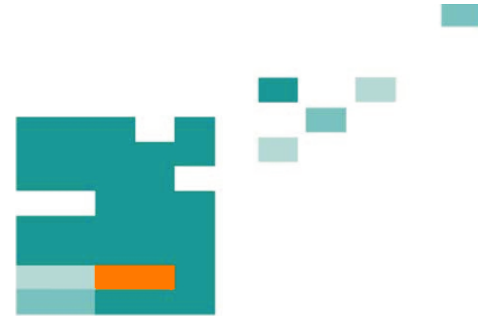


## 55. IWK

Internationales Wissenschaftliches Kolloquium  
International Scientific Colloquium



13 - 17 September 2010

# Crossing Borders within the **ABC**

**A**utomation,

**B**iomedical Engineering and

**C**omputer Science



Faculty of  
Computer Science and Automation

[www.tu-ilmenau.de](http://www.tu-ilmenau.de)

*th*  
TECHNISCHE UNIVERSITÄT  
ILMENAU

Home / Index:

<http://www.db-thueringen.de/servlets/DocumentServlet?id=16739>

## **Impressum Published by**

Publisher: Rector of the Ilmenau University of Technology  
Univ.-Prof. Dr. rer. nat. habil. Dr. h. c. Prof. h. c. Peter Scharff

Editor: Marketing Department (Phone: +49 3677 69-2520)  
Andrea Schneider (conferences@tu-ilmenau.de)

Faculty of Computer Science and Automation  
(Phone: +49 3677 69-2860)  
Univ.-Prof. Dr.-Ing. habil. Jens Haueisen

Editorial Deadline: 20. August 2010

Implementation: Ilmenau University of Technology  
Felix Böckelmann  
Philipp Schmidt

## **USB-Flash-Version.**

Publishing House: Verlag ISLE, Betriebsstätte des ISLE e.V.  
Werner-von-Siemens-Str. 16  
98693 Ilmenau

Production: CDA Datenträger Albrechts GmbH, 98529 Suhl/Albrechts

Order trough: Marketing Department (+49 3677 69-2520)  
Andrea Schneider (conferences@tu-ilmenau.de)

ISBN: 978-3-938843-53-6 (USB-Flash Version)

## **Online-Version:**

Publisher: Universitätsbibliothek Ilmenau  
[ilmedia](#)  
Postfach 10 05 65  
98684 Ilmenau

© Ilmenau University of Technology (Thür.) 2010

The content of the USB-Flash and online-documents are copyright protected by law.  
Der Inhalt des USB-Flash und die Online-Dokumente sind urheberrechtlich geschützt.

## **Home / Index:**

<http://www.db-thueringen.de/servlets/DocumentServlet?id=16739>

# THE ADVANTAGE OF SEGMENT-BASED CLASSIFICATION OF MULTI-MODAL, MULTI-TEMPORAL AND HIGH-RESOLUTION SATELLITE IMAGES

*Torsten Koch, Rico Nestler, Tim Kubertschak*

Zentrum für Bild- und Signalverarbeitung e.V. (ZBS)  
Werner-von-Siemens Str. 10  
D-98693 Ilmenau (Germany)

Contact:

torsten.koch@zbs-ilmenau.de, (+49-(0)-3677-689768-4),  
rico.nestler@zbs-ilmenau.de, (+49-(0)-3677-689768-5),  
tim.kubertschak@zbs-ilmenau.de, (+49-(0)-3677-689768-6)

## ABSTRACT

The automatic generation of land use / land cover information from multimodal satellite image data is a high challenging task. Due to the high temporal variability of several phenological and biological parameters between different vegetation types, a multi-temporal investigation arise better discernibility among various classes. Furthermore, using the information of various modalities in a synergetic manner the quality of a land-use classification will be increased substantially.

Only related to the classification step the challenge results from complex shaped class regions in feature spaces of high dimension. The needed classification algorithm should allow both a high degree of freedom to describe the classes or the inter class separation in the feature domain and a sufficient generalize property. The last mentioned is because of the limited amount of training samples per class that are obtainable.

The approach, presented in this paper, is based on a per-segment classification. A suitable classification method is closely connected with an efficient and powerful algorithm for the foregoing segmentation. Those segmentation and classification methods are components of an automated processing-pipeline, which will be developed inside the project ENVILAND-2<sup>1</sup>. The interim results, arises from that project, where presented and discussed.

*Index Terms* - segment-based classification, multimodal, multi-temporal, land-use classification

---

<sup>1</sup> ENVILAND2 is funded by the German Aerospace Centre and the Federal Ministry of Economics and Technology (of Germany)  
FKZ 50 EE 0844 -- 50 EE 0847

## 1. INTRODUCTION

Apart from the automated land-use-classification, the automated change-detection is also an important goal of the project Enviland-2. The developed processing-pipeline covers all necessary steps (from raw data preprocessing of all sensor modalities, e.g. geocoding, registration, restoration, until the more application dependent segmentation and classification).

The required precision may be increased significantly by using a segment-based classification. This is because of the integral features that are the basis of the classification, which reduces the influence of several disturbances, on the one hand. On the other hand, information leaks may be countervailed by the synergetic usage of multi-modal and multi-temporal remote sensing data (i.e. SAR and optical data). In particular, the soil humidity is a good example, since this important feature is only inferable from SAR-data.

Additionally, the fundamental multi-channel processing concept allows the processing of time series too. This will be used to identify different kinds of field crops or other vegetation types, by multiple observations during their growing period.

The multi-channel segmentation step just produces an over-segmented map, which shows concatenated areas, characterized by low variance over all channels of the used image stack [6]. Later on, the classification associates those homogeneous areas with specific classes.

## 2. STRUCTURE OF CLASSIFICATION

According to practical oriented investigations and evaluation<sup>2</sup> of methods and structures on land-use-classification [7], a generic classification concept originated. Furthermore, an conceptual approach to a solution for selection and parametrisation of important classification-methods (Mahalanobis, Mixed-Distribution, K-Nearest-Neighbour, Support-Vector-Machine (SVM) [3][4], partitional cluster analysis) was created. At present, a segment-based SVM-classification scheme will be used.

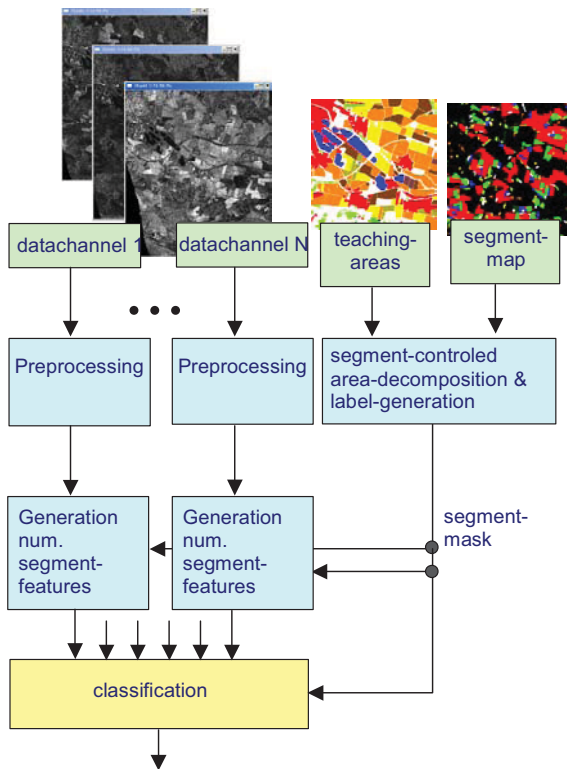


Figure 1 principle of teaching

An registered image stack (beside appropriate optional mask images) and a segment-label-image will feed into those classification. This image stack may consist of images from several satellites, sensors and recording times and is generated in the processing-stages registration [8] and segmentation [6]. Differences in resolution will be equalized by appropriate resampling methods.

Besides, the classifier has to be trained and therefore a teaching data set is required. In fact, this data set consists of a label-image, which maps certain regions to its corresponding classes. This teaching datasets are still being defined by ground-truth-maps generated by land register datasets from

<sup>2</sup> using the VIP-Toolkit (product of ZBS e.V.)

administrative offices, field inspections or image-interpretation by experts. However, it is intended to derive the teaching data set by an automated approach using unsupervised classification methods.

An optional preprocessing may be used as a first stage in order to perform a additional noise- and speckle-reduction. Beside established methods (i.e Kuan, Lee, Enhanced Lee, ...), also geometric filtering [1] and some methods, refined by the ZBS (i.e. adapted anisotropic diffusion filtering (AD), based on [5]), can be applied.

After this, numerical segment-features will be extracted. Thus in the simplest case it originates a vector per segment, with each component representing the mean intensity value of the specific channel within the segment.

All these numerical segment-features are the input of the classifier [3][4][7].

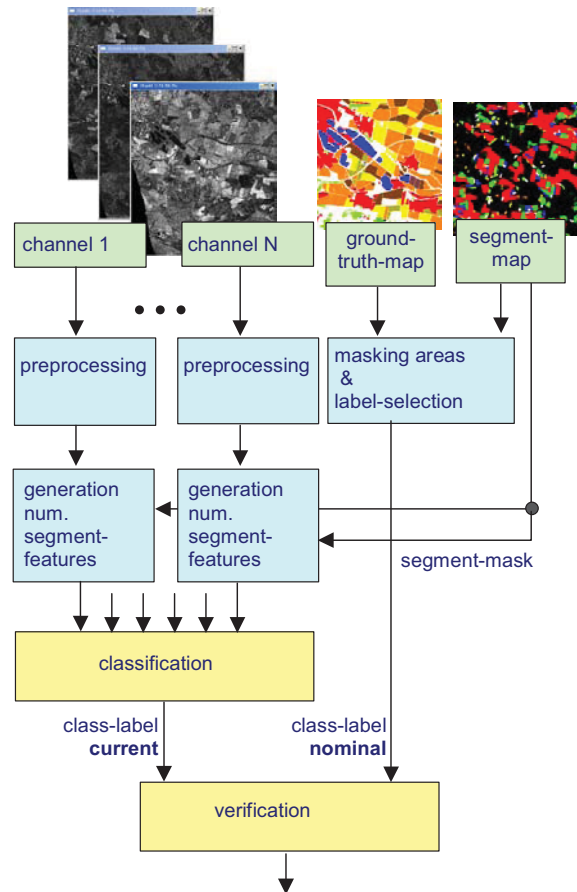


Figure 2 principle of verification

## 3. EVALUATION RATIO

To compare and evaluate the used classification methods and -approaches a comparison between the ground-truth-map and the classification-results (referring to the segment map and the image stack) will be done. In order to perform objective comparisons, the quality measures *Overall Accuracy (OA)* and *KAPPA Accuracy (KHAT)* are used

[1][2][3]. Both measures based on a N-Class-Confusion-Matrix  $H$  (Equation 5). To ensure uniqueness, below it is assumed that the confusion matrix  $H$  is just made up of segments, which cover a specific class (in the ground-truth-map) with at least 80% area rate.

$$OA = \frac{\sum_{i=1..N} H_{i,i}}{S}, \quad S = \sum_{i=1..N} \sum_{j=1..N} H_{i,j}$$

$$OA_{Opt} \rightarrow 1$$

Equation 1 Overall Accuracy ( $S$ =count of data-samples)

While  $OA$  represents the whole classification accuracy with respect to correct class-assignments only,  $KHAT$  describes the deterministic connection between classification-result and reference-data. The latter depends on the *Class Probability (CP)* and the *Decision Probability (DP)*.

$$KHAT = \frac{OA - \sum_{i=1..N} DP_i \cdot CP_i}{1 - \sum_{i=1..N} DP_i \cdot CP_i}$$

Equation 2 KAPPA Accuracy

$$CP_j = \frac{\sum_{i=1..N} H_{i,j}}{S} \quad CP_j \in [0,1] \forall j$$

Equation 3 Probability of occurrence of class  $j$  in test-data-set

$$DP_i = \frac{\sum_{j=1..N} H_{i,j}}{S} \quad DP_i \in [0,1] \forall i$$

Equation 4 Probability of classifier - decision to class  $i$

$$H = \{H_{i,j} = H(i|j)\}$$

Equation 5 Confusion-Matrix  $H$ ;  $N$ =class complexity; classifier-decision  $i \in [1,N]$ ; reference  $j \in [1,N]$

$OA$  as well as  $KHAT$  relates to one specific pre-classified test-data-set, which consist of  $S$  nearly equal sized segments. So they are suitable for comparisons, while using the same test-data-set.

Because of various land-use-regions disintegrates into segments of different size due to the specific inner class inhomogeneity, the evaluation of classification-results should also regard to the ratio of segment- to total-segments-size. If not, the result of the overall accuracy evaluation will be dominated by the influence of small regions (especially in urban areas). Therefore,  $OA$  and  $KHAT$  will be presented in both ways area-weighted and not weighted.

That ensures comparability between a segment-based and a pixel-based classification strategy.

#### 4. EXAMPLES

The functionality will now be demonstrated using the following examples. All of them classify the scenes with the 9 classes listed in Table 1. The study area is located near Nordhausen in northern Thuringia, Germany. The data basis is represented by an image-stack consisting of 5 channels of one RapidEye-scene and 6 Terra-SAR-X-scenes (with 2 polarizations each).

All TerraSAR-X-channels were preprocessed by geometric filtering [1]. Features being classified are mean segment-specific spectral values of the actual channel. Invalid areas, e.g. disturbed by clouds in the optical scenes, are wiped out by manual masking. (This manual masking will be replaced by an automatic algorithm during the project.)










i	description	legend
0	urban area	
1	coniferous forest	
2	water / wetland	
3	deciduous forest, mixed forest	
4	grassland	
5	summer cereal	
6	rapeseed	
7	maize	
8	root crops	

Table 1 list of all classes with legend and column/row-index of confusion-matrix  $H$

date of data acquisition	instrument	channel
25.05.2009	Rapid-Eye	1 (440-510 nm)
25.05.2009	Rapid-Eye	2 (520-590 nm)
25.05.2009	Rapid-Eye	3 (630-685 nm)
25.05.2009	Rapid-Eye	4 (690-730 nm)
25.05.2009	Rapid-Eye	5 (760-850 nm)
16.07.2009	TerraSAR-X	SM-VV
16.07.2009	TerraSAR-X	SM-HV
29.06.2009	TerraSAR-X	SM-HV
29.06.2009	TerraSAR-X	SM-HH
28.03.2009	TerraSAR-X	SM-HV
28.03.2009	TerraSAR-X	SM-HH
24.06.2009	TerraSAR-X	SM-HV
24.06.2009	TerraSAR-X	SM-HH
21.07.2009	TerraSAR-X	SM-HV
21.07.2009	TerraSAR-X	SM-HH
02.04.2009	TerraSAR-X	SM-HV
02.04.2009	TerraSAR-X	SM-HH

Table 2 overview of used scenes



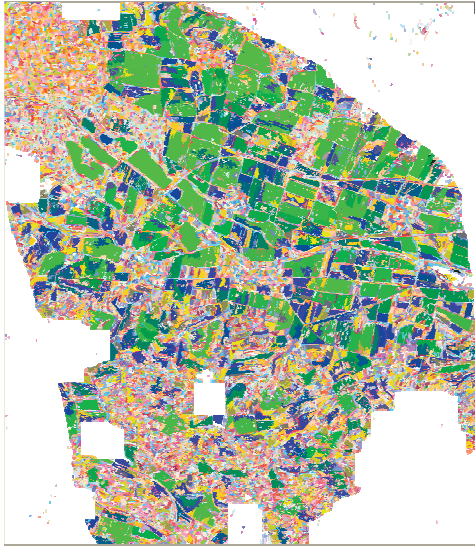


Figure 3 segment-image, each segment is colored individually (clouds are wiped out manually)

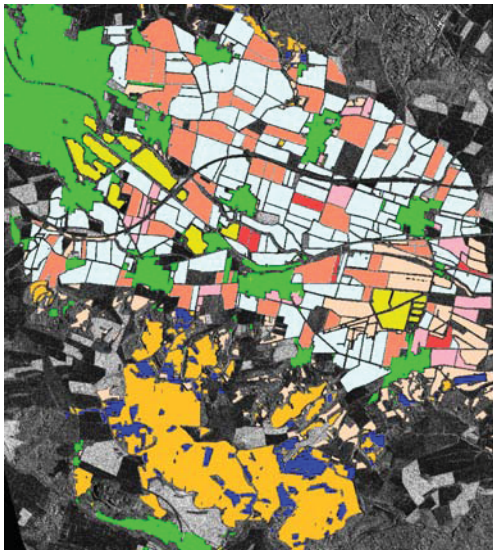


Figure 4 Reference-map (ground-truth) over TS-X-scene (legend: see Table 1)  
underlying TS-X-scene: © DLR, 2009

#### 4.1. multimodal and multi-temporal segment-based classification

Due to time constraints, only the 8000 largest segments are processed by SVM-classification. The comparison between classification-result and ground-truth-map results in the following confusion-matrix:

$$H = \begin{pmatrix} 2155 & 4 & 2 & 4 & 94 & 8 & 0 & 0 & 3 \\ 0 & 47 & 1 & 11 & 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 32 & 0 & 0 & 2 & 0 & 0 & 0 \\ 11 & 65 & 0 & 774 & 34 & 0 & 0 & 0 & 0 \\ 12 & 0 & 3 & 0 & 150 & 2 & 0 & 0 & 0 \\ 8 & 1 & 0 & 0 & 13 & 568 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 3 & 1 & 159 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 & 1 & 0 & 113 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 34 \end{pmatrix}$$

The derived OA and KHAT-measures are:

$$\begin{aligned} \text{OA (normal / area weighted)} &= 0.93 / \mathbf{0.94} \\ \text{KHAT (normal / area weighted)} &= 0.90 / \mathbf{0.93} \end{aligned}$$

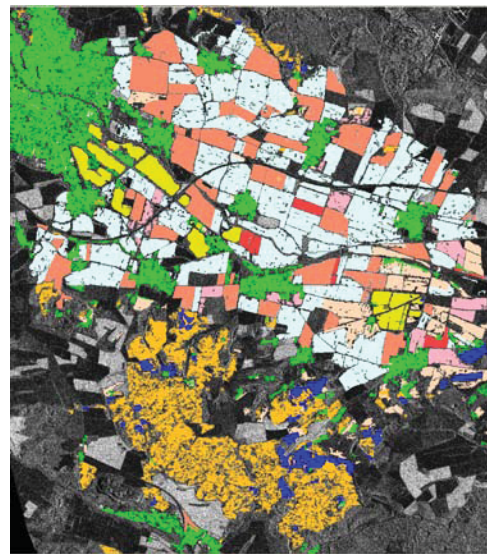


Figure 5 Result of segment-based SVM-classification on TS-X and RE  
underlying TS-X-scene: © DLR, 2009

#### 4.2. multi-temporal segment-based classification

Due to time constraints, only the 8000 largest segments are processed by SVM-classification. The SVM-classification under consideration of the multi-temporal SAR data only results in the following confusion-matrix:

$$H = \begin{pmatrix} 2032 & 32 & 6 & 39 & 124 & 9 & 1 & 0 & 1 \\ 17 & 32 & 0 & 4 & 19 & 4 & 0 & 0 & 0 \\ 1 & 0 & 30 & 0 & 0 & 1 & 0 & 0 & 0 \\ 43 & 54 & 0 & 745 & 17 & 5 & 0 & 0 & 0 \\ 16 & 0 & 1 & 0 & 130 & 8 & 0 & 0 & 0 \\ 8 & 0 & 1 & 0 & 9 & 545 & 0 & 0 & 0 \\ 32 & 0 & 0 & 1 & 2 & 0 & 158 & 0 & 0 \\ 22 & 0 & 0 & 0 & 3 & 10 & 0 & 113 & 0 \\ 23 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 36 \end{pmatrix}$$

Its corresponding quality-measures are:

$$\begin{aligned} \text{OA (normal / area weighted)} &= 0.88 / \mathbf{0.91} \\ \text{KHAT (normal / area weighted)} &= 0.83 / \mathbf{0.89} \end{aligned}$$

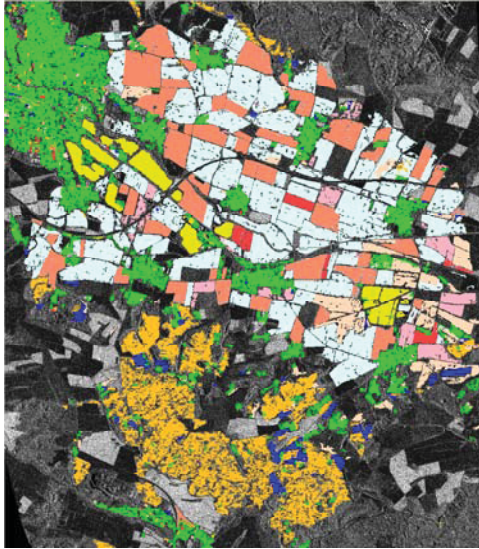


Figure 6 Result of segment-based SVM-classification on TS-X only  
underlying TS-X-scene: ©DLR, 2009

#### 4.3. multimodal and multi-temporal pixel-based classification

The pixel-based SVM-classification was processed based on the multi-modal and multi-temporal image stack mentioned above. All valid (unmasked) pixels with counterpart in the ground-truth-map were respected by evaluation. This results in the following confusion-matrix:

$$H = \begin{pmatrix} 596377 & 5881 & 4752 & 20593 & 58883 & 10922 & 3107 & 2445 & 1328 \\ 1366 & 36837 & 1239 & 12254 & 1554 & 53 & 1 & 0 & 6 \\ 688 & 164 & 110536 & 14 & 5 & 1080 & 0 & 0 & 0 \\ 45774 & 44031 & 286 & 490458 & 22823 & 3388 & 30315 & 214 & 62 \\ 16466 & 126 & 453 & 400 & 143899 & 3977 & 323 & 37 & 1 \\ 5093 & 22 & 36 & 367 & 7331 & 977100 & 520 & 11 & 7 \\ 2553 & 101 & 0 & 954 & 1669 & 2432 & 376802 & 830 & 0 \\ 5813 & 149 & 989 & 8 & 14773 & 1304 & 88 & 87062 & 295 \\ 2725 & 11 & 0 & 3 & 1 & 0 & 8 & 1474 & 26816 \end{pmatrix}$$

The following measures can be derived from it:

$$\begin{aligned} \text{OA (normal / area weighted)} &= \mathbf{0.89} \\ \text{KHAT (normal / area weighted)} &= \mathbf{0.86} \end{aligned}$$

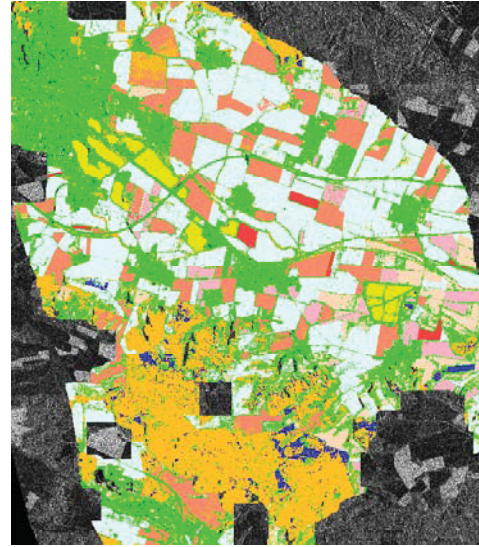


Figure 7 Result of pixel-based SVM-classification on TS-X and RE  
underlying TS-X-scene: ©DLR, 2009

## 5. DISCUSSION

A precise analysis using the quality-measures mentioned above shows that the segment-based concept for segmentation **and** classification works well, albeit class-pair-specific deficits exists (i.e. between *urban area* and *grassland* or *deciduous forest*, *mixed forest* and *coniferous forest*; see all experiments in chapter 4). This may be clarified by the unequal representation of several landcover-classes in the scene (*deciduous forest*, *mixed forest*=18.2% versus *coniferous forest*=2.7% plus *grassland*=7% versus *urban area*=50,6%) as well as the interpenetration of classes in feature domain.

Furthermore the amount of scenes of the temporal data set isn't optimal adapted to the vegetation classes in the case under consideration.

In the examples presented here, an increase in quality (OA: around 3 to 5%; KHAT: around 4 to 7%) was achieved in the case of multimodal classification.

Due to the lack of sufficiently cloud-free optical scenes, the effect of multi-temporal processing of Rapid-Eye-scenes can't be tested. With regard to the phenomenological differences in the backscattering-behavior within the vegetation period this should be evaluated in future.

Compared to the segment-based-approach, the pixel-based classification yield in slightly worse results. Some plots are very conspicuous, whose pixels assigned to different classes (i.e. *rapeseed* to *deciduous forest*, *mixed forest* ). This shows the integrative averaging property of the segment-based approach.

## 6. CONCLUSION

The results presented here, shows the benefit of a segment-based multimodal and multi-temporal SVM-classification compared to a pixel-based approach. Furthermore, it was shown, that processing with just one single modality (in this case: TS-X), results in loss of classification-accuracy (see chapter 4.2). The reason for this is the use of synergies while multimodal processing.

A further optimization may be achieved by the use of training sets from multiple scenes for teaching the classifier.

## 7. ACKNOWLEDGEMENTS

Thanks and Acknowledgements to:

- German Aerospace Center (DLR) and German Federal Ministry of Economy and Technology (BMWi) to support Envidand2;
- RapidEye AG and DLR for the provision of RapidEye data;
- DLR for the provision of TerraSAR-X data
- all project-partners for cooperation:
  - Jena-Optronik GmbH
  - Friedrich-Schiller-Universität Jena
  - Zentrum für Fernerkundung der Landoberfläche (ZFL) - Universität Bonn

## 8. REFERENCES

- [1] Thomas R. Crimmins, "Geometric filter for speckle reduction" APPLIED OPTICS, Vol. 24, No. 10, pp. 1438, 15 May 1985
- [2] R. G. Pontius, Jr., "Quantification error versus location error in the comparison of categorical maps", Photogrammetric Engineering & Remote Sensing 66(8) p.1011-1016, 2000
- [3] Brandt Tso, Paul M. Mather, Classification methods for remote sensed data, CRC press, chapter 2.7, pp. 69-74, 2009
- [4] Vladimir N. Vapnik, The nature of statistical learning theory, Springer, 1995
- [5] R. Kikinis K.Krissian, C.F. Westin and K. Vosburgh, "Oriented speckle reducing anisotropic diffusion", IEEE Trans Image Process. 2007 May;16(5):1412-24. May 2007.
- [6] T. Koch, T. Kubertschak, K.-H. Franke, "Aspekte der Segmentierung von Bildmaterial aus

multispektralen Merkmalsräumen am Beispiel von Satellitendaten", FarbBV2009, Tagungsband, 15. Workshop Farbbildverarbeitung, ISBN 978-3-9812158-4-7, Berlin, 2009

[7] C. Lucht, R. Nestler, K.-H. Franke, "Landnutzungs-klassifikation auf multisensorischen und multi-spektralen Bilddaten - Anforderungen und Lösungsansätze", FarbBV2009, Tagungsband, 15. Workshop Farbbildverarbeitung, ISBN 978-3-9812158-4-7, Berlin, 2009

[8] T. Koch, K.-H. Franke, "Aspekte der strukturbasierten Fusion multimodaler Satellitendaten und der Segmentierung fusionierter Bilder", 52. Internationales Wissenschaftliches Kolloquium, Proceedings, Volume II, ISBN 978-3-939473-17-6, Ilmenau, 2007