HEIDELBERG UNIVERSITY
FACULTY OF CHEMISTRY AND EARTH SCIENCES
INSTITUTE OF GEOGRAPHY

Bachelor´s Thesis

Improving Classification of Very-High-Resolution Satellite Imagery

---

Combining Invariant Support Vector Machines and Object-Based Image Analysis to Tackle Limited Information Input

February 14th, 2017

AUTHOR:
Lukas Blickensdörfer
Blickensdoerfer.lukas@gmail.com

REVIEWERS:

Dr. Susanne Schmidt
Heidelberg University

Dr. Christian Geiß
German Aerospace Center (DLR), Weßling

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre hiermit, dass ich diese Bachelorarbeit selbstständig, ohne die Hilfe Dritter und ohne die Nutzung anderer Quellen und Hilfsmittel als angegebenen, verfasst habe. Alle den genutzten Quellen wörtlich oder sinngemäß entnommenen Stellen, sind als solche einzeln kenntlich gemacht. Gleiches gilt auch für beigegebene Abbildungen und Kartenskizzen.

Diese Arbeit ist bislang keiner anderen Prüfungsbehörde vorgelegt worden und auch nicht veröffentlicht worden.

Heidelberg, 14.02.2017

ABSTRACT

Addressing environmental and socioeconomic challenges in the context of climate change or urbanization, often requires monitoring of large spatial areas. As remote sensing can provide such information, it evolved to be a standard tool to work on related subjects. Image classification often forms the basis for used workflows and derived products. The emergence of new sensor technologies which provide very high spatial and spectral resolution data, made the consideration of objects at finer scales possible and broadened the scope of potential applications of remote sensing. Novel image processing and classification methods such as object-based image analysis and support vector machines, are introduced to effectively exploit the information provided by improved resolutions. Nevertheless, especially for supervised approaches, classification results still depend strongly on the amount and distribution of available ground truth data as information input for training of classifiers. This thesis aims to address the issue by proposing a generic method capable of coping with small ground truth data sets to classify very high spatial resolution data. This is done by transferring invariant support vector machines to the methodology of object-based image analysis. Resulting classifiers appear invariant to scale or geometry representation in ground truth data sets and thus achieve better classification accuracies on limited information input. Experiments on a very-high-resolution image of a complex urban land cover composition – cologne city center – suggest that the proposed method has much scope for future developments. Results show 0.2-0.03 points of κ accuracy improvement (to reach 0.57-0.79) on small ground truth data sets (20 or less training samples per class) compared to a state of the art classification system for a binary classification problem. Although less pronounced, multi-class settings resemble those tendencies. However, in order to ensure general validity of the results, further research is needed.

ZUSAMMENFASSUNG

Aktuelle Umweltprobleme und sozioökonomische Herausforderungen, wie sie im Kontext von Klimawandel oder Urbanisierung auftreten, erfordern häufig die Beobachtung und Untersuchung großflächiger Gebiete. Die Fernerkundung hat sich in diesem Zusammenhang als effiziente Methode bewährt, deren Ergebnisse maßgeblich auf Bildklassifizierungen aufbauen. Durch technische Weiterentwicklungen werden immer höhere räumliche und spektrale Auflösungen erzielt. Gleichzeitig etablieren sich neue Methoden der Bildverarbeitung und Klassifizierung, darunter objektbasierte Verfahren und Support Vector Machines, die den Informationsgehalt hoher Auflösungen nutzbar machen sollen. Dennoch hängen Klassifikationsergebnisse, insbesondere für überwachten Klassifikationsmethoden, stark von verfügbaren Trainingsgebieten und in-situ Daten ab, da diese Referenzdatensätze für das Ableiten von Entscheidungsfunktion des Klassifikators nötig sind. Diese Arbeit stellt eine Methode vor, die auf das Klassifizieren von räumlich hochaufgelösten Satellitenbildern anhand von minimal verfügbaren Testgebieten ausrichtet ist. Objektbasierte Klassifizierung wird dafür mit der Virtual Support Vector Machine in einen Arbeitsablauf integriert. Erzeugte Klassifikatoren werden dadurch invariant zu spezifischen Größen- oder Geometrieeigenschaften von Bildobjekten aus verfügbaren Trainingsgebieten. Dadurch verbessert sich vor allem für Situationen mit limitierter Verfügbarkeit von Testflächen die Qualität des Klassifikationsergebnisses. Die an einer räumlich hochaufgelösten Szene des Stadtzentrums von Köln durchgeführten Experimente liefern vielversprechende Ergebnisse. Verglichen mit einem objektbasierten State-of-the-Art-Klassifikator können für Situationen mit wenigen verwendeten Testgebieten (20 oder weniger Samples pro Klasse) Verbesserungen zwischen 0.2-0.03 Punkte κ Genauigkeit (auf insgesamt 0.57-0.79) eines binären Klassifikationsergebnisses erzielt werden. Diese Tendenzen bestätigen sich auch unter Einbezug mehrerer Klassen, wenn auch weniger ausgeprägt. Um jedoch zu allgemeingültigen Ergebnissen zu kommen, sind weitere Experimente erforderlich.

ACKNOWLEDGEMENTS

I would like to thank Dr. Susanne Schmidt for her support and supervision on this work which includes thematic supervision during the working process but also the provision of a workspace for long-lasting computer calculations at the South Asian Institute of Heidelberg University.

I would like to thank Dr. Christian Geiß for his constant availability for professional advice and willingness to discuss challenges that emerged along the way. My gratitude also goes to Patrick Manuel Aravena Pelizari. Thank you both for giving me the opportunity to work on your idea of combining OBIA and VSVM, during my internship and as basis of this thesis. Also, thank you and the entire DLR team “Modelling and Geostatistical Methods”, at the Department of Geo-Risks and Civil Security, for providing satellite data of Cologne as well as the corresponding ground truth data set.

CONTENT

FIGURES

ABBREVIATIONS

VHR	Very High Resolution
OBIA	Object-Based Image Analysis
SVM	Support Vector Machine
MRIS	Multi-Resolution Image Segmentation
NDVI	Normalized Differenced Vegetation Index
GLCM	Grey-Level Cooccurrence Matrix
SV	Support Vectors
RBF	(Gaussian) Radial Basis Function
VSV	Virtual Support Vectors
VSVM	Virtual Support Vector Machine
MS	Margin Sampling

SYMBOLS USED FOR MODELS AND MODEL COMPONENTS

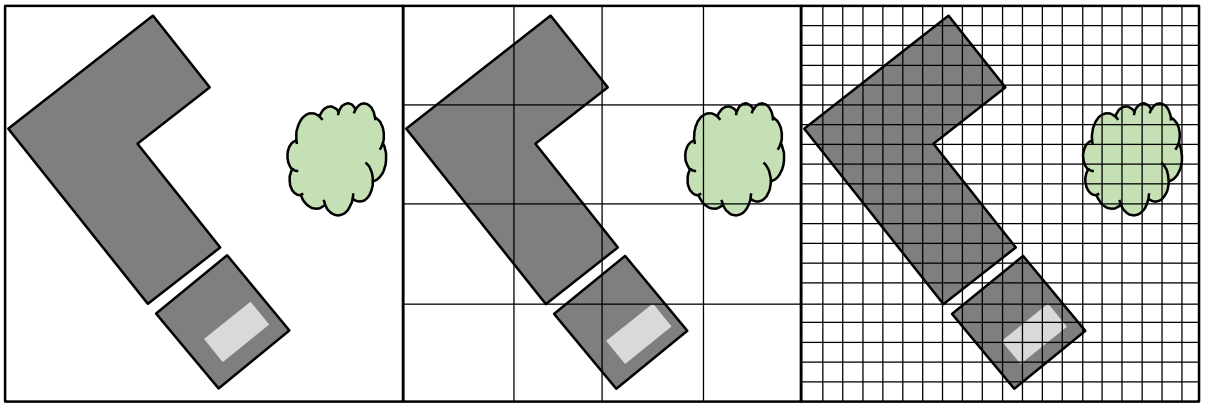
$trainSet$	Ground truth data to train the SVM during hyperparameter optimization.
$testSet$	Ground truth data to test the SVM during hyperparameter optimization.
seg_{base}	Segmentation map on optimal parameters, well modelled image-objects.
$segSet_{scale/geom}$	Segmentation maps encoding scale/geometry variance in image-objects.
SV_{base}	Set of all support vectors of initial SVM.
$pVSV_{scale/geom}$	Set of all potential virtual support vectors encoding scale/geometry variance.
$sVSV_{scale/geom}$	Set of save representative virtual support vectors encoding scale/geometry variance.
$VSV_{scale/geom}$	Set of representative and non-redundant virtual support vectors encoding scale/geometry variance.
SVM_{base}	Initial SVM model on one optimal segmentation map.
$pVSVM_{scale/geom}$	Virtual SVM considering all potential virtual support vectors.
$VSVM_{scale/geom}$	Final virtual SVM invariant to scale/geometry representation.
$SVM_{MLscale/MLgeom}$	Multilevel object-based reference model.

1 INTRODUCTION

Managing challenges of climate change, biodiversity or the complexity of urban environments often requires monitoring of large spatial areas as well as extensive spatial features. (Lu and Weng, 2007: 824f; Alpin and Smith, 2011: 869f). Remote sensing constitutes a pragmatic and cost-effective method to provide this kind of information (Momeni et al., 2016: 1f). It can serve the development of thematic maps at different scales of detail, large-scale estimation of parameters such as population distribution (Schöpfer et al., 2015, Taubenböck and Wurm, 2015) or vulnerability assessment (Stumpf and Kerle, 2011; Geiss et al., 2016a). Image classification often forms the basis for such derivatives (Lu and Weng, 2007: 823).

Remote sensing data typically consists of airborne or satellite images (Richards and Jia, 2006: 1ff; Albertz, 2009: 9ff). Sensors record different electromagnetic wavelength ranges which can lie within the spectrum of human vision, blue, green and red, but which are often supplemented by up to hundreds more for hyperspectral sensors. Each wavelength range is recorded on a so-called spectral image band. The number of these bands determines the spectral resolution. Remote sensing exploits the fact that different materials have characteristic reflection properties in specific wavelength ranges. By means of those characteristic spectral profiles, pixels of different groups can be distinguished and assigned to classes of interest. This method is generally referred to as pixel-based classification (Alpin and Smith, 2011: 870).

The spatial resolution or ground sampling distance is the size of the earth surface covered by one image pixel. For latest sensor systems like IKONOS, QuickBird, GeoEye-1, WorldView-2, or WorldView-3 it can reach up to a few meters or sub meter resolution (Momeni et al., 2016: 2). An object of interest can only be identified in the image domain, if the ground sampling distance is smaller or roughly equals its size (Blaschke, 2010: 3f). Generally, a scene is referred to be *under-sampled* when objects of interest are much smaller than the ground sampling distance, and *over-sampled* when the object of interest is much greater than the ground sampling distance (Fig. 1) (Momeni et al., 2016: 2f). In Very-High-Resolution (VHR) data, objects are most commonly over-sampled. This avoids mixed pixels and makes identification of objects at finer spatial scales possible. It therefore bares potential to produce more accurate classification maps and creates the possibility to investigate on smaller objects of interest. However, new challenges arise, as the highly detailed representation of objects increases intraclass- and reduces interclass heterogeneity (Alpin and Smith, 2011: 870; Momeni et al., 2016: 2f). Typical urban examples are ventilation or air-conditioning systems on rooftops. As they cover several pixels showing spectral profiles that differ from the rest of roof areas, they introduce intraclass variance in a potential roof class. In addition, spectral resolution of VHR sensors is rather limited. Thematic classes thus tend to show similar spectral



(a) (b) (c)

Fig. 1: Relationship of ground sampling distance with object of interest. (a) Objects of interest appear under sampled. (b) Ground sampling distance corresponds roughly the size of the object of interest. (c) Objects of interest appear over sampled. *Adapted from Blaschke (2010: 3).*

profiles if they are of similar materials, e. g. road asphalt and some rooftop materials (Hamedianfat and Shafri, 2015: 3381).

Both these challenges are addressed by the Object-Based Image Analysis (OBIA) approach (Blaschke, 2010: 3ff; Stumpf and Kerle, 2011: 2565; Blaschke et al., 2014: 81ff). By replacing the pixels as element to be classified by pixel groups – so called *image-objects* – intraclass heterogeneity can be reduced (Alpin and Smith, 2011: 870; Momeni et al. 2016: 3). Additionally, indicators such as texture, shape, size, pattern and association derived for each image-object can supplement spectral information and increase interclass differences (Hamedianfat and Shafri 2015: 3381). Besides the shift to image-objects, non-parametric classifiers such as Support Vector Machines (SVM) increasingly substitute the statistical parametric methods like the Maximum Likelihood classifier (Melgani and Bruzzone, 2002, 2004; Foody and Mathur, 2004; Mountrakis et al., 2011: 249ff; Camps-Valls and Bruzzone, 2005, 2009; Salcedo-Sanz et al., 2014). Methods like SVM do not imply statistical assumptions on the data and therefore perform better on rather noisy data of complex environments (Momeni et al., 2016: 3).

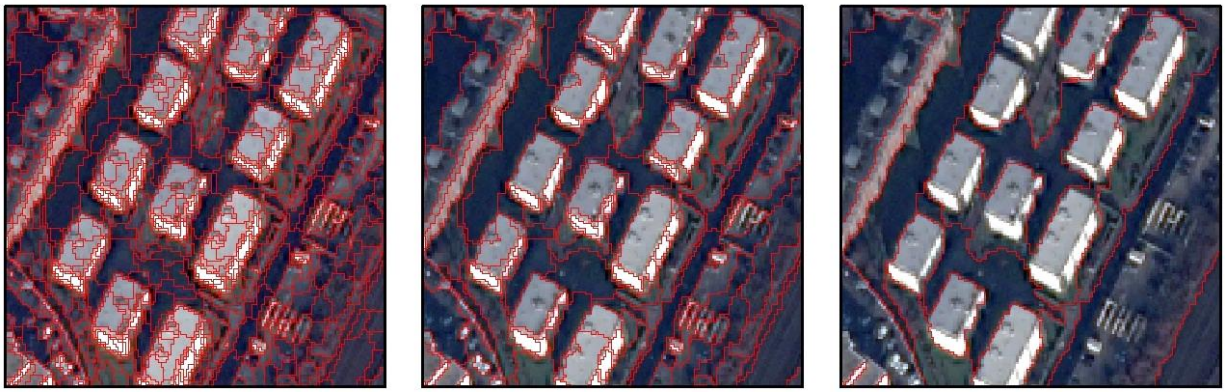
Despite these proceeding concepts, still the quality of classification appears often strongly connected to user interaction (Lu and Weng, 2007: 825). This is especially true for supervised classification, where labeled samples need to be provided for learning a classification model (Fernandez et al., 2014: 4690). The labeled samples, referred to as ground truth data, are often sparse and can be costly as they might need to be generated through manual photointerpretation. Also, they can be limited by external sources, for instance availability of in-situ data (Dópido et al., 2013: 4032; Izquierdo-Verdiguier et al., 2013: 981). When working on complex and heterogeneous areas, gathering sufficient ground truth data can therefore become difficult (Lu and Weng, 2007: 825). This problem is confronted by several strategies.

The semi-supervised classification exploits unlabeled data (Dópido et al., 2013; Li and Zhou, 2015; Lu et al. 2016) while active- and relearning produce intermediate classification results which are reused for enhancement of the thematic map accuracies (Tuia et al., 2009b, 2011; Geiss and Taubenböck, 2015). They improve the performance in numerus settings but can potentially lead to a decrease in performance for other cases (Li and Zhou, 2015: 175). In the field of pattern recognition DeCoste and Schölkopf (2002) introduce so-called invariant SVM to confront related problems in the context of handwritten digit identification. Classification outcome is improved by including artificial samples into model training, which carry variations of characteristics not present in ground truth data but expected to occur in the data to classify. Izquierdo-Verdiguier et al. (2013) successfully transferred this approach to remote sensing using patch-based classification. However, determination of the variances appears challenging and time consuming. Yet, they are substantial to the approach as basis for encoding additional information, adapted to each class- and scene-specific task (Izquierdo-Verdiguier et al., 2013: 982).

Against this background, this thesis introduces a classification strategy to confront presented challenges linked to classification of VHR images and limited ground truth data. It aims on providing assistance where large ground truth data sets are not available or limited through barriers like time or cost effort of in-situ surveys. The approach combines OBIA and invariant SVMs and benefits from characteristic improvements each method implies. Therefore, section 2 introduces OBIA and its principle components, while section 3 presents the theoretical background and application of SVMs in general but also for invariant SVMs specifically. Section 4 gives a detailed explanation of the proposed approach and its functionality which is tested on a set of experiments presented in section 5. The results are summarized in section 6 and potential findings discussed in section 7. This thesis closes with final conclusions and an outlook for further investigation.

2 OBJECT-BASED IMAGE ANALYSIS

The OBIA approach, inspired by the human perception and interpretation of images, is founded on the assumption that an image is composed by interrelated objects of different size and shape (Bruzzone and Carlin, 2006: 2588; Lang, 2008: 6). The idea is to model such image-objects within the image domain to represent objects of interest with high accuracy. Subsequently a classifier labels the whole image-object instead of a single pixel. This bears the advantage that in the classification process spectral information can be supported by the mentioned object indicators describing the texture, shape or size of the image-objects. As this enhances the distinguishability of different land cover classes, better classification results can be achieved. Notably the nature of objects of interest can differ greatly, depending on the aim of the application and scene under investigation. Classification tasks might focus on single building types (Momeni et al., 2016) or tree species (Bunting and Lucas, 2006) while others aim on classifying urban agglomerations (Jacquin et al., 2008) or forest areas (Dorren et al., 2003). Castilla and Hay (2008) define the basic units of OBIA. The 'image-object' is considered as "*a discrete region of a digital image that is internally coherent and different from its surroundings"* (Castilla and Hay, 2008: 94). However, they extend this definition regarding semantic meaning. Seeing the image-object in relationship to the object of interest, there are generally three cases considered: First, oversegmentation, where the object of interest is represented by more than one image segment (Fig. 2a). Second, the ideal case, where the object of interest appears well modelled by the image segment (Fig. 2b) and third, undersegmentation, where more than one object of interest is represented by only one image segment (Fig. 2c) (Liu and Xia, 2010: 187f). For an oversegmented object of interest potentially characteristic information might remain unexploited, however produced information is valid and is likely to contribute to the enhancement of the classification result. This is not the case for undersegmentation. The image segment than carries information of two or more different objects of interest and hence two or more different information classes. If this segment is used



(a) (b) (c)

Fig. 2: Image segments in relation with objects of interest. (a) oversegmentation. (b) well modelled image segments. (3) undersegmentation. *Own figure. Subset of Satellite Image: DigitalGlobe, (2014).*

as entity to be classified and gets assigned to a land cover class, by default some part of the image-object gets misclassified and thereby corrupts the classification result (Liu and Xia, 2010). This consideration leads to the extension of the 'image-object' definition by the constraint that image-objects show barely oversegmentation and no undersegmentation regarding the objects of interest (Castilla and Hay, 2008: 96f).

The next two sections will focus on how those meaningful image-objects are generated through image segmentation and which characteristic indicators, so called features, can be derived from them to enhance distinguishability of different classes.

2.1 Image Segmentation

The image segmentation evolved to an essential part of OBIA as it generates the building blocks of the analysis – the image-objects (Blaschke, 2010: 3; Stumpf and Kerle, 2011: 2567). The purpose of segmentation algorithms is to partition the image into segments by merging neighboring homogenous pixels together and therefore making a differentiation between heterogeneous neighboring regions possible (Schiewe, 2002; Taubenböck et al., 2010: 121). In different segmentation strategies, homogeneity can be defined by different pixel or region properties, like spectral or spatial characteristics. For instance, pixels of water bodies typically show similar spectral properties, especially in the wavelength ranges of near-infrared. A segmentation procedure which gives the spectral characteristics on this band more weight concerning the homogeneity measure, will consequently perform well in modelling lakes or water-ways. In this sense, resulting segments are image regions created by one or several homogeneity criteria in one or several dimensions of feature space (Schiewe, 2002; Blaschke, 2010: 3). The outcome of the procedure must correspond to the stated requirements of image-objects as their information is subsequently used to enhance the image classification (Blaschke, 2010: 4). Segmentation algorithms used in remote sensing applications form four families: point-based, edge-based, region-based and combined approaches (Schiewe, 2002). For a more general overview of basic principles and application of the different methods Schiewe (2002) as well as Dey et al. (2010), among others, provide further readings.

This work makes use of a bottom-up region-merging segmentation algorithm, namely Multi-Resolution Image Segmentation (MRIS), which can be accounted to the region-based approaches (Benz et al., 2004). The algorithm has been applied successfully in numerous studies (Dey et al., 2010: 38). It allows the user to regulate the relative size of generated image segments through the definition of the maximal tolerated heterogeneity within the image segment by setting the so-called scale parameter. Additionally, by the parameters shape and compactness, it is defined to which extend spectral or spatial attributes influence the partition (Fig. 3) (Benz et al., 2004: 246f).

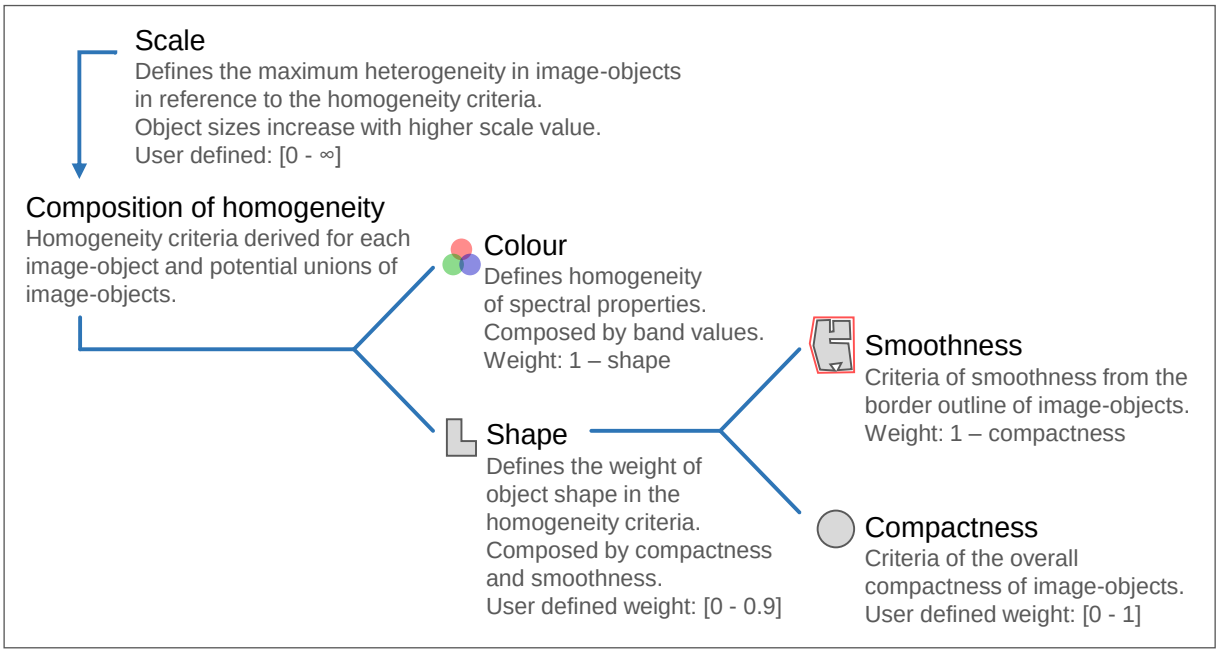


Fig. 3: Concept of Multiresolution Image Segmentation. Composition of object homogeneity and influence of the scale parameter. *Adopted from eCognition (2016: 71).*

The algorithm starts at the pixel level, considering each pixel as a separate image segment. A pairwise clustering process then merges the two segments that produce the minimum global growth of heterogeneity. Step by step this merging process generates larger segments until the smallest growth of heterogeneity surpasses the user defined scale parameter. A formal definition of the procedure is provided by Benz et al. (2004) and Bruzzone and Carlin (2006).

Merging of image segments follows the bottom-up approach, where segments and their borders are never created entirely new but only formed by the union of two already existing segments. This proceeding makes it possible to create not only one segmentation map, but to define a set of scale parameters and create a set of unambiguous hierarchically related segmentation maps for the same image. They satisfy the constraint

$$\bigcup_{O_i^s \subseteq O_j^{s+1}} O_i^s = O_j^{s+1} \tag{1}$$

where at the segmentation level s the image is subdivided in N^s objects O_i^s $(i = 1, 2, \dots, N^s)$ (Bruzzone and Carlin, 2006: 2590; Geiss and Taubenböck, 2015: 2337). In other words, the constraint guarantees that any object at level $s-1$ cannot be part of more than one object in level s. This modelling of hierarchical relation between objects of different segmentation levels makes it possible to not only substitute the pixel as image analysis element with the object, but to add object information from different levels to hierarchically related analysis elements. The feature vector of the image-object in the lowest level gets extended by feature vectors of the objects enclosing it at higher levels. Additionally, further object indicators concerning pattern

and association become available. Bruzzone and Carlin (2006: 2588) state that by considering such hierarchical sets of segmentation maps in the classification system, it is possible to analyze each real-world object at its optimal representation level. Also, it takes into account that objects are logically interrelated within the same level and hierarchically related to those in higher or lower levels. This concept is referred to as multilevel object-based classification.

2.2 Image-Object Features

Object features, also referred to as object metrics, carry the information that describes each object and builds the foundation to determine the classification rules. They construct the so-called feature space. The term refers to a space, where each feature measure is considered a dimension. If objects are, for example, characterized by only three features, the pixel values blue, green and red, the feature space is 3-dimentional and each object has its definite position within it – specified by the coordinates blue, green and red. Therefore, the number of considered features defines the dimensions of feature space.

Combining aspects like shape, size, pattern, tone, texture, shadow and association makes object identification by human vision possible (Olson, 1960; Blaschke et al., 2014: 182). In this sense, a diverse pool of object features arose concerning spectral and geometry-related properties, including texture, but also encoding topological information like neighborhood and hierarchical relation (Bruzzone and Carlin, 2006: 2591; Blaschke, 2010:10; Geiss et al., 2016b: 5952). In the following, the focus is on different feature types that encode spectral, textural and spatial characteristics. After a brief introduction for each group, the measures selected for this study are discussed in more detail.

2.2.1 Spectral

Spectral information refers to the reflection behavior of different real-world objects. Therefore, spectral information can be exploited by analyzing directly the numerical values of each available spectral band (Bruzzone and Carlin, 2006: 2591) and additionally calculated band ratios like the Normalized Differenced Vegetation Index (NDVI, $\frac{(near\ infrared-red)}{(near\ infrared+red)}$) (Rouse et al., 1973). For this study, the statistical measures of central tendency and spread, *mean* and *standard deviation,* of pixels included in an image-object are calculated for each spectral band and the stacked NDVI correspondingly and used for spectral characterization of the image-object.

2.2.2 Texture

Texture refers to the frequency of tone variance, i.e. spectral band values and the spatial arrangements of those variances (Hay and Niemann, 1994; Pacifici et al., 2009: 1277; Blaschke et al., 2014: 183). Especially in scenarios with limited spectral resolution, as it is the case for VHR or panchromatic imagery, it has been demonstrated that the use of textural features has great potential for the improvement in classification accuracies (Zhang et al., 2003; Carleer and Wolff, 2006; Laliberte and Rango, 2009; Pacifici et al., 2009).

The Grey-Level Co-occurrence Matrix (GLCM) method (Haralick et al., 1973; Haralick, 1979) is well established for texture characterization in remote sensing and successfully used in many studies (Zhang et al., 2003; Pacifici et al., 2009; Stumpf and Kerle, 2011). All GLCM measures are based on a symmetric matrix, which counts grey-level co-occurrences of directly neighboring pixels for all pixels within the image-object. This approach makes it possible to recognize specific patterns or arrangements of reflectance intensity. By summarizing those matrices using different functions, a variety of texture measures for the image-object can be created. In order to keep the feature set compact, but also in consideration of the high computational burden and strong correlations between many GLCM measures (Cossu, 1988; Laliberte and Rango, 2009), the set of GLCM textural features for this study is limited to *homogeneity*, *dissimilarity* and *mean*. *Homogeneity* returns high values whenever an object shows low variance in its pixel grey-levels, hence is high for homogeneous objects (Pacifici et al., 2009: 1281). *Dissimilarity* is a measure of contrast within the image-object, while *mean* is the mean of the co-occurrence matrix and thus the mean co-occurrence of all present grey-levels. Their formal definitions are provided in the following according to the implementation in the used software eCognition (eCognition, 2016: 412ff):

$$homogeneity = \sum_{i,j=0}^{G-1} \frac{P_{i,j}}{1 + (i-j)^2}, \tag{2}$$

$$\text{dissimilarity} = \sum_{i,j=0}^{G-1} P_{i,j}|i-j|, \tag{3}$$

$$mean = \frac{\sum_{i,j=0}^{G-1} P_{i,j}}{G^2}. \tag{4}$$

G is the number of grey-level values present in the image-object and therefore the amount of rows or columns in the corresponding co-occurrence matrix. i and j are the coordinates in the matrix and $P_{i,j}$ is the normalizes value at the position i, j within the matrix, hence the normalized frequency of the co-occurrence of grey-level pair i, j within the image-object.

2.2.3 Shape

Shape features are measures that describe the outline or general form of the individual image-object (Blaschke et al., 2014: 182). Especially in applications aiming to detect man-made objects, like the classification of urban environments, this group has great potential to improve classification results (Sun et al., 2015: 3737). This is due to the fact that man-made objects often show regular boundaries, which are rarely found with natural objects. For this study, shape related properties are therefore considered by the indices *rectangular fit*, *elliptic fit*, *roundness*, *shape index* and *compactness*. The *rectangular fit* and *elliptic fit* are based on the same principle:

$$rectangular\ fit = \frac{A_R}{A_O}, \tag{5}$$

$$elliptic\ fit = \frac{A_E}{A_O}, \tag{6}$$

with A_O being the area of the image-object and A_R its intersection with a rectangle having identical area and its proportion rotated to the original objects shape moments. A_E is the intersection with an ellipse having similar properties respectively (Sun et al., 2015: 3738ff). *Roundness* is calculated by

$$roundness = \varepsilon^{max} - \varepsilon^{min} \tag{7}$$

where ε^{max} is the radius of the smallest, the image-object enclosing ellipse and ε^{min} the radius of the largest, by the image-object enclosed ellipse. *Shape index* describes the smoothness of the image-objects borders and is mathematically expressed as

$$shape\ index = \frac{L_O}{\sqrt[4]{A_O}} \tag{8}$$

where L_O is the border length of the image-object. *Compactness* is defined by

$$compactness = \frac{4 * \pi * A_O}{{L_O}^2} \tag{9}$$

with L_O being the perimeter of the image-object. Therefore, for the most compact shape – the circle – returns the highest value: one.

Sun et al. (2015) point out that the extraction of shape information from image-objects can be improved in reference to these traditional shape features. Blaschke et al. (2014: 182) however argue that shape features may often not be inherently distinctive, but can appear as important factor in diverse composed feature sets.

3 Support Vector Machine Classification

In this section, the SVM algorithm is briefly reviewed and its principal functionality is discussed as the method proposed in chapter 4 builds on some SVM specific properties. However, for a more extensive, general introduction it is referred to Cortes and Vapnik (1995), Burges (1998), Vapnik (1998), Schölkopf and Smola (2002) and Foody and Mathur (2004), while remote sensing specific literature is provided by Melgani and Bruzzone (2004), Camps-Valls and Bruzzone (2005, 2009), Mountrakis et al. (2011) and Salcedo-Sanz et al. (2014).

3.1 Theoretical Background

The Support Vector Machine, first introduced by Cortes and Vapnik (1995) in the field of machine learning, constitutes a group of non-parametric supervised classification and regression approaches.

For a binary supervised classification problem, the SVM seeks to find the optimal decision surface, also called optimal separating hyperplane, which separates the instances of the two classes in feature space (Burges, 1998). As input, labeled samples for each class need to be provided. On the basis of this set of labeled samples, the SVM is learned. A great number of hyperplanes might be able to separate the samples of the two classes (Fig. 4a). However, by choosing the one that lies in the center of the widest sample free corridor, it is assumed that the separation is more likely to be valid for unseen data points. In this sense the optimal decision surface is the one that maximizes a margin formed by two additional surfaces lying parallel to the decision surface and intersecting the samples closest to it (Fig.4b). These samples lying on the margin border are called Support Vectors (SV) (Burges, 1998; Leinenkugel et al., 2011; Geiss et al., 2016a: 1922f). Unlabeled samples are located at either side of the optimal separating hyperplane and can thereby be labeled correspondingly

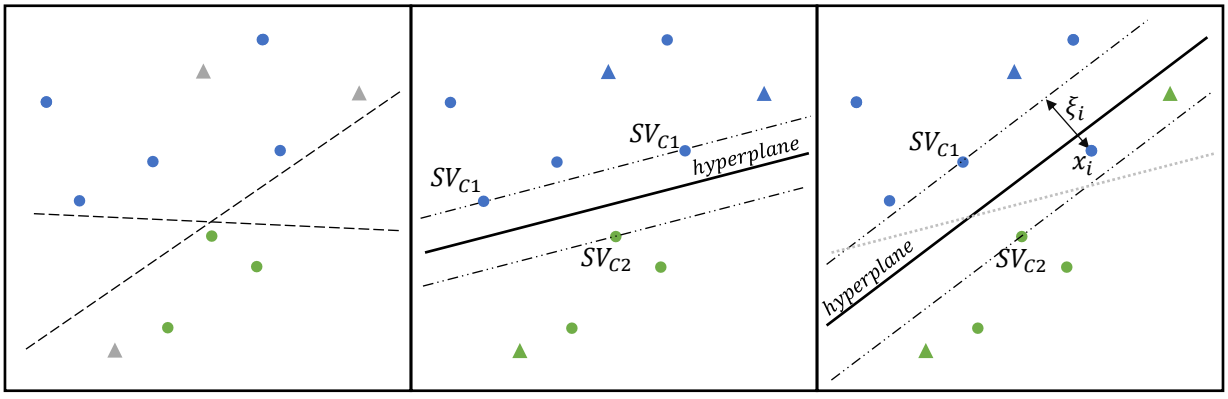


Fig. 4: Idealized process of locating the optimal separating hyperplane. Dots represent labelled training data of two classes (blue and green). Triangles represent new unseen instances (grey: unclassified, blue/green: classified correspondingly). (a) Available labelled instances for training. (b) Separation of instances by the hyperplane of the SVM. (c) Adaption to soft margin SVM, e.g. C-SVM. *Adapted from Melgani and Bruzzone (2004: 1781); Izquierdo-Verdiguier et al. (2013: 982).*

(Fig. 4b). It is important to stress that the SVs do carry all the information needed to define the optimal hyperplane and hence to build the decision function. Robust models which generalize well can therefore already be derived from relatively small training set sizes (Geiss et al., 2016a: 1922f).

Considering a set of labeled training samples $S = \{X, Y\}$, where $X = \{x_l\}_{l=1}^{n} \in \mathbb{R}^d$ are n d-dimensional feature vectors associated with the labels $Y = \{y_l\}_{l=1}^{n} \in \{-1, 1\}$, suitable parameters are found to define the optimal separating hyperplane during training of the SVM (Camps-Valls and Bruzzone, 2005: 1353; Geiss et al., 2016a: 1930ff). As data is rarely linear separable, meaning an accurate linear separation by a hyperplane of information classes is not possible, a nonlinear transformation $\phi(\cdot)$ maps the labeled samples from the original feature space $\mathcal{X}$ into a space of higher dimensionality $\mathcal{H}$. This is helpful as linear separation in $\mathcal{H}$ can be easier to achieve and matches a potentially high complex nonlinear separation in $\mathcal{X}$ (Fig. 5). In this sense, mapping by an appropriate $\phi(\cdot)$ makes instances of the two classes more likely to appear separable. Yet, if the data to classify is very noisy or classes appear to have great overlap in feature space, it can be convenient to allow misclassification of some samples with the aim of widening the margin and the generalization capacity. This is implemented by a so-called slack variable ξ and a penalty factor C (Fig.4c) (Cortes and Vapnik, 1995: 280ff; Geiss et al., 2016a: 1931f). In these cases, literature speaks of soft margin classifiers implemented as C-SVM because the margin is “softened” as it allows samples to appear on the wrong side of the hyperplane, but each misclassification gets penalized by a defined factor C. In other words: if factor C is chosen very large the model tempts to overfit on training data, because misclassification is highly penalized and therefore rare. While in the contrary case, the model may allow too many misclassified samples which cause underfitting (Ben-Hur and Weston, 2010: 7ff). As both cases produce poor results on unseen data, the

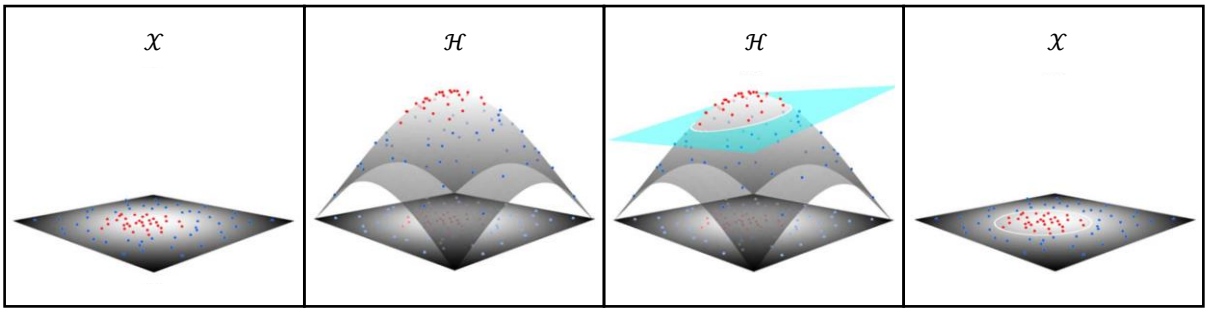


Fig. 5: Idealized process of defining a nonlinear decision function by SVMs. (a) Samples of two classes, red and blue dots, that are not linear separable in $\mathcal{X}$. (b) Samples are mapped by the nonlinear transformation $\phi(\cdot)$ to the higher dimensional space $\mathcal{H}$. This makes linear separation by a hyperplane (cyan) possible, which is fitted by maximising the margin (c). In input space $\mathcal{X}$ this separation corresponds a nonlinear decision function (d). *From Geiss et al. (2016a: 1931).*

penalty factor C must be determined carefully. On this foundation, the method delivers a decision function optimized on the training data in the form of:

$$f(x_*) = sgn\left(\sum_{i=1}^{n} y_i\, \alpha_i K(x_i, x_*) + b\right) \tag{10}$$

with x_* being an instance of unidentified class membership, x_i being the ith of n support vectors with corresponding label y_i and support vector coefficient α_i. Those coefficients and the bias of the hyperplane b are defined during the optimization process, hence the training of the classifier on the labeled samples, and encode also the influence of C on the decision function (Cortes and Vapnik, 1995; Camps-Valls and Bruzzone, 2005; Gehler and Schölkopf, 2009). The decision function $f(x_*)$ depends on the underlying data through the dot product of the mapped instances. The dot product can be replaced by a kernel function $K(x_i, x_j) = \phi(x_i), \phi(x_j)$ that returns the similar result and thereby the explicit calculation of any mapping $\phi(\cdot)$ can be avoided. This is known as the kernel trick (Gehler and Schölkopf, 2009: 27). This property allows the SVM algorithm to efficiently compute the decision function also for high dimensional data, as costly mapping by $\phi(\cdot)$ can be spared. While there are several different kernels with a variety of characteristics, the one commonly used for environmental applications is the Gaussian Radial Basis Function (RBF) kernel which takes the form of

$$K(x_i, x_j) = \exp\left(-\gamma \left|\left|x_i - x_j\right|\right|^2\right) \tag{11}$$

(Bruzzone and Carlin, 2006: 2592; Volpi et al, 2013: 80). $\gamma > 0$ thereby controls the width of the Gaussian. In this sense, γ determines the flexibility of the SVM in fitting on the training data. A high γ value can be seen as great flexibility and therefore 'tight' fitting on labeled samples, which in consequence can lead to overfitting. But for a too low γ value the SVM might not be flexible enough to model a complex classification setting (Ben-Hur and Weston, 2010: 7ff). Again, the parameter needs to be set according to the classification problem. Consequently, regarding the application of the C-SVM in combination with the RBF kernel, it is essential for good performance – meaning good generalization on unseen data – that the two hyperparameters C and γ are carefully chosen.

In remote sensing, classification tasks often consider a variety of land cover classes. The SVM, originally designed to solve binary classification, therefore needs to be effectively extended to multiclass problems. The one-against-one approach is a suitable strategy to solve this problem (Hsu and Lin, 2002: 425). For k information classes a total number of $k(k-1)/2$ binary classifiers are trained, representing each possible binary class combination. A label of an unclassified sample is then predicted by each classifier and subsequently keeps the label with the highest count of assignments (Hsu and Lin, 2002; Foody and Mathur, 2004: 1337).

3.2 SVMs in Remote Sensing

Mountrakis et al. (2011) point out benefits and challenges of SVMs in remote sensing: A key characteristic is the relatively high classification accuracy on small training data sets, compared two traditional methods (Mantero et al., 2005; Mountrakis et al., 2011: 248). In various settings, the limited number of training instances is combined with a very high dimensionality of the same. This is especially true for hyperspectral datasets (Melgani and Bruzzone, 2004; Camps-Valls and Bruzzone, 2005), but also applies for OBIA (Bruzzone and Carlin, 2006; Tuia et al., 2009a; Cánovas-García and Alonso-Sarría, 2015). In these settings, classification results are prone to suffer the so-called Hughes phenomena (Hughes, 1968). It describes the problem of training sets in which the feature dimensionality is much greater than the number of samples. This typically leads to a decrease in accuracy with increasing feature dimensionality. However, the SVM classifier has shown relatively high robustness to the Hughes phenomena and hence, feature reduction analyses which is needed in other approaches, can be spared (Melgani and Bruzzone, 2004: 1779; Gualtieri, 2009). SVMs are non-parametric classifiers, meaning there is no assumption made on the statistical distribution of the underlying data and all function parameters are derived in connection with provided training data. Burges (1998) demonstrated that, since the distribution of the remote sensing data is usually unknown (Mountrakis et al., 2011: 248), this can be an advantage over parametric classifiers, like the Maximum Likelihood Estimation. Another advantage quoted by Mountrakis et al. (2011) concerns the problem of overfitting on the training data. The method has shown good balancing of accuracy achieved on the training patterns and the capacity to generalize well on unseen instances. Opposed to these advantages, there are several barriers that hinder the application of SVMs. The greatest involve the choice of the used kernel as well as a deeper understanding of many model variants and adaptations for more specific tasks (Mountrakis et al., 2011: 248). To tackle such difficulties a wide range of SVM tutorials is provided in literature (Cortes and Vapnik, 1995; Burges, 1998; Ben-Hur and Weston, 2010).

Nevertheless, SVM has become a standard tool in processing and classification of remote sensing data (Izquierdo-Verdiguier et al., 2013: 981). Relevant to this thesis, is the effectiveness of the method in combination with on one side OBIA, on the other semi-supervised approaches that deal with artificial samples to overcome restrictions of sparse ground truth data on VHR imagery.

Concerning the first, Momeni et al. (2016) performed an extensive experiment comparing different spatial resolutions, classifiers and feature sets for a complex land cover classification task. They recommend a combination of OBIA with SVMs for spatially high resolution settings. Geiss et al. (2016a) report good results by combining OBIA with SVM for a multisource approach based on multispectral, elevation, spatial-temporal and in-situ data for seismic

vulnerability assessment. Fernandez et al. (2014) found that the SVM and Nearest Neighbor method produce the most accurate and robust classification of impervious surface areas by means of OBIA. Yet, it is stated that distribution and size of the training data sets still play a key role in classification results (Foody and Mathur, 2004: 1340; Fernandez et al., 2014: 4690). In the field of semi-supervised classification Bruzzone et al. (2006), Gómez-Chova et al. (2011) and Li and Zhou (2015) modify the basic SVM formulation to deal with unlabeled data, while Dópido et al. (2013) and Lu et al. (2016) include the principal of self-learning in their approaches, where the most informative unlabeled samples are selected by the machine learning algorithm itself. Li and Zhou (2015) also point out that in some cases semi-supervised learning methods appear to show worse performances as supervised classification. Their work focuses on this issue and proposes SVM variants of higher reliability while handling unlabeled data. An alternative approach to sparse training data is to incorporate prior knowledge of the data representation into the classification model, namely in form of invariant SVM.

3.3 Invariances in SVMs

The term ‘invariance’ refers to a specific property of a mathematical function. An algorithm implementing such a function is called ‘invariant’. It means that the algorithm, let it be a classifier, is robust to changes in the data representation (Izquierdo-Verdiguier et al., 2013: 981). To give a simple example, a classifier with the task to detect buildings is considered. Let the training data only include instances of buildings with quadratic shape. However, it is desirable that also buildings with a rectangular shape are recognized, as this might occur in the data to classify. If this prior knowledge about changes within the data representation, is incorporated in the classifier and enables it to manage the classification task, the classifier could be considered invariant to the shape of the buildings. Originally, encoding invariance in SVM classifiers was presented by Schölkopf et al. (1996), DeCoste and Schölkopf (2002) and Chapelle and Schölkopf (2002). DeCoste and Schölkopf (2002) show the effectiveness of two main approaches on handwritten digit recognition, but also by the identification of volcanos in preselected satellite image patches. While one exploits engineering of kernel functions that results in invariant SVMs, the second focuses on the training data. Due to its effectiveness and simplicity, this work concentrates on the latter approach. DeCoste and Schölkopf (2002) state that it is essential to have access to prior knowledge about desired invariances or in other words, the variance that might occur in data. On the basis of this knowledge, instances from the training set can be transformed to encode desired characteristics and reincluded in model training. Such a model trained on transformed and original samples should appear invariant to the included characteristics. For the given example, instances that represent quadratic houses are modified to rectangular shape and reincluded in the model. The resulting classifier, trained on original quadratic and artificial rectangular houses, will then recognize both shapes during classification and therefore improve the quality of classification. As the calculated decision

function of a SVM only relies on the SVs, it is sufficient to exclusively apply the transformation of instances to the model SVs. Artificial samples, resulting out of the transformation process, are called Virtual SVs (VSVs). The invariant SVM, trained on SVs of the initial model and generated VSVs, is referred to as Virtual SVM (VSVM). In summary, this leads to the following general process (DeCoste and Schölkopf, 2002:165):

1. train an initial SVM on some training set and extract its SVs (Fig. 6a, b)
2. transform SVs to VSVs so that they carry the information to which the model shall be invariant
3. train a second VSVM on the set of SVs and VSVs (Fig. 6c)

VSVs should appear close to the separating hyperplane as they derive from samples lying on the margin borders. Therefore, they are likely to become SVs during the second training and shift the hyperplane to be robust to the encoded change in data representation (Fig. 6c) (DeCoste and Schölkopf, 2002).

Izquierdo-Verdiguier et al. (2013) transfer this approach to patch-based classification for VHR remote sensing data. They encode invariance to rotation and scale of real-world objects as well as to appearance of shadow in the image. VSVs are generated by rotating or up- and downscaling the image patches and by remodeling reflectance values for samples of the shadow class. All three experiments showed the effectiveness of the method with an improvement of 5-7% in κ accuracy. However, the authors note that prior knowledge of variances that may appear in unseen data and the appropriate encoding of this knowledge into VSVs is of paramount importance.

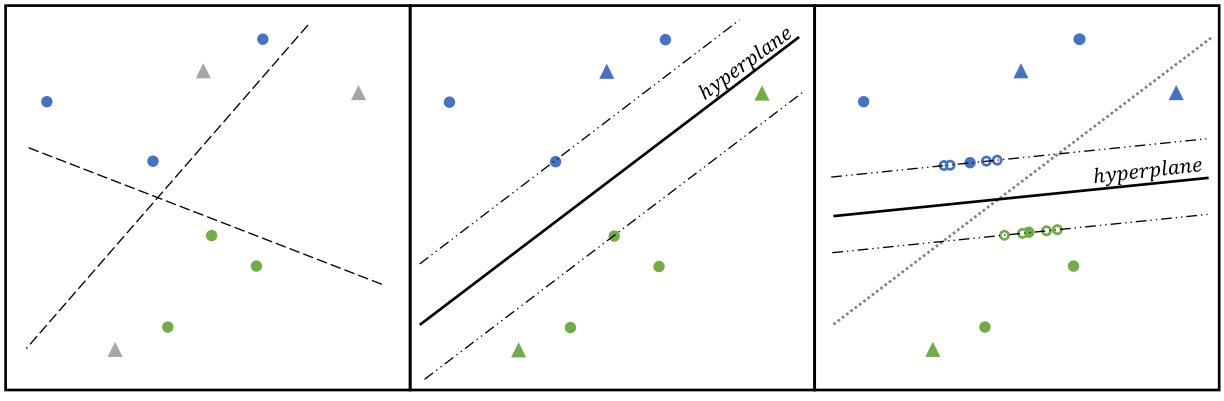


Fig. 6: Simplified functionality of VSVMs. Dots represent labelled training data of two classes (blue /green). Triangles represent new unseen instances (grey: unclassified, blue/green: classified). Circles represent VSVs of each class (blue/green). (a) Available labelled instances for training. (b) Initial trained SVM model. (c) VSVM model with VSV and shifted optimal separating hyperplane. *Adapted from Izquierdo-Verdiguier et al. (2013: 982).*

4 Invariances in supervised Object Based Classification

In aforementioned approaches for VSVMs, the encoding of prior knowledge is realized by strong interaction with the expert user. Next to the identification of the desired invariance, the degree of its manifestation needs to be set. Izquierdo-Verdiguier et al. (2013) defines a realistic scale variance of tree sizes from 50-120% of the labeled samples and encodes those values in the up- and downscaling of the image-patches. It means that prior knowledge needs to be existent or generated by the expert user and implemented according to each setting.

On this basis, in this section the principal structure is introduced that combines OBIA with VSVM classification. It is proposed in a way that is generally applicable to different classification tasks and data sources, if feature sets and information classes are adapted. Redetermination and reimplementation of possible invariance characteristics for each classification task are not necessary. While the invariance type is pre-set to scale and geometry, definition and encoding of the actual range of variance in data is replaced by the use of image segmentation. While section 4.1 specifies understanding of scale and geometry invariance in this work, section 4.2 presents the actual implementation.

4.1 Invariance in Scale and Geometry

Considering OBIA in complex, heterogeneous scenes, and for the case of sparse ground truth data, it is probable that some classes are only represented with a subset of its existing object sizes in the training set. Potential reasons include a limited number of labeled instances or over- and undersegmentation within the same segmentation map. Both conditions make optimal representation of all real-world objects impossible (Schiewe 2002: 3). Especially in an urban environment, where man-made objects can differ significantly in size – from a parking lot of a detached house to a city square (potential impervious surfaces land cover class) or from a single tree to a city park complex (potential vegetation land cover class) – scale invariance is expected to improve classification results. It means that these different scale variation of real-world objects are recognized by the classifier even if they are not present in available ground truth data. The same accounts for the geometry of objects. Houses can build stretched forms if they are built in closed rows or compact forms for single housing. Impervious surfaces are formed by stretched or compact shapes, from long and slim roadways to more compact polygons of open places. Also for scenes with unbalanced geometry characteristics – in which some classes show great geometric variety, while others are rather homogenous, geometry invariance implemented for each class accordingly is expected to enhance results.

In OBIA, segmentation determines the image-objects and therefore scale and geometry attributes. A logical consequence is to approach variations of scale and geometry by using a segmentation procedure. Through a variety of parameter settings for the segmentation

algorithm, a collection of scale and geometry representations of different objects can be achieved and used to encode invariance.

4.2 Proposed Methodology

The procedure follows the principal steps of VSVMs introduced in section 3.3. Those are transferred to the remote sensing context as follows (Fig. 7): A typical OBIA segmentation procedure (MRIS) partitions the image domain in well modelled image-objects (section 4.2.2) for the training of an initial SVM (section 4.2.3). On the base of its SVs and by the use of segmentation maps created additionally by modifying the parameter of the MRIS used before, VSVs are created (section 4.2.4). The VSVs are evaluated by their quality of information to select only representative and non-redundant VSVs for the VSVM (section 4.2.5). Finally, the VSVM is trained and used for the classification of unseen data (section 4.2.6). For a good classification result of the SVMs, a strict separation of the labeled samples in training and testing data is necessary, which is explained and presented in section 4.2.1.

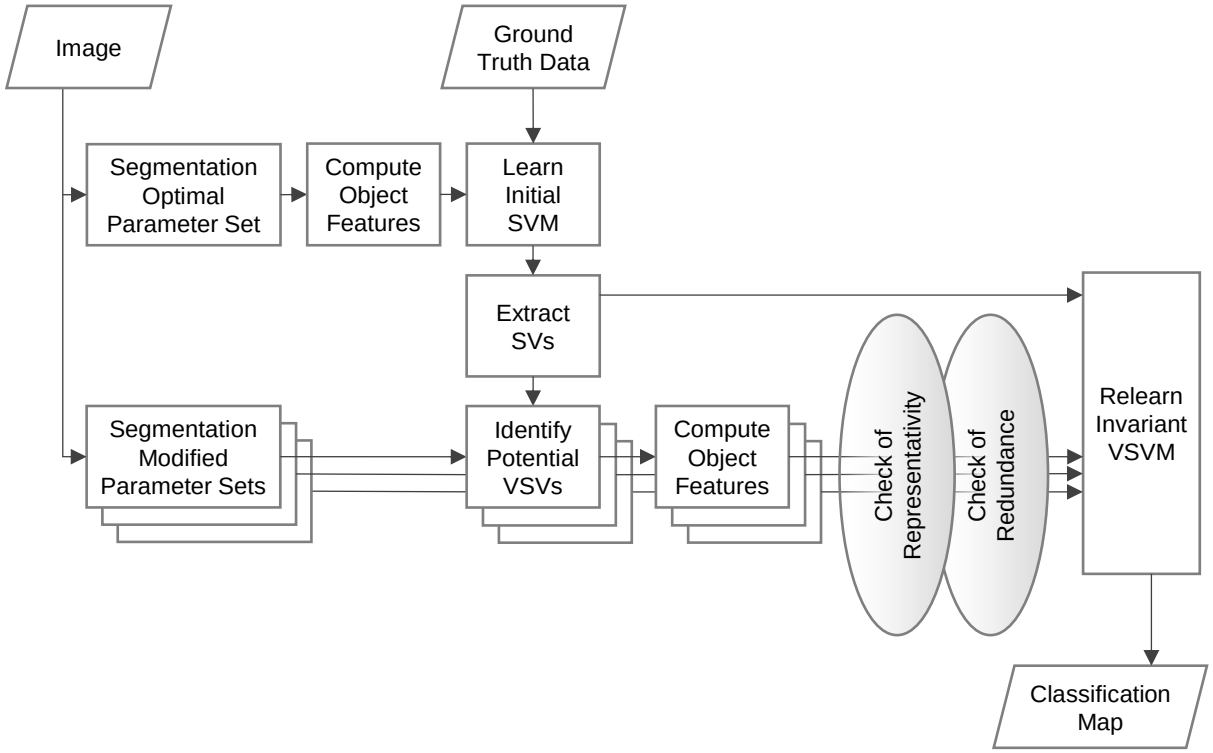


Fig. 7: Block scheme of the proposed method. *Own figure.*

4.2.1 Ground Truth Data Sets

To learn the most accurate C-SVM with a RBF kernel the cost-parameter C as well as the kernel-width parameter γ need to be defined. Generally, the optimization of this parameter combination can be solved by a grid-search strategy (Foody, 2009: 90). This means that for C and for γ sequences of potential parameter values are defined. From these sequences, all parameter combinations are used to train a classifier on a subset of available labeled samples, while the remaining samples simulate unseen data to be classified and to evaluate

performance of parameter combination. The best performing parameter combination qualifies for the classification task and is used to train the final SVM. As it is desirable to use all the information carried by labeled samples, cross-validation is frequently and successfully applied in this step (Foody and Mathur, 2004; Tzotsos and Argialas, 2008; Foody, 2009; Gehler and Schölkopf, 2009: 38; Izquierdo-Verdiguier et al., 2013; Geiss et al., 2016a; Hsu et al., 2016). However, for this VSVM the number of VSVs can easily exceed the number of original samples by its multiple. This entails the danger of strong dominance of VSVs in the parameter setting process and therefore the risk of overfitting the model on the artificial samples, while the original SVs lose on influence. Additionally, VSVs are likely to show feature characteristics that resemble their original SV. Hence, a consistent separation and simulation of unseen data might be violated when using cross-validation on a sample set that contains SVs from the initial model and their VSV derivations. In other applications, similar problems are known as data leakage. As a result, the classifier produces high accuracies during the iterations of cross-validation, but performs poorly on unseen data.

This is avoided by using the hold-out method (Foody, 2009: 90) which implies splitting labeled samples into two spatially disjunct subsets. During grid-search procedure, the first subset is used for training and can be enriched by VSVs. The second subset consists of original samples and is used to estimate the accuracy measure for each parameter combination of C and γ. In the following, those distinct sets are referred to as $trainSet$ and $testSet$.

4.2.2 Segmentation Procedure

The image segmentation used for the initial SVM requires a manual setting of parameters of the MRIS algorithm (scale, shape and compactness; Fig. 3). These are adjusted in reference to the criteria for image-objects presented in section 2 concerning over- and undersegmentation. Image-objects of this segmentation map will serve as the basis of the subsequent procedure. Therefore, the segmentation map will be referred to as seg_{base}.

For the purpose of introducing invariance, two additional sets of segmentation maps are derived from seg_{base}. They are denoted by $segSet_{scale}$ and $segSet_{geom}$ for scale and geometry invariance respectively. Selected objects of those segmentation maps will serve as VSVs in the invariant model. $segSet_{scale}$ is generated by keeping parameters of shape and compactness constant while altering the parameter of scale. It means that the weights for shape/color and compactness/smoothness, which define the composition of the heterogeneity in the image segments, are kept constant. Thereby their geometry is defined similar to the geometry of segments in seg_{base}. However, the scale parameter defining the threshold for the composed maximum heterogeneity is altered, which influences the size of resulting segments. The sizes of image segments generated by different scale parameters may generally move between an upper and a lower bound. The latter fulfils good representation of the smallest and

most homogeneous real-world objects while accepting oversegmentation for others (typically small scale parameter). The upper bound is a good representation of the largest and very heterogeneous real-world objects while accepting undersegmentation for others (typically large scale parameter). $segSet_{geom}$ is generated by modification of the shape and compactness parameter, while keeping the scale parameter of seg_{base} constant. This leads to segments of roughly the same size but different geometries compared to those of seg_{base}. The variation of shape and compactness may be realized up to the point where generated segments lose the relation to outlines of real-world objects.

The aim of defining those bounds quite broad is to encode the entire spectrum of object-scale or object-geometry variety present in data for different land cover classes into image segments. As those image segments are used to make the classifier invariant, this appears crucial to the approach. The presence of over- and undersegmented objects is accepted, because the optimization procedure (introduced in section 4.2.5) chooses only valid instances from $segSet_{scale}$ and $segSet_{geom}$. The number and intervals of parameter variations can be regarded as a balance between computational burden and exploration of potential information. However, the usage of up to nine segmentation maps for $segSet_{scale}$ and eight maps for $segSet_{geom}$ proved to be appropriate in performed experiments.

Note that $segSet_{scale}$ forms a typical multilevel representation of the image and meets equation (1), while $segSet_{geom}$ violates equation (1) as it does not build a hierarchic structure. Image segments of one segmentation map cross borders of image segments of a second segmentation map as they do not differ greatly in size but in geometry. Consequences resulting out of this are discussed in section 7.

4.2.3 Basis SVM Classification

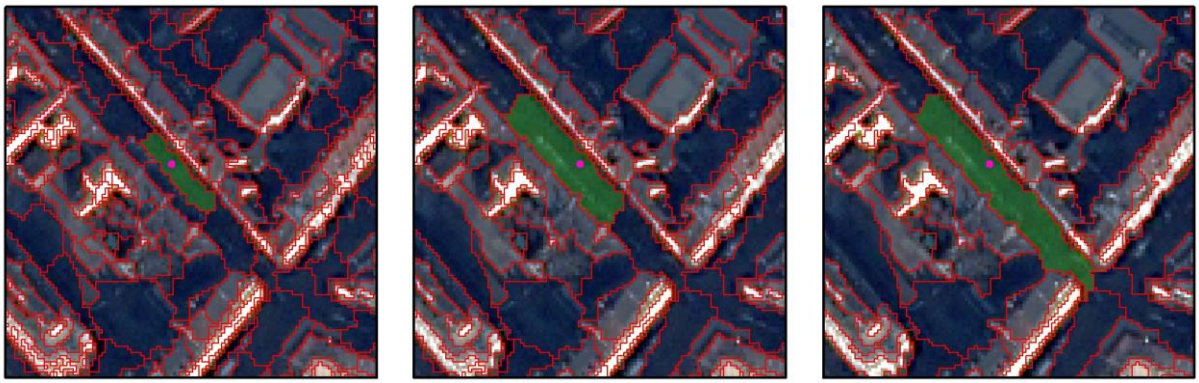
Although this work addresses invariance to scale and geometry, this section presents the method concerning scale invariance. Both procedures rely on the same principles and manner of implementation. Only the used set of segmentation maps, namely $segSet_{scale}$ and $segSet_{geom}$ distinguish the methods variation.

As the initial SVM on seg_{base} structures the foundation of the subsequent process, it will be referred to as SVM_{base}. SVM_{base} is trained on given labeled image-objects of seg_{base} that are spitted in $trainSet$ and $testSet$ for the hyperparameter optimization (C, γ). The feature set for the image-objects is specified according to the sensor, scene and classification task. Numerical values of each feature typically move in different ranges. While shape features may adopt values in the range [0,1], spectral features, depending on the bit-depth of the imagery, can appear in [0,>100] or even be negative for band ratios. This entails the problem that features showing greater numeric ranges dominate those in small ranges (Gualtieri, 2009: 69;

Hsu et al., 2016: 4). Linear scaling to the range of [0, 1] for each feature needs to be performed to avoid this issue. As result of this initial step, the decision function of SVM_{base} is trained on the optimal segmentation level seg_{base} and encodes the information of well modelled image-objects.

4.2.4 VSV Generation and Integration

In correspondence with the VSVM introduced in section 3.3, manipulation of data instances to encode the invariance is only applied to SVs. Therefore, SVs of SVM_{base} are extracted from the model. This set of SVs will be referred to as SV_{base}. Those instances are then located in the image domain and in each segmentation map of $segSet_{scale}$ segments are identified that include an instance of SV_{base} (Fig. 8). If $segSet_{scale}$ is composed by N different segmentation maps and SV_{base} contains M instances, a total number of $M * N$ segments are nominated during this step. As the image-objects of $segSet_{scale}$ are modelled to represent the entire spectrum of scale characteristics present in data, image-objects selected as VSVs should encode broad scale variance of all classes respectively. The same set of features used in SVM_{base} is calculated for each identified segment and scaled with the same transformations applied for image-objects of $segSet_{base}$. Otherwise the value distributions of identical features in different segmentation maps are decoupled and incomparable. The resulting feature vectors are labeled similar to the SVs of SV_{base} from which they are derived. They construct a set of potential VSVs, referred to as $pVSV_{scale}$. To address the relationship between an instance x^{OrgSV} of SV_{base} and instances derived from x^{OrgSV} – $x_1^{VSV}, x_2^{VSV}, \dots, x_N^{VSV}$ of $pVSV_{scale} - x^{OrgSV}$ is called *parent* of $x_1^{VSV}, x_2^{VSV}, \dots, x_N^{VSV}$. Summarized $pVSV_{scale}$ should introduces further scale characteristics which are likely to appear in data, but are absent in the original labeled instances. Consequently, using $pVSV_{scale}$ should make the classifier invariant to specific scale representations and thereby increase the potential for good generalization.



(a) (b) (c)

Fig. 8: Exemplary generation of VSVs. Pink: labeled sample of SV_{base}. Green: selected image-object. (b) Labeled sample of SV_{base} and selected image-object of segmentation map $segSet_{base}$ considered in SVM_{base}. (a), (c) Labeled sample of SV_{base} located in segmentation maps of $segSet_{scale}$ and image-objects identified as VSVs. *Own figure. Subset of Satellite Image: DigitalGlobe, (2014).*

4.2.5 Optimization Procedure

At this stage, information for the classification by SVM_{base} is encoded in SV_{base}, while potential VSVs are included in $pVSV_{scale}$. However, guidelines for generating $segSet_{scale}$ (stated in section 4.2.2) do allow over- and undersegmentation. In consequence, depending on the objects spectral and spatial homogeneity, $pVSV_{scale}$ does also contain misleading information introduced by those over- and undersegmentations. Lu et al. (2016) approach a related difficulty within the context of active learning by an optimization procedure. By introducing a set of rules concerning feature similarity, they exclude mixed pixels from a selection of unlabeled samples that subsequently enrich training data. In the following section, this ruleset is adopted and integrated in the proposed method to exclude instances from $pVSV_{scale}$ that strongly over- or undersegment objects of interest and therefore are not representative for their land cover class.

In addition, considering the computational burden of the classification, it is not desirable to enlarge the training dataset with instances that only carry information already encoded in the original data (Tuia et al., 2011: 606). Therefore, the principle of the active learning Margin Sampling (MS) strategy (Tuia et al., 2009b: 2221) is used to exclude redundant instances from $pVSV_{scale}$ (section 4.2.5.2).

4.2.5.1 Measure of Feature Similarity

Lu et al. (2016) use a Euclidian distance measure d, which determines the distance in feature space of some labeled SVs and an unlabeled sample selected as potential candidate to enrich the training data. Mixed pixels are expected to show great spectral differences to non-mixed pixels. Therefore, they are located far apart in feature space: the larger their distance the less their similarity. In this sense d can be used as a similarity measure between samples, which allows identification and consequently exclusion of mixed pixels. Transferring these insights to object-based classification, it is assumed that objects of the same class do not only show alike spectral features, but also alike geometric and textual features. Potential VSVs that encode information resulting of over- or undersegmentation do not share this similarity with their parent. More precisely, d measures the feature similarity between a potential VSV and its parent, so that

$$d_{ij} = \sqrt{\sum_{m} \left(x_{im}^{VSV} - x_{jm}^{OrgSV} \right)^2} . \tag{12}$$

x_j^{orgSV}, $j = \{1, 2, \dots, M\}$ is the jth instance of SV_{base}. x_i^{VSV}, $i = \{1, 2, \dots, N\}$ denotes the ith potential VSV derived from x_j^{orgSV} and m denotes the number of features per instance. In other words, the distance between each potential VSV and its parent in feature space is calculated.

By introducing a maximum distance δ and using it as threshold, a potential VSV lying in large distance to its parent and therefor is likely to encode over- or undersegmentation, can be excluded from $pVSV_{scale}$ (Fig. 9a). As intraclass variances – and therefor potential distances in feature space – is highly dependent on the scene and the class of interest, δ must be adjusted for each setting and land cover class. To avoid manual thresholding, the step of adapting δ for each class was automated as follows:

$$\delta_C = \frac{2}{N_C\,(N_C\,-1)} \sum_{i=1}^{N_C-1} \sum_{j=i+1}^{N_C} \sqrt{\sum_m \left(x_{im}^{OrgSV_C} - x_{jm}^{OrgSV_C}\right)^2}\,. \tag{13}$$

Hereby N_C is the number of SVs in SV_{base} for the class C. $x_i^{OrgSV_C}$ and $x_j^{OrgSV_C}$ hence denote the ith or jth SV of SV_{base} for the class C. Simply put, δ_C is the mean distance in feature space between all instances of SV_{base} that belong to the same class C. Moreover, it is of interest to downscale δ_C for situations with already great information content in SV_{base}, typically high number of SV in SV_{base}, or to upscale it for the contrary case. Multiplying δ_C with a factor k introduces this flexibility:

$$\delta_{Ck} = k * \delta_C\,. \tag{14}$$

By excluding instances of $pVSV_{scale}$ according to

$$sVSV_{scale} = pVSV_{scale} \cap \{x_i^{VSV_C} | d_{ij} \leq \delta_{Ck}\} \tag{15}$$

$sVSV_{scale}$ only contains VSVs that lie within the radius of δ_{Ck} around their parent from SV_{base} and can be seen as save VSVs (Fig.9a).

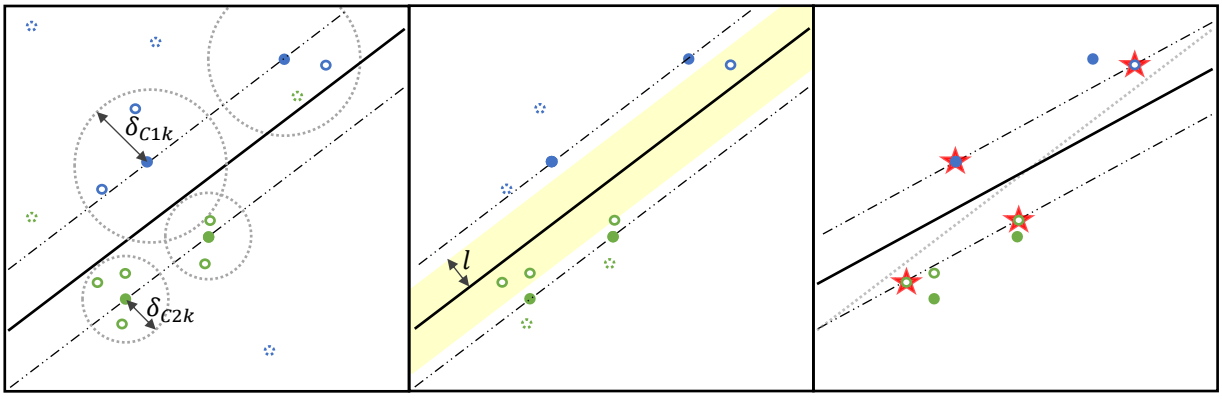


Fig. 9: Principle functionality of the optimization procedure for potential VSVs. Dots represent labelled training data of two classes (blue /green). Circles represent VSV of each class (blue/green), if dashed they get excluded in corresponding step. Stars mark selected SV in the final model (a) Initialy trained model with VSVs and feature similarity measure δ_{Ck} for each class. (b) Adapted Margin Sampling Strategy with corresponding threshold l. (c) Final model with shifted hyperplane and new selected SVs. *Adapted from Lu et al. (2016: 4921).*

4.2.5.2 Adapted Margin Sampling Strategy

The dataset $sVSV_{scale}$ only contains VSVs that properly represent the class of their parent. However, by only selecting VSVs that carry for the modelbuilding relevant information, $sVSV_{scale}$ can be further reduced. Referring to the MS strategy (Tuia et al., 2009b: 2221, 2011: 608f), only those VSVs that are located close to the margin of SVM_{base} are likely to become SVs in the final model and thereby contribute to the classification result. Consequently, only VSVs in close distance to the hyperplane are selected for the new model. With the similar motivation that leads to the factor k included in equation (14) – the adjustment to the amount of already included information in the original training data –, a second threshold l is introduced which determines the margin of acceptance along the hyperplane. In detail, l determines the maximum distance an instance of $sVSV_{scale}$ may have to the decision hyperplane of SVM_{base} to be considered in training of the final model (Fig. 9b). The distance of an instance to the hyperplane can be determined by the adaption of the decision function (9) to

$$f_d(x_*) = \left| \sum_{i=1}^{n} y_i \, \alpha_i K(x_i, x_*) + b \right| \tag{16}$$

for a two-class SVM (Tuia et al., 2011: 608). It is recalled that the one-against-one scheme is used in multiclass problems. Correspondingly, a VSV of $sVSV_{scale}$ is selected for a final set VSV_{scale}, if its distance to the hyperplane is in at least one of the SVMs its class constructs, smaller than the threshold l. In consequence, the VSVs contained in VSV_{scale} have passed through the optimization procedure and qualify therefore as: save, in the sense that no misleading information through over- or undersegmentation is introduced, and informative, as they are located close to the decision hyperplane and therefore are likely to shift its position (Fig 9c).

4.2.6 Final Model

To train the final model, which can be characterized as invariant to encoded characteristics and is expected to produce higher classification accuracy of unseen data, the VSVs are included in the training process. The new training set $trainSet_{scale}$ is formed through the union of SV_{base} and VSV_{scale}. On this set and the use of $testSet$ for the hyperparameter optimization (C, γ) the invariant model $VSVM_{scale}$ is learned (Fig 9c). However, the performance still relies on the parameters k and l introduced in section 4.2.5.1 and 4.2.5.2. For further decoupling the system from thresholding, setting k and l can be seen as another optimization problem. Therefore, two sets of parameters $K = \{0.3, 0.6, 0.9\}$ and $L = \{0.5, 1.0, 1.5\}$ are defined. The hyperparameter optimization (C, γ) is executed for each combination of $K \times L$ and the one that produces the highest classification accuracy is nominated as final model $VSVM_{scale}$. This final invariant model is used for the actual classification task. Recall that the procedure formulated

for scale invariance can be adapted for geometry invariance by using the set of segmentation maps $segSet_{geom}$ instead of $segSet_{scale}$.

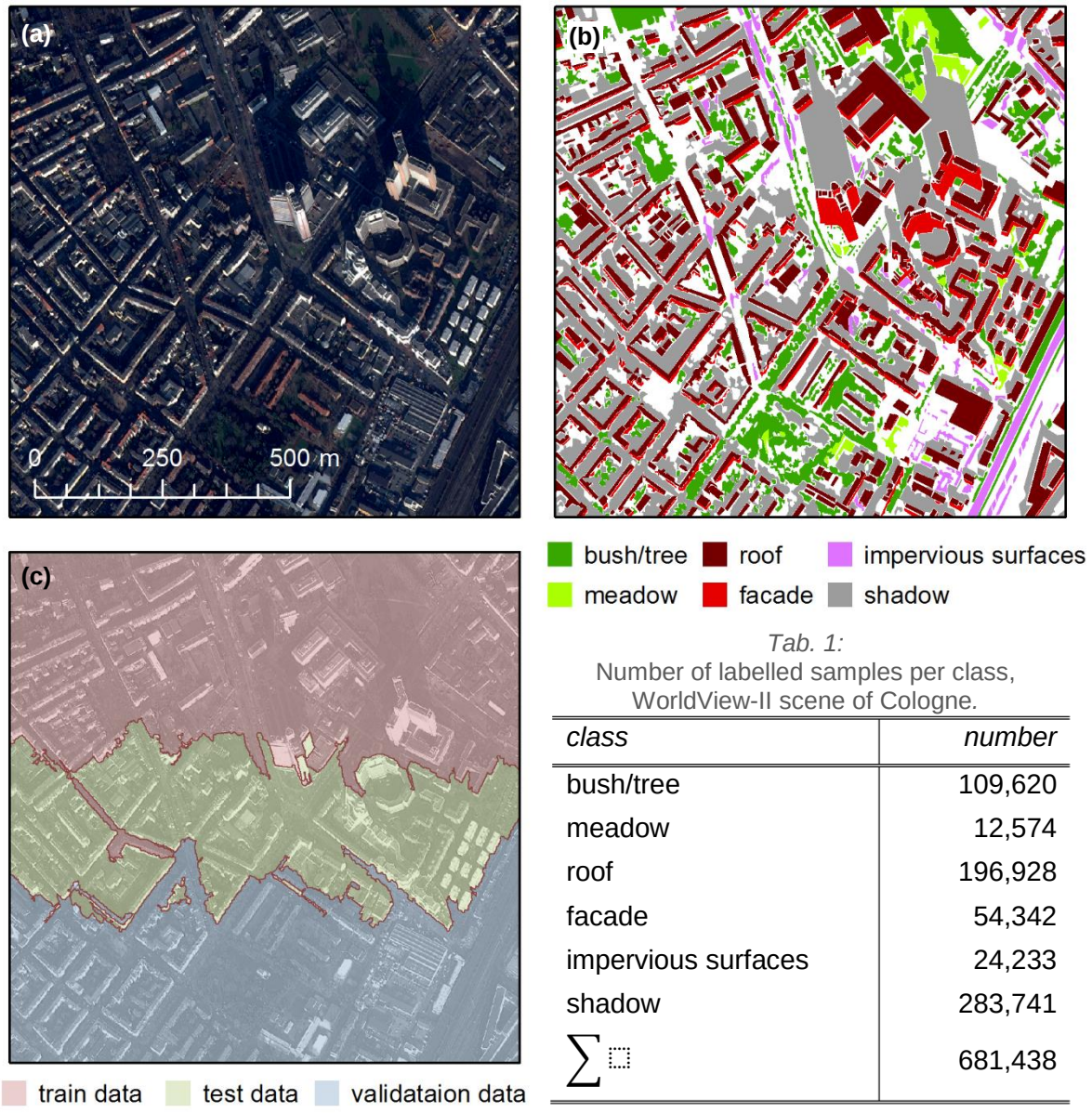
5 EXPERIMENTS

For assessing the proposed method, a set of experiments are realized. The aim is to 1) evaluate its potential for binary and multiclass settings on complex VHR data, which is typically subject to OBIA approaches. 2) To explore the behavior of the method on different amounts of ground truth data. The focus is on very small training set sizes, which are naturally not capable to cover all object variances potentially present in unseen data and which are prone to suffer the Hughes phenomena. And 3) to assess the effectiveness of the optimization procedure to exclude misleading samples due to under- and oversegmentation and, resulting out of this, the sensitivity of the method to the segmentation maps included in $segSet_{scale}$ or $segSet_{geom}$. The last point is motivated by the idea, that a wide range of segmentation maps could be given as input while the algorithm chooses the useful samples generated from the segmentation maps itself. Then, time-intensive evaluations of the segmentation maps could be spared, and reduced to a ruff maximum and minimum estimation of parameters, while the process excludes misrepresentations. The used data foundation is presented in section 5.1 while the general experimental setup is presented in section 5.2.

5.1 WorldView-II Scene of Cologne, Germany

The image used in the experiments was acquired by the WorldView-II satellite sensor over the city of Cologne in Germany on January 31, 2014 and is subset to an extend of 1000 x 1000 pixels (Fig. 10a) (DigitalGlobe, 2014). It captures the complex land cover composition of the city center, composed by buildings of commercial use, different housing types, parks or vegetation dominated regions and impervious areas like roads, parking lots or public squares. Shadows appear mostly adjacent to buildings and in some cases extensively cover objects of all land cover classes. The off-nadir acquisition makes it possible to identify also the facades of buildings. For reduction of the computational burden, the pan-sharpened (Hyperspherical Color Sharpening) multispectral image with a geometrical resolution of 0.65m is scaled by a nearest neighbor interpolation to a resolution of 1.00m. The set of features used to characterize the image-objects and constitute the information input and foundation of the classification, is composed by measures introduced in section 2.2. Namely, *mean* and *standard deviation* of the four by the sensor recorded bands in the wavelength ranges of blue (450-520 nm), green (520-600 nm), red (630-690 nm) and near-infrared (NIR) (790-900 nm), supplemented by an NDVI form the spectral features. For texture information, the GLCM measures *homogeneity*, *dissimilarity* and *mean* are computed while shape information is covered by the indices *rectangular fit*, *elliptic fit*, *roundness*, *shape index* and *compactness*. The ground truth data consists in 681,438 labelled pixels of six classes: bush/tree, meadow, roof, facade, impervious

surfaces and shadow (Fig. 10b; Tab. 1). This data, generated through photointerpretation, was provided by the team Modelling and Geostatistical Methods of the German Remote Sensing Data Center at the German Aerospace Center. As the hyperparameter optimization (C, γ) requires the two spatially disjunct data sets $trainSet$ and $testSet$, and as for validation of the method a third set of unseen instances is needed, the image domain is spitted according to Fig. 10c. Note that the image splitting is performed in correspondence to object outlines of the largest objects in $segSet_{scale}$, however any object of $segSet_{geom}$ surpassing the splitting border is excluded from the process.



Tab. 1:
Number of labelled samples per class, WorldView-II scene of Cologne.

class	*number*
bush/tree	109,620
meadow	12,574
roof	196,928
facade	54,342
impervious surfaces	24,233
shadow	283,741
∑	681,438

Fig. 10: Data foundation, WorldView-II scene of Cologne, Germany. (a) Scene subset under investigation. (b) Ground truth data. (c) Dataset splitting in disjunct training, testing, and validation data. *Own figure. Satellite Image: DigitalGlobe, (2014).*

5.2 Experimental Setup

Both types of invariance, scale and geometry, are tested separately on one two- and one multi-class setting. The two-class setting discriminates the class bush/tree from the rest of the land cover classes, while the multi-class problem considers all six land cover classes.

To simulate the different quantities of labeled instances in $trainSet$ and $testSet$, a stratified random sampling, is used to select the same number of samples for each class from the pool of ground truth data. This random set of labeled samples is iteratively subset to narrow information input along the process. Doing so, the number of the training and testing samples respectively ranges from 67 to 2 per considered class in the two-class setting and from 500 to 5 for the multi-class setting. Those ranges are chosen as in those settings the proposed method shows the greatest performance difference in comparison with state of the art classification methods. Also, sample sizes are large enough to identify the plateau effect of reported accuracies, where the quality of classification stabilizes at some maximum level. To produce reliable and comparable results, the average classification accuracy and its standard deviation are calculated over a total of 20 runs for each setting. The accuracy of the classification is estimated by Cohen´s kappa statistic κ (Foody, 2004), since it accounts for bias class sample sizes. Additionally, the number of SVs is reported to estimate the complexity of each model generated. As a reference for classification quality, a multilevel object-based classification (section 2.1) is calculated that uses the same information input as considered for the VSVM procedure, i.e. the image-objects of the different segmentation maps and the calculated feature set. It aims on pointing out which performance advantages and disadvantages the proposed method has, compared to a state of the art method, while both rely on the same input data. For every model, the grid search to determine C and γ is run over the sequences $C = \{2^{-4}, 2^{-3.5}, \dots, 2^{12}\}$ and $\gamma = \{2^{-5}, 2^{-4,5}, \dots, 2^{3}\}$, following the recommendations of Hsu et al. (2016: 5) that proved to be suitable for similar settings (Geiss and Taubenböck, 2015; Geiss et al., 2016a).

To make detailed evaluation of the performance possible and the contribution of different method components accountable, the κ coefficient is reported of four models in each experiment. First, the multilevel object-based classification as reference ($SVM_{MLscale}$/ SVM_{MLgeom}). Second, the model SVM_{base} as foundation of the proposed invariant model. Third, a model that relies on the union of SV_{base} and $pVSV_{scale}$ or $pVSV_{geom}$ for training ($pVSVM_{scale}/pVSVM_{geom}$). It can be seen as sub-step during the calculation process of the invariant model, as the VSVs are integrated but not filtered by the optimization procedure. This is of interest as the model uses the exact same information as the multilevel object-based reference classification. However, it is included as invariant instances instead of extended feature dimensions. Also, it makes the evaluation of the optimization measures possible, as

they filter the VSVs in the next processing step before building the final model ($VSVM_{scale}/VSVM_{geom}$), which completes the set of reported accuracies.

All four experiments use the same SVM_{base} model on the optimal image-objects of seg_{base}. For generating seg_{base}, more emphasis is given to the shape component in the MRIS algorithm, since man-made elements like urban structures show typically regular shapes and sizes. The components compactness and smoothness are balanced equally. Therefore, the parameters of the MRIS algorithm to derive seg_{base} are set to shape: 0.7 and compactness: 0.5, while optimal scale parameter is found to be 20.

In the experiments exploring scale invariance, the VSVs are derived from segmentation maps of $segSet_{scale}$, which encode the scale invariance within the image segments as they are generated by altering the scale parameter of the MRIS. $segSet_{scale}$ is composed by nine segmentation maps of the image, which result from the nine scale parameters, 10, 15, 25, 30, 35, 40, 50, 60 and 80 combined with the constant shape (0.7) and compactness parameter (0.5) from seg_{base}. In contrast, the set of segmentation maps $segSet_{geom}$ used for the experiments on geometry invariance, is generated by constant scale parameter of 20, while altering the shape and compactness parameter. I.e. the eight parameter combinations (shape; compactness) (0.1;0.9), (0.3;0.7), (0.3;0.5), (0.5;0.7), (0.5;0.5), (0.5;0.3), (0.7;0.3) and (0.9;0.1) are used with the constant scale parameter of 20.

Experiments are realized using two main software solutions. For image segmentation, the ‘multiresolution segmentation’ algorithm (eCognition, 2016: 67-72) implemented in eCognition Developer 9.2.1 software by Trimble is used. eCognition is also utilized for feature calculation (eCognition, 2016: 357-372, 409-420). However, any application that fulfils the remakes made on those steps can substitute the program. The rest of the method is implemented as stated in section 4 in the software environment R using of the package ‘caret’ and ‘kernlab’ for classification related tasks (Karatzoglou, 2004; Tillé and Matei, 2015; Kuhn et al. 2016; R Core Team, 2016a; R Core Team, 2016b).

6 RESULTS

Obtained results are summarized in two plots for each experimental setting. Of these, one reports the κ statistics, while the second provides the number of used SVs in each model. Both are reported as function of the used ground truth data in the models' training. In account with the experiments, the results are presented for scale and geometry invariance separately.

6.1 Scale Invariance

For the two-class setting, estimated measures for performance evaluation are reported in Fig. 11a, b. It can be observed that κ statistics for the $VSVM_{scale}$ show the highest accuracies throughout all sizes of ground truth data sets and outperform $SVM_{MLscale}$. This tendency is much stronger in small training sets (between 0.2 and 0.09 points increase for under 20 training samples per class) and decreases with larger training sets (0.02 points increase for 134 samples per class). The $VSVM_{scale}$ reports already with 10 training samples per class κ values above 0.7 (i.e. 0.72) which is achieved by $SVM_{MLscale}$ only with more than 26 samples per class

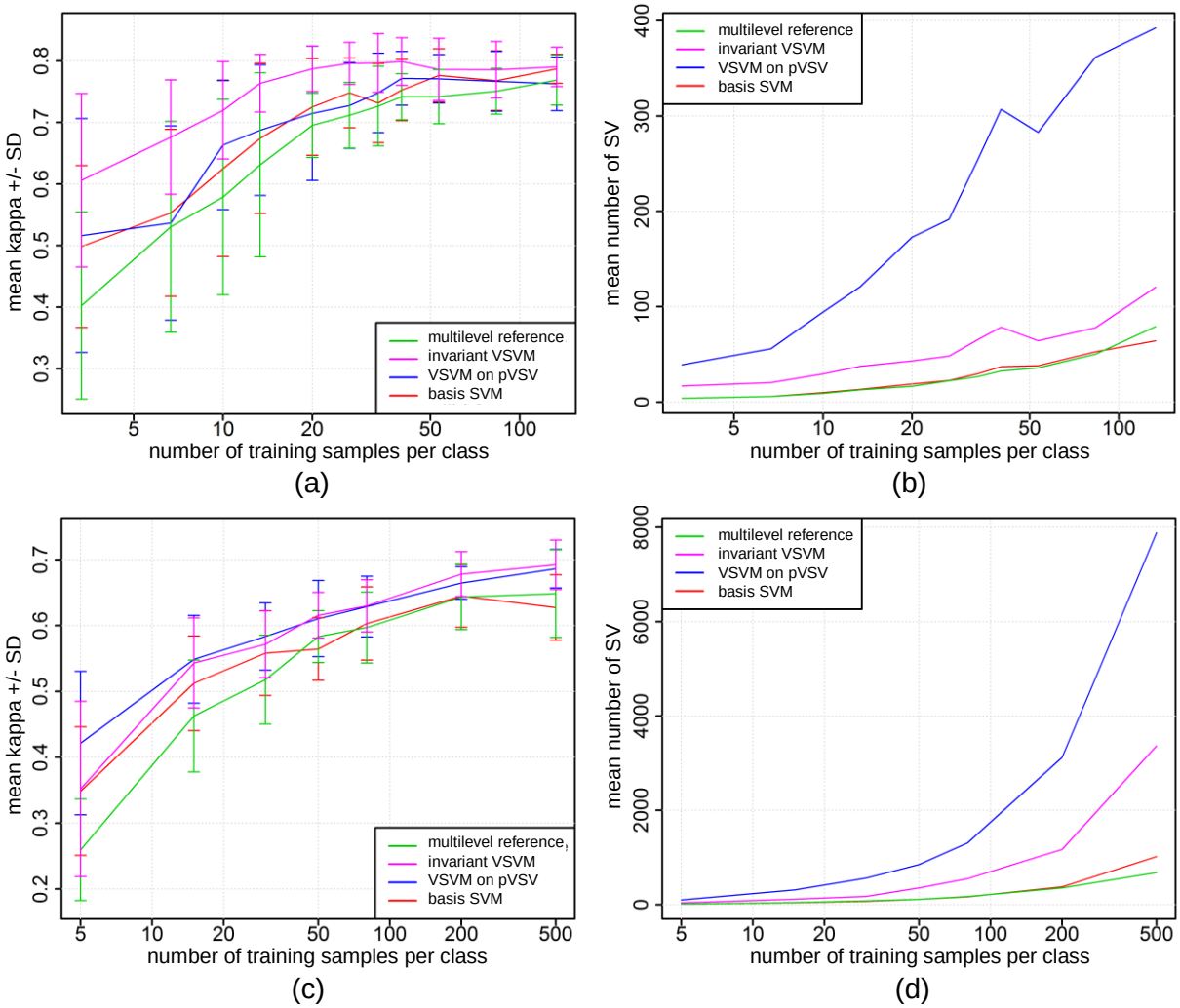


Fig. 11: Classification results for experiments considering scale invariance. (a), (b): two-class setting. (c), (d): multi-class setting. Mean and standard deviation of κ statistics and mean number of SVs for each model and amount of ground truth data used in training. Number of realizations: 20. *Own figure.*

(i.e. 0.71). The highest accuracy of 0.8 is obtained by the $VSVM_{scale}$ on 54 samples, followed by a slight decrease of accuracy, while the highest accuracy obtained by $SVM_{MLscale}$ lies with 0.77 at 134 training samples per class. Also, the standard deviation of κ statistics for $VSVM_{scale}$ is smaller compared to $SVM_{MLscale}$ in settings with sparse training data. However, with large training data sizes their difference gets negligible and partly reversed. Considering the performance of $SVM_{MLscale}$ it must also be noted that even the initial SVM_{base}, only considering one segmentation level, shows slightly higher κ accuracys. Fig. 11b shows very clearly that $pVSVM_{scale}$ is built on a multiple number of SVs compared to the rest of reported models. $VSVM_{scale}$ uses roughly double the number of SVs in reference to $SVM_{MLscale}$.

In general, the multi-class setting of the scale invariance generates lower κ accuracies compared to the two-class setting (Fig. 11c). As observed for the binary classification problem, on limited information input the $VSVM_{scale}$ performs better than the $SVM_{MLscale}$. However, despite the improvement, the model´s accuracies stay under 0.61 for less than about 30 samples per class. It reaches between 0.35 and 0.62 κ accuracy on 50 or less training samples per class, which constitutes an improvement of 0.09 to 0.04 points in reference to $SVM_{MLscale}$. Although $VSVM_{scale}$ and $SVM_{MLscale}$ approximate each other on large information input $VSVM_{scale}$ still achieves about 0.03 points of improvement in κ accuracy for the largest tested sampling sizes (500, 200 and 80 samples per class). In contrast to the two-class setting however, results show stronger performance of the $pVSVM_{scale}$. Its κ accuracy ranges in the same level as achieved by $VSVM_{scale}$, but shows almost always at least double the number of used SVs (Fig. 11d). The highest accuracy is produced by $VSVM_{scale}$ with 0.69 on 500 samples per class. The $SVM_{MLscale}$ also reaches its maximum at 500 samples with an κ accuracy of 0.65.

6.2 Geometry Invariance

For the two-class setting of geometry invariance results are reported in Fig. 12a, b. The $VSVM_{geom}$ model shows 0.2-0.03 points κ accuracy improvement for less than 20 samples per class compared to SVM_{MLgeom}, which appears highly unsteady in these settings. The first κ accuracy greater than 0.7 produces $VSVM_{geom}$ on only 10 samples per class (0.72) while SVM_{MLgeom} exceeds this level the first time on 20 samples per class (0.73). However, from about 40 training samples per class, SVM_{MLgeom} outperforms $VSVM_{geom}$, both in mean κ accuracy as well as in reported standard deviation. Therefore, the highest κ accuracy is reached by SVM_{MLgeom} (0.82) on 84 samples per class, while the $VSVM_{geom}$ reaches its plateau at 26 samples and a κ of about 0.79. The number of used SVs (Fig. 12b) for small training sizes resembles the characteristics of the two-class problem considering scale invariance.

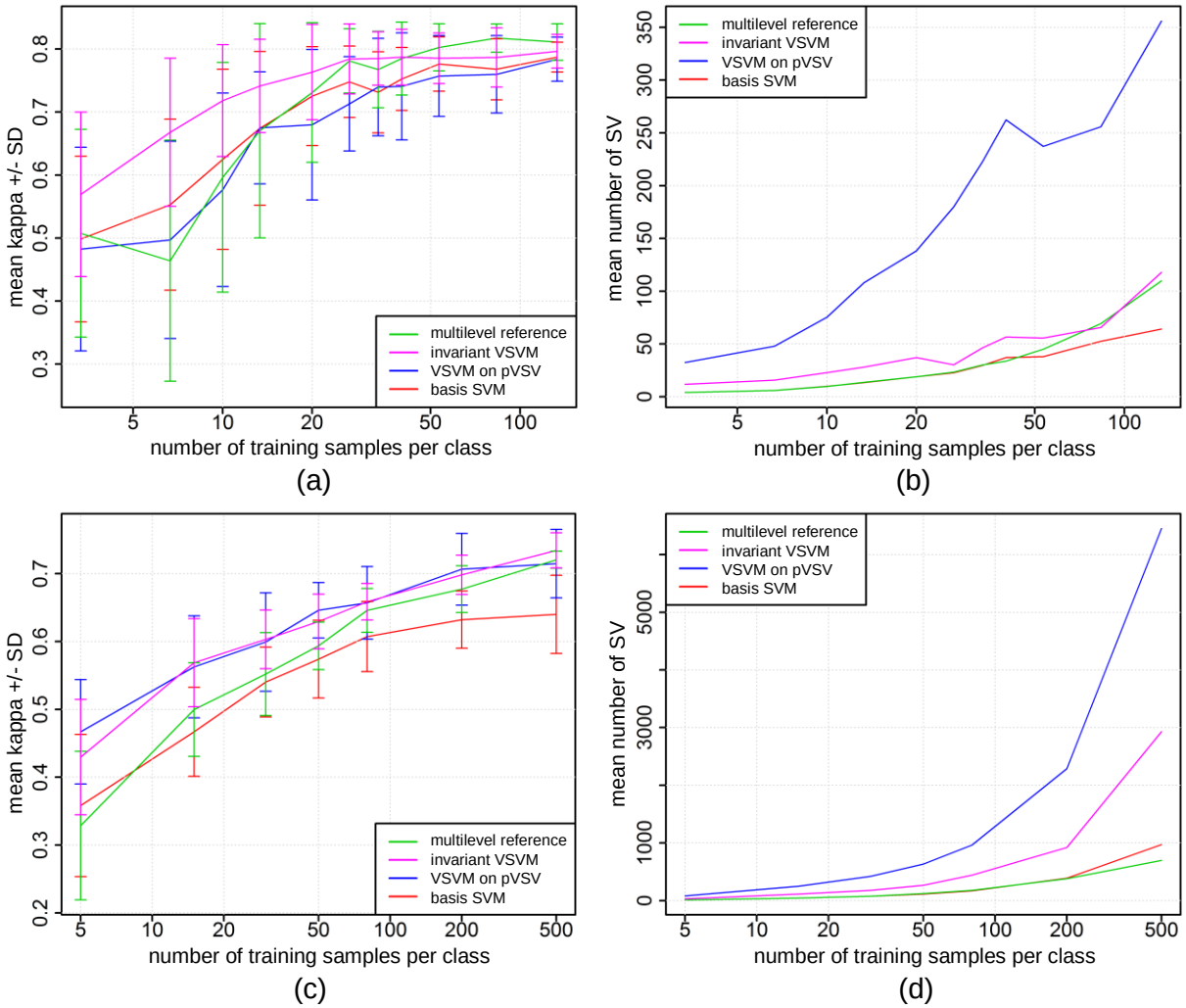


Fig. 12: Classification results for experiments considering geometry invariance. (a), (b): two-class setting. (c), (d): multi-class setting. Mean and standard deviation of κ statistics and mean number of SVs for each model and amount of ground truth data used in training. Number of realizations: 20. *Own figure.*

The results of the multi-class setting for geometry invariance (Fig. 12c, d) resemble tendencies of the multi-class setting for scale invariance. Nevertheless, the models in general show slightly better performance. SVM_{MLgeom} appears to minimize the accuracy difference to $VSVM_{geom}$ even further on large training sets. Yet again, the invariant models stay under 0.6 κ accuracy for under 30 sample per class. SVM_{MLgeom} reaches this level only between 50 and 80 samples per class. Accuracy increase achieved by $VSVM_{geom}$ in reference to SVM_{MLgeom} lies between 0.11 and 0.04 points (to reach between 0.43 and 0.63 κ accuracy on 50, or less training samples per class). In this setting the highest classification quality is reached by $VSVM_{geom}$ with 0.73 κ accuracy on 500 samples per class, whereas SVM_{MLgeom} reaches its highest accuracy of about 0.72 κ accuracy on the same sample size.

7 DISCUSSION

Results proof that the reported accuracies for the proposed invariant VSVMs in general correspond to assumptions on their performance. In settings where training data and thus information input of the model is limited, the invariant VSVM is capable to enhance the classification outcome and build more robust models. However, if the available ground truth data is vast, improvement is reduced and even turns to accuracy decrease for the binary geometry invariance setting (Fig. 12a) in comparison to the multilevel reference classification (Fig.11a, c, 12a, c).

Next to the introduced invariance, the improvement of accuracy for small sample sizes might be connected to effects related to the Hughes phenomena. This can be assumed as, with exception of the two-class setting of geometry invariance (Fig. 12a), $pVSVM_{scale}$ and $pVSVM_{geom}$ show higher accuracies compared to the corresponding multilevel classification (Fig. 11a, c, 12c). Yet all these models rely on the exact same information input. However, the multilevel representation considers object information of different scales by extending the feature vector. In contrast, $pVSVM_{scale}$ and $pVSVM_{geom}$ include this information as artificial samples in the training set and thereby avoid the high dimensional feature vectors that lead to problems connected to the Hughes phenomena.

Considering the optimization procedure (section 4.2.5) different tendencies can be observed for two-class problems and multi-class problems. For the binary classification tasks (Fig. 11a, b, 12a, b), the optimization procedure clearly works well. While reducing the number of considered SVs to less than a third in most cases, the achieved accuracy is improved clearly. In multi-class settings (Fig. 11c, d, 12c, d), it can be observed that the $pVSVM_{scale}$ and $pVSVM_{geom}$ produce comparable (most settings) or higher (very small training sets) accuracies as $VSVM_{scale}$ and $VSVM_{geom}$. The reason for those differences between two- and multi-class settings cannot be finally determined without further experiments. However, there are two main factors to consider: First, the information content within the training set in relation to the overall complexity of the classification task. Even though, the multi-class settings consider larger training sets this relation might depend on further factors. And second, that different information classes might respond differently on the optimization procedure and the method as a whole. Nevertheless, it can be noted that in no case the optimization procedure impairs accuracies below the level of the reference multilevel classification while $pVSVM_{scale}$ or $pVSVM_{geom}$ produce better accuracy. This is important as therefore no accuracy decrease in reference to the multilevel model can be observed which is implied by the optimization procedure.

Maximal achieved accuracies present another aspect of discussion. They need to be interpreted as a tool for relative comparison of the different models, yet carefully judged in

terms of maximal achievable accuracy. This is due to the fact, that object-based classification accuracy must always be regarded in connection to the quality of segmentation maps (Liu and Xia, 2010). Also, characterizing features are chosen on the basis of literature recommendations, however might not be the optimal set for the scene under investigation. In addition, the plateau-effect which appears with approximation of maximal accuracies is, especially for multi-class settings, only indicated and by far not well established. In this context, performance of the initial SVM SVM_{base} also needs to be discussed. The reasons of it good performance compared to the reference classification could be connected to the Hughes phenomena as well. This is supported by the observation that in three of the four settings (Fig. 11c, 12a, c) the multilevel classifier outperforms SVM_{base} on large sample sizes. Yet, this tendency appears rather weak. On one hand this illustrates the limited potential of multilevel models for small sample sizes. On the other hand, it relativizes the reported numerical differences of κ values between the invariant VSVM and the reference multilevel classification.

The fact that the objects of $segSet_{geom}$ do not meet constraint (1) – hence do not build a hierarchal structure – and that this might influence generated accuracies was already mentioned in section 4.2.2. Precisely it means that objects of different segmentation maps overlap each other which leads to the following problem: If an labeled instance is chosen for training which is located in the edge area of an image-object, e.g. a corner of a building, for a different segmentation map of $segSet_{geom}$ this instance might be included in an object modelling a different class, e.g. the sidewalk next to it. Consequently, image-objects which model different classes carry the same label. In case of the multilevel model this leads to feature vectors entailing information of different real-world objects. This could contribute to the unsteady accuracy of SVM_{MLgeom} and $pVSVM_{geom}$ in the two-class geometry invariance experiment (Fig. 12a). However, $VSVM_{geom}$ in the same setting seems to exclude this misrepresentation during the optimization procedure as accuracy is enhanced and standard deviation reduced.

Next to increasing the accuracy of segmentation maps, the method aims at further decupling results from parameter setting with reference to existing models, especially as proposed by Izquierdo-Verdiguier et al. (2013). For this scene, it could be observed that the optimization procedure is able to prevent from poor accuracies in reference to the multilevel classifier. Consequently, it can be assumed that the additional presence of over- and undersegmentation in segmentation maps of $segSet_{scale}$ and $segSet_{geom}$ does not strongly affect the accuracy of the invariant model as their exclusion by the optimization procedure appears effective. Thus, the strongest variations in object size and geometry can be roughly estimated and do not require great effort in determination. Consequently, their definition for different data appears relatively easy and effective. The optimal range of the factors k and l is not expected to vary

strongly on other scenes as they are factors to scene and class specific measures derived from the data itself. Extra effort and dependency related to thresholding appears therefore reasonable compared to the multilevel object-based classification where included segmentation maps need to be chosen carefully and influence directly the classification results (Liu and Xia, 2010). However, to generalize these findings, experiments on other scenes need to confirm those observations.

8 Conclusion and Outlook

This work introduces a classification approach to address challenges occurring by classification of VHR imagery on limited ground truth data. The OBIA method, which is extensively used for VHR data classification, is combined with VSVMs, commonly used in disciplines like pattern recognition to build invariant classifiers. This property of a classifier to be invariant, hence to be robust to changes in data representation, is implemented for shape and geometry characteristics of image-objects used in the OBIA. To evaluate the proposed method, different classification settings are tested on a WorldView-II scene and compared to the state of the art multilevel object-based classification method. Experiments demonstrate a great potential of the introduced approach. For two-class settings classification κ accuracy can be enhanced about 0.2-0.03 points on small training sets (to reach 0.57-0.79 κ accuracy on 20 or less training samples per class), while improvement for multiclass settings lies around 0.11-0.04 points (to reach 0.43-0.63 κ accuracy on 50, or less training samples per class). For large training sets the method shows generally less to no improvement in classification accuracy.

Considering limits of maximal achieved accuracies, the sole focus on small training sets and differences in two- and multi-class problems the experiments presented in this thesis only constitute a set of initial tests on the proposed method to generally explore its potential. As the experiments only consider one scene, stated results need to be confirmed on further classification tasks and settings. To generate more universal findings and ensure general validity of the proposed method, further research should investigate:

- The method´s performance on further scenes with different landcover classes and classification tasks.
- The validation of the significance of reported classification accuracy differences by statistical measures.
- The behavior of the method if invariance is exclusively encoded for a subset of land cover classes – due to reported differences in binary and multi-class classification problems.

- Challenging the computational burden as optimization over k and l requires to compute $k * l$ models.

Nevertheless, this thesis points out that the invariant VSVM introduced to the context of remote sensing image classification using OBIA has great potential to enhance classification results and reduce barriers related to limitations of ground truth data.

SOURCE OF SATELLITE DATA

DigitalGlobe (2014), WorldView-II scene form January 31, 2014, DigitalGlobe, Longmont, Colorado. Provided by: Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Deutsches Fernerkundungsdatenzentrum, Georisiken und zivile Sicherheit.

Pre-processed (radiometric correction, sensor correction, geometric correction, pan-sharpening) by supplier according to: http://www.pasco.co.jp/eng/products/worldview-2/ (Accessed 8 December, 2017)

Subset to: Top: 5643586.5, Left: 354145.0, Right: 355145.0, Bottom: 5642586.5

REFERENCES

Albertz, J., 2009. Einfühung in die Fernerkundung. Grundlagen der Interpretation von Luft- und Satellitenbildern. Edition 4. WBG, Darmstadt, 254 p.

Aplin, P., Smith, G.M., 2011. Introduction to object-based landscape analysis. International Journal of Geographical Information Science 25 (6), 869–875.

Ben-Hur, A., Weston, J., 2010. A user's guide to support vector machines. In: Carugo, O., Eisenhaber, F. (Eds.), Data Mining Techniques for the Life Sciences. Methods in molecular biology 609, Springer, New York, pp. 223–239.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS Journal of Photogrammetry and Remote Sensing 58 (3-4), 239–258.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65 (1), 2–16.

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic Object-Based Image Analysis - Towards a new paradigm. ISPRS Journal of Photogrammetry and Remote Sensing 87 (100), 180–191.

Bruzzone, L., Carlin, L., 2006. A Multilevel Context-Based System for Classification of Very High Spatial Resolution Images. IEEE Transactions on Geoscience and Remote Sensing 44 (9), 2587–2600.

Bruzzone, L., Chi, M., Marconcini, M., 2006. A Novel Transductive SVM for Semisupervised Classification of Remote-Sensing Images. IEEE Transactions on Geoscience and Remote Sensing 44 (11), 3363–3373.

Bunting, P.J., Lucas, R.M., 2006. The delineation of tree crowns in Australian mixed species forests using hyperspectral Compact Airborne Spectrographic Imager (CASI) data. Remote Sensing of Environment 101 (2), 230-248.

Burges, C.J., 1998. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2 (2), 121–167.

Camps-Valls, G., Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 43 (6), 1351–1362.

Camps-Valls, G., Bruzzone, L., 2009. Kernel Methods for Remote Sensing Data Analysis. John Wiley & Sons, Ltd, Chichester, UK, 403 p.

Cánovas-García, F., Alonso-Sarría, F., 2015. Optimal Combination of Classification Algorithms and Feature Ranking Methods for Object-Based Classification of Submeter Resolution Z/I-Imaging DMC Imagery. Remote Sensing 7 (4), 4651–4677.

Carleer, A.P., Wolff, E., 2006. Urban land cover multi-level region-based classification of VHR data by selecting relevant features. International Journal of Remote Sensing 27 (6), 1035–1051.

Castilla, G., Hay, G.J., 2008. Image objects and geographic objects. In: Blaschke, T., Lang, S., Hay, G.J. (Eds.), Object-Based Image Analysis. Springer, Berlin, Heidelberg, pp. 91–110.

Chapelle, O., Schölkopf, B., 2002. Incorporating invariances in nonlinear support vector machines. In: Dietterich, T.G., Becker, S., Ghahramani Z. (Eds.), Advances in Neural Information Processing Systems. MIT Press, Cambridge, MA, pp. 609-616.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20 (3), 273–297.

Cossu, R., 1988. Segmentation by means of textural analysis. Pixel 1 (2), 21-24.

DeCoste, D., Schölkopf, B., 2002. Training Invariant Support Vector Machines. Machine Learning (46), 161–190.

Dey, V., Zhang, Y., Zhong, M., 2010. A review on image segmentation techniques with remote sensing perspective. In: Wagner W., Székely, B. (Eds.), Proceedings of the International Society for Photogrammetry and Remote Sensing Symposium (ISPRS10), vol. XXXVIII (Part 7A), Vienna, Austria, 5–7 July.

Dópido, I., Li, J., Marpu, P.R., Plaza, A., Bioucas Dias, J.M., Benediktsson, J.A., 2013. Semisupervised Self-Learning for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing 51 (7), 4032–4044.

Dorren, L.K., Maier, B., Seijmonsbergen, A.C., 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. Forest Ecology and Management 183 (1-3), 31-46.

eCognition, 2016. eCognition Developer 9.2 Reference Book. Document Version 9.2.1. Trimble, Munich, Germany.

Fernandez, I., Aguilar, F.J., Aguilar, M.A., Alvarez, M.F., 2014. Influence of Data Source and Training Size on Impervious Surface Areas Classification Using VHR Satellite and Aerial Imagery Through an Object-Based Approach. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (12), 4681–4691.

Foody, G.M., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. Photogrammetric Engineering and Remote Sensing 70 (5), 627-633.

Foody, G.M., 2009. On Training and Evaluation of SVM for Remote Sensing Applications. In: Camps-Valls, G., Bruzzone, L. (Eds.), Kernel Methods for Remote Sensing Data Analysis. John Wiley & Sons, Ltd, Chichester, UK, pp. 85–109.

Foody, G.M., Mathur, A., 2004. A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on Geoscience and Remote Sensing 42 (6), 1335–1343.

Gehler, P.V., Schölkopf, B., 2009. An introduction to kernel learning algorithms. In: Camps-Valls, G., Bruzzone, L. (Eds.), Kernel Methods for Remote Sensing Data Analysis. John Wiley & Sons, Ltd, Chichester, UK, pp. 25–47.

Geiss, C., Taubenböck, H., 2015. Object-Based Postclassification Relearning. IEEE Geoscience and Remote Sensing Letters 12 (11), 2336–2340.

Geiss, C., Jilge, M., Lakes, T., Taubenbock, H., 2016a. Estimation of Seismic Vulnerability Levels of Urban Structures with Multisensor Remote Sensing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (5), 1913–1936.

Geiss, C., Klotz, M., Schmitt, A., Taubenbock, H., 2016b. Object-Based Morphological Profiles for Classification of Remote Sensing Imagery. IEEE Transactions on Geoscience and Remote Sensing 54 (10), 5952–5963.

Gómez-Chova, L., Camps-Valls, G., Bruzzone, L., Calpe-Maravilla, J., 2010. Mean Map Kernel Methods for Semisupervised Cloud Classification. IEEE Transactions on Geoscience and Remote Sensing 48 (1), 207–220.

Gualtieri, J.A., 2009. The Support Vector Machine (SVM) Algorithm for Supervised Classification of Hyperspectral Remote Sensing Data. In: Camps-Valls, G., Bruzzone, L. (Eds.), Kernel Methods for Remote Sensing Data Analysis. John Wiley & Sons, Ltd, Chichester, UK, pp. 49–83.

Hamedianfar, A., Shafri, H.Z.M., 2015. Detailed intra-urban mapping through transferable OBIA rule sets using WorldView-2 very-high-resolution satellite images. International Journal of Remote Sensing 36 (13), 3380–3396.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics 3 (6), 610–621.

Haralick, R.M., 1979. Statistical and structural approaches to texture. Proceedings of the IEEE 67 (5), 786–804.

Hay, G.J., Niemann, K.O., 1994. Visualizing 3-D Texture: A Three Dimensional Structural Approach to Model Forest Texture. Canadian Journal of Remote Sensing 20 (2), 90–101.

Hsu, C.-W., Lin, C.-J., 2002. A comparison of methods for multiclass support vector machines. IEEE transactions on neural networks 13 (2), 415–425.

Hsu, C.W., Chang, C.C., Lin, C.J. 2016. A practical guide to support vector classification. Department of Computer Science, National Taiwan University. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf (Accessed 12 December, 2017)

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory 14 (1), 55–63.

Izquierdo-Verdiguier, E., Laparra, V., Gomez-Chova, L., Camps-Valls, G., 2013. Encoding Invariances in Remote Sensing Image Classification With SVM. IEEE Geoscience and Remote Sensing Letters 10 (5), 981–985.

Jacquin, A., Misakova, L., Gay, M., 2008. A hybrid object-based classification approach for mapping urban sprawl in periurban environment. Landscape and Urban Planning 84 (2), 152-165.

Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. kernlab - An S4 Package for Kernel Methods in R. Journal of Statistical Software 11(9), 1-20. http://www.jstatsoft.org/v11/i09/ (Accessed 10 February, 2017)

Kuhn, M., Contributions from Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y. and Candan C., 2016. caret: Classification and Regression Training. R package version 6.0-71. https://CRAN.R-project.org/package=caret (Accessed 10 February, 2017)

Laliberte, A.S., Rango, A., 2009. Texture and Scale in Object-Based Analysis of Subdecimeter Resolution Unmanned Aerial Vehicle (UAV) Imagery. IEEE Transactions on Geoscience and Remote Sensing 47 (3), 761–770.

Leinenkugel, P., Esch, T., Kuenzer, C., 2011. Settlement detection and impervious surface estimation in the Mekong Delta using optical and SAR remote sensing data. Remote Sensing of Environment 115 (12), 3007–3019.

Lang, S., 2008. Object-based image analysis for remote sensing applications: modeling reality - dealing with complexity. In: Blaschke, T., Lang, S., Hay, G.J. (Eds.), Object-Based Image Analysis. Springer, Berlin, Heidelberg, pp. 3-27.

Li, Y.-F., Zhou, Z.-H., 2015. Towards Making Unlabeled Data Never Hurt. IEEE transactions on pattern analysis and machine intelligence 37 (1), 175–188.

Liu, D., Xia, F., 2010. Assessing object-based classification. Advantages and limitations. Remote Sensing Letters 1 (4), 187–194.

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing 28 (5), 823–870.

Lu, X., Zhang, J., Li, T., Zhang, Y., 2016. A Novel Synergetic Classification Approach for Hyperspectral and Panchromatic Images Based on Self-Learning. IEEE Transactions on Geoscience and Remote Sensing 54 (8), 4917–4928.

Mantero, P., Moser, G., Serpico, S.B., 2005. Partially supervised classification of remote sensing images through SVM-based probability density estimation. IEEE Transactions on Geoscience and Remote Sensing 43 (3), 559–570.

Melgani, F., Bruzzone, L., 2002. Support vector machines for classification of hyperspectral remote-sensing images. IEEE International Geoscience and Remote Sensing Symposium. IGARSS 2002, pp. 506–508.

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on Geoscience and Remote Sensing 42 (8), 1778–1790.

Momeni, R., Aplin, P., Boyd, D., 2016. Mapping Complex Urban Land Cover from Spaceborne Imagery. The Influence of Spatial Resolution, Spectral Band Set and Classification Approach. Remote Sensing 8 (2), 88.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing. A review. ISPRS Journal of Photogrammetry and Remote Sensing 66 (3), 247–259.

Olson, C.E., 1960. Elements of photographic interpretation common to several sensors. Photogrammetic Engineering 26 (4), 651–656.

Pacifici, F., Chini, M., Emery, W.J., 2009. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. Remote Sensing of Environment 113 (6), 1276–1292.

R Core Team, 2016a. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (Accessed 10 February, 2017)

R Core Team, 2016b. foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ... R package version 0.8-67. https://CRAN.R-project.org/package=foreign (Accessed 10 February, 2017)

Richards, J.A., Jia, X., 2006. Remote Sensing Digital Image Analysis. An Introduction. Edition 4. Springer, Berlin, Heidelberg, 439 p.

Rouse, J.W., Haas, R.H. Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. In: Freden, S.C., Mercanti, E.P., Becker, M.A. (Eds), Third Earth Resources Technology Satellite-1 Symposium. Volume 1: Technical Presentations, section A. NASA, USA, pp. 309-317.

Salcedo-Sanz, S., Rojo-Álvarez, J.L., Martínez-Ramón, M., Camps-Valls, G., 2014. Support vector machines in engineering. An overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 4 (3), 234–267.

Schiewe, J. (2002) Segmentation of high-resolution remotely sensed data – concepts, applications and problems. In: Armenakis, C., Lee, Y.C., (Eds), Symposium on Geospatial Theory, Processing and Applications, vol. XXXIV (Part 4), Ottawa, Canada, 9-12 July.

Schölkopf, B., Burges, C., Vapnik, V., 1996. Incorporating invariances in support vector learning machines. In: Malsburg, C. von der, Seelen, W. von, Vorbrüggen, J.C., Sendhoff, B. (Eds), Artificial neural networks - ICANN 96. 6th International Conference, Bochum, Germany, July 16 - 19, 1996. Proceedings. Springer, Berlin, Heidelberg, pp. 47–52.

Schölkopf, B., Smola, A.J., 2002. Learning with kernels. Support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge, MA, 626 p.

Schöpfer, E., Aravena Pelizari, P., Spröhnle, K., 2015. Temporäre Siedlungen: Wenn aus Flüchtlingslagern Städte werden. In: Taubenböck, H., Wurm, M., Esch, T., Dech, S. (Eds.), Globale Urbanisierung. Springer, Berlin, Heidelberg, pp. 71–81.

Stumpf, A., Kerle, N., 2011. Object-oriented mapping of landslides using Random Forests. Remote Sensing of Environment 115 (10), 2564–2577.

Sun, Z., Fang, H., Deng, M., Chen, A., Yue, P., Di, L., 2015. Regular Shape Similarity Index. A Novel Index for Accurate Extraction of Regular Objects from Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing 53 (7), 3737–3748.

Taubenböck, H., Esch, T., Wurm, M., Roth, A., Dech, S., 2010. Object-based feature extraction using high spatial resolution satellite data of urban areas. Journal of Spatial Science 55 (1), 117–132.

Taubenböck, H., Wurm, M., 2015. Ich weiß, dass ich nichts weiß – Bevölkerungsschätzung in der Megacity Mumbai. In: Taubenböck, H., Wurm, M., Esch, T., Dech, S. (Eds.), Globale Urbanisierung. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 171–178.

Tillé, Y. and Matei, A., 2015. sampling: Survey Sampling. R package version 2.7. https://CRAN.R-project.org/package=sampling (Accessed 10 February, 2017)

Tuia, D., Pacifici, F., Kanevski, M., Emery, W.J., 2009a. Classification of Very High Spatial Resolution Imagery Using Mathematical Morphology and Support Vector Machines. IEEE Transactions on Geoscience and Remote Sensing 47 (11), 3866–3879.

Tuia, D., Ratle, F., Pacifici, F., Kanevski, M.F., Emery, W.J., 2009b. Active Learning Methods for Remote Sensing Image Classification. IEEE Transactions on Geoscience and Remote Sensing 47 (7), 2218–2232.

Tuia, D., Volpi, M., Copa, L., Kanevski, M., Munoz-Mari, J., 2011. A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. IEEE Journal of Selected Topics in Signal Processing 5 (3), 606–617.

Tzotsos, A., Argialas, D., 2008. Support Vector Machine Classification for Object-Based Image Analysis. In: Blaschke, T., Lang, S., Hay, G.J. (Eds.), Object-Based Image Analysis. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 663–677.

Vapnik, V.N., 1998. The nature of statistical learning theory. Edition 2. Springer, New York, 314 p.

Volpi, M., Tuia, D., Bovolo, F., Kanevski, M., Bruzzone, L., 2013. Supervised change detection in VHR images using contextual information and support vector machines. Earth Observation and Geoinformation for Environmental Monitoring 20, 77–85.

Zhang, Q., Wang, J., Gong, P., Shi, P., 2003. Study of urban spatial patterns from SPOT panchromatic imagery using textural analysis. International Journal of Remote Sensing 24 (21), 4137–4160.