, χ^2 or Kolmogorov-Smirnov), as shown in the second row of figure 2c. Finally, rotation invariance is achieved by computing an integral invariant over low-pass filter orientations (e.g. mean, median, maximum and similar) as shown in the third row. The aim of this design is to retain rotation invariance, which is deemed important in remotely sensed imagery, while providing a large number of degrees of freedom to ensure a highly diverse set of features for use in binary decisions.

Individual nodes are trained using the techniques described in [2], including advanced techniques such as *sample bagging*. Individual binary decisions are trained by selecting a random feature from the pool and then setting the threshold to minimize the entropy of the foreground/background training sample distributions propagated to descendents. The overall cascade is constructed as described in [5]. Interestingly, this process guarantees a maximum false alarm as well as a minimum hit rate, and therefore a user specified classifier performance with respect to the training data, upon completion.

4. SPATIAL CONTEXT AND OBJECT GEOMETRY

The cascade classifier of the preceding section makes decisions that rely primarily on local image contents (depending on the size of the low-pass filters used in binary decision nodes). An extension that explicitly incorporates non-local context and object geometry entails letting binary decision nodes access the entire Gaussian scale space, as illustrated in the left column of figure 3a, of a given input feature. In this approach, each binary decision is trained by selecting a decision threshold that is optimal over all considered scales

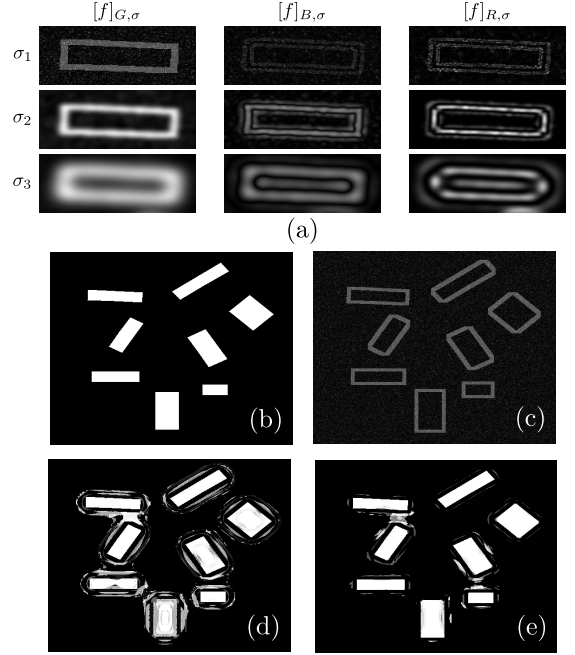


Fig. 3. Learning contextual and geometric patterns via scale space analysis: a) Features extracted from Gaussian scale space, from left to right: the scale space itself, the Hessian trace and the Hessian determinant. b) Training mask used in a synthetic example. c) Simulated feature for the binary decision nodes (in this example all binary decision nodes use the same input feature). d) Class posterior density for a single decision tree. e) Class posterior averaged over five decision trees.

of observation.

In addition to the scale space itself, nodes may also base decisions on the trace and determinant of the associated Hessian, as illustrated in the second and third columns of figure 3a. These rotation invariant quantities are directly related to the presence of *blob* and *ridge* like structures in the underlying feature and were originally introduced for multi-scale geometric structure detection in [8].

As illustrated in figure 3b-e, this approach adds valuable discriminative power to the classifier: although the input signal only distinguishes the outline of the rectangular objects considered, the classifier can correctly discriminate the object interior on the basis of contextual information.

5. PRELIMINARY RESULTS

The preliminary results in this section illustrate classification/segmentation performance for a classifier trained to detect buildings in high resolution, dual-polarized TerraSAR-X imagery. The classifier was trained and evaluated on distinct regions of the same data acquisition. Training was carried out on the basis of manually annotated foreground and

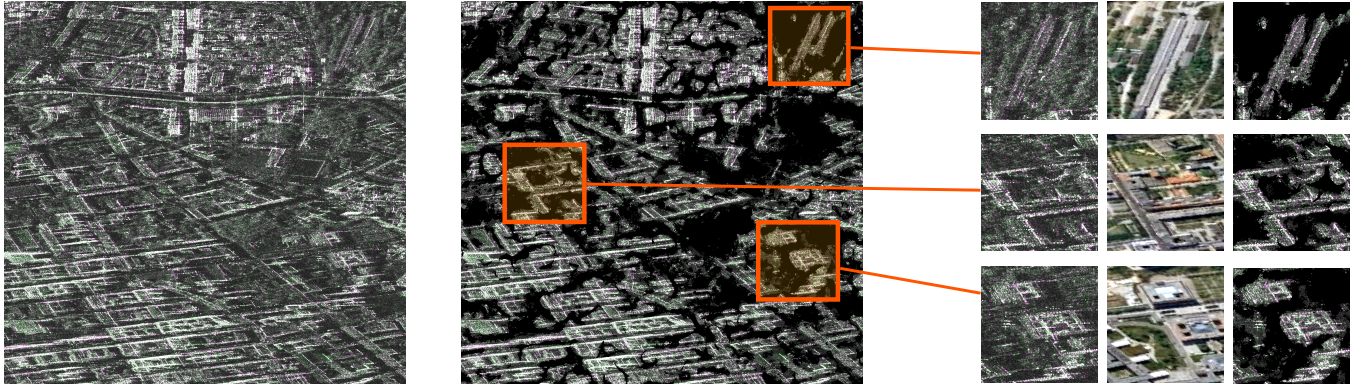


Fig. 4. The results of applying the classifier to unseen data after training for the category 'building'. Left: High-resolution TerraSAR-X image (dual-polarized) acquired over Berlin, Germany, in 2009. Middle: Segmentation result with pixels in the 'background' class set to black. Right: Detail views of three regions with reference optical imagery.

background masks created with the help of optical imagery back-geocoded into the slant-range geometry.

A visual inspection of the results illustrated in figure 4 suggests that the classifier is, by and large, able to accurately localize buildings despite pronounced geometric distortions and high levels of clutter in the underlying dataset. Encouragingly, the classifier output in some areas is difficult judge, even to an experienced user, without reference to the corresponding optical data.

Although these preliminary results appear promising, further work is required to quantitatively validate the technique with respect to a larger corpus of SAR data and other state-of-the art approaches.

6. CONCLUSION

The preceding sections outline a novel approach to object segmentation and classification for multi-channel SAR acquisitions. The classifier architecture, featuring several techniques adapted from related research in computer vision, is intended to tackle challenging problems involving highly heterogeneous foreground and background classes. The classifier incorporates scale space analysis techniques which are shown, on the basis of simulated data, to provide valuable, discriminative information in terms of object context and geometry.

Preliminary results concerning building detection in dual-polarized TerraSAR-X data are presented and appear promising. A more quantitative evaluation and a comparison with alternative state of the art approaches, however, remains to be carried out.

7. REFERENCES

- [1] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texon forests for image categorization and segmentation," in

Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, June 2008, pp. 1–8.

- [2] Leo Breiman, "Random forests," in *Machine Learning*, 2001, pp. 5–32.
- [3] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Texonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in *In ECCV*, 2006, pp. 1–15.
- [4] M. Varma and A. Zisserman, "A Statistical Approach to Texture Classification from Single Images," *International Journal of Computer Vision*, vol. 62, pp. 61–82, April 2005.
- [5] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 1, pp. 511, 2001.
- [6] Rainer Lienhart, Er Kuranov, and Vadim Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," in *In DAGM 25th Pattern Recognition Symposium*, 2003, pp. 297–304.
- [7] Jäger M., Reigber A., and Hellwich O., "A Non-Parametric Texture Descriptor for Polarimetric SAR Data with Applications to Supervised Classification," in *To appear in Proc. EuSAR2014*, 2014.
- [8] Tony Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, pp. 79–116, 1998.