Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Αγρονόμων Τοπογράφων Μηχανικών
Τομέας Τοπογραφίας

Διερεύνηση τεχνικών κατάτμησης εικόνας, υπολογιστικής νοημοσύνης και εμπείρων συστημάτων στην Τηλεπισκόπηση

Διδακτορική Διατριβή

του

ΑΓΓΕΛΟΥ Κ. ΤΖΩΤΣΟΥ

Επιβλέπων: Δημήτρης Αργιαλάς
Καθηγητής Ε.Μ.Π.

Εργαστήριο Τηλεπισκόπησης
Αθήνα, Ιούνιος 2014

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 2η Ιουνίου 2014.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Δημήτρης Αργιαλάς	Δημήτρης Ρόκος	Βασιλεία Καραθανάση
Καθηγητής Ε.Μ.Π.	Ομ.Καθηγητής Ε.Μ.Π.	Αν.Καθηγήτρια Ε.Μ.Π.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Στέφανος Κόλλιας	Κων/νος Κουτρούμπας	Κων/νος Καράντζαλος
Καθηγητής Ε.Μ.Π.	Ερευνητής Β' Ε.Α.Α.	Λέκτορας Ε.Μ.Π.

(Υπογραφή)

.....
Χαράλαμπος Ιωαννίδης
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....

Άγγελος Τζώτσος

Διπλωματούχος Αγρονόμος Τοπογράφος Μηχανικός Ε.Μ.Π.

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Αγρονόμων Τοπογράφων Μηχανικών
Τομέας Τοπογραφίας
Εργαστήριο Τηλεπισκόπησης

NATIONAL TECHNICAL UNIVERSITY OF ATHENS

SCHOOL OF RURAL AND SURVEYING ENGINEERING

DEPARTMENT OF TOPOGRAPHY

Investigation of Image Segmentation, Machine Learning and Knowledge-Based Expert System Methods in Remote Sensing

Angelos Tzotsos

Thesis for the degree of

Doctor of Philosophy

Advisor: Demetre Argialas
Professor at NTUA

Remote Sensing Laboratory

Athens, June 2014

I would like to dedicate this thesis to my family for their support and understanding during all those years.

Acknowledgements

I would like to thank my advisor, Professor Demetre Argialas for his guidance through the years and for his insightful comments and review of this manuscript.

I would like to thank my better half, Sophia Valanis for editing my English in this manuscript.

I would also like to thank all the members of the dissertation committee and my colleagues at Remote Sensing Laboratory for their useful comments and corrections.

Last but not least, I would like to thank the entire Free and Open Source Software community for providing the shoulders for me to stand on, in order to implement my research.

Abstract

The objective of this research was to research and implement state-of-the-art computer vision and machine learning methods for Object-Based Image Analysis (OBIA), as well as integration with knowledge-based expert systems. The first contribution was the development of a generic image segmentation algorithm, as a low level processing part of an integrated object-oriented image analysis system. The implemented algorithm is called MSEG and can be described as a region merging procedure. The first primitive object representation is the single image pixel. Through iterative pairwise object fusions, which are made at several iterations, called passes, the final segmentation is achieved. The criterion for object merging is a homogeneity cost measure, defined as object heterogeneity, and computed based on spectral and shape features (indices) for each possible object merge. The heterogeneity is then compared to a user defined threshold, called scale parameter, in order for the decision of the merge to be determined. The processing order of the primitive objects is defined through a procedure (Starting Point Estimation), which is based on image partitions, statistical indices and dithering algorithms. MSEG provides several parameters to be defined by the end user. MSEG offers a multi-resolution algorithm which performs segmentations at several levels, and at the same time provides automatic topology of objects within each level and among levels. The algorithm was implemented in C++ and was tested on remotely sensed images of different sensors, resolutions and complexity levels. The results were satisfactory since the produced primitive objects, were compared with other segmentation algorithms and are capable of providing meaningful objects through a follow up classification step.

The second contribution of this research involved the design and development of a region-based multi-scale segmentation algorithm with the inte-

gration of complex texture features. The implemented algorithm is called Texture-based MSEG and it is also a region merging procedure. The first object representation is the single pixel and through several iterations, the final segmentation is achieved. As with simple MSEG, a homogeneity cost measure is defined and computed based on spectral and shape features for each possible object merge. An integration of texture features to the region merging segmentation procedure was implemented through an Advanced Texture Heuristics module. Towards this texture-enhanced segmentation method, complex statistical measures of texture had to be computed based on objects, however, and not on rectangular image regions. The approach was to compute grey level co-occurrence matrices for each image object and then to compute object-based statistical features. The Advanced Texture Heuristics module, integrated new heuristics in the decision for object merging, involving similarity measures of adjacent image objects, based on the computed texture features. The algorithm was implemented in C++ and was tested on remotely sensed images of different sensors, resolutions and complexity levels. The results were promising and comparable to those of other segmentation algorithms. A comparison between the simple algorithm and the texture-based algorithm results showed that in addition to spectral and shape features, texture features did provide good segmentation results.

The third contribution of this research involved the integration of Support Vector Machines (SVM) with OBIA. The Support Vector Machine is a theoretically superior machine learning methodology with great results in pattern recognition. Especially for supervised classification of high-dimensional datasets and has been found competitive with the best machine learning algorithms. In the past, SVMs were tested and evaluated only as pixel-based image classifiers. During recent years, advances in Remote Sensing occurred in the field of Object-Based Image Analysis (OBIA) with combination of low level and high level computer vision techniques. Moving from pixel-based techniques towards object-based representation, the dimensions of remote sensing imagery feature space increases significantly. This results to increased complexity of the classification process, and causes problems to traditional classification schemes. The objective of this study was to evaluate

SVMs for their effectiveness and prospects for object-based image analysis as a modern computational intelligence method. Here, an SVM approach for multi-class classification was followed, based on primitive image objects provided by a multi-resolution segmentation algorithm. Then, a feature selection step took place in order to provide the features for classification which involved spectral, texture and shape information. After the feature selection step, a module that integrated an SVM classifier and the segmentation algorithm was developed in C++. For training the SVM, sample image objects derived from the segmentation procedure were used. The proposed classification procedure followed, resulting in the final object classification. The classification results were compared to the Nearest Neighbor object-based classifier results, and outperformed them.

In the forth contribution of this research, an object-oriented image classification framework was developed which incorporates nonlinear scale space filtering into the multi-scale segmentation and classification procedures. Morphological levelings, which possess a number of desired spatial and spectral properties, were associated with anisotropically diffused markers towards the construction of nonlinear scale spaces. Image objects were computed simultaneusly at various scales and were connected to a kernel-based learning machine for the classification of various earth-observation data from both active and passive remote sensing sensors. Unlike previous object-based image analysis approaches, the developed approach does not require the tuning of any parameter – of those which control the multi-scale segmentation and object extraction procedure, like shape, color, texture etc. – since the scale hierarchy is implicitly derived from scale space representation properties. The developed object-oriented image classification framework was applied on a number of remote sensing data from different airborne and spaceborne sensors including SAR images, high and very high resolution panchromatic and multispectral aerial and satellite datasets. The very promising experimental results along with the performed qualitative and quantitative evaluation demonstrate the potential of the proposed approach.

The fifth contribution of this research involved the development of a multi-

scale object-oriented image analysis framework, which incorporated a region merging segmentation algorithm enchanced by advanced edge features and nonlinear scale space filtering. For the region merging procedure the MSEG multiscale segmentation algorithm was extended. Initialy, edge and line features were extracted from remote sensing imagery at several scales using scale-space representations. These features were used by the enchanced segmentation algorithm as constraints for the growth of image objects at various scales. The first primitive object representation was the single image pixel. Through iterative pairwise object merging, done at several iterations, the final segmentation was achieved. The borders of the images were not permitted to intersect with the egde features thus primitive objects were bounded by the edge features. Image objects were computed at various scales and were passed on to a kernel-based learning machine for their classification into thematic information. The developed object-oriented image classification framework was applied on very high resolution remotely sensed imagery acquired by various airborne and spaceborne panchromatic, multispectral, hyperspectral and microwave sensors. The experimental results showed that this segmentation algorithm outperformed previous MSEG versions. The very promising results along with the performed qualitative and quantitative evaluation demonstrated the contribution of the developed approach.

The sixth contribution of this research involved the implementation of an object-based image classification method, incorporating the Relevance Vector Machine framework for image object classification. The Relevance Vector Machine (RVM) is a kernel classification method, extending the Support Vector Machine (SVM) through Bayessian theory, dealing with uncertainty within the kernel-based classification framework. Recent studies showed very promising classification results in pixel-based classification problems for remote sensing imagery. The goal of this research was to introduce the RVM classification method to Object-based Image Analysis (OBIA) framework and evaluate its effectiveness and prospects. A RVM multi-class classification scheme was developed, based on image primitives, provided by an edge-enhanced multi-resolution segmentation algorithm. The feature space for the classification was defined from a feature selection step, involving

spectral, texture and shape information of the image objects. Then, given a object sample set, the RVM classifier was trained without the need of parameter tuning. The implemented classification step followed resulting in the final object classification. For comparison, a modified version of the SVM classifier was implemented, incorporating probabilistic output to SVM classification. The RVM classification results were compared to both multi-class binary SVM and probabilistic SVM methods. RVM was found to be comparable in accuracy, more sparce and superior in the case of a small training set of image objects. The experimental results were found very satisfactory and the qualitative and quantitative evaluation demonstrated the contributions made through the proposed framework.

Finally, in the seventh contribution of this research, a multimodal object-based image classification approach was developed and evaluated. The goal of this research was to integrate machine learning classification with knowledge-based expert systems, to extend the Object-Based Image Analysis methodology and to evaluate its effectiveness and prospects. The previously developed framework, based on scale-space representation and a state-of-the-art multi-scale image segmentation algorithm provided the primitive image objects, while their spatial and spectral properties formed a multidimensional feature space for the following classification steps. Then, a sparse kernel-based supervised classification method was developed, based on Relevance Vector Machines, which provided the initial spectral classification of the image primitives. This sparse classifier was able to perform with high accuracy even with a small number of sample objects while maintaining the accuracy performance of SVM methods. The classified image objects and their spatial and spectral features were represented and stored in a spatial database, using a vector representation in order to perform spatial analysis and reduce the memory footprint of the system. Then, a knowledge-based expert system was integrated into the spatial database to provide rule-based classification support through a pure expert system tool: CLIPS. The expert system library was wrapped using the Python programming language to the spatial database, and the image objects were bound to the rule-based system for further refinement of their classification. Using expert knowl-

edge representation and the CLIPS rule based environment, a knowledge base was formulated, resulting to the final image object classification. The advantage of this multimodal approach is that the kernel-based and the rule-based classifiers are loosely coupled, interacting on the same data within the spatial database, performing iterative classification refinements as needed in order to achieve superior classification results. This approach was evaluated against other object-based classification methods and some very promising experimental results were demonstrated.

Overall, this disertation contributed in all parts of the Object-Based Image Analysis methodology, introducing and implementing state-of-the-art computer vision, image segmentation, machine learning and knowledge-based methods in the OBIA methodology, thus integrating new tools towards automating the classification and feature extraction tasks in remote sensing. This need for more accurate and automated classification is crucial in remote sensing today, since photointerpretation of big earth observation data is a tedious process. Future prospects of this research include the implementation of the proposed methods in various remote sensing applications.

ΠΕΡΙΛΗΨΗ

Ο στόχος της παρούσας διατριβής ήταν η διερεύνηση και υλοποίηση καινοτόμων μεθόδων Όρασης Υπολογιστών και Υπολογιστικής Νοημοσύνης στα πλαίσια της μεθοδολογίας της Αντικειμενοστρεφούς Ανάλυσης Εικόνας (OBIA). Επίσης στόχος ήταν η ολοκλήρωση των μεθόδων αυτών με τεχνικές βασισμένες στη γνώση, δηλαδή με Έμπειρα Συστήματα. Η πρώτη συνεισφορά της διατριβής αφορούσε στην υλοποίηση ενός πολυκλιμακωτού αλγορίθμου κατάτμησης εικόνας, ο οποίος μπορεί να ενσωματωθεί σε μεθοδολογίες Αντικειμενοστρεφούς Ανάλυσης Εικόνας. Ο αλγόριθμος που αναπτύχθηκε, ονομάστηκε MSEG και είναι ένας αλγόριθμος της κατηγορίας ένωσης περιοχών (region merging). Αρχικά η αναπαράσταση αντικειμένων στην εικόνα ξεκινάει από τα εικονοστοιχεία και μετά απο διαδοχικές συνενώσεις γειτονικών αντικειμένων σε πολλές επαναλήψεις, επιτυγχάνεται η τελική κατάτμηση εικόνας. Το κριτήριο συνένωσης των γειτονικών αντικειμένων, βασίζεται στον υπολογισμό ενός κόστους ομοιογένειας, που ονομάζεται ετερογένεια αντικειμένου και υπολογίζεται από φασματικά και γεωμετρικά χαρακτηριστικά του αντικειμένου που προκύπτει από κάθε πιθανή συνένωση. Η ετερογένεια συγκρίνεται με ένα κατώφλι που ορίζει ο χρήστης και ονομάζεται συντελεστής κλίμακας, ώστε να αποφασιστεί αν θα γίνει ή όχι κάθε συνένωση αντικειμένων. Η σειρά με την οποία ελέγχονται τα αντικείμενα καθορίζεται από έναν αλγόριθμο που βασίζεται σε στατιστικά εικόνας, σε μεθόδους χρωματικής αντιπαράθεσης και ορθογωνιου διαχωρισμού της εικόνας σε μικρότερες περιοχές (blocks). Ο αλγόριθμος MSEG προσφέρει τη δυνατότητα στον χρήστη να ορίσει μια σειρά παραμέτρων που καθορίζουν το σχήμα και το μέγεθος των αντικειμένων της εικόνας. Ταυτόχρονα ο αλγόριθμος παρέχει τη δυνατότητα κατάτμησης σε πολλαπλά επίπεδα, διατηρώντας την τοπολογία των αντικειμένων εσωτερικά ενός επιπέδου αλλά και μεταξύ των διαδοχικών επιπέδων. Ο αλγόριθμος υλοποιήθηκε σε C++ και είναι ελεύθερο λογισμικό, ενω αξιολογήθηκε σε τηλεπισκοπικές απεικονίσεις διαφορετικών δεκτών, διακριτικών ικανοτήτων και διαφορετικής πολυπλοκότητας σημασιολογικού περιεχομένου. Τα αποτελέσματα κρίθηκαν

ικανοποιητικά διότι συγκρίθηκαν επιτυχώς με άλλες μεθόδους κατάτμησης εικόνας, και διότι τα πρωτογεννή αντικείμενα που προέκυψαν δεν περιείχαν λάθη υπο-κατάτμησης πράγμα που υποδηλώνει την δυνατότητα επιτυχούς αντικειμενοστρεφούς ταξινόμησης.

Η δεύτερη συνεισφορά της διατριβής ήταν η διερεύνηση και υλοποίηση ενός πολυκλιμακωτου αλγορίθμου κατάτμησης εικόνας, βασισμένου σε αύξηση περιοχών, με την ολοκλήρωση προηγμένων τεχνικών υφής. Ο αλγόριθμος ονομάστηκε Texture-based MSEG, και αποτελεί επέκταση του αρχικού αλγορίθμου MSEG. Και πάλι η διαδικασία κατάτμησης ξεκινάει από τα εικονοστοιχεία και με διαδοχικές επαναλήψεις και συνενώσεις αντικειμένων προκύπτει το τελικό αποτέλεσμα της κατάτμησης. Η διαδικασία της επιλογής των κατάλληλων συνενώσεων βασίζεται σε χαρακτηριστικά σχήματος, φάσματος και στην συγκεκριμένη περίπτωση σε χαρακτηριστικά υφής δευτέρου βαθμού. Για το λόγο αυτό ο αλγόριθμος υλοποιεί πίνακες συνεμφάνισης (grey level co-occurrence matrices) εντός των αντικειμένων της εικόνας και στην συνέχεια υπολογίζονται τα χαρακτηριστικά υφής από στατιστικά των πινάκων που προκύπτουν. Από τα στατιστικά προκύπτουν μέτρα ομοιότητας με τα οποία ο αλγόριθμος αξιολογεί το αν θα πρέπει να συνενώσει ή όχι δύο ή περισσότερα γειτονικά αντικείμενα. Ο αλγόριθμος υλοποιήθηκε σε C++ και αξιολογήθηκε σε πλήθος τηλεπισκοπικών απεικονίσεων, βελτιώνοντας τα αποτελέσματα κατάτμησης του αρχικού αλγορίθμου MSEG. Ταυτόχρονα τα αποτελέσματα αξιολογήθηκαν σε σύγκριση και με άλλους αλγορίθμους κατάτμησης και κρίθηκαν ικανοποιητικά.

Η τρίτη συνεισφορά της διατριβής ήταν η ολοκλήρωση ενός αλγορίθμου υπολογιστικής νοημοσύνης, των Μηχανών Διανυσματικής Υποστήριξης (Support Vector Machines) στα πλαίσια της Αντικειμενοστρεφούς Ανάλυσης Εικόνας. Οι Μηχανές Διανυσματικής Υποστήριξης θεωρούνται μια άριστη μέθοδος υπολογιστικής μάθησης με εξαιρετικά αποτελέσματα στην Αναγνώριση Προτύπων. Ειδικά σε προβλήματα επιβλεπόμενης ταξινόμησης σε μεγάλους χώρους προτύπων, έχει αποδειχθεί ότι είναι μια από τις αποτελεσματικότερες μεθόδους με πολύ καλά αποτελέσματα. Πριν από αυτή την έρευνα, οι Μηχανές Διανυσματικής Υποστήριξης

είχαν χρησιμοποιηθεί στην τηλεπισκόπηση σαν ταξινομητές για εικονοστοιχεία, με αποτελέσματα που υπερτερούν των κλασικών μεθόδων. Με την μετατόπιση των τεχνικών ανάλυσης εικόνας από τις κλασικές μεθόδους προς τις Αντικειμενοστρεφείς μεθόδους, οι διαστάσεις του χώρου προτύπων για τις τηλεπισκοπικές απεικονίσεις αυξήθηκε δραματικά. Αυτό οδήγησε σε αυξημένη πολυπλοκότητα, στην οποία οι υπάρχουσες μεθοδολογίες επιβλεπόμενης ταξινόμησης δεν μπορούσαν να αντεπεξέλθουν. Σε αυτή τη διατριβή για πρώτη φορά υλοποιήθηκε Αντικειμενοστρεφής Ταξινόμηση Εικόνας μέσω της τεχνολογίας των Μηχανών Διανυσματικής Υποστήριξης με σκοπό την διερεύνηση της αποτελεσματικότητάς τους και της προοπτικής τους σαν μια μεθοδολογία αιχμής. Στην παρούσα έρευνα υλοποιήθηκε μια τεχνική ταξινόμησης πολλαπλών τάξεων με βάση τις Μηχανές Διανυσματικής Υποστήριξης. Από τα πρωτογεννή αντικείμενα που προκύπτουν από την κατάτμηση εικόνας, πραγματοποιείται εξαγωγή χαρακτηριστικών (φασματικά, γεωμετρικά, χωρικά κλπ) και στη συνέχεια σχηματίζεται ο χώρος προτύπων. Για την εκπαίδευση των Μηχανών Διανυσματικής Υποστήριξης, επιλέγονται δείγματα εκπαίδευσης με βάση επίγειες παρατηρήσεις και ο ταξινομητής εκπαιδεύεται. Στην συνέχεια, όλα τα αντικείμενα της εικόνας ταξινομούνται παρέχοντας το τελικό αποτέλεσμα. Τα αποτελέσματα της μεθοδολογίας βρέθηκε να υπερτερούν όλων των προηγούμενων μεθόδων αντικειμενοστρεφούς ταξινόμησης, επαληθεύοντας το ότι η μέθοδος SVM θεωρείται μια από τις πιο επιτυχείς τα τελευταία χρόνια στην Υπολογιστική Νοημοσύνη.

Η τέταρτη συνεισφορά της διατριβής ήταν μια μεθοδολογία Αντικειμενοστρεφούς Ανάλυσης Εικόνας με ενσωμάτωση προηγμένων τεχνικών μη ισοτροπικής διάχυσης και φιλτραρισμάτων χώρου-κλίμακας. Τα φιλτραρίσματα χώρου-κλίμακας ενσωματώθηκαν στη διαδικασία κατάτμησης εικόνας βελτιώνοντας τα αποτελέσματα τόσο της κατάτμησης όσο και της μετέπειτα ταξινόμησης. Συγκεκριμένα, τα μορφολογικά επιπεδοσύνολα τα οποία διαθέτουν ιδιότητες διατήρησης πληροφορίας στην εικόνα (όπως οι ακμές) παρά την ομαλοποίηση την οποία επιφέρουν, υλοποιήθηκαν σαν ένα βήμα προεπεξεργασίας των τηλεπισκοπικών απεικονίσεων, πρίν από το βήμα της κατάτμησης. Στην συνέχεια ακολούθησε κατάτμηση

σε πολλαπλές κλίμακες και ταξινόμηση με Μηχανές Διανυσματικής Υποστήριξης. Σε αντίθεση με προηγούμενες μεθόδους, στην προτεινόμενη μεθοδολογία δεν απαιτείται βελτιστοποίηση των παραμέτρων κατάτμησης από το χρήστη (π.χ. χρώμα, σχήμα, κανονικότητα κλπ) καθώς η ιεραρχία κλίμακας προκύπτει έμμεσα από τις ιδιότητες των μορφολογικών επιπεδοσυνόλων μέσω της ελλαχιστοποίησης της τοπικής ετερογένειας των αντικειμένων της εικόνας. Η προτεινόμενη μεθοδολογία εφαρμόστηκε σε όλα τα είδη των τηλεπισκοπικών απεικοίσεων, περιλαμβανομένων αερομεταφερόμενων σαρωτών, δορυφορικές εικόνες, εικόνες SAR, υπερφασματικές και πολυφασματικές εικόνες. Τα πολύ ικανοποιητικά αποτελέσματα από την ποιοτική και ποσοτική αξιολόγηση, δείχνουν την αποτελεσματικότητα της μεθόδου και την δυνητική της χρήση για αυτόματη εξαγωγή χαρακτηριστικών από τηλεπισκοπικά δεδομένα.

Η πέμπτη συνεισφορά της διατριβής αφορούσε στην ολοκλήρωση προηγμένων τεχνικών ανίχνευσης ακμών στην διαδικασία κατάτμησης εικόνας για την υλοποίηση μεθοδολογίας Αντικειμενοστρεφούς Ανάλυσης Εικόνας. Ο συνδιασμός των μορφολογικών επιπεδοσυνόλων, της πολυκλιμακωτής κατάτμησης και της πληροφορίας ακμών της εικόνας έδωσε μια νέα υβριδική προσέγγιση στην κατάτμηση εικόνας. Για την πολυκλιμακωτή κατάτμηση εικόνας χρησιμοποιήθηκε και επεκτάθηκε ο αλγόριθμος MSEG. Αρχικά από την εικόνα πραγματοποιείται εξαγωγή γραμμικών χαρακτηριστικών, με την υλοποίηση της μεθόδου LSD ενώ οι αρχικές εικόνες έχουν εξομαλυνθεί με ανισοτροπική διάχυση σε διάφορες κλίμακες. Τα γραμμικά χαρακτηριστικά στην συνέχεια ενσωματώθηκαν στον αλγόριθμο MSEG σαν περιορισμοί συνένωσης περιοχών. Αρχικά ο αλγόριθμος ξεκινά από τα εικονοστοιχεία και με διαδοχικές συνενώσεις παράγει την τελική κατάτμηση της εικόνας. Κατά τη διαδικασία αυτή αν ανιχνευτεί ακμή ανάμεσα από δυο αντικείμενα της εικόνας, η συνένωση εμποδίζεται από τον τροποποιημένο αλγόριθμο κατάτμησης με αποτέλεσμα τα αντικείμενα που βρίσκονται εκατέρωθεν ακμών να μην συνενώνονται. Στην συνέχεια, τα πρωτογεννή αντικείμενα της εικόνας στις διαφορετικές κλίμακες ταξινομούνται με Μηχανές Διανυσματικής Υποστήριξης και επιλέγεται ως τελική ταξινόμηση αυτή που παρέχει την μεγαλύτερη ακρίβεια, με

βάση ποσοτικά κριτήρια. Η μεθοδολογία αυτή υπερτερεί στην κατάτμηση της εικόνας από τις προηγούμενες όπως υπέδειξε η ποιοτική και ποσοτική αξιολόγηση. Τα πρωτογεννή αντικείμενα των εικόνων ήταν εμφανώς βελτιωμένα ώς προς τα όρια και το σχήμα τους και αυτό το υπέδειξε και η βελτιωμένη ακρίβεια στα αποτελέσματα της ταξινόμησης. Η μέθοδος κατάτμησης εικόνας με ενσωμάτωση των ακμών είναι και η τελικά προτεινόμενη μεθοδος κατάτμησης εικόνας από την παρούσα έρευνα.

Η έκτη συνεισφορά της διατριβής αφορούσε την υλοποίηση Αντικειμενοστρεφούς Ταξινόμησης με μεθόδους που βρίσκονται στην αιχμή της επιστήμης της Υπολογιστικής Νοημοσύνης. Για πρώτη φορά προτάθηκε στην διατριβή αυτή η υλοποίηση αντικειμενοστρεφούς μεθόδου ταξινόμησης με βάση τις Μηχανές Διανυσμάτων Συνάφειας (Relevance Vector Machines). Οι Μηχανές Διανυσμάτων Συνάφειας είναι μια μέθοδος ταξινόμησης βασισμένη σε πυρήνες (kernel-based) που επεκτείνει την μεθοδολογία των Μηχανών Διανυσματικής Υποστήριξης σύμφωνα με την θεωρία του Bayes, χρησιμοποιώντας την έννοια των πιθανοτήτων σαν αποτέλεσμα της ταξινόμησης προτύπων ενσωματώνοντας έτσι την έννοια της αβεβαιότητας στην μεθοδολογία. Η μεθοδολογία αυτή είχε πρόσφατα χρησιμοποιηθεί για ταξινόμηση τηλεπισκοπικών απεικονίσεων με βάση τα εικονοστοιχεία με επιτυχή αποτελέσματα, αλλά δεν είχε χρησιμοποιηθεί για αντικειμενοστρεφή ταξινόμηση στο παρελθόν. Στην παρούσα έρευνα υλοποιήθηκε μεθοδολογία κατά την οποία πραγματοποιείται κατάτμηση της αρχικής εικόνας, ενώ στην συνέχεια υπολογίζεται ο χώρος προτύπων της ταξινόμησης με βάση τα φασματικά, γεωμετρικά και χωρικά χαρακτηριστικά των πρωτογεννών αντικειμένων. Οι Μηχανές Διανυσμάτων Συνάφειας εκπαιδεύονται με βάση κάποια δείγματα και στην συνέχεια πραγματοποιείται η τελική ταξινόμηση. Το πλεονέκτημα της μεθόδου είναι ότι σε αντίθεση με τις Μηχανές Διανυσματικής Υποστήριξης δεν απαιτείται η αναζήτηση παραμέτρων του εσωτερικού μοντέλου μάθησης. Πραγματοποιήθηκε σύγκριση της προτεινόμενης μεθόδου με την μέθοδο των Μηχανών Διανυσματικής Υποστήριξης και τα αποτελέσματα δεν είχαν μεγάλες αποκλίσεις. Οι Μηχανές Διανυσματικής Υποστήριξης απέδωσαν οριακά καλύτερα αλλά οι Μηχανές Διανυσμάτων Συνάφειας χρειάστηκαν πολύ λιγότερα δείγματα

και δεν απαιτούσαν την επίπονη διαδικασία της παραμετροποίησης. Μετά από ποσοτική και ποιοτική αξιολόγηση, τα αποτελέσματα της καινοτόμου μεθόδου κρίθηκαν ικανοποιητικά.

Τέλος η έβδομη συνεισφορά της διατριβής αφορούσε στην ολοκλήρωση της Αντικειμενοστρεφούς Ανάλυσης Εικόνας με συστήματα που βασίζονται στη γνώση. Στα πλαίσια αυτής της έρευνας, στόχος ήταν η διασύνδεση των προηγμένων μεθόδων Υπολογιστικής Νοημοσύνης με έμπειρα συστήματα που βασίζονται στη γνώση ώστε να διερευνηθεί και να αξιολογηθεί η χρησιμότητά τους σε εφαρμογές Τηλεπισκόπησης. Με βάση τις προηγούμενες μεθοδολογίες της διατριβής, χρησιμοποιήθηκε ο αλγόριθμος κατάτμησης εικόνας με βάση τις ακμές για την εξαγωγή πρωτογεννών αντικειμένων, ενώ μετά τον ορισμό του χώρου προτύπων από τον υπολογισμό χαρακτηριστικών των αντικειμένων, χρησιμοποιήθηκαν οι Μηχανές Διανυσμάτων Συνάφειας για την αρχική ταξινόμηση των αντικειμένων με βάση δείγματα εκπαίδευσης. Στην συνέχεια τα αποτελέσματα της κατάτμησης και της ταξινόμησης των αντικειμένων ενσωματώθηκαν σε μια χωρική βάση δεδομένων με τον μετασχηματισμό τους από εικονιστικά σε διανυσματικά δεδομένα και αποθηκεύοντας τα χαρακτηριστικά τους με τη μορφή πινάκων που ενσωματώνουν πληροφορία γεωμετρίας. Αυτό βελτίωσε την αναπαράσταση των δεδομένων μέσω τεχνολογίας Γεωγραφικών Συστημάτων Πληροφοριών (GIS) και ελλάτωσε τις απαιτήσεις αποθήκευσης των δεδομένων. Επιπλέον η χωρική βάση δεδομένων έδωσε την δυνατότητα εφαρμογής χωρικών ερωτημάτων για την ανάλυση των αντικειμένων της εικόνας. Στην συνέχεια η χωρική βάση δεδομένων διασυνδέθηκε με ένα εργαλείο εμπείρου συστήματος (CLIPS) ώστε να παρέχει τη δυνατότητα ταξινόμησης των αντικειμένων της εικόνας με κανόνες και την αναπαράσταση της ανθρώπινης γνώσης (φωτοερμηνευτών) που μπορεί να ενσωματωθεί σε ένα τέτοιο σύστημα. Για την διασύνδεση χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python. Η βασικη συνεισφορά του Εμπείρου Συστήματος στην Αντικειμενοστρεφή Ανάλυση Εικόνας είναι η δυνατότητα που προσδίδει στον χρήστη να εκφράζει εμπειρικούς κανόνες ή κανόνες που προέρχονται από τη βιβλιογραφία ώστε να προσομοιώνει την γνώση του έμπειρου φωτοερμηνευτή και να διορθώνει τα αποτελέσματα της ταξινόμησης με επαναληπτικό τρόπο.

Για την αξιολόγηση της μεθόδου, πραγματοποιήθηκαν αρχικές ταξινομήσεις με αλγορίθμους υπολογιστικής νοημοσύνης και στη συνέχεια με τη χρήση κανόνων τα αποτελέσματα βελτιώθηκαν τόσο ποιοτικά όσο και ποσοτικά. Το πλεονέκτημα της μεθόδου είναι ότι η σειρά των μεθόδων είναι στην κρίση του έμπειρου χρήστη του συστήματος και μπορεί να υλοποιήσει σύνθετες μεθόδους για την λύση δύσκολων τηλεπισκοπικών προβλημάτων, που ήταν αδύνατη χωρίς την ενσωμάτωση της γνώσης στην διαδικασία της ταξινόμησης. Η προτεινόμενη μεθοδολογία αξιολογήθηκε μέσα από τη σύγκριση με παλιότερες μεθόδους και τα αποτελέσματά της βρέθηκαν πολύ ικανοποιητικά.

Σε γενικές γραμμές η διατριβή είχε συνεισφόρα σε όλες τις διαδικασίες της Αντικειμενοστρεφούς Ανάλυσης Εικόνας και η ενσωμάτωση μεθόδων όρασης υπολογιστών, κατάτμησης εικόνας, υπολογιστικής νοημοσύνης και εμπείρων συστημάτων ολοκληρώνει τις δυνατότητες ενός σύγχρονου τηλεπισκοπικού συστήματος, προσδίδοντάς του πολλές δυνατότητες αυτοματοποίησης επίπονων διαδικασιών για τον φωτοερμηνευτή. Μελλοντικά, ανοίγονται οι ορίζοντες για χρήση της προτεινόμενης μεθοδολογίας σε πιο συγκεκριμένες εφαρμογές Τηλεπισκόπησης.

Contents

Contents xxi

List of Figures xxv

Nomenclature xxx

1 Introduction 2

1.1 Recent developments in Remote Sensing 2

1.2 Object-Based Image Analysis 3

1.3 Research Objectives 6

1.4 Contributions 7

1.5 Dissertation Overview 10

2 A Generic Region-Based Multi-Scale Image Segmentation Algorithm for Remote Sensing Imagery 12

2.1 Introduction 12

2.2 Methodology 13

2.2.1 Algorithm Requirements 13

2.2.2 The MSEG specification (Simple Profile) 14

2.2.2.1 Data Input and Macroblock Estimation modules . . . 15

2.2.2.2 Starting Point Estimation module 15

2.2.2.3 Definition of heterogeneity 18

2.2.2.4 First Pass module 20

2.2.2.5 Second Pass module 22

2.2.2.6 Nth Pass module 22

2.2.3 The MSEG Advanced Profile . . . 23
2.2.3.1 Multi-scale Algorithm module . . . 23
2.2.3.2 Global Heterogeneity Heuristics module . . . 25
2.3 Discussion of results . . . 26
2.4 Conclusions and Future Developments . . . 28

3 A hybrid texture-based and region-based multi-scale image segmentation algorithm 32
3.1 Introduction . . . 32
3.1.1 Recent developments in Remote Sensing . . . 32
3.1.2 Texture-based Image Segmentation and Object-based Image Analysis . . . 33
3.1.3 Research Objectives . . . 33
3.2 Methodology . . . 34
3.2.1 MSEG algorithm overview . . . 34
3.2.2 Advanced Texture Heuristics . . . 36
3.2.3 Implementation . . . 40
3.3 Discussion of Results . . . 41
3.4 Conclusions and future work . . . 46

4 Support Vector Machine Classification for Object-Based Image Analysis 48
4.1 Introduction . . . 48
4.1.1 Knowledge-based image classification and Object Oriented Image Analysis . . . 48
4.1.2 Computational Intelligence methods in Remote Sensing . . . 49
4.1.3 Research objectives . . . 50
4.2 Methodology . . . 50
4.2.1 Multi-scale segmentation . . . 50
4.2.2 Support Vector Machines . . . 50
4.2.3 SVM Multi-class Classification . . . 53
4.2.4 Implementation . . . 53
4.3 Discussion of Results . . . 55
4.4 Conclusions . . . 61

5 Object-based image analysis through nonlinear scale space filtering **64**
5.1 Introduction . . . 64
5.2 Related Work . . . 67
5.2.1 Scale space representations for remote sensing image analysis . 67
5.2.2 Object-based Image Analysis . . . 69
5.3 Methodology . . . 71
5.3.1 Anisotropic Morphological Levelings . . . 73
5.3.2 Multiscale Segmentation . . . 75
5.3.3 Support Vector Machines . . . 76
5.4 Evaluation and Discussion . . . 78
5.4.1 Radar satellite imagery . . . 79
5.4.2 Very high spatial resolution airborne imagery . . . 79
5.4.3 Multispectral remote sensing data . . . 82
5.5 Conclusions and Future Perspectives . . . 88

6 Multiscale segmentation and classification of remote sensing imagery with advanced edge and scale-space features **94**
6.1 Introduction . . . 94
6.2 Related Work . . . 96
6.2.1 Multiscale object-based image analysis . . . 96
6.2.2 Scale space remote sensing data representations . . . 98
6.2.3 Edge-based segmentation . . . 100
6.3 Methodology . . . 101
6.3.1 Scale-Space Filtering . . . 103
6.3.2 Multiscale Segmentation based on Advanced Edge Features . . 103
6.3.3 Kernel-based Classification . . . 107
6.4 Evaluation and Discussion . . . 108
6.4.1 Very high spatial resolution airborne imagery . . . 110
6.4.2 Radar satellite imagery . . . 112
6.4.3 Multispectral remote sensing data . . . 114
6.4.4 Hyperspectral remote sensing data . . . 117
6.5 Conclusions and Future Perspectives . . . 120

7 Relevance Vector Machines for Object-Based Image Analysis 122
7.1 Introduction . . . 122
7.2 Related Work . . . 124
7.2.1 Kernel-based Classification in Remote Sensing . . . 124
7.2.2 Support Vector Machines and Object-based Image Analysis . . . 126
7.3 Methodology . . . 127
7.3.1 Multiscale Image Segmentation . . . 128
7.3.2 Relevance Vector Machines . . . 130
7.3.3 Probabilistic Support Vector Machines . . . 133
7.4 Evaluation and Discussion . . . 134
7.4.1 Very high spatial resolution airborne imagery . . . 136
7.4.2 Multispectral satellite imagery . . . 138
7.5 Conclusions and Future Perspectives . . . 144

8 Integrating knowledge-based expert systems and advanced machine learning for object-based image analysis 146
8.1 Introduction . . . 146
8.2 Related Work . . . 148
8.2.1 Knowledge-based Expert Systems in Remote Sensing . . . 148
8.2.2 Machine Learning in Remote Sensing . . . 150
8.3 Methodology . . . 152
8.3.1 Multiscale Image Segmentation . . . 153
8.3.2 Relevance Vector Machine Classification . . . 155
8.3.3 Knowledge-based Expert System Classification . . . 157
8.4 Evaluation and Discussion . . . 162
8.4.1 Very high spatial resolution airborne imagery . . . 163
8.4.2 Multispectral satellite imagery . . . 170
8.5 Conclusions and Future Perspectives . . . 171

9 Conclusions 176

References 182

List of Figures

2.1 The MSEG flowchart with module strategy 16

2.2 The initial Landsat image (a) (eCognition User Guide 2005) with starting points from the HSI module overlaid and (b) Starting Points as estimated from the Floyd & Steinberg error diffusion algorithm (dithering) [Hawley, 1990; Ulichney, 1987] 17

2.3 Starting points as estimated from original image (a) and starting points as estimated from a subset of the original image (b). Small differences are visible and some are marked with red arrows. 18

2.4 The segmentation levels of the Multi-scale Algorithm cases presented graphically. Colors represent object ID's (non duplicate integers inside levels). (a) The bottom-up approach. (b) The top-down approach. (c) The interpolation approach. 25

2.5 The original Dessau image (eCognition User Guide 2005)(a) and (b) A first segmentation result of MSEG algorithm (Simple Profile) using scale parameter 400. 29

2.6 (a) Original Landsat TM dataset in RGB composite. (b) MSEG result with scale parameter 100, (c) eCognition's result with scale parameter 10, (d) MSEG result with scale parameter 400, (e) eCognition's result with scale parameter 20 and (f) MSEG result with scale parameter 700. 30

2.7 (a) Original aerial scan data (Toposys 2005) (b) Segmentation results with MSEG algorithm for scale parameter of 400. 31

3.1 Nth pass general flowchart 37

3.2 Left: An example of GLCM computation at 0 degree angle for a 4x4 window. The empty GLCM gets filled by adding co-occurrences symmetrically. Right: A 3-dimensional representation of the Co-occurrence matrices that have to be computed for a given orientation. Ng is the number of grey levels and N is the total number of primitive image objects 39
3.3 Segmentation result as provided by eCognition for scale parameter 10 (left) and MSEG for scale parameter 400 (right) 42
3.4 Segmentation result as provided by texture-based MSEG for scale parameter 400 and texture parameter 2.0 (left) and for scale parameter 400 and texture parameter 1.0 (right) 42
3.5 Segmentation result by eCognition for scale parameter 20 (left) and texture-based MSEG for scale parameter 2500 and texture parameter 3.0 (right) 43
3.6 Segmentation result by eCognition for scale parameters 15 (left) and 25 (right). 44
3.7 Segmentation result by simple MSEG for scale parameters 400 (left) and 700 (right). 45
3.8 Segmentation result by Texture-based MSEG for scale parameters 400 (left) and 2500 (right). 45

4.1 Left: The case of linear separable classes. Right: The case of non linear separable classes. ξ measures the error of the hyperplane fitting. 51
4.2 (a) the original Landsat TM image (source: eCognition User Guide 2005). (b) The training set of class samples(blue=Water, red=Impervious, green=Woodland and yellow=Grassland. (c) The cross-validation plot diagram for selecting the optimal values of C and γ for the SVM training. (d) The ground-truth dataset used to evaluate results 56
4.3 (a) eCognition classification result with Nearest Neighbor. (b) MSEG classification result with SVM (c) eCognition classification result with Nearest Neighbor after introducing sample errors. (d) MSEG classification result with SVM. Errors have been introduced to the training sets for generalization evaluation. 57

4.4 (a) the original aerial scanner image (source: Toposys GmbH) (b) The training set of class samples (white=bright roofs, grey=asphalt like materials, green=vegetation and orange=tile roofs). (c) The cross-validation plot diagram for selecting the optimal values of C and γ for SVM training. (d) The ground-truth dataset used to evaluate results. 60
4.5 Left: eCognition classification result with Nearest Neighbor. Right: MSEG classification result with SVM. Training sample overlap with objects set to 50% 61

5.1 Comparing the scale space representation of linear and non-linear techniques. Two scales from each technique are presented, along with the corresponding image contours (isophotes)(c,m,n,o,p) and the coarser scale (i,j,k,l) segments (q,r,s,t). For the multi-resolution (MR) approach the scales (Sc) of 0.8 and 0.4 are shown computed by a down-sampling of the initial image. For the Gaussian smoothing the scales (standard deviation values) of 2 and 7 are presented and for the non-linear approaches ALM and AML the scales (iterations) of 50 and 500. One can observe that both non-linear approaches (AML, ALM) do not blur image edges, do not produce new extrema and more effectively and accurately preserve image contours. 72
5.2 The developed framework embeds nonlinear scale space filtering into the object-based image classification procedure. 73
5.3 AML filtering results on DMC aerial multispectral image with 5 cm pixel-size (©Intergraph Corp.) 74
5.4 Applying the ALM and AML non-linear scale spaces to a TerraSAR-X (©DLR) dataset (3 meters ground resolution, StripMap mode, polarisation HH). At all scales the AML simplifies the initial image and stays closer to the initial intensity values. 80
5.5 TerraSAR-X (©DLR) segmentation results with and without the AML filtering using the MSEG and the eCognition software 81
5.6 Segmentation tests as performed on the DMC (©Intergraph Corp.) aerial image, with and without the application of nonlinear filtering 83

5.7 Segmentation tests as performed on the DMC (©Intergraph Corp.) aerial image, with and without the application of non-linear filtering 84
5.8 A segmentation test as performed on a Landsat TM image, with and without the application of non-linear filtering 85
5.9 Original multispectral Landsat TM image (Dessau, Germany) and classification results with and without AML filtering. Blue: Water Bodies, Yellow: Grassland, Green: Woodland, Red: Impervious 86
5.10 Sensitivity analysis regarding the accuracy of the classification result for different scales of filtering and segmentation. 89
5.11 Original multispectral aerial scan image (©Toposys) and classification results with and without AML filtering. Green:Vegetation, Grey: Asphalt materials, Orange: Tile roofs, White:Bright roofs 90

6.1 Comparing region merging segmentation results using scale space representations and advanced edge features. Four scales from various steps of the developed methodology are presented. The first column (a,f,k,p) presents the initial aerial image with spatial resolution of 5cm, along with three gaussian scales. The second column (b,g,l,q) presents the scale-space representation at various selected scales. The third column (c,h,m,r) presents initial image objects using the MSEG algorithm applied to the scale-space representation. The forth column (d, i ,n ,s) presents the extracted edge and line features used in the following step. The final column (e,j,o,t) shows the results of the edge enhanced MSEG algorithm developed in this research. 106
6.2 Comparing various segmentation algorithms on a DMC aerial multispectral image with 5 cm pixel-size (©Intergraph Corp.). (a) Standard MSEG with scale parameter 100. (b) Mean-Shift segmentation with default parameters. (c) Multiresolution segmentation (eCognition) with default parameters. (d) Edge constrained MSEG, without merging the edge objects in last pass, for demonstration purposes (e) Edge constrained MSEG, employing the Canny edge features. (f) Edge constrained MSEG, employing the LSD edge features. 111

6.3 Comparing various segmentation algorithms on a TerraSAR-X (©DLR) dataset (3 meters ground resolution, StripMap mode, polarisation HH). (a) The initial image. (b) Mean-Shift segmentation with default parameters. (c) Watershed segmentation with default parameters. (d) Canny edge detection applied on the AML scale-space representation. (e) Standard MSEG results with scale parameter 400. (f) Edge constrained MSEG, employing the Canny edge features. 115

6.4 Comparing various segmentation algorithms on a Landsat TM dataset (Dessau, Germany). (a) Standard MSEG results with scale parameter 100. (b) Mean-Shift segmentation with default parameters. (c) Multiresolution segmentation (eCognition) with default parameters (scale 10, shape 0.1). (d) Canny edge detection applied on the AML scale-space representation. (e) Edge constrained MSEG, with Canny edge features used and scale parameter 100. (f) Edge constrained MSEG, with Canny edge features used and scale parameter 400. 116

6.5 Comparing various segmentation algorithms on a QuickBird dataset (Eastern Attika, Greece). (a) Original image. (b) Mean-Shift segmentation with default parameters. (c) Watershed segmentation with default parameters. (d) Multiresolution segmentation (eCognition) with default parameters (e) Canny edge detection applied on the AML scale-space representation. (f) LSD line features extracted from the original image. (g) Standard MSEG results with scale parameter 100. (h) Edge constrained MSEG, with Canny edge features used and scale parameter 100. (i) Edge constrained MSEG, with LSD line features used and scale parameter 100. 118

6.6 Comparing various segmentation algorithms on a CASI Hyperspectral dataset (©Remote Sensing Laboratory, NTUA) with 95 spectral bands (Axios river, Thessaloniki, Greece). (a) Original image. (b) Mean-Shift segmentation with default parameters. (c) Watershed segmentation with default parameters. (d) Standard MSEG results with scale parameter 900. (e) Standard MSEG on AML scale-space representation and scale parameter 900. (f) Edge constrained MSEG, with Canny edge features used and scale parameter 900. 119

7.1 Original multispectral aerial scan image (©Toposys), simplification with scale-space AML algorithm, edge extraction with Line Segment Detector and image segmentation with edge-enhanced MSEG algorithm, producing the primitive image objects for further machine learning classification 137
7.2 Machine Learning classification of multispectral aerial scan image (©Toposys). 139
7.3 Original multispectral Landsat TM image (Dessau, Germany), simplification with scale-space AML algorthm, edge extraction with Line Segment Detector and image segmentation with edge-enhanced MSEG algorithm, producing the primitive image objects for further machine learning classification 141
7.4 Machine Learning classification of multispectral Landsat TM image (Dessau, Germany). 143

8.1 Original multispectral aerial scan image (©Toposys) simplified by AML algorthm, edge extraction with Line Segment Detector, image segmentation with edge-enhanced MSEG algorithm and ground truth data for evaluation of further machine learning classification 164
8.2 Image objects and their spectral and shape attributes visualized in QGIS. 165
8.3 Machine Learning and knowledge-based classification of multispectral aerial scan image (©Toposys) with the developed OBIA framework. . 167
8.4 Original multispectral Landsat TM image (Dessau, Germany) simplified with scale-space AML algorithm, edge extraction with Canny edge detector, image segmentation with edge-enhanced MSEG algorithm and ground truth data for classification evaluation. 172
8.5 Machine Learning and knowledge-based classification of multispectral Landsat TM image (Dessau, Germany) with the developed OBIA framework. 173

Chapter 1

Introduction

1.1 Recent developments in Remote Sensing

In recent years, the evolution of remote sensing sensors and technologies, along with the advancements in computer science has lead into an explosion of availability and size of earth observation data [Baumann, 2010]. Remotely sensed data are becoming more accurate and detailed, every time a new satellite sensor is launched. From big satellite sensors to smaller airborn sensors, to minicopters and UAVs, big geospatial data are produced every day, increasing the need to automate the processing and the classification of large datasets [Berni et al., 2009]. To that direction, considerable research is targeted in remote sensing methods including state-of-the-art computer vision and artificial intelligence techniques in order to tackle the size and resolution of the produced "big data".

At the same time, remote sensing has achieved great progress in image analysis algorithms [Blaschke, 2010]. Due to very high resolution imagery, such as IKONOS and Quick Bird, traditional classification methods which are tailored for lower resolution imagery, have become less effective given the magnitude of heterogeneity appearing in the spectral feature space of such imagery. The spectral heterogeneity of imaging data has increased rapidly, and the traditional methods exhibit the "salt and pepper" phenomenon in their classification results. Such problems occur also to medium resolution satellite data, such as Landsat TM, SPOT etc.

Another issue of traditional classification methods is that they do not use information related to shape, site and spatial relation (context) of the objects of the scene. Context

information is a key element to photo-interpretation, and a key feature used by all photo-interpreters because it encapsulates expert knowledge about the image objects [Argialas and Harlow, 1990]. Such knowledge is not explicit and needs to be represented and used for image analysis purposes. Shape and texture features are used extensively as context descriptors in photo-interpretation.

Research has progressed in computer vision methods applied to remotely sensed images such as segmentation, object oriented and knowledge-based methods for classification of high-resolution imagery [Kanellopoulos et al., 1999]. In Computer Vision, image analysis is considered in three levels: low, medium and high. Such approaches were usually implemented in separate software environments since low and medium level algorithms are procedural in nature, while high level is inferential and thus for the first procedural languages are best suitable while for the second an expert system environment is more appropriate.

The current need for automated image analysis and computer vision technological tools requires a processing scheme able to encapsulate effectively the content of remote sensing data. However, earth's landscape structure is complex, the context varies and so does the appearance of the images being a combination of many different intensities, representing natural features such as vegetation, geomorphological and hydrological features, man-made objects (e.g buildings, roads) and artifacts caused by variation in illumination of the terrain (e.g shadows).

Furthermore, roads, infrastructure, vegetation, landforms and other land features appear in different sizes and geographical scales in images (e.g. country road versus interstate, tree stands versus forest, maisonette versus polygon building, rill versus river, etc). Only in a few special circumstances the objects of interest belong to a certain scale while the remaining ones, to be discarded, to another. In most cases such a global scale threshold is not possible since the desired information is present at several scales [Lindeberg, 1994; Meyer and Maragos, 2000a; Weickert, 1998; Witkin, 1983].

1.2 Object-Based Image Analysis

Along with the gradual availability of earth observation data with higher spatial and spectral resolution, research efforts in classifying remote sensing data have been shifting in the last decade from pixel-based approaches to object-based ones [Blaschke, 2010; Myint

et al., 2011; Tzotsos et al., 2011]. Assigning land cover classes to individual pixels can be intuitively proper and functional for low resolution data. However, this is not the case for the emerging applications which arise from the continuously improving remote sensing sensors [Aplin and Smith, 2008; Blaschke et al., 2008]. This is mostly because at higher resolutions it is a connected group of pixels that is likely to be associated with a land cover class and not just an individual pixel [Tzotsos et al., 2011].

In addition, the earth surface exhibits various regular and irregular structures which are represented with a certain spatial heterogeneity in images composing their intensity, scale and texture. Several important aspects of earth surface can not be analyzed based on pixel information, but can only be exploited based on contextual information and the topologic relations of the objects of interest [Liu et al., 2008] through a multiscale image analysis [Benz et al., 2004; Blaschke and Hay, 2001; Dragut et al., 2010; Duarte Carvajalino et al., 2008; Hall and Hay, 2003; Hay et al., 2002; Jimenez et al., 2005; Ouma et al., 2008; Stewart et al., 2004; Tzotsos et al., 2011]. Starting with the observed spatial heterogeneity and variability, meaningful spatial aggregations (primitive object) can be formed at certain image scales configuring a relationship between ground objects and image objects. Ground objects refer to real-world objects or areas of specific land-cover class, connected by complex spatial and contextual relations. Image objects, on the other hand, refer to knowledge-free areas (primitive objects) of an image that a segmentation algorithm provides, based on various criteria. In order to configure a relationship between ground and image objects, a multiscale knowledge representation is needed, which is usually provided by the object-oriented paradigm. With such an object-based multiscale representation and analysis, which is based on certain hierarchically structured rules, the relationship between the different scales of the spatial entities is described.

The last decade, the Object-Based Image Analysis (OBIA) paradigm has been the subject of very active research and has developed into a state-of-the-art methodology in remote sensing [Blaschke, 2010]. This new methodology integrated low level image analysis methods, such as multiresolution segmentation procedures and algorithms [Baatz and Schape, 2000], with high level methods, such as Artificial Intelligence (knowledge-based expert systems and fuzzy systems) and Pattern Recognition methods. Within this approach, the low level image analysis produces primitive image objects, while the high level processing classifies these primitives into meaningful domain objects [Benz et al., 2004]. The result of this paradigm shift was to move away from pixel-based methods,

leading to advancements in classification results and accuracy, bringing remote sensing information closer to GIS systems and spatial databases [Baatz and Schape, 2000; Benz et al., 2004; Blaschke, 2010].

The main advantage of this new approach is that the digital image is no longer considered as a grid of pixels, but as a group of primitives and homogeneous regions, called image objects. Using this kind of representation, one can tackle the problem of using local criteria for classification, as this kind of features can be computed within the boundaries of an image object. Also, context information can easily be represented through topologic features between the objects. Objects can be more intelligent than pixels, in a sense of knowing their neighbours and the spatial or spectral relations within and among them. Because, primitive objects are knowledge-free, as a result of low level procedures, in order to accomplish thematic classification, one has to use higher level (knowledge-based) techniques.

The semantic objects of an image do not belong to a single but to various spatial scales. The use of scale space image representations is thus of fundamental importance for a number of image analysis and computer vision tasks. Following the ideas of Witkin [1983], Koenderink [1984] and Lindeberg [1994], many methods have been introduced to derive linear scale spaces and respectively many isotropic multi-scale operators have been developed. Either through Gaussian filtering or through isotropic multi-resolution analysis (e.g. by down-sampling the initial data), all linear scale space approaches present the same important drawback: image edges are blurred and new non-semantic objects may appear at coarse filtering scales [Ouma et al., 2008; Paragios et al., 2005; Witkin, 1983]. Under a hierarchical multi-scale segmentation or an object-based classification framework, the thematic information to be extracted is directly related with the primitive image objects computed at every segmentation scale. The better these primitive objects represent real-world entities, the better they can describe the semantics of the image [Blaschke et al., 2008; Hay and Castilla, 2006; Hofmann et al., 2008]. Therefore, the selection of the appropriate approach for constructing the multi-scale image and the hierarchical object representation is of great importance.

In this dissertation, the developed framework is based on the Object-Based Image Analysis (OBIA) approach, which involves low, medium and high level image processing sub-tasks:

- Preprocessing steps (geometric and radiometric corrections, filtering, image sim-

plification, edge detection, band math expression computations, etc.),

- Image Segmentation (in order to produce single-level or multi-level hierarchies of primitive objects within the image space),
- Computation of image object properties based on spectral, shape, topological and context features (feature space),
- Definition of object-oriented class hierarchy and representation of knowledge through rules or through training with samples,
- Classification (learning techniques or rule-based systems to perform the classification task),
- Accuracy assessment in order to derive the quality of the resulting classification,
- Vectorization steps (create the output to spatial databases and integrate the information to thematic maps).

1.3 Research Objectives

The main objective of this research effort was to implement a novel image classification framework being able to process effectively a wide variety of remote sensing data, such as hyperspectral and multispectral data from ground, aerial and spaceborne sensors, radar data, digital elevation models, etc.

Another objective of this research was to introduce state-of-the-art computer vision (such as scale-space, advanced edge features etc.) and machine learning methods (such as Support Vector Machines, Relevance Vector Machines etc.) to the OBIA approach, improving the OBIA image segmentation and object classification steps.

An additional objective of this research was the development and implementation of a new object-oriented image segmentation algorithm as a low level processing part of an object-oriented image analysis system so that to be applied at multiple image resolutions and to produce objects of multiple scales (sizes), minimizing the need of user-customizable parameters. Another goal of this research was the integration of a

Knowledge-based Expert System framework to the developed segmentation and classification framework so that to improve the knowledge representation for the OBIA classification task.

The main motivation for this research was to provide an Object-Based Image Analysis system in the form of Free and Open-Source Software (FOSS).

1.4 Contributions

Generally, this dissertation was focused on improving the OBIA methodology by incorporating state-of-the-art methods and techniques in the segmentation and classification of remotely sensed imagery.

In short, the contribution of this dissertation was:

- Three (3) published chapters in Remote Sensing books [Tzotsos and Argialas, 2008; Tzotsos et al., 2008, 2014].
- One (1) publication in highly rated Remote Sensing journal [Tzotsos et al., 2011]
- Two (2) pending (or in press) publications in highly rated Remote Sensing journals
- Three (3) peer reviewed publications in conferences [Argialas and Tzotsos, 2006; Tzotsos, 2006; Tzotsos and Argialas, 2006]
- More than 125 citations, especially for the introduction of Support Vector Machines to the Object-Based Image Analysis methodology (50+ citations)

More specifically, this research introduced seven main contributions: three (3) algorithms in image segmentation, three (3)methods in object-based machine learning classification and one (1) method in expert systems for remote sensing. Those contributions are described in detail in the following chapters in chronological order.

The first contribution was the investigation and implementation of a new region merging multi-scale image segmentation algorithm, named MSEG. This algorithm was implemented in 2006 [Tzotsos and Argialas, 2006] and introduced several enhancements to the multi-scale segmentation algorithm as proposed by Baatz and Schape [2000]. A main advantage of the MSEG algorithm was that it was available as Free and Open Source Software, while there was no reference implementation available freely for the Baatz and

Schape [2000] algorithm. The results of the proposed algorithm were 100% reproducable and they were evaluated as comparable to the ones provided by the commercial software.

The second contribution was the design and implementation of a region-based multi-scale segmentation algorithm with the integration of complex texture features [Tzotsos et al., 2008]. In this effort, the MSEG algorithm was extended to include second order texture features that were computed within the boundaries of an image primitive object instead of the whole image, as in previous studies. The approach was to compute grey level co-occurrence matrices for each image object and then to compute object-based statistical features in order to perform iterative object merging until the final image segmentation was achieved. This approach improved the MSEG algorithm results, as it was suggested by the qualitative results. The drawback was the computational cost for deriving the second order texture statistics.

The third contribution was the introduction of Support Vector Machines to the methodology of Objec-Based Image Analysis [Tzotsos and Argialas, 2008]. Until 2006, SVMs were applied only in pixel based studies in Remote Sensing. Commercial and proprietary OBIA software did not include kernel-based classification methods at that time and there was a need for evaluating the use of sparse methods in classification of image objects for remote sensing applications. In 2006, the first SVM application to OBIA was published [Tzotsos, 2006] and it was followed by a more detailed presentation in 2008 [Tzotsos and Argialas, 2008]. In that effort, MSEG algorithm was used to derive a set of image objects and then a set of features were derived from the objects representing spectral, spatial and shape characteristics of the objects. Then training samples were used to form the training and cross-validation of the SVM leading to classification of all objects from the kernel-based classifier. The application of SVMs to OBIA provided very promising results and more proprietary and free software integrated this functionality the following years.

The forth contribution was the implementation of an OBIA methodology, introducing scale-space representation and anisotropic filtering in the preprocessing steps of OBIA, but also in the segmentation procedure [Tzotsos et al., 2011]. By transforming the original remote sensing imagery into the scale-space domain, it was shown that the tuning of segmentation parameters was not needed in the following steps, thus making the segmentation procedure more robust and user friendly. In previous efforts, an exhaustive

parameter selection was needed for image segmentation, leading to the representation of image primitives into object hierarchies. The proposed method provided native scale-space hierarchies where only one scale parameter is involved. The proposed methodology provided very good classification results combined with a SVM classifier.

The fifth contribution was the integration of advanced edge features in the OBIA segmentation process [Tzotsos et al., 2014]. A state-of-the-art computer vision algorithm was integrated to the MSEG region merging procedure, acting as a boundary/constraint of the merging process. The LSD line detector was implemented, but other edge detectors can also be integrated easily, like the Canny algorithm. This integration helped maintain better shape properties for primitive objects, while raising the scale parameter and producing larger image objects. This was also combined with the scale-space algorithm in order to eliminate heterogeneity of the image. The algorithm was improved overall, outperforming existing multiscale segmentation solutions.

The sixth contribution of this dissertation was the integration of Relevance Vector Machines in object-based classification. This methodology was introduced for the first time, combining state-of-the-art machine learning methods in the OBIA framework (pending publication). The method showed very promising results, in the sense that smaller training set is needed to derive a classification of similar accuracy as the one produced by SVM. Also, since RVM is based on Bayessian theory a probabilistic output was an additional advantage of the machine learning framework.

The last contribution of this research was the development and the evaluation of a Knowledge-Based Expert System tool, within the OBIA framework, as a post-classification step to improve the results provided by machine learning techniques and integrate human knowledge before the production of a thematic map. This process moves from the raster space to the vector space, using spatial databases as the base technology and exporting the image objects and their features to spatial vector representation. Then, SVM and RVM were used to classify geospatial database objects, updating the assigned class of each object. Then, the CLIPS expert system software was integrated to the spatial database in order to be able to parse the image objects using the Rete algorithm and production rules, based on spectral, spatial and context information. The obtained results were evaluated against previous OBIA solutions and were very promising (pending publication in OBIA 2014).

1.5 Dissertation Overview

The above contributions are presented in detail in the following chapters. Each chapter of the dissertation is derived from a published work as described previously and is presented in chronological order instead of thematic order. In Chapter 1, an introduction of the dissertation is presented and the contributions are described. In Chapter 2, the original MSEG implementation is presented without the texture and edge enhancements. Then in Chapter 3, the texture enhancements of the MSEG algorithm are discussed and evaluated. In Chapter 4, the machine learning part of the dissertation is introduced with the integration of Support Vector Machines to the OBIA framework. Then in Chapter 5, the scale-space integration is discussed and the framework to avoid parameter tuning during segmentation is introduced. In Chapter 6, the edge enchancements to the segmentation algorithm are presented, leading to the final version of the MSEG algorithm for this research. In Chapter 7, the Relevance Vector Machines are introduced and evaluated for OBIA classification. In Chapter 8 the integration of Knowledge-Based Expert Systems is presented, as a post-classification (or even simple classification) step for deriving semantic objects using properly represented image interpretation knowledge. Finally in Chapter 9 the conclusions of the dissertation are provided, along with the prospects for future research in the OBIA field.

Chapter 2

A Generic Region-Based Multi-Scale Image Segmentation Algorithm for Remote Sensing Imagery

2.1 Introduction

During the last few years, a new approach, called Object-Based Image Analysis, integrated low level image analysis methods, such as multiresolution segmentation procedures and algorithms [Baatz and Schape, 2000], with high level methods, such as Artificial Intelligence (knowledge-based expert systems and fuzzy systems) and Pattern Recognition methods. Within this approach, the low level image analysis produces primitive image objects, while the high level processing classifies these primitives into meaningful domain objects [Benz et al., 2004].

The main advantage of this new approach is that the digital image is no longer considered as a grid of pixels, but as a group of primitives and homogeneous regions, called image objects. Using this kind of representation, one can tackle the problem of using local criteria for classification, as this kind of features can be computed within the boundaries of an image object. Also, context information can easily be represented through topologic features between the objects. Objects can be more intelligent than pixels, in a sense of knowing their neighbours and the spatial or spectral relations within and among them. Because, primitive objects are knowledge free, as a result of low level procedures, in or-

der to accomplish thematic classification, one has to use higher level (knowledge-based) techniques.

Various segmentation algorithms and methods have been proposed in the literature over the last decades, with promising results e.g. [Pal and Pal, 1993; Sonka et al., 1998]. The decision for a segmentation technique to be included into an Object-Oriented Image Analysis System is based on the particular use of the system and the problems it has to deal with. In remote sensing, a preferable choice would be a multi-scale image segmentation algorithm. Remote sensing applications, deal with data of various resolutions and complexity, so a generic and less computation-intensive segmentation algorithm is a better choice [Baatz and Schape, 2000]. The main purpose of segmentation in Object Oriented Image Analysis is not the extraction of semantic objects, but the extraction of image primitives. There are, however, applications where complex algorithms, with more specific and less generic use, can result into semantic information [Chen et al., 2002; Fauzi and Lewis, 2003; Havlicek and Tay, 2001; Liapis et al., 1998].

The main objective of this research was the development and implementation of a new object-oriented image segmentation algorithm as a low level processing part of an object-oriented image analysis system so that to be applied at multiple image resolutions and to produce objects of multiple scales (sizes), according to user-customizable parameters.

A further objective was the ability for the produced algorithm to be generic and produce good and classification-ready results to as many remote sensing data as possible. Remote sensing data are, in general, difficult to process, with complex texture and spectral information. Therefore, there was a need for the algorithm to be able to handle shape information and context features in order to produce better results as perceived by image interpreters.

2.2 Methodology

2.2.1 Algorithm Requirements

Because the main objective of the new algorithm was to be used as a pre-classification processing step for remote sensing imagery, several requirements were set during the design stage of the algorithm. Since remote sensing images do not follow specifications for spectral or spatial resolution, the algorithm should be able to process all possible reso-

lutions of data and produce primitive objects which are capable of providing successful classification results upon further steps of high-level image analysis.

Another basic requirement for the segmentation algorithm was to provide primitive objects at different scales. When a photo interpretation task is carried out by an expert, the scale of imagery is specified by the nature of image semantics to be recognized. During the higher level image processing steps, there will be a need to have primitive objects of different sizes and preferably, on different scales of abstraction. Using several layers of abstraction, the algorithm should provide the classification task with features which encapsulate higher level of knowledge (context information) about the objects to be identified. To implement this, the algorithm should be able to produce multiple level hierarchies of objects with a known topology.

Another requirement for typical object oriented image analysis tasks is the ability of the segmentation procedure to have reproducible results for the same image, with the same parameters. If the way, that primitive objects are produced, changes dramatically with small parameter changes, then the result is prone to errors and highly unstable. This could result in non-robust knowledge-bases, working only in very specific cases.

Due to the significance of the primitive objects to the final classification result, the segmentation algorithm should be planned in a way, which does not oppose the edge information of image semantics. This means that the final boundaries of the primitive objects should not be intersecting in any way with virtual boundaries of semantic regions that an expert photo interpreter would have digitized. To that direction, as will be explained on the following sections, measures of heterogeneity were used, and hybrid segmentation approaches were included in the specification of the segmentation algorithm.

Finally, secondary requirements for the algorithm were computational efficiency, customization of parameters for optimization purposes, use of non spectral segmentation criteria and flexibility of implementation standards.

2.2.2 The MSEG specification (Simple Profile)

In order to comply with the above requirements, the MSEG algorithm was designed to be a region merging technique, since region merging techniques are fast, generic and can be fully automated (without the need of seed points) [Pal and Pal, 1993; Sonka et al., 1998]. Given that existing Object-Oriented Image Analysis systems [Benz et al., 2004]

have used such methods was also a strong argument for the effectiveness of the region merging techniques.

In order to combine all the required features into an algorithm, the object-oriented programming model was adopted in order to build interactive modules, with many parameters for optimizing the final result. In Figure 2.1, the general organization of the MSEG modules is presented.

2.2.2.1 Data Input and Macroblock Estimation modules

After a data input stage, an image partitioning procedure was applied to the dataset resulting into rectangular regions of variable dimensions, called macroblocks. The selected partitioning method has been applied in the past to model local variations of images and local features like texture and homogeneity [Fauzi and Lewis, 2003; Liapis et al., 1998; Sonka et al., 1998] as well as for splitting and merging. In the MSEG algorithm, image partitioning was applied for computation of local statistics and for starting points. Starting points were then used for initialization of the algorithm and for the reproducibility of segmentation results.

2.2.2.2 Starting Point Estimation module

After the Macroblock Estimation, the macroblocks were pushed to the Starting Point Estimation module (SPE) to calculate local statistics and provide starting points for initialization of the region merging process. It should be stretched that starting points are not used as seed points (as in region growing techniques) but are used to keep track of the order in which all pixels are initially processed.

There are two basic methods implemented within MSEG for starting point estimation. The first is a statistical method, aimed to produce one starting point per macroblock. For a default macroblock size of 16x16 pixels, a histogram is computed as well as image statistics. Then, scanning the macroblock from the top left corner to the bottom right, the first pixel with the maximum spectral distance from the median value is selected. This pixel is selected ad hock without any previous knowledge about the image or its statistical properties. Using this statistical method, for a given macroblock, the same pixel will be selected if the above procedure is repeated as many times as it is desirable, thus ensuring reproducibility of starting points. There is also a choice of where this algorithm

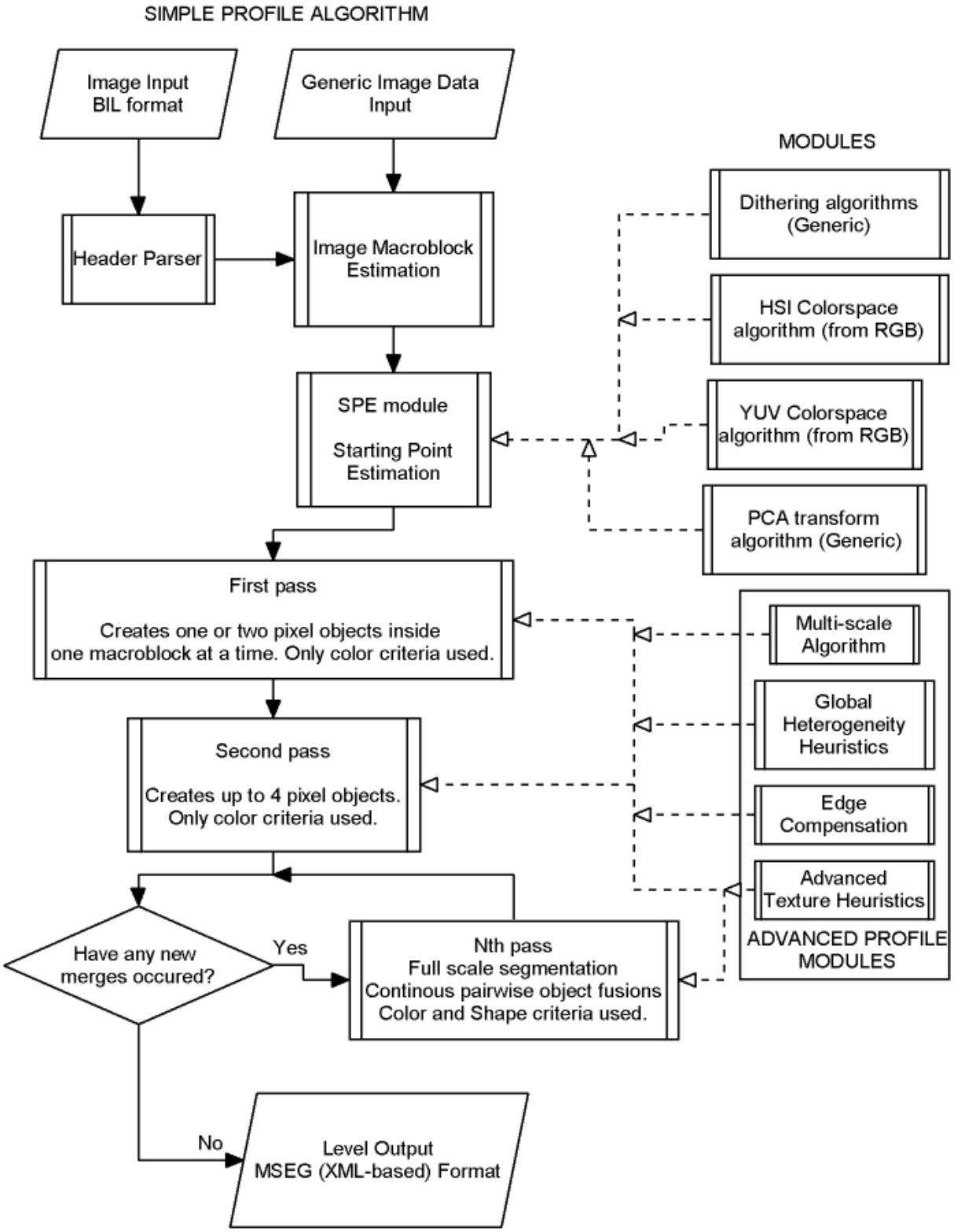


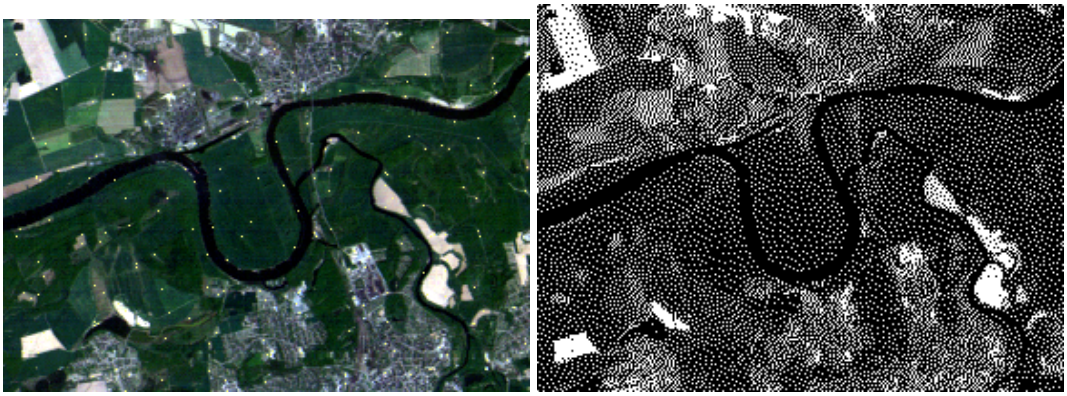
Figure 2.1: The MSEG flowchart with module strategy

is applied to. Using three modules, the statistical method for starting points estimation can be applied either to the intensity (I) band (after the original image is transformed to the HSI colorspace), either to the intensity (Y) band of the YCbCr colorspace or to one of the principal components of the original image (preferably to the first component). In case of a one band image or a non RGB image, the statistical procedure can be applied to one selected band.

The second method implemented for starting point estimation, was based on a dithering or halftoning algorithms, transforming the original image into a binary image through statistical procedures [Hawley, 1990; Ulichney, 1987].

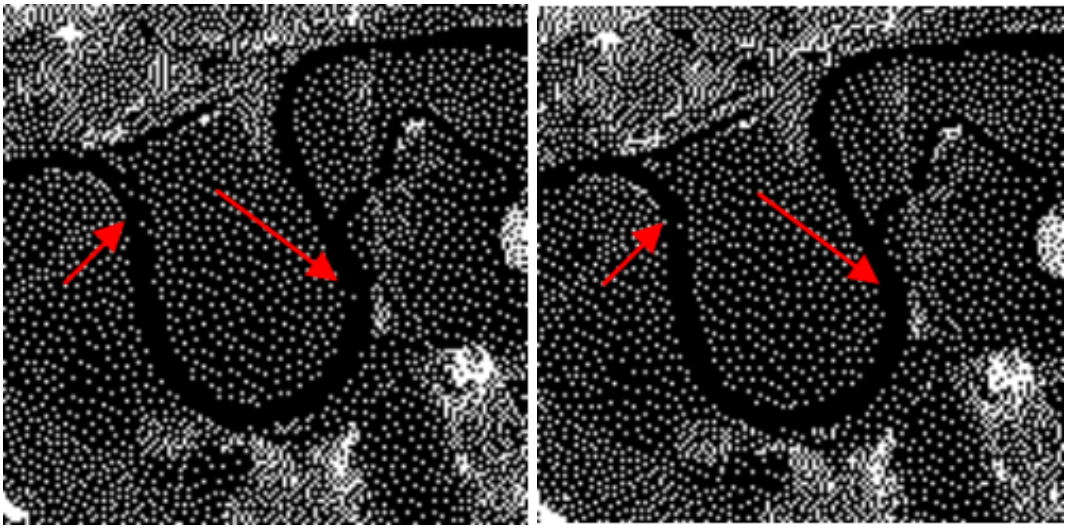
The binary image was used to determine the starting points within each macroblock. Thus this method produced more than one starting points per macroblock, increasing the subsequent computational cost for initialization, and also provided a very good starting point estimator alternative when tested similar (but not the same) images as it provided similar starting points (Figure 2.3). For example if it is needed to test the segmentation algorithm on a subset of the original image, the starting points will be approximately the same, however, even in this case the results will not be 100% identical.

In Figure 2.2, the SPE module starting points are presented while in Figure 2.3, are being compared to those produced in a subset of the original image, where dithering has produced almost identical results.



(a) (b)

Figure 2.2: The initial Landsat image (a) (eCognition User Guide 2005) with starting points from the HSI module overlaid and (b) Starting Points as estimated from the Floyd & Steinberg error diffusion algorithm (dithering) [Hawley, 1990; Ulichney, 1987]



(a) (b)

Figure 2.3: Starting points as estimated from original image (a) and starting points as estimated from a subset of the original image (b). Small differences are visible and some are marked with red arrows.

2.2.2.3 Definition of heterogeneity

MSEG is basically defined as a region merging technique. Like all algorithms of this kind, it is based on several local or global criteria and heuristics, in order to merge objects in an iterative procedure, until no other merges can occur [Sonka et al., 1998]. In most cases, it is typical to select a cost function, to define how good and stable an object is after a merging procedure, or to make the decision regarding that merge according to the cost function. As a result, various definitions of homogeneity (energy minimization measures or measures of similarity within an object) have been defined [Pal and Pal, 1993; Sonka et al., 1998]. Recently, a very successful segmentation algorithm, embedded in the Object Oriented Image Analysis Software eCognition [Baatz and Schape, 2000], implemented such measures of homogeneity, for making the merging decision between neighbouring objects, with very good results. Some spectral and spatial heuristics were also used to further optimize the segmentation.

In the proposed segmentation algorithm, similar homogeneity measures were used so that to simulate eCognition's results, and then to apply extra heuristics (as part of the advanced profile of MSEG). Basically, heterogeneity is defined as a cost measure

of homogeneity for each primitive object of the image. The heterogeneity h simulates the homogeneity cost of the whole scene as a result of merging two image sub-regions. This cost is computed for every possible merge and if it is relatively small, defined by a threshold (scale parameter f), the two regions are marked as compatible, and therefore they are characterized as an object match.

Let us consider an image object (sub-region) within a k-dimension feature space. A heterogeneity measure, in this case, can be defined by a simple function in such a way that it does not depend on the feature values themselves, but rather it depends on the difference of the feature values for each dimension of the feature space. A statistical measure with this property is the variance or standard deviation. For example, for a k-band digital image, a simple measure of heterogeneity could be:

$$h = \sum_k \sigma_k \tag{2.1}$$

If it is desirable for every feature dimension (band) to contribute to the heterogeneity unequally, different weights can be applied to the above equation. Equation 2.1 treats every object in the same manner, without being able to incorporate the information of their size and shape. An easy and effective way to include the size of the object to the above equation, and to constrain the object from growing without final size limit, is to include its area n to the equation. The so derived heterogeneity cost function depends only on spectral values, thus it is defined as spectral heterogeneity:

$$h_s = \sum_k w_k \cdot n \cdot \sigma_k \tag{2.2}$$

To define heterogeneity, not only based on spectral measures, but also using shape indices, based on the boundary shape of objects, Baatz and Schape [2000] used two shape criteria and those were included in the MSEG algorithm. The first shape heterogeneity criterion is introduced to model the compactness of an image object. It is defined as the ratio of the perimeter of the object (l) to the perimeter of a square, with the same area as the object. If n is the area of the object, then compactness heterogeneity is defined according

to the following equation:

$$h_{comp} = \frac{l}{4 \cdot \sqrt{n}} \tag{2.3}$$

The second shape heterogeneity criterion is defined as the ratio of the perimeter of the object to the perimeter of the bounding rectangle of the object. In this case, if l is the perimeter of the object and b is the perimeter of the rectangle, $l \geq b$ is always true. When $l \to b$, then the object's perimeter is smooth and the heterogeneity is approximating the value 1.

$$h_{smooth} = \frac{l}{b} \tag{2.4}$$

Now, let's assume that O1 and O2 are two initial objects, and Om is the object created by their merge. The color heterogeneity of the merge is computed as:

$$h_{color} = h_s^{O^m} - (h_s^{O^1} + h_s^{O^2}) \tag{2.5}$$

In the same manner, the smoothness and compactness heterogeneities are calculated. Then the shape heterogeneity is computed by the following equation:

$$h_{shape} = w_{comp} \cdot h_{comp} + w_{smooth} \cdot h_{smooth} \tag{2.6}$$

Finally, the total heterogeneity h of the merge is computed as:

$$h_{total} = w_{color} \cdot h_{color} + w_{shape} \cdot h_{shape} \tag{2.7}$$

where w are weights, defined by the user. The total heterogeneity h_{total} is compared with the maximum allowed value of heterogeneity, which is the scale parameter threshold f, and if $h_{total} \leq f$ then, the virtual merge is successful (object match).

2.2.2.4 First Pass module

The purpose of the first segmentation pass was to initialize image objects and to provide the first over-segmentation, in order for the algorithm to be able to begin region merging at following stages. Initially, the objects of the image are the single pixels. During first pass,

the algorithm is merging single pixels-objects pair wise, inside each macroblock. The merging is made only inside macroblocks for speed efficiency, since the image is over-segmented and two-pixel objects at most can be created from the merging procedures (Figure 2.1). The macroblock spatial restriction will not be used, during later stages. It is a good speed to quality tradeoff, while it is only used for the first pass.

For each macroblock the starting point (or starting points in the dithering case) is pushed inside a structure called priority list. This priority list tracks the sequence of the objects that have been processed in the past, and must be processed in the future. Each point that enters the priority list, causes a domino effect to its neighbors, adding them to the list and providing them with a lower priority, in order for them to be processed in the near future. For the object with the highest priority occurs a virtual merge with each one of its neighbors. The merging heterogeneities are computed for every virtual merge. If there isn't a single match, the object is pushed back to the list with lower priority, to be processed again later. Should one or more matches occur, then the winner is the one with the lowest merging heterogeneity, and from now on will be referred as "best match". Then, the above procedure is repeated for the "best match" object. After merging virtually, this object with all it's neighbors, it is checked whether it is a case of mutual best match with the initial object (this time a match is certain to occur, at least with the initial object). If there is a mutual best match, a new object is being created and the old objects get replaced by the new one. If there is not a best match, both initial and best match objects are pushed back in the priority list with lower priorities and wait to be treated again later. If an object is treated for a number of times without achieving merging, then it is deactivated for the first pass, and stays as single pixel object waiting the next pass to be activated again.

The above procedure is repeated until there is no object in the priority list for the macroblock and is repeated for all macroblocks, until all pixels are processed for the first pass. Thus, after the first pass, it is possible to have one- or two-pixel objects, as one merge per object is allowed. Shape criteria are not computed for the first pass since the shape criteria of a single-pixel object or a two-pixel object, even if they are computed, they can not differentiate the merging result. Within the first pass (and later passes), the topology of the objects is known at all times, using a dynamic pointer structure.

2.2.2.5 Second Pass module

For the second pass of the merging algorithm, the objects created by the first pass are used in a new pair wise merging procedure. Again, the same strategy of merging is used, finding the best match for each object, and then checking if there is a mutual best match in order to merge the two objects. In the second pass, since the objects can grow up to four pixel size, only the color heterogeneity is computed and used.

During the second pass, the search of the matching objects is not restricted within the macroblock structure, thus it is possible to merge objects from different macroblock areas. In order to control the growth of objects across the image, there is a small modification to the priority list for tracking the sequence of the object treatment. At this time the priority list is using an extra priority feature to control spatial distribution of the objects across the image. The new priority list consists of two priority mechanisms, one for distance priority and one for the sequence priority. Thus, it simulates multiple different threads of merging across the image, which have a controlled distance between them, in order for the image objects to grow equally across the scene.

When the second pass begins, all objects containing a starting point, are given the highest priority within this modified priority list. Thus, the initial state of the merging is controlled and can be fully reproduced as the starting points don't change from one pass to another. All neighbors, during the virtual merging step, are pushed within the priority list, and thus, all objects will be certainly treated by the algorithm.

2.2.2.6 Nth Pass module

The Nth pass module, is called this way, because there is no knowledge of how many times (n times) it will iterate until the algorithm converges. The algorithm is considered finished, when during the nth pass no more merges occur and the algorithm converges (Figure 2.1). Then, the objects are exported and marked as final primitives. Compared to the second pass module, only small changes have been introduced in the nth pass module. The main difference was that all heterogeneity criteria were used (color and shape) for merging selection since large objects with complex shape can occur.

2.2.3 The MSEG Advanced Profile

The core of the MSEG algorithm (Simple Profile) included, the pass modules, as the basic elements of a region merging segmentation procedure. The extension of the Simple Profile was used to include extra functionality algorithms and innovative techniques for improving the results. The Advanced Profile, as implemented at present, included the following modules:

- the Multi-scale Algorithm (MA), and
- the Global Heterogeneity Heuristics (GHH)

2.2.3.1 Multi-scale Algorithm module

The Multi-scale Algorithm module was designed to give to the MSEG algorithm the ability to create multiple instances of segmentations for an image, with different scale parameters. Thus, the produced primitive objects could vary in size and therefore, the multi-scale representation could model large image entities, as well as small ones. In order to include multiple instances of segmentation, inside an object-oriented image analysis system, those instances must be properly constrained to be integrated and used together.

The first problem, that has to be addressed, when dealing with multiple segmentations, is the compatibility between scales, in order to combine information and objects. One simple way to deal with this problem is to create a multi-level representation, and incorporate the multiple segmentations within this representation, hierarchically. A level can be defined as a grid with the same size as the image grid, but with object ids (representing object primitives at each scale) instead of pixel values. The level hierarchy keeps the levels of segmentation sorted by object size. In this approach, a hierarchy of levels can hold the information of topology in different scales, and produce intra-level context information. Along with inter-level topology that is created among objects during passes, the context information is fully supported for the classification step.

The above solution, may give flexibility for combining segmentation levels, but has a drawback. For the intra-level topology, to be stable and easily manipulated, there has to be a rule of compatibility between object boundaries across levels. This means that larger scale objects, have to include an integer number of sub-objects, and their boundaries must

not intersect with sub-object boundaries. If this is not the case, then the context information of "part-of" relationships, as defined in an object-oriented knowledge-based system, cannot be activated, thus bringing the expert system classification to a disadvantage. In the present Multi-scale segmentation Algorithm, every new level depends only from the nearest (scale-wise) super-level or the nearest sub-level, or both. This approach produces three cases for level creation:

The bottom-up approach. Let's suppose a segmentation level is created (for example with scale parameter 10), and another level is to be created with larger objects (for example with scale parameter 50). In order to keep compatibility between these two levels inside a hierarchy structure, the second level is initialized from the first. This means that only nth passes, with initial objects those of the first level, are executed to create the larger objects of the second level (Figure 2.4a). If a larger level is to be created afterwards (with scale parameter 100), it will be initialized by the second level (Figure 2.4a).

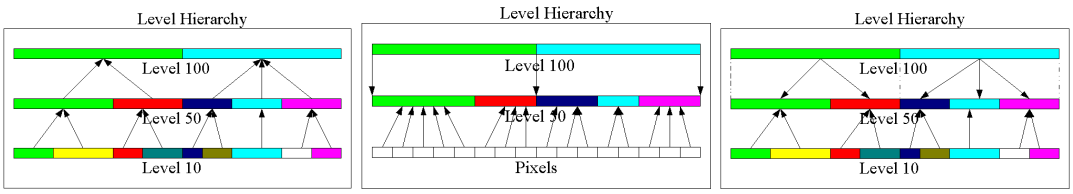
The top-down approach. Let's assume a large segmentation level has been created (for example with scale parameter 100) and a smaller level is to be created (for example with scale parameter 50). In that case, the first pass initiates from pixel objects, but before each merge is finalized, the Multi-scale Algorithm applies a constraint, so that the two pixels must have the same super-object at the larger level. Only objects with same super-object are virtually merged to find mutual best matches. The same applies for the second, and the nth passes (Figure 2.4b). The final level will obey the boundaries of the super-level and will have smaller sized objects, due to the scale parameter.

The interpolation approach. Let's assume a large segmentation level and a small segmentation level have been created (with scale parameters 100 and 10 respectively). If a medium size level is to be created (scale parameter 50), then a combination of the first two approaches is used to interpolate the new level between the other two. The new level will be initialized by the lower level, and the virtual merges during the nth passes will occur only between objects with common super-objects. The final level, will obey both super-level and sub-level boundaries and will be compatible with both (Figure 2.4c).

The Multi-scale Algorithm module uses these three approaches to create potential hierarchy trees according to user specification. The implementation of this module had as a result a better speed of the MSEG algorithm, as it constrained the number virtual merges (due to super-objects), and reduced passes (due to initialization from sub-levels).

Finally, an innovation of the MSEG algorithm was that multiple hierarchies have been

supported since there was no restriction to the combinations of levels. Thus, one level can also be direct sub-level or direct super-level to many levels at the same time, leading to multi branch hierarchies.



(a) (b) (c)

Figure 2.4: The segmentation levels of the Multi-scale Algorithm cases presented graphically. Colors represent object ID's (non duplicate integers inside levels). (a) The bottom-up approach. (b) The top-down approach. (c) The interpolation approach.

2.2.3.2 Global Heterogeneity Heuristics module

Up to now, the Simple Profile passes were based on the merging criterion of the mutual best match between neighboring objects. This criterion, sometimes is sub-optimal, due to computation cost (many virtual merges occur and few of them are producing real merges). This heterogeneity heuristic was found optimal at minimizing the scene heterogeneity after region merging procedures [Baatz and Schape, 2000]. But there are cases where the global heterogeneity can be sacrificed for speed gain (especially in images with less high frequencies, where there are large areas of the same spectral signature, and best match is not of high importance). For such cases, or for large image datasets applications where optimization in speed is necessary, additional global heterogeneity heuristics for merging objects inside a pass have been introduced to the algorithm.

For the above reasons, an accuracy-to-speed ratio module has been implemented, including the global heterogeneity heuristics. The accuracy refers to the global heterogeneity cost that is added to the image with each merge that occurs during segmentation. The three modes (heterogeneity heuristics) implemented in this module are:

Fastest mode. This mode uses a merging procedure that is based on the simple matching criterion. This means that for any object extracted from the priority list, not all virtual merges with its neighbors take place. Instead, virtual merges stop when the first match is found. Then, without any other search mechanism, the merge occurs. Even if

no virtual merges occur afterwards, it is important to notice that all neighbors must be placed in the priority list in order to ensure that all objects will be treated by the algorithm. This mode, although it is the fastest, has major drawbacks: it does not minimize the heterogeneity from the merge, and it provides unstable results. Since, no full virtual merging is taking place, the sequence of treatment for objects is now important. This makes the segmentation procedure unstable and error prone. Also, this method can not guarantee reproducibility if the order of treatment for neighbor objects is random.

Fast mode. This mode uses a merging procedure that is based on the best match criterion. This means that for any extracted object from the priority list, all virtual merges could take place, and the object is merged with the best match, without any other searching. Again, all neighbors are placed in the priority list. Thus, the random neighbor search will not give reproducibility problems. The starting point estimation is, however, the weak link of this heuristic, because its results are of great significance to the final segmentation result. On the positive site, heterogeneity is reduced compared to the fastest mode.

Standard mode. This mode is the default mutual best match procedure as it was described in the above sections.

2.3 Discussion of results

The implemented version of the MSEG algorithm was tested on a variety of image data, in order to assess the quality of the results, its ability to be generic and its speed. In the following, only remote sensing examples are shown from a variety of sensors.

Evaluating the results of a segmentation algorithm does not depend on the delivery of semantic objects, but rather on the generation of good object primitives useful to further classification steps. Thus, over-segmentation was not an issue of great importance. Indeed, the algorithm was designed to provide over-segmentation so that merging of segments, towards the final image semantics, to be achieved by a follow up classification procedure. Boundary distinction (primitive boundaries can not intersect two or more semantic object boundaries) and full-scene segmentation were of great significance. Since the eCognition software [Benz et al., 2004] is greatly used for object oriented image analysis purposes the evaluation of results was mainly based on comparison with outputs from eCognition.

The first image tested, was a Landsat TM satellite image of the town of Dessau, in Germany. In Figure 2.5a the whole dataset available for testing is shown. Both the MSEG and the eCognition segmentation algorithms were used with the same parameters wherever possible. In Figure 2.5b, is shown a first segmentation result, provided from the MSEG algorithm with scale parameter 400, color weight 0.7, shape weight 0.3, compactness weight 0.4 and smoothness weight 0.6. A satisfying result was obtained, since the edges of the thematic categories do not seem to intersect the primitive object boundaries.

For evaluation and visualization purposes, a subset of the original image was segmented both by eCognition and MSEG. Those results are presented in Figure 2.6. In Figure 2.6b, the segmentation of the image was provided by MSEG, with scale parameter value of 100. The weights as described by equations (2.6) and (2.7) had the following values: $w_{color} = 0.8$, $w_{shape} = 0.2$, $w_{comp} = 0.4$ and $w_{smooth} = 0.6$. Using a small value (100) for the scale parameter, did not allow for scene heterogeneity to grow, and as a result, merging of objects was restricted and the primitive object size was small. On the impervious regions of the image (blue-gray areas, where cultural objects are visible), the heterogeneity was large. This resulted in smaller than the average primitives for that area. The algorithm simply did not allow highly textured regions to be merged with such small value of scale parameter, thus resulting to primitive objects with really small amount of merged pixels. On vegetated areas, this restriction is of less importance, due to low contrast in spectral values, and this resulted to larger primitive objects. A small noise stripping within the vegetation areas (Figure 2.6a, red arrow), seems to have resulted to a kind of elongated polygons on the segmentation result (Figure 2.6b, red arrow), and that was caused by the small scale parameter used. It should be noticed that this kind of noise was also embedded to other spectral channels of Landsat TM imagery as well. Taking a closer look at the results of Figure 2.6b, it was observed that no edges have been violated by the object boundaries, hence a classification step is possible on those objects. A disadvantage of the present form of visualization is that dense vectorization of the boundaries limits the observing capacity of the results at this scale of segmentation. On the other hand, if no vectors were drawn, the distinction between the objects would not be possible for the reader.

In Figure 2.6c, a segmentation of the same image was produced using the eCognition software with scale parameter 10 - which corresponds to the image segmented by the MSEG with scale parameter 100 (Figure 2.6b), as the scale parameter of the eCogni-

tion algorithm is squared internally (eCognition User Guide 2005). The color and shape parameters were used with exactly the same values. The result of this test, shows, that the two segmentations are incompatible with the same scale parameters. eCognition produces larger objects, is less sensitive to spectral heterogeneity, but the boundaries of it's primitive objects seemed to be more well shaped (the reader should bear in mind that a better vectorization software is used in eCognition).

In order to reproduce a result compatible to eCognition, MSEG was given a higher value (400) for scale parameter (Figure 2.6c and 2.6d). For the new scale parameter of 400, the MSEG seems to be less sensitive to spectral heterogeneity, but still more sensitive than eCognition. Both algorithms keep good alignment with the image edges and both provide usable over-segmentations of the initial image. Still, despite of the better vectorization, eCognition has better shaped boundaries, which is a clue, that the shape weights of this algorithm are valued un-equally in the eCognition implementation, or that other unknown heuristics are involved. In both systems, MSEG and eCognition , the thematic category boundaries are well respected by the segmentations.

In the last step of evaluation, segmentations were produced from both algorithms using larger scale parameters: eCognition's result for the scale value of 20 is presented in Figure 2.6e while MSEG's result for scale parameter of 700 is presented in Figure 2.6f.

eCognition produced better shaped objects, but is less sensitivity to spectral homogeneity. MSEG gave equally good results around the edges of the river, or other bright areas. MSEG in general gave better results within the impervious areas of the image. The reader should also note the similarities among some image objects (red arrows in Figures 2.6e and 2.6f). The MSEG algorithm was tested on a large variety of remote sensing imagery, such as IKONOS, Quick-Bird, Spot, Lidar imagery as well as radar images and aerial photos and has produced successful results on all those datasets. Further down, aerial scan digital imagery is presented along with its segmentation results in Figure 2.7.

2.4 Conclusions and Future Developments

Overall, the proposed image segmentation algorithm MSEG, gave very promising segmentation results for remote sensing imagery. It has proved to be a generic segmentation solution for remote sensing platforms (Ikonos, Quick Bird, Landsat, Spot, Lidar, aerial scanner data, radar images and aerial photographs). The boundaries of the primitive ob-

(a) (b)

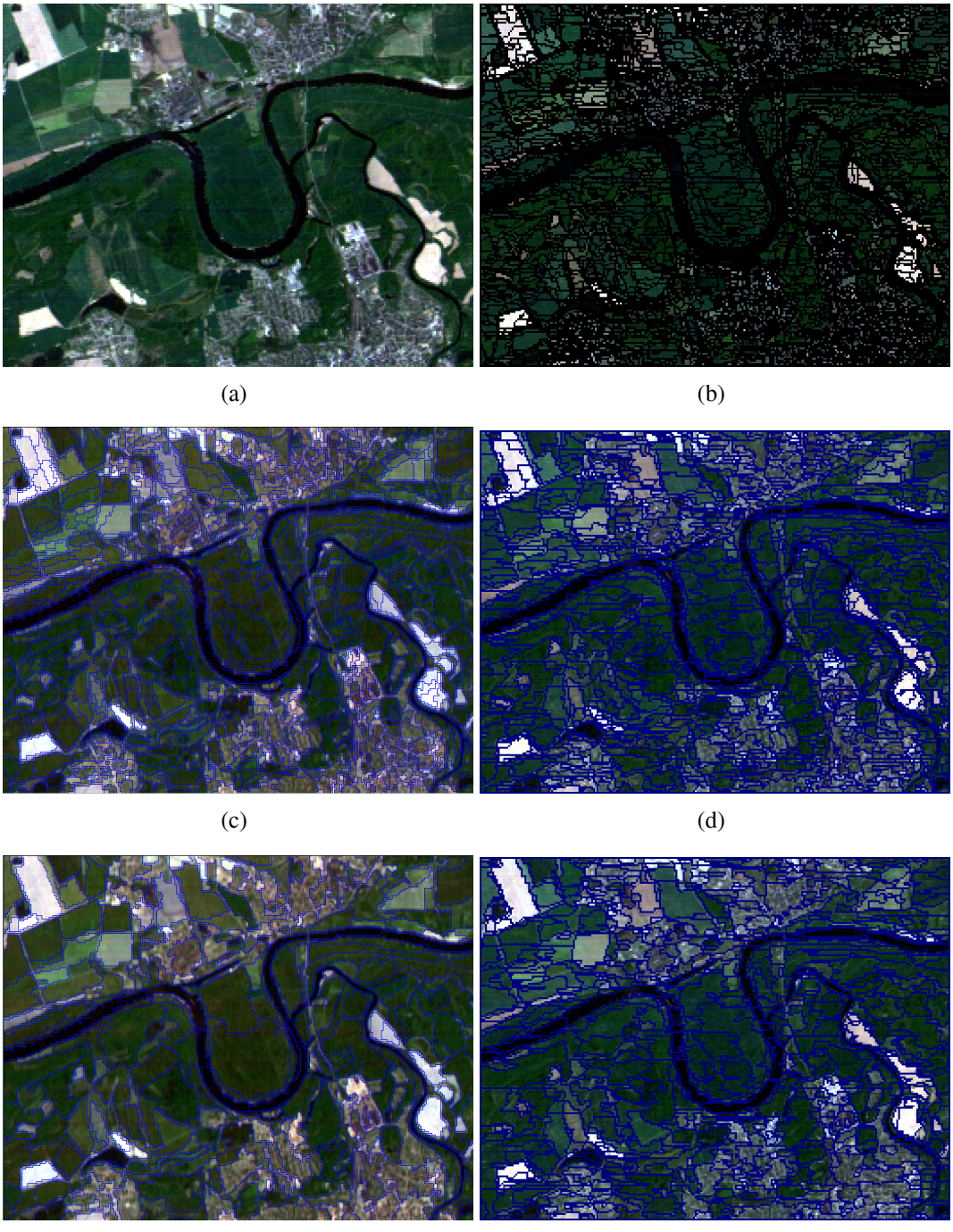
Figure 2.5: The original Dessau image (eCognition User Guide 2005)(a) and (b) A first segmentation result of MSEG algorithm (Simple Profile) using scale parameter 400.

jects that were extracted in each case were compatible with the semantic object edges. Thus, for object oriented image analysis, MSEG is qualified as a successful low level processing algorithm.

MSEG has however some disadvantages that have to be further investigated. Its shape heterogeneity criteria are not quite effective and must be further tested and optimized to provide better results. At the same time, MSEG gave worse results in comparison to those of the eCognition algorithm for large scale parameters. For low scale parameters, MSEG was more sensitive to spectral homogeneity and provided better results (less merging errors). It must be noted that all comparisons made, were using only the Simple Profile of MSEG, as the various Advanced Profiles are under development and testing.

Generally, the algorithm meets its expectations in computational speed and memory efficiency. The reference software is still not optimized for speed, but provides viable execution times. Speed can also be improved using the GHH module as described in the Advanced Profile section.

The initial requirements of the algorithm were accomplished since it can produce multi-scale results for all remote sensing data. Also, with the Multi-scale Algorithm module, several levels of segmentation can be used for classification efficiency. At the



(a) (b)

(c) (d)

(e) (f)

Figure 2.6: (a) Original Landsat TM dataset in RGB composite. (b) MSEG result with scale parameter 100, (c) eCognition's result with scale parameter 10, (d) MSEG result with scale parameter 400, (e) eCognition's result with scale parameter 20 and (f) MSEG result with scale parameter 700.

(a) (b)

Figure 2.7: (a) Original aerial scan data (Toposys 2005) (b) Segmentation results with MSEG algorithm for scale parameter of 400.

same time, the MSEG provided the innovative solution of multiple and branch-based hierarchies for level organization. The need for reproducibility of results was covered 100% but only for the case of same image and same parameter application.

Chapter 3

A hybrid texture-based and region-based multi-scale image segmentation algorithm

3.1 Introduction

3.1.1 Recent developments in Remote Sensing

Recently, remote sensing has achieved great progress both in sensor resolution and image analysis algorithms. Due to very high resolution imagery, such as IKONOS and Quick Bird, traditional classification methods, have become less effective given the magnitude of heterogeneity appearing in the spectral feature space of such imagery. The spectral heterogeneity of imaging data has increased rapidly, and the traditional methods tend to produce "salt and pepper" classification results. Such problems occur also to medium resolution satellite data, such as Landsat TM, SPOT etc.

Another disadvantage of traditional classification methods is that they do not use information related to shape, site and spatial relation (context) of the objects of the scene. Context information is a key element to photo-interpretation, and a key feature used by all photo-interpreters because it encapsulates expert knowledge about the image objects [Argialas and Harlow, 1990]. Such knowledge is not explicit and needs to be represented and used for image analysis purposes. Shape and texture features are used extensively as

context descriptors in photo-interpretation.

3.1.2 Texture-based Image Segmentation and Object-based Image Analysis

Texture initially was computed as standard deviation and variance. Haralick proposed texture features computed from co-occurrence matrices [Haralick, 1979; Haralick et al., 1973]. These second order texture features were used in image classification of remote sensing imagery with good results [Materka et al., 1998]. Recently, even more complex texture models were used for classification and segmentation, such as Hidden Markov Models, Wavelets and Gabor filters [Materka et al., 1998] with very good results in remote sensing and medical applications. Several methods were proposed for texture-based image segmentation, taking advantage of the latest texture modeling methods [Chen et al., 2002; Fauzi and Lewis, 2003; Havlicek and Tay, 2001; Liapis et al., 1998]. At the same time, image classification moved towards artificial intelligence methods [Benz et al., 2004; Sukissian et al., 1994].

During the last few years, a new approach, called Object-Oriented Image Analysis, integrated low level image analysis methods, such as segmentation procedures and algorithms [Baatz and Schape, 2000], with high level methods, such as Artificial Intelligence (fuzzy knowledge-based systems) and Pattern Recognition methods. Within this approach, the low level image analysis produces primitive image objects, while the high level processing classifies these primitives into meaningful domain objects [Benz et al., 2004].

In order to extract primitive objects from a digital image, a segmentation algorithm can be applied. Various segmentation algorithms and methods have been proposed over the last decades, with promising results [Pal and Pal, 1993; Sonka et al., 1998]. In remote sensing, a multi-scale image segmentation algorithm is aiming not to the extraction of semantic objects, but to the extraction of image primitives [Baatz and Schape, 2000].

3.1.3 Research Objectives

The main objective of this research was the integration of complex texture features into an object-oriented image segmentation algorithm [Tzotsos and Argialas, 2006] to be used

as a low level processing part of an object-oriented image analysis system so that to be applied at multiple image resolutions and to produce objects of multiple scales (sizes), according to user-customizable parameters.

Another objective was the ability of the produced algorithm to be generic and produce satisfying and classification-ready results to as many remote sensing data as possible. Remote sensing data with complex texture and spectral information are, in general, difficult to process. Therefore, there was a need for the algorithm to be able to handle texture information and context features in order to produce better segmentation results.

This research aimed to further develop recent technologies and provide integration of new features to object based image analysis methodology. Motivation for this research is to provide an Object-Based Image Analysis system in the form of Free and Open-Source Software.

3.2 Methodology

3.2.1 MSEG algorithm overview

The MSEG algorithm [Tzotsos and Argialas, 2006] was designed to be a region merging technique, since region merging techniques are fast, generic and can be fully automated (without the need of seed points) [Pal and Pal, 1993; Sonka et al., 1998]. Given that commercial Object-Oriented Image Analysis system [Benz et al., 2004] has used such methods was also a strong argument for the effectiveness of the region merging techniques.

Like all algorithms of this kind, MSEG is based on several local or global criteria and heuristics, in order to merge objects in an iterative procedure, until no other merges can occur [Sonka et al., 1998]. In most cases, a feature of some kind (mean spectral values, texture, entropy, mean square errors, shape indices etc.) or combination of such features computes the overall "energy" of each object. Then, the merging algorithm uses heuristics and "trial and error" methods in order to minimize the overall energy of the segmentation that is produced. In other words, it is typical to select a cost function, to define how good and stable an object is after a merging procedure, or even to make the decision regarding that merge.

Various definitions of homogeneity (energy minimization measures within an object)

have been defined [Pal and Pal, 1993; Sonka et al., 1998]. Recently, a very successful segmentation algorithm, embedded in the Object Oriented Image Analysis Software eCognition [Baatz and Schape, 2000], implemented such measures of spectral and spatial homogeneity, for making the merging decision between neighboring objects, with very good results. In the proposed segmentation algorithm, similar homogeneity measures were used, and then complex texture features were implemented in later stages.

The proposed algorithm was initialized through the application of an image partitioning method to the dataset resulting into rectangular regions of variable dimensions, called macroblocks. Image partitioning was applied for computation of local statistics and starting points. It should be pointed that starting points were not used as seed points (as in region growing techniques) but were used to keep track of the order in which all pixels were processed initially. Having this order computed and stored for each macroblock, the whole segmentation algorithm can be reproduced accurately and provide 100% identical results for the same parameters and image. There are two basic methods implemented within MSEG for starting point estimation [Tzotsos and Argialas, 2006]. The first is a statistical method, producing one starting point per macroblock. The second method was based on dithering algorithms, transforming the original image into a binary image through statistical procedures [Ulichney, 1987]. The binary image was used to determine the starting points within each macroblock.

In order for the MSEG algorithm to provide primitive objects, several steps of region merging (passes) were followed. The initial objects of the image are the single pixels. The purpose of a first segmentation pass was to initialize the image objects and to provide the first over-segmentation, in order for the algorithm to be able to begin region merging at following stages. During first pass, the algorithm merged single pixels-objects pair wise. The criterion for object merging was a homogeneity cost measure, defined as object heterogeneity, and computed based on spectral and shape features for each possible object merge. The heterogeneity was then compared to a user defined threshold, called scale parameter, in order for the decision of the merge to be determined.

For the following pass of the algorithm, the objects created by the previous pass were used in a new pair wise merging procedure. The merging strategy included finding the best match for each object, and then checking if there was a mutual best match in order to merge the two objects [Tzotsos and Argialas, 2006]. Passes were executed iteratively until the algorithm converged. The algorithm was considered finished, when during the

last pass no more merges occurred. Then, the objects were exported and marked as final primitives (Figure 3.1).

In order to extend the basic elements of the region merging segmentation procedure, a multi-scale algorithm was designed to give to the MSEG algorithm the capability to create multiple instances of segmentations for an image, each with different scale parameters. The problem when dealing with multiple segmentations is the compatibility between scales, in order to combine information and objects. One simple way to deal with this problem is to create a multi-level representation, and incorporate the multiple segmentations within this representation, hierarchically. A single-level hierarchy is not flexible, when dealing with remote sensing classification problems [Argialas and Tzotsos, 2004]. A multi-level hierarchy approach or a branch-based hierarchy model can represent more complex spatial relations. Thus, in the present multi-scale algorithm, every new level depends only from the nearest (scale-wise) super-level or the nearest sub-level, or both. More details on MSEG implementation can be found in [Tzotsos and Argialas, 2006].

3.2.2 Advanced Texture Heuristics

The basic objective of the Advanced Texture Heuristic module was to build upon MSEG algorithm, in order to improve segmentation results. Since texture is a key photo-interpretation element, it was decided to use more complex texture features, than standard deviation (used in eCognition), variance, or other first order texture features. Since second order texture features have been used as good and practical classification features [Haralick et al., 1973; Materka et al., 1998], there was a need to test those measures for segmentation purposes and specifically as an add-on to the region merging multi-scale algorithm that was previously developed. The basic idea was that when making a merging decision between adjacent image objects, there should be a texture similarity measure, provided by complex texture computations that could help this decision. This way, primitive objects with similar texture could be merged, even if color or shape criteria are not in favor of this merge.

Given that MSEG is a region merging algorithm, not all state of the art methods for modeling texture are compatible for a hybrid segmentation solution. The recent literature has shown that Markov Random Fields, wavelets and Gabor filters, have great potential for

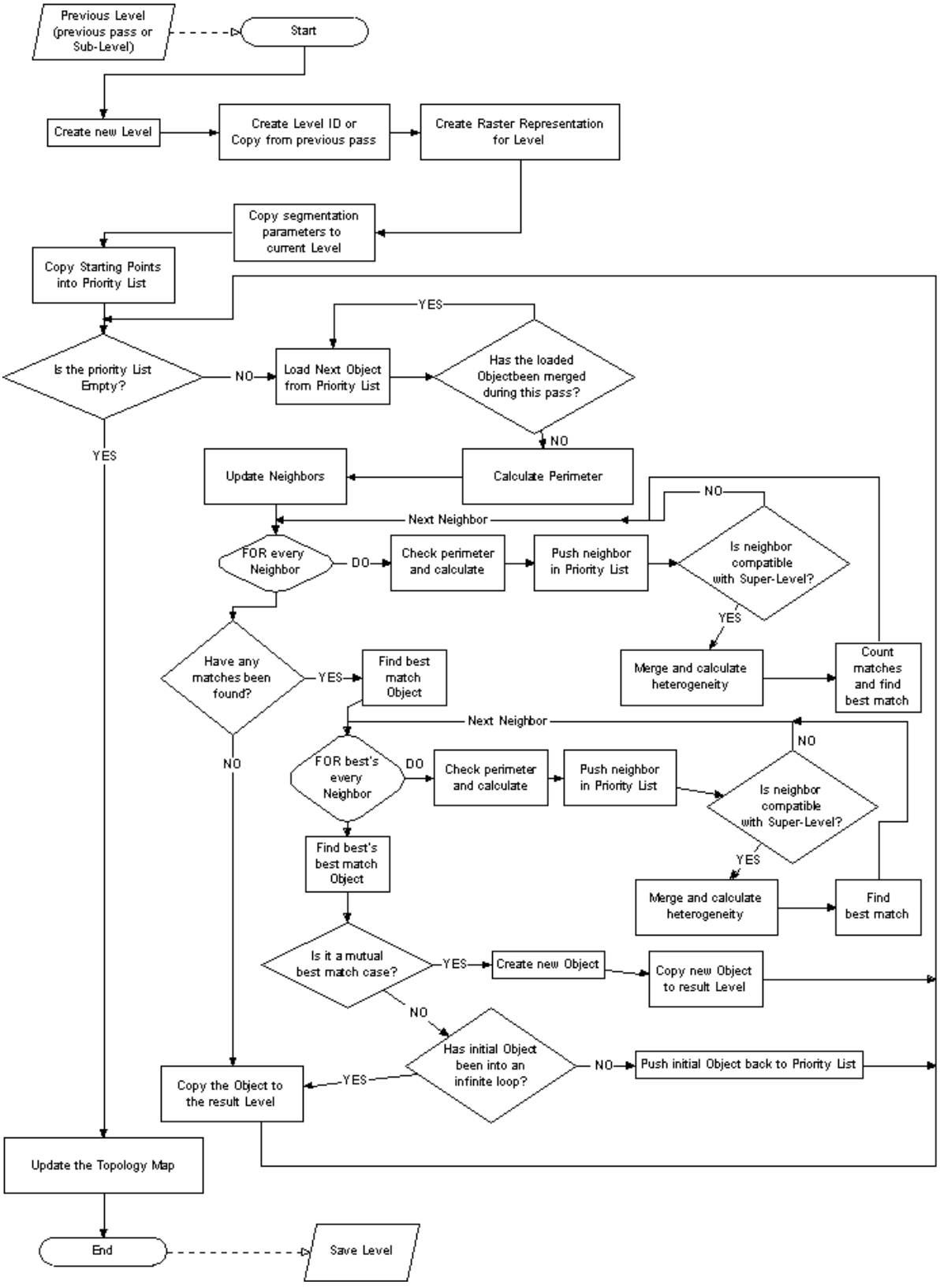


Figure 3.1: Nth pass general flowchart

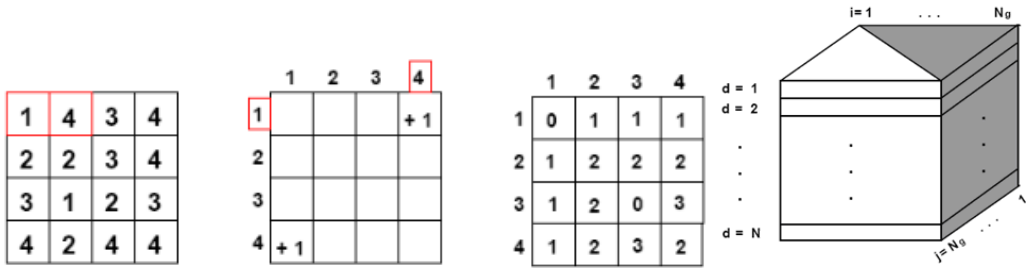
texture analysis [Materka et al., 1998]. Their disadvantage is that they are very complex and time consuming to use with a merging procedure, computing thousands of virtual merges during a full object merging search. At the same time, wavelets and Gabor filters are computationally inefficient to be used locally, within the boundaries of a single – and sometimes very small - primitive object. Markov Random Fields are easier to adopt for region-based texture segmentation, but they were found incompatible with the current merging search method, since they are based on Bayesian reasoning.

A traditional method for modeling texture, which was proved to be very good for practical purposes in supervised classification [Haralick et al., 1973; Schröder and Dimai, 1998], is based on the Grey Level Co-occurrence Matrix (GLCM) features. GLCM is a two dimensional histogram of grey levels for a pair of pixels that are separated by a fix spatial relationship. The Grey Level Co-occurrence Matrix approximates the joint probability distribution of this pair of pixels. This is an insufficient approximation for small windows and a large number of grey levels. Therefore the image data have to be pre-scaled to reduce the number of grey levels in the image. Directional invariance can be obtained by summing over pairs of pixels with different orientations [Schröder and Dimai, 1998].

From the GLCM, several texture measures can be obtained, such as homogeneity, entropy, angular second moment, variance, contrast etc [Haralick et al., 1973]. To compute the GLCM, several optimization methods have been introduced. Most applications of GLCM for remote sensing images, at pixel-level, included computation of the co-occurrence matrix less often for the whole image, and more often for a predefined image sliding window of fixed size.

On a pixel-based texture analysis with the use of GLCM, for each direction (0, 45, 90, 135 degrees) a different co-occurrence matrix is formed by adding co-occurrences to the grey level pair position (Figure 3.2). If Ng is the number of grey levels after the grey level reduction, then each co-occurrence matrix will be of size Ng x Ng.

In order to use the second order texture features into MSEG, four GLCMs should be computed for each primitive image object, during the merging procedure. The computation of so many GLCMs can be extremely intensive, and would significantly slow down the performance of the region merging algorithm. Furthermore, as multiple virtual merges occur, before an object merge could be decided, the GLCM computations can actually be much more than theoretically expected. Thus, a decision was made to



(a) (b)

Figure 3.2: Left: An example of GLCM computation at 0 degree angle for a 4x4 window. The empty GLCM gets filled by adding co-occurrences symmetrically. Right: A 3-dimensional representation of the Co-occurrence matrices that have to be computed for a given orientation. Ng is the number of grey levels and N is the total number of primitive image objects

optimize the use of GLCM features only for the initial objects and not for the virtual merged objects, so that to limit the co-occurrence matrix computation to a maximum of 4N (where N is the number of objects).

The Advanced Texture Heuristic module aimed to implement texture similarity measures in order to contribute to object merging search. Haralick states as good texture similarity measures, the Homogeneity (Equation 3.1) and the Angular Second Moment (Equation 3.2) features [Haralick, 1979]. Both were implemented in the module.

$$homogeneity = \sum_{i,j=0}^{N_g-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{3.1}$$

$$ASM = \sum_{i,j=0}^{N_g-1} P_{i,j}^2 \tag{3.2}$$

When an object was treated from the MSEG algorithm during a pass, the texture features were computed and the mutual best match search procedure compared neighbor objects to the selected one. Before the color and shape heterogeneity criteria were computed and involved to the scale parameter comparison, texture heterogeneity was computed, as the difference of the values of the texture similarity features. These val-

ues, one for each direction, were then compared with a threshold called texture parameter which is defined by the user. If the two objects were compatible by the texture parameter, then the computation of the spectral and shape heterogeneity took place, in order to fulfill the mutual best match criterion, and the merge to occur.

The described heuristic, uses the texture parameter, to reduce the heterogeneity computations. This means that, when activated, the Advanced Texture Heuristic module has greater priority than the scale parameter, but cannot perform any merging, without color and shape compatibility of image objects. If one wishes to perform segmentation using only texture features, the scale parameter can be set to a very large value, so not to constrain the merging by the color and shape criteria. In the following section, an optimization procedure for the GLCM computation is described.

3.2.3 Implementation

Having to compute thousands of co-occurrence matrices, during a region merging segmentation procedure can be computationally intense. Optimization algorithms for the computation of GLCMs have been proposed in literature [Argenti et al., 1990] but only for the pixel-based case. In order to tackle this problem, the GLCM computation should be optimized to be used with objects, rather than pixels. A modification to the traditional methods was performed, so that to make the procedure faster but still accurate. At first, image band selection took place. If the computation of the GLCM was to be performed for each band separately, the whole segmentation process would not be optimal for performance. So, instead of using all bands, the Advanced Texture Heuristic module can use the intensity band of the HSI colorspace, or the Y band of the YCbCr colorspace (used as default), or a principal component band of the image, or finally a single image band. After the band selection, a grey level reduction was performed at the selected band. The final number of grey levels can be selected by the user, with a quantizer parameter. The default value, as used by many other GLCM implementations, was set to 32 grey levels.

It was determined that the optimal procedure to compute the GLCMs was to perform some kind of global initialization, so that to speed up the inter-object GLCM computation. For each of the image pixels, a direction search was performed, to evaluate the grey level pair co-occurrences. For the 4 different directions, a vector was designed to hold the overall co-occurrence information and was used in a way similar to a database index.

Thus, no direction search was performed twice during the pass stages. Each time an object co-occurrence matrix had to be used, it was computed very fast within the object boundaries.

This procedure was not tested for algorithmic complexity, but was compared to a simple GLCM implementation and was found more stable and faster. The implementation of the Advanced Texture Heuristic module was performed in C++. The modified version of the algorithm was called Texture-based MSEG.

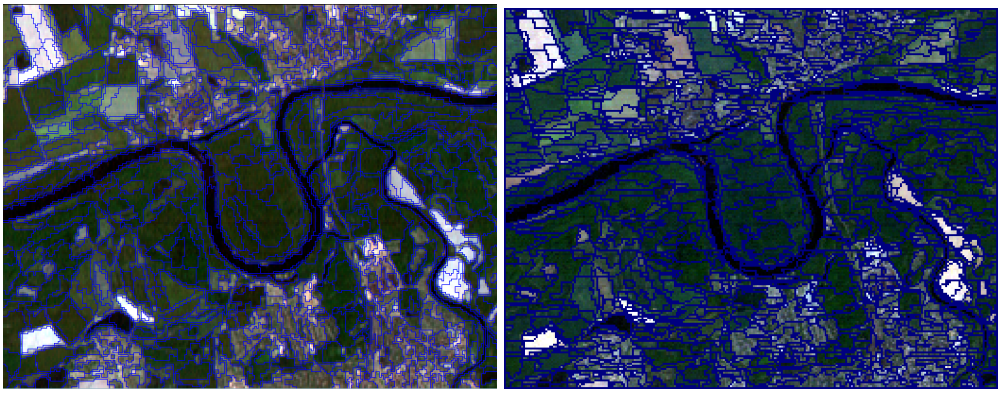
3.3 Discussion of Results

The implemented version of the MSEG algorithm was tested on a variety of image data, in order to assess the quality of the results, its generalization and speed. Evaluating the results of a segmentation algorithm does not depend on the delivery of semantic objects, but rather on the generation of good object primitives useful to further classification steps.

The algorithm was designed to provide (a) over-segmentation so that merging of segments, towards the final image semantics, to be achieved by a follow up classification procedure [Tzotsos and Argialas, 2008] and (b) boundary distinction and full-scene segmentation. Since eCognition [Benz et al., 2004] is greatly used for object oriented image analysis purposes, the evaluation of results was mainly based on comparison with outputs from eCognition. Also, a comparison was made to results of simple MSEG, to show how the texture features perform with region merging segmentation.

For the evaluation of the algorithms a Landsat TM image was used. For all tests, the color criterion was used with a weight of 0.7 and the shape criterion with weight 0.3.The eCognition software was used to provide segmentation with scale parameter 10. Then, the simple MSEG was used to provide segmentation with scale parameter 400 (through trial and error) to simulate the mean object size of eCognition's results. It should be noted that scale parameters are implementation dependent. The results are shown in Figure 3.3. In Figure 3.4 the results from the texture-based MSEG for scale parameter 400 are shown.

Comparing results between Figures 3.3 and 3.4, shows that similar sized object can be obtained by all 3 segmentation algorithms. For the scale parameter of 400, the MSEG seems to be more sensitive to spectral heterogeneity than the eCognition results with scale parameter 10. Both algorithms keep good alignment with the image edges and both provide usable over-segmentations of the initial image [Tzotsos and Argialas, 2008]. The

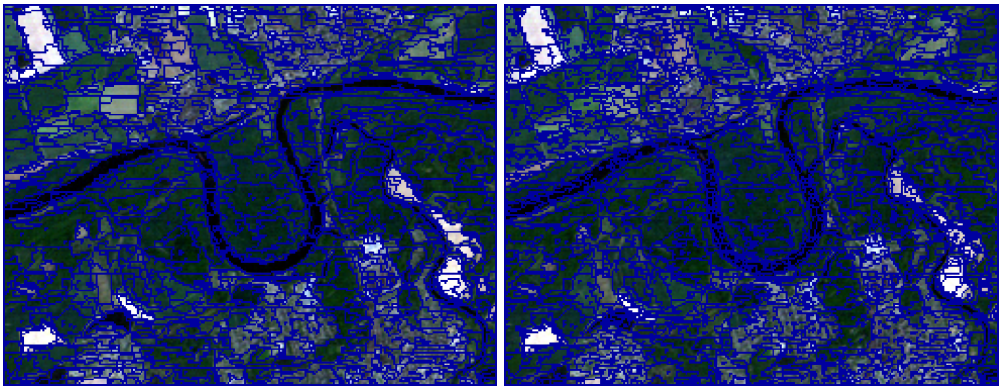


(a) (b)

Figure 3.3: Segmentation result as provided by eCognition for scale parameter 10 (left) and MSEG for scale parameter 400 (right)

texture-based MSEG also provides good segmentation of the image, improving the simple MSEG result by creating texturally homogenous regions, but at the same time, working against the shape criterion, providing less compact or smooth boundaries for objects.

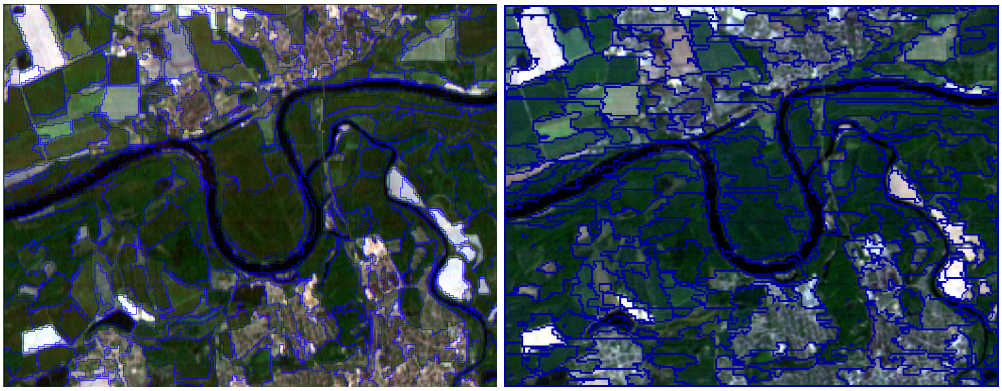
Further to its better vectorization, eCognition has better shaped boundaries, which is a clue that the shape weights of this algorithm are valued un-equally or that additional embedded heuristics are involved. In both systems, MSEG (with or without texture) and eCognition, the thematic category boundaries are well respected by the segmentations.



(a) (b)

Figure 3.4: Segmentation result as provided by texture-based MSEG for scale parameter 400 and texture parameter 2.0 (left) and for scale parameter 400 and texture parameter 1.0 (right)

In a further step of evaluation (Figure 3.5), the result of eCognition for the scale value of 20 is comparable to the result provided by the texture-based MSEG when a very large (2500) scale parameter was used (so that the scale parameter would not significantly interfere with the final mean object size) and the texture parameter was set to 3.0. The texture-based MSEG result is very good especially inside the urban areas, where there are complex texture patterns. There, it merged the object primitives in such a way, so that to provide larger homogenous objects in comparison to eCognition or the simple MSEG.



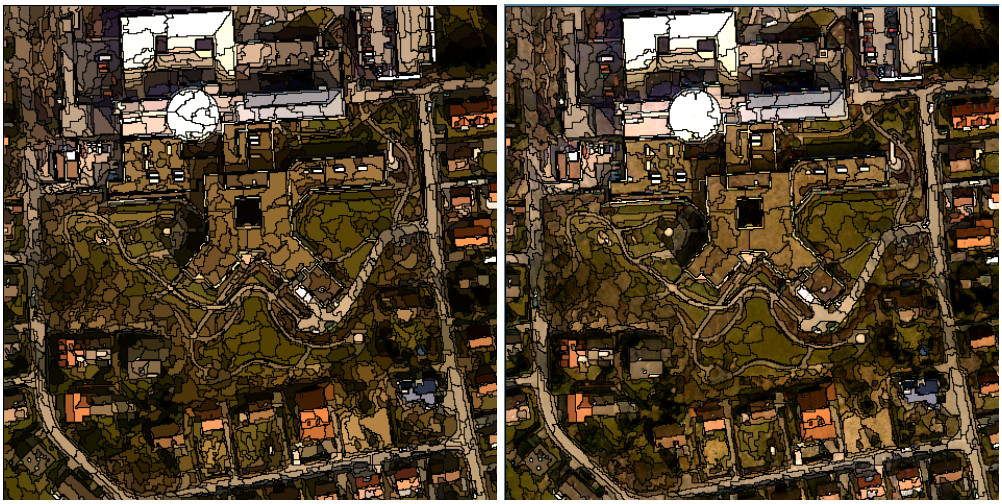
(a) (b)

Figure 3.5: Segmentation result by eCognition for scale parameter 20 (left) and texture-based MSEG for scale parameter 2500 and texture parameter 3.0 (right)

Also, in Figure 3.4, the difference of segmentation results is shown, when only the texture parameter is changed. Smaller texture parameter provides smaller primitive objects. It should be noticed that texture free objects, like the very bright white areas in the image, don't get affected by the texture parameter change, as expected. A final step of evaluation included testing the segmentation algorithms with a very high resolution remotely sensed image. A digital multispectral image from an aerial scanner (Toposys GmbH) with resolution of 0.5m was used. This image was selected because outperforms in resolution all commercial satellite data available today.

In Figure 3.6, the results of the eCognition algorithm with scale parameters 15 and 25 are presented. These results are very good, especially across the road outlines. The areas of interest in this test were those with complex texture, e.g. the mixed grasslands. The results of eCognition for these areas are the most over-segmented.

In Figure 3.7, the simple MSEG results are shown for scale parameters 400 and 700.

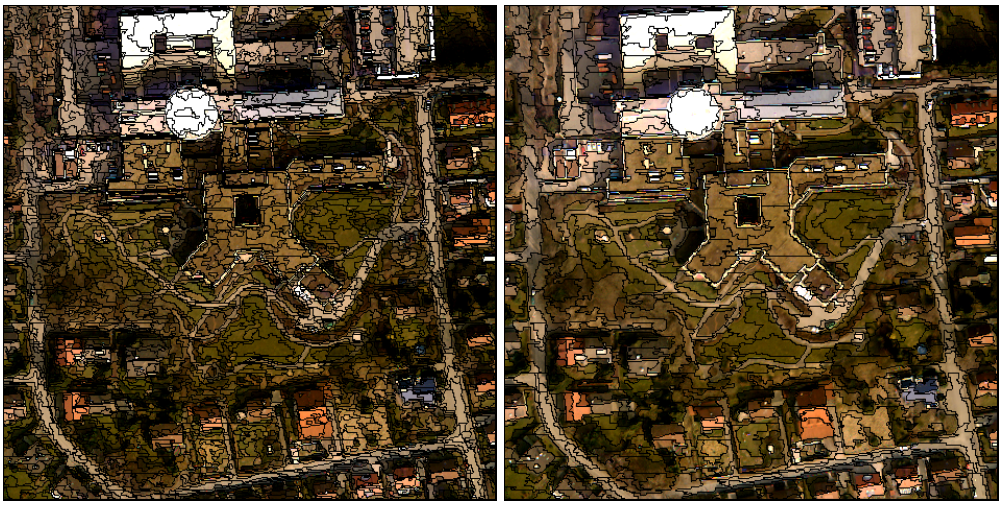


(a) (b)

Figure 3.6: Segmentation result by eCognition for scale parameters 15 (left) and 25 (right).

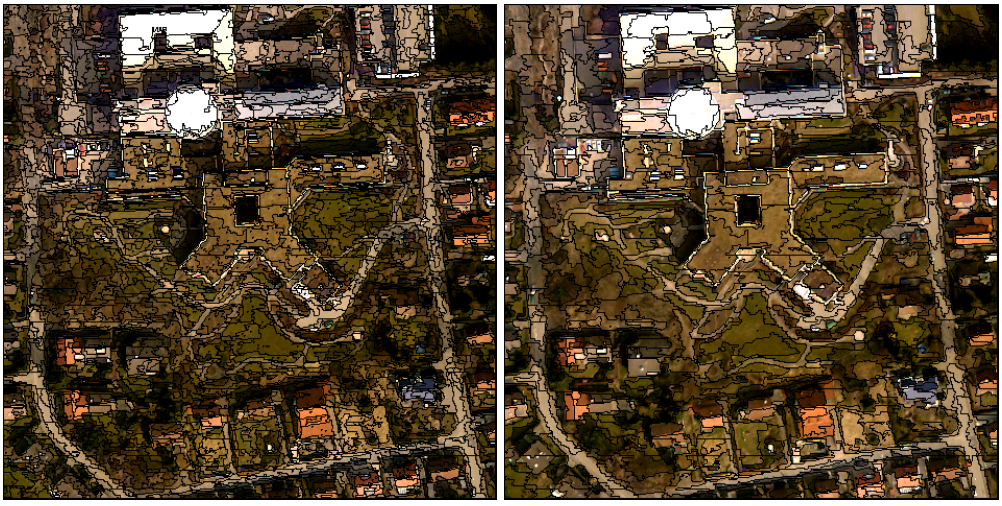
As before, the results are good especially across the building and road edges. Scale parameter 400 provides very over-segmented result and seems to be the most sensitive to color heterogeneity. For scale parameter 700 the result is better, especially in the complex texture areas, but eCognition has better results on road edges. It can be concluded after extensive testing of both algorithms, that the color and shape criteria are not optimized in the same way, thus many differences occur in the shapes of primitive objects.

In Figure 3.8, the Texture-based MSEG results are shown for scale parameters of 400 and 2500. For the scale parameter 400 test, a strict texture parameter of 1.0 was used, so that to demonstrate the capability of the algorithm to locate even the smallest differences in texture. This can be observed in the left-center area of the image, where there is a mixture of bare land with grassland. Thus a small texture parameter provides significant over-segmentation of the region. For the scale parameter 2500 test, this high value was used in order to provide freedom to the algorithm to reach to a result that was based only to texture parameter (3.0). The results were better than those of eCognition and of the simple MSEG in the complex textured areas. Again, the eCognition segmentation algorithm outperforms MSEG on the road edges, but results are very close.



(a) (b)

Figure 3.7: Segmentation result by simple MSEG for scale parameters 400 (left) and 700 (right).



(a) (b)

Figure 3.8: Segmentation result by Texture-based MSEG for scale parameters 400 (left) and 2500 (right).

3.4 Conclusions and future work

Overall, the designed image segmentation algorithm, gave very promising segmentation results for remote sensing imagery. With the addition of the Advanced Texture Heuristic module, it was shown to be a good and generic segmentation solution for remote sensing imagery. The boundaries of the primitive objects extracted in each case were compatible with those of the semantic objects. Thus, for object oriented image analysis, the texture-based MSEG is qualified as a successful low level processing algorithm.

MSEG has however some disadvantages that have to be further investigated. Its shape heterogeneity criteria are not quite effective and must be further tested and optimized to provide better results.

Future developments for the MSEG include integration of the algorithm with higher level artificial intelligence and pattern recognition methods for classification. Part of this integration with Support Vector Machines is presented in [Tzotsos and Argialas, 2008].

Chapter 4

Support Vector Machine Classification for Object-Based Image Analysis

4.1 Introduction

4.1.1 Knowledge-based image classification and Object Oriented Image Analysis

In recent years, research has progressed in computer vision methods applied to remotely sensed images such as segmentation, object oriented and knowledge-based methods for classification of high-resolution imagery [Argialas and Harlow, 1990; Kanellopoulos et al., 1999]. In Computer Vision, image analysis is considered in three levels: low, medium and high. Such approaches were usually implemented in separate software environments since low and medium level algorithms are procedural in nature, while high level is inferential and thus for the first procedural languages are best suitable while for the second an expert system environment is more appropriate.

New approaches were proposed, during recent years in the field of Remote Sensing. Some of them were based on knowledge-based techniques in order to take advantage of the expert knowledge derived from human photo-interpreters [Argialas and Goudoula, 2003; Yoo et al., 2005]. Especially within an Expert System environment, the classification step can be implemented through logic rules and heuristics, operating on classes and features, which are implemented by the user through an object-oriented representation

[De Moraes, 2004; Moller-Jensen, 1997].

Very recently a new methodology called Object Oriented Image Analysis was introduced, integrating low-level, knowledge-free segmentation with high-level, knowledge-based fuzzy classification methods. This new methodology was implemented through commercial software, eCognition, which included an object-oriented environment for the classification of satellite imagery [Baatz and Schape, 2000; Benz et al., 2004].

4.1.2 Computational Intelligence methods in Remote Sensing

Other fields of Artificial Intelligence have been developed and applied recently. Computers became more capable of performing calculations and thus, a new field of A.I. called Computational Intelligence and Machine Learning evolved. In this field, techniques like Neural Networks, Fuzzy Systems, Genetic Algorithms, Intelligent Agents and Support Vector Machines [Negnevitsky, 2005] are included. Machine learning is an integral part of Pattern Recognition, and its applications such as classification [Theodoridis and Koutroumbas, 2003]. Digital remote sensing used in the past pattern recognition techniques for pixel classification purposes, while recently modern machine learning techniques have been implemented to achieve superior classification results [Brown et al., 2000; Fang and Liang, 2003; Foody and Mathur, 2004; Huang et al., 2002a; Theodoridis and Koutroumbas, 2003].

The Support Vector Machine (SVM) is a theoretically superior machine learning methodology with great results in classification of high-dimensional datasets and has been found competitive with the best machine learning algorithms [Foody and Mathur, 2004; Huang et al., 2002a]. In the past, SVMs were tested and evaluated only as pixel based image classifiers with very good results [Brown et al., 2000; Foody and Mathur, 2004; Huang et al., 2002a; Melgani and Bruzzone, 2004].

Furthermore, for remote sensing data, it has been shown that Support Vector Machines have great potential, especially for hyperspectral data, due to their high-dimensionality [Melgani and Bruzzone, 2004; Mercier and Lennon, 2003]. In recent studies, Support Vector Machines were compared to other classification methods, such as Neural Networks, Nearest Neighbor, Maximum Likelihood and Decision Tree classifiers for remote sensing imagery and have surpassed all of them in robustness and accuracy [Foody and Mathur, 2004; Huang et al., 2002a].

4.1.3 Research objectives

The objective of this study was to evaluate the effectiveness and prospects of SVMs for object-based image analysis, as a modern computational intelligence method. A secondary objective was to evaluate the accuracy of SVMs compared to a simpler and widely used classification technique in object-based image analysis such as the Nearest Neighbor. Also, the computational efficiency and training size requirements of SVMs were addressed.

4.2 Methodology

4.2.1 Multi-scale segmentation

Image segmentation is an integral part of Object-Based Image Analysis methodology [Benz et al., 2004]. The digital image is no longer considered as a grid of pixels, but as a group of primitive and homogeneous regions, called image objects. The object oriented representation provides the classification process with context and shape information that could not be derived from single pixels. These are very important factors to photo-interpretation and image understanding [Biederman, 1985; Lillesand and Kiefer, 1987; Sonka et al., 1998]. Objects can be more intelligent than pixels, in a sense of knowing their "neighbours" and the spatial or spectral relations with and among them [Baatz and Schape, 2000].

In order to perform object based classification, a segmentation algorithm is needed to provide knowledge-free primitive image objects. For this research effort the MSEG multi-scale segmentation algorithm was used [Tzotsos and Argialas, 2006; Tzotsos et al., 2008]. The main reason for this choice was that it has an open architecture to implement new features in C++. For evaluation purposes, the Multiresolution Segmentation algorithm from eCognition was also used [Baatz and Schape, 2000].

4.2.2 Support Vector Machines

Recently, particular attention has been dedicated to Support Vector Machines as a classification method. The SVM approach seeks to find the optimal separating hyperplane between classes by focusing on the training cases that are placed at the edge of the class

descriptors (Fig.4.1). These training cases are called support vectors. Training cases other than support vectors are discarded. This way, not only an optimal hyperplane is fitted, but also less training samples are effectively used; thus high classification accuracy is achieved with small training sets [Mercier and Lennon, 2003]. This feature is very advantageous, especially for remote sensing datasets and more specifically for Object-based Image Analysis, where object samples tend to be less in number than in pixel based approaches.

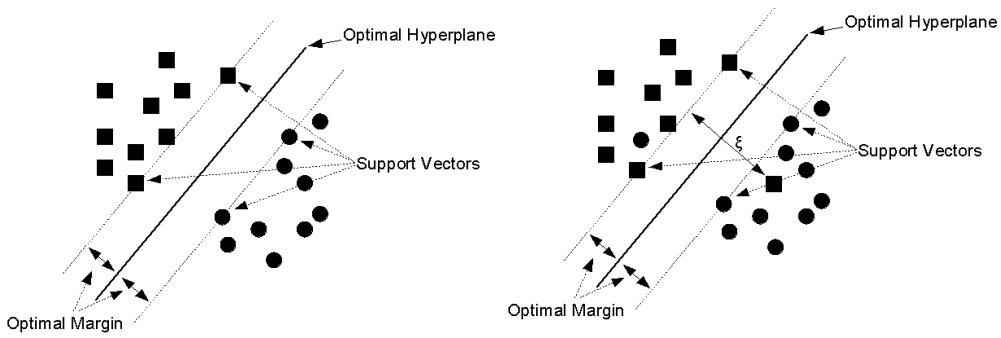


Figure 4.1: Left: The case of linear separable classes. Right: The case of non linear separable classes. ξ measures the error of the hyperplane fitting.

A complete formulation of Support Vector Machines can be found in a number of publications [Cortes and Vapnik, 1995; Theodoridis and Koutroumbas, 2003; Vapnik, 1995, 1998]. Here, the basic principles will be presented and then their implementation and application to Object Based Image Analysis will be evaluated.

Let us consider a supervised binary classification problem. If the training data are represented by $\{xi, yi\}, i = 1, 2, \ldots, N$, and $yi \in \{-1, +1\}$, where N is the number of training samples, $yi = +1$ for class ω_1 and $y_i = -1$ for class ω_2. Suppose the two classes are linearly separable. This means that it is possible to find at least one hyperplane defined by a vector w with a bias w0, which can separate the classes without error:

$$f(x) = w \cdot x + w_0 = 0 \tag{4.1}$$

To find such a hyperplane, w and w_0 should be estimated in a way that $y_i(w \cdot x_i + w_0) \geq +1$ for $y_i = +1$ (class ω_1) and $y_i(w \cdot x_i + w_0) \leq -1$ for $y_i = -1$ (class ω_2). These two, can

be combined to provide equation:

$$y_i(w \cdot x_i + w_0) - 1 \geq 0 \tag{4.2}$$

Many hyperplanes could be fitted to separate the two classes but there is only one optimal hyperplane that is expected to generalize better than other hyperplanes (Fig.4.1). The goal is to search for the hyperplane that leaves the maximum margin between classes. To be able to find the optimal hyperplane, the support vectors must be defined. The support vectors lie on two hyperplanes which are parallel to the optimal and are given by:

$$w \cdot x_i + w_0 \pm 1 \tag{4.3}$$

If a simple rescale of the hyperplane parameters w and w_0 takes place, the margin can be expressed as $\frac{2}{\|w\|}$. The optimal hyperplane can be found by solving the following optimization problem:

$$Minimize \quad \frac{1}{2} \parallel w \parallel^2 \tag{4.4}$$

Subject to $y_i(w \cdot x_i + w_0) - 1 \geq 0, i = 0, 1, ..., N$ Using a Lagrangian formulation, the optimal hyperplane discriminant function becomes:

$$f(x) = \sum_{i \in S} \lambda_i y_i (x_i x) + w_0 \tag{4.5}$$

where λ_i are the Lagrange multipliers and S is a subset of training samples that correspond to non-zero Lagrange multipliers. These training samples are called support vectors.

In most cases, classes are not linearly separable, and the constrain of equation 4.2 cannot be satisfied. In order to handle such cases, a cost function can be formulated to combine maximization of margin and minimization of error criteria, using a set of variables called slack variables ξ (Fig. 4.1). To generalize the above method to non-linear discriminant functions, the Support Vector Machine maps the input vector x into a high-dimensional feature space and then constructs the optimal separating hyperplane in that space. One would consider that mapping into a high dimensional feature space would add extra complexity to the problem. But, according to the Mercer's theorem [Theodoridis and Koutroumbas, 2003; Vapnik, 1998], the inner product of the vectors in the mapping

space, can be expressed as a function of the inner products of the corresponding vectors in the original space.

The inner product operation has an equivalent representation:

$$\Phi(x)\Phi(z) = K(x, z) \tag{4.6}$$

where K(x,z) is called a kernel function. If a kernel function K can be found, this function can be used for training without knowing the explicit form of Φ.

4.2.3 SVM Multi-class Classification

The SVM method was designed to be applied only for two class problems. For applying SVM to multi-class classifications, two main approaches have been suggested. The basic idea is to reduce the multi-class to a set of binary problems so that the SVM approach can be used.

The first approach is called "one against all". In this approach, a set of binary classifiers is trained to be able to separate each class from all others. Then each data object is classified to the class for which the largest decision value was determined [Hsu and Lin, 2002]. This method trains N SVMs (where N is the number of classes) and there are N decision functions. Although it is a fast method, it suffers from errors caused by marginally imbalanced training sets.

The second approach is called "one against one". In this, a series of classifiers is applied to each pair of classes, with the most commonly computed class kept for each object. Then a max-win operator is used to determine to which class the object will be finally assigned. The application of this method requires N(N-1)/2 machines to be applied. Even if this method is more computationally demanding than the "one against all" method, it has been shown that can be more suitable for multi-class classification problems [Hsu and Lin, 2002], thus it was selected for SVM object-based image classification.

4.2.4 Implementation

In order to apply the SVM methodology for Object-Based Image Analysis, it was necessary to perform a segmentation task. The MSEG algorithm was selected to perform segmentation at multiple scales [Tzotsos and Argialas, 2006; Tzotsos et al., 2008] and to

produce primitive image objects to be used for SVM classification.

For the primitive objects to be usable by a classification algorithm there was a need to implement an interface between image objects and the classifier. An extra module was implemented into the MSEG core library to add the functionality of selecting sample objects. Because a comparison was to be made with the Nearest Neighbor classifier used in eCognition, a TTA Mask [Benz et al., 2004] import module was also implemented, so that the training object selection process would be as transparent and objective as possible. For the object feature representation, XML was selected, so that open standards are followed.

A widely used SVM library called LIBSVM [Chang and Lin, 2011] was then modified to be able to handle XML files as well as training samples from the MSEG algorithm. A classifier module was then implemented as a modified version of LIBSVM.

The proposed Object-based Image Analysis system worked as following. A segmentation procedure was carried out with parameters like scale, color and shape. The properties of the primitive objects were then computed and exported to XML. A TTA Mask file along with its attribute table was imported to the system and training object samples were defined. A training set of feature vectors was exported from the MSEG algorithm and was used for training the SVM module.

The SVM module is capable to use 4 types of kernels for training and classification: Linear, Polynomial, Radial Basis Function and Sigmoid. All the above kernels follow Mercer's theorem and can be used for mapping the feature space into a higher dimensional space to find an optimal separating hyperplane. In literature, there have been many comparison studies between the most common kernels [Huang et al., 2002a; Mercier and Lennon, 2003]. For pixel-based classification of remotely sensed data, it has been shown that local kernels such as RBF can be very effective and accurate. Also, the linear kernel is a special case of the RBF kernel, under specific parameters [Hsu and Lin, 2002]. Based on the above, for the current study only RBF kernels were used.

For the training of the SVM classifier, an error parameter C and a kernel parameter γ had to be obtained. In order to find the optimal parameters for the RBF kernel function a cross-validation procedure was followed. First the training set was scaled to the range of $[-1, +1]$ to avoid features in greater numerical ranges dominating those in smaller ranges [Negnevitsky, 2005]. Then, the training set was divided to many smaller sets of equal size. Sequentially each subset was tested using the classifier trained by the remaining

subsets. This way each image object was predicted once during the above process. The overall accuracy of the cross-validation was the percentage of correctly classified image objects.

After the cross-validation delivered the optimal parameters for the SVM classifier, the training set was used to train the SVM. Then the classifier was provided with all image primitive objects so to derive the final object based classification. The output of the above procedure was a classification map as well as an updated XML representation of the segmentation level.

4.3 Discussion of Results

For the evaluation of the above methodology, initially a Landsat TM image was used. For comparison purposes, an object-based classification of the same image was performed in eCognition. The training samples in both cases were the same (a TTA mask file) and were obtained by the eCognition user guide (2005) for objective evaluation. The original Landsat TM image and the training samples are presented in Figure 4.2. A reference dataset was also derived from photo-interpretation and was used to compute confusion matrices.

First, the training samples were applied to small primitive objects that were derived by eCognition with scale parameter 10 and by MSEG with scale parameter 100. Scale parameters are implementation dependent [Tzotsos and Argialas, 2006]. For the export of training samples, the minimum overlap for sample object was set to 50% [Benz et al., 2004]. The overall accuracy of the Nearest Neighbor (NN) method, based on the reference dataset was 85.6%. The overall accuracy of the object-based SVM classification was 90.6% (Figure 4.3, Tables 4.1 and 4.2).

Then, in order to test the generalization ability of both classifiers, an error was introduced into the training samples, in the form of not using a minimum overlap restriction for sample object selection. This way, more training objects were selected with errors derived from the segmentation procedures. An interesting observation was that the SVM behaved better than the NN at the second training set and provided better classification results (Tables 4.3 and 4.4) giving an overall accuracy of 86.0% against 84.1% for the NN. Both classification results are presented in Figure 4.3.

The evaluation of the methodology also included testing on very high resolution data

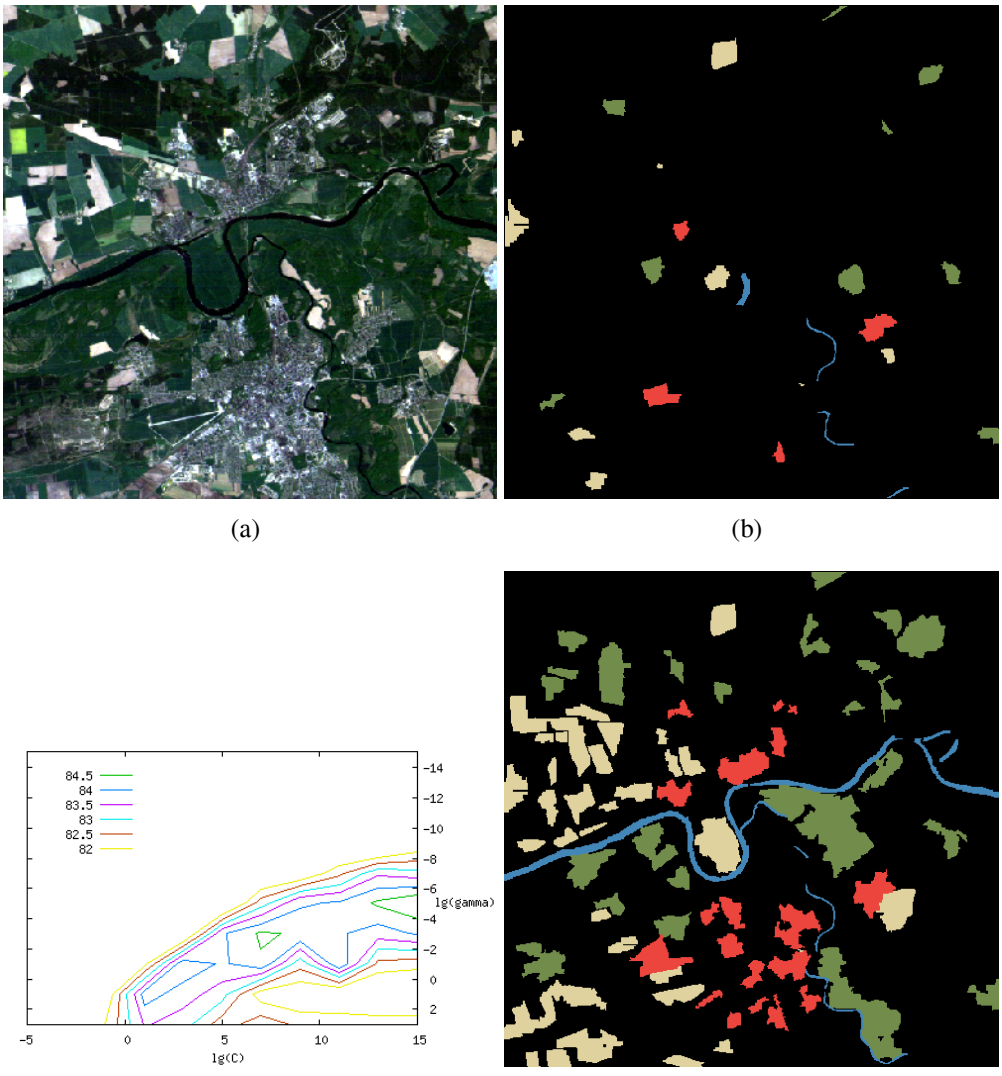
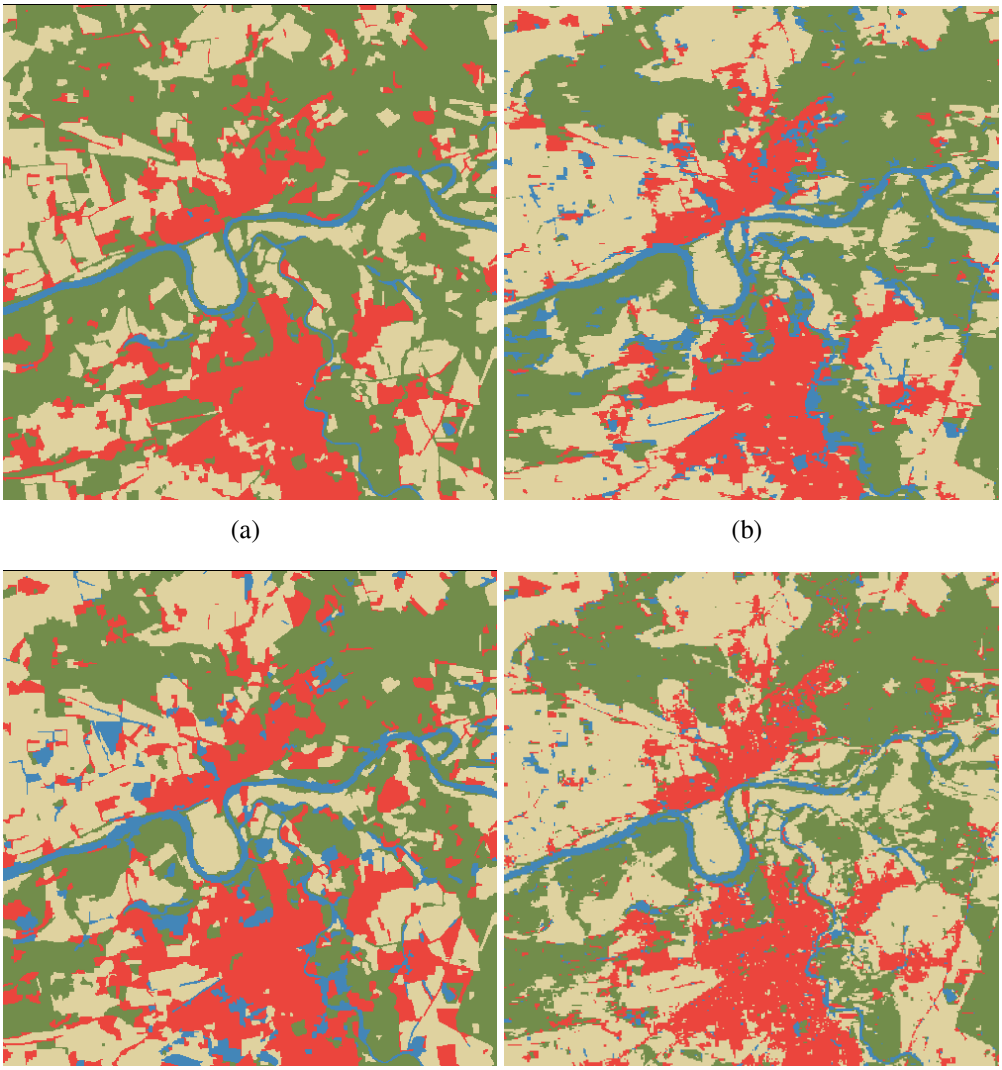


Figure 4.2: (a) the original Landsat TM image (source: eCognition User Guide 2005). (b) The training set of class samples(blue=Water, red=Impervious, green=Woodland and yellow=Grassland. (c) The cross-validation plot diagram for selecting the optimal values of C and γ for the SVM training. (d) The ground-truth dataset used to evaluate results



(a) (b)

(c) (d)

Figure 4.3: (a) eCognition classification result with Nearest Neighbor. (b) MSEG classification result with SVM (c) eCognition classification result with Nearest Neighbor after introducing sample errors. (d) MSEG classification result with SVM. Errors have been introduced to the training sets for generalization evaluation.

	Woodland	Grassland	Impervious	Waterbodies
Woodland	17922	3381	280	0
Grassland	2578	12854	195	0
Impervious	139	770	8539	0
Waterbodies	80	0	0	4740
Overall Accuracy: 85.6%				

Table 4.1: Nearest Neighbor confusion matrix. The overall accuracy was 85.6%

	Woodland	Grassland	Impervious	Waterbodies
Woodland	17846	2088	45	740
Grassland	767	15937	210	91
Impervious	231	215	8305	263
Waterbodies	180	13	10	4537
Overall Accuracy: 90.6%				

Table 4.2: SVM confusion matrix. The overall accuracy was 90.6%

	Woodland	Grassland	Impervious	Waterbodies
Woodland	16080	1470	0	0
Grassland	2195	13891	195	0
Impervious	899	314	8605	0
Waterbodies	1545	1330	214	4740
Overall Accuracy: 84.1%				

Table 4.3: Nearest Neighbor confusion matrix after introducing sample error. The overall accuracy was 84.1%

	Woodland	Grassland	Impervious	Waterbodies
Woodland	16816	3458	207	238
Grassland	1262	15506	178	59
Impervious	249	325	8315	125
Waterbodies	349	755	1	3635
Overall Accuracy: 86.0%				

Table 4.4: SVM confusion matrix after introducing sample error. The overall accuracy was 86.0%.

(Toposys GmbH). An aerial scanner image with resolution of 0.5m was used for classification purposes. This image was selected because outperforms in resolution all commercial satellite data available today. Again, the SVM classification method was compared to the Nearest Neighbour classifier of eCognition software. In both cases 4 basic land cover classes were selected and the same samples were used to train the classifiers. In Figure 4.4, the original dataset is presented. A reference dataset was also created through photo-interpretation in order to compute confusion matrices (Figure 4.4).

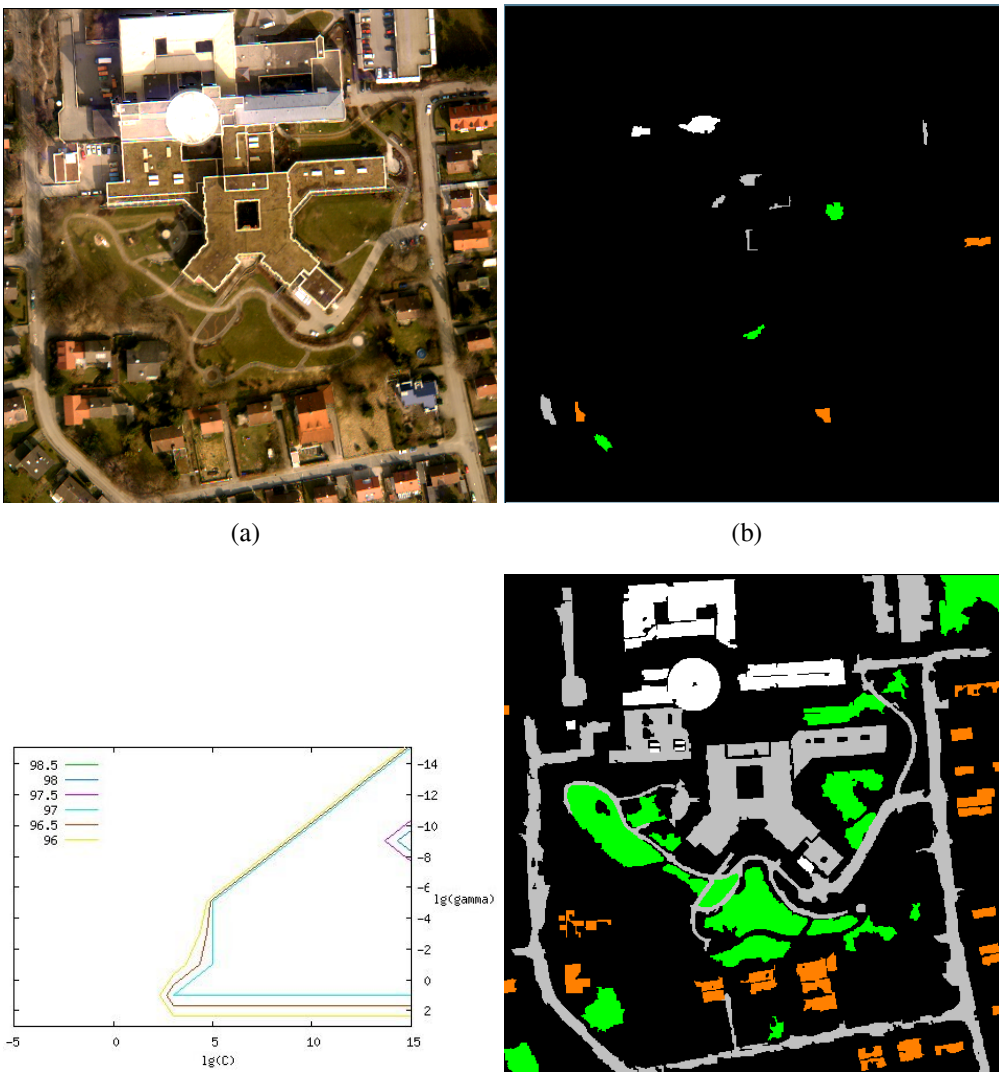
The eCognition software provided primitive objects through the segmentation algorithm with scale parameter 15. Then the sample data were imported to the NN classifier and the classification took place. The minimum overlap between samples and primitive objects was set to 50%. The classification result is presented in Figure 4.5. After the classification step, a confusion matrix was computed (Table 4.5) and the overall accuracy was 87.4%.

The MSEG algorithm provided primitive objects for testing the object-based SVM classifier. The scale parameter was set to 300. The sample data were given as input to the SVM classifier and a cross validation procedure was followed to provide the best C and γ parameters for the SVM classifier. After the classification step, a confusion matrix was computed (Table 4.6) The overall accuracy of the object-based SVM classification was 87.6%.

Although Table 4.5 and 4.6 overall accuracies seem to imply similar classification results, Figure 4.5 shows that there is not great similarity. A closer look at the confusion matrices reveals that both classifiers include classification errors, but in different class combinations. This behaviour is caused by the fact that object-based NN is treating internally each sample as a separate class to calculate distances, while SVM uses only the objects that form the support vectors for learning.

	Vegetation	Tile Roofs	Bright Roofs	Asphalt Like
Vegetation	15708	1528	0	887
Tile Roofs	0	2672	0	238
Bright Roofs	0	12	8408	387
Asphalt Like	2078	1706	2019	34847
Overall Accuracy: 87.4%				

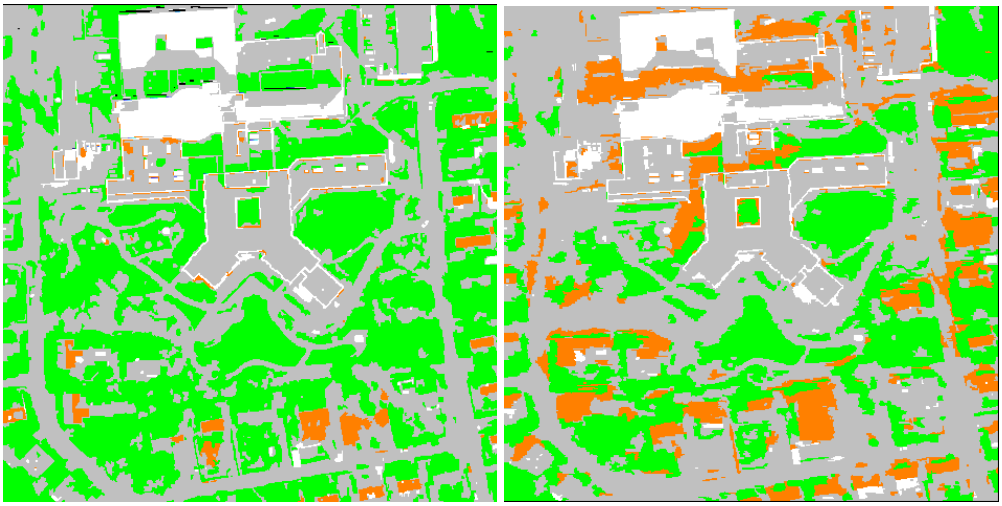
Table 4.5: Nearest Neighbor confusion matrix. The overall accuracy was 87.4%.



(a) (b)

(c) (d)

Figure 4.4: (a) the original aerial scanner image (source: Toposys GmbH) (b) The training set of class samples (white=bright roofs, grey=asphalt like materials, green=vegetation and orange=tile roofs). (c) The cross-validation plot diagram for selecting the optimal values of C and γ for SVM training. (d) The ground-truth dataset used to evaluate results.



(a) (b)

Figure 4.5: Left: eCognition classification result with Nearest Neighbor. Right: MSEG classification result with SVM. Training sample overlap with objects set to 50%

	Vegetation	Tile Roofs	Bright Roofs	Asphalt Like
Vegetation	14760	39	0	2987
Tile Roofs	36	4229	480	1173
Bright Roofs	0	18	8377	2032
Asphalt Like	45	1493	421	34400
Overall Accuracy: 87.6%				

Table 4.6: SVM confusion matrix. The overall accuracy was 87.6%

4.4 Conclusions

Overall, the SVM classification methodology was found very promising for Object-Based Image Analysis. It has been shown that it can produce better results than the Nearest Neighbor for supervised classification.

The computational efficiency of SVM was great, with only a few seconds of runtime necessary for training. This was theoretically expected but also, the implementation in C++ is extremely fast. This performance result occurred on test images up to 1000x1000 pixels size. However, very large remote sensing datasets were not tested.

A very good feature of the SVM is that only a small sample set is needed to provide very good results, because only the support vectors are of importance during training. Future work will include comparison of many SVM kernels for object oriented image classification. Also, an integration of SVM classifiers with rule-based classifiers will be implemented for context-based classification.

Chapter 5

Object-based image analysis through nonlinear scale space filtering

5.1 Introduction

Along with the gradual availability of earth observation data with higher spatial and spectral resolution, research efforts in classifying remote sensing data have been shifting in the last decade from pixel-based to object-based approaches. Assigning land cover classes to individual pixels can be intuitively proper and functional for low resolution data. However, this is not the case for the emerging applications which arise from the continuously improving remote sensing sensors [Aplin and Smith, 2008; Blaschke, 2010; Blaschke et al., 2008]. This is mostly because, at higher resolutions, it is a connected group of pixels that is likely to be associated with a land cover class and not just an individual pixel.

In addition, the earth surface exhibits various regular and irregular structures which are represented with a certain spatial heterogeneity in images. This heterogeneity appears with variations in intensity, scale and texture. Several important aspects of earth observation data can not be analyzed based on pixel information, but can only be exploited based on contextual information and the topologic relations of the objects of interest [Argialas and Harlow, 1990; Liu et al., 2008] through a multi-scale image analysis [Benz et al., 2004; Blaschke and Hay, 2001; Duarte Carvajalino et al., 2008; Hall and Hay, 2003; Hay et al., 2002; Jimenez et al., 2005; Ouma et al., 2008; Stewart et al., 2004]. Starting

with the observed spatial heterogeneity and variability, meaningful spatial aggregations (objects) can be formed at certain image scales configuring a relationship between ground objects and image objects. With such an object-based multi-scale analysis, which is based on certain hierarchically structured rules, the relationships between the different scales of the spatial entities are being described.

The semantic objects of an image do not belong to a single but to various spatial scales. The use of scale space image representations is thus of fundamental importance for a number of image analysis and computer vision tasks. It dates back to sixties and was first introduced by Iijima [Weickert et al., 1999]. Following the ideas of Witkin [1983], Koenderink [1984] and Lindeberg [1994], many methods have been introduced to derive linear scale spaces and respectively many isotropic multi-scale operators have been developed. Either through Gaussian filtering or through isotropic multi-resolution analysis (e.g. by down-sampling the initial data), all linear scale space approaches present the same important drawback: image edges are blurred and new non-semantic objects may appear at coarse filtering scales [Ouma et al., 2008; Paragios et al., 2005; Witkin, 1983]. Under a hierarchical multi-scale segmentation or an object-based classification framework, the thematic information to be extracted is directly related with the primitive image objects computed at every segmentation scale. The better these primitive objects represent real-world entities, the better they can describe the semantics of the image [Blaschke et al., 2008; Hay and Castilla, 2006; Hofmann et al., 2008]. Therefore, the selection of the appropriate approach for constructing the multi-scale image and the hierarchical object representation is of great importance.

The motivation, here, was to embed into an object-based processing scheme an advanced scale space formulation, which possesses suitable qualitative properties desired in image analysis and remote sensing. The introduced image classification framework incorporates advanced morphological scale space filtering and therefore enforces the multi-scale segmentation and class separation procedures to be constrained by hierarchically simplified image representations. Morphological levelings, a kind of advanced self-dual morphological operators, were selected possessing a number of desired spatial and spectral properties [Karantzalos et al., 2007; Meyer, 2004; Meyer and Maragos, 2000b] and were associated with anisotropically diffused markers. Based on these advanced multi-scale image representations, image objects were computed at various segmentation scales and were connected to a kernel-based learning machine for the classification of various

earth-observation data from both active and passive remote sensing sensors.

Previous object-based image analysis approaches [Baatz and Schape, 2000; Benz et al., 2004; Blaschke et al., 2004; Carleer et al., 2005; Dragut et al., 2009; Hay et al., 2002; Ouma et al., 2008; Tzotsos and Argialas, 2006; Tzotsos et al., 2008; Zhou et al., 2009], require the tuning of parameters (such as shape, color, segmentation scale, texture etc) that define the multi-scale object representation. By contrast, in this approach, the scale hierarchy is implicitly derived from scale space representation principles. Furthermore, the developed framework does not incorporate $i)$ any linear scale space filtering (like in Blaschke and Hay 2001; Hay et al. 2002, 2003; Stewart et al. 2004) or $ii)$ any multi-resolution image representation by down-sampling the image at different spatial resolutions (like in Hall and Hay 2003). Such a process actually performs in a similar way with the isotropic filtering, possessing the same qualitative drawbacks. In the proposed approach, advanced morphological scale space representations have been embedded in the object-based image analysis framework and thus, the construction of multi-scale hierarchical object representations was adequately constrained by a refined edge-preserving geometric image simplification (Fig. 5.1). Last but not least, unlike other research efforts which exploited the use of anisotropic diffusion for pixel-based remote sensing data classification [Camps-valls and Bruzzone, 2005; Lennon et al., 2002; Plaza et al., 2009], the developed methodology introduces the use of anisotropic morphological levelings [Karantzalos, 2008; Karantzalos et al., 2007] under an object-oriented classification scheme.

The remainder of the chapter is structured as follows. In Section 5.2, the related work on the use of scale space techniques and object-based classification schemes for remote sensing applications is briefly described. The developed object-based classification framework is detailed in Section 5.3, along with a description and detailed analysis of its different processing steps. Experimental results, the performed quantitative evaluation and the discussion of results are given in Section 5.4. Finally, conclusions and perspectives for future work are in Section 5.5.

5.2 Related Work

5.2.1 Scale space representations for remote sensing image analysis

There are linear and non-linear scale space representations. Since linear scale space approaches, by acting isotropically in the image domain, delocalize and blur image edges, non-linear operators and non-linear scale spaces have been studied and applied in various image processing and computer vision applications. Following the pioneering work of Perona and Malik [1990] there has been a flurry of activity in partial differential equation approaches and anisotropic diffusion filtering techniques [Weickert, 1998]. For remote sensing applications, a number of anisotropic diffusion schemes have been proposed and applied to aerial and satellite datasets [Camps-valls and Bruzzone, 2005; Duarte-Carvajalino et al., 2007; Karantzalos and Argialas, 2006; Lennon et al., 2002; Ouma et al., 2008; Plaza et al., 2009], combined, in most cases, with pixel-based classification techniques. These scale space formulations were based either on diffusions during which the average luminance value is preserved or on geometrically driven approaches formulated under a variational framework. Although these formulations may reduce the problems of isotropic filtering, they do not eliminate them completely: spurious extrema and important intensity shifts may still appear [Karantzalos et al., 2007; Meyer and Maragos, 2000b].

Another approach to produce non-linear scale spaces is through mathematical morphology and, in particular, with morphological levelings, which have been introduced by Meyer [1998a] and further studied by Matheron [1997] and Serra [2000]. Morphological levelings overcome the drawback of spurious extrema or important intensity shifts and possess a number of desired properties for the construction of advanced scale space representations. Levelings, which are a general class of self-dual morphological operators, do not displace contours through scales and are characterized by a number of desirable properties for the construction of non-linear scale space representations. They satisfy the following spatial and spectral properties/principles [Karantzalos et al., 2007; Meyer, 2004; Meyer and Maragos, 2000b]:

- invariance by spatial translation,
- isotropy, invariance by rotation,

- invariance to a change of illumination,
- the causality principle,
- the maximum principle, excluding the extreme case where the image is completely flat.

In addition, levelings:

- do not produce new extrema at larger filtering scales,
- enlarge smooth zones,
- they also create new smooth zones
- they are particularly robust (strong morphological filters)
- they do not displace edges

Following the definitions from Meyer [2004] and Karantzalos et al. [2007], by comparing the values of neighboring pixels in the image domain, levelings are a particular class of images with fewer contours than a given image f. One can define a function g as a leveling of another function f if and only if the following inequality holds:

$$f \wedge \delta g \leq g \leq f \vee \varepsilon g \tag{5.1}$$

where δ is an extensive operator ($\delta g \geq g$) and ε an anti-extensive one ($\varepsilon g \leq g$).

For the construction of levelings, a class of functions h is defined, which separates function g from the reference function f. This type of function is known as a marker function [Meyer and Maragos, 2000b] and can be defined with the following formulation $g \wedge f \leq h \leq g \vee f$. Algorithmically, one can interpret the above equation and construct levelings with the following pseudo-code: in cases where $\{h < f\}$, replace the values of h with $f \wedge \delta h$ and in cases where $\{h > f\}$, replace the values of h with $f \vee \varepsilon h$. Equally and in a single parallel step we have

$$h = (f \wedge \delta h) \vee \varepsilon h \tag{5.2}$$

The algorithm is repeated until the above equation has been satisfied everywhere. This convergence is certain, since the replacements on the values of h are pointwise monotonic. Hence, levelings can be considered as transformations $\Lambda(f, h)$ where a marker h is transformed to a function g, which is a leveling of the reference signal f. Where $\{h < f\}$, h is increased as little as possible until a flat zone is created or function g reaches the reference function f and where $\{h > f\}$, h is decreased as little as possible until a flat zone is created or function g reaches the reference function f. This makes function g to be flat on $\{g < f\}$ and $\{g > f\}$ and the procedure continues until convergence.

Different types of levelings can be constructed based on different types of extensive δ and anti-extensive ε operators. Based on a family of extensive dilations δ_i and the corresponding family of adjunct erosions ε_i, where $\delta_i < \delta_j$ and $\varepsilon_i > \varepsilon_j$ for $i > j$, multi-scale levelings (a hierarchy of levelings) can be constructed [Meyer and Maragos, 2000b]. Multi-scale levelings can also be constructed when the reference function f is associated with a series of marker functions $\{h_1, h_2, ..., h_n\}$. The constructed levelings are respectively,

$$g_1 = f,\ g_2 = \Lambda(f, h_1),\ g_3 = \Lambda(f, h_2), ...,\ g_{n+1} = \Lambda(f, h_n) \tag{5.3}$$

Thus, a series of simpler and simpler images, with fewer and fewer smooth zones are produced.

5.2.2 Object-based Image Analysis

Instead of classifying individual pixels into discrete land cover classes, object-based classification approaches construct a hierarchical object representation of an image and the classifier is responsible for associating them with a land cover class. Therefore, it is not just the spectral signature of each pixel, but the statistical, geometric and topological characteristics of each object that play a key role during classification.

The commercial availability of such an object-based image analysis software [Baatz and Schape, 2000] enabled the accomplishment of several studies for various engineering and environmental remote sensing applications [Benz et al., 2004; Dragut et al., 2009; Zhou et al., 2009, and references therein]. To this end, during the last decade, the chal-

lenge was to construct an efficient object representation through certain multi-scale (region merging or other) segmentation techniques [Blaschke et al., 2004; Carleer et al., 2005; Jimenez et al., 2005; Neubert et al., 2006; Tzotsos and Argialas, 2006], which partition the image on several regions/objects based on the spectral homogeneity in a local neighborhood. In addition to the spectral homogeneity criterion, shape parameters are used to define geometric properties that the segmentation algorithm must take into account when computing the overall homogeneity (scale parameter) of each image object during the search for optimal merges. A texture optimization procedure was introduced for the MSEG algorithm [Tzotsos et al., 2008] integrating grey level co-occurrence matrices and introducing an object-based cost measure for texture homogeneity as an additional parameter to the segmentation procedure. Such an integration of spatial and spectral information can produce a multi-scale object representation but only through an iterative and exhaustive tuning (based on trial and error investigation) of certain parameters, like shape, scale, texture, etc. [Baatz and Schape, 2000; Benz et al., 2004; Blaschke et al., 2004; Carleer et al., 2005; Dragut et al., 2009; Hay et al., 2005; Ouma et al., 2008; Tzotsos and Argialas, 2006; Zhou et al., 2009].

Other research efforts were based on the construction of linear scale spaces for the multi-scale analysis of several landscape structures [Blaschke and Hay, 2001; Hay et al., 2002, 2003; Stewart et al., 2004] or on the construction of multi-scale representations through object-specific analysis and up-scaling, through the computation of a number of coarse and fine scales by sampling the initial image [Hall and Hay, 2003].

The object-based classification framework developed here is based on *i)* the construction of advanced morphological scale space representations (Fig. 5.1) and *ii)* an advanced kernel classifier, the Support Vector Machine (SVM). SVM [Theodoridis and Koutroumbas, 2003; Vapnik, 1998] is a state-of-the-art machine learning methodology that can be used for pattern classification and has proven very successful in pixel-based remote sensing applications [Foody and Mathur, 2004; Huang et al., 2002a; Melgani and Bruzzone, 2004; Mercier and Lennon, 2003]. Recently, an object-based image classification scheme was proposed using Support Vector Machines [Tzotsos and Argialas, 2008].

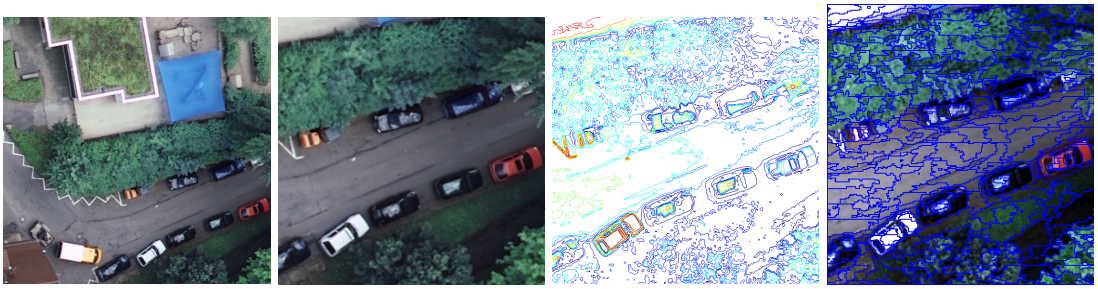
5.3 Methodology

The developed object-based framework is integrating certain computer vision and machine learning methods for performing common Object Based Image Analysis (OBIA) tasks, such as image segmentation, object hierarchy representation, classification etc. This section is divided into three sub-sections which describe the three major components of the framework: scale-space filtering, image segmentation and kernel-based object classification. An overview of the developed approach is shown in Fig.5.2. Before going into a detailed account of each of the three sections, a general presentation of the developed methodology is presented.

The proposed algorithm was designed in such a way that can take as an input any type of remote sensing imagery (Multispectral, Panchromatic, Radar, Hyperspectral etc). Every initial image is decomposed into its n separate bands in order to achieve the proper non-linear filtering, depending on the scale parameter of the scale space filter. For each band, a scale space representation is created based on Anisotropic Morphological Leveling (AML) formulation. Each scale space cube is a 3D representation of the initial band at m successive filtering scales (1, 2, 3, ... m). From such a band-oriented representation, a scale-oriented one is constructed by merging bands of the same filtering scale. Thus, a series of m simplified images (of m successive filtering scales) is constructed, from which the multi-scale object representation will be derived.

A multi-scale region merging segmentation algorithm is then applied to each simplified image. During this procedure, the tuning of the standard segmentation parameters is of less importance, and only the scale parameter of the merging plays a key role. In contrast with most OBIA implementations, the size and shape of objects are constrained primarily by the morphological scale space filtering. In Fig.5.1 it can be observed that for the same segmentation parameters, different segment sizes are produced, with different shape each time. The segmentation algorithm is applied to each simplified image, without the need of tuning the standard parameters (shape, color, texture etc.). The multilevel object representation is derived only using the segmentation scale parameter. Consequently, the final object hierarchy is dependent on the filtering scale and segmentation scale parameters.

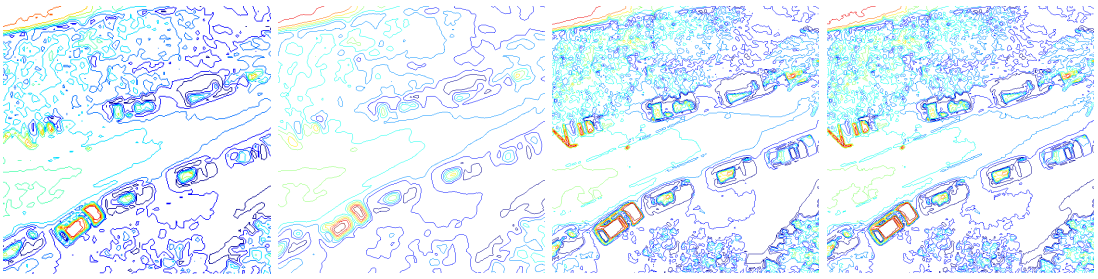
The final step of the algorithm includes the definition of the class hierarchy, according to the semantics of the image, and the classification, which is performed by a support



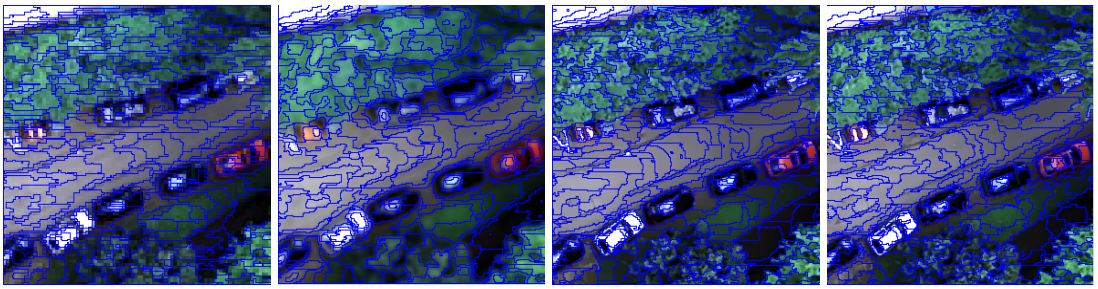
(a) Initial Image (b) Crop (c) Contours (d) Segments

(e) MR, Sc: 0.8 (f) Gauss, Sc: 2 (g) ALM, Sc: 50 (h) AML, Sc: 50

(i) MR, Sc: 0.4 (j) Gauss, Sc: 7 (k) ALM, Sc: 500 (l) AML, Sc: 500



(m) Contours MR (n) Contours Gauss (o) Contours ALM (p) Contours AML



(q) Segments MR (r) Segments Gauss (s) Segments ALM (t) Segments AML

Figure 5.1: Comparing the scale space representation of linear and non-linear techniques. Two scales from each technique are presented, along with the corresponding image contours (isophotes)(c,m,n,o,p) and the coarser scale (i,j,k,l) segments (q,r,s,t). For the

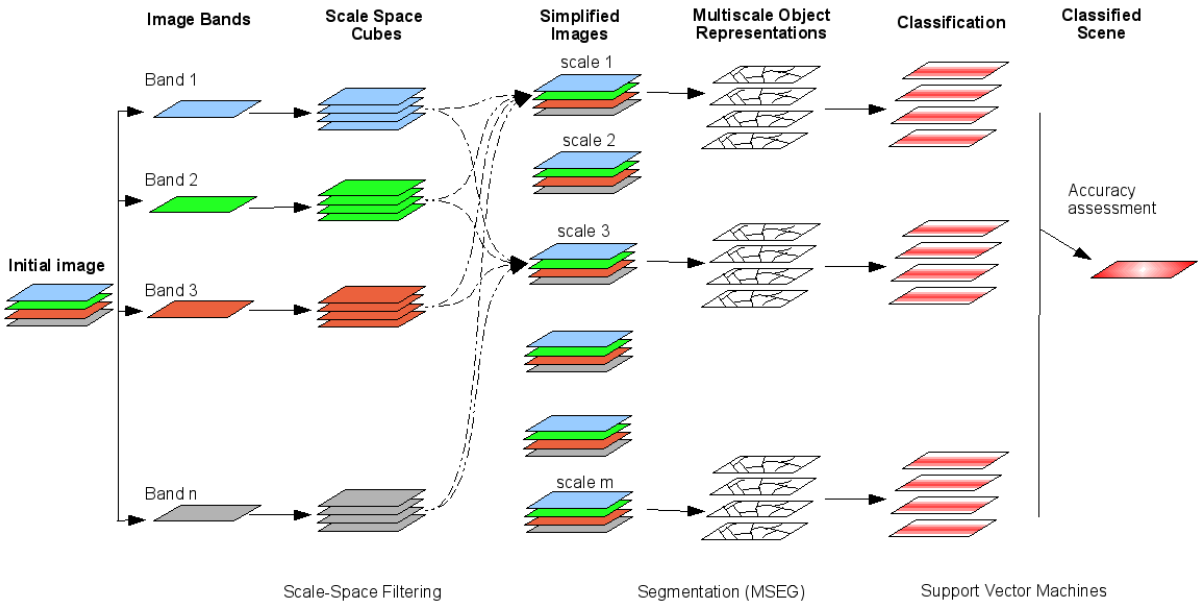


Figure 5.2: The developed framework embeds nonlinear scale space filtering into the object-based image classification procedure.

vector machine classifier using training samples for each class. For each level of objects created from the segmentation step and for all image filtering scales, a machine learning procedure is executed, providing object classifications of equal number. Finally, an accuracy assessment step is performed and the selected as optimal classification result is obtained.

5.3.1 Anisotropic Morphological Levelings

As in many digital remote sensing applications and methods, the first step in the developed approach is a low-level preprocessing of the original image. This initial filtering serves the purpose of removing noise, as well as simplifying the complexity of the information included in the original data. Especially in high spatial resolution remote sensing imagery, it is of great importance to reduce the heterogeneity of the original dataset in order to achieve better medium (segmentation) and high level (classification) results. Image semantic objects tend to incorporate great value of spectral heterogeneity as the resolution gets higher with the recent available sensors. For example, in Fig.5.3 an ultra high spatial resolution image (5 cm pixel size) is presented acquired from an aerial multispec-

tral scanner. One can observe the tiles on the roof of the building, the road marking signs and the fine texture of the rooftop material. It is obvious that this kind of detail cannot be addressed properly either through pixel-based procedures or through a simple OBIA scheme.

(a) Original Image (b) AML, Scale: 100 (c) AML, Scale: 500

Figure 5.3: AML filtering results on DMC aerial multispectral image with 5 cm pixel-size (©Intergraph Corp.)

Therefore, non-linear scale space filtering was selected to be the first step towards properly simplifying the initial image, towards addressing the usual over-segmentation and misclassification problems. As stated previously, state-of-the-art filtering methods have moved away from linear models and tend to preserve the edge information like Anisotropic Diffusion (ADF). Morphological Levelings are another powerful tool for obtaining scale-space information [Karantzalos and Argialas, 2006; Meyer and Maragos, 2000b]. A combination of morphological levelings with anisotropic markers, known as Anisotropic Morphological Levelings (AML), have given better results, by possessing the ability for image simplification and at the same time by preserving important image properties [Karantzalos and Argialas, 2006; Karantzalos et al., 2007].

In this study, an AML implementation has been incorporated as a preprocessing tool for object-based image analysis. Its native multi-scale representation ability was exploited along with the use of a multi-scale segmentation method to derive better image primitive objects, and thus better classification results. The AML filtering was applied to every single band of the original data. At each iteration, anisotropic markers were constructed based on the formulations of Alvarez et al. [1992]. The greater the number of filtering scales used, the stronger the simplification, since an equal number of diffusions are per-

formed in the original data in order to provide the markers for the reconstruction of the leveling.

In the following sections, results from the use of AML filtering will be presented as well as its contribution to the overall developed approach for a variety of remote sensing data.

5.3.2 Multiscale Segmentation

For the multi-scale segmentation procedure, the MSEG algorithm was employed for providing the primitive image objects at different scale parameters. MSEG is a region-based multi-scale segmentation algorithm recently developed for object-oriented image analysis [Tzotsos and Argialas, 2006]. It can be described as a region merging procedure, starting from a pixel representation and creating objects through continuous pair-wise object fusions, executed in iterations, called *passes* of the algorithm. For each pass, every object is evaluated in relation with its neighboring objects towards the optimal pair of objects adequate for fusion. In every pass, an image object can be merged only once, aiming at a balanced object growth.

In order to achieve reproducibility of results a heuristic procedure was introduced, called Starting Point Estimation (SPE). Using image statistics and color-space transformations, the image is partitioned in tiles and then starting points are computed for each tile [Tzotsos and Argialas, 2006]. These points are not used as seeds, but are used to determine the order in which object merging evaluation will be performed. In this way, MSEG produces exactly the same result for the same parameters and the same initial image.

Like many other region-based segmentation algorithms [Baatz and Schape, 2000; Pal and Pal, 1993], the MSEG algorithm defines a cost function for each object merge and then implements various optimization techniques to minimize this cost. The cost function is implemented using the measure of homogeneity (color and shape) in the same way with other approaches [Baatz and Schape, 2000]. The threshold of the allowed merging cost for the segmentation procedure is called scale parameter, since it implicitly dictates the area growth of the image objects. In order to achieve a multi-scale object representation, several scale parameters must be defined during several executions of the segmentation algorithm. For the integrity of the topological relations of the objects, a set of cross-level

constraints can be activated [Tzotsos and Argialas, 2006].

Given that the goal, here, was to design a generic processing framework, several filtering and segmentation scales were examined and validated through quantitative and qualitative evaluation based on ground truth data. In particular, a sensitivity analysis was more than important in order to demonstrate that different types of remote sensing imagery respond differently to the same filtering and segmentation methods and, therefore, require different parameter settings for most OBIA procedures. For example, a 32-bit lidar dataset is segmented into smaller objects than an 8-bit multispectral image at the same scale parameter due to different radiometry, magnitude of edges, number of bands, etc.

A novelty of the developed segmentation process is that, by employing an elegant scale space filtering, the tuning of the region-merging parameters is of less importance to the final segmentation result. There was no need for an exhaustive manual search to obtain the best parameter settings for multilevel segmentation. It is the scale space filtering that primarily constrained the construction of the multi-scale object representation and not the region-merging procedure. A sensitivity analysis was performed to demonstrate this concept and it is discussed in the evaluation section bellow (Fig.5.10, Section 4.3).

Using non-linear scale-space filtering before segmentation lead to improvements as shown in Fig.5.1. To further demonstrate this, having all parameters (except scale parameter) set to their default values during testing, several segmentation tests were performed on the initial images, as well as on the filtered images. Furthermore, tests were performed using the Baatz and Shape's segmentation algorithm producing improved results. The improved segmentation results can be implicitly proved by the increase of classification accuracy, when using the same segmentation parameters for the same image, the same class samples and the same classification parameters. Class samples were initially provided as external vector files and then a percentage of overlap between samples and objects was computed in order to select the sample objects for training.

5.3.3 Support Vector Machines

For the developed approach, the SVM classifier by Tzotsos and Argialas [2008] was employed. After image segmentation, image objects were extracted and object properties were computed forming the feature space of the classification problem. The computed

properties are bound to each object by a unique identifier within the object hierarchy of the image. Some of the objects were selected as samples and their properties formed a training set for the SVM.

In general, the SVM seeks to find the optimal separating hyperplane between classes by focusing on the training data that are placed at the edge of the class descriptors. These training data are called support vectors. Training data other than support vectors are discarded. Thus, not only an optimal hyperplane is fitted, but also less training samples are effectively used [Tzotsos and Argialas, 2008]. This method works very well for classes that are linearly separable. In the case that image classes are not linearly separable, the SVM maps the feature space into a higher dimensionality using kernels [Theodoridis and Koutroumbas, 2003; Vapnik, 1998] and then separates classes in that new feature space forming the support vectors.

Since the SVM classification method was initially designed for binary classification problems, a heuristic one-against-one strategy was employed for multiclass classification [Hsu and Lin, 2002]. Many binary classifiers were applied for each pair of classes and for every object of the image and then a max-win operator determined the final classification of the object.

Since SVM is not a parameter-free method, the proper learning parameters had to be determined for each learning procedure. A cross validation scheme was implemented. The training set was divided into several training subsets. Subsequently, some subsets were used for training and others for testing, while changing the training parameters. This process is iterative and the best parameters are determined after several iterations.

After determining the proper learning parameters with cross validation, the training set was used to finally train the SVM. Then, using the one-against-one strategy, all image objects were classified and the class identification tag was applied to the object database. The final step of this procedure was the quantitative quality assessment, using ground truth data, for the formulation of confusion matrices and accuracy measurements.

To sum up, the initial dataset is simplified and a successive series of simplified images were constructed forming a non-linear scale space. A multi-scale object representation was then computed from the scale space cubes without the tuning of any standard parameter (like shape, color, texture, etc). Finally, the classifier based on the statistical properties of each object and their hierarchy associated each one with a land cover class.

5.4 Evaluation and Discussion

As stated earlier, the objectives of the developed approach were: (a) to formulate an advanced multi-scale object representation under an object oriented framework, (b) to construct a processing system with the minimum tuning parameters and (c) to evaluate its performance regarding its qualitative and quantitative behavior in various remote sensing datasets. In particular, of special interest are the very high spatial resolution data, which take classification methods to their limit, due to very high heterogeneity.

In Fig.5.1 a comparison of linear and non-linear scale space representations is presented. The filtering result from a fine and a coarse scale is shown along with the corresponding image contours (isophotes) and segments. One can observe that linear approaches like Gaussian or downsampling do not respect the edge information of the initial image and they degrade or blur the final image result. By observing the object contours after linear image simplification it appears that a segmentation procedure may produce a smaller number of primitive image objects. This desired result, though, is not achieved for small objects of interest or image objects with strong edges. More specifically, in downsampling methods, the boundary of the derived objects is less smooth due to the pixel effect near the edges. The Gaussian filtering is mixing semantic objects after just a few scales of filtering.

On the other hand, anisotropic diffusion filtering (ALM) [Alvarez et al., 1992] preserves edge information even after a significant number of iterations (Fig.5.1). Thus, the image semantics are protected from simplification. The main disadvantage of the ALM algorithm is that it alters the spectral signatures of image objects significantly which is not desired and acceptable for most remote sensing applications. This effect can be observed in Fig.5.4, where a quantitative comparison of ALM and Anisotropic Morphological Levelings takes place in the form of a spectral signature plot. The results were obtained from the application of non-linear filtering on a TerraSAR-X dataset (Fig.5.4). As shown in the plots, the ALM method creates new spectral values for the simplified pixels of the image that are far from the original spectral values. By contrast, the AML method used in this paper, simplifies the image while keeping the new values as close to the original as possible. In addition to edge preservation property, AML emerges as a superior method for scale-space representation. The simplified SAR images are shown in Fig.5.4, while the evolution of the AML simplification across filtering scales is also

shown.

In order to validate the developed algorithm's experimental results and demonstrate its behaviour under several type of datasets and settings, two standard multi-scale segmentation techniques were employed: a multi-scale region-merging approach (MSEG) [Tzotsos and Argialas, 2006]) and the corresponding one embedded in the eCognition software [Baatz and Schape, 2000].

5.4.1 Radar satellite imagery

The MSEG segmentation algorithm was first applied to the TerraSAR-X image. As shown in Fig.5.5 (a) and (b), the obtained primitive image objects are less in number after the simplification step. Moreover, the objects in (b) are more compact in shape and their boundaries are smoother than the ones in (a). It must be noted that both segmentation results were derived by the same segmentation parameters (scale parameter 100, color 0.8, shape 0.2, compactness 0.5 and smoothness 0.5). Furthermore, in (c) and (d), the same test was performed using the eCognition segmentation algorithm with similar results. Therefore, the non-linear simplification addresses the over-segmentation problem and at the same time improves the shape of the resulting image objects by describing them more consistently with the real world objects.

5.4.2 Very high spatial resolution airborne imagery

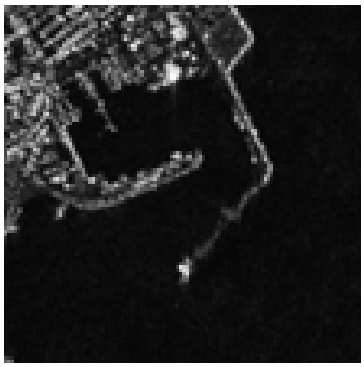
The developed algorithm was also applied to an ultra high spatial resolution dataset, with a pixel size of 5cm from the DMC airborne digital scanner. The goal was to evaluate the behavior of the introduced object-based framework for such a demanding task, i.e classifying objects at such a large scale, processing a huge amount of information (hundreds of megabytes), construct thousands scale space representations. This dataset (Fig.5.1,5.3) shows that heterogeneity inside image objects of a specific thematic class can pose quite a challenge to any classification or segmentation algorithm available today.

In Fig.5.3, steps from the simplification process are presented, while in Fig.5.6 results from segmentation step and the construction of the multilevel object representations are shown. One can clearly observe that, without the tuning of any region merging parameters, the segmentation result is superior when using the AML filtering. In order to validate

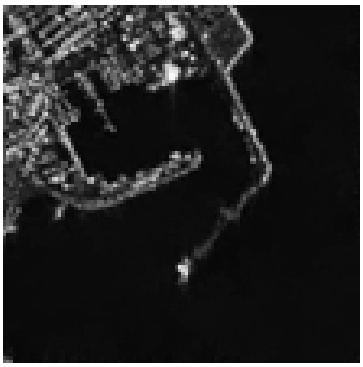
(a) Original Image

(b) AML, Scale: 100

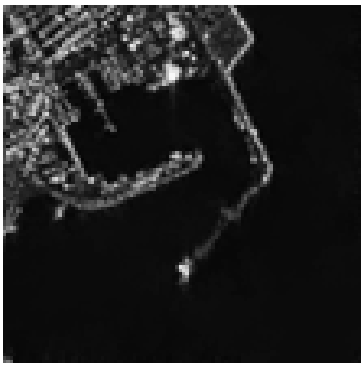
(c) AML, Scale: 500



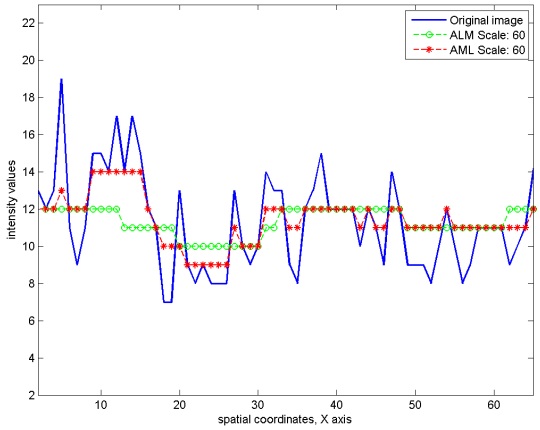
(d) Original Image



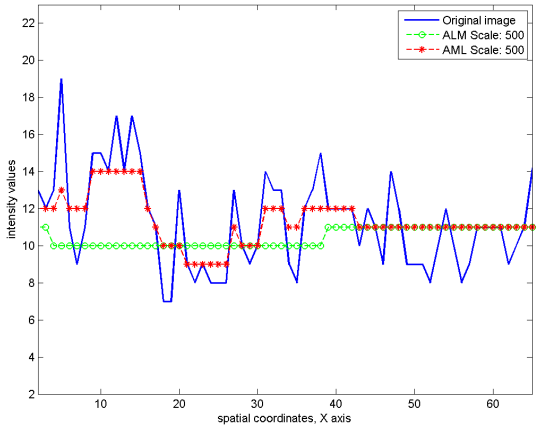
(e) AML, Scale: 100



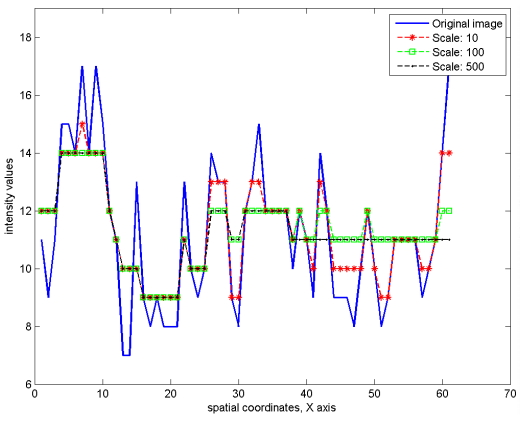
(f) AML, Scale: 500

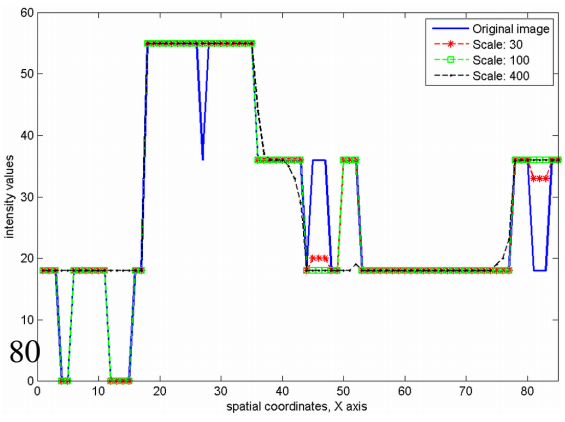


(g) ALM and AML at Scale 60

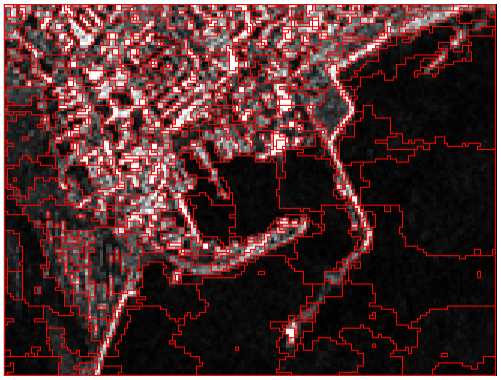


(h) ALM and AML at Scale 500

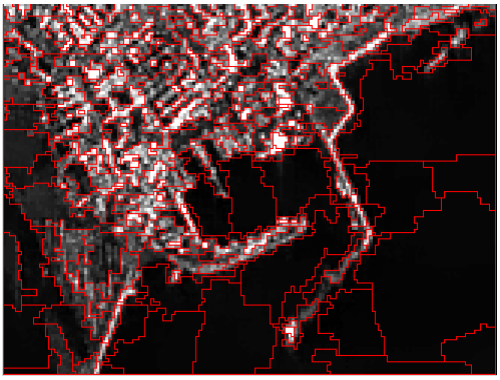




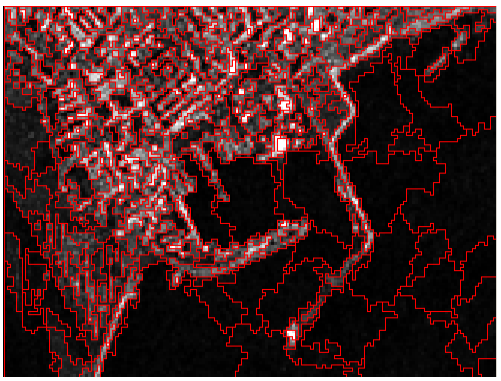
(i) Spatial Cross Sections at Different AML Scales



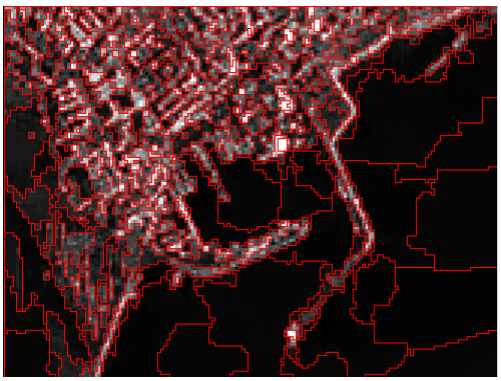
(a) Original image segmented using MSEG



(b) AML filtered image (Sc:500) segmented using MSEG



(c) Original image segmented using eCognition



(d) AML filtered image (Sc:500) segmented using eCognition

Figure 5.5: TerraSAR-X (©DLR) segmentation results with and without the AML filtering using the MSEG and the eCognition software

this observation with an other region merging algorithm, the same experiments were performed with the eCognition segmentation algorithm. While the tiles of the roof on the initial image result into an over-segmented roof when using the standard merging processes, after the application of the AML simplification algorithm the tile heterogeneity is reduced and merging algorithms produce better outcomes. At the same time, those regions of the image which hold edge information are preserved (like dormer windows/roof windows). For those objects, the segmentation result does not change significantly, which is a desired behaviour. Once again, the smoothness of the resulting object boundaries is improved without changing the shape parameters of the segmentation algorithms.

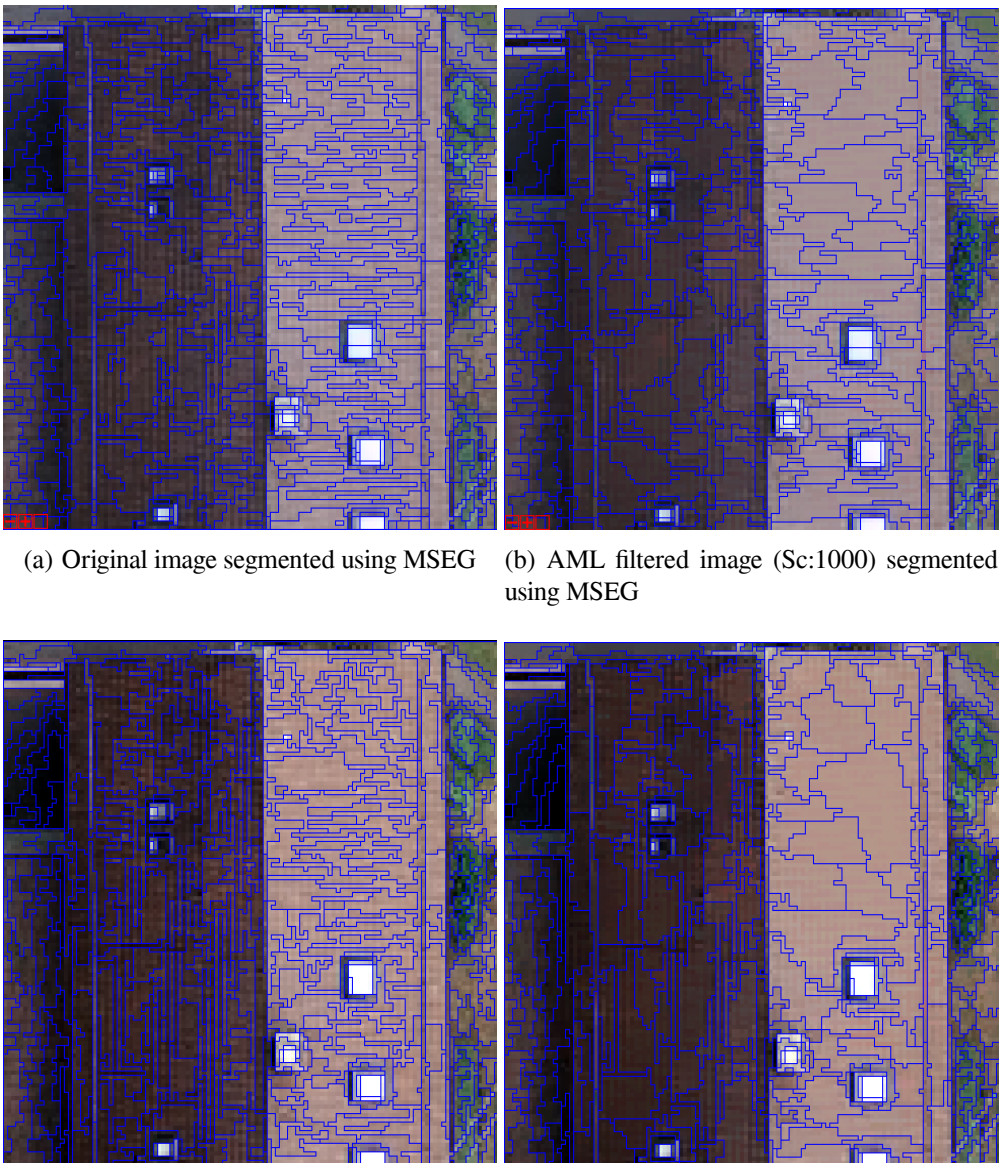
As it is demonstrated in Fig.5.7, similar conclusions can be derived when examining the behavior of the developed algorithm at another region of the ultra high spatial resolution dataset. It can be observed that segments in the car area/regions, where preserving the edge information is crucial in all filtering scales, the computed segments at finer and coarser scales did not change significantly. The same applies to objects that compose the street linings. On the other hand, street and vegetation segments show great improvement, since texture information is simplified in those regions of the image.

5.4.3 Multispectral remote sensing data

The next series of tests was performed on medium and high spatial resolution multispectral remote sensing data. For this, a Landsat TM image with pixel spatial resolution of 30m was used, as well as an aerial scan with half a meter ground resolution and four spectral bands.

After a close look at the results (Fig.5.8), one can observe the superior qualitative object representation acquired under the AML scale space filtering. This observation is validated by the classification results obtained using the same parameter settings when comparing with the ground truth data. In Fig.5.9, the result of the classification comparison is demonstrated. Using exactly the same training samples in both cases, the SVM classification was improved as shown in Fig.5.9 and Table 5.1. Moreover, the same test was performed using the Nearest Neighbor classifier. The overall classification accuracy was still improved, but could not match the accuracy of the advanced SVM classifier (Table 5.2).

At the same time, using the ground truth data, the proposed approach was tested in



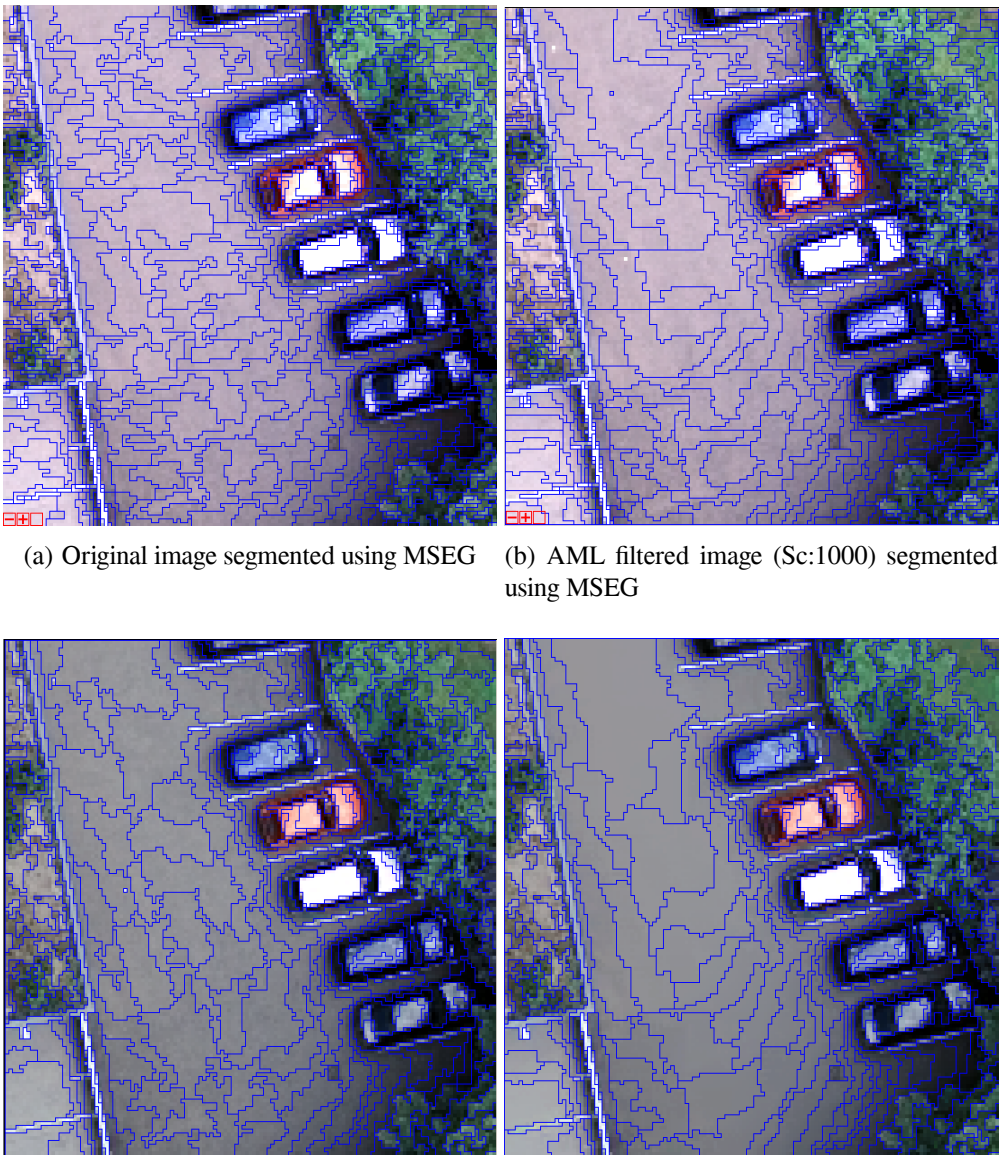
(a) Original image segmented using MSEG

(b) AML filtered image (Sc:1000) segmented using MSEG

(c) Original image segmented using eCognition

(d) AML filtered image (Sc:1000) segmented using eCognition

Figure 5.6: Segmentation tests as performed on the DMC (©Intergraph Corp.) aerial image, with and without the application of nonlinear filtering



(a) Original image segmented using MSEG

(b) AML filtered image (Sc:1000) segmented using MSEG

(c) Original image segmented using eCognition

(d) AML filtered image (Sc:1000) segmented using eCognition

Figure 5.7: Segmentation tests as performed on the DMC (©Intergraph Corp.) aerial image, with and without the application of non-linear filtering

(a) Original image zoom

(b) AML, Scale: 1000

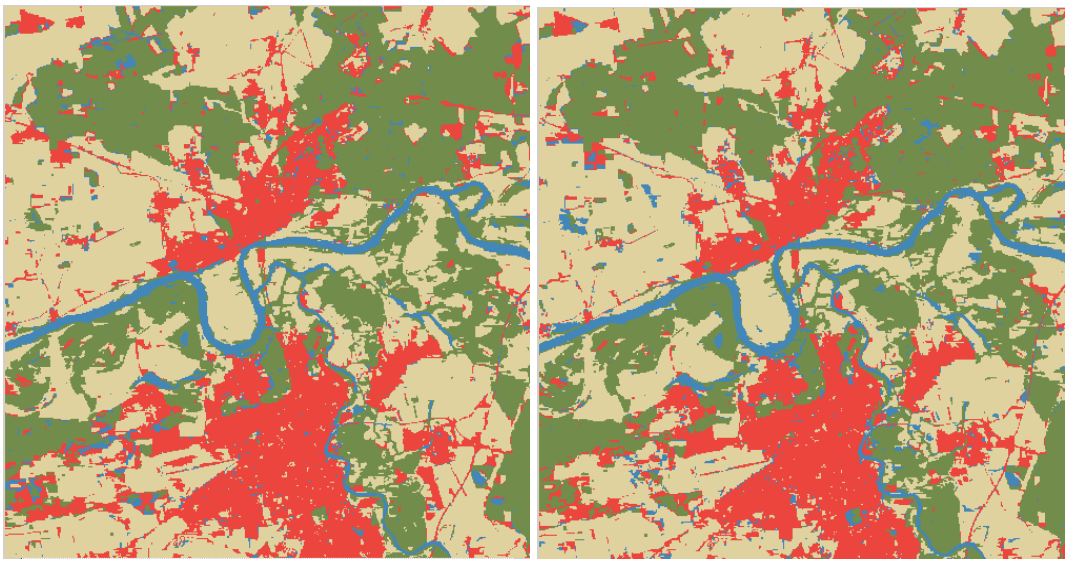
(c) Original image segmented using MSEG

(d) AML filtered image (Sc:1000) segmented using MSEG

Figure 5.8: A segmentation test as performed on a Landsat TM image, with and without the application of non-linear filtering

(a) Initial Image

(b) Ground truth data



(c) Classification Result (MSEG+SVM): Accuracy=87.58%

(d) Classification Result (AML+MSEG+SVM): Accuracy=88.35%

Figure 5.9: Original multispectral Landsat TM image (Dessau, Germany) and classification results with and without AML filtering. Blue: Water Bodies, Yellow: Grassland, Green: Woodland, Red: Impervious

order to evaluate the obtained results. This result is presented in the form of a sensitivity analysis using the accuracy results from the SVM classifier (Fig.5.10 (a) and (b)). The best classification result obtained by the proposed approach for the Landsat image was at segmentation scale 900 and filtering scale 100 with an accuracy of 91.96% (Table 5.1).

The same evaluation test was performed on the multispectral image provided by Toposys. In Fig.5.11, classification results are presented. The performance of the classifier was improved by the non-linear scale-space filtering as shown also in Table 5.2. The sensitivity analysis performed for various segmentation and filtering scales is presented in Fig.5.10 (c) and (d). This test also included anisotropic filtering using ALM for comparison of results. The AML result is superior, especially on higher filtering scales, where the classification accuracy seems to be less sensitive to the scale variations.

Again, the developed framework was executed using the available ground truth data and the best classification result was found to be at segmentation scale 2500, AML scale 500 with overall accuracy of 92.53% (Table 5.2).

It can be observed that the classification through the AML scale-space representation has a more stable/consistent behaviour in both datasets, scoring higher accuracy rates at most filtering scales (Fig.5.10). Although in approximately all cases the use of AML ameliorates significantly the classification accuracy, one can observe that from medium to coarser segmentation scales, the resulting improvement of the developed method is more than 10% compared to earlier efforts (eCognition, MSEG etc).

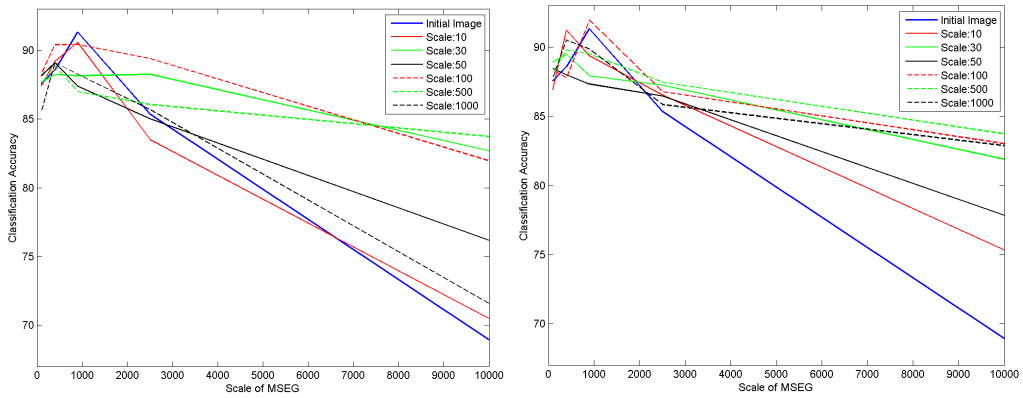
For further validation of the above results, the same number of tests were conducted on the multispectral datasets, after applying linear filtering instead of non-linear filtering. As presented in Fig.5.10 (e and f), the SVM classification algorithm was performed on the same scales of the Gaussian filtered images, using the same training samples. Then, accuracy assessment was performed using the same ground truth data, obtaining new accuracy results. It can be observed (Fig.5.10e and f) that the degradation of the linearly filtered datasets caused a decrease of the maximum classification accuracy for each filtering scale and for both datasets. Also, while scale was increasing, the object-based classification process provided less accurate results.

As it can be observed in Fig.5.10, the application of the non-linear filtering resulted to a more accurate classification for almost all filtering scales. Especially, for the AML filter the classification accuracy increased and it reached its peak (more than 90%) for the filtering scales of 100 and 500. In contrast, the application of the linear filtering

in all cases and for almost all filtering scales impaired the accuracy of the classification (Fig.5.10e and f).

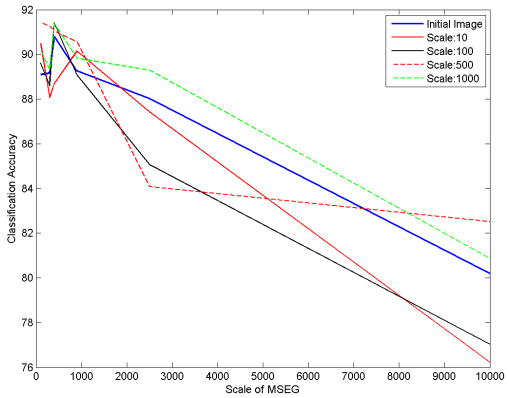
5.5 Conclusions and Future Perspectives

A new object-based classification method was developed based on advanced scale space representations, multilevel object representations and a support vector machine classifier. In contrast to previous efforts, we constructed the multilevel object representation based primarily on the advanced simplification procedure and not on the region merging process. The employed AML scale space formulation was designed and which implicitly possesses a number of desired qualitative properties and thus eliminated the need for tuning several parameters during segmentation. The performed qualitative and quantitative evaluation reported that the developed algorithm outperformed previous efforts, both regarding the construction of the object representations and the classification results. The algorithm is stable, fast and can efficiently account for various classification tasks in various types of remote sensing data. Some of the topics for further research and development are: solutions for object-specific extraction tasks based on the developed framework, incorporating unsupervised and knowledge-based classification approaches and optimizing the algorithm for real time applications.

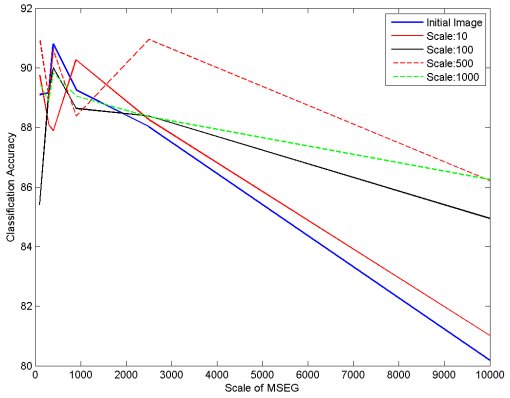


(a) ALM at Landsat TM image

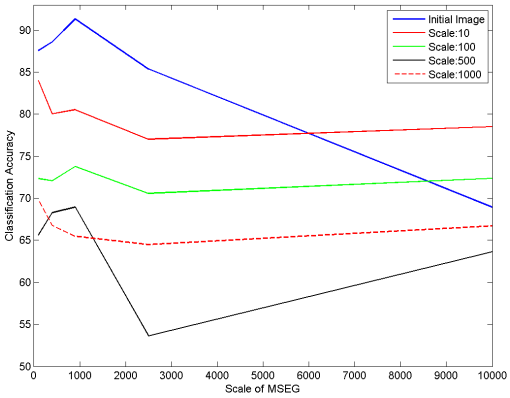
(b) AML at Landsat TM image



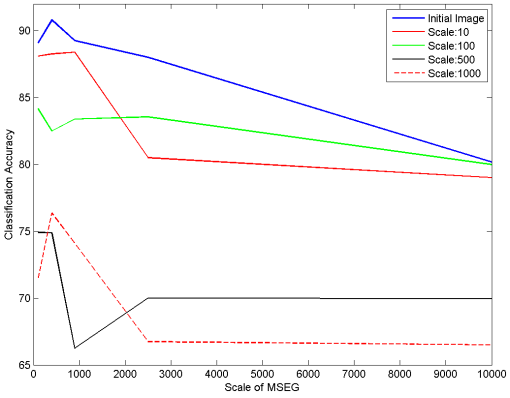
(c) ALM at Toposys aerial scan image



(d) AML at Toposys aerial scan image



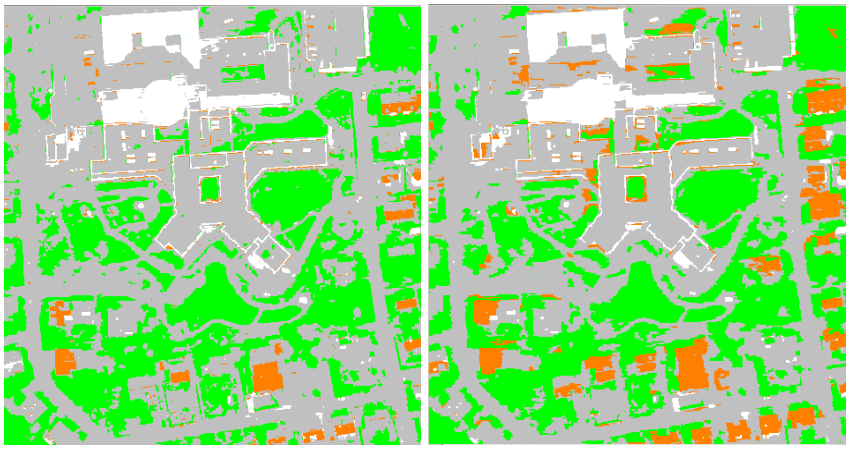
(e) Gaussian at Landsat TM image



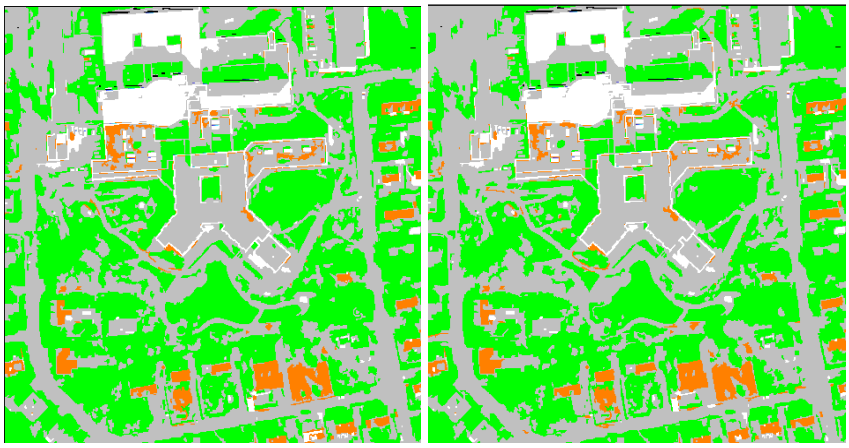
(f) Gaussian at Toposys aerial scan image

Figure 5.10: Sensitivity analysis regarding the accuracy of the classification result for different scales of filtering and segmentation.

(a) Original image (b) Ground truth data



(c) SVM classification result (original image + MSEG). Accuracy=89.1% (d) SVM classification result (AML scale 100 + MSEG).Accuracy=90.16%



(e) Nearest Neighbor classification result (original image). Accuracy= 86.72% (f) Nearest Neighbor classification result (AML scale 100). Accuracy=87.35%

Figure 5.11: Original multispectral aerial scan image (©Toposys) and classification results with and without AML filtering. Green:Vegetation, Grey: Asphalt materials, Orange: Tile roofs, White:Bright roofs

Classification Accuracy without the AML (eCognition, Sc:10)				
	Woodland	Grassland	Impervious	Waterbodies
Woodland	17686	1351	200	113
Grassland	1524	14033	119	4
Impervious	178	1344	8563	6
Waterbodies	1331	277	132	4617
Overall Accuracy: 87,22%				

Classification Accuracy with the AML (eCognition, Sc:10)				
Woodland	16670	48	197	184
Grassland	3528	15624	179	21
Impervious	195	1332	8577	11
Waterbodies	326	1	61	4524
Overall Accuracy: 88,18%				

Classification Accuracy without the AML (MSEG, Sc:100)				
	Woodland	Grassland	Impervious	Waterbodies
Woodland	16479	3551	352	337
Grassland	449	15614	904	38
Impervious	189	263	8369	193
Waterbodies	111	2	3	4624
Overall Accuracy: 87,58%				

Classification Accuracy with the AML (MSEG, Sc:100)				
Woodland	16661	3264	298	496
Grassland	304	15798	668	235
Impervious	93	2	1	4644
Waterbodies	326	1	61	4524
Overall Accuracy: 88,35%				

Best Classification Result with the AML and MSEG at Sc:900				
Woodland	15864	39	0	1883
Grassland	54	5262	10	592
Impervious	0	107	8276	2044
Waterbodies	151	635	154	35419
Overall Accuracy: 91,96%				

Table 5.1: Quantitative results regarding the classification accuracy for the multispectral LANDSAT TM Image. The developed algorithm scored better in all cases, indicating that the combination of the advanced scale space representation (AML) and the kernel classifier (SVM) outperform earlier approaches.

Classification Accuracy without the AML (eCognition, Sc:10)				
	Vegetation	Tile Roofs	Bright Roofs	Asphalt Like
Vegetation	15597	1561	0	1033
Tile Roofs	33	3409	0	1255
Bright Roofs	0	118	8431	381
Asphalt Like	2156	830	1996	33690
Overall Accuracy: 86,72%				

Classification Accuracy with the AML (eCognition, Sc:10)				
Vegetation	15437	1501	0	809
Tile Roofs	22	3581	0	1014
Bright Roofs	0	17	8426	410
Asphalt Like	2327	819	2001	34126
Overall Accuracy: 87,35%				

Classification Accuracy without the AML (MSEG, Sc:100)				
	Vegetation	Tile Roofs	Bright Roofs	Asphalt Like
Vegetation	15731	0	0	2055
Tile Roofs	169	3072	81	2596
Bright Roofs	0	5	8333	2089
Asphalt Like	229	74	383	35673
Overall Accuracy: 89,10%				

Classification Accuracy with the AML (MSEG, Sc:100)				
Vegetation	15603	285	0	1898
Tile Roofs	19	5454	9	436
Bright Roofs	0	17	8355	2055
Asphalt Like	365	1496	358	34140
Overall Accuracy: 90,16%				

Best Classification Result with the AML and MSEG at Sc:2500				
Vegetation	15721	16	0	2049
Tile Roofs	61	5450	19	388
Bright Roofs	0	2	8350	2075
Asphalt Like	120	217	317	35705
Overall Accuracy: 92,53%				

Table 5.2: Quantitative results regarding the classification accuracy for the high spatial resolution airborne multispectral TOPOSYS dataset. The developed algorithm scored better in all cases, indicating that the combination of the advanced scale space representation (AML) and the kernel classifier (SVM) outperform earlier approaches.

Chapter 6

Multiscale segmentation and classification of remote sensing imagery with advanced edge and scale-space features

6.1 Introduction

The current need for automated image analysis and computer vision technological tools requires a processing scheme able to encapsulate effectively the content of remote sensing data. However, earth's landscape structure is complex, the context varies and so does the appearance of the images being a combination of many different intensities, representing natural features such as vegetation, geomorphological and hydrological features, man-made objects (e.g buildings, roads) and artefacts caused by variation in illumination of the terrain (e.g shadows).

Furthermore, roads, infrastructure, vegetation, landforms and other land features appear in different sizes and geographical scales in images (e.g. country road versus interstate, tree stands versus forest, maisonette versus polygon building, rill versus river, etc). Only in a few special circumstances the objects of interest belong to a certain scale while the remaining ones, to be discarded, to another. In most cases such a global scale threshold is not possible since the desired information is present at several scales [Lindeberg,

1994; Meyer and Maragos, 2000a; Weickert, 1998; Witkin, 1983].

Scale space representations and multiscale image analysis provide the framework to explore the entire image content by detecting the scale(s) at which objects or patterns appear and are most distinctly identified. Towards this end and parallel to the human visual system, several multiscale low level processes (i.e. filtering, segmentation, etc) have been developed during which a series of representations of the same image are computed (from fine to coarse) and used for the recognition of earth surface objects [Benz et al., 2004; Blaschke and Hay, 2001; Duarte Carvajalino et al., 2008; Hall and Hay, 2003; Hay et al., 2002; Jimenez et al., 2005; Karantzalos and Argialas, 2006; Ouma et al., 2008; Stewart et al., 2004]. The mathematical models and the manner for constructing these scale space representations is of fundamental importance.

In addition, during the last decade the way of classifying remotely sensed imagery is been changing and instead of classifying individual pixels into discrete land cover classes, object-based classification approaches construct a hierarchical object representation of an image and the classifier is responsible for associating them with a land cover class [Blaschke, 2010]. Therefore, it is not just the spectral signature of each pixel, but the statistical, geometric and topological characteristics of each object that play a key role during classification. Object-based classification is considered optimal for the analysis of very high resolution remote sensing imagery (with spatial resolution of 5m per pixel or less), since this kind of imagery is introducing more complexity to the classification tasks, due to increased heterogeneity and the increased number of land-cover classes that can be observed. The statistical, geometric and contextual characteristics of the image primitives are considered by the object-based methods (much like in photointerpretation), in contrast to pixel-based approaches [Blaschke, 2010]. Recent studies are highlighting that the determination of one or more optimal filtering scales for image segmentation is still a challenge and that a multiscale object-based classification is a significantly better approach than the classical per-pixel classification procedure [Myint et al., 2011; Tzotsos et al., 2011].

In this chapter, an object-based image analysis framework was developed which integrates advanced scale space representations, edge and line feature detection, multiscale segmentation and a kernel-based classification. The contributions of this approach are twofold:

- a generic framework able to process any remotely acquired raster data (satellite/airborne

data, multispectral/ hyperspectral data and radar data of any spatial resolution) without the need of tuning any parameter (scale, color, texture, etc) and

- a new robust multi-scale segmentation procedure (replacing an earlier segmentation algorithm in the framework) which is constrained by advanced edge-based features.

The paper is structured as follows. In Section 2, the related work on scale space representations, multiscale object-based analysis and edge-based image segmentation is briefly presented. The developed object-based image analysis framework is detailed in Section 3, along with a description and detailed analysis of its different processing steps. Experimental results and the performed quantitative evaluation are given in Section 4. Finally, conclusions and perspectives for future work are in Section 5.

6.2 Related Work

6.2.1 Multiscale object-based image analysis

Along with the gradual availability of earth observation data with higher spatial and spectral resolution, research efforts in classifying remote sensing data have been shifting in the last decade from pixel-based approaches to object-based ones [Blaschke, 2010; Myint et al., 2011; Tzotsos et al., 2011]. Assigning land cover classes to individual pixels can be intuitively proper and functional for low resolution data. However, this is not the case for the emerging applications which arise from the continuously improving remote sensing sensors [Aplin and Smith, 2008; Blaschke et al., 2008]. This is mostly because at higher resolutions it is a connected group of pixels that is likely to be associated with a land cover class and not just an individual pixel [Tzotsos et al., 2011].

In addition, the earth surface exhibits various regular and irregular structures which are represented with a certain spatial heterogeneity in images composing their intensity, scale and texture. Several important aspects of earth surface can not be analyzed based on pixel information, but can only be exploited based on contextual information and the topologic relations of the objects of interest [Liu et al., 2008] through a multiscale image analysis [Benz et al., 2004; Blaschke and Hay, 2001; Dragut et al., 2010; Duarte Carvajalino et al., 2008; Hall and Hay, 2003; Hay et al., 2002; Jimenez et al., 2005; Ouma et al., 2008; Stewart et al., 2004; Tzotsos et al., 2011]. Starting with the observed spatial hetero-

geneity and variability, meaningful spatial aggregations (primitive object) can be formed at certain image scales configuring a relationship between ground objects and image objects. Ground objects refer to real-world objects or areas of specific land-cover class, connected by complex spatial and contextual relations. Image objects, on the other hand, refer to knowledge-free areas (primitive objects) of an image that a segmentation algorithm provides, based on various criteria. In order to configure a relationship between ground and image objects, a multiscale knowledge representation is needed, which is usually provided by the object-oriented paradigm. With such an object-based multiscale representation and analysis, which is based on certain hierarchically structured rules, the relationship between the different scales of the spatial entities is described.

During the last decade, a number of object-based image analysis software was developed in the form of proprietary software (eCognition, ENVI FX, ERDAS Objective) [Baatz and Schape, 2000] or in the form of free software (Orfeo Toolbox, EDISON, MSEG) [Christophe and Inglada, 2009; Inglada and Christophe, 2009; Tzotsos and Argialas, 2006; **?**] enabling the broad application on various engineering and environmental remote sensing studies [Benz et al., 2004; Blaschke, 2010; Dragut et al., 2009; Mladinich et al., 2010; Zhou et al., 2009]. In all cases, the challenge was to construct an efficient scale-space object representation through certain multiscale (region merging or other) segmentation techniques [Blaschke et al., 2004; Carleer et al., 2005; Jimenez et al., 2005; Neubert et al., 2006; Tzotsos and Argialas, 2006], which partition the image on several regions/objects, based on the spectral homogeneity in a local neighborhood. In addition to the spectral homogeneity criterion, shape parameters are used to define geometric properties that the segmentation algorithm must take into account when computing the overall homogeneity (scale parameter) of each image object during search for optimal merges.

In 2008, a texture optimization procedure was introduced for the MSEG algorithm [Tzotsos et al., 2008] integrating grey level co-occurrence matrices and introducing an object-based cost measure for texture homogeneity as an additional parameter to the segmentation procedure. Such an integration of spatial and spectral information can produce a multiscale object representation but only through an iterative and exhaustive tuning (based on trial and error investigation) of certain parameters, like shape, scale, texture, etc. [Baatz and Schape, 2000; Benz et al., 2004; Blaschke et al., 2004; Carleer et al., 2005; Dragut et al., 2009; Hay et al., 2005; Ouma et al., 2008; Tzotsos and Argialas,

2006; Zhou et al., 2009].

Other research efforts were based on the construction of linear scale spaces for the multiscale analysis of several landscape structures [Blaschke and Hay, 2001; Hay et al., 2002, 2003; Stewart et al., 2004] or on the construction of multiscale representations through object-specific analysis and up-scaling, through the computation of a number of coarse and fine scales by sampling the initial image [Hall and Hay, 2003]. Furthermore, other studies employed unsupervised classification algorithms for both optical and radar data [Derrode and Mercier, 2007; Jung, 2007] or multiple hierarchical segmentations [Akcay and Aksoy, 2008].

More recent research efforts are focusing on optimizing the segmentation procedure through a data-driven thresholding approach [Martha et al., 2011] and on constructing advanced non-linear scale space representations for efficient supervised classification [Tzotsos et al., 2011] and change detection over urban areas [Doxani et al., 2012].

6.2.2 Scale space remote sensing data representations

Earth surface objects cannot be represented to a single scale but rather to many. The use of scale space image representations is thus of fundamental importance for a number of image analysis and computer vision tasks. It dates back to sixties and was first introduced by Iijima [Weickert et al., 1999]. In western literature many linear scale-space methods were introduced [Koenderink, 1984; Lindeberg, 1994; Witkin, 1983], and respectively many isotropic multiscale operators were developed. Either through Gaussian filtering or through isotropic multi-resolution analysis (by down-sampling the initial data), all linear scale space approaches present the same important drawback: image edges are blurred and new non-semantic objects may appear at coarse scales [Ouma et al., 2008; Paragios et al., 2005; Witkin, 1983]. Under a hierarchical multiscale segmentation or an object-based classification framework, the thematic information to be extracted is directly related with the primitive image objects computed at every scale. The better these primitive objects represent real-world entities, the better they can describe the semantics of the image [Blaschke et al., 2008; Hay and Castilla, 2006; Hofmann et al., 2008; Tzotsos et al., 2011]. Therefore, the selection of the appropriate approach for constructing the multiscale image and hierarchical object representation is of great importance.

Since linear scale space approaches, by acting isotropically in the image domain, de-

localize and blur image edges, non-linear operators and non-linear scale spaces have been studied and applied in various image processing and computer vision applications. Following the pioneering work of Perona and Malik [1990] there has been a flurry of activity in partial differential equation and anisotropic diffusion filtering techniques [Weickert, 1998]. For remote sensing applications, a number of anisotropic diffusion schemes have been proposed and applied to aerial and satellite datasets [Camps-valls and Bruzzone, 2005; Duarte-Carvajalino et al., 2007; Karantzalos and Argialas, 2006; Lennon et al., 2002; Ouma et al., 2008; Plaza et al., 2009], combined, in most cases, with pixel-based classification techniques. All their scale space formulations, though, were based either on diffusions during which the average luminance value is preserved or on geometrically driven ones formulated under a variational framework. Although these formulations may reduce the problems of isotropic filtering, they do not eliminate them completely: spurious extrema and important intensity shifts may still appear [Karantzalos et al., 2007; Meyer and Maragos, 2000a; Tzotsos et al., 2011].

Therefore, another way to produce non-linear scale spaces is through mathematical morphology and, in particular, with morphological levelings, which have been introduced by Meyer [1998b] and further studied by Matheron [1997] and Serra [2000]. Morphological levelings overcome the drawback of spurious extrema or important intensity shifts and possess a number of desired properties for the construction of elegant scale space representations. Levelings, which are a general class of self-dual morphological operators, do not displace contours through scales and are characterized by a number of desirable properties for the construction of non-linear scale space representations. They satisfy the following spatial and spectral properties/axioms [Karantzalos et al., 2007; Meyer, 2004; Meyer and Maragos, 2000a; Tzotsos et al., 2011]:

- invariance by spatial translation,
- isotropy, invariance by rotation,
- invariance to a change of illumination,
- the causality principle,
- the maximum principle, excluding the extreme case where g is completely flat.

In addition, levelings:

- do not produce new extrema at larger scales,
- enlarge smooth zones,
- they also create new smooth zones
- they are particularly robust (strong morphological filters)
- they do not displace edges

Designing and formulating an optimal scale space framework in still an active area of research. Recent efforts include studies on certain scale space formulation [Nilufar et al., 2012; Ouzounis et al., 2012], studies on a varying stopping time [Gilboa, 2008] and studies on the behavior on corner and other local descriptors [Jiang et al., 2011; Kimmel et al., 2011; Xu et al., 2012; Zhong et al., 2009].

6.2.3 Edge-based segmentation

Extracting primitives (i.e. contours, edges, lines, etc) is a basic low-level operation in the human visual system. Along with the aim to understand and simulate the human vision, the importance of building up computational models for the perception of primitives is a major component in many applications of computer vision, such as object/pattern recognition, robot vision, remote sensing and medical image analysis.

These primitives give important information about the geometric content of images. Most of the earth surface objects and in particular most man-made objects are made of flat surfaces with certain geometric features. In addition, many shapes can be described roughly or in detail with edge and line primitives. Therefore, edge or line segments can be used as a low-level feature description in order to extract information from images and can serve as the basic tool to analyze and detect more complex shapes [Chia et al., 2012; Papari and Petkov, 2011; Von Gioi et al., 2010; Wang and Oliensis, 2010]. In the context of scale space representations, image primitives may support more stable and efficient representations since their description can be independent of the object size. Edge and line primitives, defined mainly by the object geometrical properties, allow the robust and

efficient feature comparison in various scales. The later is of major importance due to large appearance variations of object instances belonging to the same class.

However, even the recent more sophisticated edge and line detectors cannot produce connected segments and suffer from the earth surface complexity pictured in images, shadows, occlusions, etc. Therefore, recent efforts are trying to merge the advantages of edge/line detection and image segmentation techniques in order to produce connected object contours/boundaries and a comprehensive object description [Cufi et al., 2003; Kermad and Chehdi, 2002; Pavlidis and Liow, 1990]. Certain primitive combinations have been proposed in order to describe more efficiently object boundaries [Chia et al., 2012; Klonus et al., 2012]. Another recent study proposed a region-based unsupervised segmentation and classification algorithm which included the computation of an edge strength model [Yu et al., 2012]. This edge penalty model improved segmentation performance by preserving segment boundaries.

6.3 Methodology

The main objective, here, was to design the overall framework in order to be generic, robust and able to process effectively a wide variety of remote sensing data, such as hyperspectral and multispectral data from ground, aerial and spaceborne sensors, radar data, digital elevation models, etc. It is based upon the Object-Based Image Analysis (OBIA) approach, which generally includes low, medium and high level image processing subtasks:

- Preprocessing steps (geometric and radiometric corrections, filtering, scale space image simplification, edge detection, band math expression computations, etc.),
- Image Segmentation (in order to produce single-level or multi-level hierarchies of primitive objects within the image space),
- Computation of image object properties based on spectral, shape, topological and context features (feature space),
- Definition of object-oriented class hierarchy and representation of knowledge through rules or through training with samples,

- Classification (learning techniques or rule-based systems to perform the classification task),
- Accuracy assessment in order to derive the quality of the resulting classification,
- Vectorization steps (create the output to spatial databases and integrate the information to thematic maps).

The developed approach is integrating certain advanced computer vision and machine learning methods for implementing the above tasks. Although the proposed methodology is following the above general OBIA approach, which is also available in several proprietary and free software solutions, the originality of the developed framework lies in the novel approach that excludes the parameter tuning step, in the robustness of the multi-scale hybrid (edge-based and region-based) segmentation algorithm, in the extensive use of non-linear scale-space representations for image simplification purposes and in the integration with kernel-based machine learning methods for classification.

Briefly the developed framework consists of the following steps: Firstly, for every band of the initial image, a scale-space representation is generated, using the Anisotropic Morphological Leveling (AML) formulation [Karantzalos et al., 2007]. The supported type of the imagery can be up to double precision and of any number of bands. A feature extraction step is then applied on the scale-space stack. For this step two algorithms were tested and are presented here: the Canny edge detector [Canny, 1986] and the Line Segment Detector (LSD) [Von Gioi et al., 2010]. A multi-scale segmentation algorithm is applied afterwards which is able to integrate the simplified scale space stack along with the corresponding edge information. During this step primitive image objects are formed. The procedure starts from single-pixel, and through pairwise merges bounded by edge information, several levels of image objects are produced. The multiscale object hierarchies is been constructed without any parameter tuning. In a similar way with [Tzotsos et al., 2011], here the edge features are produced without tuning the edge extraction parameters (Fig. 1). Last in the processing order comes a dual classification procedure using a support vector machine classifier. The first classification is performed on all scale-space representations and their corresponding segmentations, while the second optimal one is performed after an interim accuracy assessment.

6.3.1 Scale-Space Filtering

The first step in the developed approach is the construction of the non-linear scale-space representation in order to elegantly simplify raw data. Anisotropic diffusion methods are used widely in computer vision applications to simulate the filtering procedures that are performed in the human vision system. Such methods provide robust simplification of images without the loss of important information such as edges, that are of high importance for higher level processing algorithms. Especially in OBIA applications, where very high resolution data are usually processed, it is very important to simplify the complexity of the initial data and provide a multiscale representation, since different features of the image reside in different scales.

For this preproccesing step, the Anisotropic Morphological Levelings (AML) [Karantzalos et al., 2007] were incorporated in the processing scheme. Anisotropic Morphological Levelings are a combination of morphological levelings with anisotropic markers and are employed in order to achieve better segmentation results, reduce the heterogeneity of image data, reduce over-segmentation and create accurate image objects. Fig.6.1 shows that the scale space filtering by creating a series of simplified data leads to a multi-scale segmentation (without tuning any segmentation parameters like texture, color, shape, etc).

Starting from the initial image and for every available band, a scale-space representation was generated, using the AML formulation. Using iterative anisotropic morphological operations, and increasing scales (10,50,100,500,1000) a scale-space cube (3D) representation was constructed from each initial band. The result of this step was a scale-space stack with simplified versions of the raw data. Note that during this process edge information was preserved in all scales, contrary to isotropic (e.g. Gaussian) filtering that loses edge information as scale increases (Fig.6.1 f,g,k,l,p,q). In the following sections further analysis of Figure 6.1 will follow.

6.3.2 Multiscale Segmentation based on Advanced Edge Features

Edge and line features were computed for every image in the scale space stack. Edge information was obtained from the standard Canny detector [Canny, 1986] and the recent Line Segment Detector (LSD) [Von Gioi et al., 2010, 2012].

The Canny edge detector was employed in order to provide primitive edge features

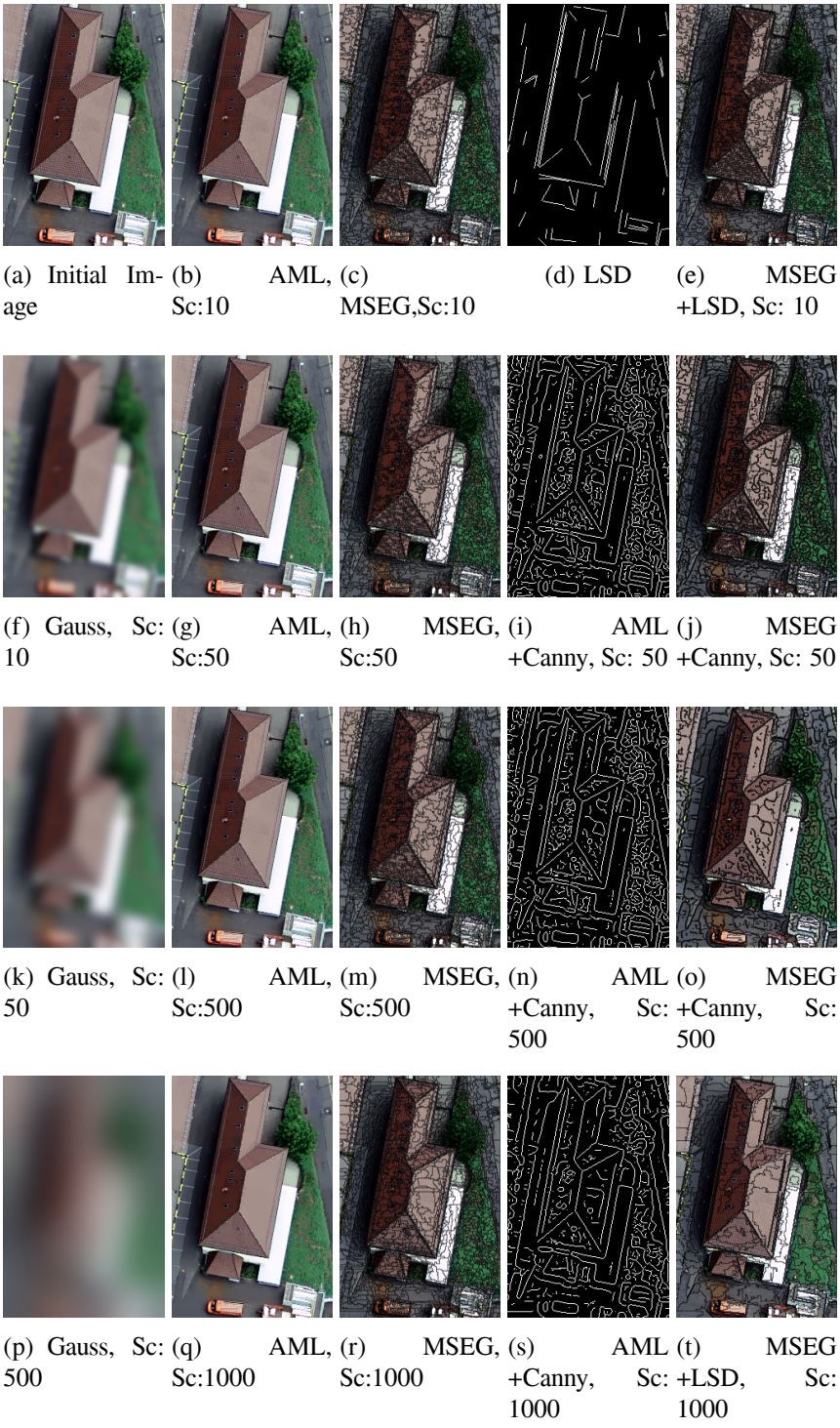
that were integrated to the implemented region merging algorithm. Throughout this research, the variance parameter of the Canny detector was set stable to 5. In Fig.6.1(e,n,s) results from the application of the Canny edge detection on scale-space images is shown. LSD is a linear-time Line Segment Detector giving subpixel accurate results. The LSD algorithm starts by computing the level-line angle at each pixel to produce a level-line field, i.e., a unit vector field such that all vectors are tangent to the level line going through their base point. Then, this field is segmented into connected regions of pixels that share the same level-line angle up to a certain tolerance. These connected regions are called line support regions [Von Gioi et al., 2012]. Each line support region (a set of pixels) is a candidate for a line segment. The principal inertial axis of the line support region is used as main rectangle direction. After examining and validating line support regions, and test that they are aligned properly, a selection of meaningful rectangles is provided as the final result. In Fig.6.1d the application of LSD on a very high resolution aerial scanner image is demonstrated. LSD has been designed to be automated and includes an internal filtering and simplification procedure with constant scale of 0.8.

For the multi-scale segmentation procedure, an improved version of the initial MSEG algorithm [Tzotsos and Argialas, 2006] was implemented. The initial MSEG is a region-based multi-scale segmentation algorithm recently developed for object-oriented image analysis. Briefly, starting from a pixel representation it creates objects through continuous pair-wise object fusions, executed in iterations (passes). For each pass, every object is evaluated in relation with its neighboring objects towards the optimal pair of objects adequate for fusion. In every pass, an image object can be merged only once, aiming at a balanced object growth. MSEG algorithm defines a cost function for each object merge and then implements various optimization techniques to minimize this cost. The cost function is implemented using the measure of homogeneity (color and shape) in the same way with other approaches [Baatz and Schape, 2000]. The threshold of the allowed merging cost for the segmentation procedure is called scale parameter, since it implicitly dictates the area growth of the image objects. Results from the application of the MSEG algorithm are shown in Fig.6.1(c,h,m,r). Through this research, the parameters of the MSEG algorithm were set stable, the color parameter was set to 0.8 and the shape parameter was set to 0.2. The goal was to allow the simplified data (from the scale space stack) to control the way that image segments and objects are being created and not the region merging procedure. In a previous study [Tzotsos et al., 2011] is was shown that

there is no need for tuning the segmentation parameters when the approach includes a reliable edge-preserving formulation for the scale space computation.

The MSEG algorithm was improved in order to be able to integrate edge information (as a constrain) during the segmentation procedure. The goal was to design a more robust and generic segmentation procedure that would be able to take into account advanced edge and line features. In particular, the region merging algorithm starts by selecting initialization points throughout the image using the SPE (Start Point Estimation) module [Tzotsos and Argialas, 2006] and a queue of pixels is created in order to be able to achieve reproducibility. Then, iterative pairwise fusions occur within the image space, starting from single pixel objects, in a way that local heterogeneity is minimized (for color and shape criteria). During this pairwise merging of image objects, the edge information is used as a boundary. Two adjacent pixels will not be merged into an object if one or both reside on top of an edge. After the first pass of the region merging procedure, image objects of one or two pixels exist, with edge pixels being constrained and not merged to each other. During the following passes edge objects (still single pixels) are not merged, thus not permitting object merging between image regions that are separated by a line or a continuous edge feature. After several passes (iterations) converging of the algorithm occurs and no more object merging is performed, due to scale parameter. At this point the edge objects are still intact by the region merging procedure, thus binding the procedure into respecting edge features. Finally a last iteration of the algorithm is forced on edge objects only, and a selection is made, to which neighboring object they should be merged, based on local heterogeneity. This step is taking advantage to the fact that both Canny and LSD features are one-pixel wide, thus edge objects are always capable of merging with non edge objects.

A certain novelty of the developed segmentation process is that it does not use an edge penalty model for the edge compensation as presented in other approaches [Cufi et al., 2003; Kermad and Chehdi, 2002; Yu et al., 2012], but a topological constraint, effective throughout the region merging procedure. The results of this enhancement is presented in Fig.6.1(e,j,o,t) showing very promising segmentation results. A scale parameter of value 100 was used for all tests on Fig.6.1, showing that scale-space representation effectively provides the scale of the obtained objects. More results are presented and discussed in following sections.



(a) Initial Image (b) AML, Sc:10 (c) MSEG,Sc:10 (d) LSD (e) MSEG +LSD, Sc: 10

(f) Gauss, Sc: 10 (g) AML, Sc:50 (h) MSEG, Sc:50 (i) AML +Canny, Sc: 50 (j) MSEG +Canny, Sc: 50

(k) Gauss, Sc: 50 (l) AML, Sc:500 (m) MSEG, Sc:500 (n) AML +Canny, Sc: 500 (o) MSEG +Canny, Sc: 500

(p) Gauss, Sc: 500 (q) AML, Sc:1000 (r) MSEG, Sc:1000 (s) AML +Canny, Sc: 1000 (t) MSEG +LSD, Sc: 1000

Figure 6.1: Comparing region merging segmentation results using scale space representations and advanced edge features. Four scales from various steps of the developed methodology are presented. The first column (a,f,k,p) presents the initial aerial image with spatial resolution of 5cm, along with three gaussian scales. The second column (b,g,l,q) presents the scale-space representation at various selected scales. The third column (c,h,m,r) presents initial image objects using the MSEG algorithm applied to the scale-space representation. The forth column (d, i ,n ,s) presents the extracted edge and line features used in the following step. The final column (e,j,o,t) shows the results of the edge enhanced MSEG algorithm developed in this research.

6.3.3 Kernel-based Classification

For the developed approach, an SVM classification scheme [Tzotsos, 2006; Tzotsos and Argialas, 2008; Tzotsos et al., 2011; Vapnik, 1998] was employed. After the multiscale segmentation which is constrained by edge information, image objects were extracted and object properties were computed forming the feature space of the classification step. For each primitive image object, spectral, shape and spatial properties (such as mean band values, standard deviation, GLCM texture features, area, perimeter, compactness, number of neighbors etc) were extracted by the topological model used to handle object topology. This model was proposed by [Lehmann, 2008] but also developed independently in MSEG [Tzotsos and Argialas, 2006]. The computed properties are bound to each object by a unique identifier within the object hierarchy of the image. Some of the objects are selected as samples and their properties formed a training set for the SVM [Tzotsos and Argialas, 2008].

The SVM classifier seeks to find the optimal separating hyperplane between classes by focusing on the training data (support vectors) that are placed at the edge of the class descriptors. Training data other than support vectors are discarded. Thus, not only an optimal hyperplane is fitted but less training samples are effectively used as well [Tzotsos and Argialas, 2008]. This method works very well for classes that are linearly separable. In the case that image classes are not linearly separable, the SVM maps the feature space into a higher dimensionality using kernels [Theodoridis and Koutroumbas, 2003; Vapnik, 1998] and then separates classes in that new feature space forming the support vectors.

Since the SVM classification method was initially designed for binary classification problems, a heuristic one-against-one strategy was employed for multiclass classification [Hsu and Lin, 2002]. Many binary classifiers were applied for each pair of classes and for every object of the image and then a max-win operator determined the final classification of the object. A n-fold cross validation scheme was also used in order to define the parameters needed for the training procedure. After specification of parameters, the training set was used to train the classifier. Primitive image object were classified using this trained SVM algorithm using the one-against-one strategy. Finally a quality assessment took place, using ground truth data, that were not used during the training procedure.

The above classification procedure was repeated for all scales (values starting from

1 to a maximum value defined by the user) in the scale-space representation in order to determine the best classification accuracy, as proposed in [Tzotsos et al., 2011]. The best scale was then selected based on the best classification accuracy. After determining the best scale to perform classification, a final classification step took place to produce the optimal results.

To sum up, the initial dataset was simplified and a successive series of simplified images were constructed forming a non-linear scale space. The simplified imagery that was derived was then used to extract edge and line features using advanced methods. An edge enhanced multi-scale image segmentation algorithm was employed to provide primitive image objects from the scale space images without the tuning of any standard parameter. Finally, a classification step was performed to complete the OBIA tasks and to evaluate the developed method.

6.4 Evaluation and Discussion

As already stated, the overall objective of the present research was: (a) to introduce a generic and robust framework able to process any kind of remote sensing data without tuning any parameters (like scale, texture, color, etc) during the computation, (b) to introduce a multi-scale segmentation algorithm which is constrained by advanced edge-based features at various scales and (c) to evaluate the developed methodology in various remote sensing datasets.

In Fig.6.1 a general overview of the developed method is presented. Starting from the initial image (Fig.6.1a), Gaussian filtering at different scales demonstrates the loss of edge information due its isotropic character (Fig.6.1 f,k,p). These results are directly compared with the AML scale-space representations at equivalent scales. One can observe in Fig.6.1 (b,g,l and q) that edge information is preserved while the initial image is simplified. For example, in the tile roof of the building the single tiles are more difficult to distinguish as scale increases. The results from the application of the standard MSEG algorithm on the simplified images using the same scale parameter value (100) are shown as well Fig.6.1 (c,h,m and r). The standard MSEG algorithm performs well across object boundaries but produces over-segmented results and the mean object size is increasing along with scale. The edge and line feature extraction at various scales is demonstrated in Fig.6.1, as well. The result from the application of the LSD in the orig-

inal image is shown in Fig.6.1 (d). This is an impressive result, demonstrating that LSD is robust and works well for man-made objects, even if not all of the building sides have been detected correctly. The results of the Canny edge detector are also presented at different scales Fig.6.1 (i,n and s). It can be observed that due to the simplified data through the AML scale space computation, more clear edge features are detected which describe accurately object boundaries. Less false detections have been, also, detected inside homogeneous regions as for example in the tile roof region. Furthermore, results from the application of the improved MSEG algorithm are presented in Fig.6.1 (e,j,o and t). The first result (Fig.6.1e) shows how the developed algorithm is constrained by the detected LSD line features. On the homogeneous regions there is not much difference which is normal since the same AML scale is used for both (c) and (e). The second result (Fig.6.1j) shows how the Canny edges are preserved inside the roof segments and how the improved segmentation method is been robustly constrained by edge information. This result is better than (h), where image objects are oversegmented and arbitrarily set inside a homogeneous region of the image. The third result in Fig1.(o) shows that the combination of edge information with region merging in higher scales is outperforming the standard MSEG algorithm (Fig.6.1m) at the same AML scale and segmentation scale parameter. Moreover, the result in Fig.6.1(t) shows how the scale space in combination with the edge-constrained segmentation tackles the oversegmentation issue shown in Fig.6.1r.

These aforementioned results demonstrate that the developed method outperforms earlier efforts [Baatz and Schape, 2000; Tzotsos and Argialas, 2008; Tzotsos et al., 2011]. In addition, in order to further validate the developed algorithm's experimental results and demonstrate its performance under several type of datasets and settings, a variety of remotely sensed data has been selected with different spatial and spectral characteristics. In the following sub-sections, the developed method was compared against previous research efforts [Tzotsos et al., 2011] and other standard OBIA implementations implemented in Orfeo Toolbox [Inglada and Christophe, 2009] and eCognition [Benz et al., 2004].

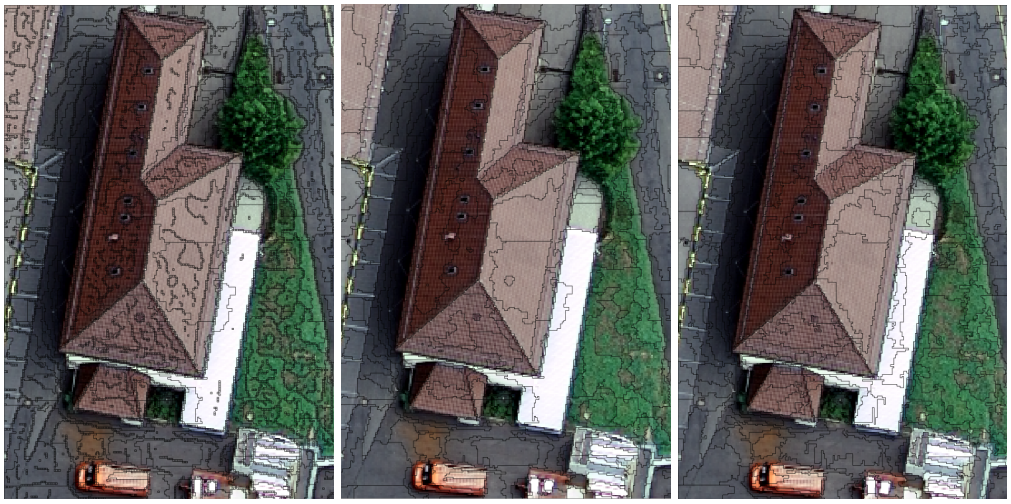
6.4.1 Very high spatial resolution airborne imagery

The developed methodology was applied to a variety of very high and ultra high resolution remote sensing imagery. At first a 5cm resolution image from a DMC airborne digital scanner was tested. This kind of data is practically impossible to handle using traditional pixel-based classification and image analysis approaches. As shown in Fig.6.2 it is possible to segment this image into primitive objects in order to construct a feature space for OBIA classification. In this figure, a comparison of various segmentation methods is performed. Initially, the standard MSEG segmentation algorithm is tested in (Fig.6.2 a) at a scale of 100. The MSEG algorithm is applied on a scale-space AML representation and the result is achieved without any parameter tuning.

In Fig.6.2(d) results from the application of the developed edge-constrained MSEG algorithm are demonstrated. The edge objects are not merged to the rest of the image objects and they remain unmerged until a final step concludes the segmentation procedure and produces the result in Fig.6.2e. A comparison of the developed algorithm with the Mean-Shift algorithm [Comaniciu and Meer, 2002] (Fig.6.2 b) shows that while in some image regions Mean Shift can merge large parts of the image into one object, it fails to do so in other areas of the same texture. This behavior is not optimal and can lead to problems for classification steps since the mean object size varies. On the other hand multiresolution segmentation, as provided by eCognition (Fig.6.2 c), manages to segment the image with a homogeneous object size, but it suffers from over-segmentation problems, especially on the roof objects. For both algorithms the default values were used to avoid parameter tuning. The developed method on the other side (Fig.6.2 e,f), manages to obtain similar objects in size, which can be very applicable in a multiscale OBIA approach to classification. The main difference between (Fig.6.2e) and (Fig.6.2f) is that image edges are larger in number as derived from the Canny algorithm and this leads to better results, in areas that LSD (Fig.6.2f) has not detected any straight lines in the image. So the LSD method is not very suitable for curved or small edges.

In order to evaluate the developed method and to show the advantages of edge-constrained segmentation algorithm, a test similar to the one performed in [Tzotsos et al., 2011] was deployed. For this test, a very high resolution aerial scanner image, was used with 4 spectral bands, in order to perform full scale Object-Based Image Analysis tests. The initial image was segmented using a simple MSEG algorithm, without parameter

(a) Standard MSEG (b) Mean-Shift (c) eCognition



(d) Edge MSEG, no merge of edge objects (e) Edge MSEG Canny (f) Edge MSEG LSD

Figure 6.2: Comparing various segmentation algorithms on a DMC aerial multispectral image with 5 cm pixel-size (©Intergraph Corp.). (a) Standard MSEG with scale parameter 100. (b) Mean-Shift segmentation with default parameters. (c) Multiresolution segmentation (eCognition) with default parameters. (d) Edge constrained MSEG, without merging the edge objects in last pass, for demonstration purposes (e) Edge constrained MSEG, employing the Canny edge features. (f) Edge constrained MSEG, employing the LSD edge features.

tuning (default values of scale parameter 100, color 0.8 and shape 0.2 were selected). After primitive objects were obtained, a training set was given to a kernel-based classifier (SVM) to perform learning, based on the feature space introduced by object spectral and shape properties. For this test four generic land cover classes were used: Vegetation, Tile Roofs, Bright Roofs and Asphalt like materials. A set of training samples/objects was introduced to the SVM and a classification was performed. Using ground truth data, a quantitative evaluation was performed and a confusion matrix is presented in Table 6.1. The accuracy of the object-based classification was 88.07% similar to the results reported for this approach in [Tzotsos and Argialas, 2008].

A similar approach was then followed for the same image, with the same segmentation parameters and the same training and testing samples. This time, a scale-space AML representation was used to provide anisotropic diffusion and simplification of the initial dataset. After the SVM classification and evaluation of results (Table 6.1) an overall accuracy of 89.29% was achieved, similar to the accuracy reported in [Tzotsos et al., 2011].

Finally the developed method was similarly applied to the same data. A scale-space AML method was used to simplify the initial dataset. A segmentation step was then performed using the edge-constrained MSEG algorithm, and specifically the Canny edge features were used. A set of primitive objects was obtained and object properties were extracted (spectral and shape features). The same training test was given to the SVM classifier and a final classification of objects was obtained. As shown in Table 6.1, the overall accuracy of the developed method, outperformed the previous tests with an accuracy of 90.29%. This shows that edge features helped the segmentation procedure to obtain more meaningful objects, that are capable of providing very good classification results. This procedure was repeated again with some different parameters and similar results were produced. Of course the difference in accuracy is not significant, but it shows that compatible results can be produced.

6.4.2 Radar satellite imagery

Experimental results includes the application of the developed methodology at high resolution SAR data (TerraSAR-X dataset). The initial SAR image is shown in Fig.6.3 (a) and the output result from the edge-constrained segmentation are compared with the

Classification Accuracy with MSEG only				
	Vegetation	Tile Roofs	Bright Roofs	Asphalt Like
Vegetation	15247	0	0	2539
Tile Roofs	198	2856	15	2849
Bright Roofs	0	1	8362	2064
Asphalt Like	215	34	498	35612
Overall Accuracy: 88.07%				

Classification Accuracy with AML				
Vegetation	15523	0	0	2263
Tile Roofs	15	3764	124	2015
Bright Roofs	0	0	8389	2038
Asphalt Like	583	30	482	35264
Overall Accuracy: 89.29%				

Classification Accuracy with AML and Edge enhancement				
	Vegetation	Tile Roofs	Bright Roofs	Asphalt Like
Vegetation	15791	34	0	1961
Tile Roofs	244	4710	45	919
Bright Roofs	0	0	8311	2116
Asphalt Like	141	909	475	34834
Overall Accuracy: 90.29%				

Table 6.1: Quantitative results regarding the classification accuracy for a high spatial resolution airborne multispectral dataset. The developed OBIA methodology scored better, indicating that the enhancement of MSEG with advanced edge features along with advanced scale space representations (AML) and the kernel classifier (SVM) outperforms earlier approaches.

ones from the Mean-Shift and Watershed Fig.6.3 (b and c). The Mean-Shift method did not perform well, and resulted in under-segmentation around the shore line (Fig.6.3b). Even if the water area was successfully segmented into one image object, the under-segmentation is always a very poor image segmentation performance. The application of the Watershed algorithm (Fig.6.3c), on the other hand, resulted in serious over-segmentation as can be seen in Figure 6.3(c).

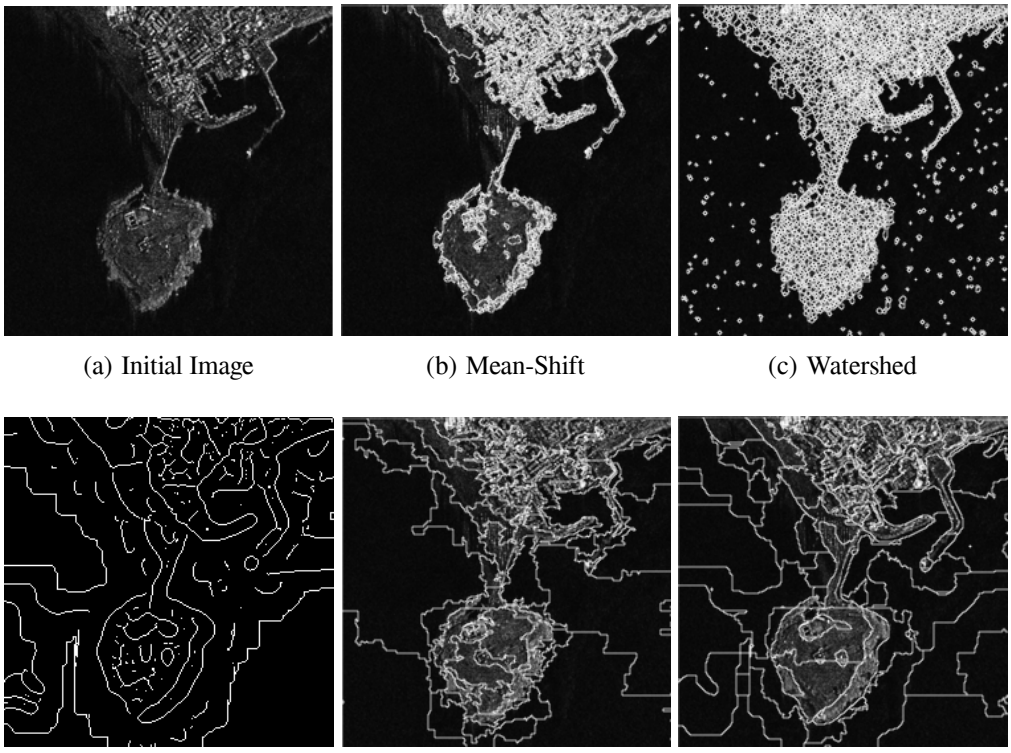
A better result was obtained with the application of the standard MSEG algorithm, although there were some problems in objects around the shore line (Fig.6.3e). For this reason a Canny edge feature extraction was performed (Fig.6.3d) and the results were imported to the edge-constrained segmentation algorithm, which outperformed all other segmentation algorithms (Fig.6.3f). The developed algorithm managed to obtain image objects of similar scale and due to the imposed edge information the output boundaries described clearly the rather more compact image objects. Again no parameter tuning was performed and default value of scale parameter 100 was used.

6.4.3 Multispectral remote sensing data

The next series of tests were performed on medium and high spatial resolution multi-spectral remote sensing data. For this, a Landsat TM image with pixel spatial resolution of 30m was used, as well as an QuickBird satellite image with 1m ground resolution and four spectral bands.

For the Landsat TM imagery, the same comparison of image segmentation methods was performed and is presented in Fig.6.4. In this situation the image had a horizontal stripping noise problem, making it more difficult for the segmentation algorithms to perform well. The application of MSEG with scale parameter 100 resulted in major over-segmentation, but still the algorithm resulted in objects of similar size and scale. The stripes of the image are obvious in this segmentation result (Fig.6.4a).

The Mean-Shift segmentation algorithm performs much better in this specific test (Fig.6.4b), since strong simplification of the image is involved internally, making the algorithm more robust in noise presence. On the other hand, still the size of image objects is variant across the image, even for the same semantic objects/areas. Multiresolution segmentation (eCognition) produced good result especially on the parcel area. Some problems were observed with objects allocated on the river sides, where oversegmenta-

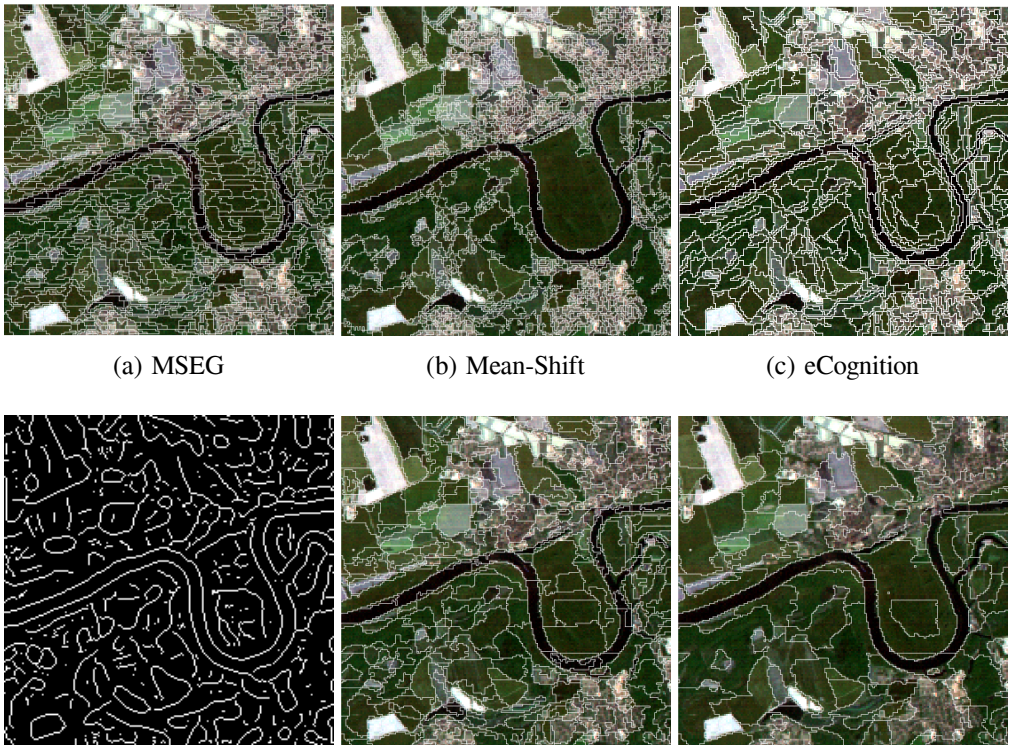


(a) Initial Image (b) Mean-Shift (c) Watershed

(d) Canny on AML (e) MSEG (f) Edge MSEG with Canny

Figure 6.3: Comparing various segmentation algorithms on a TerraSAR-X (©DLR) dataset (3 meters ground resolution, StripMap mode, polarisation HH). (a) The initial image. (b) Mean-Shift segmentation with default parameters. (c) Watershed segmentation with default parameters. (d) Canny edge detection applied on the AML scale-space representation. (e) Standard MSEG results with scale parameter 400. (f) Edge constrained MSEG, employing the Canny edge features.

tion occurred (Fig.6.4c). A Canny edge detection step (Fig.6.4d) was introduced and the edge-constrained segmentation was tested (Fig.6.4e). One can observe that the later produced much better results than the standard MSEG algorithm. Object boundaries are more clear and compact, while the mean size of the primitive objects is approximately the same across the image. A test was also performed with an increased scale parameter (Fig.6.4f). The developed edge-constrained segmentation performed even better at larger scales, while the stripping problem was less apparent (Fig.6.4f).



(a) MSEG (b) Mean-Shift (c) eCognition

(d) Canny on AML (e) Edge MSEG sc:100 (f) Edge MSEG sc:400

Figure 6.4: Comparing various segmentation algorithms on a Landsat TM dataset (Dessau, Germany). (a) Standard MSEG results with scale parameter 100. (b) Mean-Shift segmentation with default parameters. (c) Multiresolution segmentation (eCognition) with default parameters (scale 10, shape 0.1). (d) Canny edge detection applied on the AML scale-space representation. (e) Edge constrained MSEG, with Canny edge features used and scale parameter 100. (f) Edge constrained MSEG, with Canny edge features used and scale parameter 400.

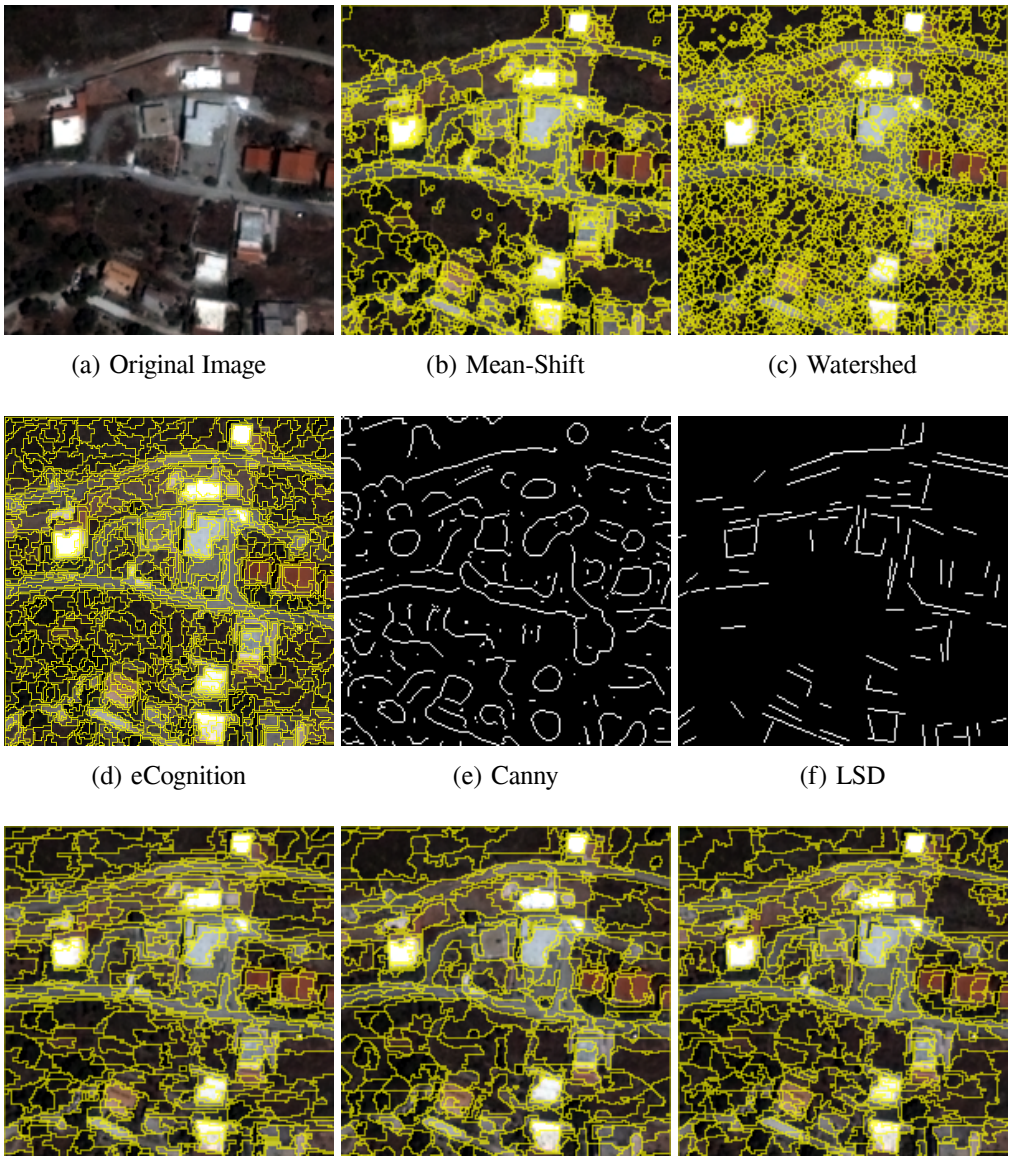
For the QuickBird imagery (Fig.6.5a) a similar test was performed and demonstrated at Fig.6.5 (e,h,f,i) within a semi-urban area. The developed algorithm was able to de-

tect building objects (Fig.6.5 h,i) and in particular when it was constrained by the LSD features (the scale parameter for the simplification was 400). Again similar problems occurred with the Mean-Shift algorithm (Fig.6.5b) obtaining objects at different scales (i.e larger objects in low contrast areas of the image). The Watershed algorithm on the other hand, produced an over-segmentation (Fig.6.5c) but kept all image objects on the same scale. The multiresolution segmentation (eCognition) algorithm provided objects of the same scale (Fig.6.5d), with less over-segmentation problems than Watershed, but did not outperform Mean-Shift and the enhanced MSEG algorithms. Both enhanced MSEG and Mean-Shift had good results in building-like objects with the developed method having a small advantage in preserving the edges of the image semantics.

6.4.4 Hyperspectral remote sensing data

The developed segmentation algorithm was also tested with hyperspectral remote sensing data, obtained by a CASI aerial scanner (Fig.6.6). The spectral resolution of the dataset was 95 bands and the spatial resolution was 5m. Again, mean-shift and watershed algorithms were tested to compare with the developed method, but in this specific test, it was impossible for those algorithms to be applied to the full spectral resolution of the hyperspectral dataset, since both algorithms are not designed to work on a large number of image bands. For this test, a subset of bands were used to derive mean-shift and watershed results (Fig.6.6 b,c). Both those algorithms produced similar results for this test. A mixture of small scaled and large scaled objects were obtained at the same time, with watershed being more accurate in this case, providing better results in edge areas (Fig.6.6c).

On the other hand, MSEG is designed to be applied to images of any spectral resolution, up to 65535 bands. It is demonstrated in Fig.6.6d that the simple MSEG algorithm is performing very well, given that the default scale parameter is easily reached (since it is a heterogeneity threshold) with a big number of bands contributing to object heterogeneity. Therefore, this over-segmentation (Fig.6.6d) cannot be considered as a major problem, rather than an effect caused by the nature of this dataset. After application of a strong simplification AML filtering, the results were improved (Fig.6.6e). In all cases of MSEG application it is shown that MSEG respects the scale of the image objects in a better way that the other algorithms tested. This is very crucial for OBIA applications,



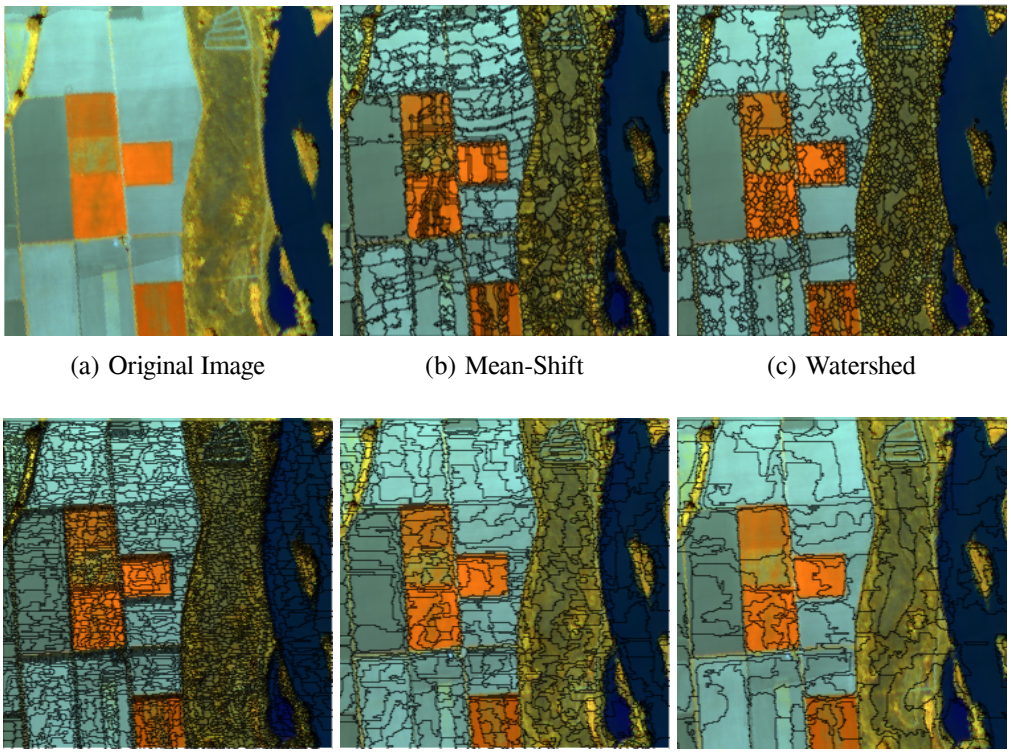
(a) Original Image (b) Mean-Shift (c) Watershed

(d) eCognition (e) Canny (f) LSD

(g) MSEG (h) MSEG Canny (i) MSEG LSD

Figure 6.5: Comparing various segmentation algorithms on a QuickBird dataset (Eastern Attika, Greece). (a) Original image. (b) Mean-Shift segmentation with default parameters. (c) Watershed segmentation with default parameters. (d) Multiresolution segmentation (eCognition) with default parameters (e) Canny edge detection applied on the AML scale-space representation. (f) LSD line features extracted from the original image. (g) Standard MSEG results with scale parameter 100. (h) Edge constrained MSEG, with Canny edge features used and scale parameter 100. (i) Edge constrained MSEG, with LSD line features used and scale parameter 100.

especially when multiscale approaches are necessary. Finally the enhanced MSEG algorithm was tested in Fig.6.6f and the results were obviously better that other approaches shown here. Edge information was preserved (Fig.6.6f) and a similar mean size of image objects was produced on this specific scale.



(a) Original Image (b) Mean-Shift (c) Watershed

(d) MSEG (e) MSEG with AML (f) MSEG with AML, Edge

Figure 6.6: Comparing various segmentation algorithms on a CASI Hyperspectral dataset (©Remote Sensing Laboratory, NTUA) with 95 spectral bands (Axios river, Thessaloniki, Greece). (a) Original image. (b) Mean-Shift segmentation with default parameters. (c) Watershed segmentation with default parameters. (d) Standard MSEG results with scale parameter 900. (e) Standard MSEG on AML scale-space representation and scale parameter 900. (f) Edge constrained MSEG, with Canny edge features used and scale parameter 900.

6.5 Conclusions and Future Perspectives

A new object-based image analysis framework was proposed and developed in this research, based on advanced edge features incorporated in a multiscale region merging algorithm. Advanced scale-space representations were used in order to avoid manual tuning of segmentation and feature extraction parameters. Kernel-based classification was implemented to complete the OBIA framework. The developed image segmentation algorithm was shown to work on any type of remote sensing data, outperforming some widely used segmentation algorithms in some cases. The improvement of the MSEG segmantation results was demonstrated, and the edge enhancements were shown to make the algorithm robust and generic for multiscale OBIA applications. The performed qualitative and quantitative evaluation reported that the developed algorithm outperformed previous efforts, both regarding the construction of object representations and classification results.

In terms of performance, the developed improvement of the MSEG algorithm had a major impact on execution times. For the very high resolution airborne image (Fig.6.1a), execution time for original MSEG algorithm was 7-8 seconds on a 3 GHz PC with 4 GB of RAM running GNU/Linux operating system. With the edge-based optimization, the number of possible object merges droped significantly per segmentation pass, and the execution time was around 4-5 seconds. Still, there is room for speed optimization improvements, since for the same image, the execution times for multiresolution segmentation (eCognition) was 2-3 seconds and for Mean Shift (Orfeo Toolbox) was 3-4 seconds.

An advantage of the developed OBIA framework is that it is composed by free (as in freedom) software. For the implementation of this research a number of Free and Open Source libraries were used (MSEG, cvAML, libSVM, OrfeoToolbox and GDAL). The coding was performed in C++ and Python programming languages. A disadvantage of the developed OBIA framework is that a knowledge-based classification solution is not yet integrated, but currently worked on.

Some of the topics for further research and development are: extension of the developed OBIA framework to integrate knowledge-based classification, solutions for object-specific extraction tasks based on scale-space shape-priors, and adaptation of the developed methodology to specific remote sensing applications.

Chapter 7

Relevance Vector Machines for Object-Based Image Analysis

7.1 Introduction

In recent years, the availability of remote sensing sensors and technologies, along with the advancements in computer science and artificial intelligence has lead into an explosion of size of earth observation data [Baumann, 2010]. At the same time, remotely sensed data become more accurate and detailed, every time a new satellite sensor is launched. From big satellite sensors to smaller airborne sensors, to minicopters and UAVs, big geospatial data are produced every day, increasing the need to automate the processing and the classification of large datasets [Berni et al., 2009]. To that direction, considerable research is targeted in remote sensing methods including state-of-the-art computer vision and artificial intelligence techniques in order to tackle the size and resolution of the produced ``big data''.

The last decade, there is a lot of research in the field of Object-Based Image Analysis (OBIA) and remote sensing. This new methodology moved away from pixel-based methods to object-oriented paradigm [Baatz and Schape, 2000; Benz et al., 2004; Blaschke, 2010], leading to great advancements in classification accuracy and semi-automated extraction of information from remotely sensed data to GIS systems and spatial databases. The OBIA methodology integrates low level image processing methods, such as image segmentation with high level recognition and classification methods such as knowledge-

based systems, fuzzy logic and machine learning.

In the field of remote sensing, classification is one of the most commonly used processes [Foody, 2008]. Research is focused on developing and testing new algorithms in order to provide more accurate results, towards minimizing the need of manual interference when creating a thematic map from remote sensing imagery. State-of-the-art methods are developed for the supervised classification problem, especially in the fields of patern recognition and machine learning [Camps-Valls, 2009]. Recently, many studies regarding Support Vector Machines and kernel-based methods have been evaluated in remote sensing data [Gómez-Chova et al., 2011; Mountrakis et al., 2011], and have also been applied and evaluated within the OBIA framework [Tzotsos and Argialas, 2008]. The SVM approach is shown as superior to alternative approaches, since it produces better generalization and less over-fitting, independently of the dimensionality of the dataset [Foody and Mathur, 2004]. Recently, a probabilistic extension of the SVM, the Relevance Vector Machine has been evaluated for remotely sensed data, showing promising results in comparison to SVM [Camps-Valls et al., 2006, 2007; Foody, 2008].

In this chapter, an object-based image analysis framework was developed integrating a recently developed hybrid edge and region-based multiresolution segmentation algorithm [Tzotsos et al., 2014], with a RVM multi-class classifier. Also, another probabilistic SVM classifier has been implemented in order to compare classification results within the OBIA framework. The contribution of this approach is twofold:

- A state-of-the-art sparse classification algorithm (RVM) is integrated in the OBIA framework able to process image objects without the need of parameter tuning and
- Two probabilistic machine learning classifiers are introduced in OBIA methodology.

The paper is structured as follows. In Section 7.2, the related work on RVM, OBIA and multiscale edge-based image segmentation is briefly presented. In Section 7.3, the developed object-based image analysis framework is presented in detail. The quantitative and quantitative evaluation of the experimental results are presented in Section 7.4, while conclusions and future work are discussed in Section 7.5.

7.2 Related Work

7.2.1 Kernel-based Classification in Remote Sensing

Kernel methods have proven effective in the analysis of images of the Earth acquired by airborne and satellite sensors. Kernel methods provide a consistent and well-founded theoretical framework for developing nonlinear techniques and have useful properties when dealing with low number of (potentially high dimensional) training samples, the presence of heterogenous multimodalities, and different noise sources in the data [Gómez-Chova et al., 2011]. Traditional classifiers such as Gaussian maximum likelihood or artificial neural networks are affected by the high dimensionality of input samples, tend to overfit data in the presence of noise, or perform poorly when a low number of training samples are available [Fukunaga and Hayes, 1989; Hughes, 1968].

Support Vector Machine (SVM) was introduced by Vapnik [1998], offering a sparse methodology for regression and binary classification, that has been shown to deliver great results, comparative with the best recent machine learning methods [Theodoridis and Koutroumbas, 2003]. The SVM approach seeks to find the optimal separating hyperplane between classes by focusing on the training cases that are placed at the edge of the class descriptors. These training cases are called support vectors, while the other training cases are discarded. This way, not only an optimal hyperplane is fitted, but also less training samples are effectively used; thus high quality classification is achieved with small training sets [Mercier and Lennon, 2003].

In general, supervised classification and pattern recognition methods use labelled groups of pixels as information about class membership and build a model able to generalize to the whole image [Camps-Valls, 2009]. At present, the most successful methods are neural networks [Del Frate et al., 2007] and support vector machines [Camps-valls and Bruzzone, 2005]. SVMs are particularly appealing in the remote sensing field due to their ability to generalize well even with limited training samples, which is a common limitation in remote sensing applications. However, they suffer from parameter assignment issues that can significantly affect obtained results [Mountrakis et al., 2011; Pal, 2009].

SVM has been used in a variety of remote sensing applications, and has been shown to outperform other classification methods [Foody and Mathur, 2004; Huang et al., 2002a].

The method integrates in the same classification procedure *i)* a feature extraction step, as samples are mapped into a higher dimensional space where a simpler (linear) classification is performed; *ii)* a regularization procedure by which model complexity is efficiently controlled; and *iii)* the minimization of an upper bound of the generalization error, thus following the Structural Risk Minimization (SRM) principle [Gómez-Chova et al., 2011]. These theoretical properties make SVM very attractive in the context of remote sensing image classification [Camps-valls and Bruzzone, 2005] since separation of land cover classes is usually a tedious task. SVMs have been applied to both multispectral [Huang et al., 2002b] and hyperspectral [Camps-Valls et al., 2008] data in a wide range of domains, including object recognition [Inglada, 2007], landcover and multi-temporal classification [Camps-Valls et al., 2008], and urban monitoring [Fauvel et al., 2008]. Another field of growing interest is that of classifier ensembles [Briem et al., 2002] and Random Forests [Ham et al., 2005].

Relevance Vector Machine (RVM) was introduced by Tipping [2001], as a novel approach to kernel-based regression and classification by extending the SVM using a probabilistic Bayessian approach. RVM, similarly to SVM, was originally developed as a tool for binary analysis [Foody, 2008]. SVM was extended to tackle multiclass classification problems, through multi-classification strategies [Hsu and Lin, 2002]. The goal of binary RVM was to predict the posterior probability of assignment for one of the classes (0 or 1) for a given input x. A case may then be allocated to the class with which it has the greatest likelihood of membership [Foody, 2008; Tipping, 2001]. RVM will be described in more detail in the following sections.

Recently, RVM has been used as a state-of-the-art machine learning method for Remote Sensing applications. It was initially used as a regression tool by Camps-Valls et al. [2006] where performance of the RVM was evaluated in terms of accuracy of the estimations, sparseness of the model, robustness to low number of training samples, and computational efficiency. Results suggested that RVMs offer an excellent trade-off between accuracy and sparsity of the solution, and become less sensitive to the selection of free parameters [Camps-Valls et al., 2006]. This research was followed by Foody [2008] who evaluated RVM as a classification method for high resolution remote sensing imagery.

More recently, RVM was evaluated as a classification method for hyperspectral datasets [Camps-Valls et al., 2007; Demir and Erturk, 2007]. Demir and Erturk [2007] has shown

that approximately the same classification accuracy is obtained using SVM and RVM-based classification, with a significantly smaller rate of relevance vectors and much faster estimation time for the RVM method. This feature indicated that the RVM-based hyperspectral classification approach is more suitable for applications that require low model complexity, i.e. small number of training samples is available.

Another effort introduced the Mahalanobis kernel in the formulation of the RVM to take into account the covariance of the features in the classification process for hyperspectral classification [Camps-Valls et al., 2007]. The results were very promising, confirming the accuracy and robustness of the RVM method, especially for the ease of free parameters tuning. A study by Braun et al. [2011], compared RVM and Import Vector Machine (IVM) [Zhu and Hastie, 2001] with SVM for hyperspectral classification. This research suggested that both IVM and RVM significantly outperformed the SVM in terms of overall accuracy, and that the RVM was most reliable in terms of lower producers accuracy, while the consumers accuracy values were considerably higher than the values of IVM and SVM [Braun et al., 2011]. More recent results also showed that the RVM was able to derive classifications of similar accuracy to the SVM for airborne and spaceborne multispectral imagery, but RVM required considerably fewer training cases [Pal and Foody, 2012].

7.2.2 Support Vector Machines and Object-based Image Analysis

In the field of Object-based image Analysis (OBIA), many new developments have been introduced the last years [Blaschke, 2010]. Given that numerous platforms have been developed, in the form of proprietary software (eCognition, ENVI FX, ERDAS Objective) [Baatz and Schape, 2000] or in the form of free software (Orfeo Toolbox, EDISON, MSEG) [Christophe and Inglada, 2009; Inglada and Christophe, 2009; Tzotsos and Argialas, 2006], a broad range of applications on various engineering and environmental remote sensing studies have been developed [Benz et al., 2004; Blaschke, 2010; Dragut et al., 2009; Mladinich et al., 2010; Zhou et al., 2009].

Recently, the SVM methodology was introduced to the OBIA framework as a classifier of primitive image objects [Tzotsos, 2006; Tzotsos and Argialas, 2008], outperforming traditional OBIA classification methods like the Nearest Neighbor classifier [Benz et al., 2004]. In an effort to provide OBIA classification results without the tuning of

image segmentation parameters, Tzotsos et al. [2011] developed a scale-space framework, along with SVM classification, providing promising OBIA classification results. This methodology was further enhanced with advanced edge-based features, incorporated into the image segmentation step [Tzotsos et al., 2014].

Other recent OBIA studies include an introduction of a boost-classifier adapted to OBIA framework and to multi-scale segmentation [Dos Santos et al., 2012]. The principle of boosting is to combine weak classifiers to build an efficient global classifier. Using the paradigm of boosting, an OBIA framework was proposed: each weak classifier was trained for one level of image segmentation and for one region descriptor. The weak classifiers were based on linear SVM and region distances provided by descriptors. It was shown that combination of weak classifiers was able to reduce training time and to produce a stronger classifier [Dos Santos et al., 2012].

7.3 Methodology

The proposed research aims to evaluate and integrate state-of-the-art machine learning methods into the OBIA framework, especially for the classification of primitive image objects. This section is divided into three sub-sections which describe the three major components of the framework: edge-enhanced region-based multiscale image segmentation, RVM object classification and and probabilistic SVM object classification. Before going into a detailed account of each of the three sections, a general presentation of the developed methodology is presented.

The developed framework consists of the following steps: Initially, the remote sensing imagery is pre-processed through scale-space simplification, in order to increase homogeneity within image objects but preserving edge information at the same time. The scale-space formulation of AML was used for this step of the pre-processing [Karantzalos et al., 2007; Tzotsos et al., 2011]. Next, an edge feature extraction step was performed by using a state-of-the-art line detector: the Line Segment Detector (LSD) [Von Gioi et al., 2010]. The edge features, as well as the scale-space representation of the initial image were integrated into a hybrid (edge and region-based) multi-scale image segmentation algorithm, which produced the image primitive objects at selected scales [Tzotsos et al., 2014]. A feature selection procedure was then performed by extracting spectral, textural and spatial features from the primitive objects, forming the feature space for the

following supervised classification steps. Then, a classification process was performed, by incorporating RVM [Tipping et al., 2003] and probabilistic SVM [Platt, 1999] (for comparison purposes) into OBIA classification step: A training set of labelled image objects was provided to the machine learning classifiers and training was performed. In the case of SVM, cross validation was needed to determine the best parameters. Then all the primitive image objects were processed by the trained classifiers, producing the final classifications (RVM and SVM). A qualitative and quantitative evaluation was then performed to measure classification accuracy and computational efficiency of the two classification algorithms.

7.3.1 Multiscale Image Segmentation

As in most OBIA applications, a first step of the image analysis includes the image preproccesing, and the image segmentation, in order to create image primitives [Blaschke, 2010; Tzotsos et al., 2011]. In the proposed framework, a noise removing step was used to simplify the complexity and heterogeneity of the initial remote sensing imagery (RSI). This step is of great importance in order to achieve better medium level (segmentation) and high level (classification) results. Image semantic objects tend to incorporate great value of spectral heterogeneity as the resolution gets higher with the recent available sensors [Tzotsos et al., 2014].

In this study, a non-linear scale space filtering method was applied, based on Morphological Levelings [Karantzalos and Argialas, 2006; Meyer and Maragos, 2000b]. Specifically, a combination of morphological levelings with anisotropic markers, known as Anisotropic Morphological Levelings (AML), has given better results, providing image simplification and at the same time preserving important image properties [Karantzalos and Argialas, 2006; Karantzalos et al., 2007]. The AML formulation, including a native multi-scale representation, has been used for the initial RSI processing, leading to a simplified version of the imagery. The simplification scale is defined by the number of iterations made by the AML algorithm. For more iterations, a stronger simplification is achieved, since an equal number of diffusions are performed in the original data in order to provide the markers for the reconstruction of the leveling [Tzotsos et al., 2011]. For this study, the scale vallue of the simplification was equal to the scale value of the upcoming segmentation step in order to produce image objects, thus avoiding the tuning

of the segmentation parameters.

The RSI simplification was followed by an edge feature extraction step. For this step the LSD algorithm was used, since it does not involve any parameter tuning for the extraction of edges, and has been popular in computer vision applications [Von Gioi et al., 2010]. The LSD algorithm starts by computing the level-line angle at each pixel to produce a level-line field, i.e., a unit vector field such that all vectors are tangent to the level line going through their base point. Then, this field is segmented into connected regions of pixels that share the same level-line angle up to a certain tolerance and is considered as a possible line segment. After examining and validating possible line segments, and test that they are aligned properly, a selection of meaningful rectangles is provided as the final result [Tzotsos et al., 2014; Von Gioi et al., 2010].

For the multi-scale segmentation procedure, an edge enhanced version of the MSEG algorithm [Tzotsos and Argialas, 2006] was implemented using the LSD features. The original MSEG algorithm is a region-merging image segmentation algorithm, using multi-level spatial relationships, in order to produce topologically correct multi-level image segmentations. Staring from a single pixel object representation, it merges image objects iteratively, until no other merge can happen, based on the parameters set by the end user. The criteria for converging is based on the formulation of image heterogeneity as proposed by Baatz and Schape [2000]. Here, as in a previous study [Tzotsos et al., 2011], there was no need for parameter tuning, since the approach included a reliable edge-preserving formulation from AML. Through this study, the parameters of the MSEG algorithm were set stable, the color parameter was set to 0.8 and the shape parameter was set to 0.2.

The edge-enhanced version of the MSEG algorithm [Tzotsos et al., 2014] was used in this proposed method. The line segment features were used as a constraint, during the region merging procedure, preventing the region merging algorithm to consider as neighbors (and thus as possible merges), image objects that lied on opposite sides of a line segment object. This constrain is applied throughout the segmentation iterations, leading to a final setup of image objects where the pixels placed on top of the line segments remain unmerged and other image pixels are merged into image primitives (objects). Then a last iteration of the algorithm (cleaning step) merging is forced on edge objects only, and a selection is made, to which neighboring objects they should be merged, based on local heterogeneity. The fact that LSD features are one-pixel wide is taken into account, thus

edge objects are merged with non edge objects at a last iteration of the merging algorithm [Tzotsos et al., 2014].

After image segmentation, image primitives (non-semantic objects) were extracted and object properties were computed forming the feature space of the classification problem as proposed in Tzotsos and Argialas [2008]. The computed properties (spectral, textural, spatial) were bound to each object by a unique identifier within the object hierarchy of the image. A set of labelled objects (ground truth) was selected, forming a training sample for the following machine learning step.

7.3.2 Relevance Vector Machines

Many problems in machine learning fall under the category of supervised learning: given a set of input vectors $X = \{x_n\}_{n=1}^{N}$ with corresponding class labels $T = \{t_n\}_{n=1}^{N}$. The goal of supervised learning is to be able to predict of t for new values of x. When t is a continous variable, the problem is called *regression*, while when t belongs to a discrete set, the problem is called *classification*. Both SVM and RVM are considered as models where the prediction of t is expressed as a linear combination of basis functions $\phi_m(x)$:

$$y(x, w) = \sum_{m=0}^{M} w_m \phi_m(x) = x^T \phi \tag{7.1}$$

with w_m denoting the model weights [Bishop and Tipping, 2000].

In the case of SVM, the basis function is of the form of kernel function K so that $\phi_m(x) = K(x, x_m)$. In the SVM case, the kernel functions are mapping the input vector x into a high dimensional feature space, where according to the Mercer's theorem [Vapnik, 1998], the inner product of the vectors in the mapping space, can be expressed as a function of the inner products of the corresponding vectors in the original space.

Recently the RVM formulation was introduced by Tipping [2001], introducing probabilistic predictions to the kernel model as in Equation (7.1). This formulation retains the robustness and performance of the SVM while improving the sparseness of the model [Bishop and Tipping, 2000].

In the case of regression, RVM models the conditional distribution of the label variable t as a Gaussian distribution of the form:

$$P(t|x, w, \tau) = N(t|y(x, w), \tau^{-1}) \tag{7.2}$$

with $N(z|m, S)$ denoting a multi-variate Gaussian distribution with mean m and covariance S. Assuming an independent, identically distributed dataset, the likelihood function can be written:

$$P(T|X, w, \tau) = \prod_{n=1}^{N} P(t_n|x_n, w, \tau) \tag{7.3}$$

Given $\alpha = \{\alpha_m\}$ as a vector of weights, and hyperparameter α_m assigned to each weight w_m, the weights w are assigned a Gaussian prior [Bishop and Tipping, 2000]:

$$P(w|\alpha) = \prod_{m=0}^{N} N(w_m|0, \alpha_m^{-1}) \tag{7.4}$$

In RVM formulation [Tipping, 2001], the values for hyperparameters are estimated using type-II maximum likelihood [Berger, 1985], with marginal likelihood maximized for α and τ, with the analytical convolution of:

$$P(T|X, \alpha, \tau) = \int P(T|X, w, \tau)P(w|\alpha)dw \tag{7.5}$$

For the classification case, RVM models the conditional distribution of labels by:

$$P(t|x, w) = \sigma(y)^t[1 - \sigma(y)]^{1-t} \tag{7.6}$$

where $\sigma(y) = (1 + exp(-y))^{-1}$ and $y(x, w)$ as in Equation (7.1). The likelihood function in classification case is given by:

$$P(T|X,w) = \prod_{n=1}^{N} \sigma(y_n)^{t_n}[1-\sigma(y_n)]^{1-t_n} \tag{7.7}$$

which cannot be solved analytically as in regression case (Equation 7.5) [Bishop and Tipping, 2000]. For the classification case Tipping [2001] proposed initially a local Gaussian approximation to the posterior distribution of the weights. The RVM algorithm is initialized with all M basis functions included in the model (1) and updates hyperparameters iteratively, until model convergence. In a later study, a fast marginal likelihood maximisation algorithm was proposed [Tipping et al., 2003], where RVM is initialized with an empty set of basis functions and the maximization of the marginal likelihood is achieved through efficient additions and deletions of candidate basis functions in the sparse model (Equation 7.1). In order to increase the marginal likelihood, basis functions are added to the model, the model is re-estimated (until convergence) and if a basis function becomes redundant ($\alpha_i = \infty$), it is removed from the model.

For this study, the efficient RVM algorithm was implemented in C++ in order to be integrated to the OBIA framework. The objects obtained from the RSI multi-scale segmentation step, along with their computed features from the feature selection step, formed the input x of the RVM classification. A labeled set of image objects (ground truth) was spatially overlaid on the final image segmentation. A selection of training objects X was obtained, using an object overlap rule of 75% [Benz et al., 2004; Tzotsos and Argialas, 2008], along with their labels T.

In order to apply a multi-class RVM classification on x to obtain t, a one-against-all strategy was applied, since it has been shown to be very efficient in multiclass SVM classification cases [Hsu and Lin, 2002] and has also been applied successfully in OBIA [Tzotsos and Argialas, 2008]. The training set X was set as an input to the RVM training module, providing a trained RVM model which included the relevance vectors - the samples for which the parameters are not discarded from the sparse model ($\alpha \neq \infty$). Relevance vectors are the equivalent of the support vectors in the SVM formulation. Finally, the complete set x of image primitives was given as input to the trained RVM, providing prediction set t (classification). The proposed classification method was applied to RSI within the OBIA framework. Evaluation of results are provided in a following section.

7.3.3 Probabilistic Support Vector Machines

Uncertainty handling is usually a desired feature of a machine learning classifier, since it provides the ability to combine results with other classifiers, like knowledge-based expert systems, hybrid decision systems or ensembles [Theodoridis and Koutroumbas, 2003]. Even in a simple multiclass classification problem, the classification result can be validated and evaluated using a metric between the ``classification score'' of the winning class against the runner-up classes [Benz et al., 2004]. It has been suggested that producing a posterior probability $P(class|input)$, and choosing the class based on the maximal posterior probability over all classes is the Bayes optimal decision [Platt, 1999].

Support Vector Machines provide a raw decision value, given an unlabelled vector x as an input:

$$f(x) = \sum_i y_i \alpha_i k(x_i, x) + b \tag{7.8}$$

with $k(x_i, x)$ being the kernel function. Vapnik [1998] proposed a method to create posterior probability as an output to the decision value by decomposing the feature space into a direction orthogonal to the separating hyperplane and all of the other dimensions grouped together. Then the posterior probability depends on a scaled version t of $f(x)$ and a vector u for the group of other dimensions. The proposed formulation by Vapnik [1998] was:

$$P(y = 1|t, u) = \alpha_0(u) + \sum_{n=1}^{N} \alpha_n(u) cos(nt) \tag{7.9}$$

A more recent approach included mapping the SVM output to posterior probability through a sigmoid function [Platt, 1999]. Given that the output of a binary SVM is $y = \pm 1$, a Gaussian can be fit to the densities $p(f|y = 1)$ and $p(f|y = -1)$, leading to a posterior probability rule of $P(y = 1|f)$ being a sigmoid [Hastie and Tibshirani, 1998]. A parametric model was used by Platt [1999] to fit the posterior $P(y = 1|f)$ into the SVM model:

$$P(y = 1|f) = \frac{1}{1 + exp(Af + B)} \tag{7.10}$$

A similar sigmoid fit to the output of the Nearest Neighbor classifier was proposed to map classification output to a fuzzy membership value within the OBIA framework [Benz et al., 2004].

In the present study, the sigmoid fit formulated by Equation (7.10) was implemented in C++ and applied to the previously implemented SVM classification method [Tzotsos and Argialas, 2008], in order to compare RVM and SVM classification results within the OBIA framework.

7.4 Evaluation and Discussion

The objective of the developed approach was to introduce the RVM sparse classification algorithm in the OBIA framework in order to classify image primitives without tuning of parameters and to compare the results against previously proposed kernel-based (SVM) classification methods. A secondary objective was to implement a probabilistic SVM approach in order to compare OBIA classification results between the two state-of-the-art machine learning methods.

In this section, a comparison between three classifiers is presented: RVM, SVM and probabilistic SVM. Since SVM method was previously compared to standard remote sensing classifiers [Tzotsos and Argialas, 2008] like Nearest Neighbor, this comparison was not considered for the paper. In Tables 1 and 2, a quick look shows that SVM outperforms both probabilistic methods in classification accuracy, but the results are very close and can be considered comparable. The quantitative results of this study, confirm the comparison between SVM and RVM in pixel-based approach, as presented by Foody [2008].

For the multi-class SVM classification, the class assignment is a binary operation, leading to object classification based on a voting mechanism. This is applied either with a one-against-all or a one-against-one strategy [Hsu and Lin, 2002]. After multiple SVM classifications, a post-processing step, involving voting, assigns each image object to the class with larger number of votes. The output of such procedure is a binary decision. The

probabilistic approach followed here, involved the classification decision being made by the maximum posterior probability from the sparse classifier. This is a great advantage of the proposed approach, especially in the case of machine learning classification being followed by a rule-based refinement of classification results using class-related features, which is very common within the OBIA framework [Benz et al., 2004].

Another advantage of the RVM algorithm is that is does not include any parameter tuning. SVM classification usually involves a selection step for model and kernel parameters. For this study, the RBF kernel was used for both SVM classifiers (probabilistic and multi-class), which involved two parameters: the penalty term C and RBF parameter gamma. The model parameters are usually determined by a cross-validation process, where the training set is divided in training and validating parts. Cross validation was performed here too in order to evaluate (in a rotational manner) the values of the parameters that lead to higher accuracy. This step was not needed in RVM training, due to the properties of the sparse model and the integration of the Bayessian theory. In addition, the basis function used in RVM was a non-Mercer kernel [Tipping, 2001]. Here, this feature was not used and the RBF kernel was also selected in RVM for comparison purposes.

RVMs are known to be more sparse than SVMs [Tipping, 2001], that is when the final model is trained RVMs keep a smaller number of relevance vectors in order to predict the class of a new vector x than the number of support vectors needed by SVMs. For this study and for the case of small training sample, after the training step, for each class a range of [2-6] relevance vectors were used for RVM, while a range of [12-29] support vectors were used for SVM classifiers. This had an impact on decision time: when both RVMs and SVMs are trained, the decision of the class of a new x is much faster for RVM, since the relevance vectors involved in the decision are much fewer in number than support vectors that are placed on the margin boundaries. On the other hand, RVM needed almost double time to train, since the model re-evaluation is more complex than SVM.

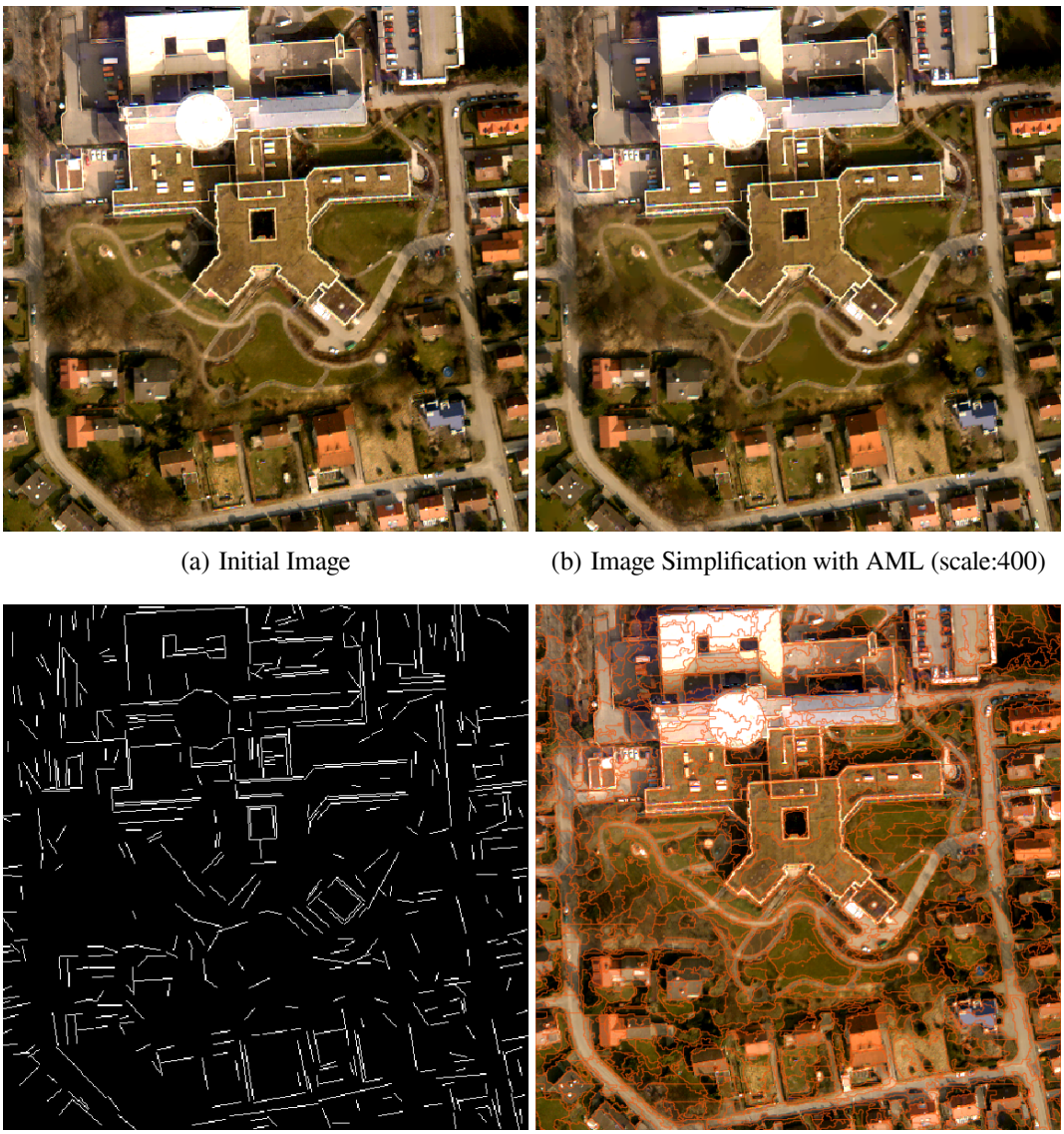
For the evaluation of results, two case studies are presented, involving both qualitative and quantitative evaluation criteria. First, a very high spatial resolution (0.5m) dataset from an airborne scanner was used to evaluate the classification accuracy, since OBIA approaches are known to perform better on high resolution data [Blaschke, 2010]. Secondly, a high resolution Landsat TM imagery was used to evaluate the proposed method.

7.4.1 Very high spatial resolution airborne imagery

The developed methodology was applied to a very high resolution (50cm) airborne imagery. The initial image is presented in Fig.7.1(a). A scale space representation was constructed using the AML algorithm at a scale of 400. It is shown in Fig.7.1(b) that the initial image is simplified, in a non-linear manner, preserving the edge information of the dataset. This step was performed in order to avoid detection of false edges due to heterogeneity and noise. Then, the LSD edge extraction algorithm was applied on the simplified image and the result is shown in Fig.7.1(c). Those edge features were given as input in the MSEG algorithm and the segmentation scale was set to 400 (same as scale-space parameter). The resulting segmentation is shown in Fig.7.1(d). For each image object, a feature extraction procedure was performed, computing spectral and texture features, based on the pixel values belonging to each object. In addition, shape features were also computed for each object. All object features were then exported to a feature database forming the prediction (or testing) set for the machine learning algorithms.

For training the supervised learners, a set of ground truth vector data was converted to a raster file, with class identification number encoded to it. Based on this dataset, a set of image objects was selected from the feature database, as long as the spatial overlap was at least 75% to avoid selection and training errors. This percentage was set to this default value in order to provide a vector to segmentation overlap tolerance: image object boundaries are not aligned to the ground truth vector boundaries. Thus, a total of 70 sample objects was extracted for 4 classes as can be seen in Fig.7.4.1 (d). Based on this set both RVM and SVM classifiers were trained. For the SVM the number of support vectors were: 23 for *Vegetation*, 12 for *Tile Roofs*, 29 for *Bright Roofs* and 14 for *Asphalt Like* classes. Each of these numbers involve the total support vectors for each iteration of the one-against-all procedure, before the voting decision for the final classification. For the RVM model, the number of relevance vectors were: 2 for *Vegetation*, 6 for *Tile Roofs*, 3 for *Bright Roofs* and 5 for *Asphalt Like* classes. Thus it is shown that RVM model in this case is far more sparse, leading to better performance (execution time) in prediction of class for the total of 4891 image primitives extracted from MSEG.

After the training of the classifiers, a prediction was made for all image objects. A quantitative evaluation was performed using a second set of ground truth data and producing pixel-based confusion matrices (Table 7.1) The classification results are also



(a) Initial Image

(b) Image Simplification with AML (scale:400)

(c) Egde extraction with LSD algorithm

(d) Image segmentation result with edge-enhanced MSEG (scale:400)

Figure 7.1: Original multispectral aerial scan image (©Toposys), simplification with scale-space AML algorithm, edge extraction with Line Segment Detector and image segmentation with edge-enhanced MSEG algorithm, producing the primitive image objects for further machine learning classification

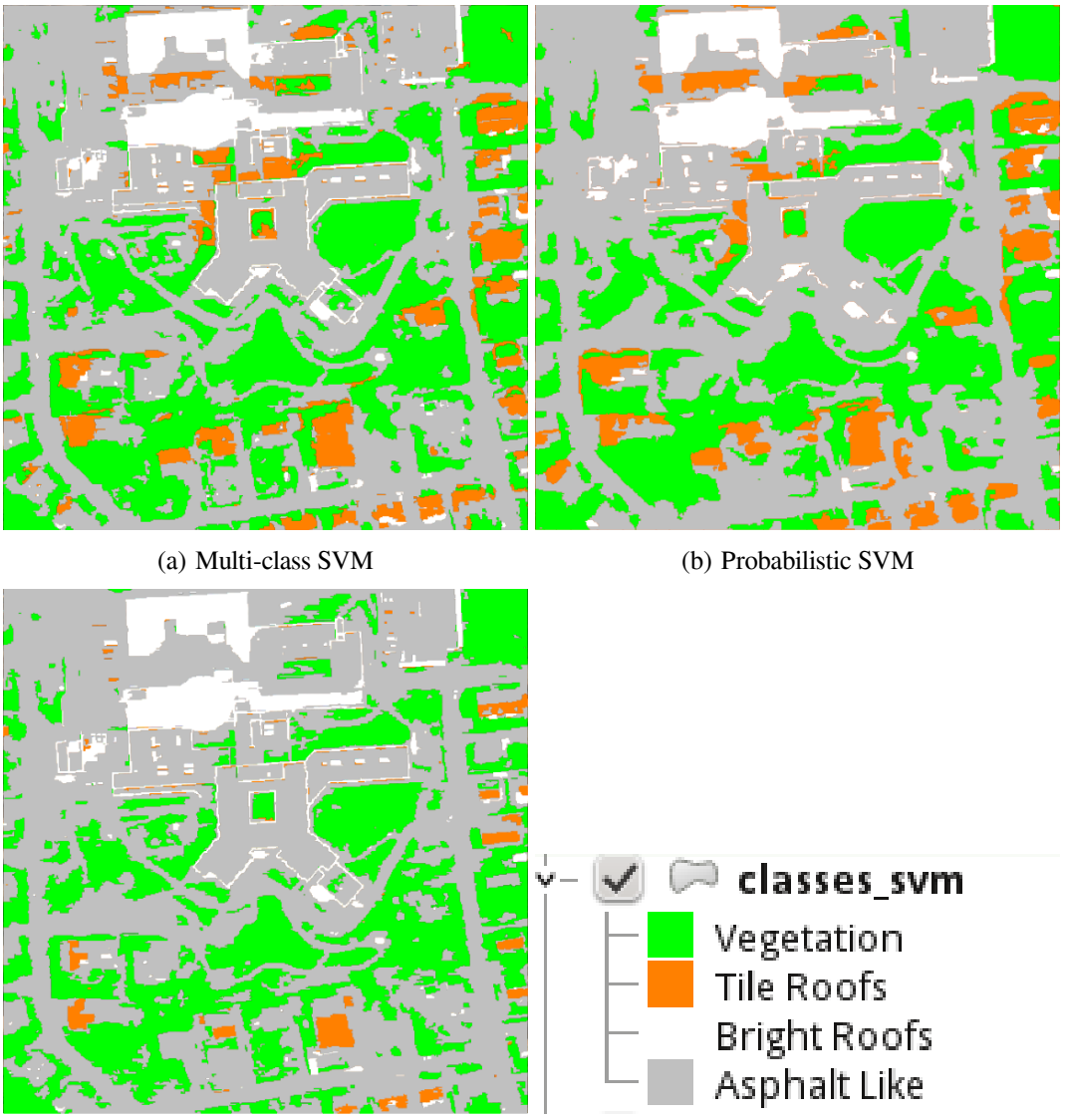
presented Fig.7.4.1. for qualitative evaluation.

From Table 7.1 it is shown that plain SVM algorithm outperformed probabilistic SVM and RVM with an accuracy of 91.34%, 89.50% and 85.75% respectively. The accuracy differences were not very significant and confirm the results of the pixel-based comparison by Foody [2008]. From the quantitative evaluation (Table 7.1) it is also apparent that RVM performed best for 2 out of 4 classes, but showed having issues with Vegetation objects being classified as *Asphalt Like* objects. From the qualitative evaluation (Fig.7.4.1), it is confirmed the RVM (Fig.7.4.1c) was more efficient in detecting the rectangular shape of *Tile Roofs* and seemed to delineate the *Bright Roofs* objects correctly. Both probabilistic SVM (Fig.7.4.1b) and multi-class SVM (Fig.7.4.1a) were more successful detecting all *Tile Roofs*, although their extracted shape was not accurate. In all three classifiers, there was a confusion between *Bright Roofs* and *Asphalt Like* objects.

7.4.2 Multispectral satellite imagery

The developed methodology was also applied to a high resolution satellite Landsat TM multispectral image to confirm the previous results. The initial multispectral dataset is presented in Fig.7.3(a). In Fig.7.3(b) the AML filtering is shown at scale 100. Then, LSD was applied to extract line segments as shown in Fig.7.3(c). Finally, the edge-enhanced MSEG segmentation algorithm was used with scale 100 to produce the image primitives (Fig.7.3d). As described before, a feature extraction step (as performed in all OBIA applications) was performed and object properties were computed to create the classification feature space. Then an object selection was performed based on ground truth data (again the overlap tolerance was set to 75%) leading to 765 sample objects for 4 classes (Fig.7.4d).

Based on this set of samples RVM and SVM classifiers were trained. For the SVM the number of support vectors were: 58 for *Woodland General*, 15 for *Grassland General*, 12 for *Impervious General* and 14 for *Waterbodies* classes. For the RVM model, the number of relevance vectors were: 15 for *Woodland General*, 20 for *Grassland General*, 15 for *Impervious General* and 19 for *Waterbodies* classes. In this case, due to the large sample number, SVM is very sparse and similar to RVM, while RVM training time was very long (aprox. 5 minutes) compared to SVM training (aprox.30 seconds). Classification times for all classifiers were similar, few seconds for the total of 26505 image primitives



(a) Multi-class SVM (b) Probabilistic SVM

(c) RVM (d) Class Legend

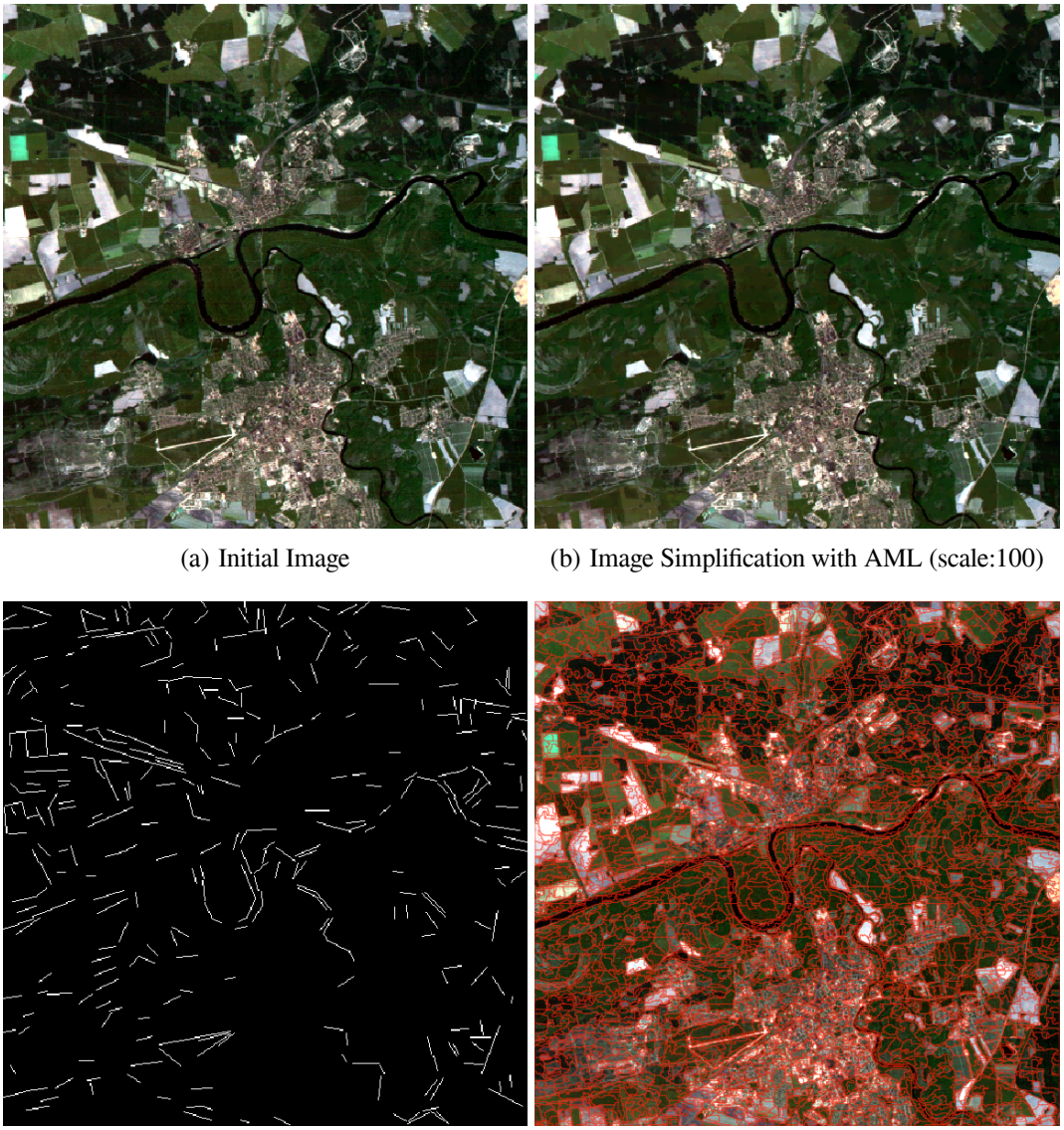
Figure 7.2: Machine Learning classification of multispectral aerial scan image (©Toposys).

Multi-Class SVM Classification Accuracy				
	Vegetation	Tile Roofs	Bright Roofs	Asphalt Like
Vegetation	16113	51	0	1622
Tile Roofs	53	4701	99	1065
Bright Roofs	0	2	8353	2072
Asphalt Like	425	248	468	35218
Overall Accuracy: 91.34%				

Probabilistic SVM Classification Accuracy				
Vegetation	14912	15	0	2859
Tile Roofs	137	4765	36	980
Bright Roofs	0	1	8260	2166
Asphalt Like	227	432	549	35151
Overall Accuracy: 89.50%				

RVM Classification Accuracy				
Vegetation	12564	0	0	5222
Tile Roofs	0	3669	384	1865
Bright Roofs	0	1	8363	2063
Asphalt Like	81	21	407	35850
Overall Accuracy: 85.75%				

Table 7.1: Quantitative results regarding the classification accuracy of SVM, probabilistic SVM and RVM for a very high spatial resolution airborne multispectral dataset.



(a) Initial Image

(b) Image Simplification with AML (scale:100)

(c) Egde extraction with LSD algorithm

(d) Image segmentation result with edge-enhanced MSEG (scale:100)

Figure 7.3: Original multispectral Landsat TM image (Dessau, Germany), simplification with scale-space AML algorithm, edge extraction with Line Segment Detector and image segmentation with edge-enhanced MSEG algorithm, producing the primitive image objects for further machine learning classification

extracted from MSEG.

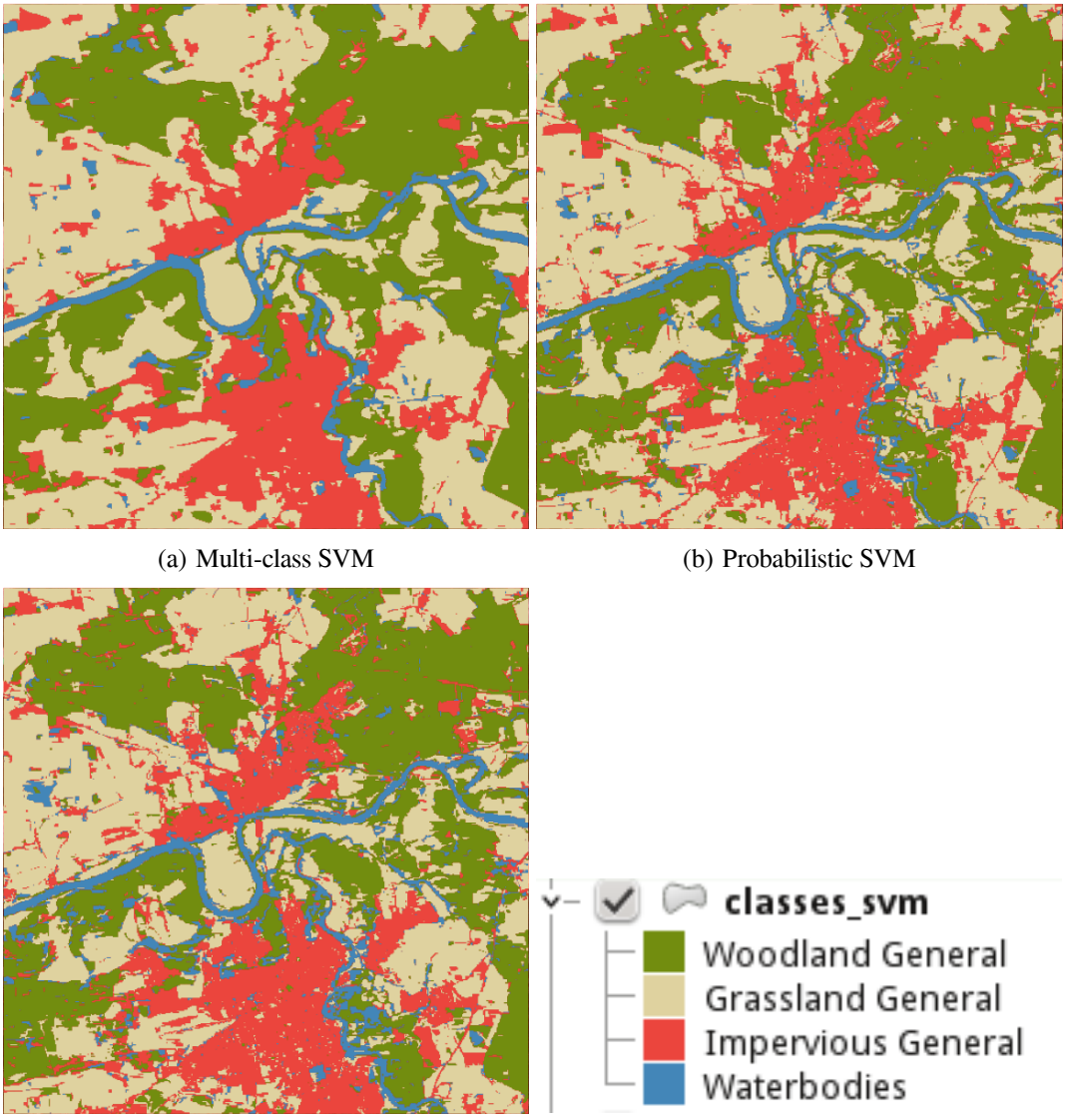
As can be observed from Table 7.2, multi-class SVM again outperformed probabilistic SVM and RVM with an accuracy of 89.60%, 88.25% and 85.79% respectively. Again, the accuracy differences were marginal. From the qualitative evaluation (Fig.7.4), it can be observed that probabilistic SVM (Fig.7.4b) performed better for the *Waterbodies* class, while multi-class SVM (Fig.7.4a) performed better for the *Impervious General* class. RVM, again had more issues with one of the classes, and specifically *Waterbodies* (Fig.7.4c), resulting in small variation from the SVM accuracy.

Multi-Class SVM Classification Accuracy				
	Woodland	Grassland	Impervious	Waterbodies
Woodland	17218	3028	47	426
Grassland	851	15876	235	43
Impervious	182	207	8529	96
Waterbodies	229	8	2	4501
Overall Accuracy: 89.60%				

Probabilistic SVM Classification Accuracy				
Woodland	17099	2771	126	723
Grassland	1081	15498	327	99
Impervious	356	298	8247	113
Waterbodies	115	38	1	4586
Overall Accuracy: 88.25%				

RVM Classification Accuracy				
Woodland	16183	2815	155	1566
Grassland	1187	15130	533	155
Impervious	236	296	8313	169
Waterbodies	124	79	1	4536
Overall Accuracy: 85.79%				

Table 7.2: Quantitative results regarding the classification accuracy of SVM, probabilistic SVM and RVM for a high spatial resolution satellite multispectral dataset.



(a) Multi-class SVM (b) Probabilistic SVM

(c) RVM (d) Class Legend

Figure 7.4: Machine Learning classification of multispectral Landsat TM image (Dessau, Germany).

7.5 Conclusions and Future Perspectives

In this study, a state-of-the-art Relevance Vector Machine classification method was introduced and evaluated within the Object-Based Image Analysis framework, showing promising results, comparable with the Support Vector Machine classifier. Previous RVM studies in remote sensing involved only pixel-based RVM classification and regression. As a secondary contribution, a probabilistic SVM implementation was also evaluated within the OBIa framework, with very promising results, almost similar with the binary SVM classifier.

Based on the above evaluation and given that RVM was always faster in execution times, while being more sparse and being independent of model parameter tuning, RVM can be considered as an excellent choice as an OBIA classifier, especially in cases when the sample/training size is very small. This is a valuable feature since OBIA supervised classification usually involves a small set of training samples, because ground truth data are evaluated as objects and not as pixels. A major feature of both proposed algorithms is the handling of uncertainties through posterior probabilities computation. On the other hand, long training time for RVM is an obvious disadvantage, although various optimizations have already been proposed and implemented.

The performed qualitative and quantitative evaluation reported that machine learning classification results are very promising and can be easily combined with other classification methods (e.g. rule-based). Some of the topics for further research and development are: comparison of RVM and SVM methods with other state-of-the-art machine learning algorithms, evaluation of fuzzy SVMs, evaluation of unsupervised learning methods and integration with knowledge-based classification approaches for complex remote sensing applications.

Chapter 8

Integrating knowledge-based expert systems and advanced machine learning for object-based image analysis

8.1 Introduction

The last decade, the Object-Based Image Analysis (OBIA) paradigm has been the subject of very active research and has developed into a state-of-the-art methodology in remote sensing [Blaschke, 2010]. This new methodology uses the object-oriented paradigm, moving away from pixel-based methods, leading to advancements in classification results and accuracy, bringing remote sensing information closer to GIS systems and spatial databases [Baatz and Schape, 2000; Benz et al., 2004; Blaschke, 2010]. The OBIA methodology integrates low level image processing methods with high level classification methods, such as knowledge-based systems, in order to achieve classification results of high quality. Knowledge-based expert systems are used to represent human expertise into machine readable code in the form of production rules. Those rules are then used to extract image semantics.

At the same time, the availability of remotely sensed data is increasing (satellite data, airborn sensors, UAVs, micro-satellites etc.) and the need for fast, accurate and auto-

mated classification procedures is now more profound than ever. Towards this, a significant progress is being made the last years, incorporating machine learning methods into classification of remotely sensed data [Gómez-Chova et al., 2011; Mountrakis et al., 2011]. In the field of knowledge-based systems, recent research is focused on the representation of knowledge but also in the efficiency of reasoners and integration with other artificial intelligence systems [Arvor et al., 2013; Crevier and Lepage, 1997; Sahin et al., 2012].

Recently, advanced machine learning methods have been evaluated and integrated into the OBIA framework [Tzotsos, 2006; Tzotsos and Argialas, 2008] outperforming previous object-based classification methods on various remote sensing data. Support Vector Machine (SVM) classification has also been evaluated within an OBIA framework, through integration with scale-space representations, showing very promising results and eliminating the need of parameter tuning within the OBIA framework [Tzotsos et al., 2011].

In this paper, an object-based image analysis framework was developed integrating:

- advanced machine learning classifiers such as Support Vector Machine and Relevance Vector Machines (RVM)
- a Hybrid edge-based and region-based image segmentation algorithm
- a non-linear scale-space simplification algorithm and
- a Knowledge-Based Expert System (KBES) shell.

The contributions of this approach are twofold:

- A state-of-the-art sparse machine learning classification algorithm (RVM) is integrated in the OBIA framework able to perform supervised classification on primitive image objects and
- An object-based image classification is achieved using a well-established Expert System tool (CLIPS) [Giarratano and Riley, 1998], integrating OBIA with Free and Open Source Geospatial Software (FOSS4G).

The paper is structured as follows. In Section 2, the related work on knowledge-based expert systems, OBIA and machine learning is briefly presented. In Section 3, the

developed object-based image analysis framework is presented in detail. The evaluation of the experimental results are presented in Section 4, while conclusions and future work are discussed in Section 5.

8.2 Related Work

8.2.1 Knowledge-based Expert Systems in Remote Sensing

In Artificial Intelligence, an Expert System (also called a Knowledge Based Expert System) is a computer system designed to emulate the decision-making and problem-solving ability of a human expert in a narrow domain or discipline [Giarratano and Riley, 1998; Jackson, 1990; Prasad and Babu, 2008]. Expert Systems are composed by: *i)* the knowledge base, which is a collection of facts (data) and rules derived from the human expert, typically represented as IF-THEN statements, *ii)* the inference engine, which is the main processing element of the expert system that applies the rules to the known facts to deduce new facts, and *ii)* the user interface which is the method by which the expert system interacts with a user [Giarratano and Riley, 1998]. Expert Systems are designed to solve complex problems by reasoning about given knowledge, while asking for user input when no other information is available.

Some properties/advantages of Expert Systems [Crevier and Lepage, 1997; Prasad and Babu, 2008] include:

- they simulate human reasoning about a problem domain, rather than simulating the domain itself.
- they perform reasoning over representations of human knowledge
- they solve problems by heuristic or approximate methods
- they provide ease of prototyping: rules can be added easily without modifying existing rules
- their ability to allow the separate representation, but combined use, of knowledge pertaining to different domains

Given the above properties, Expert Systems have found many applications in various domains especially medical, military, education etc. [Jackson, 1990; Sahin et al., 2012]. The first efforts in the use of declarative knowledge for image processing occurred in the late 1970s and early 1980s [Crevier and Lepage, 1997]. In Geosciences and particularly in Remote Sensing, the potential of Expert Systems became apparent in the early days, since image interpretation was (and still is) a difficult and time consuming task performed by a small number of experts [Argialas and Harlow, 1990]. Image interpretation can be defined as the extraction of the image semantic and it consists in obtaining useful spatial and thematic information on the objects by using human knowledge and experience [Forestier et al., 2012]. Expert systems are considered to have a great potential for image classification especially where the combined use of multiple sources of data such as satellite imagery and GIS data are necessary [Weng, 2012]. Remote sensing experts, usually observe the differences between the visual interpretation of the spectral information and the semantic interpretation of the pixels at different levels of abstraction [Forestier et al., 2012].

In the remote sensing literature, expert systems have mostly been proposed as a multistage classification method. In general, remote sensing knowledge bases include structural knowledge and procedural knowledge in the form of a spatio-temporal ontology of the world. This knowledge is usually represented in a semantic network or conceptual graph [Baraldi and Boschetti, 2012]. Object-Based Image Analysis (OBIA) has evolved in the early 2000s based on the notion of semantic networks. A semantic network consists of: *i)* a hierarchical representation of image classes (semantics) represented as nodes; *ii)* spatial (e.g. topological) and non-spatial relations (e.g., is-a, part-of) between classes represented as arcs between nodes [Baatz and Schape, 2000; Benz et al., 2004; Blaschke, 2010]. Semantic networks are applied on a set of image objects (primitives) with their computed properties (feature space) providing classification results in a form of vector datasets and spatial databases.

During the last decade, the use of OBIA has been drastically popularized because of its ability to use semantics based on descriptive assessment and knowledge, an approach tightly related to the Expert Systems paradigm [Arvor et al., 2013]. OBIA builds upon older methods and concepts such as image segmentation, edge-detection, feature extraction and classification that have been used in remote sensing image analysis for decades, but their integration with expert knowledge provided a new paradigm for pro-

cessing remote sensing images [Blaschke, 2010]. Still, OBIA has some limitations: *i)* each expert has his own conceptualization of the reality he intends to represent on the image, thus two experts analyzing the same data will obtain two different results because of their different experiences and *ii)* the processing chain to achieve a classification is not entirely controlled and documented [Arvor et al., 2013]. For the first problem, the use of ontologies (as a formal, explicit specification of a shared conceptualization) is being actively researched as a solution of global representation [Arvor et al., 2013]. For the second problem, knowledge-based expert systems are becoming more and more important in image processing due to their ability to document and provide full reports of their reasoning [Jackson, 1990]. Even though acquiring and representing the knowledge of a domain is often a tedious process, the advantages of it are undeniable [Forestier et al., 2012]. In this paper we propose such an integration between OBIA and traditional expert system tools for object-based image classification.

8.2.2 Machine Learning in Remote Sensing

Lots of current efforts in remote sensing involve development of new algorithms to make pixel-based classification tasks more efficient. With increasing volumes of available imagery and increasing variety of sensors (SAR, hyperspectral, multispectral, thermal etc.), robust classification algorithms are needed for remote sensing scientists to be able to effectively process the imagery available [Camps-Valls, 2009]. The state-of-the-art in remote sensing pixel-based classification evolves around unsupervised and supervised machine learning methods. Especially kernel methods and ensembles have been shown to provide the most accurate results, while at the same time requiring small user interaction [Gómez-Chova et al., 2011; Maulik and Chakraborty, 2011; Mountrakis et al., 2011]. Various classification schemes have also been developed and evaluated in the direction of multimodal classification, i.e. various classifiers working in parallel to iteratively improve image classification results [Givens et al., 2013; Guillaumin et al., 2010].

Support Vector Machine (SVM) is a sparse binary regression and classification methodology that has been shown to deliver great results, comparative with the best recent machine learning methods [Mountrakis et al., 2011; Vapnik, 1998]. Various steps are integrated in the SVM classification procedure: *i)* a feature extraction step, where samples are mapped to a higher dimensional space and a simpler (linear) classification is performed;

ii) a regularization procedure by which the model's complexity is efficiently controlled; and *iii)* the minimization of an upper bound of the generalization error, thus following the Structural Risk Minimization (SRM) principle [Gómez-Chova et al., 2011].

Due to the theoretical property of sparseness, SVM is very attractive in the context of remote sensing image classification [Camps-valls and Bruzzone, 2005]. SVM has been used in a variety of Remote Sensing applications, and has been shown to outperform other classification methods [Foody and Mathur, 2004; Huang et al., 2002a]. SVM classification has been applied to both multispectral [Huang et al., 2002b] and hyperspectral [Camps-Valls et al., 2008] data in a wide range of domains, including object recognition [Inglada, 2007], landcover and multi-temporal classification [Camps-Valls et al., 2008], and urban monitoring [Fauvel et al., 2008]. Recently SVM classification has also been evaluated within the OBIA framework [Tzotsos and Argialas, 2008; Tzotsos et al., 2011] giving promising results.

Relevance Vector Machine (RVM) was introduced by Tipping [2001], introducing a novel approach to kernel-based regression and classification by extending the SVM using a probabilistic Bayessian approach. RVM, similarly to SVM, was originally developed as a tool for binary analysis [Foody, 2008]. Recently, RVM has been used as a state-of-the-art machine learning method for Remote Sensing applications. It was initially used as a regression tool by Camps-Valls et al. [2006] where results suggested that RVMs offer an excellent trade-off between accuracy and sparsity of the solution, and become less sensitive to the selection of the free parameters. This research was followed by Foody [2008] who evaluated RVM as a classification method for high resolution remote sensing imagery. RVM has also been evaluated as a classification method for hyperspectral datasets with good results [Camps-Valls et al., 2007; Demir and Erturk, 2007]. Demir and Erturk [2007] has shown that approximately the same classification accuracy is obtained using SVM and RVM-based classification, with a significantly smaller relevance vector rate and much faster testing time for the RVM method. This feature indicated that the RVM-based hyperspectral classification approach is more suitable for applications that require low complexity.

8.3 Methodology

For this research a multimodal object-based image classification approach was designed and evaluated. The goal of this research was to integrate state-of-the-art machine learning classification methods with knowledge-based expert systems, incorporating them into the Object-Based Image Analysis methodology and to evaluate their effectiveness and prospects. This section is divided into three sub-sections which describe the three major components of the framework: edge enhanced region-based multiscale image segmentation, RVM object-based classification and Expert System object-based classification. First, a general presentation of the developed methodology is presented here.

The developed OBIA framework initially applies non-linear scale-space simplification, to decrease object heterogeneity, especially in cases of very high spatial resolution, while maintaining the edge information of the image semantics. For this step, the ALM and AML formulations have been implemented [Karantzalos et al., 2007; Perona and Malik, 1990] and integrated in the OBIA pre-processing step [Tzotsos et al., 2011]. The scale-space representation of the initial imagery is then processed by a state-of-the-art line detector: the Line Segment Detector (LSD) [Von Gioi et al., 2010] so to derive the edge information of the filtered image. Next, the simplified image and the edge information are given as input in the edge-enhanced MSEG multiscale segmentation algorithm [Tzotsos et al., 2014] which produces a set of primitive objects, at the given scale parameter. After deriving the primitive objects, a feature selection step follows which involves computation of spectral, textural and spatial features. Thus, a classification feature space is formed, to be used in the following classification steps. The segmentation and feature extraction steps result in a vector representation of primitive objects which are stored in a spatially-enabled Relational Database Management System (RDBMS). Then a supervised machine learning algorithm was used as a first classification step.

In this paper we evaluated the employment of RVM in the context of OBIA, while previous studies have evaluated RVM as a pixel-based classifier Foody [2008]. A training set of labelled image objects was provided to the machine learning classifier and training was performed using cross validation [Tzotsos et al., 2011]. Then all the primitive image objects were classified by the trained RVM classifier, producing the first classification result and updating the assigned class column in the spatial RDBMS.

In order to produce a refined knowledge-based classification result, a three-tier soft-

ware framework was developed integrating *i)* the spatial RDBMS holding the vector information, *ii)* the object-oriented environment of the Python programming language (as a intermediate geospatial processing level) and *iii)* a Pythonic wrapper around the CLIPS Expert System C library. The image objects and their properties are loaded from the database into Python as native objects with attributes. Those objects are then represented into CLIPS objects and their attributes (facts) are loaded in the CLIPS agenda. The user provides a knowledge base, using the CLIPS programming language, representing the rules to be used for the classification. The CLIPS inference engine is instantiated within Python and the rules are automatically fired using the RETE algorithm [Giarratano and Riley, 1998] to provide a class assignment for each object. Finally the result (assignment to a class) is submitted back to Python object and the spatial database is updated with the new class identifier. The classification result is then evaluated against ground truth GIS information and the final accuracy is derived.

8.3.1 Multiscale Image Segmentation

In order to implement OBIA classification, the first step is image segmentation so that to produce primitive objects. Before image segmentation, it is a common practice to preproccess image data so that segmentation algorithms can be more effective. In this approach, the preproccesing stage involved the construction of the non-linear scale-space representation in order to elegantly simplify image data. Such methods provide robust simplification of images without the loss of important information such as edges, that are of high importance for higher level processing algorithms, especially in OBIA applications, where high resolution data are usually involved. For the preproccesing step, the Anisotropic Morphological Levelings (AML) [Karantzalos et al., 2007] were implemented. The AML algorithm has already been proposed as a part of an OBIA methodology with very promising results since it leads to better segmentation results [Tzotsos et al., 2011]. In the filtering process, only the scale parameter of the scale-space formulation is used (number of filtering iterations), in order to provide the level of abstraction wanted by the end user.

After the image simplification step, the edge information of the image semantics was preserved and it was detected using a state-of-the-art line detector, Line Segment Detector (LSD) [Von Gioi et al., 2010]. LSD is a linear-time Line Segment Detector

giving subpixel accurate results without the need of parameter tuning in the process. The LSD algorithm starts by computing the level-line angle at each pixel to produce a level-line field, i.e., a unit vector field such that all vectors are tangent to the level line going through their base point. Then, this field is segmented into connected regions of pixels that share the same level-line angle up to a certain tolerance. These connected regions are called line support regions [Von Gioi et al., 2010]. Each line support region (a set of pixels) is a candidate for a line segment. The principal inertial axis of the line support region is used as main rectangle direction. After examining, validating the line support regions, and testing that they are aligned properly, a selection of meaningful rectangles is provided as the final result.

For the creation of primitive image objects, an improved version of the initial MSEG image segmentation algorithm [Tzotsos and Argialas, 2006] was implemented. MSEG is a region-based multi-scale segmentation algorithm developed for object-oriented image analysis. Briefly, starting from a pixel representation it creates objects through continuous pair-wise object fusions, executed in iterations (passes). For each pass, every object is evaluated in relation with its neighboring objects towards the optimal pair of objects adequate for fusion. MSEG algorithm defines a cost function for each object merge and then implements various optimization techniques to minimize this cost. The cost function is implemented using the measure of homogeneity (color and shape) in the same way with other approaches [Baatz and Schape, 2000]. Through this research, the cost parameters of the MSEG algorithm were set to the default values, while setting the same scale parameter with the scale-space filtering: the larger the scale, the more image simplification is applied and larger image segments are created. In a previous study [Tzotsos et al., 2011] is was shown that there is no need for tuning the segmentation parameters when combined with the appropriate edge-preserving scale-space formulation.

In order to take into advantage the edge detection of the LSD algorithm, a modified version of MSEG was used to derive image objects [Tzotsos et al., 2014]. The edge-enhanced MSEG is able to integrate edge information (as a constraint) during the segmentation procedure. During the iterative region-merging, starting from single pixel objects, the edge information is used as a boundary. Two adjacent segments (pixels at first iteration) will not be merged into a new object if one or both reside on top of an edge. This way during following iterations, edge objects remain unmerged preserving the edge boundaries from intersecting a single image object. After several passes (iter-

ations) converging of the algorithm occurs and no more object merging is performed, due to scale parameter. Finally a last iteration of the algorithm is forced on edge objects only, and a selection is made, to which neighboring object they should be merged, based on heterogeneity of the resulting object. In this paper, the resulting image objects were then exported both as labelled image (segmented image) and as vector data and imported into a spatial RDBMS for further classification purposes. For every segmentation level obtained, a single spatial RDBMS table is created. This reduces the memory footprint of the classification procedures, while at the same time the raster representation (labelled image) is maintained in storage for extracting more features or representing the final result as raster if needed. After image segmentation, object properties were computed forming the feature space of the classification problem as proposed in Tzotsos and Argialas [2008]. The computed properties (spectral, textural, spatial) were bound to each object by a unique identifier within the object hierarchy of the image and stored into the RDBMS as column records.

8.3.2 Relevance Vector Machine Classification

Many problems in machine learning fall under the category of supervised learning. Given a set of input vectors $X = \{x_n\}_{n=1}^{N}$ with corresponding class labels $T = \{t_n\}_{n=1}^{N}$, the goal of supervised learning is to be able to predict of t for new values of x. When t is a continous variable, the problem is called *regression*, while when t belongs to a discrete set, the problem is called *classification*. Both SVM and RVM are considered as models where the prediction of t is expressed as a linear combination of basis functions $\phi_m(x)$:

$$y(x, w) = \sum_{m=0}^{M} w_m \phi_m(x) = x^T \phi \tag{8.1}$$

with w_m denoting the model weights [Bishop and Tipping, 2000].

In the case of SVM, the basis function is of the form of kernel function K so that $\phi_m(x) = K(x, x_m)$. In the SVM case, the kernel functions are mapping the input vector x into a high dimensional feature space, where according to the Mercer's theorem [Vapnik, 1998], the inner product of the vectors in the mapping space, can be expressed as a function of the inner products of the corresponding vectors in the original space.

Recently the RVM formulation was introduced by Tipping [2001], introducing probabilistic predictions to the kernel model as in Equation (1). This formulation retains the robustness and performance of the SVM while improving the sparseness of the model [Bishop and Tipping, 2000].

In the case of classification, RVM models the conditional distribution of the label variable as a Gaussian distribution of the form:

$$P(t|x,w) = \sigma(y)^t[1-\sigma(y)]^{1-t} \tag{8.2}$$

where $\sigma(y) = (1+exp(-y))^{-1}$ and $y(x,w)$. The likelihood function in the classification case is given by:

$$P(T|X,w) = \prod_{n=1}^{N} \sigma(y_n)^{t_n}[1-\sigma(y_n)]^{1-t_n} \tag{8.3}$$

which cannot be solved analytically [Bishop and Tipping, 2000].

A fast marginal likelihood maximisation algorithm was proposed [Tipping et al., 2003], where RVM is initialized with an empty set of basis functions and the maximization of the marginal likelihood is achieved through efficient additions and deletions of candidate basis functions in the sparce model (Equation 1). In order to increase the marginal likelihood, basis functions are added to the model, the model is re-estimated (until convergence) and if a basis function becomes redundant ($\alpha_i = \infty$), it is removed from the model.

For this study, the fast RVM algorithm was implemented in C++ in order to be integrated to the OBIA framework. The objects obtained from the edge-enhanced MSEG algorithm, along with their computed features from the feature selection step, formed the input x of the RVM classification. A labeled set of training image objects was spatially overlaid on the final image segmentation. A selection of training objects X was obtained, using a rule of 75% overlap [Benz et al., 2004; Tzotsos and Argialas, 2008], along with their labels T. To implement multi-class RVM classification on x to obtain t, a one-against-all strategy was applied, since it has been shown to be very efficient in multiclass SVM classification cases [Hsu and Lin, 2002] and has also been applied successfully in

OBIA [Tzotsos and Argialas, 2008]. The training set X was set as an input to the RVM training module, providing a trained RVM model which included the relevance vectors - the samples for which the parameters are not discarded from the sparse model ($\alpha \neq \infty$). Relevance vectors are the equivalent of the support vectors in the SVM formulation. Finally, the complete set x of image primitives was given as input to the trained RVM, providing prediction (classification) set t. The predicted values t were imported into the RDBMS table as a class assignment, into a specifically defined column acting as current classification result, to be refined in following knowledge-based classification.

8.3.3 Knowledge-based Expert System Classification

It is a well-established process in OBIA methodology to refine classification results based on initial spectral classification (usually through Nearest Neighbor) and then using fuzzy rules as a knowledge-based classifier [Benz et al., 2004]. Usually this is implemented using classification-based segmentation and iterative classification steps as to extract the image semantics in single object representation (e.g. a mountain object, a house object, a building block object) [Argialas and Tzotsos, 2006; Hofmann, 2001].

In this effort, a similar process is implemented, by integrating *i)* a state-of-the-art machine learning classifier (RVM) instead of the Nearest Neighbor classifier, *ii)* a full Expert System environment instead of fuzzy classification and *iii)* a spatial RDBMS serving as the spatial engine for the spatial analysis of image objects. All these components are integrated through an object-oriented representation using the Free and Open Source (FOSS) programming language Python.

After the RVM classification (described above) the RDBMS included at least a spatial table, representing segmentation level at given scale. Also, the results of the RVM classifier were included as one record per image object. This can be represented in a GIS environment to provide spatial visualization of the current classification. In this effort the QGIS desktop environment [Sherman, 2008] has been used and both Spatialite and PostGIS [Obe and Hsu, 2011] systems have been tested as RDBMS back-ends.

In order to perform knowledge-based image classification for the proposed OBIA method, a Python application was developed, linking the vector results of the MSEG segmentation algorithm with the CLIPS Expert System library. CLIPS is a public domain software tool for building expert systems. The name is an acronym for "C Language

Integrated Production System". The syntax and name was inspired by OPS (Official Production System). The first versions of CLIPS were developed starting in 1985 at NASA-Johnson Space Center. The original name of the project was NASA's AI Language (NAIL) [Giarratano and Riley, 1998]. CLIPS is considered the reference implementation of the fast RETE search algorithm [Forgy, 1982], used in most modern reasoners and expert system shells. CLIPS supports only forward chaining strategy. The two main methods of reasoning when using an inference engine are Forward chaining and Backward chaining. Forward chaining starts with the available data (image objects and their properties) and uses inference rules to extract more data (class to which the object belongs) until a goal is reached (no more objects to classify). Backward chaining starts with a list of goals (or a hypothesis) and tries to find data to prove the hypothesis [Giarratano and Riley, 1998]. CLIPS and Forward chaining fit very well to the classification tasks in remote sensing since the initial data (pixels) are all available and can be evaluated to obtain the classification result.

The developed approach includes a three-tier software framework integrating:

- the spatial RDBMS holding the vector information,
- the object-oriented environment of the Python programming language and
- a Pythonic wrapper (PyCLIPS) around the CLIPS Expert System C library

After the image segmentation step, the feature space of the classification is formed, based on the spectral, shape, texture and spatial properties of the derived image objects. Those are stored in the vector database, but also a generic CLIPS class is automatically defined by the MSEG algorithm as output, in order to define those object properties for the CLIPS environment. For example, for a 4-band remote sensing image the following code is automatically generated:

```
(defclass IMG_OBJECT (is-a USER) (role concrete)
        (slot id (type INTEGER))
        (slot area (type INTEGER))
        (slot perimeter (type INTEGER))
        (slot compactness (type FLOAT))
        (slot mean-1 (type FLOAT))
```

```
        (slot mean-2 (type FLOAT))
        (slot mean-3 (type FLOAT))
        (slot mean-4 (type FLOAT))
        (multislot neighbors)
        (slot class-id (type INTEGER))
        (slot classification (type SYMBOL))
)
```

defining the image object class and the available slots. Then, the developed algorithm reads all image objects (records) from the RDBMS tables creating a list of Python objects with their attributes in working memory. Then PyCLIPS [Garosi, 2008; Löwe, 2013] is used to instantiate a CLIPS *instance* and populate the above *slots* per image object in working memory:

```
def load_objects_from_database():
    conn = db.connect("segments.db")
    conn.row_factory = db.Row
    cur = conn.cursor()
    SQL = "select * from Level"
    cur.execute(SQL)
    while True:
        r = cur.fetchone()
        if not r:
            conn.close()
            break
        res = {
            'id': clips.Integer(r['ID']),
            'area': clips.Integer(r['AREA']),
            'perimeter': clips.Integer(r['PERIMETER']),
            'compactness': clips.Float(r['COMPACTNESS']),
            'mean-1': clips.Float(r['MEAN1']),
            'mean-2': clips.Float(r['MEAN2']),
            'mean-3': clips.Float(r['MEAN3']),
```

```
            'mean-4': clips.Float(r['MEAN4']),
            'class-id': clips.Integer(r['CLASS'])
        }
        c = clips.FindClass("IMG_OBJECT")
        i = clips.BuildInstance(str(r['ID']), c)
        for k in res.keys():
            i.Slots[k] = res[k]
```

After the class instances are created in CLIPS, the knowledge base is loaded in the CLIPS environment and the the rules that need to be fired for all available objects are determined and loaded in the working memory.

```
clips.Load("rules.clp")
clips.Reset()
create_objects_from_database()
clips.Run()
save_classification_results_to_database()
```

As a final step the resulting instances are then imported to the Python environment and the classification column is updated in the database with the new value from CLIPS. An example of a rule is presented, that will classify all vegetation objects by computing the NDVI index:

```
(defrule find-vegetation
        ?seg <- (object
                        (is-a IMG_OBJECT)
                        (mean-4 ?nir)
                        (mean-3 ?red)
                )
        (test (> (/ (- ?nir ?red) (+ ?nir ?red 0.0001)) 0.0))
        =>
        (send ?seg put-classification "Vegetation")
        (send ?seg put-class-id 1)
)
```

A great advantage of the proposed methodology is the ability to perform native, multi-level spatial queries to the spatial RDBMS and then use the result for inference within the expert system environment. In general CLIPS is able to link to external functions in C language, but PyCLIPS extends this functionality, providing bindings for the Python language as well [Garosi, 2008]. The following rule will detect if vegetation is present on an other level of image objects (segmentation level above or bellow) intersecting with the current object:

```
(defrule find-vegetation-intersecting
        ?seg <- (object
                        (is-a IMG_OBJECT)
                        (id ?id)
                )
        (test (> (python-call py_find_vegetation ?id) 0))
        =>
        (send ?seg put-classification "Vegetation")
)
```

and the Python function for intersecting in PostGIS:

```
def py_find_vegetation(id):
    con = psycopg2.connect(database='testdb', user='user')
    cur = con.cursor()
    SQL = "SELECT ST_Intersection(cur.the_geom,ab.the_geom), cur.id "
    "FROM level_one as cur, level_two as ab WHERE cur.id = '%s'"
    "AND ab.class = 'Vegetation'" % (str(id))
    cur.execute(SQL)
    r = cur.fetchone()
    if r:
        return 1
    else:
        return 0
```

8.4 Evaluation and Discussion

The objective of the proposed research was to integrate a state-of-the-art machine learning clasifier such as the Relevance Vector Machines (RVM) in the OBIA framework as a method for supervised object classification. The second objective was to implement object-based image classification using an Expert System tool. Another goal of this research was to implement all the OBIA methodology using Free and Open Source Geospatial Software.

In this section, the results of the proposed methodology are presented, and a qualitative and quantitative evaluation is performed to determine how the proposed classification methods performed. In previous studies, the SVM classifier was previously compared to standard remote sensing classifiers [Tzotsos and Argialas, 2008] like Nearest Neighbor, this comparison was not considered for the paper. Instead, a comparison of SVM and RVM classification was performed to determine if previously reported pixel-based classification results Foody [2008] would be confirmed in the object-based case.

An advantage of the RVM classification method is the lack of parameter tuning. In SVM methods, usually a cross validation step is needed in order to determine the best parameters to perform the learning procedure by dividing the training set in training and validating parts. This step was not needed in RVM training, due to the integration of the Bayessian theory. Another advantage of the RVM classifier proved to be its sparseness. The final RVM model keeps a smaller number of relevance vectors in order to predict the class of a image object than the number of support vectors needed by SVMs. In this study, after the training step, for each class a range of [4-8] relevance vectors were used for RVM, while a range of [10-30] support vectors were used for SVM classifiers. This lead to better performance in classification decision for the RVM classifier: the decision of the class of an evaluated image object is much faster for RVM, since the relevance vectors involved in the decision are much fewer in number than support vectors. On the other hand, SVM models are less complex in training and perform better during that stage.

After the RVM classification, the classification results were stored in a spatial RDBMS, and the knowledge-based classifier was triggered, with a small number of rules, in order to determine if this would improve the resulting classification. The goal was not to perform classification with hundreds or thousands of rules (as a real expert system in production)

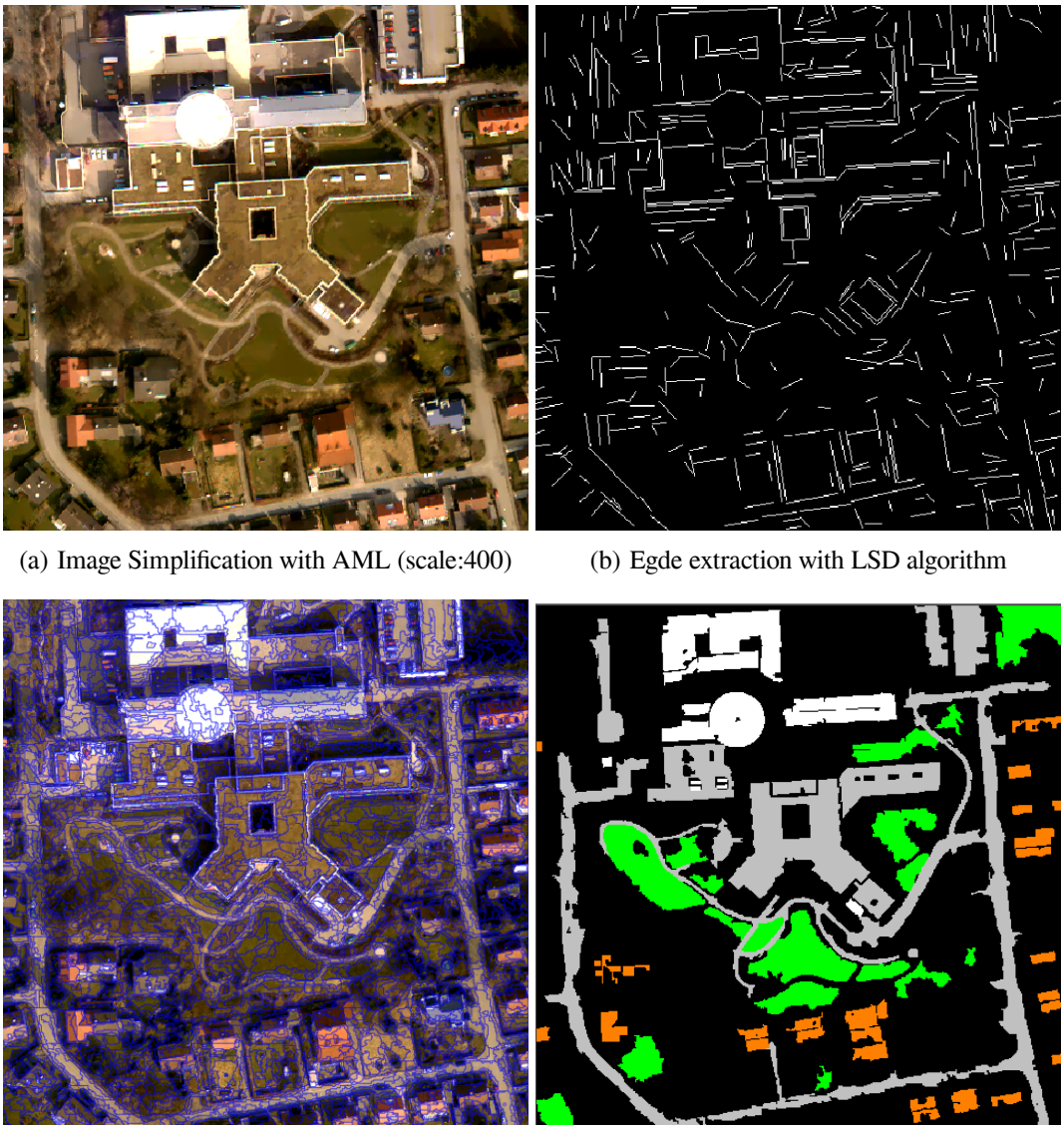
but to provide a proof-of-concept implementation to show the advantages of integrating machine learning with knowledge-based systems.

For the evaluation of results, two case studies are presented, involving both qualitative and quantitative evaluation criteria. First, a very high spatial resolution (0.5m) dataset from an airborne scanner was used to evaluate the classification robustness and accuracy, since OBIA approaches are known to perform better on very high resolution data [Blaschke, 2010]. Secondly, multispectral Landsat TM imagery was used to evaluate the proposed methodology.

8.4.1 Very high spatial resolution airborne imagery

First, the proposed OBIA methodology was applied to a very high resolution remote sensing imagery from an airborne digital scanner with 0.5m spatial resolution and 4-band spectral resolution. Then a scale-space simplification was performed, using the AML algorithm as shown in Fig.8.1(a). Due to high spatial resolution, a large value of scale parameter was selected (400) for the non-linear scale-space algorithm. Then, the LSD algorithm was applied to the simplified imagery, providing the edge information for the next segmentation step. The result of the LSD algorithm is presented in Fig.8.1(b). The detected edges are following straight lines around man-made objects, making the object detection easier for the segmentation step. The MSEG algorithm used the edge information as input and was able to provide primitive image objects at the same scale (400) as the scale-space filtering (Fig1.c). No tuning of any segmentation parameter was needed, since the heterogeneity of the image objects was reduced by the simplification step. Then, for each object, spectral and shape features were computed to form the feature space. The vector representation of image objects with their attributes were stored in a spatial database (Spatialite or PostGIS) and visualized in QGIS (Fig.8.2).

Then, a set of training data was selected and imported to both machine learning algorithms, so that training step to be preformed with the same input data. For the SVM classifier, a cross-validation step was performed, determining the best parameters by splitting the training set and use part for training and testing, until the best performance is reached. Based on the RBF kernel, the SVM model was trained and then connected with the test set (all the image objects available from the MSEG segmentation step). All objects were classified and a confusion matrix was computed based on the ground truth data



(a) Image Simplification with AML (scale:400)

(b) Egde extraction with LSD algorithm

(c) Image segmentation result with edge-enhanced MSEG (scale:400)

(d) Ground Truth data

Figure 8.1: Original multispectral aerial scan image (©Toposys) simplified by AML algorthm, edge extraction with Line Segment Detector, image segmentation with edge-enhanced MSEG algorithm and ground truth data for evaluation of further machine learning classification

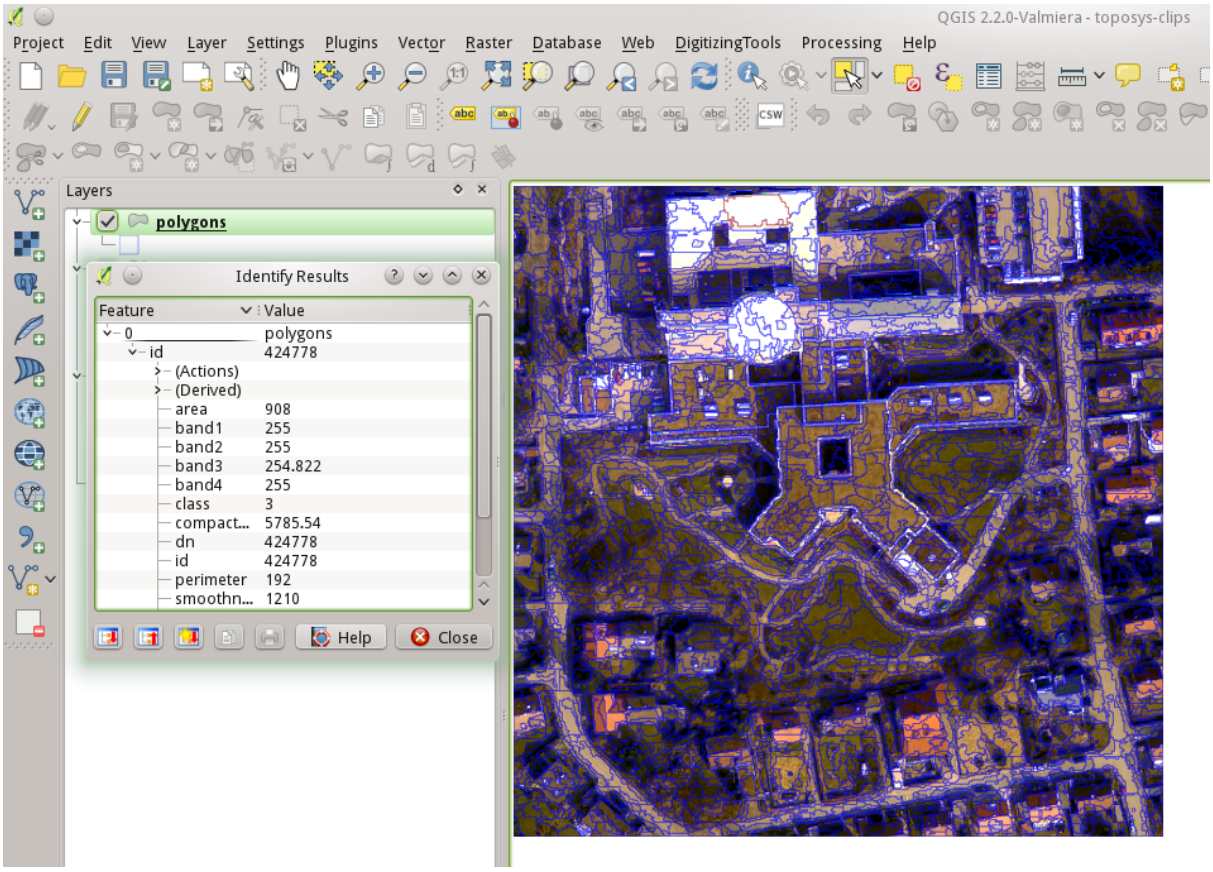
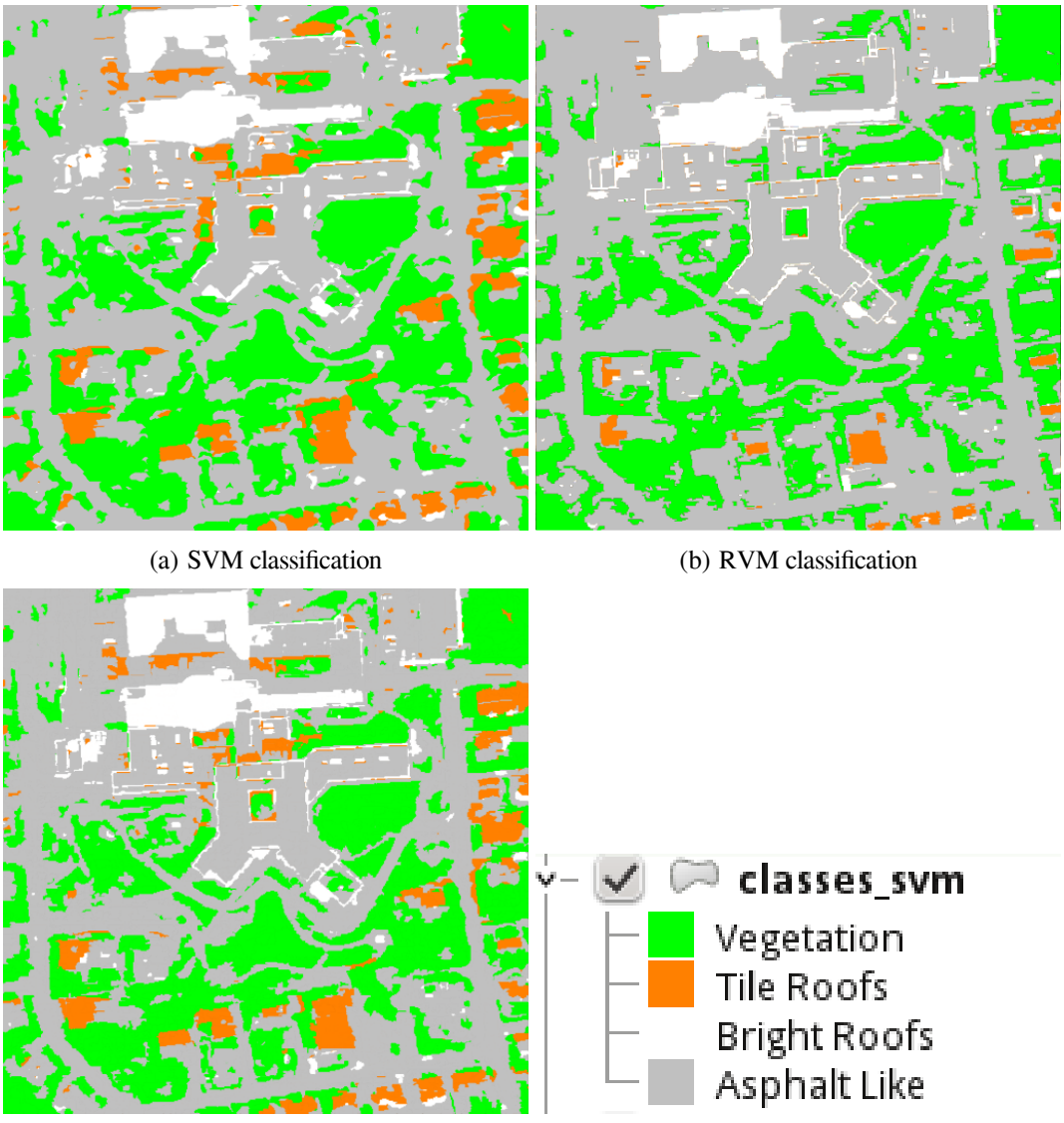


Figure 8.2: Image objects and their spectral and shape attributes visualized in QGIS.

available (Fig.8.1d). The accuracy of the SVM classifier was 91.34% showing that SVM is an excellent supervised classifier, even with a small training set. The SVM classification result is presented in Fig.8.3a. Then the RVM classifier was trained with the same data and after training, all image objects were given a class assignment (Fig.8.3b). An advantage of the RVM classifier is that due to its probabilistic nature, it provides posterior probabilities along with the class assignments for image objects, while SVM only provides a binary result. The RVM classification was compared to the ground truth data and the accuracy was determined to be 87.90%, comparative to the SVM result.

Then, the best classification result was selected to be further improved using knowledge-based refinement. The SVM classification result was imported to the CLIPS expert system and the following rules were applied:

```
(defrule fix-tile-roofs
        ?seg <- (object
                        (is-a IMG_OBJECT)
                        (mean-1 ?blue)
                        (compactness ?cmp)
                        (class-id 2)
                )
        (test (< ?blue 48.0))
        (test (> ?cmp 800.0))
        =>
        (send ?seg put-class-id 4)
)

(defrule fix-tile-roofs2
        ?seg <- (object
                        (is-a IMG_OBJECT)
                        (compactness ?cmp)
                        (class-id 2)
                )
        (test (> ?cmp 2000.0))
        =>
```



(a) SVM classification (b) RVM classification

(c) Knowledge-based classification (d) Class Legend

Figure 8.3: Machine Learning and knowledge-based classification of multispectral aerial scan image (©Toposys) with the developed OBIA framework.

```
        (send ?seg put-class-id 4)
)

(defrule fix-tile-roofs3
        ?seg <- (object
                        (is-a IMG_OBJECT)
                        (mean-1 ?blue)
                        (class-id 2)
                )
        (test (< ?blue 34.0))
        =>
        (send ?seg put-class-id 4)
)
```

The first rule uses the spectral value of the blue channel, as well as the compactness shape feature to determine if an object is elongated and dark. In that case the object is re-classified to the *Asphalt Like* class. Similarly, the other two rules are using specific values of the same spectral value and the compactness feature to do a re-classification. Those 3 rules implement a triple OR. The above rule can be expressed as: *IF Tile Roofs object is very elongated OR blue spectral value is too low OR the object is a bit elongated and has a small blue spectral value THEN re-classify as Asphalt Like*. The result of those 3 rules improves the final classification as can be seen in Fig3c. Lots of elongated objects that were previously defined as *Tile Roofs* are now fixed and re-assigned to the *Asphalt Like* class. The resulting classification accuracy was 91.76%. The improvement was due to the fact that *Tile Roofs* objects are limited in number compared to the objects of other classes. In the qualitative comparison of SVM classification result with the knowledge-based re-classification, it is shown that the *Tile Roofs* shape was improved.

Another advantage of a knowledge-based expert system is that it can provide full reports regarding the classification decisions, making it a very open tool for inspection and debugging, in contrast to machine learning methods which do not provide a way to easily inspect the internal processes. In the following text block, part of a report is presented, showing the assignment of an elongated *Tile Roof* object to *Asphalt Like*:

```
(defrule fix-tile-roofs
MSG << init ED:1 (<Instance-inst-424586>)
::= local slot perimeter in instance inst-424586 <- 222
::= local slot id in instance inst-424586 <- 424586
::= local slot mean-2 in instance inst-424586 <- 36.5566
::= local slot mean-3 in instance inst-424586 <- 2.88679
::= local slot mean-1 in instance inst-424586 <- 61.1179
::= local slot mean-4 in instance inst-424586 <- 115.349
::= local slot area in instance inst-424586 <- 212
::= local slot smoothness in instance inst-424586 <- 373.0
::= local slot compactness in instance inst-424586 <- 3232.37
::= local slot std-1 in instance inst-424586 <- 20.1962
::= local slot std-3 in instance inst-424586 <- 5.01285
::= local slot std-2 in instance inst-424586 <- 12.3595
::= local slot std-4 in instance inst-424586 <- 18.0445424586
::= local slot class-id in instance inst-424586 <- 2
==> Activation 0      fix-tile-roofs2: [inst-424586]
==> instance [inst-424071] of IMG_OBJECT
MSG >> create ED:1 (<Instance-inst-424071>)
HND >> create primary in class USER
       ED:1 (<Instance-inst-424071>)
FIRE   10 fix-tile-roofs2: [inst-424426]
MSG >> put-class-id ED:2 (<Instance-inst-424426> 4)
HND >> put-class-id primary in class IMG_OBJECT
       ED:2 (<Instance-inst-424426> 4)
::= local slot class-id in instance inst-424426 <- 4
HND << put-class-id primary in class IMG_OBJECT
       ED:2 (<Instance-inst-424426> 4)
MSG << put-class-id ED:2 (<Instance-inst-424426> 4)
```

8.4.2 Multispectral satellite imagery

The proposed OBIA methodology was also applied to a high resolution satellite Landsat TM multispectral image for evaluation. The dataset was filtered using the AML scale-space algorithm with scale parameter 400 (Fig.8.4a). Then, the Canny edge detector was applied (Fig.8.4b), due to the fact that man-made objects are not visible in a Landsat imagery, thus no straight lines were expected. Next, the edge-enhanced MSEG algorithm was applied with scale parameter 400 and no tuning of color and shape parameters. The derived image primitives are presented in Fig.8.4c. A feature extraction step was performed and object properties were computed to create the classification feature space.

Next a set of training samples was determined and RVM and SVM classifiers were trained. All objects were classified using both classifiers and confusion matrices were computed based on the ground truth data available (Fig.8.4d). The accuracy of the SVM classifier was 87.64% (Fig.8.5a) while the accuracy for RVM classifier was 85.79% (Fig.8.5b). Both classifiers are sparse, which is an excellent feature for OBIA applications, since training samples are usually small in number. Again, SVM outperformed RVM but the classification results are comparative, showing that RVM can be considered a classifier of great quality for OBIA applications. Additionaly, RVM is capable of producing probabilistic output. From Figure 8.5 it can be observed that the qualitative result of RVM tends to be better, especially for the *Waterbodies*.

The following step was to re-classify the previous machine learning classification result using a knowledge base:

```
(defrule fix-waterbodies1
        ?seg <- (object
                        (is-a IMG_OBJECT)
                        (class-id 4)
                        (mean-1 ?blue)
                        (mean-5 ?mir)
                        (mean-4 ?nir)
                )
        (test (< ?blue 70.0))
        (test (> ?mir 30.0))
```

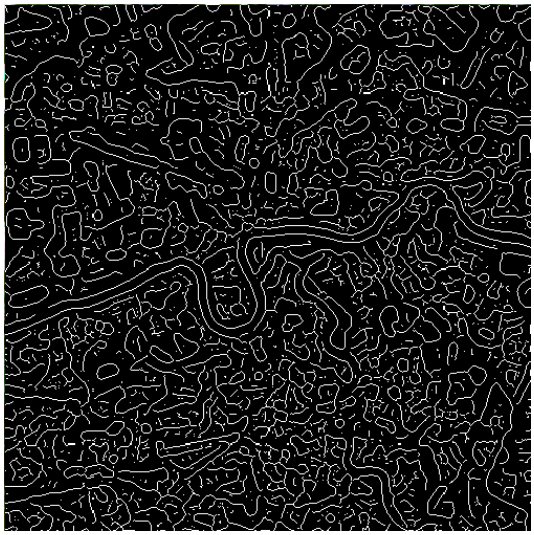
```
        =>
        (send ?seg put-class-id 3)
)

(defrule fix-waterbodies2
        ?seg <- (object
                        (is-a IMG_OBJECT)
                        (class-id 4)
                        (mean-1 ?blue)
                        (mean-5 ?mir)
                        (mean-4 ?nir)
                )
        (test (> ?nir 130.0))
        (test (> ?mir 30.0))
        =>
        (send ?seg put-class-id 2)
)
```

The above rules are used to correct classification results for the waterbodies, using spectral properties. The first rule will reassign all objects that have a range of blue band value (less than 70) and a range of middle infrared band value (more than 30) to *Impervious*. The second rule will re-classify waterbodies to grassland if the near-infrared values are high. After the knowledge-based classification step, a classification result was obtained (Fig.8.5c) and the resulting accuracy was 88.13%. It can be observed in Fig.8.5c that there are far less misclassified *Waterbodies* in the result obtained by CLIPS classification.

8.5 Conclusions and Future Perspectives

In this study, a state-of-the-art machine learning classification method, the Relevance Vector Machine was evaluated within the proposed Object-Based Image Analysis framework, showing promising results, of almost equal quality with the Support Vector Machine classifier. Previous RVM studies in remote sensing involved only pixel-based RVM

(a) Image Simplification with AML (scale:400)

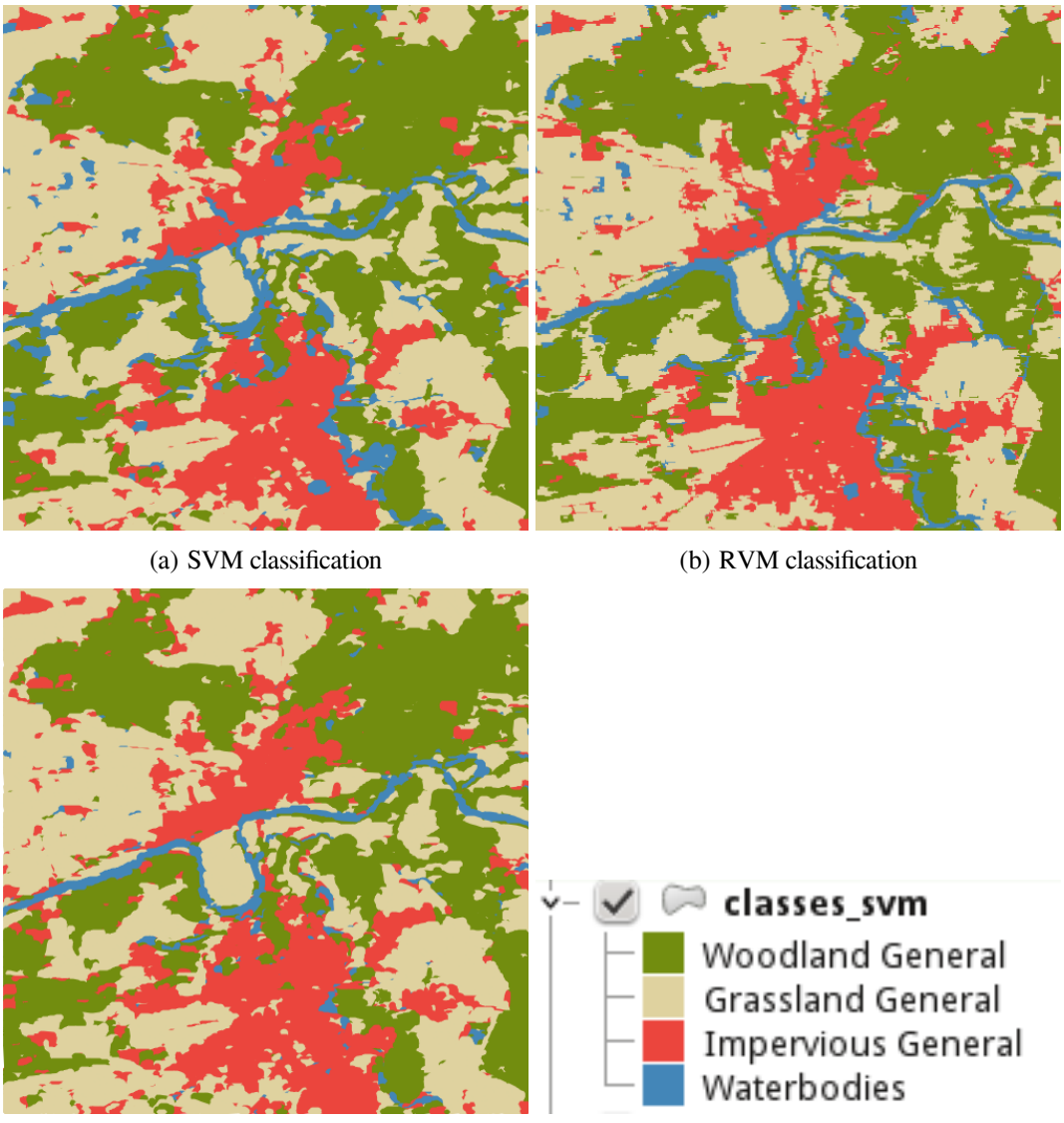


(b) Egde extraction with Canny algorithm

(c) Image segmentation result with edge-enhanced MSEG (scale:400)

(d) Ground Truth data

Figure 8.4: Original multispectral Landsat TM image (Dessau, Germany) simplified with scale-space AML algorithm, edge extraction with Canny edge detector, image segmentation with edge-enhanced MSEG algorithm and ground truth data for classification evaluation.



(a) SVM classification (b) RVM classification

(c) Knowledge-based classification (d) Class Legend

Figure 8.5: Machine Learning and knowledge-based classification of multispectral Landsat TM image (Dessau, Germany) with the developed OBIA framework.

classification and regression. The second contribution of this paper was the integration of an Knowledge-Based Expert System tool for object-based image classification.

RVM was shown to be very sparse and faster in prediction performance from SVM, while being independent of parameter tuning. An important feature of RVM is the probabilistic output, which can be integrated with other classification methods that can handle uncertainty. SVM has outperformed RVM in classification accuracy, but the margin was not large. Given the above advantages of RVM, it can be considered an excellent classifier for OBIA. A small issue of RVM is the long training times, which can be improved by recent fast implementations in the literature.

The CLIPS expert system was integrated in the developed OBIA framework, covering and completing an important missing part of OBIA methodology. A pure expert system has the advantage of being a very generic tool and easily extended. In this research, CLIPS was extended to use spatial operators, though integration with spatial database systems and an object-oriented programming language (Python).

It was shown that as a concept, the integration of human knowledge into an image analysis framework can refine and improve results, even with a small number of rules and heuristics by the expert user. The knowledge examples shown here, were a proof-of-concept of what a rule based system can do. A full expert system application would need to include far more in rule numbers and complexity. Through the performed qualitative and quantitative evaluation it was suggested that multi-modal classification systems integrating machine learning and knowledge-based systems are are very promising and can provide high quality results. It is important to mention that all the components of the proposed OBIA framework are Free and Open Source Software, promoting the notion of open science for widely available, extensible and reproducible results.

Some of the topics for further research and development are: comparison of RVM and SVM methods with other state-of-the-art machine learning algorithms such as ensembles and random forests, evaluation of fuzzy SVMs and evaluation of unsupervised learning methods. Finally there are a lot of opportunities for investigation of expert systems and OBIA for complex remote sensing applications.

Chapter 9

Conclusions

The proposed methodology, as described in the previous chapters makes significant contributions on all aspects of the Object-Based Image Analysis (OBIA) methodology:

- Preprocessing: Scale-space and anisotropic filtering was proposed and implemented for image simplification before the multiscale segmentation step.
- Image Segmentation: Three new image segmentation approaches have been proposed and implemented, with the edge-based MSEG being the most successful according to the evaluation of results.
- Computation of feature space: A run-length encoding was proposed and implemented for object representation in MSEG, minimizing the space needed to store the geometric features and optimizing the translation from raster to vector data representation.
- Definition of object-oriented class hierarchy and representation of knowledge: The CLIPS language has been used for knowledge representation and has been integrated to OBIA methodology. Automation of image object representation in CLIPS has been implemented in MSEG.
- Classification: Novel approaches have been introduced in this dissertation, such as SVM, RVM and Knowledge-based Expert Systems for object-based classification. Also an integrated classification framework using MSEG, scale-space and machine learning has been proposed, removing the need for parameter tuning.

- Accuracy assessment: Machine learning methods have been integrated and evaluated in OBIA for accuracy assessment, like the cross-validation methods in SVM classification.

- Vectorization: Computer vision, machine learning and knowledge-based methods have been integrated with a spatial database, linking the raster (image objects) and vector space (GIS objects within the spatial database).

The main objective of this research was to implement a novel image classification framework being able to process effectively a wide variety of remote sensing data. As shown in Chapters 5 and 6, a novel OBIA classification framework was proposed, implemented and evaluated for a wide variety of remote sensing data. A new object-based classification method was developed based on advanced multilevel object representations and a support vector machine classifier. In contrast to previous efforts, the multilevel object representation was based primarily on the advanced simplification procedure and not on the region merging process. The proposed OBIA framework has been shown to be stable, fast and can efficiently account for various classification tasks in various types of remote sensing data (Chapter 5 and 6).

Another objective of this research was to introduce state-of-the-art computer vision (such as scale-space, advanced edge features etc.) and machine learning methods (such as Support Vector Machines, Relevance Vector Machines etc.) to the OBIA approach, improving the OBIA image segmentation and object classification steps. Both scale-space and advanced edge features were integrated in the image segmentation process as shown in Chapters 5 and 6, leading to promising segmentation results.

The SVM classification was introduced in OBIA for the first time in this research. Overall, the SVM classification methodology was found very promising for Object-Based Image Analysis. It has been shown that it can produce better results than the Nearest Neighbor for supervised classification. A very good feature of the SVM is that only a small sample set is needed to provide very good results, because only the support vectors are of importance during classification. The computational efficiency of SVM was great, with only a few seconds of runtime necessary for training. This was theoretically expected but also, the implementation in C++ is extremely fast.

In this study, a state-of-the-art Relevance Vector Machine classification method was introduced and evaluated within the Object-Based Image Analysis framework, show-

ing promising results, comparable with the Support Vector Machine classifier. Previous RVM studies in remote sensing involved only pixel-based RVM classification and regression. As a secondary contribution, a probabilistic SVM implementation was also evaluated within the OBIA framework, with very promising results, almost similar with the binary SVM classifier. Based on the above evaluation (Chapter 7) and given that RVM was always faster in prediction execution times, while being more sparse and being free from any model parameter tuning, RVM can be considered as an excellent choice as an OBIA classifier, especially in cases when the sample/training size is very small. This is a valuable feature since OBIA supervised classification usually involves a small set of training samples/objects. A major feature of both proposed algorithms is the uncertainty handling through posterior probabilities computation. On the other hand, long training time for RVM is an obvious disadvantage, although various optimizations have already been proposed and implemented.

An additional objective of this research was the development and implementation of a new object-oriented image segmentation algorithm as a low level processing part of an object-oriented image analysis system so that to be applied at multiple image resolutions and to produce objects of multiple scales (sizes), minimizing the need of user-customizable parameters.

Overall, the designed multi-scale image segmentation algorithm (Chapter 2), gave very promising segmentation results for remote sensing imagery, providing parameters to set the desired size of image objects along with shape constraints. With the addition of the Advanced Texture Heuristic module (Chapter 3) and the Advanced Edge Features (Chapter 6), it was shown to be a good and generic segmentation solution for remote sensing imagery. The extracted boundaries of the primitive image objects, in each case, approximated the semantic object boundaries more successfully than previous segmentation efforts.

As a conclusion for image segmentation, a new object-based image analysis framework was proposed and developed, based on advanced edge features incorporated in a multiscale region merging algorithm. Advanced scale-space representations were used in order to avoid manual tuning of segmentation and feature extraction parameters. Kernel-based classification was implemented to complete the OBIA framework. The developed image segmentation algorithm was shown to work on any type of remote sensing data, outperforming some widely used segmentation algorithms in certain cases. The

improvement of the MSEG segmantation results was demonstrated (Chapter 6), and the edge enhancements were shown to make the algorithm robust and generic for multiscale OBIA applications. The performed qualitative and quantitative evaluation reported that the developed algorithm outperformed previous efforts, both regarding the construction of object representations and classification results. In terms of performance, the edge-based improvement of the MSEG algorithm had a major impact on execution times. For the very high resolution airborne images (Fig.6.1a), execution time for original MSEG algorithm was 7-8 seconds on a 3 GHz PC with 4 GB of RAM running GNU/Linux operating system. With the edge-based optimization, the number of possible object merges droped significantly per segmentation pass, and the execution time was around 4-5 seconds. Still, there is room for speed optimization improvements, since for the same image, the execution times for multiresolution segmentation (eCognition) was 2-3 seconds and for Mean Shift (Orfeo Toolbox) was 3-4 seconds. For object oriented image analysis, the edge-based MSEG is qualified as a successful low level processing algorithm.

The employed scale space formulation that was designed and implemented, possesses a number of desired qualitative properties and thus eliminated the need for tuning several parameters during segmentation. The performed qualitative and quantitative evaluation reported that the developed algorithm outperformed previous efforts, both regarding the construction of the object representations and the classification results.

Another goal of this research was the integration of a Knowledge-based Expert System to the developed segmentation and classification framework so that to improve the OBIA classification task. The CLIPS expert system was integrated in the proposed OBIA framework, covering and completing an important missing part of OBIA methodology. A pure expert system has the advantage of being a very generic and easily extended tool. In this research, CLIPS was extended to use spatial operators, though integration with spatial database systems and an object-oriented programming language (Python).

It was shown that as a concept, the integration of human knowledge into an image analysis framework can, in principle, refine and improve results, even with a small number of rules and heuristics by the expert user. The knowledge examples shown here, were a proof-of-concept of what a rule based system can do. A full expert system application would need to include larger number of rules of greater complexity. Through the performed qualitative and quantitative evaluation (Chapter 8) it was suggested that multimodal classification systems integrating machine learning and knowledge-based systems

are very promising and can provide high quality results.

Last but not least, a goal of this research was to provide an Object-Based Image Analysis system in the form of Free and Open-Source Software (FOSS). The developed OBIA framework is composed entirely by free (as in freedom) software. For the implementation of this research a number of Free and Open Source libraries were developed (MSEG, cvAML) or integrated (pyCLIPS, PostGIS, SpatiaLite, libSVM, OrfeoToolbox, QGIS and GDAL). The coding was performed in C++ and Python programming languages. It is important to mention that all the components and tools implemented or used for this dissertation were Free and Open Source Software, promoting the notion of open science for widely available, extensible and reproducible results.

Some of the topics for further research and development are: comparison of RVM and SVM methods with other state-of-the-art machine learning algorithms such as ensembles and random forests, evaluation of fuzzy SVMs and unsupervised learning methods. Finally there are a lot of opportunities for investigation of expert systems, machine learning and the proposed OBIA framework for complex remote sensing applications.

References

H.G. Akcay and S. Aksoy. Automatic detection of geospatial objects using multiple hierarchical segmentations. *Geoscience and Remote Sensing, IEEE Transactions on*, 46 (7):2097 --2111, july 2008. ISSN 0196-2892. doi: 10.1109/TGRS.2008.916644. 98

L. Alvarez, P.-L. Lions, and J.-M. Morel. Image selective smoothing and edge detection by nonlinear diffusion.II. *SIAM Journal on Numerical Analysis*, 29 (3)(3):845--866, 1992. 74, 78

P. Aplin and G.M. Smith. Advances in object-based image classification. *Interanational Archive of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37 (Part B7)(B7):725--728, 2008. 4, 64, 96

Fabrizio Argenti, Luciano Alparone, and Giuliano Benelli. Fast algorithms for texture analysis using co-occurrence matrices. In *Radar and Signal Processing, IEE Proceedings F*, volume 137, pages 443--448. IET, 1990. 40

D Argialas and V Goudoula. Knowledge-based land use classification from ikonos imagery for arkadi, crete, greece. remote sensing for environmental monitoring, gis applications, and geology ii. In *Proceedings of SPIE*, volume 4886, 2003. 48

D. Argialas and C. Harlow. Computational Image Interpretation Models: An Overview and a Perspective. *Photogrammetric Engineering and Remote Sensing*, 56 (6)(6):871--886, 1990. 3, 32, 48, 64, 149

Demetre Argialas and Angelos Tzotsos. Automatic extraction of physiographic features and alluvial fans in nevada, usa, from digital elevation models and satellite imagery

through multiresolution segmentation and object-oriented classification. *Proceedings of ASPRS 2006 Annual Conference, Reno, Nevada; May 1-5*, 2006. 7, 157

Demetre P Argialas and Angelos Tzotsos. Automatic extraction of aluvial fans from aster l1 satellite data and a digital elevation model using object-oriented image analysis. In *XXth ISPRS Congress, Istanbul, Turkey*, 2004. 36

Damien Arvor, Laurent Durieux, Samuel Andrés, and Marie-Angélique Laporte. Advances in geographic object-based image analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82:125--137, 2013. 147, 149, 150

M. Baatz and A. Schape. Multiresolution segmentation an optimization approach for high quality multi-scale image segmentation. In *Strobl, J. et al. (eds.): Angewandte Geographische Infor-mationsverarbeitung XII.*, pages 12--23. Wichmann, Heidelberg, 2000. 4, 5, 7, 12, 13, 18, 19, 25, 33, 35, 49, 50, 66, 69, 70, 75, 79, 97, 104, 109, 122, 126, 129, 146, 149, 154

Andrea Baraldi and Luigi Boschetti. Operational automatic remote sensing image understanding systems: Beyond geographic object-based and object-oriented image analysis (geobia/geooia). part 1: Introduction. *Remote Sensing*, 4(9):2694--2735, 2012. 149

Peter Baumann. The ogc web coverage processing service (wcps) standard. *Geoinformatica*, 14(4):447--479, 2010. 2, 122

U.C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58 (3-4)(3-4):239--258, 2004. 4, 5, 12, 14, 26, 33, 34, 41, 49, 50, 54, 55, 64, 66, 69, 70, 95, 96, 97, 109, 122, 126, 132, 133, 134, 135, 146, 149, 156, 157

James O Berger. *Statistical decision theory and Bayesian analysis*. Springer, 1985. 131

J Berni, Pablo J Zarco-Tejada, Lola Suárez, and Elias Fereres. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(3):722--738, 2009. 2, 122

Irving Biederman. Human image understanding: Recent research and a theory. *Computer vision, graphics, and image processing*, 32(1):29--73, 1985. 50

Christopher M Bishop and Michael E Tipping. Variational relevance vector machines. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, pages 46--53. Morgan Kaufmann Publishers Inc., 2000. 130, 131, 132, 155, 156

T. Blaschke. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65 (1)(1):2--16, 2010. 2, 3, 4, 5, 64, 95, 96, 97, 122, 126, 128, 135, 146, 149, 150, 163

T. Blaschke and G. Hay. Object-oriented image analysis and scale-space: theory and methods for modeling and evaluating multi-scale landscape structure. *Interanational Archive of Photogrammetry and Remote Sensing*, 34 (Part 4/W5):22--29, 2001. 4, 64, 66, 70, 95, 96, 98

T. Blaschke, C. Burnett, and A. Pekkarinen. Image segmentation methods for object-based analysis and classification. In *de Jong, S.M., van der Meer, F.D. (Eds.), Remote Sensing and Digital Image Analysis: Including the Spatial Domain (Chapter 12)*, pages 211--236. Kluwer Academic Publishers, 2004. 66, 70, 97

T. Blaschke, S. Lang, and G. Hay. *Object Based Image Analysis - Spatial concepts for knowledge driven remote sensing applications*. New York: Springer, 2008. 4, 5, 64, 65, 96, 98

Andreas Ch Braun, Uwe Weidner, and Stefan Hinz. Support vector machines, import vector machines and relevance vector machines for hyperspectral classification—a comparison. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, pages 1--4. IEEE, 2011. 126

Gunnar Jakob Briem, Jon Atli Benediktsson, and Johannes R Sveinsson. Multiple classifiers applied to multisource remote sensing data. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(10):2291--2299, 2002. 125

Martin Brown, Hugh G Lewis, and Steve R Gunn. Linear spectral mixture models and support vector machines for remote sensing. *Geoscience and Remote Sensing, IEEE Transactions on*, 38(5):2346--2360, 2000. 49

Gustavo Camps-Valls. Machine learning in remote sensing data processing. In *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on*, pages 1--6. IEEE, 2009. 123, 124, 150

Gustavo Camps-valls and Lorenzo Bruzzone. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43 (6): 1351--1362, 2005. 66, 67, 99, 124, 125, 151

Gustavo Camps-Valls, Luis Gómez-Chova, Jordi Muñoz-Marí, Joan Vila-Francés, Julia Amorós-López, and Javier Calpe-Maravilla. Retrieval of oceanic chlorophyll concentration with relevance vector machines. *Remote Sensing of Environment*, 105(1):23--33, 2006. 123, 125, 151

Gustavo Camps-Valls, Antonio Rodrigo-González, J Muoz-Mari, Luis Gómez-Chova, and Javier Calpe-Maravilla. Hyperspectral image classification with mahalanobis relevance vector machines. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 3802--3805. IEEE, 2007. 123, 125, 126, 151

Gustavo Camps-Valls, Luis Gómez-Chova, Jordi Muñoz-Marí, José Luis Rojo-Álvarez, and Manel Martínez-Ramón. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(6):1822--1835, 2008. 125, 151

J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679--698, 1986. 102, 103

A.P. Carleer, O. Debeir, and E. Wolff. Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering and Remote Sensing*, 71 (11):1285--1294, 2005. 66, 70, 97

Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011. 54

Junqing Chen, Thrasyvoulos N Pappas, Aleksandra Mojsilovic, and Bernice Rogowitz. Adaptive image segmentation based on color and texture. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 3, pages 777--780. IEEE, 2002. 13, 33

Alex Yong-Sang Chia, Deepu Rajan, Maylor Karhang Leung, and Susanto Rahardja. Object recognition by discriminative combinations of line segments, ellipses, and appearance features. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34 (9):1758 --1772, sept. 2012. ISSN 0162-8828. doi: 10.1109/TPAMI.2011.220. 100, 101

E. Christophe and J. Inglada. Open source remote sensing: Increasing the usability of cutting-edge algorithms. *IEEE Geoscience and Remote Sensing Newsletter*, 2009. 97, 126

Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603--619, 2002. 110

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20 (3):273--297, 1995. 51

Daniel Crevier and Richard Lepage. Knowledge-based image understanding systems: a survey. *Computer Vision and Image Understanding*, 67(2):161--185, 1997. 147, 148, 149

X. Cufi, X. Munoz, J. Freixenet, and J. Marti. A review of image segmentation techniques integrating region and boundary information. *Advances in imaging and electron physics*, 120:1--39, 2003. 101, 105

Ronei Marcos De Moraes. An analysis of the fuzzy expert systems architecture for multispectral image classification using mathematical morphology operators. *International Journal of Computational Cognition*, 2(2):35--69, 2004. 49

Fabio Del Frate, Fabio Pacifici, Giovanni Schiavon, and Chiara Solimini. Use of neural networks for automatic classification from high-resolution images. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(4):800--809, 2007. 124

Begüm Demir and Sarp Erturk. Hyperspectral image classification using relevance vector machines. *Geoscience and Remote Sensing Letters, IEEE*, 4(4):586--590, 2007. 125, 151

S. Derrode and G. Mercier. Unsupervised multiscale oil slick segmentation from sar images using a vector hmc model. *Pattern Recognition*, 40(3):1135--1147, 2007. 98

Jefersson Alex Dos Santos, P-H Gosselin, Sylvie Philipp-Foliguet, Ricardo da S Torres, and Alexandre X Falcão. Multiscale classification of remote sensing images. 2012. 127

G. Doxani, K. Karantzalos, and M. Tsakiri-Strati. Monitoring urban changes based on scale-space filtering and object-oriented classification. *International Journal of Applied Earth Observation and Geoinformation*, 15(0):38 -- 48, 2012. ISSN 0303-2434. doi: 10.1016/j.jag.2011.07.002. 98

Lucian Dragut, Dirk Tiede, and Shaun R Levick. Esp: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science*, 24(6):859--871, 2010. 4, 96

Lucianand Dragut, Thomas Schauppenlehner, Andreas Muhar, Josef Strobl, and Thomas Blaschke. Optimization of scale and parametrization for terrain segmentation: An application to soil-landscape modeling. *Computers & Geosciences*, 35 (9):1875--1883, 2009. 66, 69, 70, 97, 126

J.M. Duarte-Carvajalino, P.E. Castillo, and M. Velez Reyes. Comparative study of semi-implicit schemes for nonlinear diffusion in hyperspectral imagery. *IEEE Transactions on Image Processing*, 16 (5):1303--1314, 2007. 67, 99

J.M. Duarte Carvajalino, G. Sapiro, M. Velez Reyes, and P.E. Castillo. Multiscale representation and segmentation of hyperspectral imagery using geometric partial differential equations and algebraic multigrid. *IEEE Transactions on Geoscience and Remote Sensing*, 46 (8):2418--2434, 2008. 4, 64, 95, 96

Hongliang Fang and Shunlin Liang. Retrieving leaf area index with a neural network method: Simulation and validation. *Geoscience and Remote Sensing, IEEE Transactions on*, 41(9):2052--2062, 2003. 49

Mathieu Fauvel, Jón Atli Benediktsson, Jocelyn Chanussot, and Johannes R Sveinsson. Spectral and spatial classification of hyperspectral data using svms and morphological

profiles. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(11):3804--3814, 2008. 125, 151

Mohammad FA Fauzi and Paul H Lewis. A fully unsupervised texture segmentation algorithm. 2003. 13, 15, 33

Giles M Foody. Rvm-based multi-class classification of remotely sensed data. *International Journal of Remote Sensing*, 29(6):1817--1823, 2008. 123, 125, 134, 138, 151, 152, 162

G.M. Foody and A. Mathur. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions On Geoscience And Remote Sensing*, 42 (6):1335--1343, 2004. 49, 70, 123, 124, 151

Germain Forestier, Anne Puissant, Cédric Wemmert, and Pierre Gançarski. Knowledge-based region labeling for remote sensing image interpretation. *Computers, Environment and Urban Systems*, 36(5):470--480, 2012. 149, 150

Charles L Forgy. Rete: A fast algorithm for the many pattern/many object pattern match problem. *Artificial intelligence*, 19(1):17--37, 1982. 158

Keinosuke Fukunaga and Raymond R. Hayes. Effects of sample size in classifier design. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(8):873--885, 1989. 124

Francesco Garosi. Pyclips manual, 2008. http://pyclips.sourceforge.net/manual/pyclips.html. 159, 161

J Giarratano and G Riley. Expert systems: principles and programming. *PWS-Kent, Boston, MA*, 1998. 147, 148, 153, 158

G. Gilboa. Nonlinear scale space with spatially varying stopping time. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(12):2175 --2187, dec. 2008. ISSN 0162-8828. 100

Ryan N Givens, Karl C Walli, and Michael T Eismann. A multimodal approach to high resolution image classification. In *Applied Imagery Pattern Recognition Workshop: Sensing for Control and Augmentation, 2013 IEEE (AIPR*, pages 1--7. IEEE, 2013. 150

Luis Gómez-Chova, Jordi Muñoz-Marí, Valero Laparra, Jesús Malo-López, and Gustavo Camps-Valls. A review of kernel methods in remote sensing data analysis. In *Optical Remote Sensing*, pages 171--206. Springer, 2011. 123, 124, 125, 147, 150, 151

Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid. Multimodal semi-supervised learning for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 902--909. IEEE, 2010. 150

Ola Hall and Geoffrey J. Hay. A multiscale object-specific approach to digital change detection. *International Journal of Applied Earth Observation and Geoinformation*, 4 (4)(4):311--327, 2003. 4, 64, 66, 70, 95, 96, 98

Jisoo Ham, Yangchi Chen, Melba M Crawford, and Joydeep Ghosh. Investigation of the random forest framework for classification of hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(3):492--501, 2005. 125

Robert M Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786--804, 1979. 33, 39

Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6): 610--621, 1973. 33, 36, 38

Trevor Hastie and Robert Tibshirani. Classification by pairwise coupling. *The annals of statistics*, 26(2):451--471, 1998. 133

Joseph P Havlicek and Peter C Tay. Determination of the number of texture segments using wavelets. *Electronic Journal of Differential Equations*, pages 61--70, 2001. 13, 33

Stephen Hawley. Ordered dithering. In *Graphics gems*, pages 176--178. Academic Press Professional, Inc., 1990. xxv, 17

G. Hay and G. Castilla. Object-based image analysis: Strengths, weaknesses, opportunities and threats. *Interanational Archive of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (4/C42)(4/C42):on CD--ROM, 2006. 5, 65, 98

G. Hay, G. Castilla, M. Wulder, and J. Ruiz. An automated object-based approach for the multiscale image segmentation of forest scenes. *International Journal of Applied Earth Observation and Geoinformation*, 7 (4)(4):339--359, 2005. 70, 97

G. J. Hay, P. Dub, A. Bouchard, and D. J. Marceau. A scale-space primer for exploring and quantifying complex landscapes. *Ecological Modelling*, 153 (1-2)(1-2):27 -- 49, 2002. 4, 64, 66, 70, 95, 96, 98

G.J. Hay, T. Blaschke, D.J. Marceau, and A. Bouchard. A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57 (5-6)(5-6):327--345, 2003. 66, 70, 98

P. Hofmann, J. Strobl, and T. Blaschke. A method for adapting global image segmentation methods to images of different resolutions. In *International Conference on Geographic Object-Based Image Analysis. ISPRS Volume 38 (on CD-ROM)*, volume XXXVIII, 2008. 5, 65, 98

PCRS Hofmann. Detecting urban features from ikonos data using an object-oriented approach. In *Proceedings of the first annual conference of the remote sensing & photogrammetry society*, pages 79--91, 2001. 157

C.-W. Hsu and C.-J. Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions On Neural Networks*, 13 (2):415--425, 2002. 53, 54, 77, 107, 125, 132, 134, 156

C. Huang, L. S. Davis, and J. R. G. Townshend. An assessement of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23 (4): 725--749, 2002a. 49, 54, 70, 124, 151

C Huang, LS Davis, and JRG Townshend. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725--749, 2002b. 125, 151

G Hughes. On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on*, 14(1):55--63, 1968. 124

J. Inglada and E. Christophe. The orfeo toolbox remote sensing image processing software. In *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, volume 4, pages IV--733. IEEE, 2009. 97, 109, 126

Jordi Inglada. Automatic recognition of man-made objects in high resolution optical remote sensing images by svm classification of geometric image features. *ISPRS journal of photogrammetry and remote sensing*, 62(3):236--248, 2007. 125, 151

Peter Jackson. *Introduction to expert systems*. Addison-Wesley Longman Publishing Co., Inc., 1990. 148, 149, 150

Hao Jiang, S.X. Yu, and D.R. Martin. Linear scale and rotation invariant matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(7):1339 --1355, july 2011. ISSN 0162-8828. doi: 10.1109/TPAMI.2010.212. 100

L. O. Jimenez, J. L. Rivera-Medina, E. Rodriguez-Diaz, E. Arzuaga-Cruz, and M. Ramirez-Velez. Integration of spatial and spectral information by means of unsupervised extraction and classification for homogenous objects applied to multispectral and hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43 (4): 844--851, 2005. 4, 64, 70, 95, 96, 97

C.R. Jung. Unsupervised multiscale segmentation of color images. *Pattern Recognition Letters*, 28(4):523--533, 2007. 98

I. Kanellopoulos, G. Wilkinson, and T. Moons. *Machine Vision and Advanced Image Processing in Remote Sensing*. Springer Verlag, 1999. 3, 48

K Karantzalos. A 4D Morphological Scale Space Representation for Hyperspectral Imagery. *Interanational Archive of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37:127--132, 2008. 66

K. Karantzalos and D. Argialas. Improving edge detection and watershed segmentation with anisotropic diffusion and morphological levelings. *International Journal of Remote Sensing*, 27 (24):5427--5434, 2006. 67, 74, 95, 99, 128

K. Karantzalos, D. Argialas, and N. Paragios. Comparing morphological levelings constrained by different markers. In *ISMM, G.Banon, et al. (eds), Mathematical Morphol-*

ogy and its Applications to Signal and Image Processing, pages 113--124, 2007. 65, 66, 67, 68, 74, 99, 102, 103, 127, 128, 152, 153

C.D. Kermad and K. Chehdi. Automatic image segmentation system through iterative edge-region co-operation. *Image and Vision Computing*, 20(8):541--555, 2002. 101, 105

R. Kimmel, Cuiping Zhang, A.M. Bronstein, and M.M. Bronstein. Are mser features really interesting? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(11):2316 --2320, nov. 2011. ISSN 0162-8828. doi: 10.1109/TPAMI.2011.133. 100

S. Klonus, D. Tomowski, M. Ehlers, P. Reinartz, and U. Michel. Combined edge segment texture analysis for the detection of damaged buildings in crisis areas. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(4):1118 --1128, aug. 2012. ISSN 1939-1404. doi: 10.1109/JSTARS.2012.2205559. 101

J. Koenderink. The structure of images. *Biological Cybernetics*, 50 (5):363--370, 1984. 5, 65, 98

G. Lehmann. Label object representation and manipulation with itk. *Insight J*, pages 1--34, 2008. 107

M. Lennon, G. Mercier, and L. Hubert-Moy. Classification of hyperspectral images with nonlinear filtering and support vector machines. In *IEEE International Geoscience and Remote Sensing Symposium*, volume 3, pages 1670--1672, 2002. 66, 67, 99

S Liapis, N Alvertos, and G Tziritas. Unsupervised texture segmentation using discrete wavelet frames. In *IX European Signal Processing Conference*, pages 2529--2532, 1998. 13, 15, 33

T.M. Lillesand and R.W. Kiefer. *Remote-Sensing and Image Interpretation*. Wiley, New York, 1987. 50

T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Dordrecht, 1994. 3, 5, 65, 94, 98

Yu Liu, Qinghua Guo, and Maggi Kelly. A framework of region-based spatial relations for non-overlapping features and its application in object based image analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63 (4)(4):461 -- 475, 2008. 4, 64, 96

Peter Löwe. g. infer: A grass gis module for rule-based data-driven classification and workflow control. In *EGU General Assembly Conference Abstracts*, volume 15, page 13713, 2013. 159

T.R. Martha, N. Kerle, C.J. van Westen, V. Jetten, and K.V. Kumar. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(12): 4928 --4943, dec. 2011. ISSN 0196-2892. doi: 10.1109/TGRS.2011.2151866. 98

Andrzej Materka, Michal Strzelecki, et al. Texture analysis methods--a review. *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, pages 9--11, 1998. 33, 36, 38

G. Matheron. Les Nivellements. Technical report, Centre de Morphologie Mathematique, France, 1997. 67, 99

Ujjwal Maulik and Debasis Chakraborty. A self-trained ensemble with semisupervised svm: An application to pixel classification of remote sensing imagery. *Pattern Recognition*, 44(3):615--623, 2011. 150

F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions On Geoscience And Remote Sensing*, 42 (8):1778-- 1790, 2004. 49, 70

G. Mercier and M. Lennon. Support vector machines for hyperspectral image classification with spectral-based kernels. In *IGARSS, Toulouse, France, July 21 25*. IGARSS p.(on CD-ROM), 2003. 49, 51, 54, 70, 124

F. Meyer. From connected operators to levelings. In *Mathematical Morphology and Its Applications to Image and Signal Processing, (H. Heijmans and J. Roerdink, Eds.)*, pages 191--198, Kluwer Academic, 1998a. 67

F. Meyer. Levelings, image simplification filters for segmentation. *International Journal of Mathematical Imaging and Vision*, 20:59--72, 2004. 65, 67, 68, 99

F. Meyer and P. Maragos. Nonlinear scale-space representation with morphological levelings. *Journal of Visual Communication and Image Representation*, 11:245--265, 2000a. 3, 95, 99

Y. Meyer. The Levelings. In *Mathematical Morphology and Its Applications to Image and Signal Processing*. Kluwer Academic Publishers, 1998b. 99

Y. Meyer and P. Maragos. Nonlinear Scale-Space Representation with Morphological Levelings. *Journal of Visual Communication and Image Representation*, 11:245--265, 2000b. 65, 67, 68, 69, 74, 128

C.S. Mladinich et al. An evaluation of object-oriented image analysis techniques to identify motorized vehicle effects in semi-arid to arid ecosystems of the american west. *GIScience and Remote Sensing*, 47(1):53--77, 2010. 97, 126

Lasse Moller-Jensen. Classification of urban land cover based on expert systems, object models and texture. *Computers, environment and urban systems*, 21(3):291--302, 1997. 49

Giorgos Mountrakis, Jungho Im, and Caesar Ogole. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247--259, 2011. 123, 124, 147, 150

Soe W. Myint, Patricia Gober, Anthony Brazel, Susanne Grossman-Clarke, and Qihao Weng. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 115(5):1145 -- 1161, 2011. ISSN 0034-4257. doi: 10.1016/j.rse.2010.12.017. 3, 95, 96

M. Negnevitsky. *Artificial Intelligence, a Guide to Intelligent Systems*. Pearson Education, 2005. 49, 54

M. Neubert, H. Herold, and G. Meinel. Evaluation of remote sensing image segmentation quality - further results and concepts. *Interanational Archive of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (4/C42):6p, 2006. ISSN 1682-1777. 70, 97

S. Nilufar, N. Ray, and H. Zhang. Object detection with dog scale-space: A multiple kernel learning approach. *IEEE Transactions on Image Processing*, 21(8):3744 --3756, 2012. ISSN 1057-7149. doi: 10.1109/TIP.2012.2192130. 100

Regina Obe and Leo Hsu. *PostGIS in action*. Manning Publications Co., 2011. 157

Y.O. Ouma, S.S. Josaphat, and R. Tateishi. Multiscale remote sensing data segmentation and post-segmentation change detection based on logical modeling: Theoretical exposition and experimental results for forestland cover change analysis. *Computers & Geosciences*, 34 (7)(7):715--737, 2008. 4, 5, 64, 65, 66, 67, 70, 95, 96, 97, 98, 99

G.K. Ouzounis, M. Pesaresi, and P. Soille. Differential area profiles: Decomposition properties and efficient computation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(8):1533 --1548, aug. 2012. ISSN 0162-8828. doi: 10.1109/TPAMI.2011.245. 100

Mahesh Pal. Kernel methods in remote sensing: a review. *ISH Journal of Hydraulic Engineering*, 15(sup1):194--215, 2009. 124

Mahesh Pal and Giles M. Foody. Evaluation of svm, rvm and smlr for accurate image classification with limited ground data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(5):1344--1355, 2012. 126

N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26 (9):1277--1294, 1993. 13, 14, 18, 33, 34, 35, 75

Giuseppe Papari and Nicolai Petkov. Edge and line oriented contour detection: State of the art. *Image and Vision Computing*, 29(2–3):79 -- 103, 2011. ISSN 0262-8856. doi: 10.1016/j.imavis.2010.08.009. 100

N. Paragios, Y. Chen, and O. Faugeras. *Handbook of Mathematical Models of Computer Vision*. Springer, 2005. 5, 65, 98

T. Pavlidis and Y.T. Liow. Integrating region growing and edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(3):225--233, 1990. 101

P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (7):629--639, 1990. 67, 99, 152

John Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61--74, 1999. 128, 133

A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. Tilton, and G. Trianni. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113 (S1):S110--S122, 2009. 66, 67, 99

GNR Prasad and A Vinaya Babu. A study on expert systems in agriculture. *Extension of Technologies: From Labs to Farms*, page 297, 2008. 148

Seda Sahin, Mehmet R Tolun, and Reza Hassanpour. Hybrid expert systems: A survey of current approaches and applications. *Expert Systems with Applications*, 39(4):4609--4617, 2012. 147, 149

Michael Schröder and Alex Dimai. Texture information in remote sensing images: A case study. 1998. 38

J. Serra. Connections for Sets and Functions. *Fundamentae Informatica*, 41:147--186, 2000. 67, 99

Gary Sherman. *Desktop GIS: Mapping the Planet with Open Source Tools*. Pragmatic Bookshelf, 2008. 157

M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. PWS, Pacific Grove, CA, 1998. 13, 14, 15, 18, 33, 34, 35, 50

S.A. Stewart, G.J. Hay, P.L. Rosin, and T. J. Wynn. Multiscale structure in sedimentary basins. *Journal of Basin Research*, 16 (2):183--197, 2004. 4, 64, 66, 70, 95, 96, 98

Levon Sukissian, Stefanos Kollias, and Yiannis Boutalis. Adaptive classification of textured images using linear prediction and neural networks. *Signal processing*, 36(2): 209--232, 1994. 33

S. Theodoridis and K Koutroumbas. *Pattern Recognition*. Elsevier Academic Press, 2003. 49, 51, 52, 70, 77, 107, 124, 133

Michael E Tipping. Sparse bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, 1:211--244, 2001. 125, 130, 131, 132, 135, 151, 156

Michael E Tipping, Anita C Faul, et al. Fast marginal likelihood maximisation for sparse bayesian models. In *Proceedings of the ninth international workshop on artificial intelligence and statistics*, volume 1. Jan, 2003. 128, 132, 156

A. Tzotsos. A support vector machine approach for object based image analysis. In *Proceedings of 1st International Conference on Object-based Image Analysis, OBIA 2006, Salzburg, Austria, July 4-5*, 2006. 7, 8, 107, 126, 147

A. Tzotsos and D. Argialas. Mseg: A generic region-based multi-scale image segmentation algorithm for remote sensing imagery. In *Proceedings of ASPRS 2006 Annual Conference, Reno, Nevada, May 1-5*. ASPRS p.(on CD-ROM), 2006. 7, 33, 34, 35, 36, 50, 53, 55, 66, 70, 75, 76, 79, 97, 104, 105, 107, 126, 129, 154

A. Tzotsos and D. Argialas. *Support Vector Machine Classification for Object-Based Image Analysis*. In: Blaschke T., Lang S. and Hay G. (Eds.) Object Based Image Analysis - Spatial concepts for knowledge driven remote sensing applications, New York: Springer pp. 663-679, 2008. 7, 8, 41, 46, 70, 76, 77, 107, 109, 112, 123, 126, 130, 132, 134, 147, 151, 155, 156, 157, 162

A. Tzotsos, C. Iosifidis, and D. Argialas. *A hybrid texture-based and region-based multi-scale image segmentation algorithm*. In: Blaschke T., Lang S. and Hay G. (Eds.) Object Based Image Analysis - Spatial concepts for knowledge driven remote sensing applications, New York: Springer pp. 221-237, 2008. 7, 8, 50, 53, 66, 70, 97

A. Tzotsos, K. Karantzalos, and D. Argialas. Multiscale segmentation and classification of remote sensing imagery with advanced edge and scale-space features. *In: Q.Weng (Editor), Scale Issues in Remote Sensing, John-Wiley & Sons*, pages 170--196, 2014. 7, 9, 123, 127, 128, 129, 130, 152, 154

Angelos Tzotsos, Konstantinos Karantzalos, and Demetre Argialas. Object-based image analysis through nonlinear scale-space filtering. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1):2 -- 16, 2011. ISSN 0924-2716. doi: 10.1016/j.isprsjprs.2010.07.001. 4, 7, 8, 95, 96, 98, 99, 102, 104, 107, 108, 109, 110, 112, 127, 128, 129, 147, 151, 152, 153, 154

R. Ulichney. *Digital Halftoning*. The MIT Press, Cambridge, MA., 1987. xxv, 17, 35

V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag., 1995. 51

V.N. Vapnik. *Statistical Learning Theory*. John-Wiley and Sons, Inc., 1998. 51, 52, 70, 77, 107, 124, 130, 133, 150, 155

R.G. Von Gioi, J. Jakubowicz, J.M. Morel, and G. Randall. Lsd: A fast line segment detector with a false detection control. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(4):722--732, 2010. 100, 102, 103, 127, 129, 152, 153, 154

R.G. Von Gioi, J. Jakubowicz, J.M. Morel, and G. Randall. LSD: a Line Segment Detector. *Image Processing On Line*, 2012. doi: 10.5201/ipol.2012.gjmr-lsd. 103, 104

Hongzhi Wang and John Oliensis. Generalizing edge detection to contour detection for image segmentation. *Computer Vision and Image Understanding*, 114(7):731 -- 744, 2010. ISSN 1077-3142. doi: 10.1016/j.cviu.2010.02.001. 100

J. Weickert. *Anisotropic Diffusion in Image Processing*. ECMI Series, Teubner-Verlag, Stuttgart, Germany, 1998. 3, 67, 95, 99

J. Weickert, S. Ishikawa, and A. Imiya. Linear scale-space has first been proposed in Japan. *Journal of Mathematical Imaging and Vision*, 10 (3)(3):237--252, 1999. 65, 98

Qihao Weng. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, 117:34--49, 2012. 149

A. Witkin. Scale-space filtering. In *International Joint Conference on Artificial Intelligence*, pages 1019--1021, 1983. 3, 5, 65, 95, 98

Yong Xu, Sibin Huang, Hui Ji, and Cornelia Fermüller. Scale-space texture description on sift-like textons. *Computer Vision and Image Understanding*, 116(9):999 -- 1013, 2012. ISSN 1077-3142. doi: 10.1016/j.cviu.2012.05.003. URL . 100

Hun-Woo Yoo, Han-Soo Park, and Dong-Sik Jang. Expert system for color image retrieval. *Expert Systems with Applications*, 28(2):347--357, 2005. 48

P. Yu, A.K. Qin, and D.A. Clausi. Unsupervised polarimetric sar image segmentation and classification using region growing with edge penalty. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(4):1302 --1317, april 2012. ISSN 0196-2892. doi: 10.1109/TGRS.2011.2164085. 101, 105

Baojiang Zhong, Kai-Kuang Ma, and Wenhe Liao. Scale-space behavior of planar-curve corners. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(8):1517 --1524, aug. 2009. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.295. 100

Weiqi Zhou, Ganlin Huang, Austin Troy, and M.L. Cadenasso. Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study. *Remote Sensing of Environment*, 113 (8)(8):1769--1777, 2009. 66, 69, 70, 97, 98, 126

Ji Zhu and Trevor Hastie. Kernel logistic regression and the import vector machine. In *Advances in neural information processing systems*, pages 1081--1088, 2001. 126