12-31-2013

Advanced Earth Observation for Humanitarian Information Extraction

Chandi Witharana
University of Connecticut - Storrs, chandi.witharana@gmail.com

Advanced Earth Observation for Humanitarian Information Extraction

Chandi Witharana, Ph.D.

University of Connecticut, 2013

Abstract

Advanced Earth observation (EO) is increasingly recognized as an indispensible tool to support rigorous and robust decision making in crisis management. In crisis scenarios, EO data need to be streamed through time-critical workflows for delivering reliable and effective information to civil protection authorities. In this context, the over arching goal of this research is to closely examine the integral segments of routing EO-based rapid mapping workflows to understand prevailing shortfalls and to devise novel approaches to catalyze conditioned geoinformation delivery to cater increasing user demand. This study envisions three interconnected objectives, which are primarily fuelled by very high spatial resolution (VHSR) imagery, data fusion, image segmentation, and geographic object-based analysis (GEOBIA) framework. Focal study areas encapsulate natural and anthropogenic crises having occurred in the recent past: the 2010 earthquake-damaged areas in Haiti, the 2010 flood-impacted sites in Pakistan, and armed-conflicted areas and internally displaced persons (IDP) camps in Sri Lanka

The first objective investigated how different data fusion algorithms perform when applied to VHSR satellite images that encompass ongoing- and post-crises scenes. The evaluation entailed twelve fusion algorithms. The spatial and spectral fidelities were assessed subjectively using fourteen quality indices. Ehlers, Wavelet, and High-pass filtering (HPF) fusion algorithms had the best scores for the majority of spectral quality indices. The University of New Brunswick and Gram-Schmidt fusion algorithms had the

best scores for spatial metrics. The HPF algorithm emerged as the overall best performing fusion algorithm.

The second objective aimed to unravel the synergies of data fusion and image segmentation in the context of EO-based rapid mapping workflows. We statistically compared the quality image object candidates among twelve fused products and their original MS and PAN images. We have shown that the GEOBIA framework has the ability to create meaningful image objects during the segmentation process by compensating the fused image's spectral distortions with the high-frequency information content that has been injected during fusion. We further questioned the necessity of the data fusion step in rapid mapping context. Bypassing time-intense data fusion steps helps to intensify EO-based rapid mapping workflows.

The third objective explored the efficacy of supervised empirical discrepancy measures for optimizing multiresolution segmentation (MRS) algorithm. I selected the Euclidean distance 2 (ED2) metric, a recently proposed supervised metric that measures dissimilarity between a reference polygon and an image object candidate, as a candidate to investigate the validity and efficacy of empirical discrepancy measures for finding the optimal scale parameter setting of the MRS algorithm. The discriminative capacity of the ED2 metric across different scales groups was tested using non-parametric statistical methods. My results showed that the ED2 metric significantly discriminates the quality of image object candidates at smaller scale values but it loses the sensitivity at larger scale values. This questions the meaningfulness of the ED2 metric in the MRS algorithms parameter optimization.

Advanced Earth Observation for Humanitarian Information Extraction

Chandi Witharana

B.Sc., University of Peradeniya, Sri Lanka, 2005

M.S., University of Connecticut, USA, 2009

A Dissertation

Submitted in partial Fulfillment of the

Requirement for the Degree of

Doctor of Philosophy at the

University of Connecticut

2013

APPROVAL PAGE

Doctor of Philosophy Dissertation

Advanced Earth Observation for Humanitarian Information Extraction

Presented by

Chandi Witharana, B.Sc., M.S.

Major Advisor_____
Thomas H. Meyer

Associate Advisor_____
Daniel L. Civco

Associate Advisor_____
Thomas Blaschke

Associate Advisor_____
Marco Neubert

Associate Advisor_____
Jeffry Osleeb

University of Connecticut

2013

ACKNOWLEDGEMENTS

This work would not have been possible without the help of many individuals and organizations. I wish to express my deep gratitude and appreciation to my principal advisor, Associate Professor Thomas Meyer, for his intellectual, financial, and infrastructure assistance for a successful completion of this degree. His constant encouragement and mentorship was invaluable in shaping not only my research at UConn but also my future career as well. I deeply indebted to my co-advisors: Professor Daniel Civco, Professor Thomas Blaschke, Dr. Marco Neubert, and Professor Jeffrey Osleeb for their constant mentorship, guidance, suggestions, and constructive criticism throughout the research. I would like to acknowledge ASPRS GeoEye Award, Alexander Goetz Award, and Pleisde User Group Award for providing me necessary very high resolution image data and instrument support.

My sincere thanks to Dr. Dirk Tiede, Department of Geoinformatics (Z_GIS), University of Salzburg, Austria for his extensive theoretical and technical contribution to my research. Special thanks are extended to Jason Parent for his technical advice on Python scripting. I greatly appreciate James Hurd for keeping all GIS and Remote Sensing software alive.

Finally, I would like to thank my wife Piyumi for her incessant patience and support I rely on daily and my little daughter Hiruni for giving her weekend and evenings to the cause of my graduate education.

I dedicate this dissertation to the citizens of Sri Lanka who facilitate to sustain free healthcare and education systems in the country.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Overview

Natural hazards, man-made disasters, civil wars, and regional conflicts occur and recur causing a huge loss of human lives and properties, prolonged social and economic disruptions, and environmental degradation (Roy and Blaschke 2010; Witharana et al. 2010; Witharana, 2012; WItharan and Civco, 2012; Witharana et al., 2013). The size and complexity of these crises have escalated drastically over the past decades and will most likely amplify in the future due to the projected impacts of climate change coupled adverse socio-economic and political conditions (e.g., paucity of natural resources, poverty, population growth, weak governance, and tension among different ethnic groups) (Voigt 2007; UNEP 2011; Hagenlocher et al. 2012). In the case of flood hazards, the United Nations (UN) estimates that over the last decade annually 200 million people over in 90 countries have been affected by catastrophic flooding (UN, 2011). Recent reports of Intergovernmental Panel on Climate Change (IPCC) - Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX) revealed a link between anthropogenic climate change and frequency of heavy precipitation, and provided projections for the future suggesting that flood hazards may increase intensity and frequency (IPPC, 2012).

Although natural and anthropogenic crises are inescapable, practicing effective crisis management strategies could relieve the impact on the human and natural environments. Humanitarian crisis management (HCM) (Witharana et al. 2013), also

referred as disaster management (DM) (Van Western, 2000) or emergency management (Joyce et al. 2010), is a prescribed cyclic process (Figure 1) of how societies respond (i.e. organization and management of resources and responsibilities) to natural and human-induced disasters (Cutter, 2003; UNISDR, 2004).

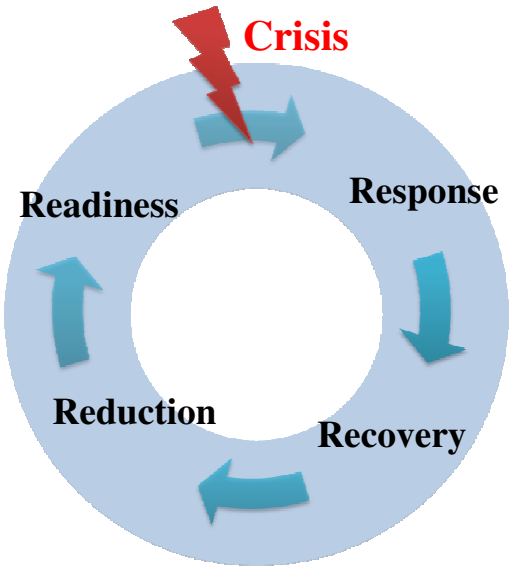


Figure 1. Humanitarian crisis management cycle (Joyce et al. 2010)

The HCM entails four key phases: 1) Response - Showing urgent response to immediately following an event such as search and rescue (SAR), humanitarian relief, evacuation, 2) Recovery - implementing long term recovery activities such as reconstruction, resettlement of returnees, establishing dismantled socioeconomic framework, 3) Reduction (mitigation) - taking measures to reduce vulnerability (hazard zonation, impose set-back zones, insurance .etc), and 4) Readiness (preparedness) - building disaster resilient communities (e.g. community awareness programs, early warning and alert systems) (UNISDR, 2004; Joyce et al., 2009; Witharana, 2010;

Witharana and Meyer, 2010). Crisis management is a time-intensive and collaborative effort that entails multiple individuals and stakeholders sharing information, expertise, and resources in support of robust situation assessment and informed decision making (MacEcheran and Cai, 2006). Effective crisis management relies on rapid and rigorous production and dissemination of pre-, ongoing- and post-crisis information (Voigt et al., 2007; Kaya et al., 2011; Hussain et al. 2011; Witharana et al. 2013)

Remote sensing (RS) is increasingly recognized as an indispensible tool to support decision making in crisis management due to its cost effectiveness and innate capability of providing global coverage, near-real-time, repeated, and objective observations encapsulating a broad spectrum of humanitarian crisis situations (Ito, 2005; Voigt et al. 2007; Taubenböck et al. 2008; Dell'Acqua and Polli, 2011). In a broad sense, information derived from RS data helps the civil protection authorities in rescue and relief, efforts, damage assessment, and the planning of remedial measure to safeguard disaster events effectively. Each phase of the HCM cycle has different demands on the remotely-sensed data.

In the recovery phase, remote sensing can contribute a great deal through the provision of objective time series analysis over large areas with both high and medium levels of spatial detail. Using post-event EO data, it is possible to temporally-variant indicators that can be used as proxies to objectively gauge and monitor the long-term disaster recovery. Some of these indicators entail removal of debris, demolition of transitional, IDP, and refugee shelters, commencement and completion of new construction or reconstruction of residential buildings and other key civil infrastructure, vegetation growth, .etc. EO-based recovery information is of high value for

humanitarian donor agencies to evaluate the progress of their recovery projects in the region. Joyce et al., (2006) used SPOT and ALOS AVNIR-2 images to monitor recovery of widespread landsliding in northern New Zealand. Hill et al., (2006) utilized time series VHSR image data to extract long-term recovery-related features, such as building construction, presence of earth moving equipment, vegetation growth, and debris in the after math of the Hurricane Katrina. Pagot and Pesaresi (2008) used bi-temporal IKONOS scenes to characterize the state of built-up structures four years after the Sierra Leone conflict and to differentiate between intact, impacted, rebuilt and new structures.

In the reduction phase, remote sensing can be used directly for hazard identification (e.g., floodplain modeling, slope stability and landslide susceptibility), hazard modeling, and mapping elements at risk (e.g., human settlements and infrastructure) (Joyce et al. 2010). Remote sensing acts as a rich data source for informed land use planning, which is a key tool practiced by authorities to mitigate hazard risk (Burby, 1998). In the case of natural hazards, EO data can intelligently utilize to analyze spatio-temporal dynamics of hazards and subsequently apply proper land use planning methods to address the potential risk. Commonly practiced planning methods include mapping hazard zones (e.g., location and range of impact) and probability of occurrence (Joyce et al. 2010). Based on EO-derived hazard maps, planning authorities stake necessary risk reduction measures, such as setback zones (e.g. proximal to active faults or inundation zones), special building codes (e.g. minimum floor heights above base flood level). EO-based (e.g., VHSR images and LiDAR data) asset information retrieval is highly beneficial for hazard consequence modeling, in which hazard data is combined with asset data and vulnerability information to estimate potential losses (Zeger et al,

2003, Witharana, 2009). For instance, building vulnerability to certain hazard type (e.g., earthquake, Tsunami, river flooding, and landslide) is based on several factors, such as construction material, building height, floor height, number of floors, proximity to other structures These type information can rigorously and cost effectively produce by complementing VHSR images and LiDAR data.

The strength and the utility of remote sensing are mostly highlighted in the response phase, which capture the "golden hour" - the narrow time window available of saving life exists primarily in the first hour following an event and decline rapidly thereafter (Gooldchild, 2006, Gregg and Houghton, 2006). In operational setting, for instance, United States's Federal Emergency Management Agency (FEMA) requires rapid assessment of the emergent crises (image acquisition and data processing) to be achieved within 48 hours (Kim et al, 2011). Earth observation (EO) data provide crucial information for on-going evacuation planning in conjunction with weather patterns, hazard behavior (e.g. water level rise, movement of fire fronts), rapid structural damage assessments, and monitoring civilians movements and ephemeral settlements. Over the past decade, EO data have been exhibiting an inextricable link to the response phase of the HCM. This is mainly attributed by the sophistication of satellite sensor technology, which has made available space-based imagery with spatial resolutions more than we ever expected. Post-IKONOS era sensors like QuickBird, WorldView-1, GeoEye-1, and WorldView-2 provide very high spatial resolution (VHSR) multi-spectral imagery that are capable of capturing crisis information at finer details, i.e. city-block to individual house or an IDP camp to an individual shelter (Lang et al., 2010; Vu et al., 2009). Due to shorter revisit times of these sensors, it is also possible to acquire real-time or near real-

time imagery over impacted areas (Kim et al., 2011). The utility of VHSR imagery in rapid damage mapping is highlighted in the disaster literature. Chiroiu (2005); Gusella et al. (2005); Vu et al. (2005); Kouchi et al. (2005); and Yamazaki et al. (2007) have successfully used Quickbird and IKONOS images for damage detection in earthquake-impacted cities in Iran (the 2003 Bam earthquake) and Algeria (the 2003 Boumerdes earthquake). Ehrlich et al. (2009) and Saito et al. (2004) investigated the use of VHSR imagery for damage assessment following the 2008 Wenchuan earthquake in China and the 2001 Gujarat earthquake in India, respectively. The 2010 Haiti earthquake has made a greater emphasis on the utility of VHSR imagery in rapid damage mapping. Number of recent studies, such as Corbane et al. (2011); Hussain et al. (2011); Stefan et al. (2011); and Tiede et al. (2011) have proven the strength of VHSR imagery complemented with the other earth-observation data (e.g. SAR images, LiDAR) for semi- and fully-automated post-crisis damage assessments. Pesaresi et al. (2007); Vu et al. (2007); and Gmaba et al. (2007) explored the use of VHRS images for rapid damage assessments in the 2004 Indian Ocean tsunami impacted areas. Al-Khudhairy et al. (2005) assessed the structural damage inflicted by armed-conflicts in the Former Yugoslav Republic of Macedonia during 2001 using post-crisis IKONOS images. Not only for rapid damage mapping, In a humanitarian perspective, the most critical parameter to be mapped and monitored is the number of people affected by a crisis (Lang et al., 2010). On-demand census on an affected population is of high value for coordinating and implementing relief operations. EO-based rapid mapping of transitional shelters can provide a rough estimate of an affected population and information on location and spatial extent of shelters and detailed information on dwelling count, type, and size. For example, during

the last stage of Sri Lanka's civil war (2007- 2009), nearly 125,000 civilians were held as hostages by rebels in a 25 sqr km area which was actually declared as a civilian safety zone (CSZ) by the Sri Lankan government. Information flow in or out of the CSZ was totally cut-off by rebels. Compared to post-disaster damage assessments, a few works have been done in the domain of EO-based civilian monitoring. Giada et al. (2003) tested different methods for extracting refugee tents and spatial extent from IKONOS imagery over the Lukole refugee camp in Tanzania. Tiede et al. (2010) developed transferable object based image analysis (OBIA) workflows for extracting dwelling structures and densities from QuickBird imagery in three IDP shelter sites in Sudan. For the same study area, Kemper et al. (2011) used GeoEye-1 images along with mathematical morphological techniques to enumerate dwellings in IDP camps.

Image acquisition is the key step in the crisis support service cycle (Figure 2) (Voigt et al., 2007, ZKI-DLR, 2010). During and immediately following a crisis, image data should be made available to image processing workflows to produce timely-information (Voigt et al., 2007). Owing to the frequent occurrence and adverse impacts of natural and man-made disasters, several global initiatives have been formulated to strengthen crisis support services. The International Charter Program, a globally functioning mechanism initiated in year 1999, coordinates the tasking of multiple satellites and archiving systems in very short time to respond to natural and man-made disasters (Voigt et al., 2007; Stryker and Jones, 2009; Kim et al., 2011) As of today, International Charter program is operated by twelve space agencies across the globe. The program has been activated over 300 times (International Charter, 2010) providing timely image products to key stakeholders involved in the crisis management cycle. The

International Charter's main activations include the 2004 Indian Ocean tsunami, the 2010 Haiti earthquake, the 2010 flooding in Pakistan, and the recent oil spill in the Gulf of Mexico (International Charter, 2010). The European Earth monitoring programme - Global Monitoring for Environment and Security (GMES) is another global initiative, led by the European Union (EU), which provides earth observation services for various policy-related issues including emergency management (Kranz et al., 2010; Lang et al., 2010).

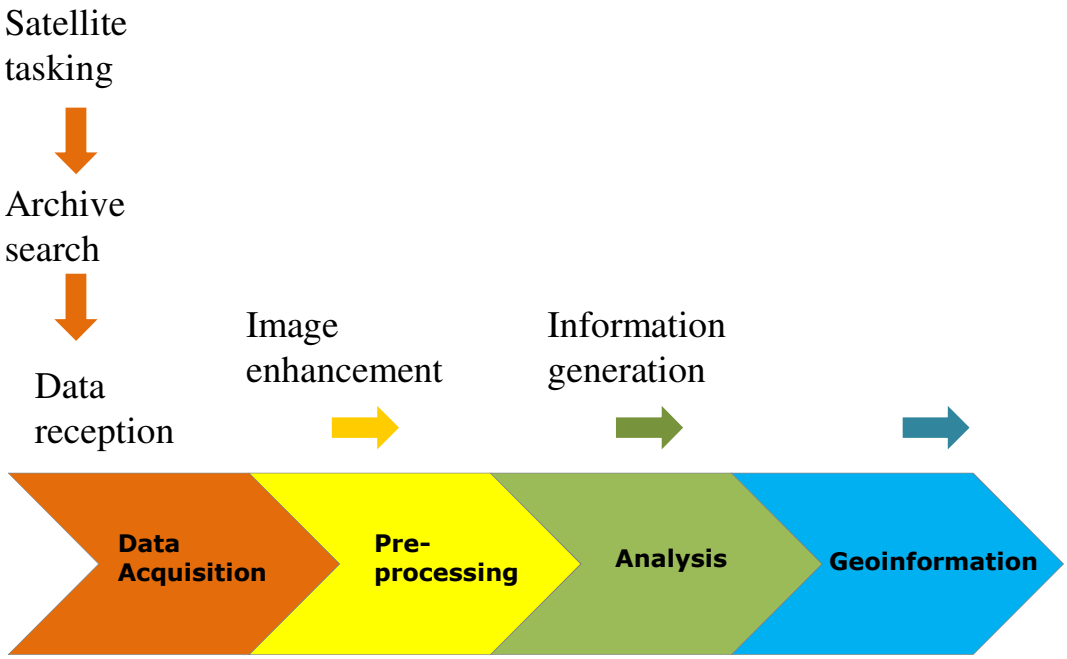


Figure 2. Earth observation based rapid mapping workflow (ZKI-DLR, 2013)

In summary, remote sensing has the potential to seamlessly integrate into the all four phases of the humanitarian crisis management cycle; however, its strength and the utility are mainly highlighted in the response phase. With respect to other phases, the response phase is highly time-critical, thus, earth observation data have to be streamed

through rapid mapping workflows to ensure the robust and rigorous delivery of ongoing- and post-crisis information to civil protection authorities. In this context, it is important to revisit and closely examine the integral segments of routing EO-based crisis information retrieval chains to understand prevailing deficiencies and to devise novel approaches to catalyze conditioned geoinformation delivery from complex image scenes to cater increasing user demand.

1.2 Research Objectives

This study envisions three main objectives in the light of data fusion and geographic-object based image analysis (GEOBIA) frame works.

These objectives are to;

1. Evaluate spectral and spatial fidelities of twelve data fusion algorithms in the context of Earth Observation based rapid mapping workflows.
2. Investigate synergies of data fusion and image segmentation in the geographic-object based image analysis framework.
3. Examine the sensitivity of empirical supervised measures for optimizing the multiresolution segmentation algorithm.

1.3 Structure of the Dissertation

Chapters 2 through 4 stand independently in a self-organized manner addressing three key research questions that I aimed to answer. These chapters comprise literature

review, research problem, methods, results, discussions, and conclusions pertaining to each objective.

CHAPTER 2

Evaluation of Pansharpening Algorithms in Support of Earth Observation based Rapid Mapping Workflows

Abstract

In order to satisfy the humanitarian information demand in ongoing- and post-crisis situations, earth observation (EO) data must be streamed through time-critical workflows. Data fusion serves as an integral segment of EO-based rapid mapping workflows. Fused images form the basis for manual, semi-, and fully-automated classification steps in the information retrieval chain. Many fusion algorithms have been developed and tested for different remote sensing applications, however, the efficacy of data fusion is weakly assessed in the context of rapid mapping workflows. In this research, we investigated how different fusion algorithms perform when applied to very high spatial resolution (VHSR) satellite images that encompass ongoing- and post-crises scenes. The evaluation entailed twelve fusion algorithms: Brovey transform, color normalization spectral sharpening (CN) algorithm, Ehlers fusion algorithm, Gram-Schmidt fusion algorithm, high-pass filter (HPF) fusion algorithm, local mean matching algorithm, local mean variance matching (LMVM) algorithm, modified intensity-hue-saturation (HIS) fusion algorithm, principal component analysis (PCA) fusion algorithm, subtractive resolution merge (SRM) fusion algorithm, the University of New Brunswick (UNB) fusion algorithm, and the wavelet-PCA fusion algorithm. These algorithms were applied to GeoEye-1 satellite images taken over three geographical settings representing natural and anthropogenic crises that occurred recently: earthquake-damaged sites in Haiti, flood-impacted sites in Pakistan, and armed-conflicted areas and internally displaced persons (IDP) camps in Sri Lanka. Fused images were assessed for spectral and spatial fidelity using a variety of quantitative quality indicators and visual inspection methods. Spectral quality metrics include correlation coefficient, root-mean-square-error

(RMSE), relative difference to mean, relative difference to standard deviation, spectral discrepancy, deviation index, peak signal-to-noise ratio index, entropy, mean structural similarity index, spectral angle mapper, and relative dimensionless global error in synthesis. The spatial integrity of fused images was assessed using Canny edge correspondence, high-pass correlation coefficient, and RMSE of Sobel-filtered edge images. Under each metric, fusion algorithms were ranked and best competitors were identified. Ehlers, WV, and HPF had the best scores for the majority of spectral quality indices. UNB and Gram-Schmidt algorithms had the best scores for spatial metrics. HPF emerged as the overall best performing fusion algorithm.

2.1 Introduction

Humanitarian crisis management is a time-critical process. Effective crisis management relies on rapid and rigorous production and dissemination of pre-, ongoing- and post-crisis information (Witharana and Civco, 2012). Remote sensing is an indispensable tool in crisis management (Cheema, 2007, Kaya et al., 2011). Earth observation (EO) data exhibit the highest demand in the response phase of the crisis management cycle (Joyce et al. 2009, Dell`Acqua and Polli, 2011). From a humanitarian perspective, the most critical parameter to be mapped and monitored is the number of people affected by a crisis (Lang et al., 2010). On-demand census on affected population is of high value for coordinating and implementing relief operations. Time-series imagery acquired during and post-crisis can assist humanitarian relief agencies to implement high priority tasks such as, monitoring civilian movements, locating transitional shelter sites (TSS) and determining dwelling counts (e.g. internally displaced persons (IDP) and refugee camps), and rapidly quantifying the extent and severity of damage to buildings and infrastructure.

Owing to the frequent occurrence and adverse impacts of natural and man-made disasters, several global initiatives have been formulated to strengthen crisis support services. The International Charter Program, a globally functioning mechanism initiated in 1999, coordinates the tasking of multiple satellites and archiving systems in a very short time to respond natural and man-made disasters (Voigt et al., 2007; Kim et al., 2011). The program has been activated in major crisis situation providing timely EO data to key stakeholders involved in the crisis management cycle. The European Earth monitoring program - Global Monitoring for Environment and Security (GMES) is another global initiative led by the European Union (EU), which provides earth observation services for various policy-related issues including emergency management (Lang et al., 2010).

Despite EO data and global ED-data dissemination initiatives, in humanitarian emergencies the timeliness of data provision and the short time window available for dispatching value-added information pose major challenges to the mapping community. Unlike other remote sensing application domains, such as land use/cover mapping, environmental monitoring, and natural resource management, in crisis scenarios, EO data need to be streamed through time-critical workflows for delivering reliable and effective information (Tiede et al., 2011). Thus, there is always a compromise among response time, analysis depth, and thematic accuracy (Voigt et al. 2011). Typically, in the context of an EO-based rapid mapping workflow, the pre-preprocessing step serves as an integral segment that stands in between data acquisition and analysis steps. In this respect, the role of data fusion cannot be overlooked since it serves as a cohesive component and routine procedure in rapid information production. However, there is an emerging

concept of challenging the necessity of data fusion in the geographic object-based analysis (GEOBIA, also called OBIA) framework. For example, Tiede et al (2011) attempted to bypass major pre-processing steps including data fusion and developed a methodology for automated extraction of damage information from very high spatial resolution (VHSR) satellite image data.

Fusion evaluation is a well-addressed research problem. There is a plethora of literature on fusion-quality assessments addressing general context (Vijayaraj et al., 2006; Karathanassi et al. 2007; Ling et al. 2007; Ling et al. 2008; Nikolakopoulos, 2008; Ehlers et al. 2010) and focusing on specific application domains (Ashraf et al. 2012, Yang et al. 2012). Despite data fusion being linked to routine rapid-mapping workflows, we have a little knowledge on the effectiveness of data fusion algorithms when applied to crisis image scenes. The choice of fusion algorithm depends on the application domain because the reflectance varies with different environmental features. Different fusion algorithms introduce spectral and spatial distortions to the resultant data depending on the scene content; therefore a careful selection of the fusion method is required. An image scene of an intact city block exhibits different spectral and spatial properties when compared to another image scene acquired over the same areas after a major disaster: perhaps flood waters the surround the buildings or they partially collapsed into rubble after an earthquake. Thus, a fusion algorithm that is designed to address high-frequency edge information of urban landscapes might not produce satisfactory results when the same area is underwater. In case of IDP camps and transitional shelters (TS) sites, even human interpreters face major challenges when extracting individual shelters because these structures are very small (compared to regular man-made dwellings), randomly

oriented (e.g., TS shelters), highly crowded, and typically disturbed by existing land cover types (e.g., tree canopies). In this respect, it is challenging to transfer the knowledge on the performances of fusion algorithms that have been tested for a different application domain (e.g., freshwater habitat mapping) to another application domain. Our contention is that fusion algorithms are scene-dependent and they should be tested with respect to the application domain in focus. Thus, the central objective of this research is to investigate how well different fusion algorithms perform when applied to VHSR images of ongoing- and post-crisis scenes with different scene contents.

Modern satellite sensor technology provides space-borne imagery whose spatial resolution rivals aerial images (Blaschke, 2010; Dey et al., 2010). Satellite sensors like IKONOS, QuickBird, GeoEye-1, and WorldView-2 provide very high spatial resolution (VHSR) multi-spectral imagery (at sub-meter level) that can capture the fine details needed for crisis information , e.g. city-block to individual house or an IDP camp to an individual shelter (Lang et al., 2010; Li et al., 2010; Vu et al., 2009). Due to shorter revisit times of these sensors, it is also possible to acquire near real-time imagery over impacted areas (Kim et al., 2011).

VHSR satellite sensors typically record image data in a low resolution multispectral (MS) mode and high resolution panchromatic (PAN) mode. The high spatial resolution is needed to accurately describe the shapes of features and structures, and the high spectral resolution is needed to classify complex land-use and land-cover types (Wald, 2000; Rachin et al., 2003; Ehlers et al., 2010, Myint et al., 2011). Humanitarian crisis management remote sensing applications require high spatial and spectral resolution images. Fusing PAN and MS images with complementary

characteristics can provide a better visualization of the observed area (Wald, 2000; Ranchin et al., 2003). Pohl and van Genderen (1998) defined image fusion as a tool to combine multisource imagery using advanced image processing techniques that can be performed at three different processing levels (pixel, feature, and decision) depending on the stage at which the fusion takes place. Image fusion can occur in different ways such as inter-sensor, intra-sensor, singled-date, and multi-date. Pansharpening, also called resolution merge (Gangkofner et al., 2008) is a pixel-level fusion technique used to increase the spatial resolution of the multispectral image while preserving the spectral information (Vijayaraj et al., 2006). The perfect pansharpening result would be the MS image that would have observed if the multispectral sensor had the spatial resolution of the panchromatic sensor (Vrabel, 1996; Wald et al. 1997; Nikolakopoulos, 2008). Many image-fusion algorithms were developed for combining complimentary characteristics of PAN and MS images to produce an enhanced multispectral image of high spatial resolution. Several classifications for grouping fusion algorithms have been proposed. Pohl and van Genderen (1998) grouped fusion algorithms into color-related methods and statistical/numerical methods. Ehlers et al. (2010) treated the latter as two separate classes (statistical and numerical) and discussed different fusion techniques under three groups. Rachin and Wald (2000) and Wald (2002) proposed grouping by (1) the projection and substitution methods, (2) the relative spectral contribution, and (3) the method relevant to the ARSIS (a French acronym: *Amélioration de la Résolution Spatiale par Injection de Structures*, which means spatial improvement by injection structures) concept. Based on the information used in a pansharpening procedure, Gangkofner et al. (2008) grouped fusion techniques as spectral substitution methods, arithmetic merging, and spatial-

domain methods. Yakhdani and Azizi (2010) further developed the classification of Pohl and van Genderen (1998) and noted four fusion algorithms groupings : 1) color-related techniques, 2) statistical/numerical methods, 3) Pyramid-based methods, and 4) hybrid methods.

A fusion algorithm that preserves the spectral properties of the MS data and the spatial properties of the PAN data would be ideal, but there is always compromise (Civco et al., 2009). Many studies report the problems and limitations associated with different fusion techniques (Chavez et al., 1991; Wald et al., 1997; Zhang, 2002). The most-encountered problem in fusion algorithms is that the fused image exhibits a notable deviation in visual appearance and spectral values from the original MS image (Ling et al., 2007, Kalpoma and Kudoh, 2007). Spectral distortions including spatial artifacts affect both manual and automated classifications because any error in the synthesis of the spectral signatures at the highest spatial resolution incurs an error in the decision (Ranchin et al., 2003). Thus, it is necessary to evaluate the quality of fused images in terms of qualitative and quantitative indices. Qualitative comparison of the fused image and the original MS and PAN images for color preservation and spatial improvements is the most simple but effective way of benchmarking different fusion algorithms (Nikolakopoulos, 2008); however, visual inspection methods are subjective and largely depend on the experience of the interpreter (Klonus and Ehlers, 2007; Ehlers et al., 2010).

A number of objective metrics have been proposed to quantify spectral and spatial distortions incurred during the fusion process. Most widely used metrics for evaluating spectral fidelity are two-dimensional correlation coefficient (CC), root mean squared error (RMSE), relative difference of means, relative variation, deviation index,

and band discrepancy. Workers like Vijayaraj et al.(2006) , Karathanassi et al. (2007), and Yakhdani and Azizi (2010) utilized peak-signal-to-noise ratio (PSNR) and entropy as spectral quality metrics in addition to common indicators. Wald (2000) proposed the ERGAS metric (from its French acronym: *erreur relative globale adimensionnelle de synthese*, which means relative dimensionless global error in synthesis), which aims to provide a quick but accurate measure of the overall quality of a fused product. Few workers used the spectral angle mapper (SAM) to assess the overall spectral quality of fused images. Universal image quality index (Q-average) is global metric that models any distortions as a combination of loss of correlation, luminance distortion, and contrast distortion (Wang and Bovik, 2002). Alparone et al. (2004) generalized the Q-average as Q-4 index, which Alparone et al. (2007) applied to assess fusion quality along with SAM and ERGAS. Alparone et al. (2006) proposed a new index called QNR based on the findings of Wang and Bovik, (2002) and Xydeas and Petrovic (2000). The correlation of the gradient information, a combined quantity to evaluate spectral consistency and information content, was developed by Weidner (2010) based on the findings of Wang and Bovik, (2002) and Xydeas and Petrovic (2000). Wang et al. (2004) proposed another metric called mean structure similarity index (MSSIM), which is an enhanced version of the Q-average. Ling et al. (2007) and Ehlers et al. (2010) adopted the MSSIM to evaluate the spectral fidelity of fused images. Compared to spectral quality indicators, only few metrics are available to evaluate the spatial fidelity of fused images. Ehlers et al. (2010), Gangkofner et al. (2008), Klonus and Ehlers (2007), and Yakhdani and Azizi (2010) used high-pass correlation and edge detection using filters like Canny, Sobel, and Perwitte. In contrast, Civco et al. (2009) and Civco and Witharana (2012) tested a new approach

based on Fast Fourier Transform (FFT). Guo et al. (2010) used another metric called average gradient index that describes the changing features of image texture and detailed information to evaluate the spatial quality of fused images.

In this study, we examined twelve fusion algorithms that are commonly encountered in literature and built into commercial software packages. The candidate algorithms were applied to GeoEye-1 images that encompass ongoing- and post-crisis scenes from three geographical settings. The spectral and spatial fidelity of fused images were assessed using a variety of quantitative quality indicators and visual inspection methods. The quantitative indicators include eleven spectral quality metrics and three spatial quality metrics. We further investigated the stability and redundancy of quality metrics, and their effect on ranking of fusion results. We aimed to identify the best discriminators especially for detecting spectral distortions in fused products

2.2 Materials and methods

2.2.1 Study area and Data

We selected three geographic areas representing natural and anthropogenic crises having occurred in the recent past: earthquake-damaged areas in Haiti, flood-impacted sites in Pakistan, and armed-conflicted areas and IDP camp in Sri Lanka (Table 1). The selection of study areas (Figure 3) was made mainly focusing on ongoing- and post-crisis humanitarian information extraction. Haiti, Pakistan, and two of four study sites from Sri Lanka serve as candidates for post-crisis information retrieval, and the other two study sites from Sri Lanka emphasize the ongoing-crisis situations.

Haiti was struck by a magnitude-7 earthquake on 12th January 2010 that caused severe infrastructure damage. The United Nations Operational Satellite (UNOSAT) program and European Commission (EC) Joint Research Center (JRC) reports reveal, of the 90,000 buildings in Port-Au-Prince, more than 25,000 were damaged. In late July 2010, heavy rains caused extensive flooding in most parts of Pakistan. The United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) estimated that 20 million people were affected and more than 12 million houses were damaged. Nowshera, the main city in Nowshera district, which is located along the flood plain of Kabul River, was exposed to severe flooding causing great damage to human settlements. Based on the extent of damage caused to the human and material capital, Port-Au-Prince and Nowshera were chosen as study areas to test the effectiveness of different fusion algorithms on post-crisis image scenes.

Sri Lanka's thirty-year civil war ended in May 2009 when government forces liberated rebel-held areas through a two-year (2007-2009) long humanitarian operation.

We selected three study sites from the conflicted zone and one beyond the conflicted zone. The two towns Killinochchi and Puthukuduirippu came under intense heavy-arm attacks during the war inflicting severe damage to their built structures. The civilian safety zone (CSZ) is the third study site located inside the conflicted zone. This 25 km^2 no-fire zone was declared in January 2009 by the government as a refuge for fleeing civilians. Menik Farm is the largest IDP camp, which is located beyond the conflicted zone. VHSR imagery taken over Killinochchi, Puthukuduirippu, and Menik Farm were selected as candidates to assess the performances of fusion algorithms under post-crisis scenarios, and the imagery over CSZ served as candidates for ongoing-crisis situations.

All image scenes used in this study (Table 1) were acquired by the GeoEye-1 sensor (Table 2), which has a spatial resolution of 0.41cm for the PAN and 1.65m for MS bands at nadir with 11-bit radiometric resolution. The images are spatially registered to the Universal Transverse Mercator (UTM) coordinate system on the WGS 84 datum.

Table 1 Study areas and general characteristics

Country	Study area	Land cover/use	Target features	Acquisition date
Sri Lanka	Civilian Safety Zone (CSZ)	Coastal	Transitional shelters	Mar. 06, 2009
	Menik Farm (MF)	Rural	IDP shelters	Feb. 28, 2010
	Killinochchi (KIL)	Townscape	War-damaged buildings	Apr. 17, 2009
	Puthukuduirippu (PUT)	Townscape	War-damaged buildings	Jun. 15, 2009
Pakistan	Nowshera (NOW)	Townscape	Flood-damaged buildings	Aug. 05, 2010
Haiti	Port-Au-Prince (PAP)	Cityscape	Earthquake-damaged buildings	Jan. 13,2010

Table 2. Basic specifications of GeoEye-1 satellite sensor

Spectral band	Spectral range (nm)	Spatial resolution (m) (nadir)
Panchromatic	450-900	0.41
Blue	450-520	1.65
Green	520-600	1.65
Red	625-695	1.65
NIR	760-900	1.65

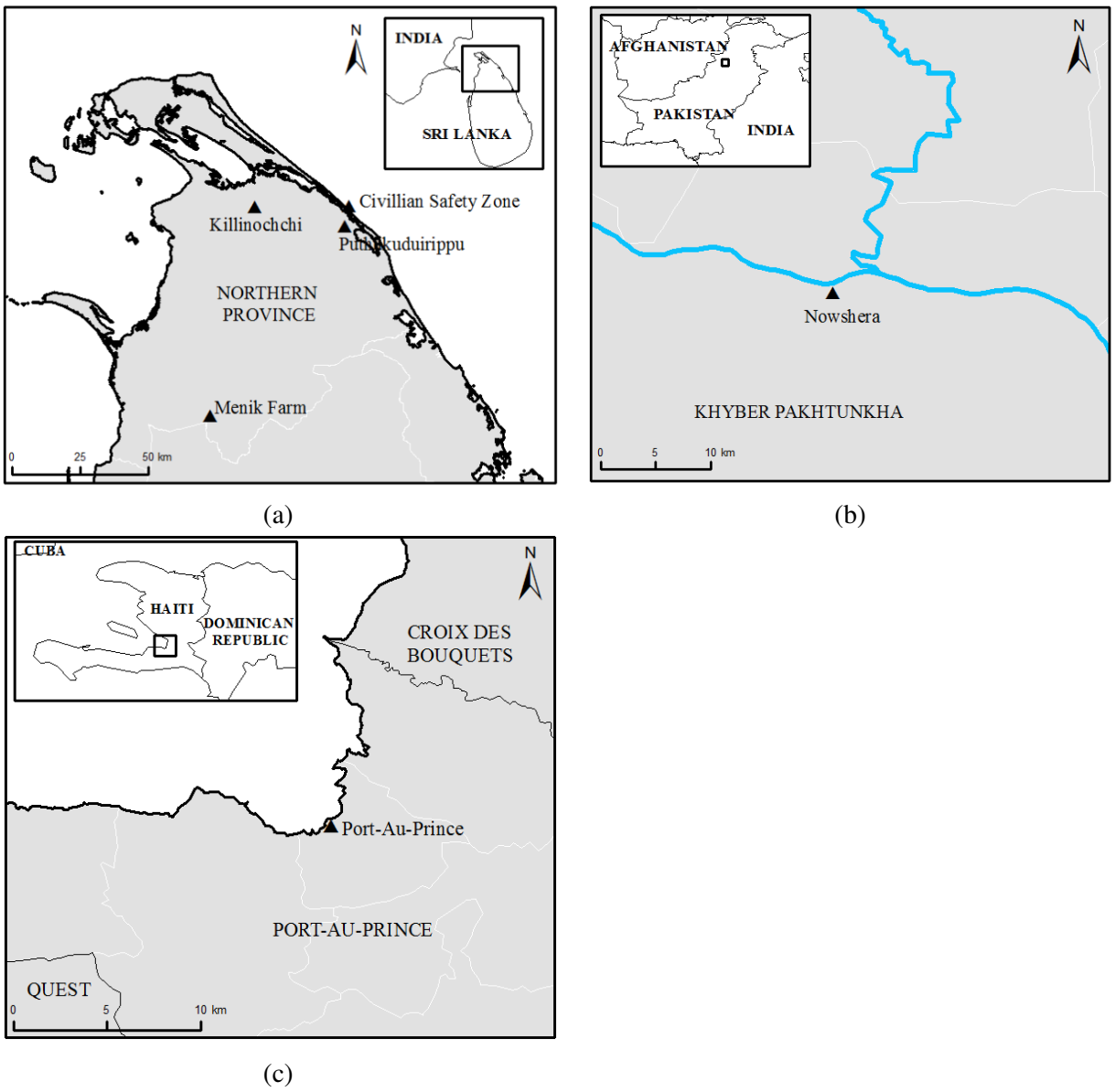
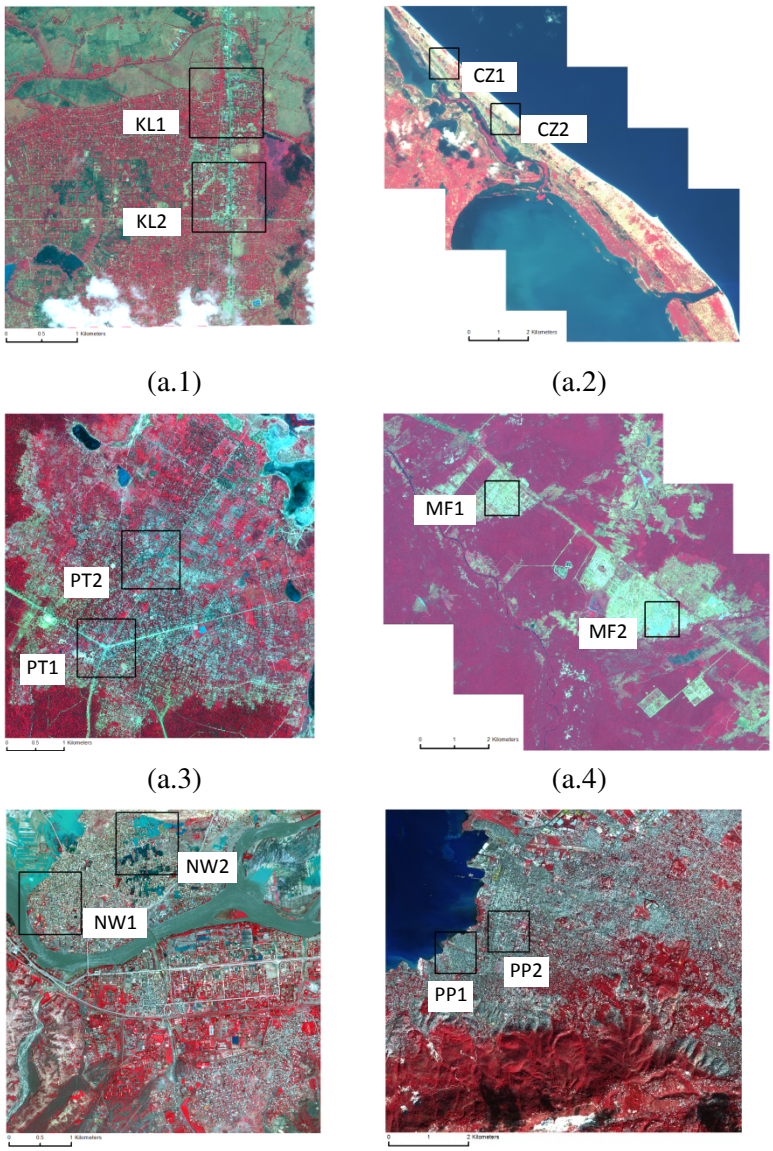


Figure 3. Index maps of the selected study sites from (a) Sri Lanka, (b) Pakistan, and (c) Haiti. Black triangles indicate the locations of GeoEye-1 images in the three countries



(a.1) (a.2)

(a.3) (a.4)

(b) (c)

Figure 4. GeoEye-1 image scenes of Sri Lanka (a.1, a.2, a.3, and a.4), Pakistan (b), and Haiti (c). All images are shown as false color composites. Black-hollowed boxes depict the candidate subsets used to evaluate fusion algorithms.

2.2.2 Methods

We extracted two subset sites (each approximately 1 km x 1 km) from each study area (Figures. 4 and 6), totaling twelve. The selection of subsets was made focusing on the features that are most likely to be extracted from ongoing- and post-crisis image

scenes (e.g., damaged buildings and infrastructure, IDP and transitional shelters). In this study, we tested twelve fusion algorithms (Table 3) that are commonly encountered in the literature and built into image processing software packages. We used ERDAS Imagine 2011, ENVI 4.8, and PCI Geomatica 2012, in conjunction with MATLAB 2008 to implement fusion algorithms. Some of the candidate fusion algorithms are proprietary (e.g., Ehlers fusion - ERDAS Imagine; the University of Brunswick fusion - PCI Geomatica). Unlike the Brovey transform and CN sharpening algorithm, which produce three-band fused images (B, G, and R or G, R, and NIR), other candidate algorithms are capable of accepting more than three bands at a time and producing four-band fused images in a single iteration. Therefore, we produced true- and false-color composites of BT and CN algorithms and layer-stacked them to create four-band pansharpened images. Fusion results were assessed using a series of quality metrics along with detailed visual inspection procedures to evaluate the spectral and spatial fidelity of fused products compared to their original MS and PAN images. Objective metrics were calculated independently for each subset and separately for each band (except for ERGAS and SAM). Mean values were calculated for all bands and for subsets in a given study area. Use of eleven spectral and three spatial metrics, totaling 14 objective quality indicators in our evaluation procedure, might be questionable because these metrics, especially spectral metrics, might be redundant and correlated (Pradhan et al. 2006; Gangkofner et al. 2008). Previous researchers used different combinations of quality indicators to rank fusion algorithms. The most common observation is that in one study a set of quality metric serves as key discriminators whereas in another study slightly or totally different quality metrics are used to discriminate fusion algorithms Our justification is that it is

important to employ a full complement of objective quality indicators and reexamine their stability and redundancy, and investigate the dependency of the ranking of fusion algorithms on quality metrics. These metrics' mathematical and statistical bases are tabularized in Table 4. The workflow's fusion-evaluation process is depicted in Figure 5.

Table 3. Candidate fusion methods, related literature, and implementation

Algorithm	Reference	Implementation
Brovey transform (BT)	Hallada and Cox (1983), Gillespie et al.(1987), Pohl and van Genderen (1998), Tu et al.(2001), Du et al. (2007), Kalpoma and Kudoh (2007), Karathanassi et al. (2007), Nikolakopoulos (2008), Wu et al. (2009), Yang et al. (2012)	ERDAS Imagine
Color normalization spectral sharpening (CN)	Vrabel et al. (2002), Klonus and Ehlers (2007), Ehlers et al.(2010)	ENVI
Ehlers fusion (EH)	Ehlers et al. (2003), Klonus and Ehlers (2007), Ling et al. (2007), Ehlers et al.(2010)	ERDAS Imagine
Gram-Schmidt fusion (GS)	Laben et al. (2000), Aiazzi et al. (2007), Karathanassi et al. (2007), Klonus and Ehlers (2007)	ENVI
High-pass filter (HPF)	Chavez et al.(1991), Vrabel (1996), Aiazzi et al. (2002), Gangkofner et al. (2008), Yang et al. (2012)	ERDAS Imagine
Local mean matching (LMM)	De Bethune et al.(1998), Karathanassi et al. (2007)	MATLAB
Local mean variance matching (LMVM)	De Bethune et al.(1998), Karathanassi et al. (2007)	MATLAB
Modified intensity hue saturation (MIHS)	Siddiqui (2003), Klonus and Ehlers (2007), Nikolakopoulos (2008), Ehlers et al.(2010), Yang et al. (2012)	ERDAS Imagine
Principle component analysis (PCA)	Chavez et al.(1991), Shettigara (1992), Vrabel (1996), Pohl and van Genderen (1998), Du et al. (2007), Kalpoma and Kudoh (2007), Karathanassi et al. (2007), Ling et al.(2008), Yang et al. (2012)	ERDAS Imagine
Subtractive resolution merge (SRM)	ERDAS (2011), Ashraf et al. (2012), Yang et al. (2012)	ERDAS Imagine
University of New Brunswick (UNB)	Zhang (2002), Zhang (2004), Alparone et al. (2007), Karathanassi et al. (2007)	PCI Geomatica
Wavelet Transform (WV)	Tu et al.(2001), Vijayaraj et al. (2006), Karathanassi et al. (2007), Gangkofner et al. (2008), Yakhdani and Azizi (2010) , Yang et al. (2012)	ERDAS Imagine

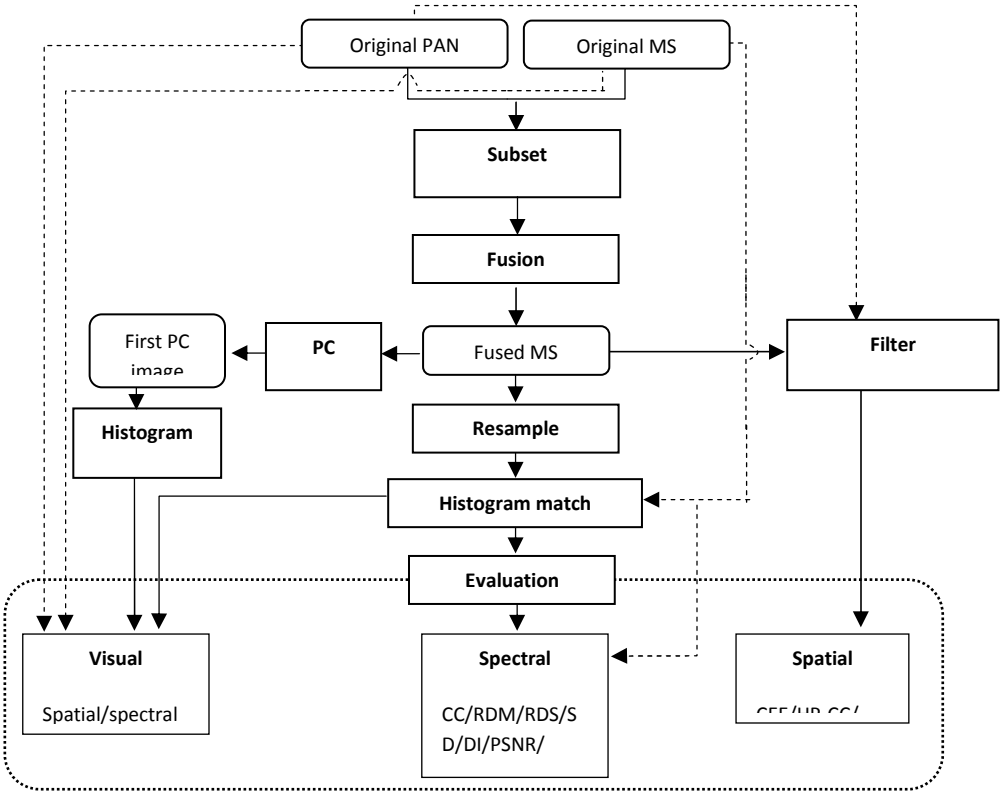
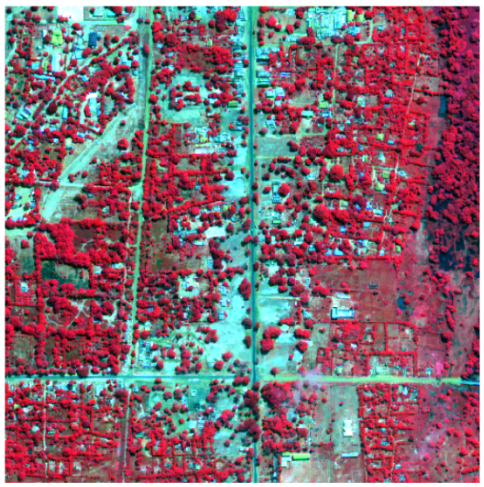
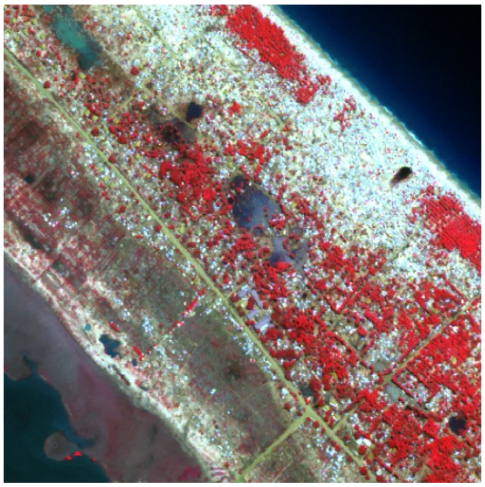


Figure 5. Fusion evaluation workflow

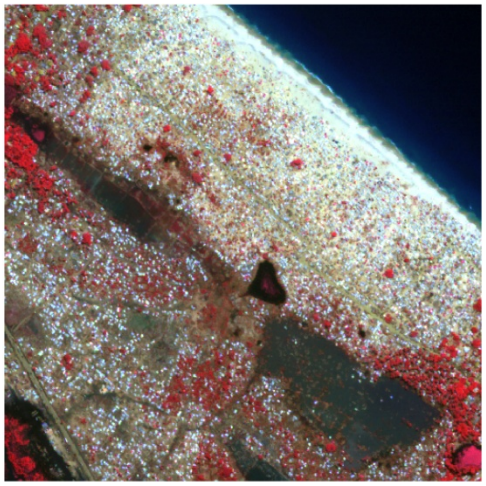


KL1

KL2

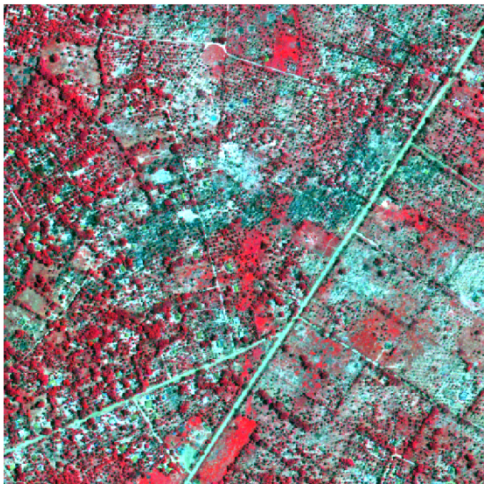


CZ1



CZ2

PT1



PT2

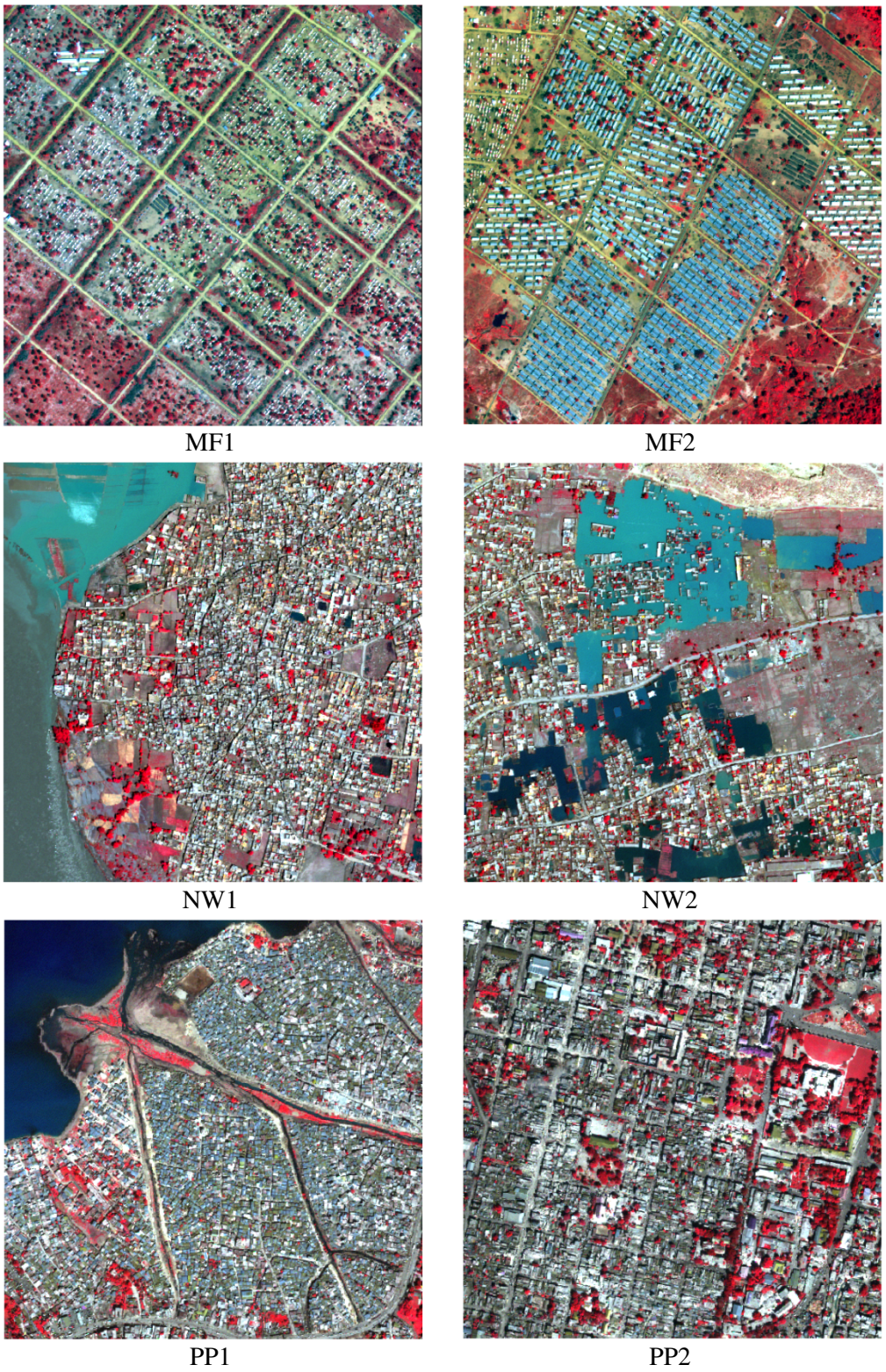


Figure 6. Subsets (1 km x 1 km) of Sri Lanka (KL1, KL2, CZ1, CZ2, PT1, PT2, MF1, and MF2), Pakistan (NW1 and NW2j), and Haiti (PP1, and PP2) shown as false color composites

2.2.3 Visual assessment

The color preservation and spatial improvement of fused images were compared to the original MS and PAN images, respectively. We selected one representative site from each country: Sri Lanka - Menik Farm (site 2), Pakistan- Nowshera (site 2), and Haiti (site - 2). In most studies, the fused image is displayed either as true-color or false-color composites and directly compared with the original MS and PAN images to assess color distortions and spatial improvement, respectively. We see two main deficiencies in this approach: 1) the high amount of spatial information in the fused MS image can mask its true spectral information and portray a pseudo-superiority of the high resolution fused MS image compared against the low resolution original MS image, 2) the high magnitude spectral information in the fused MS image can mask its true spatial information of the fused image compared against the grayscale PAN image. This kind of pseudo superiority or inferiority can mislead human interpretation. Therefore, to inspect the color similarity, fused images were degraded to the resolution of the original MS image (i.e. 0.50 m to 2.0 m) and their histograms were matched against that of the original MS image (Figure 4). We selected false-color composites (bands 2, 3, and 4) for visual inspections because this band combination is widely used for many remote sensing applications (Ehlers et al. 2010). The spatial quality inspection involved principle component (PC) analysis of fused products. The first PC of fused images (gray-scale images) was separated and their histograms were matched against the original PAN images (Figure 5). False-color composites and first-PC images along with their original images were inspected by two photo-interpretation experts to identify any spectral distortions, (e.g., brightness

reversions, saturation, a complete change of spectral characteristics, unnatural/artificial colors) and spatial improvement.

2.2.4 Spectral quality

The fused images were resampled and their histograms were matched before analyzing them with the objective quality metrics. From our literature survey, it was found that most studies overlooked the importance of matching resampled image histograms before running spectral quality metrics. We suggest it is better to match histograms after resampling of fused images because some fusion algorithms (e.g. Ehlers fusion) match histograms as the final step of the fusion workflow; however, this step is not obvious in most algorithms. Thus, we brought all the fused images into a common radiometric domain by matching their histograms against original MS images. We used eleven metrics to quantify the post-fusion color similarity (Table 4). Except two global metrics (SAM and ERGAS), other metrics were calculated band-wise and averaged over all bands.

2.2.5 Spatial quality

Reporting best values for spectral metrics does not necessarily mean that the fusion algorithm has made spatial improvements, i.e., a fusion algorithm can exhibit a high degree of spectral preservation even if no pansharpening is performed. It is therefore important to assess the spatial fidelity of fused products but only a few spatial-quality indicators have been proposed in literature. We utilized Canny edge filter (CEC), high-pass correlation coefficient (HP-CC), and RMSE of Sobel-filtered edge images (Sobel-RMSE) to quantify the quality of spatial improvement (Figure 4). We applied Canny

edge detection filter to the original PAN and each band of fused images to produce binary images in which edge-detected pixels scored 1 and others scored 0. The binary image of PAN and that of fused images were compared band-wise to calculate the edge correspondence that was reported as percent agreement. For HP-CC, we applied a high-pass (HP) filter with a 3 x 3 Laplacian kernel (Zhou et al., 2008) to the PAN image and each band of the fused images and computed correlation coefficients between the HP-filtered bands and HP-filtered PAN image were computed. The Sobel filter was applied to the original PAN and the each band of fused images. Sobel's vertical and horizontal gradient operators produced two edge images. The overall magnitude of gradient image was created by taking the Euclidean sum of the two edge images (Pradhan et al.2006). The RMSE was calculated between the PAN gradient image and gradient images of each of the fused bands.

Table 4. Summary of quantitative quality metrics

Quality metric		Mathematical expression	Addressed issue/domain/expected value	Reference
Spectral	Correlation coefficient (CC)	$$CC(B_k\|B_k^*) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\left(B_{k(m,n)} - \bar{B}_k\right)(B_{k(m,n)}^* - \bar{B}_k^*)}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}(B_{k(m,n)} - \bar{B}_k)^2 \sum_{m=1}^{M}\sum_{n=1}^{N}(B_{k(m,n)}^* - \bar{B}_k^*)^2}}$$	• Quantifies the spectral correspondence between the original MS and fused images. • domain [-1,1] • As close to 1 as possible	Gonzalez et al.(2005), Pradhan eta l.(2006), Kalpoma and Kudoh (2007), Karathanassi et al. (2007), Gangkofner et al. (2008), Nikolakopoulos (2008), Ehlers et al.(2010) , Yang et al. (2012)
	Root-mean -square-error (RMSE)	$$RMSE(k) = \sqrt{\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}(B_{k(m,n)} - B_{k(m,n)}^*)^2}{m \times n}}$$ Also defined as; $$RMSE(k) = \sqrt{bias_k^2 + \sigma_k^2}$$ Where, $\sigma_k = \sigma_{original(k)} - \sigma_{fused(k)}$, $bias_k = mean_{original(k)} - mean_{fused}$	• Measures the average amount of spectral distortion in each pixel • domain [0,inf) • Lower value	Wald (2002), Ranchin et al. (2003), Klonus and Ehlers (2007), Gangkofner et al. (2008), Ehlers et al.(2010), Yakhdani and Azizi (2010)
	Relative difference to mean (RDM)	$$RDM = \frac{\bar{B}_k^* - \bar{B}_k}{\bar{B}_k}$$	• Measure the changes in the shape of the histogram of fused image compared to original MS image. • domain (-inf, inf) • As close to 0 as possible	Wald et al. (1997), Gonzalez-Audicana et al.(2005), Vijayaraj et al.(2006), Karathanassi et al. (2007), Kim et al.(2011), Yang et al. (2012)
	Relative difference to standard deviation (RDS)	$$RDS = \frac{\sigma_{B_k^*} - \sigma_{B_k}}{\sigma_{B_k}}$$		Ranchin et al.(2003), Gonzalez-Audicana et al.(2005), Karathanassi et al. (2007), Kim et al.(2011), Yang et al. (2012)
	Spectral discrepancy (SD)	$$SD = \frac{1}{(M \times N)}\sum_{m=1}^{M}\sum_{n=1}^{N}\left\|B_{k(m,n)}^* - B_{k(m.n)}\right\|$$	• Band-wise measure of the spectral quality of the fuse image • domain [0,inf) • As close to 0 as possible	Li et al.(2002), Li et al.(2005), Ling et al.(2007), Yakhdani and Azizi (2010), Guo et al.(2010)
	Deviation index (DI)	$$DI = \frac{1}{(M \times N)}\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{\left\|B_{k(m,n)}^* - B_{k(m.n)}\right\|}{B_{k(m.n)}}$$	• Quantifies the normalized absolute difference of the fused image with the original MS image. • domain [0,inf) • As close to 0 as possible	Costantini et al.(1997), Bethune et al. (1998), Karathanassi et al. (2007), Klonus and Ehlers (2007), Ling et al.(2007), Ehlers et al.(2010)

	Peak signal-to-noise ratio (PSNR)	$$PSNR = 20log_{10}\frac{Peak}{\sqrt{MSE}}$$ Where; $$MSE = \frac{1}{I}\sum_{i=1}^{I}(B^*_{k,i} - B_{k,i})^2$$ *Peak* is the maximum possible pixel value (peak = 255, 2057, and 65535 for 8-bit, 11-bit, and16-bit images, respectively). *I* is the number of non-null pixels.	• Indicates the radiometric distortion of the fused image compared to the original MS image. • The highest possible PSNR	Li and Hu (2004), Garzelli and Nencini (2005), Karathanassi et al. (2007)
	Entropy (E)	$$E = \sum_{i=0}^{R} -p_i log_2(p_i)$$ $$Increase\ in\ information\ (entropy\ difference) = E_{B_k} - E_{B^*_k}$$ Where; p_i is the probability of existence for digital value *i*. *R* is the radiometric resolution of an image (*R* = 255, 2057, and 65535 for 8-bit, 11-bit, and16-bit images, respectively).	• Measures the additional information (spectral and spatial) available in the fused image compared to the original MS image. • The smallest possible entropy difference with the original MS image.	Bethune et al. (1998), Vijayaraj et al.(2006), Karathanassi et al. (2007), Yakhdani and Azizi (2010)
	Mean structural similarity index (MSSIM)	$$MSSIM(B_k|B^*_k) = \frac{1}{W}\sum_{j}^{W} SSIM(x_j y_j)$$ $$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2+\mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$ Where; x_i and y_i are the images contents of the original and the fused image at local window *j*; and *W* is the number of local windows of the image. μ is the mean intensity, σ is the standard deviation, C_1 and C_2 are constants.	• Reveals the spectral and structural similarity between the fused and original MS image by luminance, contrast, and structure and applying to a moving window. • domain [0,1] • As close to 0 as possible	Wang et al.(2004), Pradhan et al.(2006), Ling et al.(2007), Renyuan et al.(2009), Ehlers et al.(2010), Weidner (2010)
	Spectral angle mapper (SAM)	$$SAM(B,B^*) = \arccos\left[\frac{\sum_{k=1}^{K} B_k B^*_k}{\sqrt{\sum_{k=1}^{K} B_k^2}\sqrt{\sum_{k=1}^{K} B_k^{*2}}}\right]$$ Where; *K* is the number of bands	• Pixel-wise comparison of fused image and original MS image. The value 0 indicates low resemblance while 1 indicates a high resemblance. • domain [0,1] • As close to 0 as possible	Goetz et al.(1992), Garzelli and Nencini (2005), Jenson (2005), Aiazzi et al.(2007), Alparone et al.(2007), Du et al.(2007), Jing and Chen (2011)
	Relative dimensionless global error in synthesis (ERGAS)	$$ERGAS = 100\frac{h}{l}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\frac{RMSE(k)}{\mu(k)}\right)^2}$$ Where; *h/l* is the ratio between the pixel sizes of PAN and MS images, RMSE(*k*) and $\mu(k)$ are the root mean squared error and mean of the k^{th} band, respectively. *K* is the total number of bands.	• A global indicator that calculates the amount of spectral distortion. • domain [0,inf) • Lower value (< 3)	Wald (2000), Ranchin et al.(2003), , Gonzalez et al.(2005), Alparone et al.(2007), Kalpoma and Kudoh (2007) .
Spatial	Canny edge correspondence (CEC)	$$CEC = \left[\frac{\sum_{r=1}^{R}\sum_{c=1}^{C}(\hat{A}_{(r,c)} \times \hat{B}^{**}_{k(r,c)})}{\sum_{r=1}^{R}\sum_{c=1}^{C}\hat{A}_{(r,c)}}\right] \times 100$$ Where;	• A band-wise comparison of edges detected in the original PAN and the fused image. CES measured in percent. • domain [0,100]%	Canny (1986), Ehlers et al.(2010), Yakhdani and Azizi (2010)

		$\hat{A}, \hat{B}$ are the binary-edge images of the original PAN and degraded fused images.	• as close to 100 as possible	
	High-pass(HP) correlation coefficient (HP-CC)	$$CC(A\|B_k^{**}) = \frac{\sum_{r=1}^{R}\sum_{c=1}^{C}\left(A_{(r,c)} - \bar{A}\right)(B_{k(r,c)}^{**} - \bar{B}_k^{**})}{\sqrt{\sum_{r=1}^{R}\sum_{c}^{C}(A_{(r,c)} - \bar{A})\sum_{r=1}^{R}\sum_{c=1}^{C}(B_{k(r,c)}^{**} - \bar{B}_k^{**})^2}}$$	• Quantifies the correlation between the HP filtered bands of fused image and the HP-filtered PAN image. • domain [-1,1] • as close to 1 as possible	Zhou et al.(1998), Li et al.(2002), Gangkofner et al. (2008), Ehlers et al.(2010), Yakhdani and Azizi (2010)
	RMSE of Sobel filtered Pan and fused images (Sobel-RMSE)	$$RMSE(k) = \sqrt{\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}(SA_{(m,n)} - SB_{k(m,n)}^{*})^2}{m \times n}}$$ Where; *SA* and SB_k^{**} are the Sobel-filtered images of original PAN and fused images, respectively	• Measures the average amount of spatial distortion in each pixel • domain [0,inf) • Lower value	Pradhan et al. (2006), Klonus and Ehlers (2007), Gangkofner et al. (2008), Yang et al. (2012)

2.3 Results

2.3.1 Visual evaluation

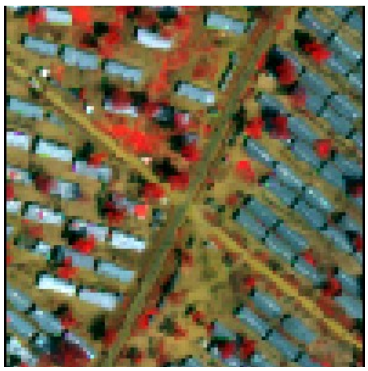
Figures 7, 8, and 9 depict the sites selected from Sri Lanka (Menik Farm - site 2), Pakistan (Nowshera - site 2), and Haiti (Port-Au-Prince - site 2), respectively. For color similarity, we choose tests areas with spectrally heterogeneous objects such as vegetation, roof tops, and roads. For assessing spatial quality, the same test area was zoomed into such that distinct objects with sharp edges (e.g. buildings, IDP shelters) were clearly visible. We selected the best two and the worst two fusion algorithms in terms of spectral and spatial fidelity. Ranking of twelve fusion algorithms (i.e., best to worst) based on their performances is a difficult task even for an expert photointerpreter; although the human brain easily discriminates among extreme variations (i.e., best fusion and worst fusion), it fails to distinguish subtle variations among images with similar spectral and spatial properties. Table 5 tabulates the ranking of fusion algorithms based on expert evaluation.

Table 5. Objective evaluation of fused images by experts

Test area	Spectral similarity		Spatial similarity	
	Best	Worst	Best	Worst
Menik Farm (Figure 4)	LMVM, Gram-Schmidt	CN sharpening, SRM	CN sharpening, UNB	Ehlers, Wavelet
Nowshera (Figure 5)	Ehlers, HPF	Brovey, SRM	Brovey, CN sharpening, UNB	LMVM, Ehlers, LMM
Port-Au-Prince (Figure 6)	Ehlers, LMVM	CN sharpening, SRM	Brovey, CN sharpening	Ehlers, LMVM

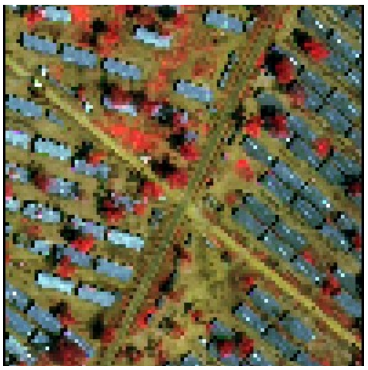
Visually comparing the original multispectral and fused images of test site-2 in Menik Farm IDP camp (Figure 7) revealed that the CN sharpening algorithm has the worst color distortions. CN seriously changed the colors of vegetation and IDP shelters.

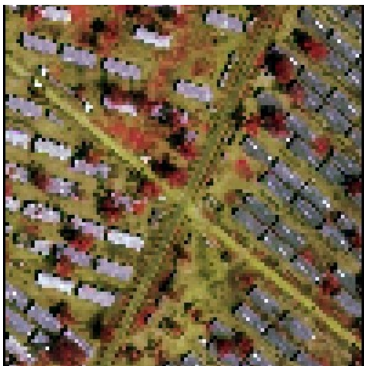
The SRM algorithm preserved basic spectral characteristic of the vegetation, IDP shelters, and bare soil; however SRM has produced a high contrast image. Based on notable spectral degradations, CN and SRM can be considered as the worst algorithms. Of the remaining ten algorithms, selecting the best-two algorithms is challenging because they preserved most of the color information of the original MS image. Focusing on IDP shelters in the original image, the shelters show two spectral appearances: the roof-tops to the right edge of the image appear dark gray whereas those to the left are light gray. Visually, all roof tops are smooth and identifiable on the original image. Among the ten contenders, LMVM, Gram-Schmidt, HPF, Ehlers, PC, and UNB showed nearly equal performances; however, of these, HPF gave a high contrast image compared to the original MS image. In Ehlers, roof-tops are not easily identifiable compared to the original and results from the other algorithms. The LMVM fusion algorithm had the best spectral agreement with the original multispectral image and Gram-Schmidt fusion algorithm can be considered as the runner-up algorithm. When assessing the spatial improvement of the same test area (Figure 7, Plate 3), the wavelet algorithm appear to give the worst spatial improvement. It produced a "fuzzy" image with almost no edge enhancements. The Ehlers algorithm performed better than Wavelet; however, when compared to other algorithms, Ehlers algorithm produced visually poor results. Therefore, Ehlers and Wavelet can be considered as the worst-two algorithms in terms of spatial improvement. In general, except Wavelet, Ehlers, LMVM, and PC, the other eight algorithms showed good spatial agreement with the original PAN image. Among this subset of contenders, CN and UNB fusion algorithms can be elected as the best two algorithms.

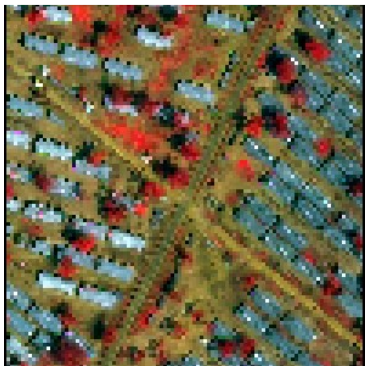


Original MS image Original PAN image

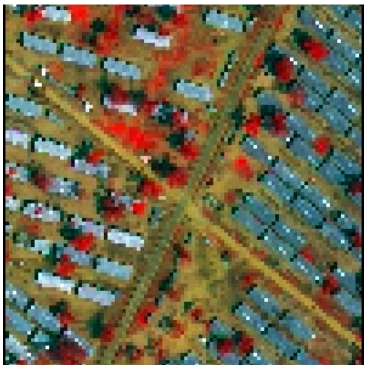
Plate 1. Original images of selected test area for visual evaluation

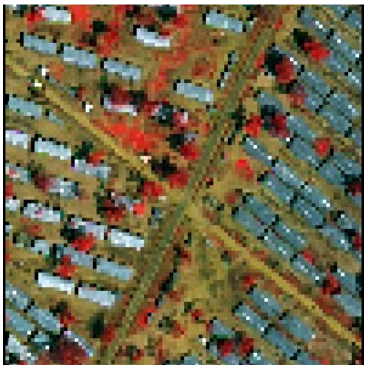


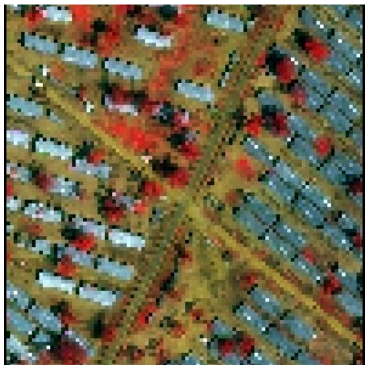




BT CN EH







GS HPF LMM

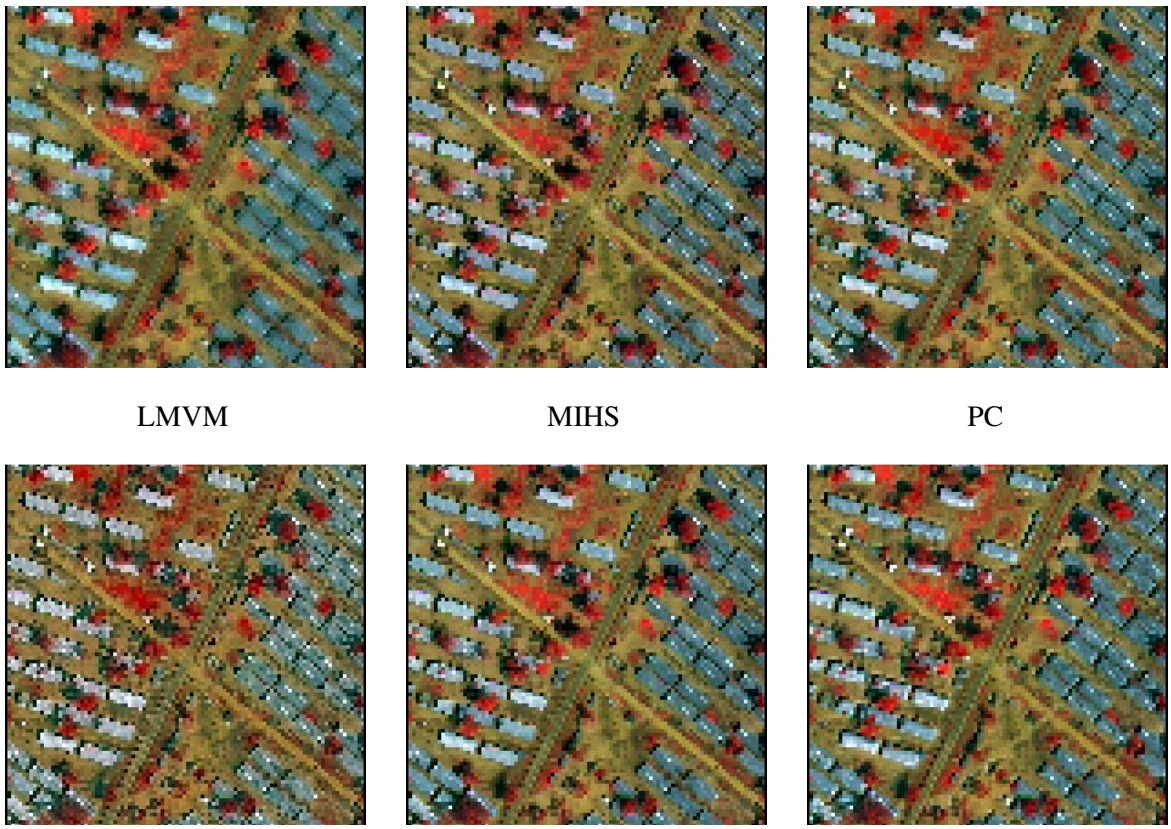


Plate 2. Resampled and histogram-matched fused images of the test area

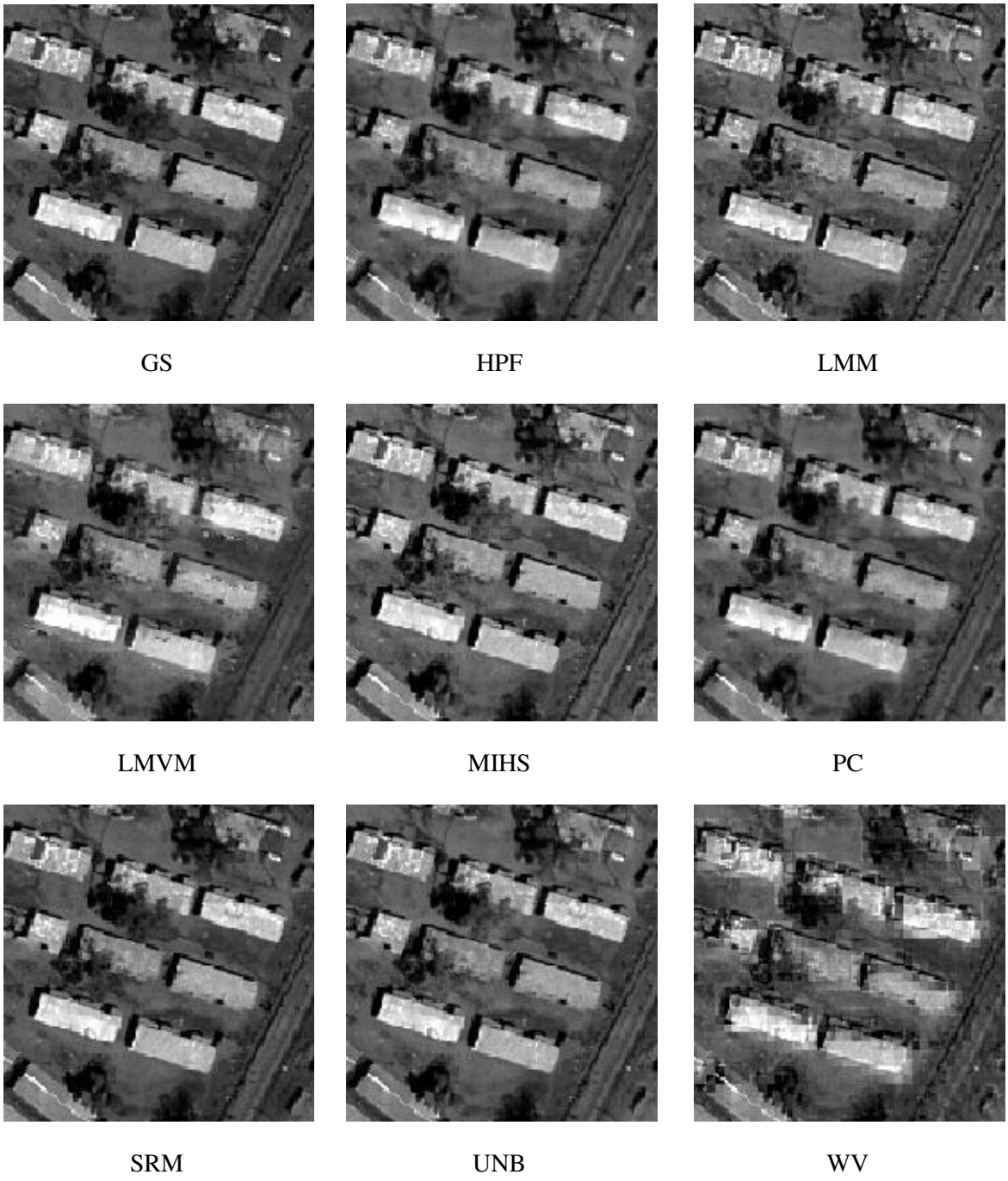
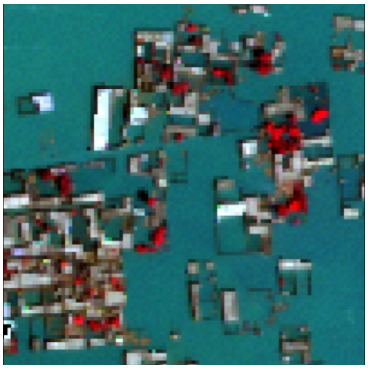


Plate 3. First-PC images of the fused data

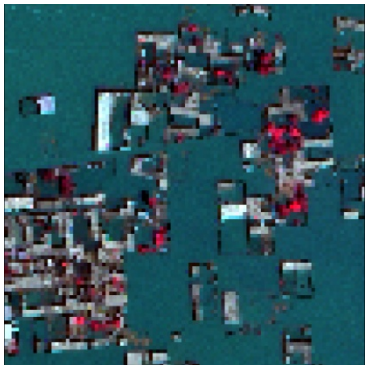
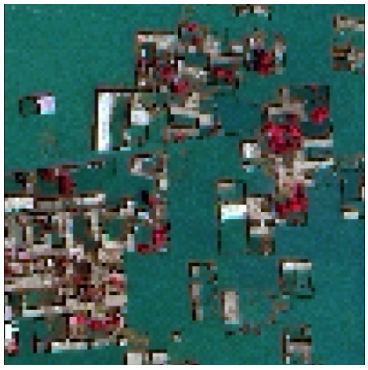
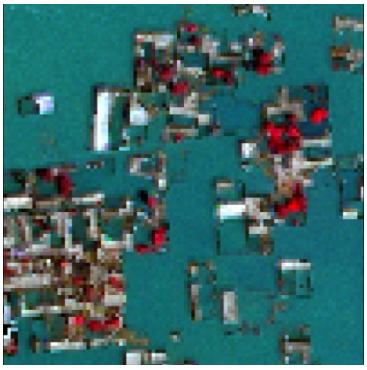
Figure 7. Original imagery and fusion results of Menik Farm IDP camp(site-2), Sri Lanka. Original MS image and fused images are shown as bands 2,3,4 composites.

Figure 8 shows the original and fused images of the flooded buildings in Nowshera, Pakistan. Water, buildings, and vegetation are the main identifiable features. When inspecting the color similarity between the original multispectral and fused images, Brovey transform and SRM algorithm produced the worst spectral preservation. The best color agreement of floodwater with the original MS image was from the Ehlers and MIHS algorithms. However, MIHS showed tree canopies in dark red whereas Ehlers showed bright red canopies as seen in the original MS image. LMVM, LMM, and WV produced poor results for inundated buildings. Despite the slight color changes in floodwater, GS, HPF, and UNB algorithms produced satisfactory fusion results. Overall, we think Brovey and SRM are the worst two algorithms whereas Ehlers and HPF are the best-two. To assess the spatial improvement, we focused on a partially inundated structure (Figure 8, Plate 1) with well emphasized edges. When assessing first-PC images, the LMVM produced the worst fusion results with poorly emphasized edges. A similar fuzziness can be seen in the fused images of Ehlers and LMM. An equal spatial improvement, which is superior to Ehlers and LMM, is observed in HPF, PC, and Wavelet algorithms. Of the remaining contenders it is difficult to select the best-two fusion algorithms because they show little difference from each other. Any two of Brovey, CN, and UNB can be elected as the best-two algorithms.

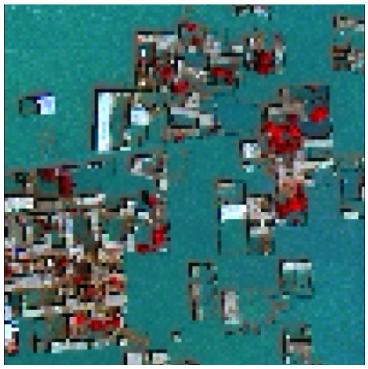
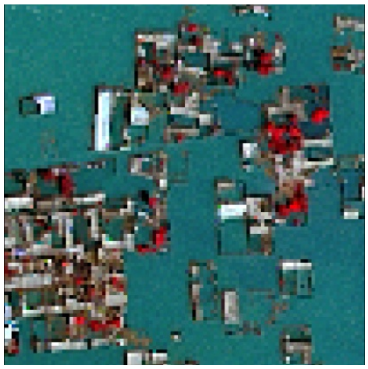
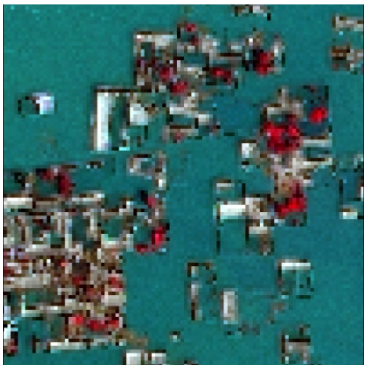



Original MS Original PAN

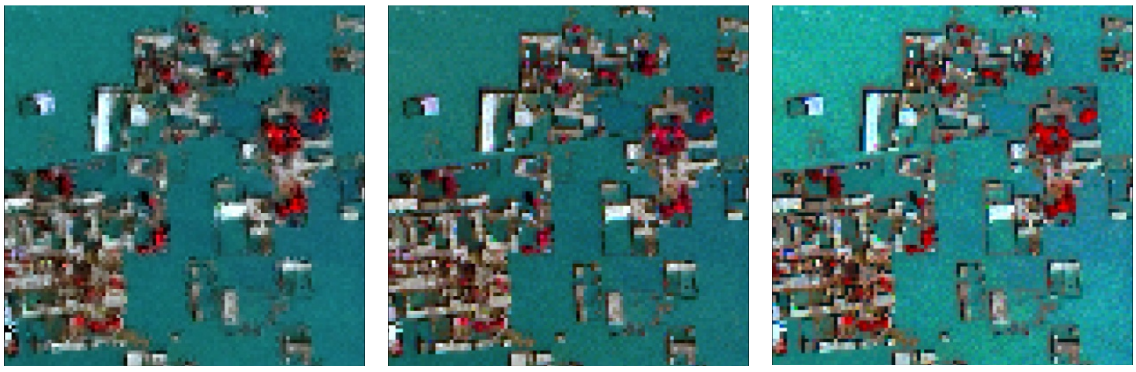
Plate 1. Original images of selected test area for visual evaluation

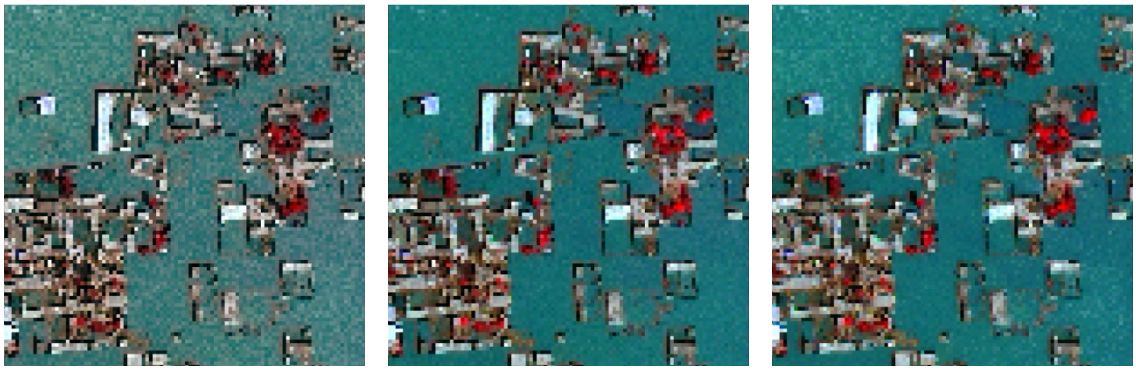
BT CN EH

GS HPF LMM



LMVM MIHS PC



SRM UNB WV

Plate 2. Resampled and histogram-matched fused images of the test area

.

BT CN EH

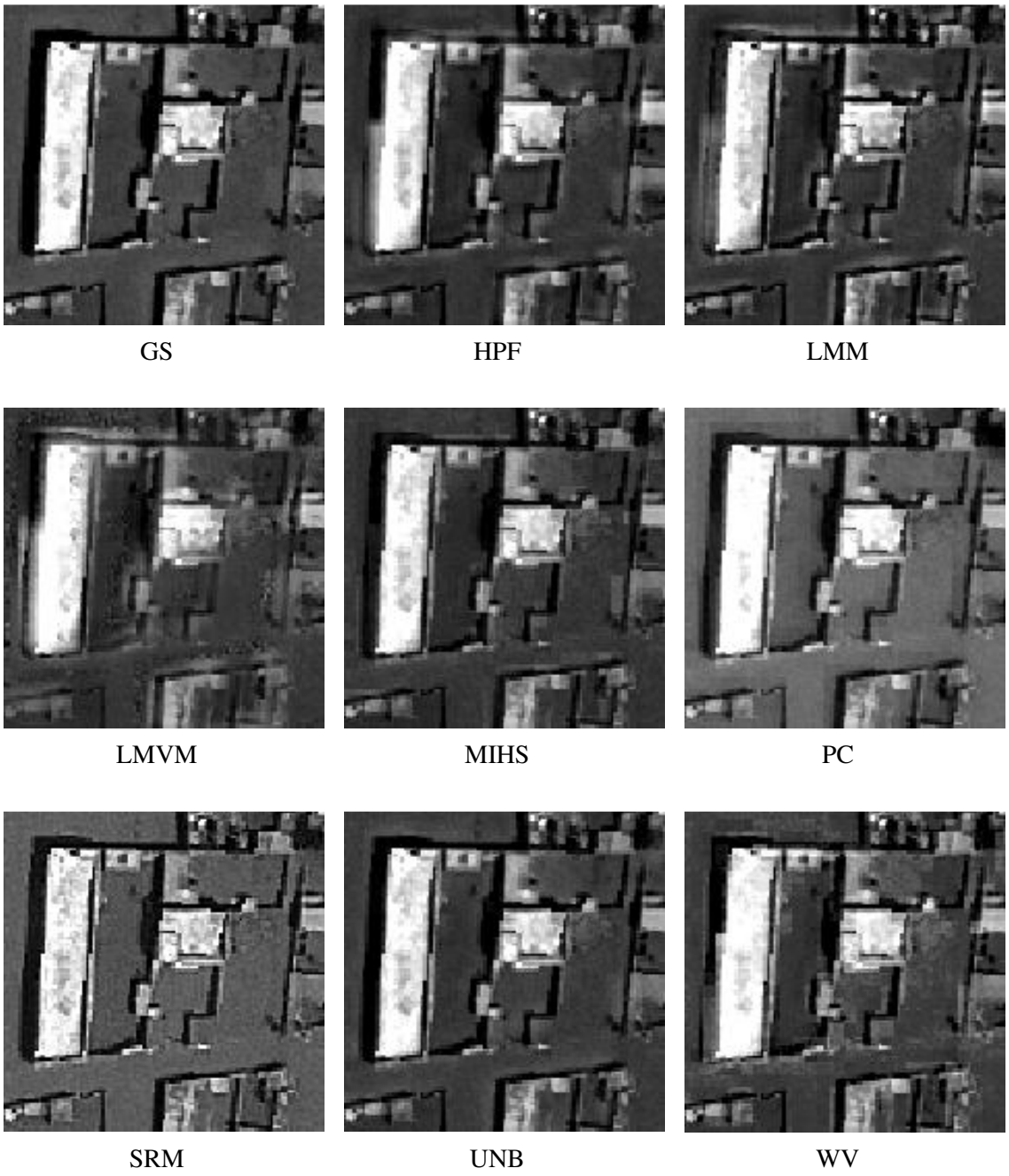


Plate 3. First-PC images of the fused data

Figure 8. Original imagery and fusion results of Nowshera (site-2), Pakistan. Original MS image and fused images are shown as bands 2,3,4 composites`

The representative site of Port-Au-Prince mainly entails intact and partially collapsed buildings, a portion of a turf field, and urban tree canopies (Figure 9). Regarding the spectral similarity, except Ehlers and LMVM algorithms, all the other fusion algorithms made severe or moderate color changes to the vegetation. Among remaining candidates, Gram-Schmidt and UNB algorithms gave an acceptable spectral agreement with the original MS image. Major color distortions can be observed in the fused images of Brovey, CN, MIHS, PC, and SRM algorithms. Of these, CN and SRM can be reported as the worst-two algorithms. The fused PC method image showed an over saturation of red in tree canopies and unnatural colors (e.g. purple) over the rooftops. In terms of spatial fidelity, Ehlers, LMVM, and Wavelet algorithms produced unsatisfactory fusion results. Edges of these fused products are not as sharp as those of the other candidates and the original PAN image. Of these three algorithms, we elected Ehlers method and LMVM algorithm as the worst-two algorithms. Brovey and CN algorithms yielded almost identical fusion results compared to each other and as well as to the original PAN image. The UNB algorithm also achieved visually-appealing spatial improvements slightly inferior to Brovey and CN algorithms but superior to HPF pansharpening technique. Based on our visual inspections, we choose Brovey and CN as the best-two algorithms for spatial fidelity.

Original MS

Original PAN

Plate 1. Original images of selected test area for visual evaluation

BT

CN

EH

GS

HPF

LMM

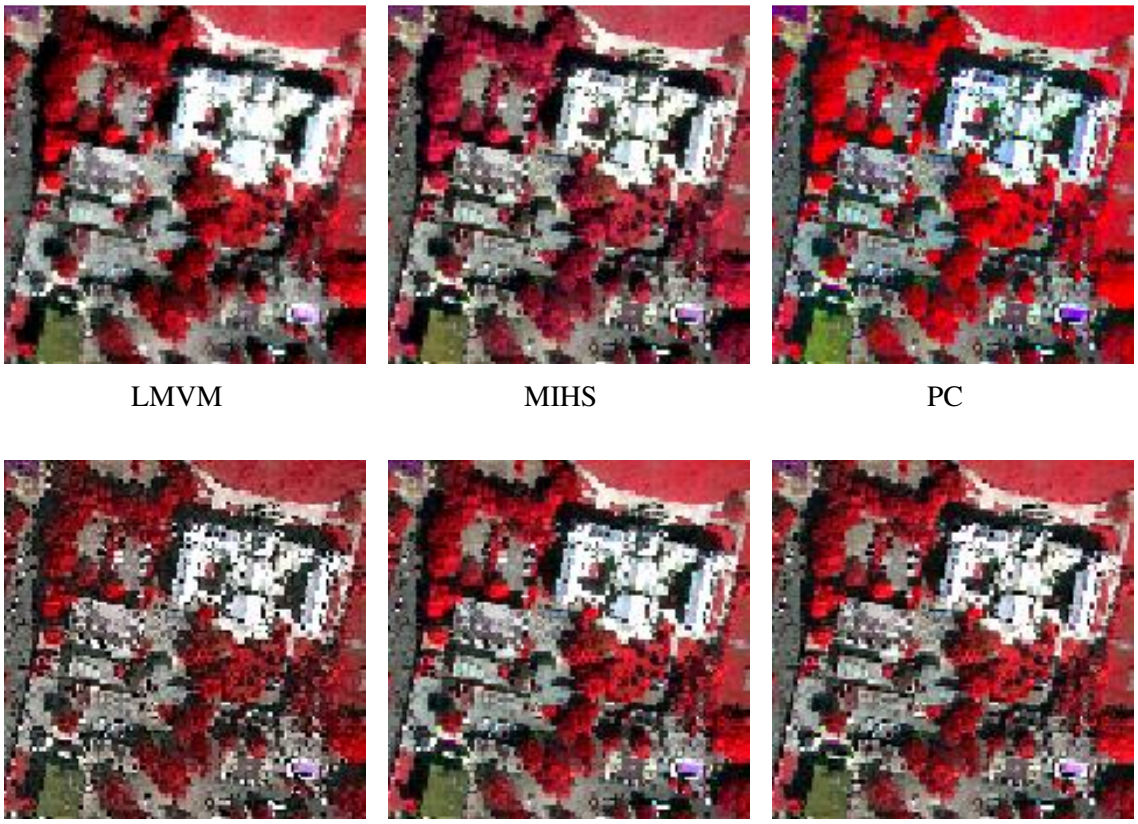
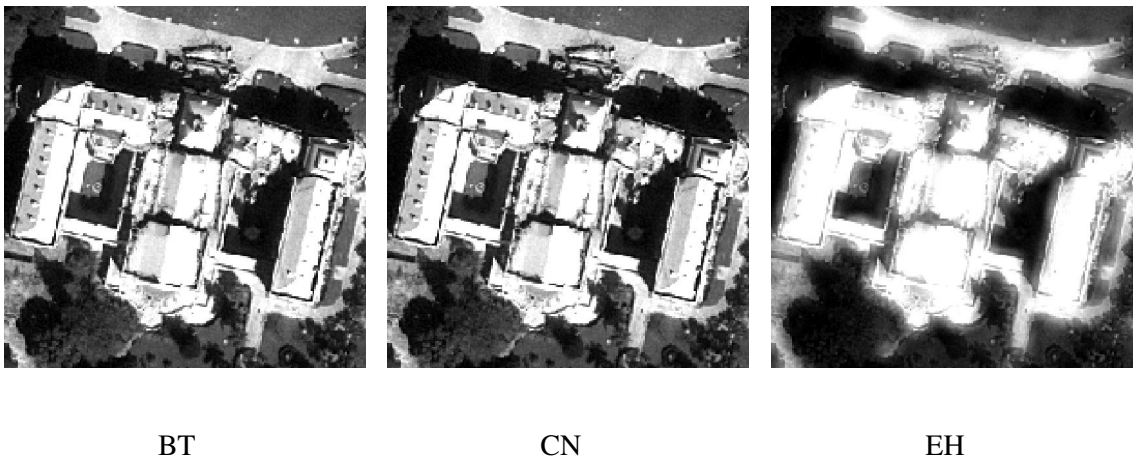


Plate 2. Resampled and histogram-matched fused images of the test area



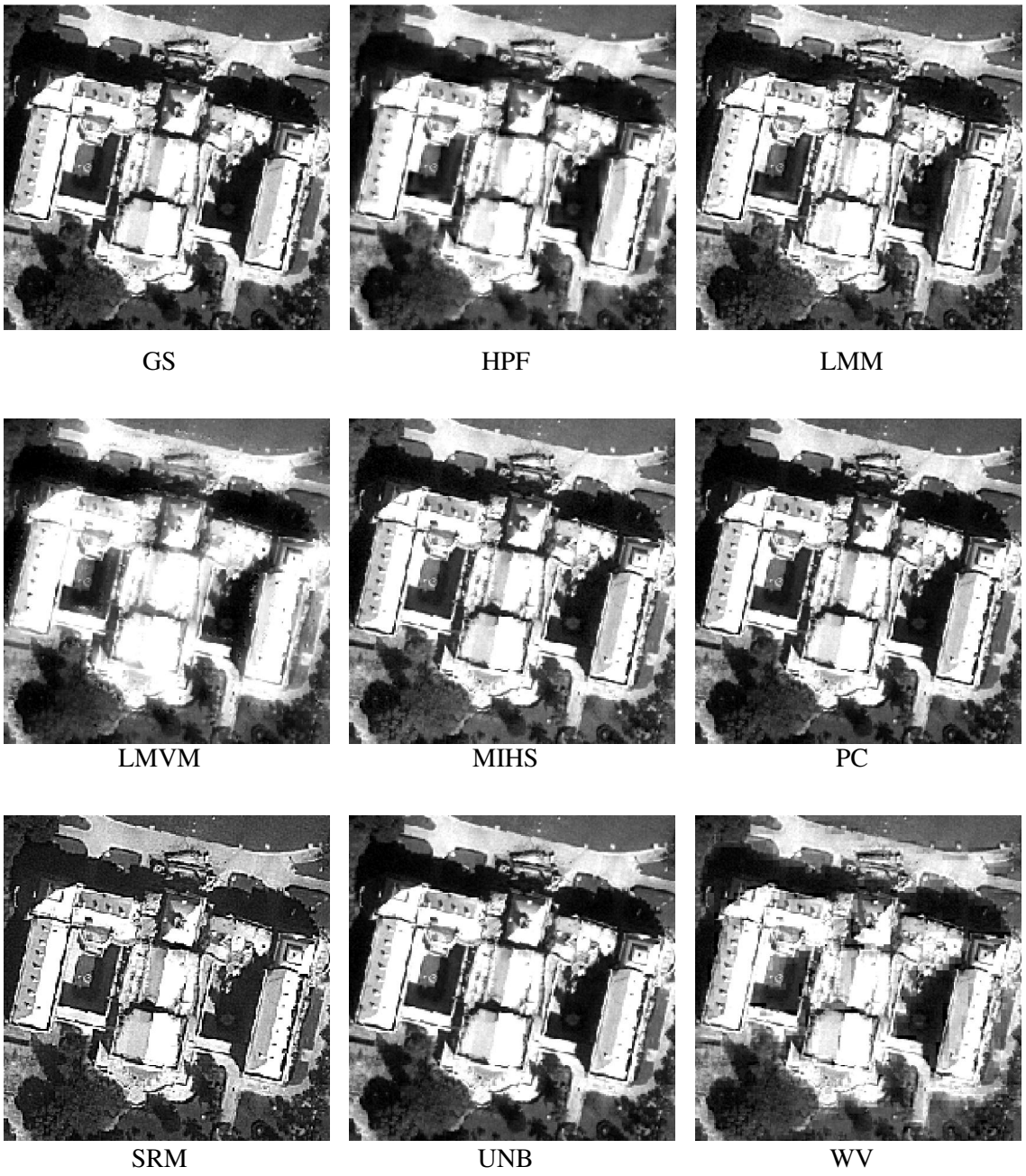


Plate 3. First-PC images of the fused data

Figure 9. Original imagery and fusion results of Port-Au-Prince (site-2), Haiti. Original MS image and fused images are shown as bands 2,3,4 composites.

2.3.2 Quantitative assessment

We corroborated visual assessment with eleven spectral metrics and three spatial metrics. Tables 6, 7, 8, 9, 10, and 11 summarize the mean scores (averaged over bands and subsets) reported by quality metrics for the five study areas. Fusion algorithms in each table are ranked by their correlation coefficient scores. We ranked fusion algorithms in each column as 1, 2, and, 3 and three gray levels as dark-, medium-, light-gray were used to highlight these three ranks, respectively. The worst value reported for a given quality metric is in bold font.

2.3.4 Spectral and spatial fidelity

Killinochchi, Sri Lanka

Table 6 (columns 2 through 11) tabulates the calculated mean values of all objective quality metrics for the two subsets (KL-1 and KL-2) of the Town of Killinochchi, Sri Lanka. Overall, Ehlers, WV, and LMVM algorithms scored the first-three ranks for the majority of quality estimators. All algorithms showed equal performance for RMD and RDS, which questions the potential of these two metrics serving further as discriminators. Ehlers algorithm received the first rank for six metrics (CC, RMSE, SD, MSSIM, ERGAS, and SAM). The Wavelet algorithm showed the second-best performances. The LMVM algorithm had slightly inferior results compared to Ehlers and WV algorithms; however, it outperformed the other nine fusion techniques. For PSNR, both Ehlers and WV algorithms reported high values but LMVM scored a value close to the worst value earned by LMM algorithm. The PC fusion method showed the worst values for six quality indices (CC, RMSE, SD, MSSIM, ERGAS, and SAM)

while SRM, LMM, and CN algorithms scored the worst values for DI, PSNR, and Entropy, respectively.

The spectral quality of Ehlers, WV, and, LMVM were superior to other algorithms, but they were not necessarily superior in their spatial performances. The CN sharpening obtained the highest values for Canny edge correspondence and high-pass correlation coefficient; however, it exhibited relatively high value for the RMSE of Sobel-filtered images (Sobel-RMSE). In contrast, the UNB algorithm consistently showed promising scores for all the three spatial metrics. The WV algorithm showed the worst values for CEC and Sobel-RMSE. The PC fusion showed the lowest value for HP-CC (0.04) and relatively poor values for the remaining two metrics. This highlights the consistent failure of the PC algorithms with respect to both spectral and spatial estimators. It is noteworthy that algorithms like GS, HPF, MIHS, and UNB produced satisfactory and consistent values for both spectral and spatial quality metrics. Overall, we observe a reversal of ranking when comparing spectral indices with spatial indices.

Table 6. Reported scores of spectral and spatial quality metrics for the Town of Killinochchi (KL-1 and KL-2), Sri Lanka

Fusion algorithm	Spectral metric											Spatial metric		
	CC	RMSE	RDM	RDS	SD	DI	PSNR	Entropy	MSSIM	ERGAS	SAM	Canny edge	HP-CC	Sobel-RMSE
Ehlers	0.96	22.68	0.00	0.00	15.80	0.04	70.15	0.15	0.96	1.62	0.80	86.35	0.90	19.78
WV	0.95	25.54	0.00	0.00	17.86	0.03	81.33	0.48	0.93	1.65	2.41	**81.08**	0.07	**34.28**
LMVM	0.95	29.88	0.00	0.00	19.57	0.05	38.55	0.04	0.94	2.07	1.56	87.58	0.80	25.53
GS	0.93	33.26	0.00	0.00	23.87	0.07	67.22	0.13	0.92	2.46	1.84	86.98	0.93	17.33
HPF	0.92	34.85	0.00	0.00	23.92	0.07	66.55	0.08	0.92	2.72	1.33	90.06	0.95	18.05
MIHS	0.91	34.09	0.00	0.00	23.96	0.07	66.69	0.10	0.92	2.40	1.26	88.23	0.87	19.43
UNB	0.90	38.03	0.00	0.00	26.47	0.08	65.87	0.10	0.90	2.96	1.34	90.18	0.96	17.80
LMM	0.88	39.97	0.00	0.00	27.22	0.07	**35.47**	0.07	0.89	2.73	1.56	89.64	0.95	18.94
Brovey	0.85	37.63	0.00	0.00	26.98	0.08	65.26	0.84	0.90	2.96	1.91	90.17	0.89	22.91
CN	0.80	52.88	0.00	0.00	39.06	0.10	63.52	**1.19**	0.85	3.46	3.79	92.49	0.98	20.98
SRM	0.77	63.75	0.00	0.00	47.58	**0.12**	62.83	0.10	0.80	4.21	3.50	89.56	0.96	28.89
PC	**0.66**	**68.58**	0.00	0.00	**53.15**	0.10	80.36	0.44	**0.70**	**4.49**	**7.25**	86.76	**0.04**	23.10

Fusion methods :CN – color normalization spectral sharpening, GS – Gram-Schmidt, HPF – High-pass filter, LMM – local mean matching, LMVM – Local mean variance matching,, MIHS –Modified intensity-hue,saturation, PC – Principal component analysis, SRM – Subtractive resolution merge, UNB – University of New Brunswick, WV – Wavelet PCA

Quality metrics: CC – Correlation coefficient, RMSE – Root-mean-square-error, RDM – Relative distance to mean, RDS – Relative distance to standard deviation, SD – Spectral discrepancy, DI – Deviation index, PSNR – Peak signal-to-noise ratio, MSSIM- Mean structure similarity index, ERGAS – relative dimensionless global error, SAM – Spectral angle mapper, HP-CC – High-pass correlation coefficient

Rank 1 Rank 2 Rank 3

**Worst value are indicated in bold

Civilian Safety Zone, Sri Lanka

Table 7 summarizes the mean scores of the objective quality indices reported for the two subsets of civilian safety zone (CSZ-1, CSZ-2) in Sri Lanka. Similar to Killinochchi, Ehlers, LMVM, and WV algorithms achieved the best results for spectral quality estimators. However, the ranking shows a slight change: although Ehlers remains first, the LMVM and WV switched their ranks as 2 and 3, respectively. Again, both RDM and RDS report zero value for all fusion techniques, showing their inability to discriminate the difference in results from fusion algorithms. The SRM fusion reported the poorest values for CC, SD, MSSIM, ERGAS, and SAM. In general UNB, HPF, MIHS, GS, and Brovey algorithms reported CCs higher than 0.90 and also showed satisfactory scores for the majority of quality indicators without reaching poor values.

The values reported for the spatial estimators by Ehlers, LMVM, and WV algorithms clearly reflect their loss of dominance when the fused products are evaluated for spatial fidelity. The Ehlers fusion scored values less than 0.90 for both CEC and HP-CC. The LMVM algorithm managed to achieve 91.09 for CEC but failed to score over 0.90 for HP-CC. Similar to Killinochchi, the WV algorithm reported the worst values for CEC and Sobel-RMSE and the second-worst value for HP-CC. The PC fusion showed the lowest value for HP-CC. For Canny edge correspondence, the CN fusion yielded the highest value. The highest value for HP-CC was achieved by both CN and UNB algorithms showing equal performances. Similarly, LMM and SRM exhibited the lowest and identical values for Sobel-RMSE (11.62).

Table 7. Reported scores of spectral and spatial quality metrics for the two study sites of Civilian safety zone (CSZ-1 and CSZ-2), Sri Lanka

Fusion algorithm	Spectral metric											Spatial metric		
	CC	RMSE	RDM	RDS	SD	DI	PSNR	Entropy	MSSIM	ERGAS	SAM	Canny edge	HP-CC	Sobel-RMSE
Ehlers	0.98	40.84	0.00	0.00	7.17	0.03	60.91	0.17	0.98	2.10	0.86	89.82	0.89	14.44
LMVM	0.96	42.02	0.00	0.00	9.74	0.05	42.70	0.06	0.96	2.48	1.39	91.09	0.87	14.34
WV	0.95	46.95	0.00	0.00	13.78	0.07	56.74	0.09	0.93	2.76	1.77	**88.46**	0.78	**15.08**
UNB	0.92	48.83	0.00	0.00	15.22	0.08	55.12	0.15	0.93	2.94	1.39	92.34	0.95	12.81
HPF	0.92	53.69	0.00	0.00	16.07	0.08	54.69	0.12	0.92	2.98	1.50	92.26	0.94	12.95
GS	0.91	54.43	0.00	0.00	17.20	0.09	54.54	0.16	0.91	3.08	1.88	92.23	0.94	12.80
MIHS	0.91	55.19	0.00	0.00	14.11	0.07	55.11	0.09	0.94	2.80	1.08	91.08	0.89	12.91
Brovey	0.90	56.44	0.00	0.00	14.66	0.07	54.85	0.84	0.93	2.85	1.23	92.36	0.92	13.99
PC	0.89	57.24	0.00	0.00	20.05	0.12	53.65	0.68	0.92	3.19	2.45	89.97	**0.76**	14.32
CN	0.88	58.66	0.00	0.00	17.08	0.08	53.90	**0.99**	0.91	3.05	1.95	93.26	0.95	14.18
LMM	0.88	64.16	0.00	0.00	15.79	0.08	**38.94**	0.06	0.91	2.90	1.38	92.13	0.94	11.62
SRM	**0.84**	**74.43**	0.00	0.00	**24.57**	**0.13**	51.81	0.11	**0.86**	**3.70**	**2.61**	92.13	0.94	11.62

Fusion methods :CN – color normalization spectral sharpening, GS – Gram-Schmidt, HPF – High-pass filter, LMM – local mean matching, LMVM – Local mean variance matching,, MIHS –Modified intensity-hue,saturation, PC – Principal component analysis, SRM – Subtractive resolution merge, UNB – University of New Brunswick, WV – Wavelet PCA

Quality metrics: CC – Correlation coefficient, RMSE – Root-mean-square-error, RDM – Relative distance to mean, RDS – Relative distance to standard deviation, SD – Spectral discrepancy, DI – Deviation index, PSNR – Peak signal-to-noise ratio, MSSIM- Mean structure similarity index, ERGAS – relative dimensionless global error, SAM – Spectral angle mapper, HP-CC – High-pass correlation coefficient

Rank 1 Rank 2 Rank 3

**Worst value are indicated in bold

The Town of Puthukuduirippu

Table 8 summarizes the average scores of spatial and spectral quality estimators for the two study sites of the Town of Puthukuduirippu (PU-1, PU-2), Sri Lanka. Regarding the spectral quality of fused images, of the eleven metrics, RDM and RDS showed no ability to discriminate between the results from different fusion algorithms. Otherwise, Wavelet, Ehlers, and HPF algorithms reported the best scores. For MSSIM, WV, Ehlers, and HPF showed equal results (0.91). The Ehlers fusion achieved the lowest values for RMSE, SD, ERGAS, and SAM. The SRM algorithm showed the worst fusion results in terms of CC, RMSE, SD, DI, MSSIM, and ERGAS. The CN fusion exhibited the lowest scores for entropy difference and the highest for SAM while LMM reported the lowest value for PSNR.

When assessing the spatial fidelity of fused images, the wavelet algorithm exhibited the worst scores for all three spatial indices. Considering all spatial indicators, the UNB algorithm shows the best performances, which reported ranks of 1, 2, and 3 for HP-CC, Sobel-RMSE, and CEC, respectively. Although the CN fusion scored the highest values for CEC and HP-CC, it exhibits a relatively high value for Sobel-RMSE (32.53). Showing the second-worst values for CEC and HP-CC along with relatively high value for Sobel-RMSE proves the consistent failure of the PC fusion method.

Table 8. Reported scores of spectral and spatial quality metrics for the Town of Puthukuduirippu (PT-1 and PT-2), Sri Lanka

Fusion algorithm	Spectral metric											Spatial metric		
	CC	RMSE	RDM	RDS	SD	DI	PSNR	Entropy	MSSIM	ERGAS	SAM	Canny edge	HP-CC	Sobel-RMSE
WV	0.93	40.62	0.00	0.00	30.01	0.08	67.19	0.04	0.91	2.65	1.90	**79.23**	**0.68**	**37.44**
Ehlers	0.92	34.39	0.00	0.00	25.50	0.08	66.45	0.14	0.91	2.52	1.24	84.05	0.94	30.24
HPF	0.92	39.22	0.00	0.00	29.31	0.10	65.47	0.05	0.91	3.12	1.61	87.72	0.95	27.26
LMVM	0.91	41.21	0.00	0.00	29.44	0.09	35.44	0.06	0.90	2.99	2.16	86.18	0.87	34.65
UNB	0.88	46.65	0.00	0.00	34.40	0.11	63.87	0.07	0.87	3.76	1.41	89.16	0.98	25.34
PC	0.88	52.95	0.00	0.00	41.67	0.11	65.35	0.47	0.87	3.37	3.44	82.25	0.60	33.55
GS	0.87	48.95	0.00	0.00	36.87	0.12	63.75	0.12	0.86	3.73	2.45	86.48	0.96	24.92
MIHS	0.83	50.92	0.00	0.00	38.08	0.11	63.15	0.07	0.85	3.67	1.69	87.39	0.91	27.30
LMM	0.82	57.11	0.00	0.00	42.29	0.12	**32.29**	0.06	0.82	4.02	2.41	88.95	0.97	27.74
Brovey	0.81	52.31	0.00	0.00	39.39	0.13	62.57	0.73	0.84	3.99	2.16	89.04	0.93	32.88
CN	0.76	65.32	0.00	0.00	50.06	0.14	61.42	**1.02**	0.79	4.49	**4.37**	91.84	0.98	32.53
SRM	**0.71**	**77.25**	0.00	0.00	**59.20**	**0.18**	60.26	0.14	**0.72**	**5.57**	3.68	89.77	0.97	36.39

Fusion methods :CN – color normalization spectral sharpening, GS – Gram-Schmidt, HPF – High-pass filter, LMM – local mean matching, LMVM – Local mean variance matching,, MIHS –Modified intensity-hue,saturation, PC – Principal component analysis, SRM – Subtractive resolution merge, UNB – University of New Brunswick, WV – Wavelet PCA

Quality metrics: CC – Correlation coefficient, RMSE – Root-mean-square-error, RDM – Relative distance to mean, RDS – Relative distance to standard deviation, SD – Spectral discrepancy, DI – Deviation index, PSNR – Peak signal-to-noise ratio, MSSIM- Mean structure similarity index, ERGAS – relative dimensionless global error, SAM – Spectral angle mapper, HP-CC – High-pass correlation coefficient

Rank 1 Rank 2 Rank 3

**Worst value are indicated in bold

Menik Farm IDP camp, Sri Lanka

Table 9 shows the mean scores of the spectral and spatial indices reported for the two subsets of Menik Fram IDP camp (MF-1, MF-2) in Sri Lanka. With respect to spectral indices, the HPF algorithm ranks as the best-candidate and LMVM and WV algorithms hold the second- and the third-best positions, respectively. The RDM reported zero value for all fusion algorithms whereas the RDS showed 0.01 shift in standard deviation for all the algorithms other than WV and LMVM. The SRM algorithm produced the worst values for CC, RMSE, SD, MSSIM, and ERGAS. The worst values for DI, entropy difference, and SAM was reported by the CN fusion algorithm. In general, algorithms like Ehlers, UNB, PC, and GS have shown average performances.

In terms of spatial indices, CN and UNB algorithms exhibited the best performances. The former reported the highest value for the CES while the latter reported the lowest value for Sobel RMSE. Both the former and the latter exhibited equal values for the HP- CC (0.98). Despite the fact that the CN fusion holds rank 1 for CEC and HP-CC, it has reported the worst value for Sobel-RMSE. Similar to other study areas, none of the algorithms producing results with high spectral agreement, achieved notable scores for spatial estimators.

Table 9. Reported scores of spectral and spatial quality metrics for the two study sites of Menik Farm (MF-1 and MF-2), Sri Lanka

Fusion algorithm	Spectral metric											Spatial metric		
	CC	RMSE	RDM	RDS	SD	DI	PSNR	Entropy	MSSIM	ERGAS	SAM	Canny edge	HP-CC	Sobel-RMSE
HPF	0.92	40.84	0.00	0.01	25.57	0.06	49.50	0.07	0.92	2.84	1.10	91.94	0.95	22.59
LMVM	0.91	42.02	0.00	0.00	25.63	0.07	34.40	0.06	0.91	2.82	1.70	90.45	0.82	33.05
WV	0.89	46.95	0.00	0.00	29.30	0.05	48.39	0.07	0.89	3.31	1.20	**86.15**	**0.75**	31.01
Ehlers	0.88	48.83	0.00	0.01	30.11	0.08	40.30	0.12	0.87	3.28	1.06	90.10	0.95	24.71
UNB	0.85	53.69	0.00	0.01	32.50	0.09	47.08	0.12	0.86	3.78	1.09	93.75	0.98	19.40
PC	0.85	54.43	0.00	0.00	34.44	0.05	47.14	0.12	0.85	3.79	1.58	92.04	0.92	20.30
GS	0.85	55.19	0.00	0.01	34.04	0.08	46.91	0.14	0.85	3.87	1.99	92.52	0.96	21.61
LMM	0.84	56.44	0.00	0.01	35.56	0.09	**31.78**	0.07	0.84	3.75	1.69	92.24	0.96	22.27
MIHS	0.83	57.24	0.00	0.01	35.53	0.09	46.66	0.11	0.84	3.83	1.49	92.31	0.92	22.74
Brovey	0.82	58.66	0.00	0.01	36.78	0.10	35.85	0.80	0.84	4.01	1.61	93.08	0.93	33.51
CN	0.79	64.16	0.00	0.01	41.87	**0.11**	35.11	**0.98**	0.81	4.23	**3.14**	95.08	0.98	**34.72**
SRM	**0.73**	**74.43**	0.00	-0.01	**49.26**	0.07	44.49	0.08	**0.74**	**5.04**	2.38	93.80	0.97	26.48

Fusion methods :CN – color normalization spectral sharpening, GS – Gram-Schmidt, HPF – High-pass filter, LMM – local mean matching, LMVM – Local mean variance matching,, MIHS –Modified intensity-hue,saturation, PC – Principal component analysis, SRM – Subtractive resolution merge, UNB – University of New Brunswick, WV – Wavelet PCA

Quality metrics: CC – Correlation coefficient, RMSE – Root-mean-square-error, RDM – Relative distance to mean, RDS – Relative distance to standard deviation, SD – Spectral discrepancy, DI – Deviation index, PSNR – Peak signal-to-noise ratio, MSSIM- Mean structure similarity index, ERGAS – relative dimensionless global error, SAM – Spectral angle mapper, HP-CC – High-pass correlation coefficient

Rank 1 Rank 2 Rank 3

**Worst value are indicated in bold

The Town of Nowshera, Pakistan

The mean scores of the spectral and spatial indices reported for the two subsets of the Town of Nowshera (NW-1, NW-2) in Pakistan are summarized in Table 10. Disregarding RDM and RDS, the spectral quality of fusion algorithms can be further discriminated based on the remaining spectral indices. The HPF algorithm seemed to be the best candidate, which exhibits rank 1 for all the indices except for entropy difference and SAM. The LMVM algorithm serves as the second-best fusion algorithm whereas Ehlers seemed to be the next best contender because most of its scores are superior to those of the WV algorithm. The SRM algorithm resulted in the worst scores for six spectral indices. Unlike previous cases, the CN fusion exhibited slightly better results, whereas Brovey transform reported the highest and the lowest values for entropy difference and SAM, respectively.

Regarding the spatial improvement of fusion results, the lowest scores for CEC and HP-CC were reported by the Wavelet algorithm proving its consistent failure. Similar to previous study sites, the UNB algorithm exhibited promising values for all the spatial indices. Interestingly, Gram-Schmidt and MIHS also showed better spatial improvement of their fusion results with respect to all spatial indices. Despite CN sharpening, Brovey, and SRM algorithms report promising values for CEC and HP-CC, all three algorithms exhibited poor RMSE values for Sobel filtered images (>42).

Table 10. Reported scores of spectral and spatial quality metrics for the Town of Nowshera (NW-1 and NW-2), Pakistan

Fusion algorithm	Spectral metric											Spatial metric		
	CC	RMSE	RDM	RDS	SD	DI	PSNR	Entropy	MSSIM	ERGAS	SAM	Canny edge	HP-CC	Sobel-RMSE
HPF	0.91	54.34	0.00	0.00	36.33	0.08	61.96	0.10	0.89	2.91	1.84	90.40	0.95	25.00
LMVM	0.89	59.87	0.00	0.00	37.39	0.08	30.90	0.07	0.86	3.15	2.02	**89.51**	0.84	33.44
WV	0.86	69.32	0.00	0.00	46.39	0.10	59.94	0.10	0.82	3.76	2.20	86.11	**0.82**	32.09
Ehlers	0.85	63.54	0.00	0.00	41.12	0.08	60.36	0.22	0.83	3.17	1.01	91.14	0.96	22.77
PC	0.79	83.49	0.00	0.00	58.19	0.12	58.31	0.36	0.77	4.52	3.00	91.85	0.91	23.70
UNB	0.78	85.56	0.00	0.00	55.25	0.12	58.00	0.16	0.75	4.56	2.42	94.71	0.99	18.15
GS	0.78	85.41	0.00	0.00	56.45	0.12	58.00	0.22	0.75	4.56	2.80	94.26	0.98	18.30
MHIS	0.77	79.74	0.00	0.00	51.62	0.10	58.36	0.12	0.76	4.01	1.54	93.01	0.94	17.64
CN	0.74	89.21	0.00	0.00	59.62	0.12	57.44	1.46	0.72	4.58	2.97	95.79	0.99	42.36
LMM	0.74	87.84	0.00	0.00	57.50	0.12	**27.45**	0.08	0.72	4.48	2.28	91.84	0.97	23.51
Brovey	0.71	96.43	0.00	0.00	66.82	0.15	57.16	1.**49**	0.71	5.24	**4.73**	95.18	0.97	42.68
SRM	**0.63**	**111.34**	0.00	0.00	**77.58**	**0.17**	55.81	0.12	**0.62**	**6.01**	4.72	95.06	0.97	**52.43**

Fusion methods :CN – color normalization spectral sharpening, GS – Gram-Schmidt, HPF – High-pass filter, LMM – local mean matching, LMVM – Local mean variance matching,, MIHS –Modified intensity-hue,saturation, PC – Principal component analysis, SRM – Subtractive resolution merge, UNB – University of New Brunswick, WV – Wavelet PCA

Quality metrics: CC – Correlation coefficient, RMSE – Root-mean-square-error, RDM – Relative distance to mean, RDS – Relative distance to standard deviation, SD – Spectral discrepancy, DI – Deviation index, PSNR – Peak signal-to-noise ratio, MSSIM- Mean structure similarity index, ERGAS – relative dimensionless global error, SAM – Spectral angle mapper, HP-CC – High-pass correlation coefficient

Rank 1 Rank 2 Rank 3

**Worst value are indicated in bold

Port-Au-Prince, Haiti

Table 11 summarizes the mean scores of the objective quality indices reported for the two subsets of Port-Au-Prince (PP-1, PP-2) in Haiti. In terms of the preservation of spectral quality relative to the original MS image, the Ehlers fusion algorithm seemed to be the best candidate whereas HPF and LMVM algorithms hold the second and third ranks, respectively. The SRM algorithm produced the worst values for the majority of spectral metrics.

With respect to spatial improvement of fused products, unexpectedly, the Ehlers algorithm exhibited poor results, which recorded the lowest value and the second-lowest value for HP-CC and CEC, respectively. The wavelet algorithm proved its continuous failure again by yielding the worst scores for CEC and HP-CC. Both UNB and Gram-Schmidt algorithms exhibited promising values for all three spatial indices. As seen in previous study sites, CN and SRM algorithms showed high values for CEC and HP-CC, however, failed to report satisfactory values for Sobel-RMSE.

Table 11. Reported scores of spectral and spatial quality metrics for Port-Au-Prince (PP-1 and PP-2), Haiti

Fusion algorithm	Spectral metric											Spatial metric		
	CC	RMSE	RDM	RDS	SD	DI	PSNR	Entropy	MSSIM	ERGAS	SAM	Canny edge	HP-CC	Sobel-RMSE
Ehlers	0.92	32.00	0.00	0.00	21.89	0.11	66.47	0.92	0.90	3.71	1.11	88.10	**0.80**	24.59
HPF	0.91	36.71	0.00	0.00	25.94	0.14	65.12	0.07	0.90	4.46	2.45	89.62	0.96	12.26
LMVM	0.89	38.99	0.00	0.00	26.01	0.14	34.48	0.06	0.87	4.73	2.96	88.41	0.84	20.62
WV	0.86	45.07	0.00	0.00	31.70	0.17	63.34	0.07	0.85	5.48	2.72	**83.53**	**0.80**	19.19
LMM	0.78	55.54	0.00	0.00	38.73	0.19	**31.39**	0.09	0.77	6.51	3.10	90.70	0.96	17.17
UNB	0.76	57.68	0.00	0.00	40.12	0.22	61.25	0.17	0.75	7.05	3.36	94.51	0.99	6.82
PC	0.76	58.12	0.00	0.00	41.52	0.22	61.15	0.61	0.76	7.10	4.49	90.78	0.91	17.98
MIHS	0.75	57.04	0.00	0.00	39.40	0.19	61.37	0.07	0.76	6.61	2.07	92.60	0.94	11.89
GS	0.75	58.75	0.00	0.00	40.66	0.21	61.06	0.18	0.74	7.16	3.87	94.63	0.99	7.04
Brovey	0.75	57.72	0.00	0.00	40.07	0.20	61.23	1.61	0.75	6.73	2.23	93.89	0.95	27.17
CN	0.73	61.16	0.00	0.00	42.68	0.21	60.70	**1.63**	0.73	7.22	4.05	95.37	0.99	26.74
SRM	**0.62**	**74.30**	0.00	0.00	**53.57**	**0.31**	59.01	0.16	**0.63**	**9.11**	**5.86**	94.39	0.98	**31.89**

Fusion methods :CN – color normalization spectral sharpening, GS – Gram-Schmidt, HPF – High-pass filter, LMM – local mean matching, LMVM – Local mean variance matching,, MIHS –Modified intensity-hue,saturation, PC – Principal component analysis, SRM – Subtractive resolution merge, UNB – University of New Brunswick, WV – Wavelet PCA

Quality metrics: CC – Correlation coefficient, RMSE – Root-mean-square-error, RDM – Relative distance to mean, RDS – Relative distance to standard deviation, SD – Spectral discrepancy, DI – Deviation index, PSNR – Peak signal-to-noise ratio, MSSIM- Mean structure similarity index, ERGAS – relative dimensionless global error, SAM – Spectral angle mapper, HP-CC – High-pass correlation coefficient

Rank 1 Rank 2 Rank 3

**Worst value are indicated in bold

Due to the involvement of a high number of spectral metrics over six study sites, we summarized the performances of the best four candidate algorithms (i.e., Ehlers, HPF, LMVM, and WV) in Figure 10. This graph offers a global picture on the performances of each algorithm across nine quality measures (CC, RMSE, SD, DI, PSNR, Entropy difference, MSSIM, ERGAS, and SAM) and over six study sites. The shading in Fig. 8 corresponds to that of Tables 6 through 11. Considering the total number of first rankings, Ehlers fusion outperformed others with respect to CC, RMSE, DI, ERGAS, and SAM. For SD and MSSIM, Ehlers and HPF algorithms were ranked first equally many times. Similarly, Ehlers, HPF, and WV were ranked first equally for PSNR. It is noteworthy that the wavelet algorithm produced poor results for ERGAS and SAM (global indicators) compared to the other three fusion techniques.

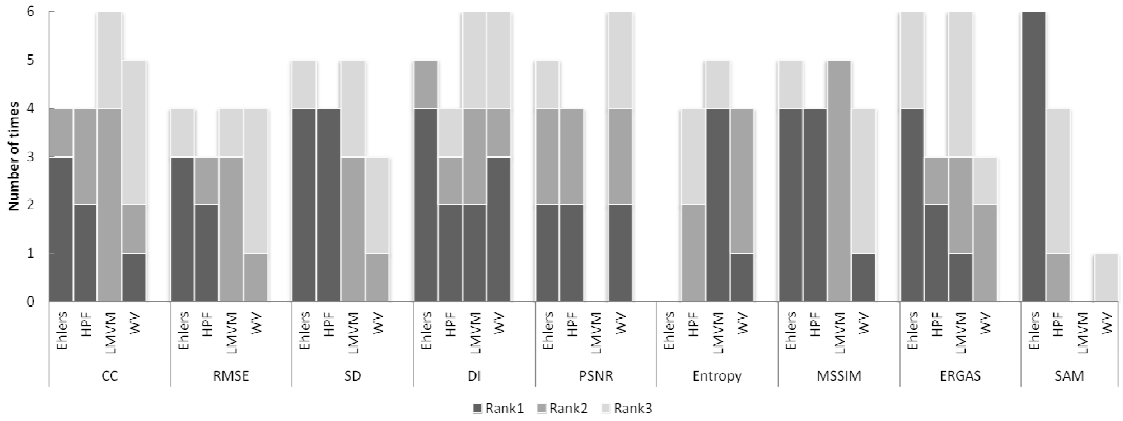


Figure 10. A summary of how each of the best-candidate algorithm perform with respect to a given spectral quality index in all six study sites. The *y*-axis depicts the total number of times a given algorithm holds rank 1, 2, or 3 for a given spectral index. The -axis depicts the spectral indices (except RDM and RDS) and the best-candidate fusion algorithms.

2.4 Discussion

This study aimed to enrich the knowledge on the efficacy of data fusion algorithms in the context of EO-based rapid mapping workflows. We applied twelve data fusion to VHSR image scenes encompassing on-going- and post-crisis scenarios and tested their performances using a series of quality metrics. We focused on key target features that are most likely to be extracted in rapid-mapping workflows such as IDP camps, transitional shelters, and damaged buildings. We also examined the stability of fusion quality indicator and how benchmarking of fusion algorithms can change with respect to the combination quality metrics that are used in the evaluation process.

From the point of visual inspections, no single algorithm was able to produce superior results by simultaneously preserving spectral and spatial properties of the original MS and PAN images. Algorithms like HPF, Gram-Schmidt, and UNB exhibited mediocre fusion results with respect to color similarity and spatial improvement. Although the validity of visual inspections has been challenged by many investigators because of subjectivity and interpreter dependency, visual inspections are necessary but alone are not sufficient; our contention is that they should always be corroborated with objective quality indices.

Taking objective spectral quality evaluation (Tables 6-11) into account it is important to examine how spectrally-superior algorithms (Ehlers fusion, HP filter, LMVM, and wavelet-PCA algorithms) behave across eleven quality metrics. With respect to two global indicators (ERGAS and SAM) and other band-wise metrics in our quality budget, the Ehlers fusion produced better-quality results (Figure 10). Ehlers and his colleagues (Klonus and Ehlers 2007; Ehlers et al. 2010; Ling et al. 2007) repeatedly

claimed the spectral quality preservation character of the Ehlers fusion algorithm not only in single-sensor fusion but also in multi-senor, and mutli-date pansharpening applications. Our multi-dimensional observations confirm the Ehlers fusion's ability to preserve spectral information of the parent MS image during pansharpening. However, it is interesting to examine the poor performances of the Ehlers algorithm in the town of Nowshera where it reported a low score for the correlation coefficient. Similarly, the high-pass filter method exhibited relatively weak performances for the two test sites of the civilian safety zone. We suspect that these anomalies are associated with the heterogeneous radiometric characteristic of bare soil (sandy coast) and floodwater covering large portions of the civilian safety zone and the town of Nowshera, respectively. Overall, scores reported for our spectral budget clearly demonstrated the superiority of spatial-domain methods (i.e., Ehlers fusion and HPF algorithm) compared to popular spectral substitution fusion techniques such as Brovey transform, CN sharpening, MIHS, and PC.

Regarding spatial quality assessment (Tables 6-11), despite the superior performances with respect to spectral similarity, Ehlers algorithm, wavelet-PCA algorithm, LMVM algorithm, and high-pass filter algorithm exhibited poor spatial improvement. Among this subset, the wavelet-PCA algorithm showed the worst scores consistently for all spatial indices where as the HPF algorithm outperformed the Ehlers algorithm and showed consistently better-quality values for spatial indices. The CN sharpening algorithm , Brovey transform algorithm, and subtractive resolution merge algorithm achieved notably better scores for spatial indices than for spectral indices. These observations emphasize the necessity of a combined approach (i.e., spectral and

spatial fidelity) for benchmarking fusion results because the best color preservation of an algorithm can be observed even if no pansharpening is performed; on the other hand, a fusion algorithm can achieve the best spatial improvement while producing results with worst color preservation. The UNB algorithm showed superior results consistently for all spatial metrics across all study sites while Gram-Schmidt and modified-IHS produced mediocre spatial improvement.

Like visual evaluation, objective assessment is much-debated because there is neither a universal index nor set of indices for quantifying spectral and spatial fidelity fusion results. There have been many metrics proposed and different combinations of metrics have been used to benchmark fusion algorithms. We intentionally included eleven spectral and three spatial indices to the objective-quality budget for examining their strength, redundancy, and effect on ranking of fusion techniques. Among spectral indices, RDM and RDS demonstrated no discriminating power among the fusion algorithms. However, some workers (e.g., Vijayaraj et al., 2006; Kim et al., 2011) used these metrics as strong detectors for capturing spectral distortions. Considerably high values can be reported for RDM and RDS when the original MS image is compared with the resampled fused image instead of resampled and histogram-matched fused image because some algorithms match histograms of the original MS and the fused image as the final step of the fusion workflow (e.g., Ehlers fusion) but this step is not obvious in most fusion techniques. We recommend histogram matching after resampling of the fused image as a standard practice in fusion evaluation because it brings all fusion results to a common datum. This is also valid for visual inspections. Considering reported scores, in general, CC, RMSE, SD, DI, and MSSIM were consistently behaved except few

disagreements exhibited by SD and DI (see Tables 6, 8, and 9). The peak signal-to-noise ratio seemed to be a valuable discriminator because of its broad range of values along with its unique capability on capturing radiometric distortions. For example, the LMVM algorithm seemed to be one of the best candidates in terms of metrics like CC, RMSE, SD, DI, and MSSIM but the PSNR was able to detect LMVM algorithm's weak radiometric performance. Similarly, SAM disqualified wavelet-PCA and LMVM algorithms despite their good scores reported for common metrics like such as CC, RMSE, and DI. As a global indicator ERGAS is supposed to give an overall picture on the quality of pansharpening. Wavelet-PCA fusion and LMVM algorithm exhibited relatively good ranks for ERGAS; however, as previously mentioned, these two algorithms do not report any rank for SAM (see Tables 6-11). Our understanding is that, even though fundamental statistical measurements like CC, RMSE, might be valid detectors of fusion artifacts, it is necessary to combine these spectral distortion parameters with radiometric distortion detectors like PSNR and SAM to achieve a comprehensive measure of fusion quality. Entropy (or entropy difference) produced inconsistent results compared to other quality indicators so we question its general dependability. For example, the LMVM fusion algorithm outperforms Ehlers fusion if the entropy difference is considered. Therefore, we suggest that entropy should be accompanied with several other quality measures.

Owing to the plethora of literature on fusion evaluation, one could draw a rational argument that there might be possibilities of transferring the knowledge on the benchmarking of fusion algorithm, which have already been tested in the general context or specific applications can be transferred to the crisis situations. As previously

mentioned, our contention is that data fusion algorithms are scene-dependent and the choice of the algorithm is a function of scene content, target features, and user needs. It is true that investigators, such as Karathanassi et al. (2007) and Ehlers et al. (2010), targetted urban landscapes in their evaluation workflows; however, in the aftermath of a crisis, features like floodwater, rubble, and debris can introduce extra complexity to the image scene. In addition, ephemeral settlements, such as IDP shelters and transitional shelters, are comparatively very small, highly crowded, and erected in landscapes where typical human dwellings are not expected. For example, a 25 km^2 narrow coastal stretch harboured more 100,000 in the civilian safety zone, Sri Lanka. In this setting, thousands of small randomly-oriented shelters were erected on the beach, which has reflectance properties similar to bright shelters. There is always a compromise between spatial and spectral properties of the objects of interest. If users (e.g., humanitarian agencies) suspect the existence of an IDP camp, there might be a high value for spectral properties because indices like NDVI could easily be utilized for discriminating man-made features. In contrast, with respect to dwelling counts for on-demand census of IDP population, it is necessary to emphasize the high frequency information content of the image to identify individual shelters. It is also important to understand how the spatial resolution of the image relates to the smallest object of interest (e.g., 2m x 2m tent). General understanding is that the spatial resolution of the image data should be at least one half the size of the smallest object of interest (Myint et al. 2011). This implies that that, for example, 1m resolution image data can be used to identify 2m x 2m shelters. This realization is only valid if the object of interest perfectly follows pixel edges, which is unlikely to happen in most instances. Therefore, we need to inject high-frequency

information to multispectral image to obtain significantly small pixels (e.g., 0.5 m) and successfully delineate target features.

Unlike the per-pixel based paradigm, the effectiveness of data fusion might be challenged in the object-based image analysis (OBIA) framework. This realization is catalyzed due to the OBIA framework's inherent nature of aggregating pixels into homogeneous segments by considering both spectral and spatial properties of pixels rather than thinking of individual per-pixel spectra (Radoux et al., 2011). Thus, whether the general view of data fusion community that is any error in the synthesis of the spectral signatures at the highest spatial resolution incurs an error in the decision, would hold true for the object-based paradigm is debatable. The OBIA framework has the ability to create meaningful image objects (Blaschke, 2010) during the segmentation process by compensating the fused image's spectral distortions with the high-frequency information content that has been injected during fusion. With respect to rapid-mapping workflows, one could argue that what is the advantage of using processor-intense and time consuming fusion techniques (e.g., spatial-domain fusion methods) over simple time-efficient fusion algorithms (e.g., arithmetic-domain fusion methods), if both methods yield meaningful image object candidates? There have been a few recent efforts (e.g., Tiede et al., 2011) to prove the possibilities of bypassing pre-processing steps (e.g., image fusion and geometric correction) in the context of rapid-mapping workflows. Undoubtedly, such approaches save processing time and help to intensify EO-based rapid mapping workflows. However, it should be noted that data fusion is not limited to VHSR image data but expands to many different combinations of multi-date, multi-sensor EO-data. There might be certain fusion scenarios where data fusion is really necessary and

highly-quality fused products are expected. Our understanding is that it is too early to either reject or justify the argument that data fusion really needed in rapid-mapping workflows. Like in fusion, where we rely on high spectral and spatial fidelity, in OBIA we expect that meaningful image objects lead to better classification. This research problem is being addressed in our ongoing work, in which we investigate the synergies of fusion and segmentation in light of object-based paradigm.

2.5 Conclusion

Data fusion algorithms are scene-dependent and the choice of the algorithm is a function of scene contents, target features, and user needs. This study aimed to enrich the knowledge on the efficacy of data fusion algorithms in the context of EO-based rapid mapping workflows. We applied twelve fusion algorithms to six GeoEye-1 images taken over three countries representing on-going and post-crisis scenarios. Benchmarking of fusion algorithms was conducted visually and quantitatively, the latter based on eleven spectral and three spatial indices. With respect to subjective and objective assessments, there is no fusion method that exhibited superior performances simultaneously for color preservation and spatial improvement. We recommend the University of New Brunswick algorithm if manual photointerpretation is involved whereas the high-pass filter fusion is recommended if semi- or fully-automated feature extraction is involved for pansharpening on-going and post crisis GeoEye-1 images. We emphasize the need of a standard workflow and a set of objective metrics for fusion evaluation. We propose deviation index, mean structure similarity index, peak-signal-to-noise-ratio, ERGAS, and spectral angle mapper as the main spectral quality discriminators. In our future research we expect to benchmark fusion algorithms when applied to multi-sensor and multi-date images covering on-going and post-crisis scenes, further investigate the stability of quality indicators, and develop a single metric that measures the combined spectral and spatial fidelity of fusion results. In the object-based paradigm, the efficacy data fusion can be challenged. This paper provides the basis for our ongoing work, in which we aim to investigate whether benchmarking of data fusion algorithms, that we observed with respect to fusion quality metrics, are manifested their image object candidates.

CHAPTER 3

Investigating synergies of data fusion and image segmentation in earth observation based rapid mapping workflows

Abstract

In humanitarian emergencies, the timeliness of data provision and the short time-window available for dispatching value-added information pose major challenges to the mapping community. We have been engaged with a continuous research effort to explore novel ways to catalyze the EO-based humanitarian crisis information retrieval chain. This paper is an exploratory study, which aimed to discover the synergies of data fusion and image segmentation in the context of EO-based rapid mapping workflows. Our approach pillared on the geographic object-based image analysis (GEOBIA) focusing on multiscale, internally-displaced persons' (IDP) camp information extraction from very high spatial resolution (VHSR) images. We applied twelve pansharpening algorithms to two subsets of a GeoEye-1 image scene that was taken over a former war-induced ephemeral settlement in Sri Lanka. A multidimensional assessment was employed to benchmark pansharpening algorithms with respect to their spectral and spatial fidelity. The multiresolution segmentation (MRS) algorithm of the eCognition Developer software served as the key algorithm in the segmentation process. The first study site was used for comparing segmentation results produced from the twelve fused products at a series of scale, shape, and compactness settings of the MRS algorithm. The segmentation quality and optimum parameter settings of the MRS algorithm were estimated by using empirical discrepancy measures. Non-parametric statistical tests were used to compare the quality of image object candidates, which were derived from the twelve pansharpened products. A wall-to-wall classification was performed based on a support vector machine (SVM) classifier to classify image objects candidates of the fused images. The second site simulated a more realistic crisis

information extraction scenario where the domain expertise is crucial in segmentation and classification. We compared segmentation and classification results of the original images (non-fused) and twelve fused images to understand the efficacy of data fusion. In light of GEOBIA framework, findings from our exploratory study have challenged the well-established view of data fusion community that any error in the synthesis of the spectral signatures at the highest spatial resolution incurs an error in the decision. We have shown that the GEOBIA has the ability to create meaningful image objects during the segmentation process by compensating the fused image's spectral distortions with the high-frequency information content that has been injected during fusion. Our findings further questioned the necessity of the data fusion step in rapid mapping context. Bypassing time-intensive data fusion helps to actuate EO-based rapid mapping workflows. We, however, emphasize the fact that data fusion is not limited to VHSR image data but expands over many different combinations of multi-date, multi-sensor EO-data. Thus, further research is needed to understand the synergies of data fusion and image segmentation with respect to multi-date, multi-sensor fusion scenarios and extrapolate our findings to other remote sensing application domains beyond EO-based crisis information retrieval.

3.1 Introduction

Natural hazards, man-made disasters, civil wars, and regional conflicts occur and recur inflicting severe damage to human environment. Although these crises are inescapable, practicing effective crisis management strategies could relieve the impact on human lives. Remote sensing (RS) is an indispensible resource in crisis management (Cheema, 2007, Joyce et al. 2009; Dell`Acqua and Polli, 2011; Kaya et al., 2011; Witharana 2012). Modern satellite sensors, such as IKONOS, QuickBird, GeoEye, WorldView, and Pléiades provide very high spatial resolution (VHSR) multi-spectral

imagery (at sub-meter level) that can capture the fine details needed for crisis information (Lang et al. 2010; Vu et al., 2009; Witharana and Civco, 012).

In humanitarian emergencies the timeliness of data provision and the short time window available for dispatching value-added information pose major challenges to the mapping community. Unlike other remote sensing application domains, such as land use/cover mapping, environmental monitoring, and natural resource management, in crisis scenarios, EO data need to be streamed through time-critical workflows for delivering reliable and effective information (Tiede et al., 2011; Witharana et al., 2013). Thus, there is always a compromise among response time, analysis depth, and thematic accuracy (Voigt et al. 2011). Typically, in the context of an EO-based rapid mapping workflow, the pre-preprocessing step serves as an integral segment that stands in between data acquisition and analysis steps (Witharana et al., 2013). Therefore, the main thrust of our work is to emphasize major steps (e.g., data fusion) involved in the pre-processing segment and explore the effect of those in the proceeding key steps (e.g., image segmentation, classification) of EO-based rapid mapping workflows.

The utility of VHSR earth observation (EO) data for rapid geospatial reporting, long-term recovery and inventory actions have been highly prominent in assessing the impacts from natural and anthropogenic disasters that occurred in the past decade. Examples include structural damage detection in the aftermath of major earthquakes; the 2003 Bam earthquake in Iran, the 2003 Boumerdes earthquake in Algeria, the 2008 Wenchuan earthquake in China, the 2011 Gujarat earthquake in India, the 2010 Haiti earthquake (Chiroiu 2005; Gusella et al. 2005; Vu et al. 2005; Kouchi et al. 2005; Yamazaki et al. 2007; Ehrlich et al. 2009; Saito et al. 2004; Corbane et al. 2011; Hussain

et al. 2011; Voigt et al. 2011; and Tiede et al. 2011), and the 2004 Indian Ocean Tsunami (Pesaresi et al. 2007; Vu et al. 2007; and Gmaba et al. 2007). In the case of anthropogenic crises, especially armed-conflicts and forced migration, the strength of EO data has been well-documented, few examples include, post-conflict damage assessments in the Former Yugoslav Republic of Macedonia (Al-Khudhairy et al. 2005) and Sierra Leone (Pagot and Pesaresi 2008), and refugee and IDP camp mapping in Sudan, Tanzania, and Kenya (Giada et al. 2003; Lang et al., 2010; Tiede et al. 2010; Kemper et al. 2011; Kim et al 2011; Hangenlocher et al. 2012).

VHSR satellite sensors typically record image data in a low resolution multispectral (MS) mode and high resolution panchromatic (PAN) mode (Nikolakopoulos 2008; Kim et al. 2011). Pansharpening is a pixel-level fusion technique used to increase the spatial resolution of the multispectral image while preserving the spectral information (Vijayaraj et al., 2006; Gangkofner et al., 2008; Makarau et al. 2012). Many data fusion algorithms (commonly categorized as spectral-domain, arithmetic merging, and spatial-domain methods (Gangkofner et al. 2008; Kim et al. 2011; Yang et al. 2012) have been developed and tested for different application domains. Fusion algorithms introduce spectral and spatial distortions to the resultant data depending on the scene content and the algorithms themselves (Ashraf et al. 2012; Witharana et al. 2013). Thus, evaluation of fusion quality is necessary. Visual inspection is the most effective way of benchmarking different fusion algorithms (Nikolakopoulos, 2008); however, it is subjective and largely dependent on the experience of the interpreter (Klonus and Ehlers, 2007; Ehlers et al., 2010). There have been many objective quality indices (Zhou et al. 1998; Wald 2000; Wang and Bovik 2002; Wang et al. 2004;

Gangkofner 2008; Ehlers 2010; Guo et al. 2010; Yakhdani and Azizi, 2010; Civco and Witharana 2012) introduced to assess the spatial and spectral fidelity of fused products.

The geographic object-based image analysis (GEOBIA) framework is a novel conceptualization of image understating that attempts to mimic innate visual cognitive abilities of humans (Hay et al. 2005; Hay et al 2008; Blaschke 2008; Marcal et al. 2008; Blaschke 2010; Marpu et al. 2010). Image segmentation, a process of partitioning of a complex image-scene into non-overlapping homogeneous regions (segments) in scene space, is the primary step within the object-based information retrieval chain (Pal and Pal 1993; Pham 2001; Schiewe 2002; Blaschke and Strobl 2001; Costa et al. 2008; Dey et al. 2010; Liu et al. 2012; Tong et al. 2012). This step is decisive because the resulting segments (image object candidates (Blaschke 2010)) form the basis for the subsequent classification, which is based on spectral, structural, topological, and semantic features (Burnett and Blaschke 2003; Hay et al. 2003; Benz et al. 2004, Lang 2008; Marcal et al. 2008; Neubert et al. 2008; Marcal and Rodrigues 2009; Sturm and Weindner 2009; Smith and Morton 2010; Tong et al. 2012). When thinking beyond perfect and optimal image object candidates, either of two artifacts are expected in segmentation, i.e. over-segmentation and under-segmentation (Marpu et al. 2010). While the former is acceptable, the latter is highly undesirable and has to be avoided (Sturm and Weidner 2009; Marpu et al. 2009). There is no universal way of evaluating segmentation; however many possible indices have been proposed in the computer-vision (Hoover et al. 1996; Charles et al. 2006; Unnikrishnan et al. 2007) and remote sensing (Meinel and Neubert 2004; Triaz-Sanz 2005; Neubert et al. 2006; Lucieer 2007; Moller 2007; Neubert et al. 2008; Weidner 2008; Marcal and Rodrigues 2009; Marpu 2009; Marpu

2010; Sturm and Weidner 2009; Liu et al. 2012) literature for assessing segmentation quality.

Humanitarian crisis management remote sensing applications require high spatial and spectral resolution images. In fact, data fusion serves as a cohesive component and routine procedure in EO-based rapid mapping workflows (Witharana, 2012; Witharana and Civco 2012). Similar to other RS applications, the GEOBIA framework is well-established in the crisis management domain due to the evident shortfalls of the pixel-based spectral-data alone model when confronted with VHSR imagery. The general view of the data fusion community is that any error in the synthesis of the spectral signatures at the highest spatial resolution causes an error in the decision. This interpretation holds true for pixel-based approaches, however, it can be challenged in the object-based paradigm because of GEOBIA's inherent nature of aggregating pixels into nested and scaled representations (Burnett and Blaschke 2003), in which image object candidates serve as main building blocks of class labeling rather than single pixels. The GEOBIA framework has the ability to create meaningful image object candidates during the segmentation process by compensating the fused image's spectral distortions with the high-frequency information content that has been injected during fusion. Data fusion and image segmentation are independently addressed in many studies; however the dependency of those two processes has not yet been addressed. We envision increased value of investigating synergies of data fusion and image segmentation in the context of EO-based rapid mapping workflows. Depending on the design goals of fusion algorithms, the pansharpening (e.g., global-scale pansharpening (Kim et al. 2011) process consumes significant time in rapid mapping workflows. For example, spatial-domain methods (e.g.

Ehlers fusion algorithm) are more time- and processor-intensive than arithmetic domain method (e.g., Brovey transform fusion algorithm). The intriguing question is "do time-intensive fusion algorithms, which are typically designed to provide high spatial and spectral fidelity of fused products, make a significant difference in the segmentation process compared to other time-efficient fusion methods?". There is another growing tendency of introducing raw image layers (i.e., PAN image layer and MS image layers) individually to segmentation algorithms (e.g., eCognition Developer's multiresolution segmentation algorithm) and refine resulting image objects during class modeling (Lang et al., 2010). For example, Tiede et al (2011) attempted to bypass major pre-processing steps, including data fusion, and developed a methodology for automated extraction of damage information from very high spatial resolution (VHSR) satellite image data. In this respect, in light of the GEOBIA framework, the requisite of pansharpening in rapid mapping workflows needs to be explored.

The overarching goal of our continuous research work is to explore novel ways to catalyze EO-based humanitarian crisis information retrieval chain. This paper is centralized on a core objective that aims to investigate the synergies of data fusion and image segmentation in a rapid mapping context. Our study is inspired by the findings of Witharana et al. (2013), in which we employed a detailed multidimensional assessment to understand the performances of twelve application-oriented data fusion algorithms when applied to ongoing- and post-crisis VHSR image scenes comprising earthquake damaged areas of Haiti, flood impacted areas of Pakistan, and armed-conflicted areas and IDP camps of Sri Lanka. Because the current study serves as an exploratory research effort that aims to maintain a greater depth of analysis, we confine our study area to an IDP

camp in Sri Lanka, which was served as a focal ephemeral settlement and decommissioned in August 2012. War-induced ephemeral settlements are highly dynamic due to continuous influx of civilians as IDPs during ongoing crisis situation and outflow of IDPs as returnees during post-crisis resettlement and livelihood restoration operations. In these circumstances ground-based information production is expensive and limited. Multiscale analysis information extraction from EO-data is of high value for stakeholders involved in humanitarian crisis services because they need timely-information of ephemeral settlements at varying scales, such as camps' extent and structure; IDP shelters' size, type, count, density, and condition; and especially indirect estimation of IDP families and individuals. In this research, we selected two study sites from the Menik Farm IDP camp in Sri Lanka. We used two candidate sites focusing on different design goals. In one study site we employed eCognition Developer's multiresolution segmentation (MRS) algorithm in a series of parameter settings (i.e., scale, shape, and compactness) and tested how the quality of image object candidates varies with respect to data fusion algorithm. The segmentation quality and optimum parameter setting of the MRS algorithm were estimated by using empirical discrepancy measures. Non-parametric statistical tests were used to compare the quality of image object candidates that were derived from different pansharpened products. We classified image object candidates from each fused image produced at the optimum parameter setting of the MRS algorithm using the support vector machine (SVM) classifier. This classification was purposely designed as a wall-to-wall classification in order to maintain a consistency in classification workflow and understand better how data fusion affects class labeling in the GEOBIA framework. In the second site we utilized domain expertise for segmentation

and classification. We introduced fused images and original images (PAN and MS) to the MRS algorithm at user-selected parameter settings and a rule-based classification was used to extract features of interest (i.e., IDP shelters). In each study site 1000 IDP shelters were randomly selected and manually extracted (i.e. digitized) for detailed visual analysis. Two remote sensing experts were tasked to inspect visually the quality of raw image objects and classified image objects produced from original and pansharpened images.

The rest of the paper is structured as follows. Section 2 describes study area, image data, data fusion algorithms and evaluation methods, image segmentation and quality assessment workflow. Section 3 reports the spatial and spectral fidelity of fused products, quality of image object candidates and statistical significance, and classification and accuracy assessment. Section 4 contains a discussion explaining the results based on the performances of fusion algorithms and the quality of their image object candidates. Finally, conclusions are drawn in Section 5.3.2

Materials and methods

3.2.1 Study area, image data, and analysis workflow

Our study area is of a former war-induced ephemeral settlement, the Menik Farm IDP camp, in Sri Lanka (Figure 11). Based on EO-data, in June 2009, the UNOSAT program estimated approximately 22,760 IDP shelters in the Menik Farm IDP camp. According to the joint humanitarian and early recovery reports of the United Nations Coordination of Humanitarian Affairs (UNOCHA, 2010), in late February 2010, the Menik Farm IDP camp hosted approximately 93,000 individuals (~28,000 IDP families).

Owing to the Sri Lankan government's well-defined post-war resettlement, recovery, and livelihood restoration strategies, this IDP camp was decommissioned within a short time-window. The selection of study areas was made mainly focusing on the value of ongoing-crisis humanitarian information extraction. Multiscale analysis information extraction from EO-data (Figure 12) is of high value for humanitarian relief and donor agencies because they need timely-information of ephemeral settlements' at varying scales, such as camps' extent and structure; IDP shelters' size, type, count, density, and condition; and especially indirect estimation of IDP families and individuals. The image scene used in this study was acquired by GeoEye-1 sensor (Table 2) in February 2010. The GeoEye-1 sensor has a spatial resolution of 0.41cm for the PAN and 1.65m for MS bands at nadir with 11-bit radiometric resolution. The images are spatially registered to the Universal Transverse Mercator (UTM) coordinate system on the WGS 84 datum. We extracted two subset sites (MF1 and MF2, each approximately 1 km x 1 km) from the whole scene (Figure 13) and introduced them to image fusion and segmentation workflows (Figure 5) and Figure 14).

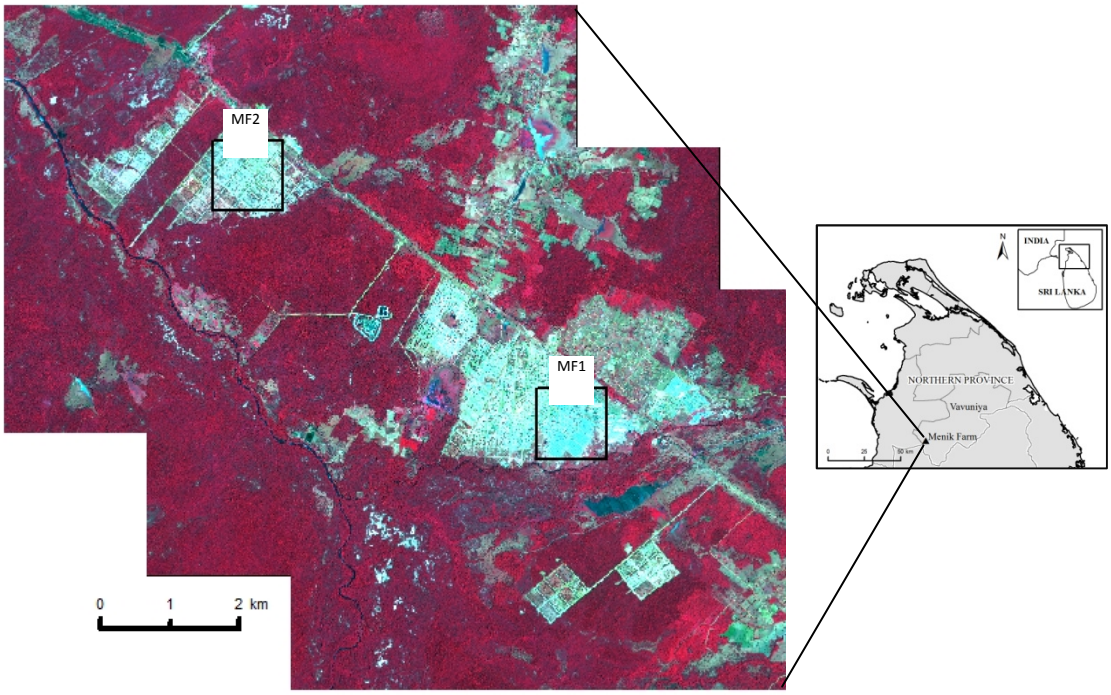


Figure 11. Index maps of the selected study area (right) and the GeoEye-1 image scenes taken over the Menik Farm IDP Camp, Sri Lanka (left). Black-hollowed boxes depict the candidate subsets.

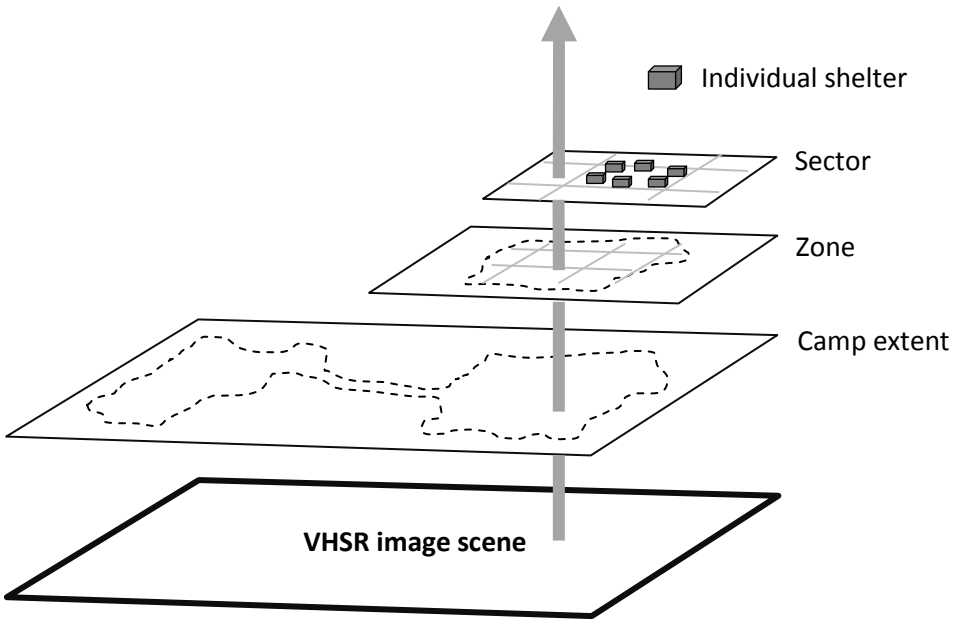


Figure 12. EO-based multiscale IDP camp information extraction

a b

Figure 13. Selected 1 km x 1 km subsets (MF1, MF2) shown as false-color composite

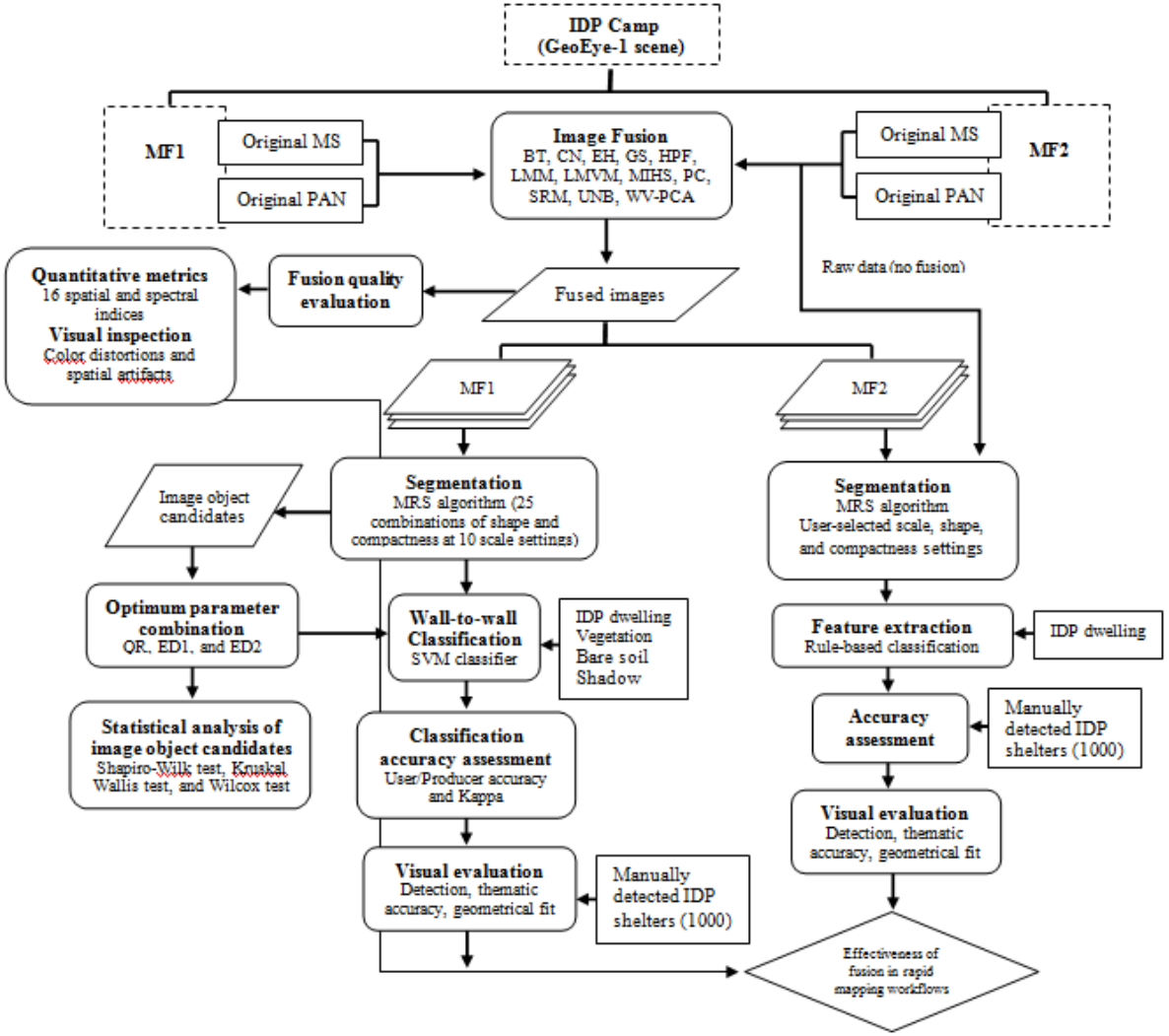


Figure 14. Flow chart showing key steps involved in the main analysis workflow

3.2.2 Methods

3.2.2.1 Image fusion

We tested twelve data fusion algorithms (Table 3), which are commonly encountered in the literature and readily accessible in image processing software packages. Three major commercial software packages: -ERDAS Imagine 2011, ENVI 4.8, and PCI Geomatica 2012-, in conjunction with MATLAB 2008 were used to implement candidate fusion algorithms, of which some are proprietary algorithms for specific image processing software packages (e.g., Ehlers fusion - ERDAS Imagine, the University of Brunswick fusion - PCI Geomatica). Unlike the Brovey transform and CN sharepening algorithm, which produce three-band fused images (B, G, and R or G, R, and NIR), other candidate algorithms are capable of accepting more than three bands at a time and producing four-band fused images in a single iteration. Therefore, we produced true- and false-color composites of BT and CN algorithms and layer stacked them to create four-band pansharpened images. Fusion results were introduced to a series of quality metrics (Table 4) along with detailed visual inspection procedures to evaluate the spectral and spatial fidelity of fused products compared to their original MS and PAN images. Objective metrics were calculated independently for each subset and separately for each band (except for ERGAS and SAM). Mean values were then calculated for all bands and for each subsets in the study area. We encourage readers to refer our previous work (Witharana, 2012; Witharana and Civco, 2012; Witharana et al., 2013) for more details on candidate fusion algorithms and quality metrics. The major steps involved in the fusion-evaluation workflow are depicted in Figure 5).

Visual assessment

A representative site was selected to inspect color preservation and spatial improvement of fused images compared to the original PAN and MS images, respectively. To assess the color similarity, fused images were resampled to the resolution of the original MS image (i.e. 0.50 m to 2.0 m) and their histograms were matched against that of the original MS image (Figure 5). The spatial quality inspection involved principal component (PC) analysis of fused products. The first PC of fused images (gray-scale images) were separated and their histograms were matched against the original PAN images (Figure 5). Finally, false-color composites and first-PC images along with their original images were inspected by two photo-interpretation experts to identify any spectral distortions, such as brightness reversions, saturation, a complete change of spectral characteristics, artificial artifacts, or unnatural/artificial colors.

Quantitative assessment

With respect to spectral fidelity, we used eleven metrics to quantify the post-fusion color similarity (Table 4). Of these indicators, except SAM and ERGAS, metrics were calculated band-wise and averaged over all bands. All fused images were resampled and histograms were matched before introducing them to spectral quality metrics. We utilized Canny edge filter (CEC), high-pass correlation coefficient (HP-CC), and RMSE of Sobel-filtered edge images (Sobel-RMSE) to quantify the quality of spatial improvement (Figure 5). A detailed discussion on the implementation of these metrics can be found in Witharana et al.(2013).

3.2.2.2 Image segmentation and quality evaluation

Many segmentation algorithms targeting RS applications have been developed and tested. In this study, we used the multiresolution segmentation algorithm (MRS) (Baatz and Schäpe 2000), which is a proprietary algorithm of the eCognition Developer (Munich, Germany) software package (Trimble GmbH, 2011). Since the inception of the GEOBIA concept, the MRS algorithm has gained a wide popularity in many RS application domains due its proven capabilities of producing the best segmentation results (Neubert et al. 2008; Blaschke 2010; Tong et al. 2012). The MRS algorithm is a bottom-up region merging technique. It aggregates individual pixels into increasingly larger size segments at multiple levels in an iterative process with respect to three parameters: 1) scale, 2) shape, and 3) compactness. At each step, merging of two adjacent segments is decided by spectral heterogeneity and shape heterogeneity whose weighted sum is used to calculate image-segment fusion values (Benz et al., 2004; Trimble GmbH, 2011). The amalgamation of two adjacent image segments occurs when the calculated fusion value (f) is less than the square of user-defined scale threshold (s) (i.e, $f < s^2$). The spectral heterogeneity determines the size of the segment. The shape heterogeneity is the weighted sum of two additional parameters; compactness and smoothness. Individual weights and weighted sums of the parameters are scaled between zero and one. A detailed review on the mathematical formulation of the MRS algorithm is beyond the scope of this paper, thus, we encourage reader to refer relevant literature (e.g., Baatz and Schäpe, 2000; Benz et al., 2004; Trimble GmbH, 2011; Tong et al., 2012) for more details.

After partitioning an image scene into non-overlapping homogenous areas, the resulting disjoint segments (image object candidates) serve as the basic building blocks for reconstructing reference objects (Figures 15). The quality of segmentation can be assessed with respect to three scenarios; optimal-, over-, and under- segmentation. The optimal segmentation serves as the best case scenario that leads to accurate classification results. The over-segmentation is acceptable but leads to over-classification while the under-segmentation is unfavorable and produces mixed classified objects. We categorize image segments into three different object types (Figure 15) as; 1) image object candidate, 2) corresponding image object, and 3) satisfactory image object. An image object candidates is realized as a corresponding image object when it intersects with a reference object (geo-object, Lang et al., 2010) that needs to be reconstructed. Being a corresponding image object does not necessarily guarantee that it will be used in the reference object reconstruction process. Therefore, a corresponding image object is further realized as a satisfactory image object when it meets the user-defined criteria (equation 3) for being elected for the reconstruction process. Figure 16 illustrates how the quality of segmentation (i.e. satisfactory image objects) affects on the quality of the classified image object during the reference object reconstruction process.

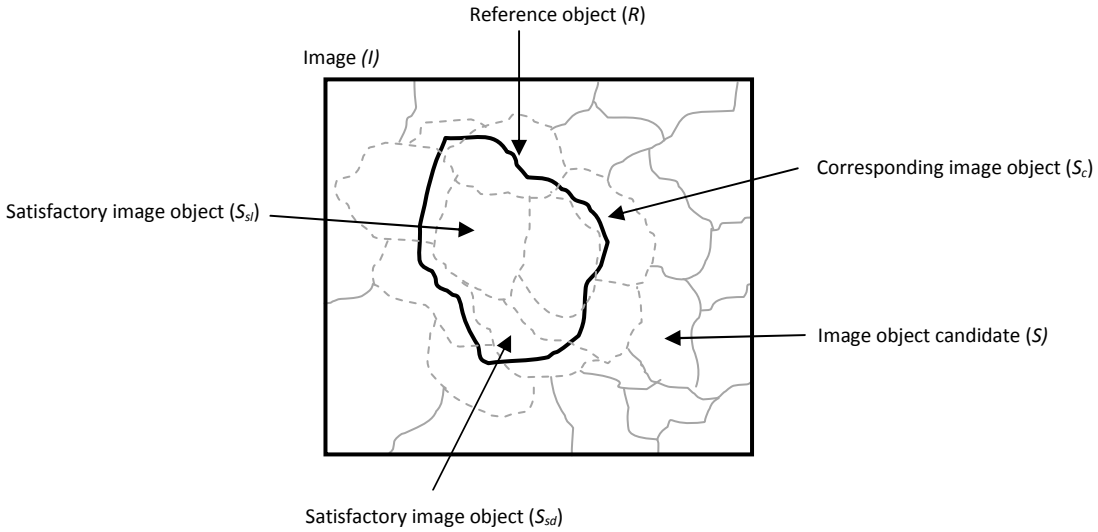


Figure 15. General classification of image segments

Let image scene be *I* and its segments as *S*.

$$I = \bigcup_{i}^{n} s_i \tag{1}$$

where, $S = \{s_i: i = 1, 2, ..., n\}$

for any s_i and s_j ;

$$s_i \cap s_j = \emptyset$$

Let *R* be reference objects of land cover class *C*. In order to reconstruct reference objects (*R*), corresponding image objects (S_c) should be identified from image object candidates (*S*).

$$R = \bigcup_{k}^{l} s_{ck} \tag{2}$$

Here s_{ck} image objects are assigned to the reference object r_p ;

$$s_{ck} \cap r_p = \emptyset$$

where, $R = \{r_p: p = 1, 2, ..., q\}$, $S_c = \{s_{ck}: k = 1,2,....,l\}$

We refine corresponding image objects (S_c) as satisfactory image objects (S_s) based on the areal overlap criteria (Clinton et al., 2010).

Here s_{su} image objects are assigned to the reference object r_p ;

$$\frac{|s_{su} \cap r_p|}{|s_{su}|} > 0.5 \quad or \quad \frac{|s_{su} \cap r_p|}{|r_p|} > 0.5 \tag{3}$$

where, $S_s = \{s_{su}: u = 1,2,....,v\}$

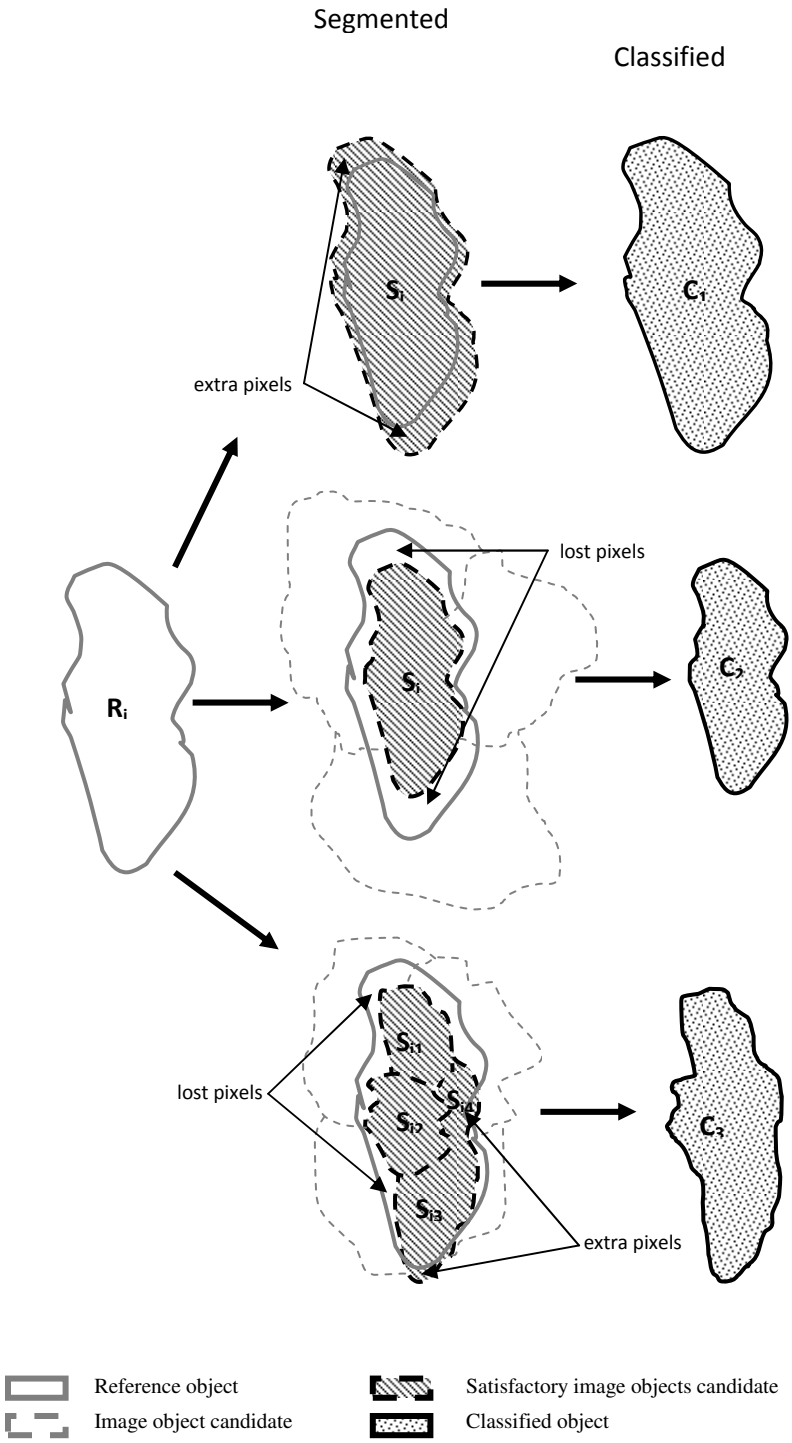


Figure 16. Reconstruction of a reference object (R) as a classified object (C) using image object candidates (S)

3.2.2.3 Segmentation quality metrics

Like segmentation itself, there is no standard method for assessing segmentation quality. This study utilizes empirical discrepancy measures (Zhang et al., 1996) for evaluating segmentation results. With respect to geometrical congruence between the reference objects (r_p) and satisfactory image objects, investigators (Sturm and Weidner 2009; Weidner 2010; Clinton 2010) defined basic quality indices: 1) quality rate (QR), 2) over-segmentation rate (OR) , 3) under-segment rate (UR), and 4) an optimization of UR and OR in the Euclidean space (ED1). By design, these metrics strictly focus on geometrical discrepancy, which can occur in three possible segmentation scenarios 1) overlap, 2) over-segmentation, and 4) under-segmentation. However, in reality, the highest geometrical fit does not necessarily manifest a good segmentation because at the highest geometrical congruency the size of satisfactory image objects can be the size of individual pixels. Figure 17 illustrates three possible arithmetic discrepancies expected in image object candidates: 1) one-to-one, 2) one-to-many, and 3) many-to-one. The first and second serve as the ideal and the most expected scenarios, respectively, while the last one is the most undesirable scenario. Liu et al. (2012) proposed three new metrics which encapsulate both geometrical (under-segmentation) and arithmetic (many-to-one correspondence) discrepancies occuring in image segmentation (equations 7, 8, and 9). We employed QR, ED1, and ED2 for assessing segmentation quality and optimum parameter setting of the MRS algorithm. For all metrics, highest segmentation quality is reported when the calculated metric value is close to zero.

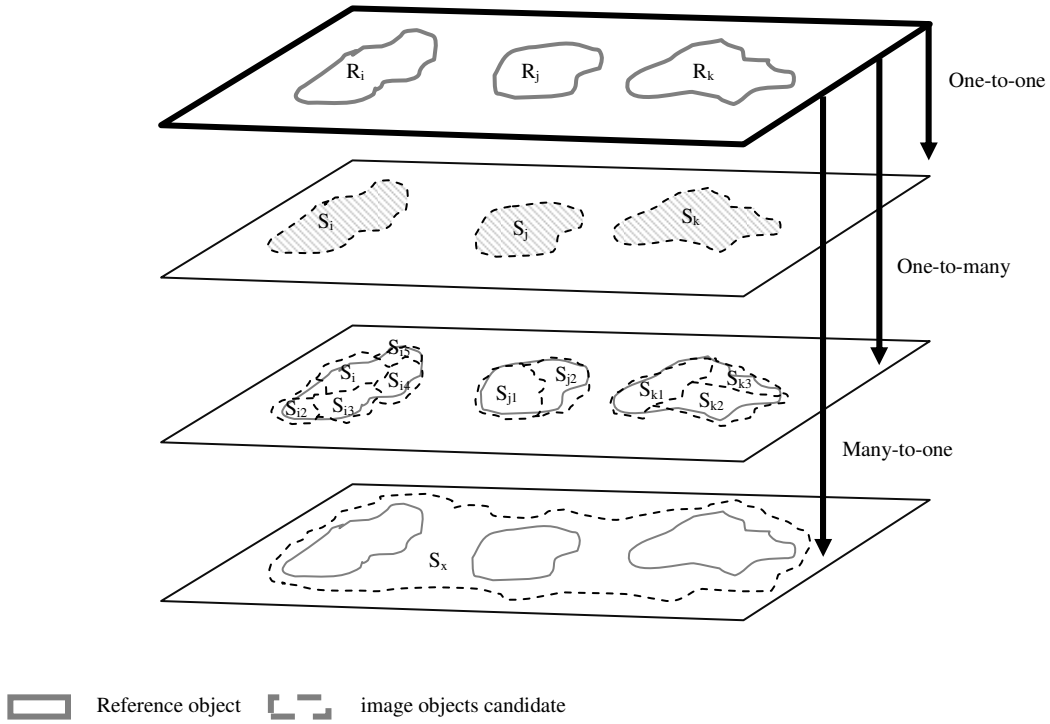


Figure 17. Possible arithmetic discrepancies among image object candidates (S) and reference objects (R); 1) one -to-one correspondence, 2) one-to-many correspondence, and 3) many-to-one correspondence.

$$QR = 1 - \frac{\sum|r_p \cap s_{su}|}{\sum|r_p \cup s_{su}|} \quad (4)$$

$$OR = 1 - \frac{\sum|r_p \cap s_{su}|}{\sum|r_p|} \quad (5)$$

$$UR = 1 - \frac{\sum|\mathrm{r_p} \cap s_{su}|}{\sum|s_s|} \quad (6)$$

$$ED1 = \sqrt{\frac{QR^2 + UR^2}{2}} \quad (7)$$

$$PSE = 1 - \frac{\sum|s_{su} - r_p|}{\sum|r_p|} \quad (8)$$

$$NSR = \frac{abs(q-v)}{q} \quad (9)$$

$$ED2 = \sqrt{PSE^2 + NSR^2} \quad (10)$$

$QR, OR, UR,$ ED1 $\in [0,1]$ and $PSE, NSR, ED2 \in [0, \infty)$

3.2.2.4 Statistical analysis

When comparing the values of objectives metrics (QR, ED1, and ED2) among different parameter setting s of a fused image itself and among different fused products, it is crucial to employ a suitable statistical technique to discriminate different groups. In general, area-based quality metrics are not normally-distributed (Marpu et al. 2010), thus, comparison of sample medians (non-parametric tests, such as Kruskal-Wallis) are more meaningful than sample means (parametric tests, such as ANOVA). Compared to parametric tests, non-parametric tests have generally low power but they are conservative, therefore, when the underlying distribution is uncertain it is more appropriate to use non-parametric tests (Siegel and Castellan 1998; Quinn and Keough 2002; Rogerson 2006). In order to verify the shape of the distribution we employed the Shapiro-Wilk test (Rogerson 2006), which is a widely-used normality test for sample size of $n < 1000$. Based on the test results of Shapiro-Wilk test, we employed two non-parametric statistical tests (Kruskal-Wallis test and Wilcoxon rank sum test) for comparing the quality image objects. The Kruskal-Walli test (Kruskal and Wallis, 1952) performs the comparison of population medians among multiple groups. When the obtained value of Kruskal-Walli test is significant (p-value > α level (e.g., 0.05)), it indicates at least one of the groups is different from at least one of the others (Sigel and Castellan, 1988). This test does indicate which groups are different or how many of the groups are different from each other. The Wilcoxon rank sum test, an alternative to two-sample t-test, entails a pair-wise comparison of groups (Gibbons and Chakraborti, 2011). This test can be performed as one-tailed (H_0: $\theta_u = \theta_v$ and H_A: $\theta_u > \theta_v$) or two-tailed (H_0: $\theta_u = \theta_v$ and H_A: $\theta_u \neq \theta_v$). The two-tailed test is more conservative than the one-tailed test

because the former takes more extreme test statistic to reject the null hypothesis. These statistical tests should be carried out under a certain user-nominated significance level (e.g., α = 0.05, 0.10, and 0.15), which specifies the probability level to accept an event that did not occur by chance. At lower significance levels, stronger evidence is needed for rejecting the null-hypothesis. In the case of multiple testing situations, special corrections, such as Bonferroni correction (Quinn and Keough 2002) and Benjamini procedure (Benjamini and Yekutieli, 2001), are needed for adjusting significance levels to control Type I error rates.

3.2.2.5 Analysis workflow

Figure 14 depicts the major processes involved in the main analysis workflow. Two candidate subsets are depicted as MF1 and MF2.As discussed earlier, original images (PAN and MS) of two study sites were fused using candidate fusion algorithms and evaluated against a series of fusion quality metrics. As seen in Figure 14, two sites are processed differently and aimed on distinct research questions. MF1 entails a detailed analysis on the segmentation quality, optimum parameter setting of the MRS algorithm, statistical significance of satisfactory image objects, and classification. Here, the main design goal is to understand how segmentation quality and classification (wall-to-wall) accuracy vary with respect to data fusion algorithms. On the other hand, the MF2 involves less complex processing steps and aims to understand how the quality of segmentation and classification (feature extraction) vary with respect to fused and original (non-fused) images.

In the case of MF1, we introduced twelve fused products to the MRS algorithm at varying parameter settings. As previously mentioned, the MRS algorithm's scale

parameter is unbounded while shape and compactness parameters are bounded between 0 and 1. Therefore, theoretically, it is possible to have an infinite number of segmentation cycles depending on the combination of values that feed into scale, shape, and compactness. The MRS algorithm is highly time- and processor-intensive, thus, we gradually increased the scale parameter at 10 unit steps from 10 to 100. Shape and compactness parameters were changed at 0.2 intervals from 0.1 to 0.9. For a given scale (e.g. scale = 10), the other two variables (shape and compactness) can yield 25 numerical combinations or 25 different segmentation cycles. In other words, a given fused product (e.g. Brovey transform fusion algorithm) entailed 250 segmentation cycles. The resulting image object candidates were exported as shapefiles and evaluated visually and using quantitative metrics (QR, ED1, and ED2). At this step, our objectives are twofold (see Figure 5): 1) investigate the statistical significance of the quality of image object candidates among fused products and 2) find out the optimum parameter setting (scale, shape, and compactness) of the MRS algorithm for each fused image and use that setting in the proceeding classification steps.

The objective comparison of segmentation results (i.e., quality of image objects candidates) of twelve fused products can occur at different parameters settings (e.g., scale-wise) of the MRS algorithm. For example, comparison of segmentation quality among fused images at scale 30 or at all scales (10 to 100). In order to minimize the plurality of the segmentation quality assessment, we selected optimum scale parameter values for each fused product based on the quality metrics (QR, ED1, and ED2). We then employed further analysis to obtain optimum settings for shape and compactness parameters at the optimum scale setting. By doing this, we obtained optimum parameter

values (scale, shape, and compactness) for each fused product with respect to each quality metric, hence the quality of image objects was compared among fused products. We used 30 randomly selected shelters for evaluating quality of image objects. The statistical tests involved in this step are already discussed under statistical analysis.

Optimum parameter setting for the MRS algorithm can be changed depending on the design goals of quality metrics. Therefore, based on segmentation quality analysis results, we used ED2-selected image objects for classification process. In order to maintain the consistency and minimum human involvement in classification workflow among twelve fused products, we used a classifier (support vector machine (SVM) classifier) available in eCognition developer software. We purposely employed a wall-to-wall classification, which is typically unlikely to be expected in EO-based rapid mapping workflows, to understand better exchange classes. We classified image objects into four candidate classes: 1) IDP dwelling, 2) vegetation, 3) bare soil, and 4) shadow. Standard accuracy assessments techniques (e.g., users, producers, and overall accuracy and kappa statistics) were employed to assess the classification accuracy. These accuracy assessment techniques are pixel-based, in fact, the validity of these quality indicators are frequently challenged in the GEOBIA framework. Therefore, we performed a detailed visual inspection (e.g., detection, thematic accuracy, and geometrical fit) of IDP dwelling class. We manually extracted 1,000 IDP shelters and used these as references to inspect the quality of classified image objects.

In the case of MF2, we introduced twelve fused products and original images (PAN and MS) to the MRS algorithm at user-selected (trial-and-error approach) parameter settings. Segmented images were then introduced to a rule-based classification

workflow. Unlike in MF1, here we performed a feature extraction focused on structures (binary classification) rather than a wall-to-wall classification approach. Similar to MF1, a detailed visual inspection of classified IDP dwellings was employed with respect to the hand-digitzed IDP shelters.

3.3 Results

3.3.1 Image fusion

3.3.1.2 Visual evaluation

Figure 7 (see Chapter 2) depicts the selected representative site for visual inspection. For color similarity, we choose tests areas with spectrally heterogeneous objects such as vegetation, roof-tops, and roads. For spatial quality, the same test area was zoomed into distinct objects with sharp edges (e.g. IDP shelters). We selected the best two and the worst two fusion algorithms in terms of spectral and spatial fidelity. Ranking of twelve fusion algorithms (i.e., best to worst) based on their performances is a difficult task even for an expert photointerpreter because, while the human brain easily capture extreme variations (i.e., best fusion and worst fusion) and exhibit reproducible results, it fails to distinguish slight variations existing between the worst and the best. Table 12 depicts the ranking of fusion algorithms based on expert evaluation.

Table 12. Objective evaluation of fused images by experts

Spectral similarity		Spatial similarity	
Best	Worst	Best	Worst
LMVM, Gram-Schmidt	CN sharpening, SRM	CN sharpening, UNB	Ehlers, Wavelet

Visually comparing the original multispectral and fused images of the test area (Figure 7) revealed that the CN sharpening algorithm has the worst color distortions. CN dramatically changed the colors of vegetation and IDP shelters. The SRM algorithm preserved basic spectral characteristic of the vegetation, IDP shelters, and bare soil;

however SRM produced an overly contrasted image. Based on notable spectral degradations, CN and SRM can be considered as the worst algorithms. Of the remaining ten algorithms, selecting the best-two algorithms is challenging because they preserved most of the color information of the original MS image. Focusing on IDP shelters in the original image, the shelters show two spectral appearances: the roof-tops to the right edge of the image appear dark gray whereas those to the left are light gray. Visually, all roof tops are smooth and identifiable on the original image. Among the ten contenders, LMVM, Gram-Schmidt, HPF, Ehlers, PC, and UNB showed nearly equal performances; however, of these, HPF gave an overly contrasted image compared to the original MS image. In Ehlers, roof-tops are not easily identifiable compared to the original and results from the other algorithms. The LMVM fusion algorithm had the best spectral agreement with the original multispectral image and Gram-Schmidt fusion algorithm can be considered as the runner-up algorithm. When assessing the spatial improvement of the same test area (Plate 3, Figure 7), the Wavelet fusion algorithm appears to give the worst spatial improvement. It produced a "fuzzy" image with almost no edge enhancements. The Ehlers algorithm performed better than Wavelet; however, when compared to other algorithms, Ehlers algorithm produced visually poor results. Therefore, Ehlers and Wavelet can be considered as the worst-two algorithms in terms of spatial improvement. In general, except Wavelet, Ehlers, LMVM, and PC, the other eight algorithms showed good spatial agreement with the original PAN image. Among this subset of contenders, CN and UNB fusion algorithms can be elected as the best two algorithms.

3.3.1.3 Quantitative assessment

We corroborated visual assessment with eleven spectral metrics and three spatial metrics. Table 9 (see Chapter 2) summarizes the mean scores (averaged over bands and subsets) reported by quality metrics for the two study sites. Fusion algorithms in the table are ranked by their correlation coefficient scores. We ranked fusion algorithms in each column as 1, 2, and, 3 and three gray levels as dark-, medium-, light-gray were used to highlight these three ranks, respectively. The worst value reported for a given quality metric is in bold font.

With respect to spectral indices, the HPF algorithm ranks as the best-candidate and LMVM and WV algorithms hold the second- and the third-best positions, respectively. The SRM algorithm produced the worst values for CC, RMSE, SD, MSSIM, and ERGAS. The worst values for DI, entropy difference, and SAM were reported by the CN fusion algorithm. In general, algorithms like Ehlers, UNB, PC, and GS have shown average performances. In terms of spatial indices, CN and UNB algorithms exhibited the best performances. The former reported the highest value for the CES while the latter reported the lowest value for Sobel RMSE. Both the former and the latter exhibited equal values for the HP- CC (0.98). Despite the fact that the CN fusion holds rank 1 for CEC and HP-CC, it has reported the worst value for Sobel-RMSE. Similar to other study areas, none of the algorithms producing results with high spectral agreement achieved notable scores for spatial quality estimators

3.3.1.4 Image segmentation and classification

Study site - MF1

In this study site we focused only on fused images and envisaged two objectives;1) statistically analyze the quality of image object candidates among different fused products and 2) assess the classification accuracy of classified image objects. MF1 mainly consisted of multi-family dwellings with rigid roofing materials.

We employed QR, ED1, and ED2 to assess the quality of image object candidates (also the optimum parameter settings of the MRS algorithm) with respect to manually extracted IDP shelters. Table 13 reports the descriptive statistics of the sampled reference objects. Figures 18, 19, and 20 depict the scatter plots of QR, ED1, and ED2, respectively for twelve fused products. For a given scale, each dot indicates a combination of shape and compactness parameters of the MRS algorithm. In the case of QR and ED1, metric values close to the zero indicate high geometric congruency between image objects and reference objects. With respect to ED2, values close to zero indicate high geometrical and arithmetic agreement between image objects and reference objects. Thus, at optimum segmentation we expect points with low values and the least scatter. On these scatter plots we used a black-hollow-box to encapsulate the point cloud at optimum segmentation. As seen on Figures 18 and 19, QR and ED1 exhibit a similar monotonous variation with increasing scale parameter. Based on these two metrics, all fused images report the optimum segmentation (i.e., best image objects) at scale 10. Compared to QR and ED1 metrics, ED2 exhibited different behavior with concave-shaped scatter plots across increasing scale parameter. With respect to ED2, fused products showed both similar and different optimum scale parameter values. For

example, Brovey transform (BT) fusion algorithm and color normalized pansharpening (CN) algorithm showed the optimum ED2 values at scale 40 whereas High-pass filter (HPF) fusion algorithm and local mean matching (LMM) fusion algorithm reported the optimum ED2 values at scale 60. It was difficult to distinguish optimum scale by analyzing scatter plots of ED2 because some fused products manifested multiple values for optimum scale. For example, (see Figure 20) the Subtractive Resolution Merge (RSM) fusion algorithm exhibited optimum segmentation at scales 60 and 70. Similarly, Gram-Schmidt (GS) fusion algorithm showed optimum segmentation at scales 50 and 60. Therefore, unlike in QR and ED1, we used black-hollowed-box to encapsulate (Figure 12) the optimum and two adjacent scales for the ease of understanding.

Table 13. General characteristics of the manually extracted IDP shelters from the MF1 study site

Variable	value
Sampled number of objects	30
Minimum area (m^2)	157.0
Maximum are a(m^2)	195.0
Mean area (m^2)	175.6
Standard deviation	10.8

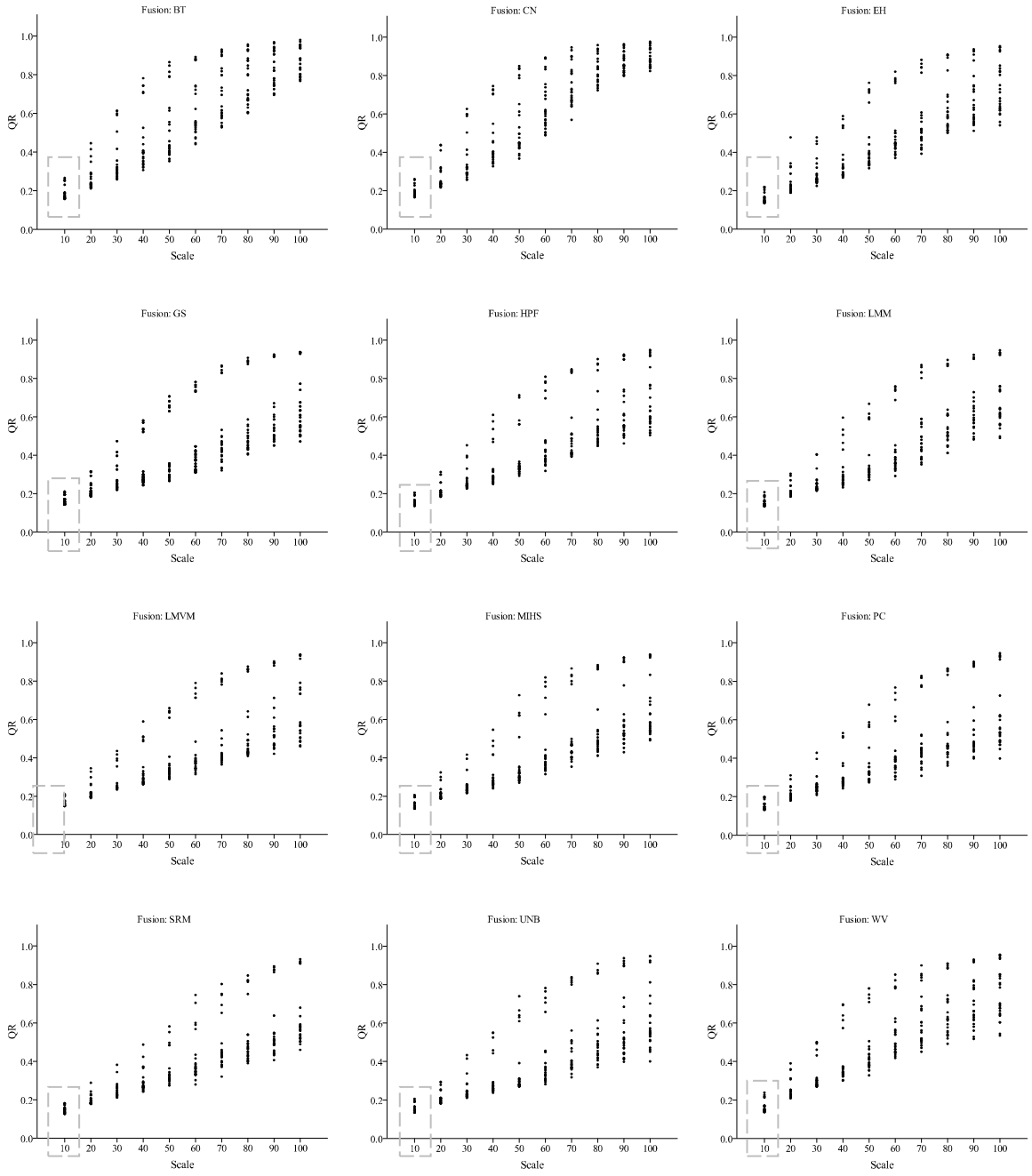


Figure 18. Scatter plots depicting the variation of the QR quality metric with respect to the multiresolution segmentation algorithm's different parameter settings across twelve fused products. The dashed-line box (scale window) encapsulated the optimum scale parameter with respect to the QR metric.

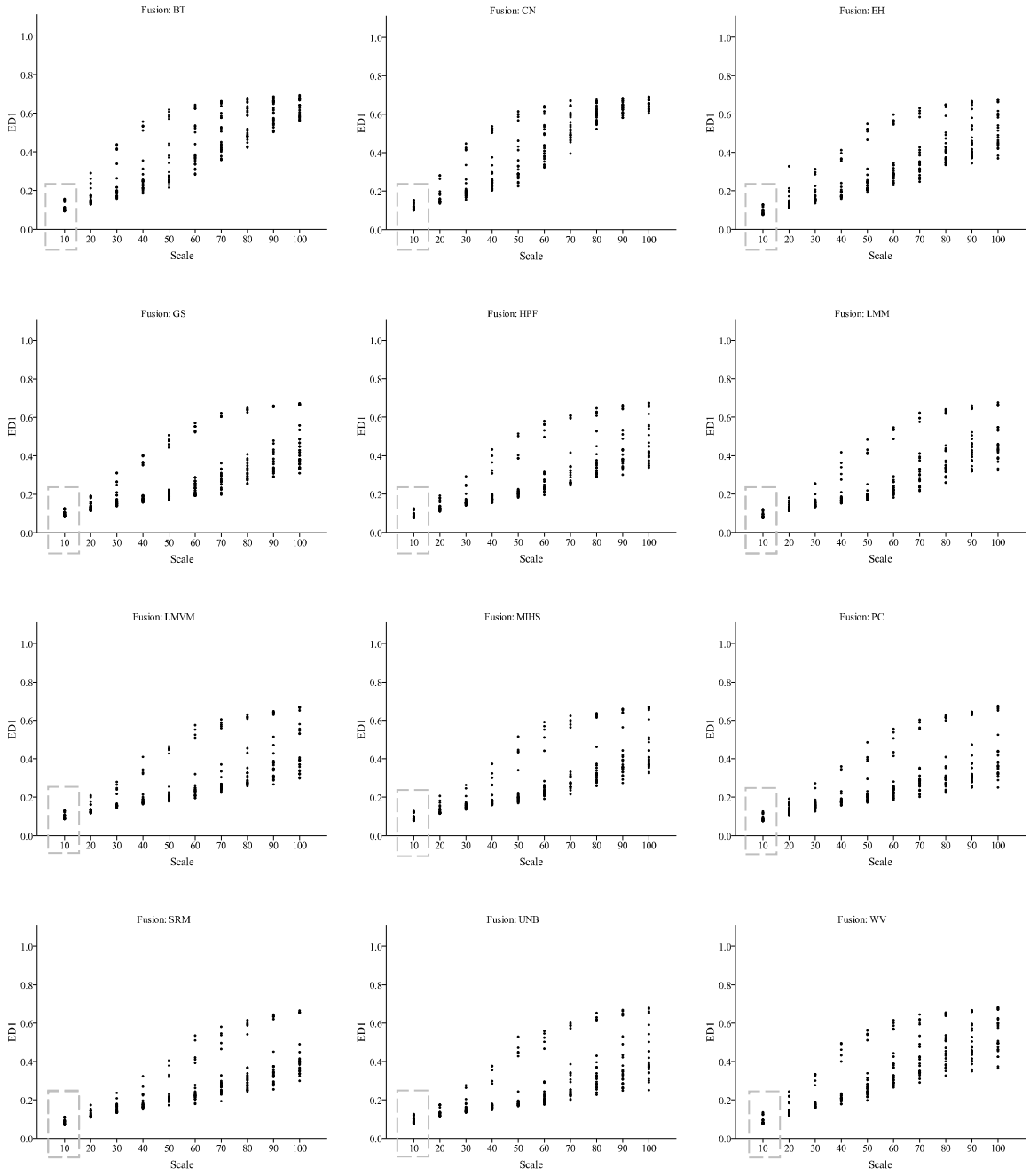


Figure 19. Scatter plots depicting the variation of the ED1 quality metric with respect to the multiresolution segmentation algorithm's different parameter settings across twelve fused products. The dashed-line box (scale window) encapsulated the optimum scale parameter with respect to the ED1 metric.

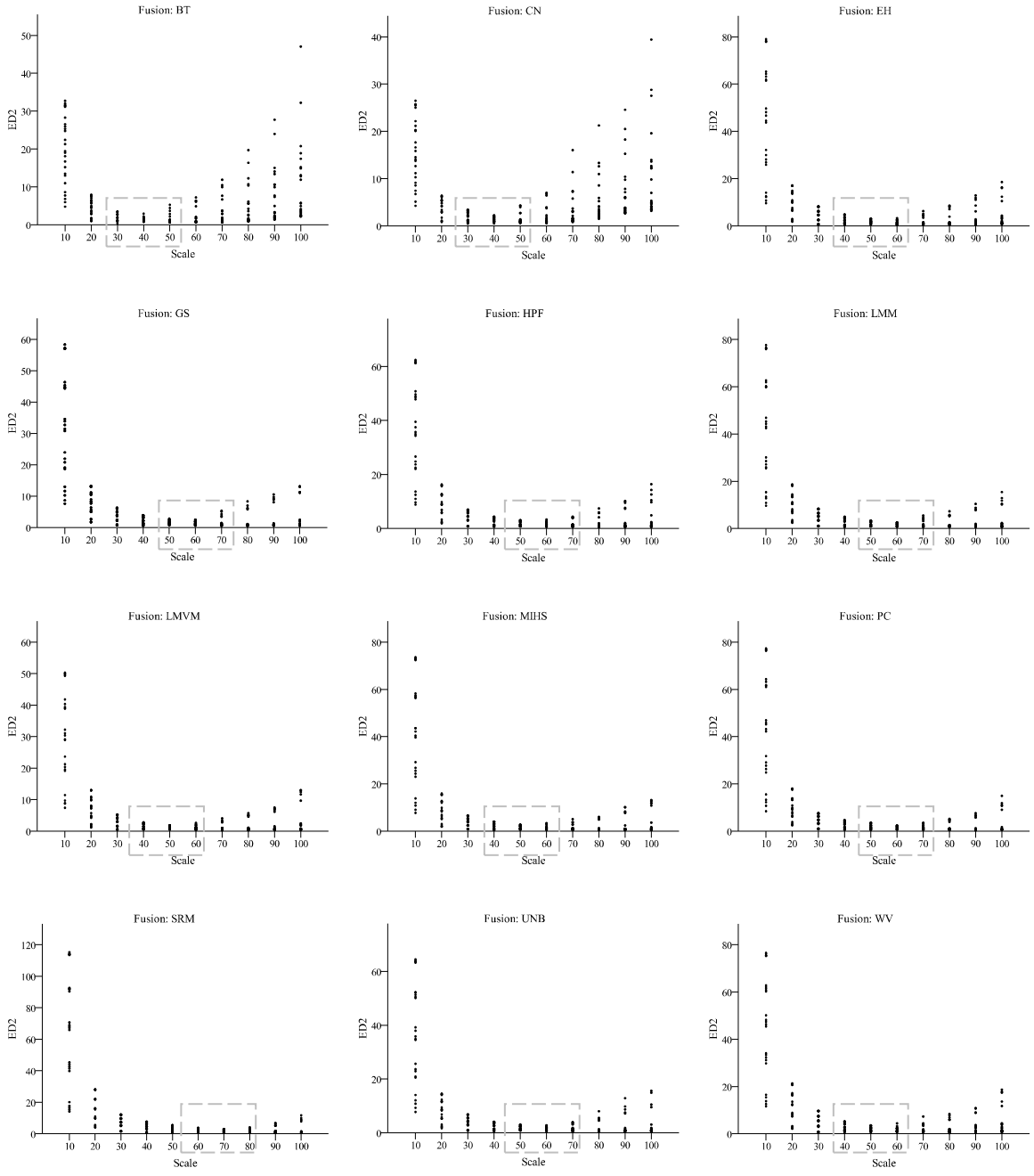


Figure 20. Scatter plots depicting the variation of the ED2 quality metric with respect to the multiresolution segmentation algorithm's different parameter settings across twelve fused products. The dashed-line box (scale window) encapsulated the optimum scale parameter with respect to the ED2 metric.

We used non-parametric statistical tests to compare the quality metric values (i.e., QR, ED1, and ED2) across different parameter settings of a fused product itself and among different fused products. As discussed in the methodology section, we aimed to compare image objects of fused images produced at the optimum parameter settings (i.e., scale, shape, and compactness) of the MRS algorithm. This comparison can be done based on each quality metric. Thus, it is decisive to focus on individual fused products and understand the optimum segmentation setting of the MRS algorithm for that particular fused image with respect to each quality metric. We applied Kruskal-Wallis test to analyze whether the quality of image objects varies across scales and different shape and compactness values. Tables 14.1 to 14.12 report the scores for the Kruskal-Wallis test. With respect to quality metrics (QR, ED1, and ED2), it is clear that all fused images produced statistically ($\alpha = 0.05$) different image objects across scales. Similarly, the change in the shape parameter significantly affected on the segmentation quality; however, the compactness parameter did not significantly influence on the quality of image objects. Based on this observation, further statistical analyses were confined to scale and shape parameters. In order to understand better the variation of the quality metrics with respect to the scale parameter, we conducted a sensitivity analysis. By design quality metrics should be sensitive to the variation of segmentation quality at different parameter settings. For example, when scale parameter changes from 10 to 20 (a coarser change in scale parameter), the quality metric (e.g., QR) should produce statistically different results for two scales. Tables 15.1 to 15.12 depict the reported scores for the Wilcoxon pairwise two-tailed test for the twelve fused products. For all fused images, QR and ED1 metrics were highly sensitive to the variation of scale

parameter and able to produce significant results at $\alpha = 0.025$ (for two-tailed test, significance level = $\alpha/2$). The ED2 metric was able capture quality variations (at $\alpha =$ 0.025) from scale 10 to 40 in most fused images; however, it failed to encapsulate e quality variation beyond optimum or near-optimum scale settings (compare Figure 20 and Tables 15.1 to 15.12). In case of the CN fusion algorithm, the ED2 metric captured segmentation quality variation beyond the optimum scale setting but was unable to distinguish quality difference at scales 40-50 and 50-60. Based on the sensitivity analysis, QR and ED1 seemed to be good quality metrics; however, they always present low values for the scale parameter. This leads to highly unfavorable over-segmentation (one-to-many correspondence, see Figure 17) scenario with very small image object candidates. Figure 21 illustrates how the segmentation quality of a fused image (Brovey transform) varies with increasing scale parameter. At lower scales, high geometrical congruency is achieved at the expense of arithmetic agreement while the opposite is observed at higher scales. Table 16 summarizes the number of corresponding image objects required with respect to three quality metrics for reconstructing 30 reference objects. For example, in case of BT fusion algorithm, QR and ED1 report 843 corresponding image objects (exceedingly high one-to-many relationships) while ED2 metric report only 43 corresponding image objects. Although the ED2 metric was not sensitive as QR and ED1, it achieved both good geometrical- and arithmetic-fit and produced meaningful results. Therefore, for the classification step, we used ED2-selected image objects. The next challenge was to determine the shape parameter at the optimum scale setting. As previously discussed, based on scatter plots we selected three scale settings, which encapsulate the optimum and near-optimum scale values (see Figure 20).

We call this as the scale-window (Figure 20, black-hollow boxes). Figure 22 depicts the variation of shape parameter for within the scale-window of each fused image. When analyzing box plots, the majority of fused products reported the lowest ED2 values at shape parameter = 0.7. We also applied the Wilcoxon pair-wise two-tailed test to analyze the statistical significance of different shape parameter settings (i.e., 0.1, 0.3,.., 0.9) within a given scale-window. Tables 17.1 to 17.12, report the Wilcoxon pair-wise two-tailed test results for all fused products. As we mentioned earlier, the compactness parameter did not produce statistically significantly different results for its different settings (i.e., 0.1, 0.3, ..., 0.9). This enabled us to select any value for the compactness parameter; however, based on visual inspections we chose 0.3 for the compactness parameter for all fused images. As the final step of the statistical analysis section, we applied Kruskal-Wallis test to compare quality metric-selected (QR selected, ED1 selected, and ED2 selected) image object candidates of fused images. Table 18 summarizes the reported *p*-values for the Kruskal-Wallis test. Test results revealed that image objects of twelve fused products are not significantly different with respect to any of the three quality metrics (QR, ED1, and ED2).

Table 14. Reported Kruskal-Wallis test results for the analysis of the individual fused product's segmentation quality (with respect to QR, ED1, and ED2) variation across scales and different shape and compactness settings. (Superscripts of ***, **, *, depict 0.05 significance level, 0.10 significance level, and 0.15 significance level, respectively)

Table 14.1 Brovey transform fusion algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9939
ED1	0^{***}	0^{***}	0.9896
ED2	0^{***}	0.0002^{***}	0.9905

Table 14.2 Color normalized pansharpening algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9982
ED1	0^{***}	0^{***}	0.9918
ED2	0^{***}	0.0049^{***}	0.9992

Table 14.3 Ehlers fusion algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9785
ED1	0^{***}	0^{***}	0.9782
ED2	0^{***}	0.0464^{***}	0.9942

Table 14.4 Gram-Schmidt fusion algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9422
ED1	0^{***}	0^{***}	0.9322
ED2	0^{***}	0.0004^{***}	0.9986

Table 14.5 High-pass filter fusion algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9999
ED1	0^{***}	0^{***}	0.9999
ED2	0^{***}	0.0083^{***}	0.9847

Table 14.6 Local mean matching fusion algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9975
ED1	0^{***}	0^{***}	0.9984
ED2	0^{***}	0.0296^{***}	0.9994

Table 14.7 Local mean variance matching fusion algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9919
ED1	0^{***}	0^{***}	0.9929
ED2	0^{***}	0.0014^{***}	0.9791

Table 14.8 Modified-IHS fusion algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9971
ED1	0^{***}	0^{***}	0.9921
ED2	0^{***}	0.0027^{***}	0.9841

Table 14.9 Principle component analysis fusion algoritm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9871
ED1	0^{***}	0^{***}	0.9868
ED2	0^{***}	0.0053^{***}	0.9620

Table 14.10 Subtractive resolution merge fusion algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9972
ED1	0^{***}	0^{***}	0.9968
ED2	0^{***}	0.0002^{***}	0.9825

Table 14.11 University of New Brunswick fusion algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9772
ED1	0^{***}	0^{***}	0.9626
ED2	0^{***}	0.0043^{***}	0.9966

Table 14.12 Wavelet-PCA fusion algorithm

Metric	Kruskal-Wallis test, *p*-value		
	Scale	Shape	Compactness
QR	0^{***}	0^{***}	0.9918
ED1	0^{***}	0^{***}	0.9903
ED2	0^{***}	0.2033	0.9942

(Superscripts ***, **, and * indicate significance levels (α) of 0.05, 0.10, and 0.15, respectively)

Table 15. Reported Wilcoxon two-tailed test results for the pairwise comparison of quality metric values (QR, ED1, and ED2) with respect to increasing scale parameter. (Superscripts of ***, **, *, depict 0.025 significance level, 0.05 significance level, and 0.075 significance level, respectively).

Table 15.1 Brovey transform fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0.0005^{***}	0.0177^{***}	0.0039^{***}	0.0208^{***}	0.0475^{**}	0.0289^{**}	0.0633^{*}
ED1	0^{***}	0^{***}	0^{***}	0.0014^{***}	0.0181^{***}	0.0037^{***}	0.0224^{***}	0.0359^{**}	0.0309^{**}	0.0618^{*}
ED2	0^{***}	0^{***}	0^{***}	0.0696^{*}	0.0559^{*}	0.9304	0.1837	0.0523^{*}	0.0409^{**}	0.0791

(Superscripts ***, **, and * indicate significance levels (α) of 0.025, 0.05, and 0.075, respectively)

Table 15.2 Color normalized pansharpening algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0.0002^{***}	0.0101^{***}	0.0015^{***}	0.0054^{***}	0.0037^{***}	0.0096^{***}	0.0915
ED1	0^{***}	0^{***}	0^{***}	0.0003^{***}	0.0114^{***}	0.0011^{***}	0.0045^{***}	0.0026^{***}	0.0066^{***}	0.0361^{**}
ED2	0^{***}	0^{***}	0.0002^{***}	0.0222^{***}	0.1467	0.6581	0.0189^{***}	0.0051^{***}	0.0189^{***}	0.0621^{*}

(Superscripts ***, **, and * indicate significance levels (α) of 0.025, 0.05, and 0.075, respectively)

Table 15.3 Ehlers fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0.0001^{****}	0.0004^{***}	0.0005^{***}	0.0010^{***}	0.0230^{***}	0.0071^{***}	0.0929	0.1007
ED1	0^{***}	0^{***}	0.0001^{***}	0.0007^{***}	0.0005^{***}	0.0009^{***}	0.0244^{***}	0.0069^{***}	0.0842	0.0932
ED2	0^{***}	0^{***}	0.0001^{***}	0.0189^{***}	0.1428	0.0114^{***}	0.8626	0.4185	0.3753	0.1522

Table 15.4 Gram-Schmidt fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0^{***}	0^{***}	0^{***}	0.0089^{***}	0.0237^{***}	0.0247^{***}	0.0567^{**}
ED1	0^{***}	0^{***}	0^{***}	0^{***}	0^{***}	0^{***}	0.0081^{***}	0.0250^{***}	0.0237^{***}	0.0567^{**}
ED2	0^{***}	0^{***}	0.0001^{***}	0.0065^{***}	0.0894	0.1428	0.2729	0.7268	0.5410	0.3930

(Superscripts ***, **, and * indicate significance levels (α) of 0.025, 0.05, and 0.075, respectively)

Table 15.5 High-pass filter fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0^{***}	0.0003^{***}	0.0004^{***}	0.0078^{***}	0.0128^{***}	0.02149^{***}	0.1888
ED1	0^{***}	0^{***}	0^{***}	0.0001^{***}	0.0003^{***}	0.0003^{***}	0.0070^{***}	0.0135^{***}	0.0204^{***}	0.1888
ED2	0^{***}	0^{***}	0.0002^{***}	0.0038^{***}	0.0711^{*}	0.4263	0.3448	0.7148	0.6581	0.4410

(Superscripts ***, **, and * indicate significance levels (α) of 0.025, 0.05, and 0.075, respectively)

Table 15.6 Local mean matching fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0^{***}	0.0015^{***}	0.0010^{***}	0.0012^{***}	0.0649^{*}	0.0641^{*}	0.111
ED1	0^{***}	0^{***}	0^{***}	0^{***}	0.0012^{***}	0.0011^{***}	0.0016^{***}	0.0681^{*}	0.0610^{*}	0.111
ED2	0^{***}	0^{***}	0.0002^{***}	0.0020^{***}	0.0773^{*}	0.2563	0.3516	0.4491	0.6555	0.4688

(Superscripts ***, **, and * indicate significance levels (α) of 0.025, 0.05, and 0.075, respectively)

Table 15.7 Local mean variance matching fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0^{***}	0.0026^{***}	0.0032^{***}	0.00283^{***}	0.0032^{***}	0.0209^{***}	0.1031
ED1	0^{***}	0^{***}	0^{***}	0^{***}	0.0018^{***}	0.0021^{***}	0.0032^{***}	0.0033^{***}	0.0214^{***}	0.1035
ED2	0^{***}	0^{***}	0^{***}	0.0069^{***}	0.1213	0.2548	0.4074	0.8538	0.6034	0.2312

(Superscripts ***, **, and * indicate significance levels (α) of 0.025, 0.05, and 0.075, respectively)

Table 15.8 Modified-IHS fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0^{***}	0.0001^{***}	0.0014^{***}	0.0020^{***}	0.0160^{***}	0.0169^{***}	0.0696^{*}
ED1	0^{***}	0^{***}	0^{***}	0^{***}	0.0039^{***}	0.0004^{***}	0.0028^{***}	0.0170^{***}	0.0179^{***}	0.0696^{*}
ED2	0^{***}	0_{***}	0.0001^{***}	0.0033^{***}	0.1467	0.2237	0.2630	0.4849	0.4377	0.7148

(Superscripts ***, **, and * indicate significance levels (α) of 0.025, 0.05, and 0.075, respectively)

Table 15.9 Principal component analysis fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0^{***}	0.0012^{***}	0.0062^{***}	0.0069^{***}	0.1838	0.1483	0.0336^{**}
ED1	0^{***}	0^{***}	0^{***}	0^{***}	0^{***}	0.0009^{***}	0.0042^{***}	0.1838	0.1429	0.0336^{**}
ED2	0^{***}	0^{***}	0.0001^{***}	0.0021^{***}	0.0456^{**}	0.1261	0.5128	0.1698	0.4643	0.5379

Table 15.10 Subtractive resolution merge fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0^{***}	0.0001^{***}	0.0023^{***}	0.0024^{***}	0.0222^{***}	0.0336^{**}	0.0110^{***}
ED1	0^{***}	0^{***}	0^{***}	0^{***}	0.0001^{***}	0.0074^{***}	0.0021^{***}	0.0244^{***}	0.0361^{**}	0.0101^{***}
ED2	0^{***}	0^{***}	0^{***}	0.0002^{***}	0.0060^{***}	0.0121^{***}	0.0878^{**}	0.3548	0.8475	0.5253

(Superscripts ***, **, and * indicate significance levels (α) of 0.025, 0.05, and 0.075, respectively)

Table 15.11 The University of New Brunswick fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0^{***}	0.0009^{***}	0.0018^{***}	0.0066^{***}	0.0288^{**}	0.1969	0.1124
ED1	0^{***}	0^{***}	0^{***}	0^{***}	0.0012^{***}	0.0027^{***}	0.0067^{***}	0.0319^{**}	0.2038	0.1218
ED2	0^{***}	0^{***}	0.0004^{***}	0.0036^{***}	0.0711^{*}	0.1936	0.3417	0.5768	0.6581	0.6271

(Superscripts ***, **, and * indicate significance levels (α) of 0.025, 0.05, and 0.075, respectively)

Table 15.12 Wavelet-PCA fusion algorithm

Metric	Kruskal-Wallis test, *p*-value	Pairwise tests (two-tailed), *p*-values								
		Scale								
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
QR	0^{***}	0^{***}	0^{***}	0^{***}	0.0005^{***}	0.0007^{***}	0.0408^{**}	0.0567^{*}	0.1310	0.2092
ED1	0^{***}	0^{***}	0^{***}	0^{***}	0.0006^{***}	0.0005^{***}	0.0406^{**}	0.0547^{*}	0.1229	0.1954
ED2	0^{***}	0^{***}	0.0003^{***}	0.0056^{**}	0.0669^{*}	0.0773^{*}	0.6305	0.7730	0.1310	0.3548

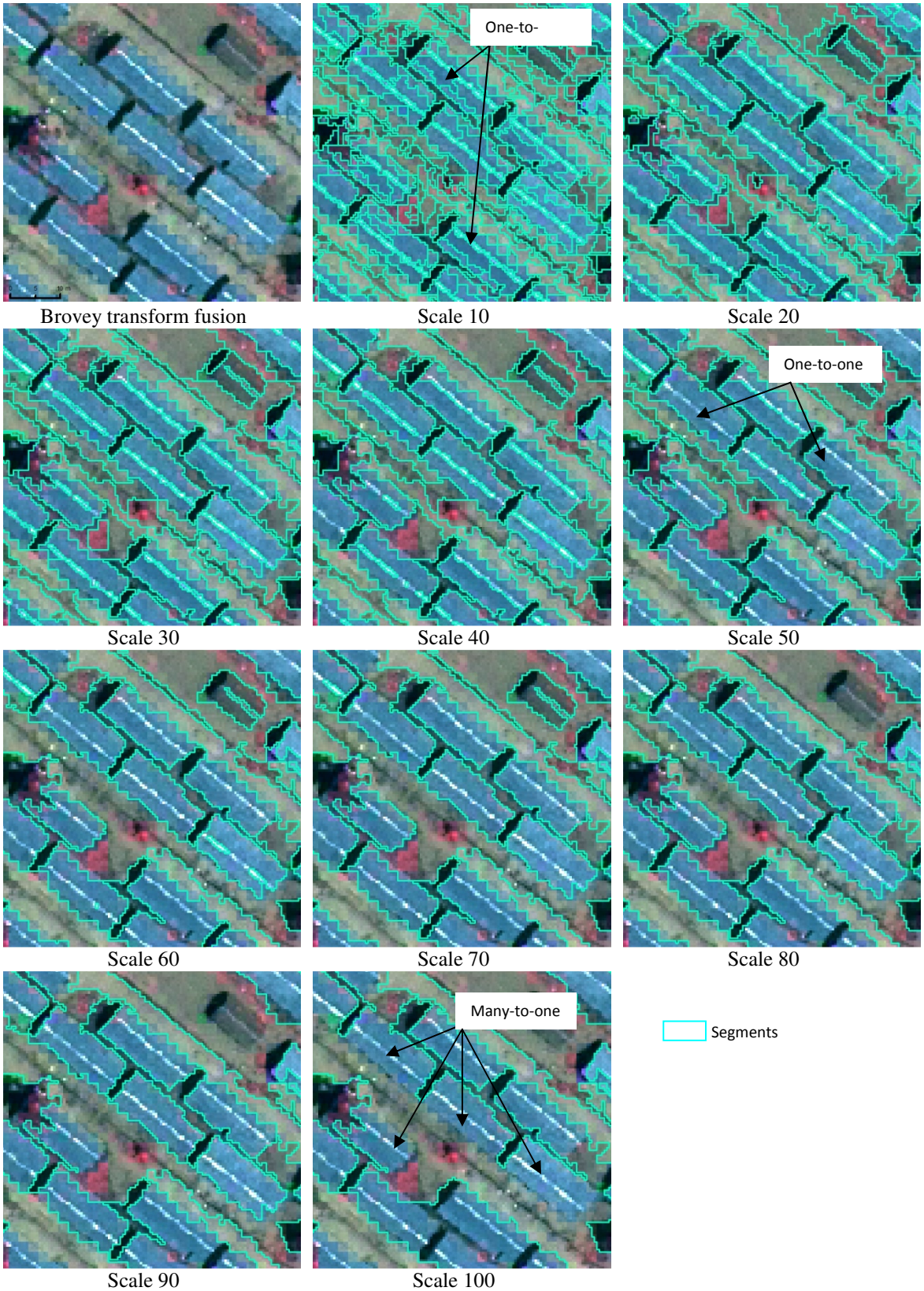
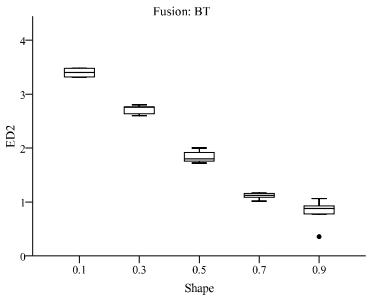


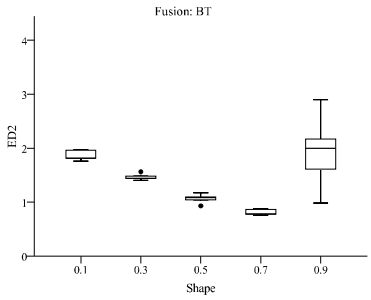
Figure 21. Shows how the segmentation quality of a fused image (Brovey transform) varies with increasing scale parameter

Table 16. Summary of the number of corresponding image objects required with respect to three quality metrics (QR, ED1, and ED2) for reconstructing 30 reference objects

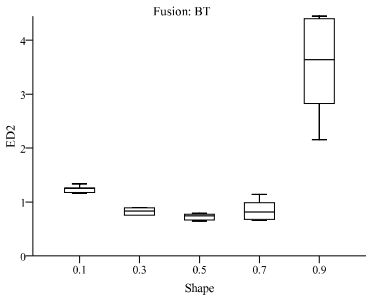
Quality metric	Fusion algorithm											
	BT	CN	EH	GS	HPF	LMM	LMVM	MIHS	PC	SRM	UNB	WV
QR	843	687	2000	1483	1585	1965	1282	1865	1957	2905	1638	1938
ED1												
ED2	43	41	44	38	46	48	40	44	40	42	41	46



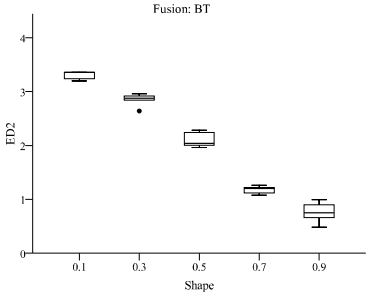
Scale 30



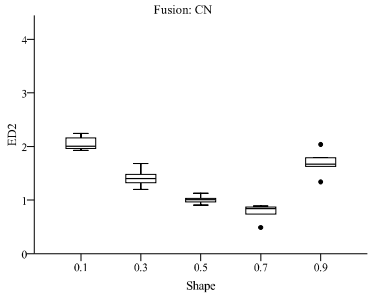
Scale 40



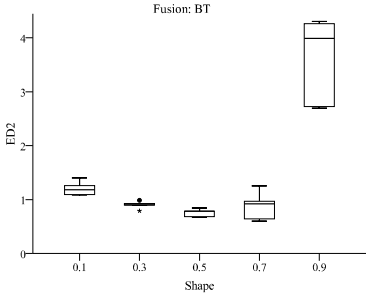
Scale 50



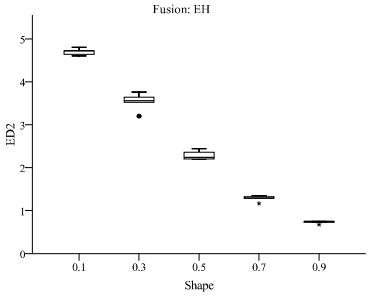
Scale 30



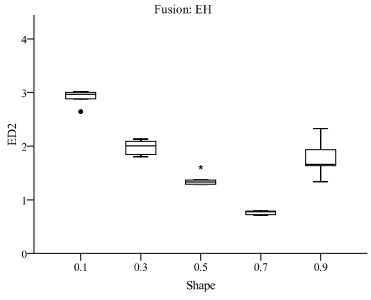
Scale 40



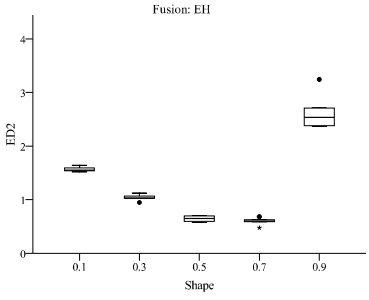
Scale 50



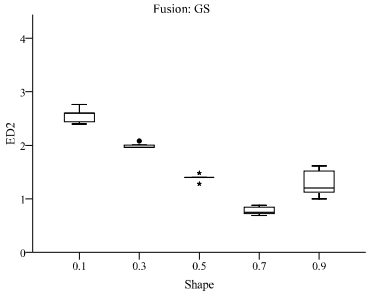
Scale 40



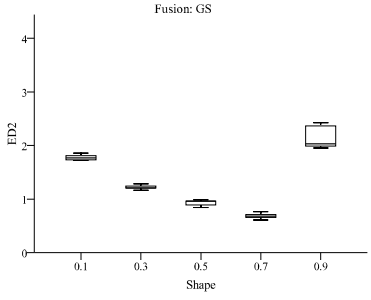
Scale 50



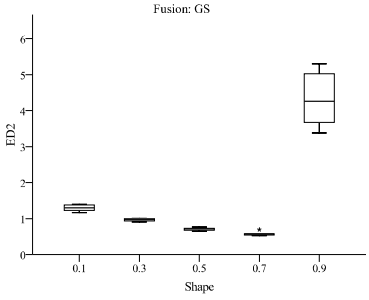
Scale 60



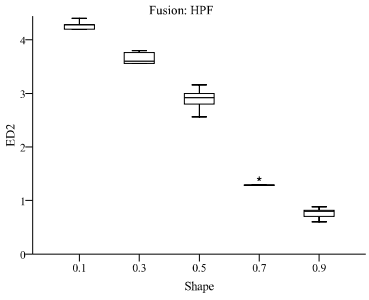
Scale 50



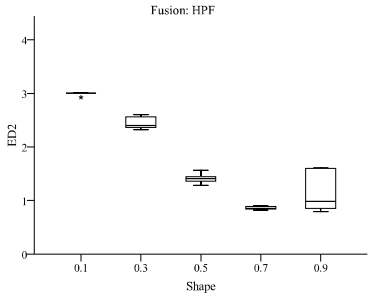
Scale 60



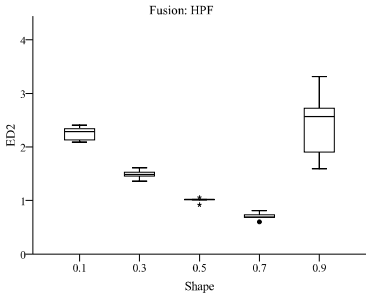
Scale 70



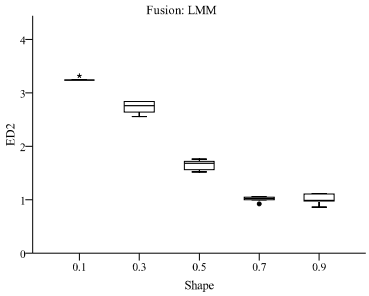
Scale 40



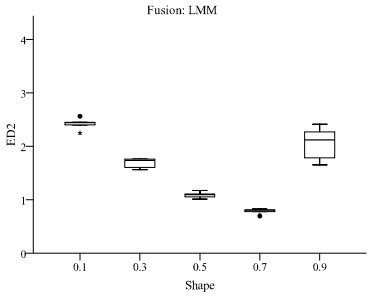
Scale 50



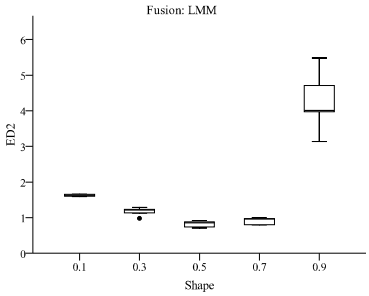
Scale 60



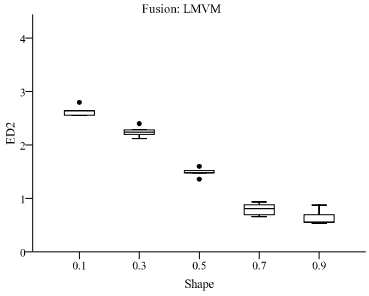
Scale 50



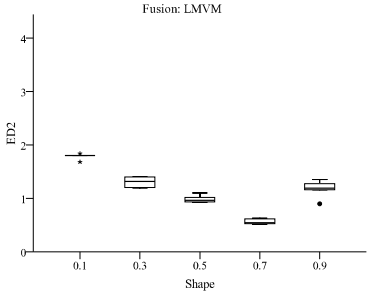
Scale 60



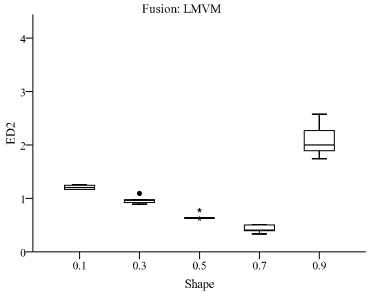
Scale 70



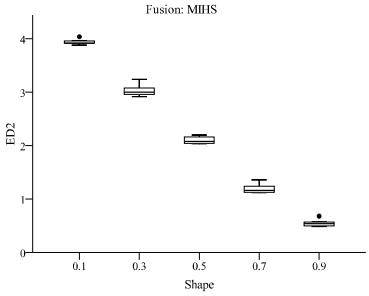
Scale 40



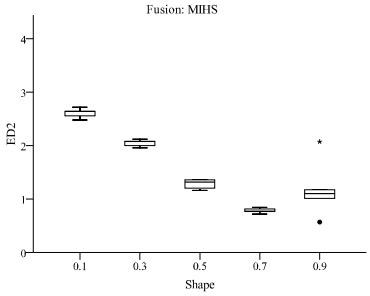
Scale 50



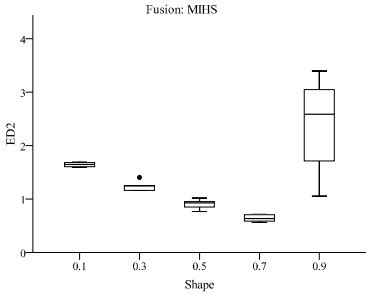
Scale 60



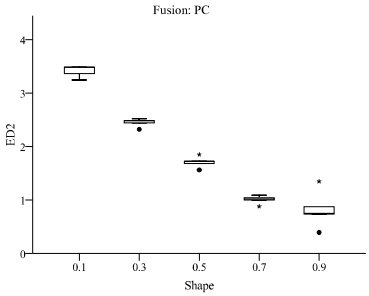
Scale 40



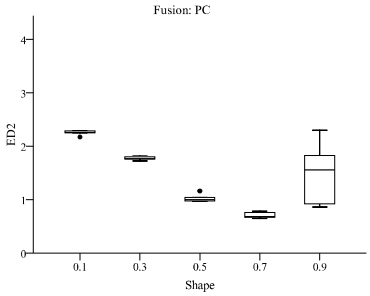
Scale 50



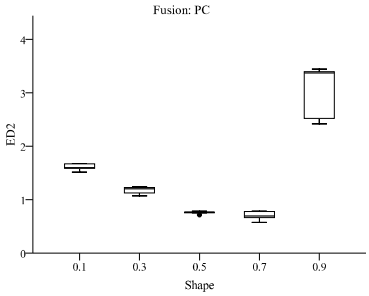
Scale 60



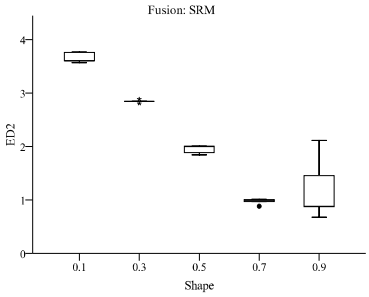
Scale 50



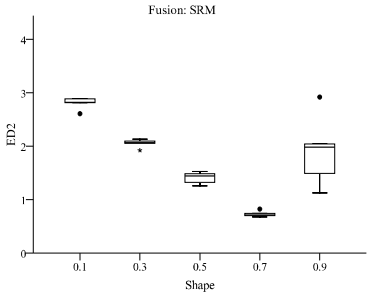
Scale 60



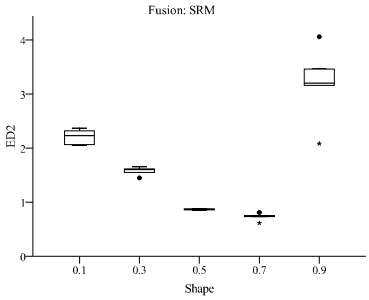
Scale 70



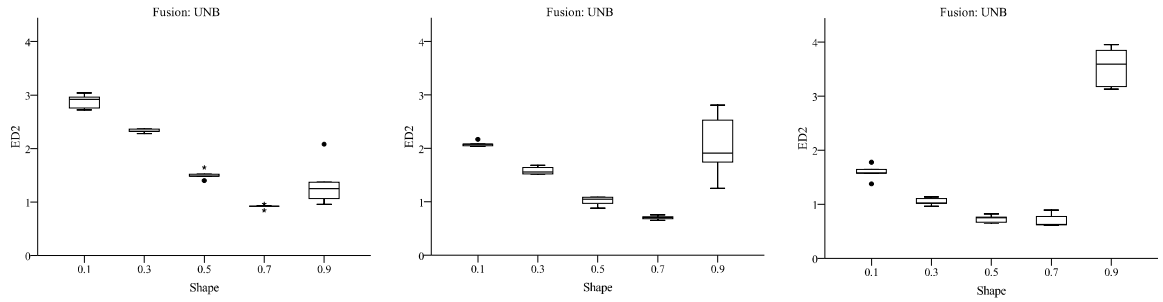
Scale 60



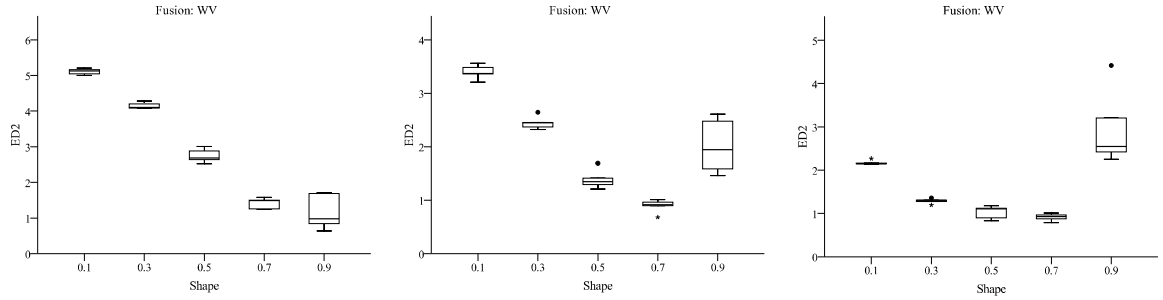
Scale 70



Scale 80



Scale 50 Scale 60 Scale 70



Scale 40 Scale 50 Scale 60

Figure 22. Box plots depicting the variation of shape parameter for within the scale-window of each fused image

Table 17. Reported Wilcoxon two-tailed test results for the pairwise comparison of the shape parameter at the quality metric-selected (QR, ED1, and ED2) optimum scale settings. (Superscripts of ***, **, *, depict 0.025 significance level, 0.05 significance level, and 0.075 significance level)

Table 17.1 Brovey transform fusion algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.1732	0.0079^{***}	0.0079^{***}	0.0079^{***}
ED1	10	0.6905	0.0079^{***}	0.0079^{***}	0.0079^{***}
ED2	30	0.0119^{***}	0.0079^{***}	0.0079^{***}	0.0159^{***}
	40	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	50	0.0079^{***}	0.0952	0.3095	0.0079^{***}

Table 17.2 Color normalized pansharpening algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0317^{**}	0.0079^{***}	0.0079^{***}	0.0079^{***}
ED1	10	0.0317^{**}	0.0317^{**}	0.0119^{***}	0.0119^{***}
ED	30	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	40	0.0079^{***}	0.0079^{***}	0.0079^{**}	0.0079^{***}
	50	0.0079^{***}	0.0158^{***}	0.6905	0.0079^{***}

Table 17.3 Ehlers fusion algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0194^{***}	0.0119^{***}	0.0079^{***}	0.0079^{***}
ED1	10	0.0158^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
ED	40	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	50	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	60	0.0079^{***}	0.0079^{***}	0.4206	0.0079^{***}

Table 17.4 Gram-Schmidt fusion algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0119^{***}	0.0119^{***}	0.0119^{***}	0.0079^{***}
ED1	10	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
ED	50	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	60	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	70	0.0079^{***}	0.0079^{***}	0.0317^{**}	0.0079^{***}

Table 17.5 High-pass filter fusion algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0119^{***}	0.0119^{***}	0.0079^{***}	0.0119^{***}
ED1	10	0.0111^{***}	0.01091^{***}	0.0116^{***}	0.0116^{***}
ED	40	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	50	0.0119^{***}	0.0079^{***}	0.0079^{***}	0.3095
	60	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}

Table 17.6 Local mean matching fusion algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
ED1	10	0.0119^{***}	0.01198^{***}	0.0079^{***}	0.0079^{***}
ED	50	0.0079^{***}	0.0079^{***}	0.0079^{***}	1
	60	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	70	0.0079^{***}	0.0079^{***}	0.2222^{**}	0.0079^{***}

Table 17.7 Local mean variance matching fusion algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0079^{***}	0.0119^{***}	0.0119^{***}	0.0079^{***}
ED1	10	0.0119^{***}	0.0119^{***}	0.0079^{***}	0.0079^{***}
ED	40	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.1508
	50	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	60	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}

Table 17.8 Modified-IHS fusion algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0119^{***}	0.0119^{***}	0.0079^{***}	0.0079^{***}
ED1	10	0.0079^{***}	0.0119^{***}	0.0119^{***}	0.0079^{***}
ED2	40	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	50	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.1508
	60	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}

Table 17.9 Principle component analysis fusion algoritm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0114^{***}	0.0119^{***}	0.0079^{***}	0.0079^{***}
ED1	10	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
ED2	50	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.1508
	60	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	70	0.0079^{***}	0.0079^{***}	0.5476	0.0079^{***}

Table 17.10 Subtractive resolution merge fusion algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0119^{***}	0.0119^{***}	0.0079^{***}	0.0079^{***}
ED1	10	0.0114^{***}	$0.0.0116^{***}$	0.0079^{***}	0.0079^{***}
ED2	60	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.6905
	70	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	80	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}

Table 17.11 University of New Brunswick fusion algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0079^{***}	0.0079***	0.0079^{***}	0.0079^{***}
ED1	10	0.0119^{***}	0.0079^{***}	0.0119^{***}	0.0119^{***}
ED2	50	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0158^{***}
	60	0.0119^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	70	0.0079^{***}	0.0079^{***}	0.5476	0.0079^{***}

(

Table 17.12 Wavelet-PCA fusion algorithm

Metric	Scale	Pairwise tests (two-tailed), *p*-values			
		Shape			
		0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9
QR	10	0.0119^{***}	0.0158^{***}	0.0079^{***}	0.0079^{***}
ED1	10	0.01167^{***}	0.02733^{**}	0.0079^{***}	0.0079^{***}
ED2	40	0.0158^{***}	0.0158^{***}	0.0079^{***}	0.6905
	50	0.0079^{***}	0.0079^{***}	0.0079^{***}	0.0079^{***}
	60	0.0079^{***}	0.0079^{***}	0.3095	0.0079^{***}

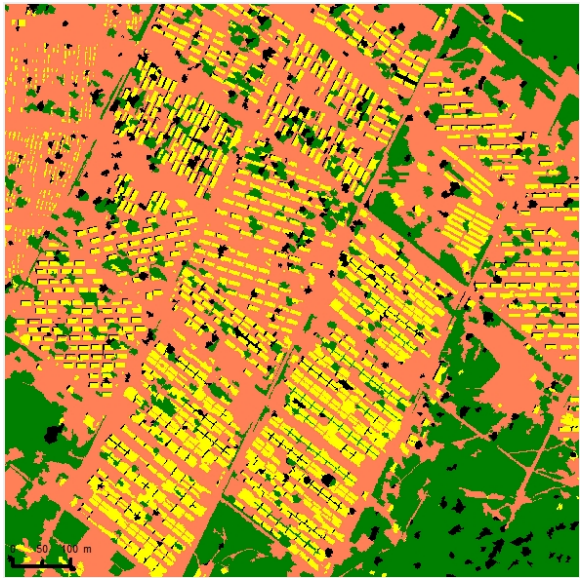
Table 18. Reported Kruskal-Wallis test results for the multiple comparison of the segmentation quality among twelve fused images at their optimum segmentation settings (based on QR, ED1, and ED2 metrics). (Superscripts of ***, **, *, depict 0.05 significance level, 0.10 significance level, and 0.15 significance level, respectively)

Metric	Kruskal-Wallis test, *p*-value
QR	0.2423
ED1	0.2340
ED2	0.4660

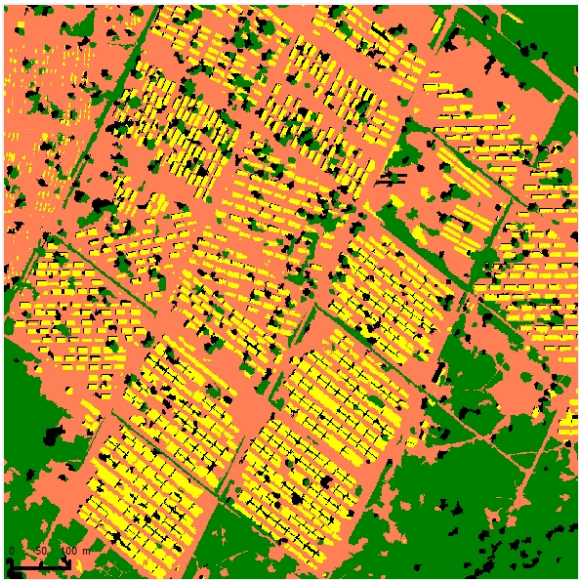
We classified ED2-selected image objects into four classes (IDP dwelling, vegetation, bare soil, and shadow) using the SVM classifier. Classification results of twelve fused products are shown in Figure 23. Table 19 summarizes classification accuracy assessment results for all fused products across four land use/cover categories. When considering overall accuracy and overall kappa, the Gram-Schmidt (GS) fusion algorithm and the color normalized (CN) pansharpening fusion algorithm reported the best and worst values, respectively. In general, compared with the other three classes, all fused products exhibited very low accuracy values (user's accuracy, producer's accuracy, and Kappa) for the shadow class. Taking into account the IDP dwelling class, which contains the focal feature of interest the principal component analysis (PCA) fusion algorithm, the modified-IHS (MIHS) fusion algorithm, the GS fusion algorithm achieved the highest producer's accuracy (78.95%). The lowest producer's accuracy was reported by the local mean matching (LMM) fusion algorithm. With respect to user's accuracy, the University of New Brunswick (UNB) fusion algorithm, the MIHS fusion algorithm, the GS fusion algorithm, and the Brovey transform (BT) fusion algorithm reported scores larger than 95 percent. Among the twelve fused products, the subtractive resolution merge (SRM) algorithm exhibited the lowest score for the user's accuracy. In terms of Kappa, the UNB fusion algorithm, the MIHS fusion algorithm, the PC fusion algorithm,

the GS fusion algorithm, and the BT fusion algorithm achieved markedly high values (> 0.95) while the SRM fusion reported the lowest Kappa value (0.75). In general, the LMM fusion algorithm, the CN fusion algorithm, and the Wavelet-PCA fusion algorithm showed comparatively low (~ 0.8) Kappa values.

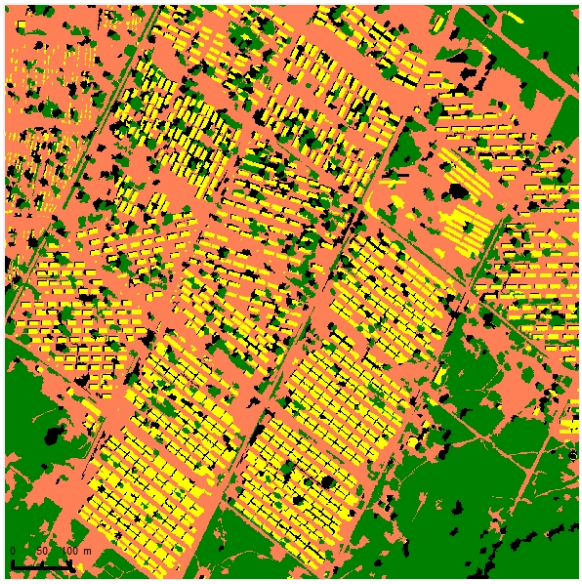
Brovey Transform



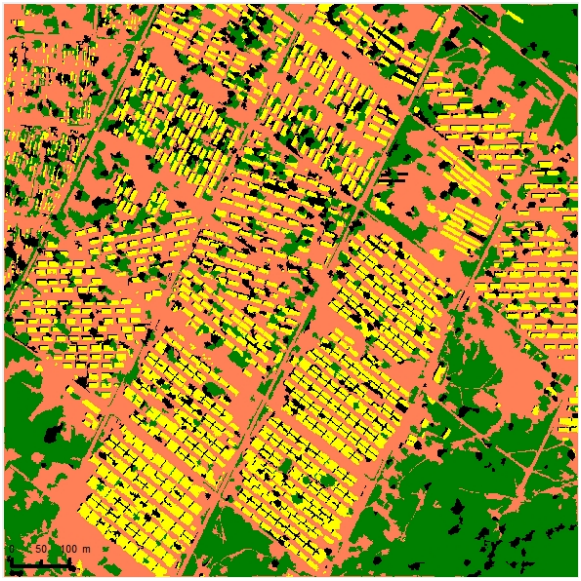
Color normalized pansharpening



Ehlers fusion

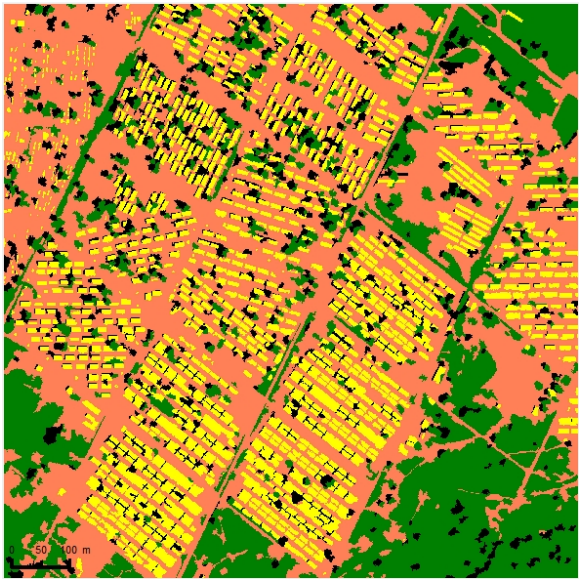


Gram-Schmidt fusion



High-pass filter fusion

Local mean matching fusion



Local mean variance matching fusion

Modified-IHS fusion

Figure 23. Support vector machine (SVM) classifier based classification results of the fused images of the MF1 study site.

Table 19. Summary of the classification accuracy assessment results of the MF1 study site for all fused products across four land use/cover categories.

Fusion algorithm	Land use/cover category												Overall accuracy (%)	Overall Kappa
	IDP dwelling			Vegetation			Bare soil			Shadow				
	Producer's accuracy (%)	User's accuracy (%)	Kappa	Producer's accuracy (%)	User's accuracy (%)	Kappa	Producer's accuracy (%)	User's accuracy (%)	Kappa	Producer's accuracy (%)	User's accuracy (%)	Kappa		
BT	76.32	96.67	0.9588	70.21	67.35	0.5732	86.60	76.36	0.5410	38.89	63.64	0.6004	76.50	0.6341
CN	71.05	84.38	0.8071	55.32	60.47	0.4832	83.51	68.07	0.3799	16.67	50.00	0.4505	68.50	0.4982
EH	73.68	87.50	0.8457	74.47	71.43	0.6265	86.60	75.68	0.5277	27.78	62.50	0.5879	76.00	0.6246
GS	78.95	96.77	0.9602	78.72	77.08	0.7004	89.69	84.47	0.6984	44.44	44.44	0.3895	81.00	0.7105
HPF	76.32	90.63	0.8843	63.83	56.60	0.4327	83.51	77.14	0.5562	27.78	50.00	0.4505	72.50	0.5757
LMM	68.42	83.87	0.8009	68.09	66.67	0.5643	88.66	76.79	0.5492	33.33	66.67	0.6337	75.00	0.6085
LMVM	76.32	90.63	0.8843	61.70	65.91	0.5544	87.63	75.22	0.5189	27.78	45.45	0.4006	74.00	0.5931
MIHS	78.95	100.00	1	65.96	67.39	0.5737	89.69	78.38	0.5802	38.89	53.85	0.4928	77.50	0.6498
PC	78.95	96.77	0.9602	85.11	64.52	0.5362	83.51	81.00	0.6311	33.33	85.71	0.8430	78.50	0.6703
SRM	73.68	80.00	0.7531	76.60	70.59	0.6155	82.47	78.43	0.5812	33.33	66.67	0.6337	75.00	0.6186
UNB	71.05	100.00	1	70.21	64.71	0.5386	89.69	77.68	0.5666	33.33	60.00	0.5604	76.50	0.6318
WV-PCA	71.05	84.38	0.8071	78.72	69.81	0.6054	85.57	79.05	0.5932	27.78	50.00	0.4505	76.00	0.6297

Certain sectors of the MF1 study site include IDP shelters with dark rooftops compared to other sectors. Figure 24 shows the PAN image of MF1 study site and zoomed-in views of bright and dark shelters. We employed a side-by-side visual inspection approach to assess simultaneously the fusion quality, segmentation quality, and classification accuracy of two zoomed-in areas comprising bright and dark shelters. Figures 25 (bright shelters) and 18 (dark shelters) depict original images (PAN and MS), fused images, segmentation results (ED2-selected image object candidates), and SVM-classified image. When examining Figure 25, in general, all fused products lead to satisfactory classification results. Pansharpened images from the CN fusion algorithm and the SRM fusion algorithm showed notable color distortions, however, they produced good segmentation and classification results with respect to the IDP dwelling class. The Ehlers fusion algorithm showed color artifacts (fuzzy appearance) around IDP shelters (see the circled area in Figure 17) and those arteficts lead to misclassification of bare soil as IDP shelters. A similar observation can be made in the pansharpened image, image segments, and classification results of the local mean variance matching (LMVM) fusion algorithm. Some color reversions (see the circled area in Figure 25) can be seen in the fused images from the MIHS fusion algorithm and the PCA fusion algorithm. However, they have produced good segmentation and classification results. Among the twelve fused images, the Wavelet-PCA showed the worst fusion quality. Image object candidates and classification results clearly manifest the poor edge enhancement associated with the Wavelet-PCA fusion algorithm. When examining Figure 26, unlike in bright shelters (Figure 25), most of the fused products failed to produce satisfactory segmentation and classifications results. This was mainly due to the close spacing of IDP

shelters and similar radiometric characteristics of dark rooftops and surrounding bare soil. In many cases, a one-to-many relationship between real world objects and image objects was observed, i.e., IDP shelters were not segmented and classified individually. For example, the panshapened image of the Ehlers fusion algorithm produced large segments containing two IDP shelters (see the circled area in Figure 25) and lead to poor classification results. Among the twelve candidate fusion algorithms, the GS fusion algorithm and the UNB fusion algorithm detected and classified IDP shelters individually. Similar to bright shelters, dark shelters exhibited the worst fusion, segmentation, and classification quality with the Wavelet-PCA fusion algorithm. The fused image showed no edge enhancement (see the circled area in Figure 25), thus, the segmentation algorithm has followed fuzzy pixel edges producing very rough object boundaries.

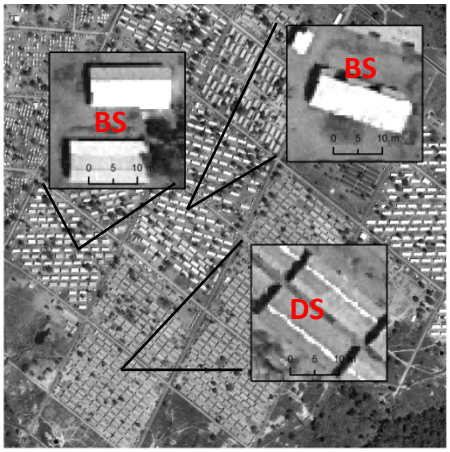
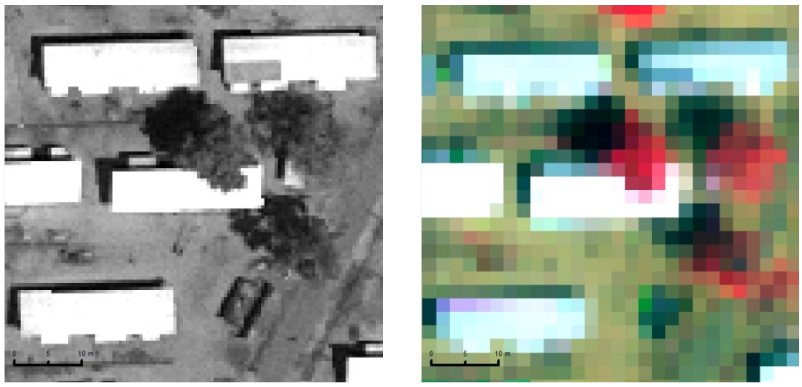
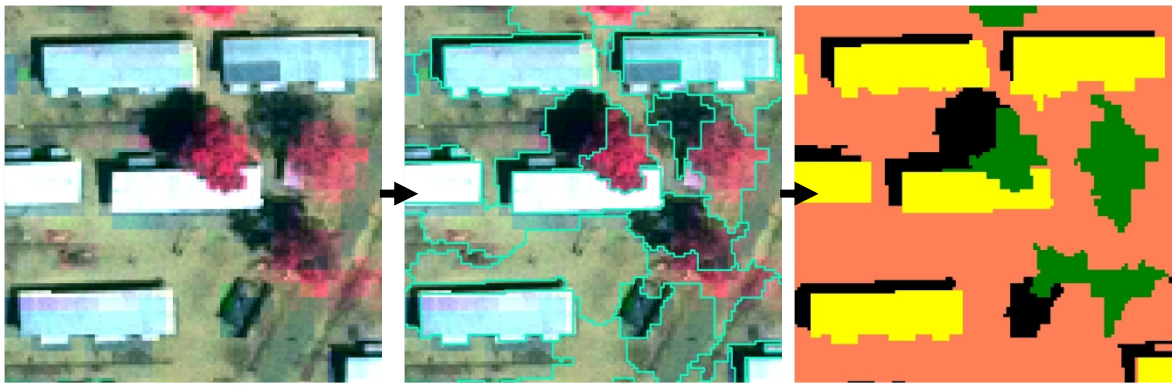


Figure 24. A zoomed-in view of the PAN image of the MF1 study site showing shows bright (BS) and dark (DS) shelters

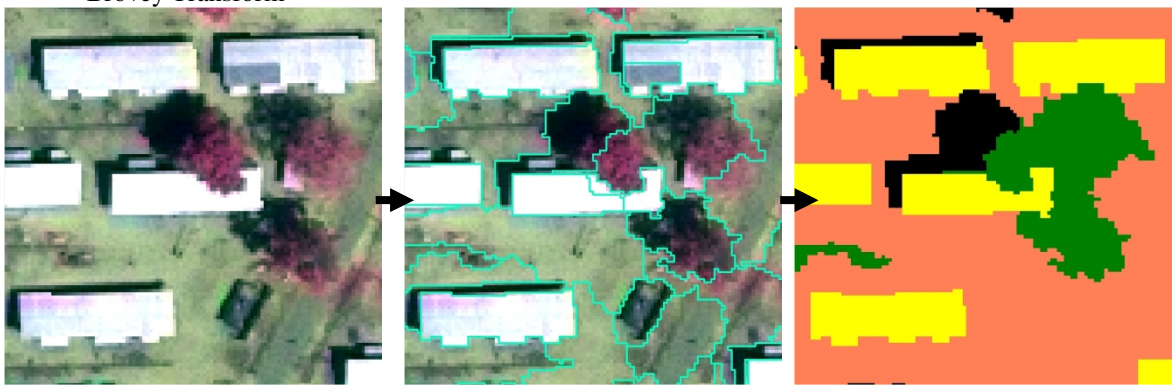


Original PAN Original MS

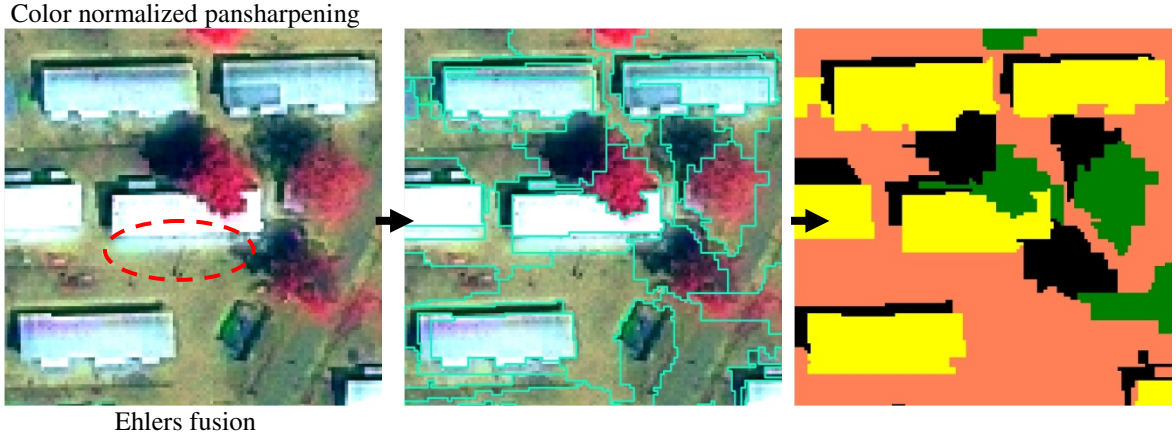
Fusion → **Segmentation** → **Classification**



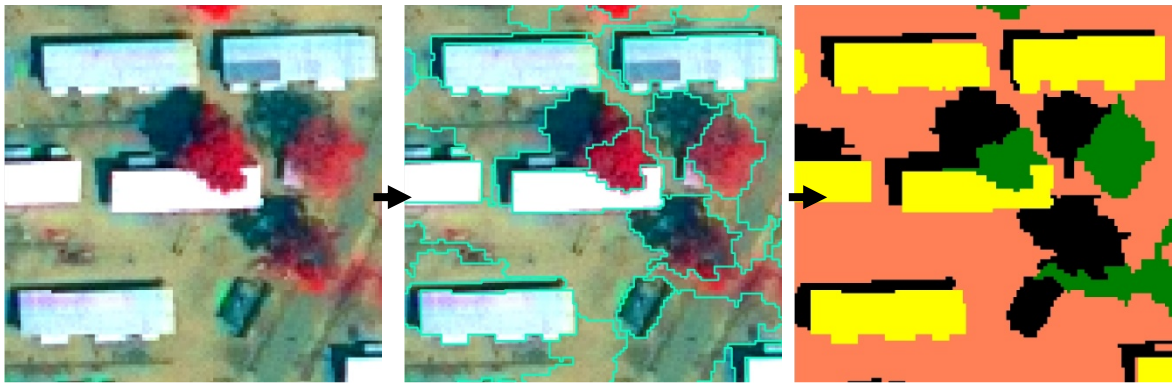
Brovey Transform



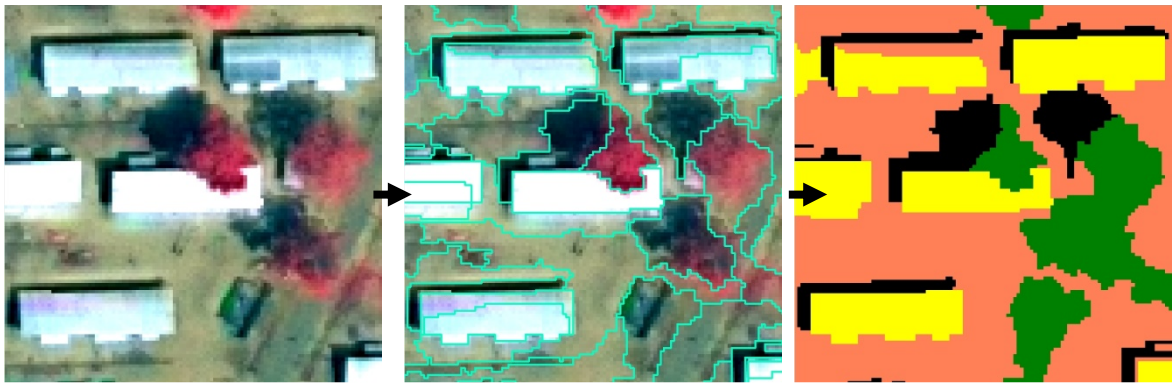
Color normalized pansharpening



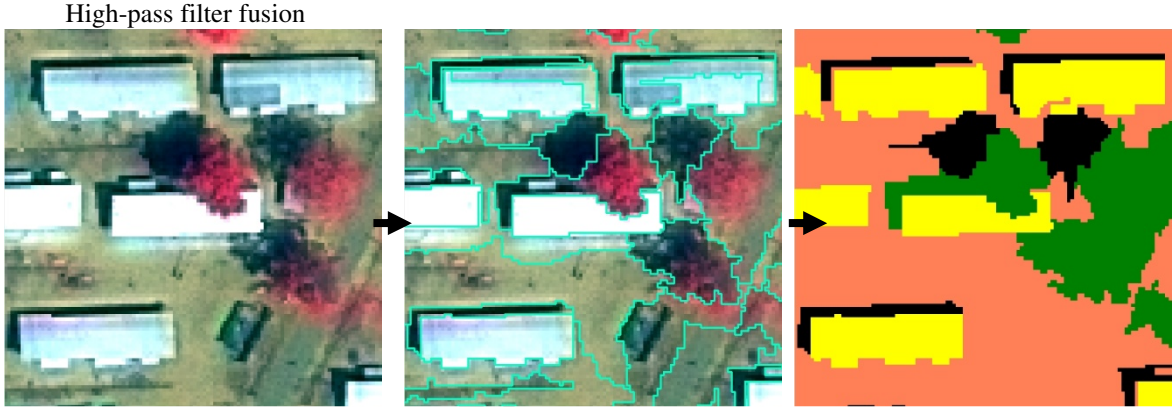
Ehlers fusion



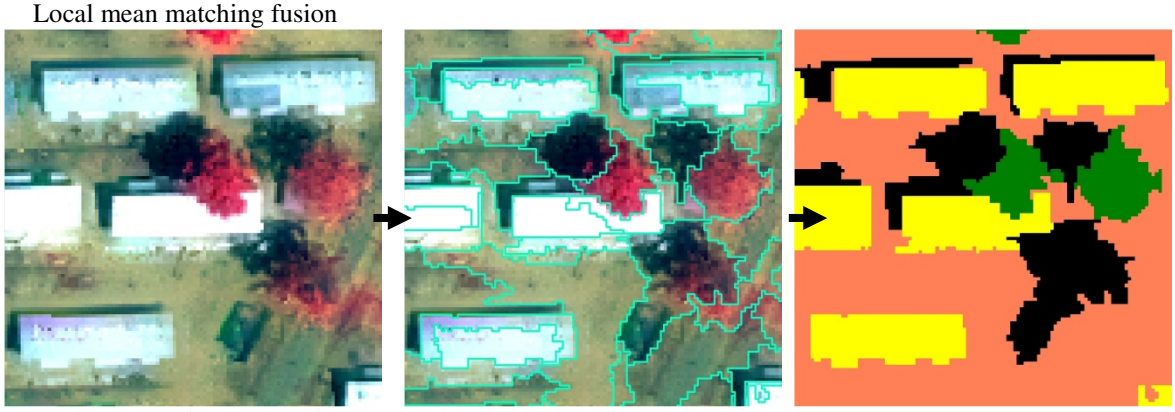
Gram-Schmidt fusion



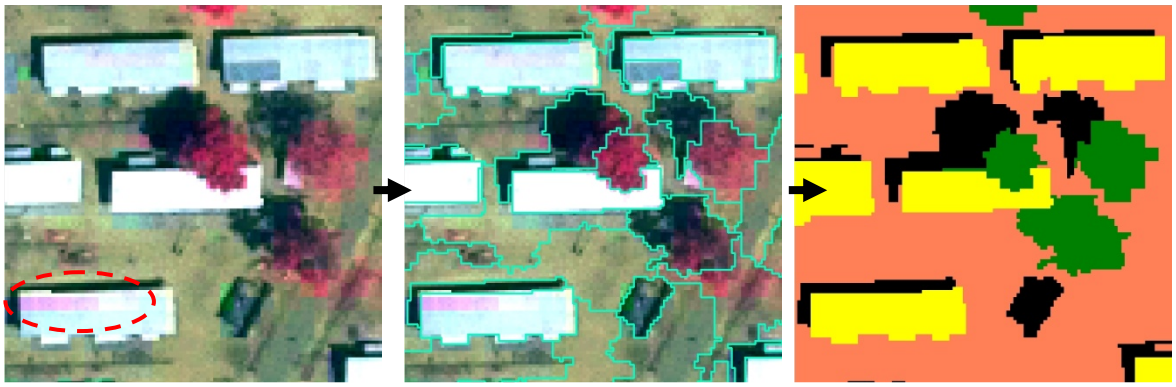
High-pass filter fusion



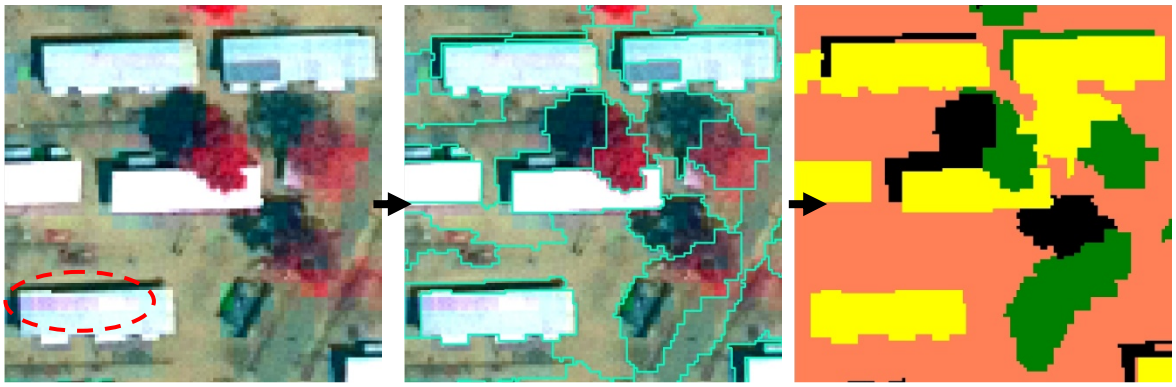
Local mean matching fusion



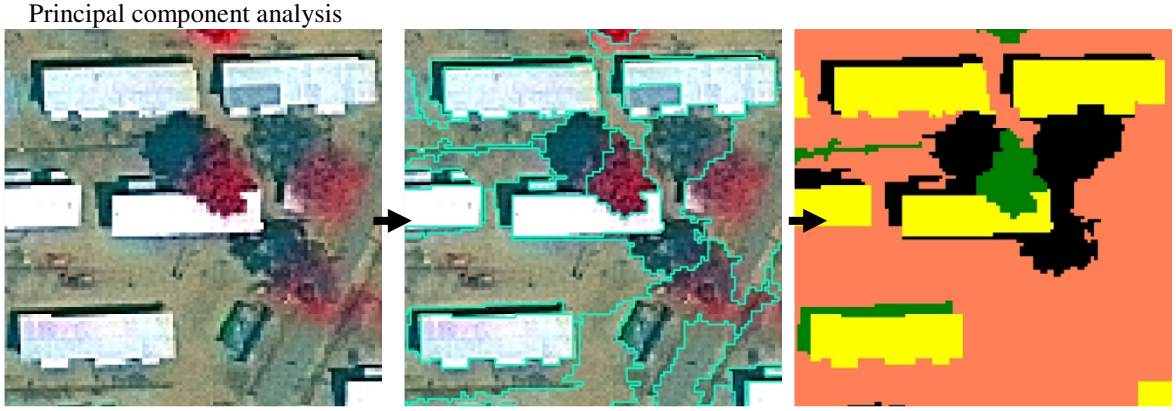
Local mean variance matching



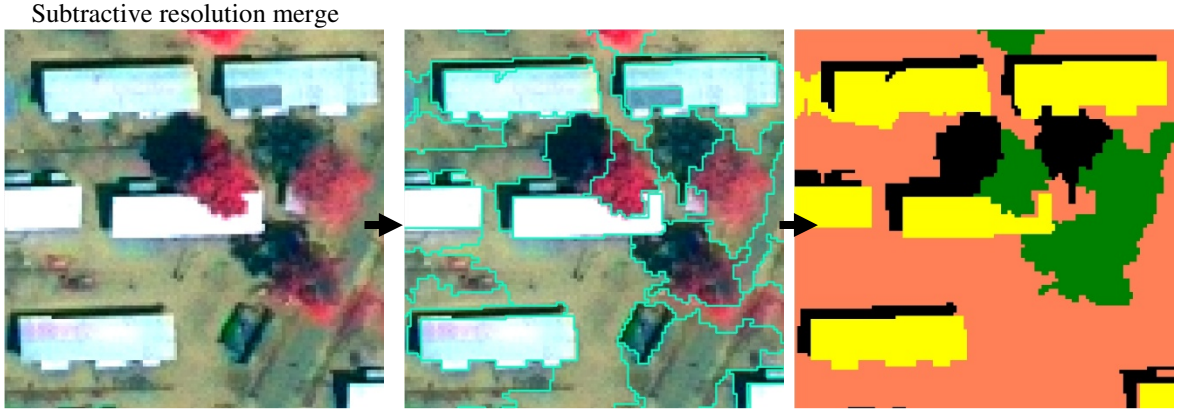
Modified-IHS fusion



Principal component analysis



Subtractive resolution merge



University of New Brunswick

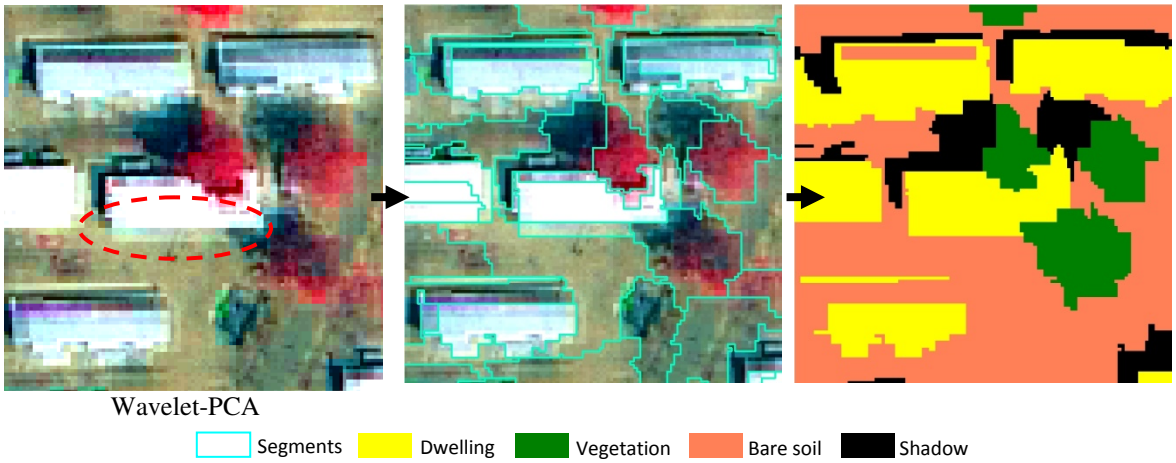
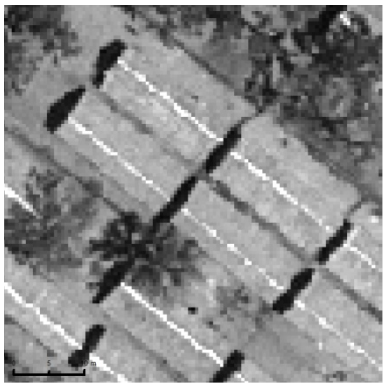
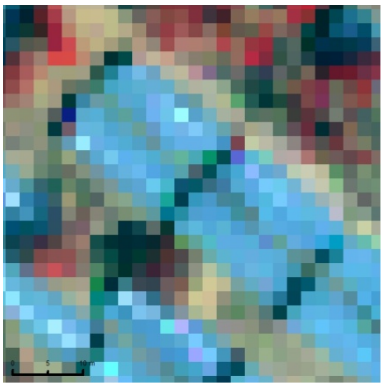


Figure 25. A representative location from a sector of the MF1 study site that encompasses bright shelters (BS). Fusion, segmentation, and classification results for each fused product are given for visual comparison.

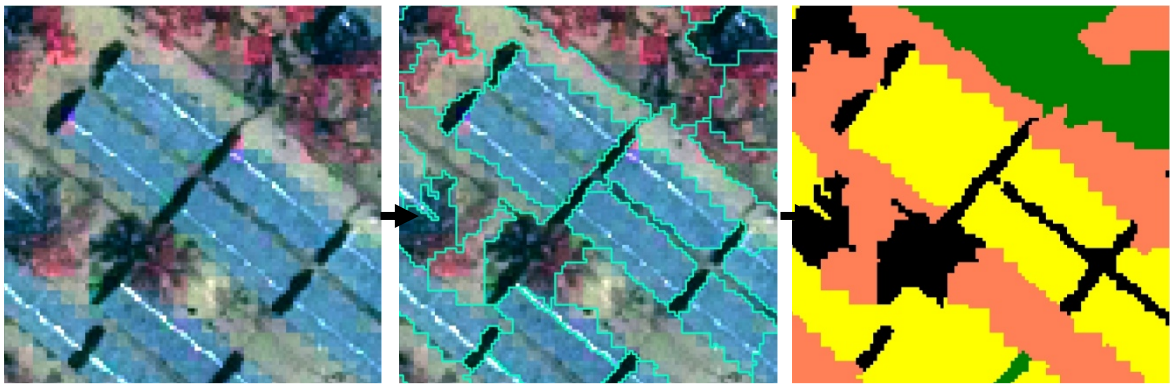


Original PAN

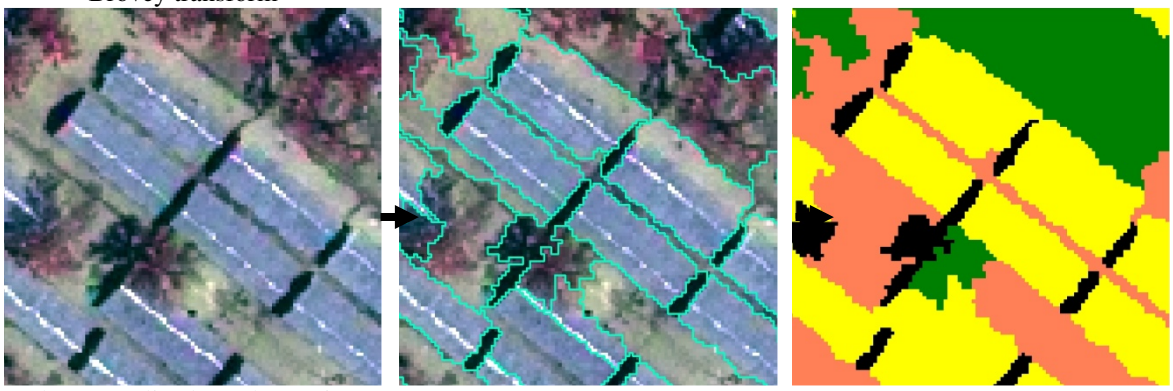


Original MS

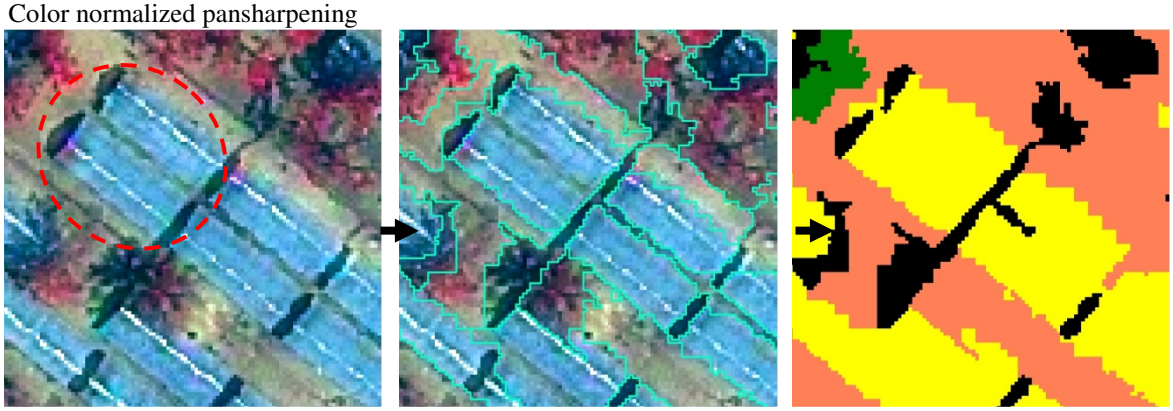
Fusion → **Segmentation** → **Classification**



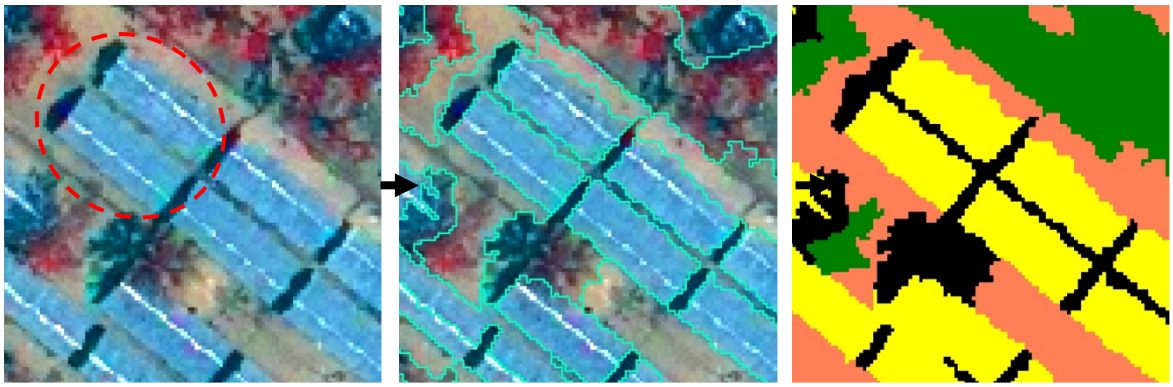
Brovey transform



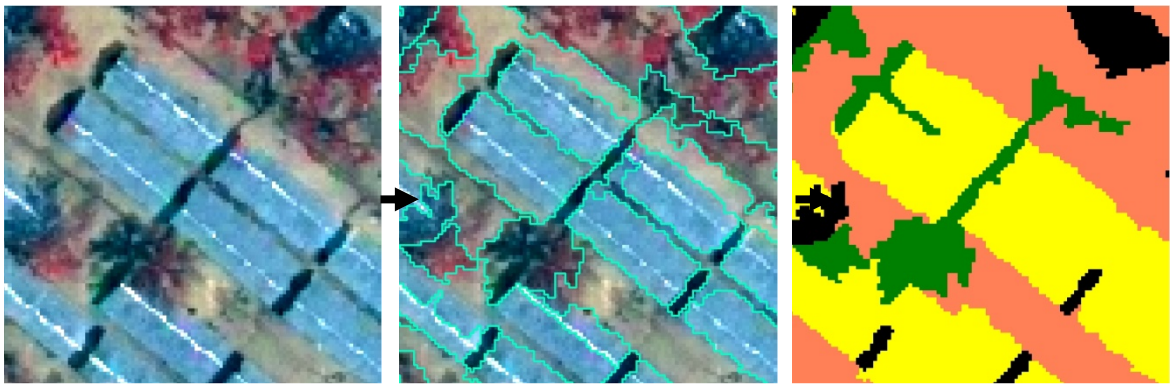
Color normalized pansharpening



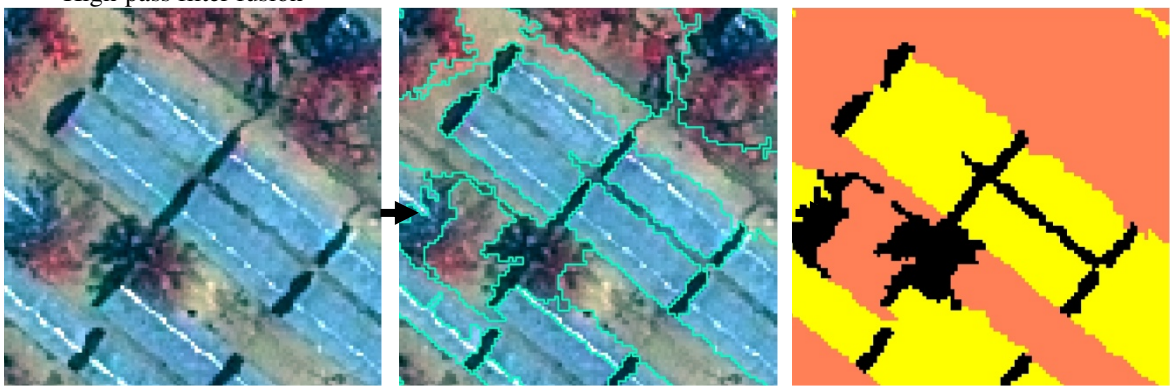
Ehlers fusion



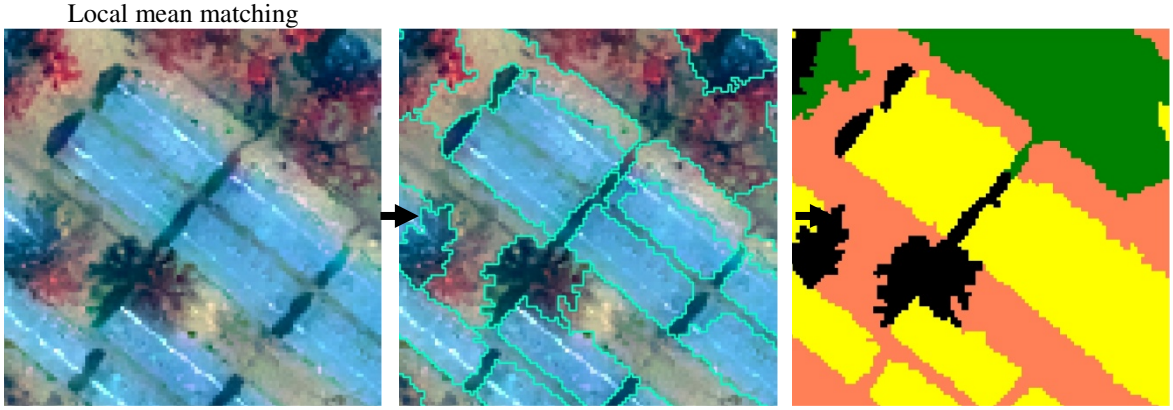
Gram-Schmidt fusion



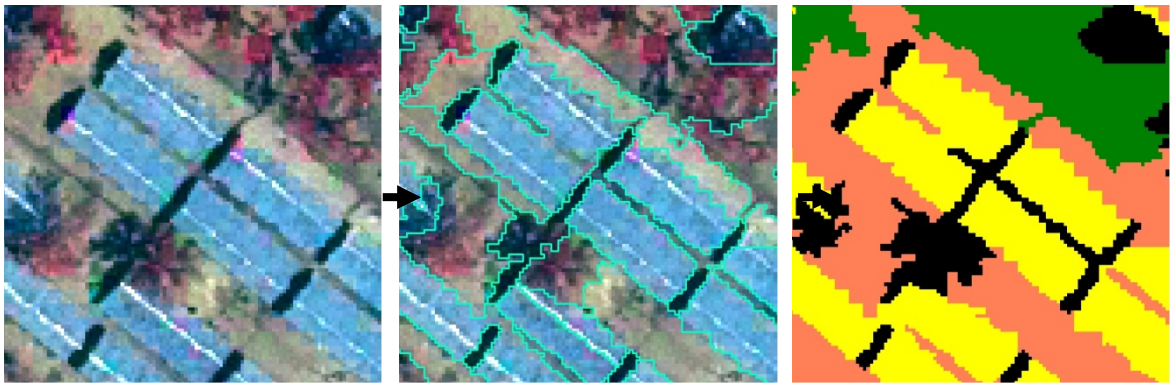
High-pass filter fusion



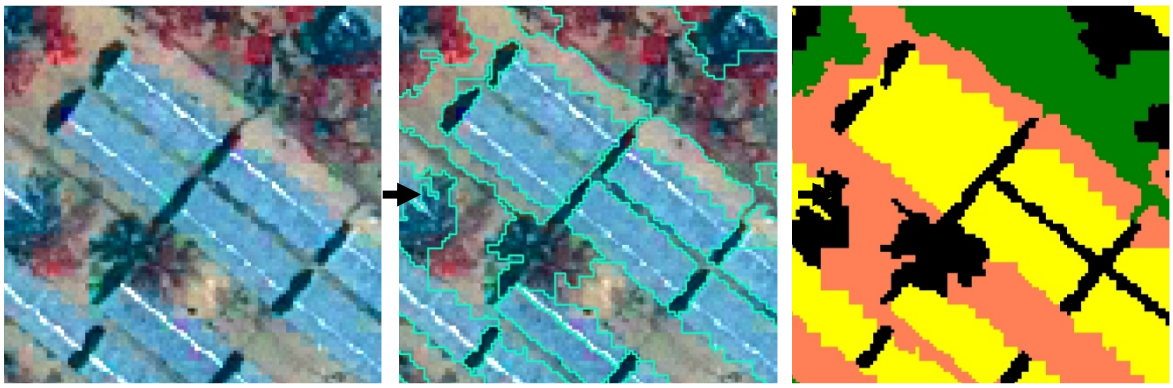
Local mean matching



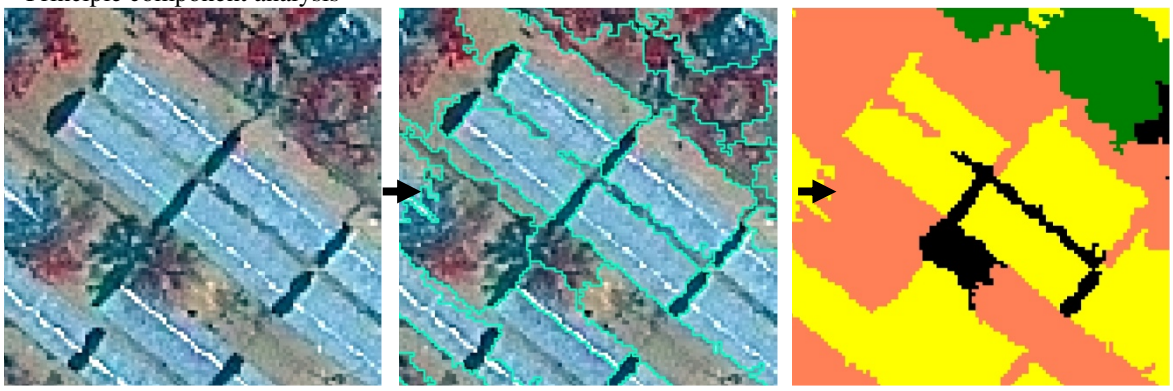
Local mean variance matching



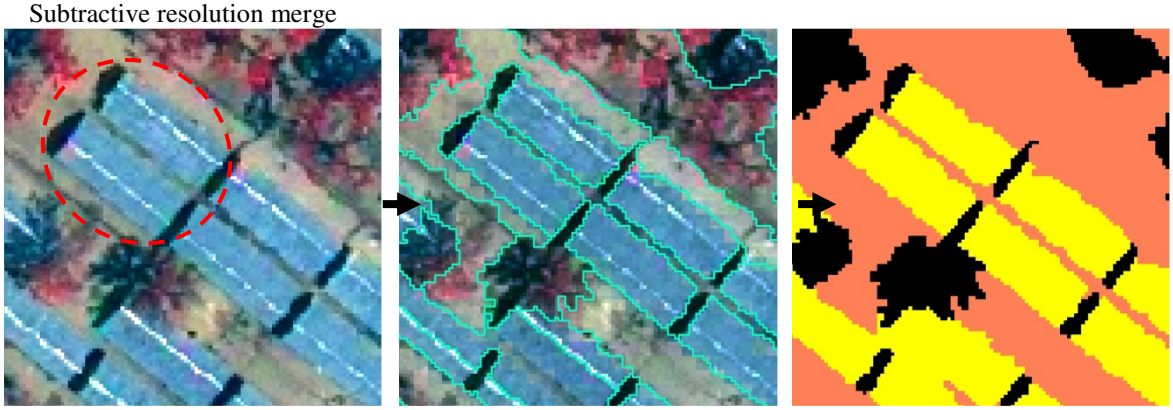
Modified-IHS



Principle component analysis



Subtractive resolution merge



University of New Brunswick

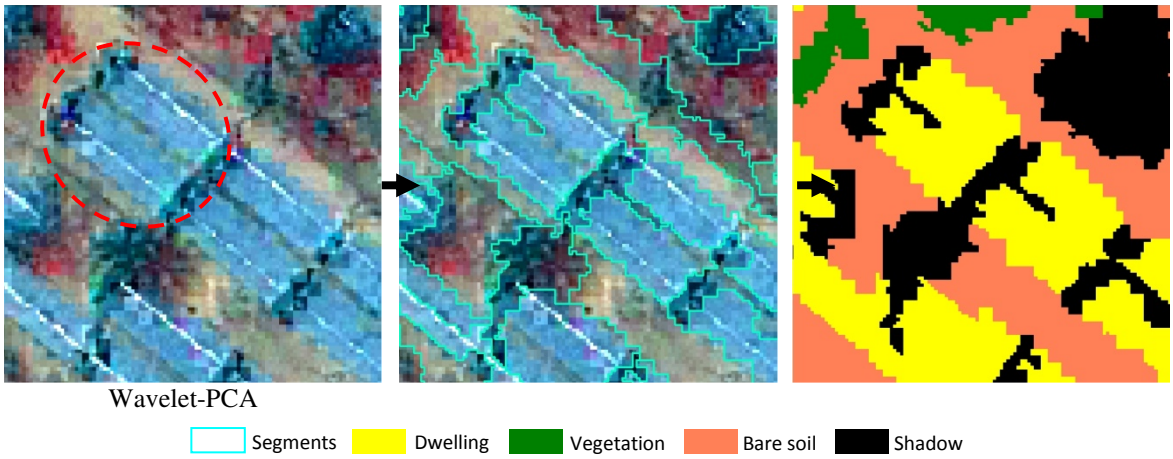


Figure 26. A representative location from a sector of the MF1 study site that encompasses dark shelters (DS). Fusion, segmentation, and classification results for each fused product are given for visual comparison.

Standard accuracy assessment methods (e.g., user's and producer's accuracy) usually rely on per-pixel validation, thus, the validity of these methods when applied to GEOBIA framework has been argued in literature. Our contention is that manual photointerpretation is the most accurate way to assess the quality of classified image objects. Thus, we performed a detailed visual inspection on the main class of interest (i.e., IDP dwelling). We randomly extracted 1000 IDP shelters and tasked two remote sensing experts to inspect the quality of classified image objects. This assessment accounted geometrical congruency (Figure 16) and arithmetic discrepancy (Figure 17) between reference objects and classified objects. We examined whether a randomly selected dwelling is individually detected in the classification (i.e., detection). If the classified object contained more than one reference dwelling (many-to-one relationship, see Figure 17), we counted such incidents as merged objects. We also counted the number of undetected reference dwellings. Figure 27 summarizes the number of reported cases for each fused products (detected, merged, and lost) as a percentage of the total reference dwellings. In multiscale information extraction, we expect an accurate detection of individual IDP shelters to estimate indirectly the IDP population. As seen on Figure 27, the GS fusion algorithm, the MIHS fusion algorithm, the PCA fusion algorithm, and the UNB fusion algorithm reported more than 75 percent of individual detection of IDP shelters. Of these, the former three achieved approximately 80 percent detection of shelters. The Wavelet-PCA fusion algorithm exhibited the lowest individual detection percentage (~ 50%). In general, the CN fusion algorithm, the Ehlers fusion algorithm, the LMVM fusion algorithm, and the SRM fusion algorithm showed mediocre detection results.

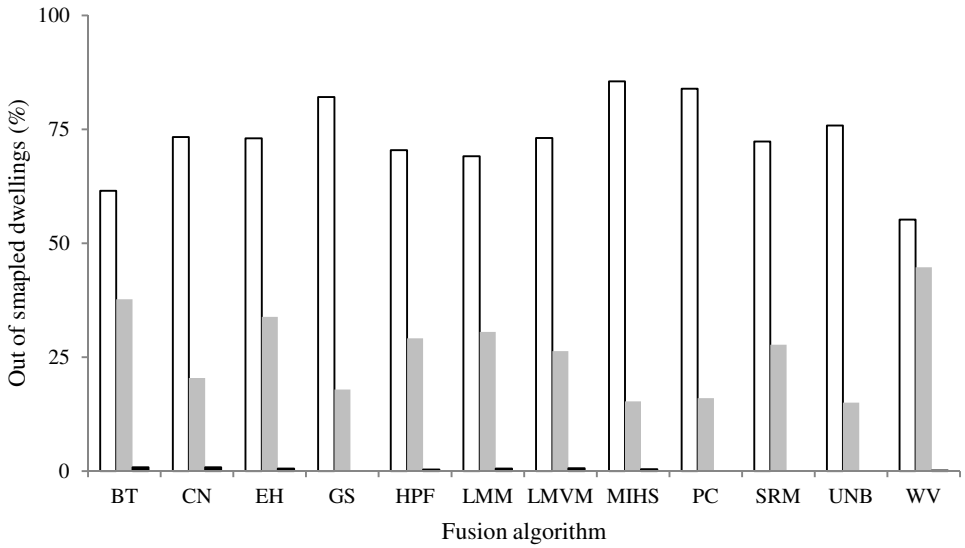


Figure 27. Summarizes the number of reported cases for each fused products (detected, merged, and lost) as a percentage of the total reference dwellings in the MF1 study site. The percentages are calculated based on a detailed visual inspection carried by two remote sensing experts with respect to a thousand randomly extracted IDP shelters.

Study site - MF2

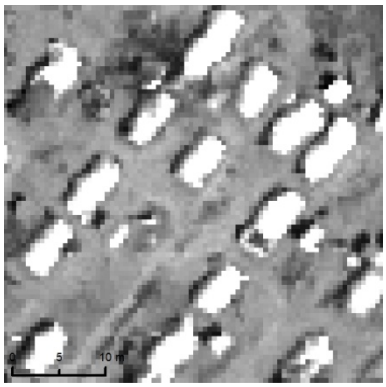
In this study site we focused on segmenting and classifying of non-fused (original PAN and MS images) and fused images. The MF2 entailed single-family hosting dwellings (tents). The main objective was to investigate the quality of image object candidates and classified objects among different fused products and non-fused image. Unlike in MF1, in MF2 we used expert knowledge to obtain optimum parameter settings of the MRS algorithm for fused and non-fused images. Here, we aimed on a feature extraction rather than a wall-to-wall classification.

For illustration purposes we selected representative areas from MF2. Figure 28 shows reference images (PAN and MS), fused images, and segmentation and classification results of fused and non-fused images. In general, the IDP tents are in varying sizes, randomly oriented, and masked by tree canopies to varying degrees.

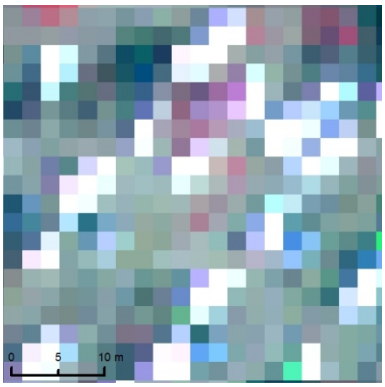
Except the Wavelet-PCA fusion algorithm, all other fusion algorithms exhibited notable improvement of spatial information in their fused products. When comparing original MS image and fused images, severe color distortions can be seen in the panshapened images from the CN fusion algorithm and the SRM fusion algorithm. Algorithms like Ehlers, LMM, and LMVM produced fuzzy appearance around IDP tents (see circled areas in Figure 28). In general, except the Wavelet-PCA fusion algorithm, remaining fused products lead to satisfactory segmentation and classifications results. In most cases, IDP tents are segmented accurately (i.e., one-to-one correspondence) and the classified individually. When comparing segmentation quality and classification results of fused images and non-fused image, it is clear that non-fused image has achieved segmentation and classification comparable (e.g., the GS fusion algorithm, the UNB fusion algorithm, and MIHS fusion algorithm) or superior (e.g., Ehlers fusion algorithm and Wavelet-PCA fusion algorithm) to those of fused images. Similar to the previous study site, we randomly extracted 1000 IDP tents and tasked two remote sensing experts to inspect the quality of classified image objects. Figure 29 reports percentages of detected, merged, and omitted IDP tents. The GS fusion algorithm, the HP fusion algorithm, the MIHS fusion algorithm, the PCA fusion algorithm, and the UNB fusion algorithm reported more than 95 percent of detection of individual IDP tents. It is interesting to note that the non-fused product achieved more than 95 % of individual detection of IDP tents. The BT fusion algorithm, the CN fusion algorithm, the Ehlers fusion algorithm, the LMM fusion algorithm, and the SRM fusion algorithm exhibited mediocre detection percentages. Compared to fused products and non-fused image, the LMVM and the Wavelet-PCA

achieved poor classification results showing low percentage of detected and a higher percentage of lost IDP tents.

As seen on Figure 22, tents are closely spaced leaving very narrow gaps between adjacent structures (see (a)) and disturbed by tree canopies (see (b)). Thus, high frequency information from the PAN image is of high value for accurately detecting individual tents. Our visual inspections revealed that a large number of tents are fully or partially located under tree canopies. Bright surfaces of these tents are partially visible through tree canopies in the PAN image but totally unidentifiable in the MS image. Figure 23 provides an ideal exemplar scenario for closely-packed and canopy-disturbed IDP tents. Outlines of the manually extracted two tents (T1, and T2) (solid red line) are overlain on the original PAN and MS images and selected fused products. Outlines of the classified image objects (solid yellow line) are also shown on corresponding non-fused and several fused images. The T1 shelter is under tree canopy and the T2 shelter is undisturbed. Two shelters are narrowly separated (~ 1m). Although the T2 tent is visible on both PAN and MS images the T1 tent is unidentifiable on the MS image. Human interpreters have the ability to detect T1 and T2 individually and digitize the outlines, especially the T1 tent based on the domain expertise. However; in automated and semi-automated IDP extraction, there is a greater possibility of detecting T1 and T2 as a single entity or leaving T1 undetected posing possible errors in IDP population estimation.

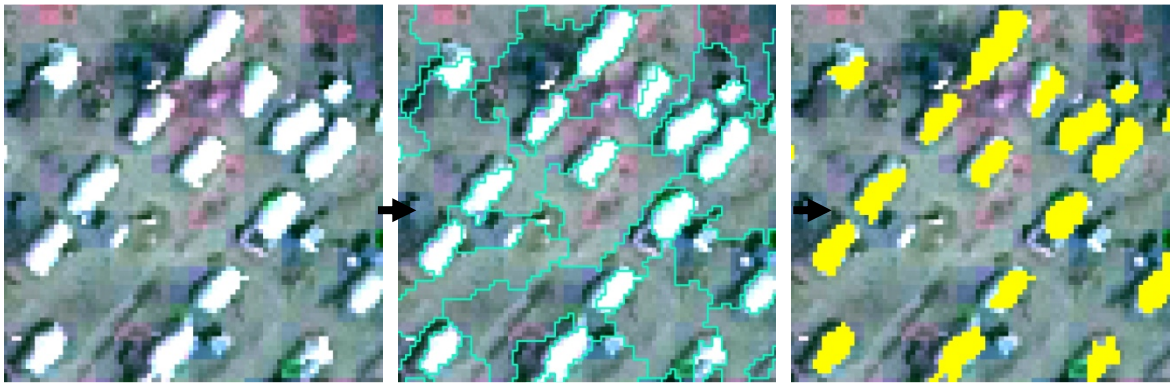


Original PAN

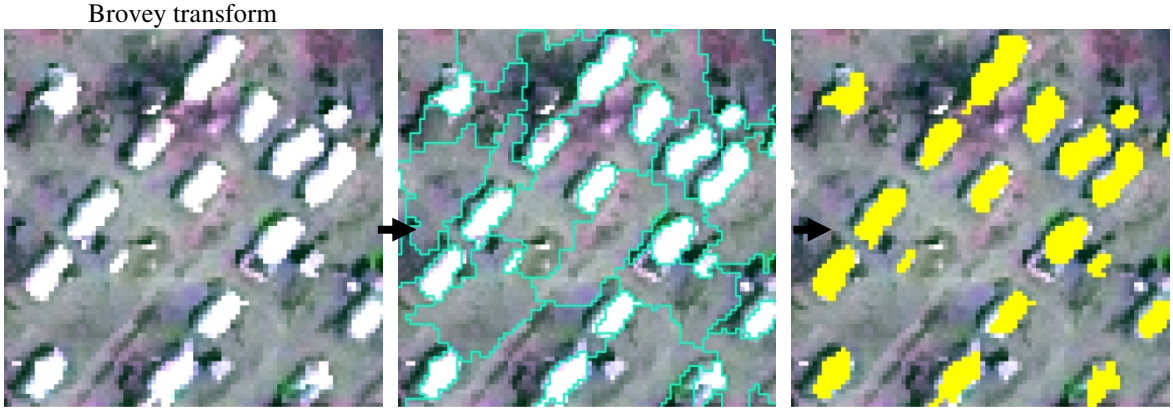


Original MS

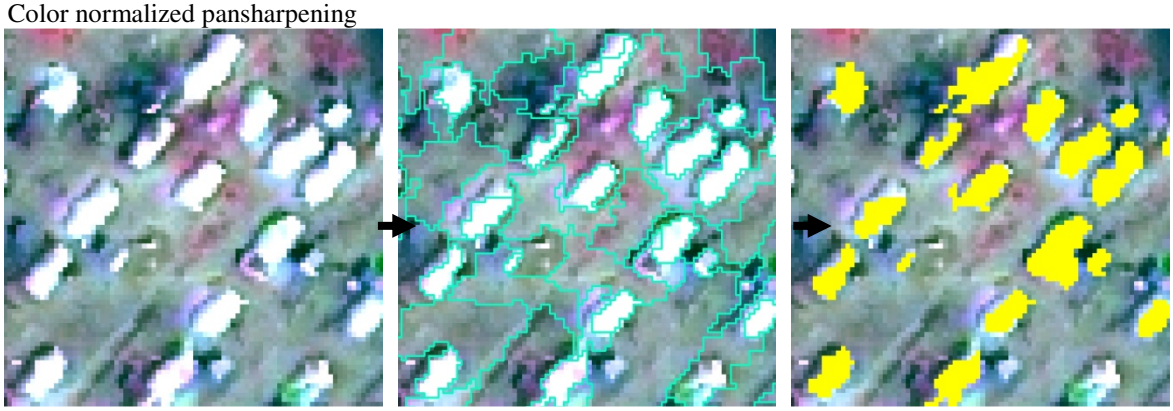
Fusion → **Segmentation** → **Classification**



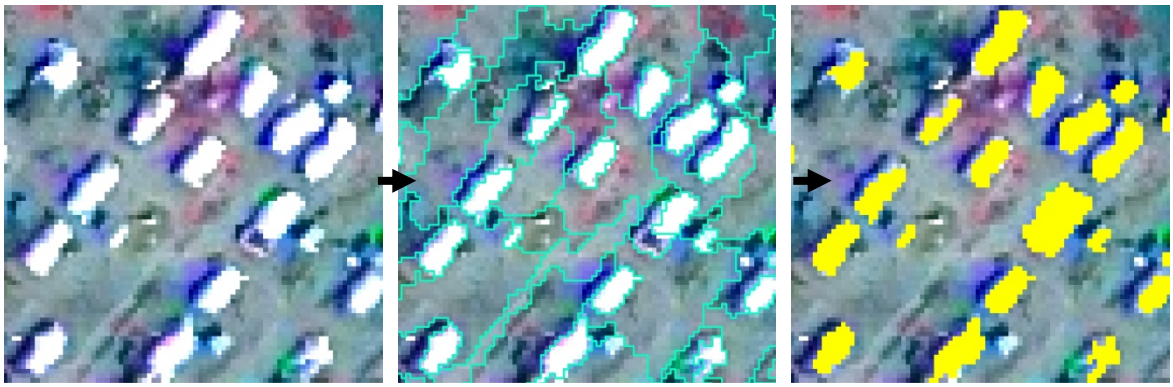
Brovey transform



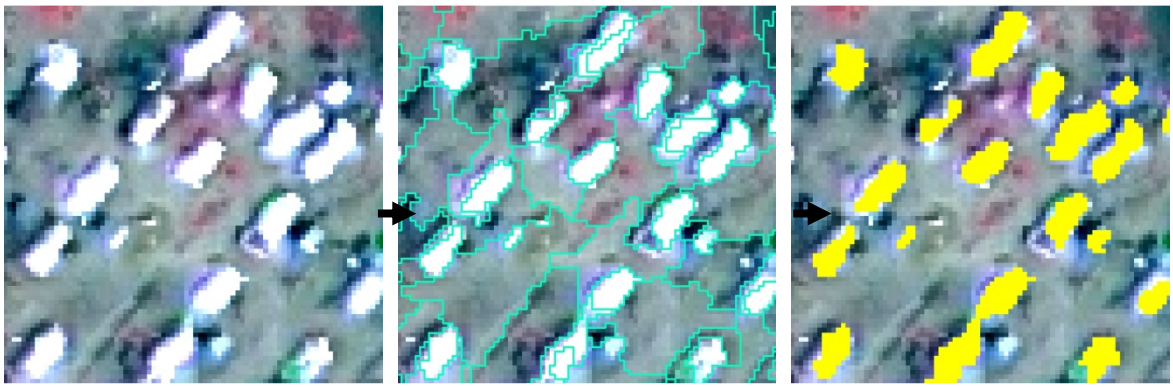
Color normalized pansharpening



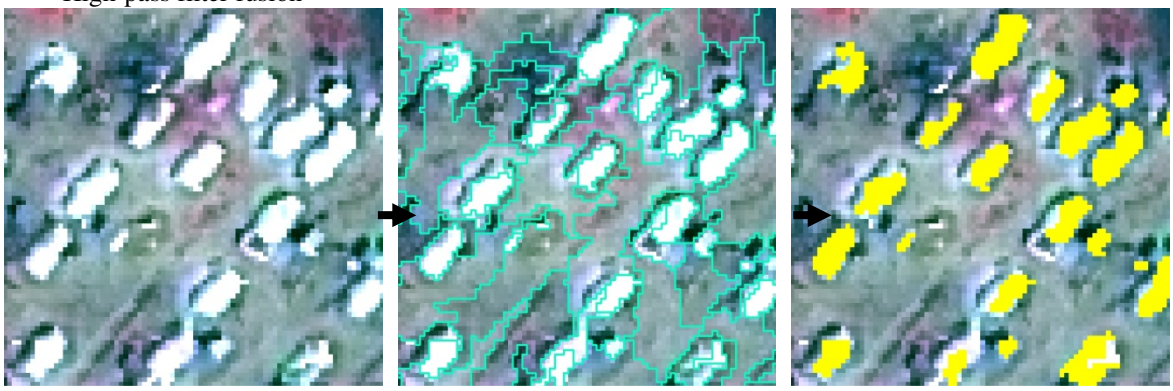
Ehler fusion



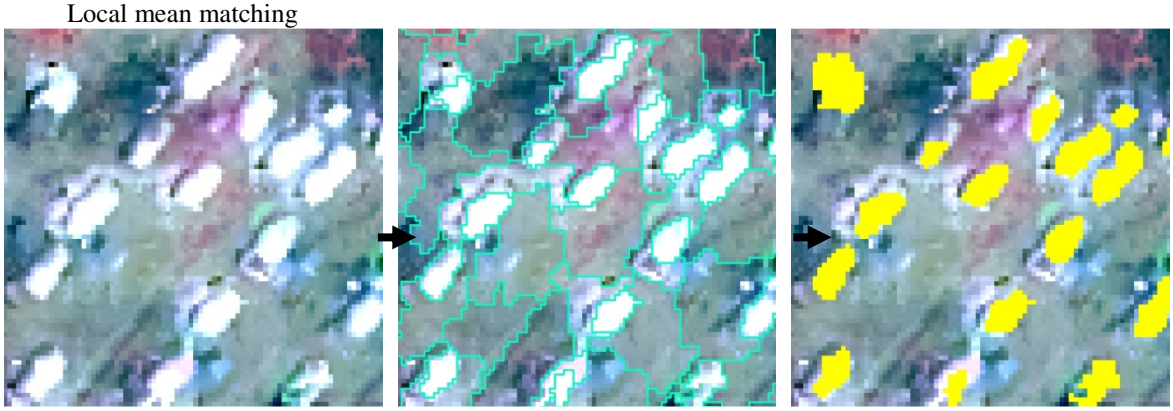
Gram-Schmidt fusion



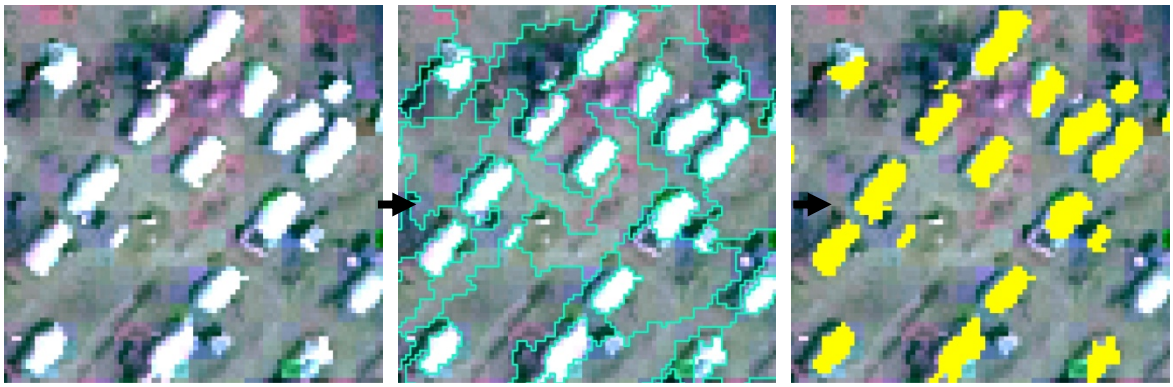
High-pass filter fusion



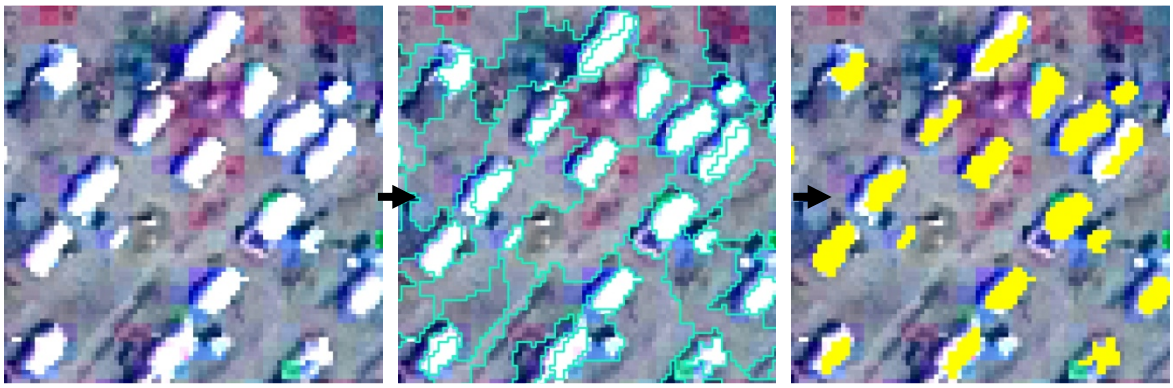
Local mean matching



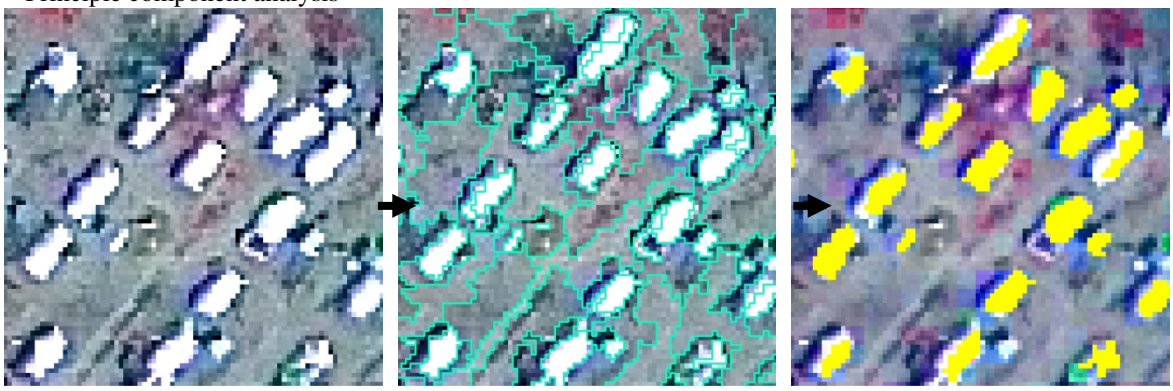
Loacl mean varaince matching



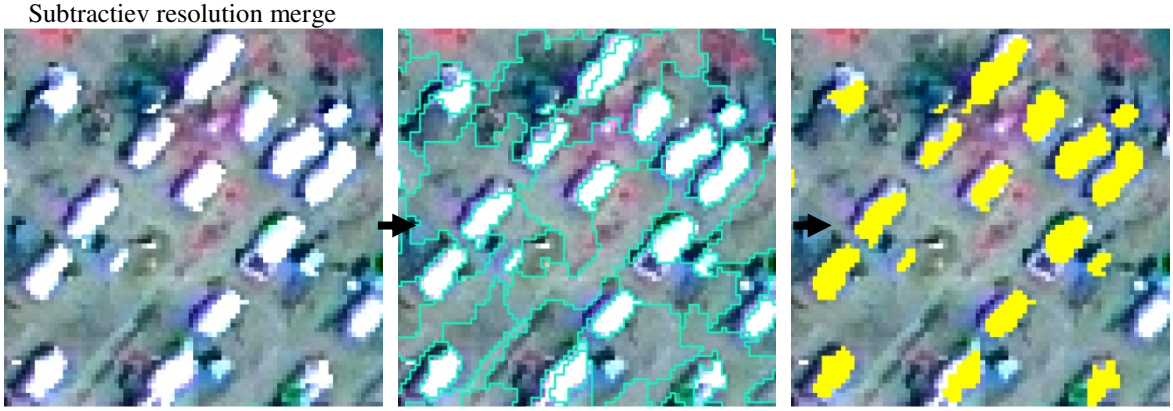
Modified-IHS



Principle component analysis



Subtractiev resolution merge



University of New Brunswick

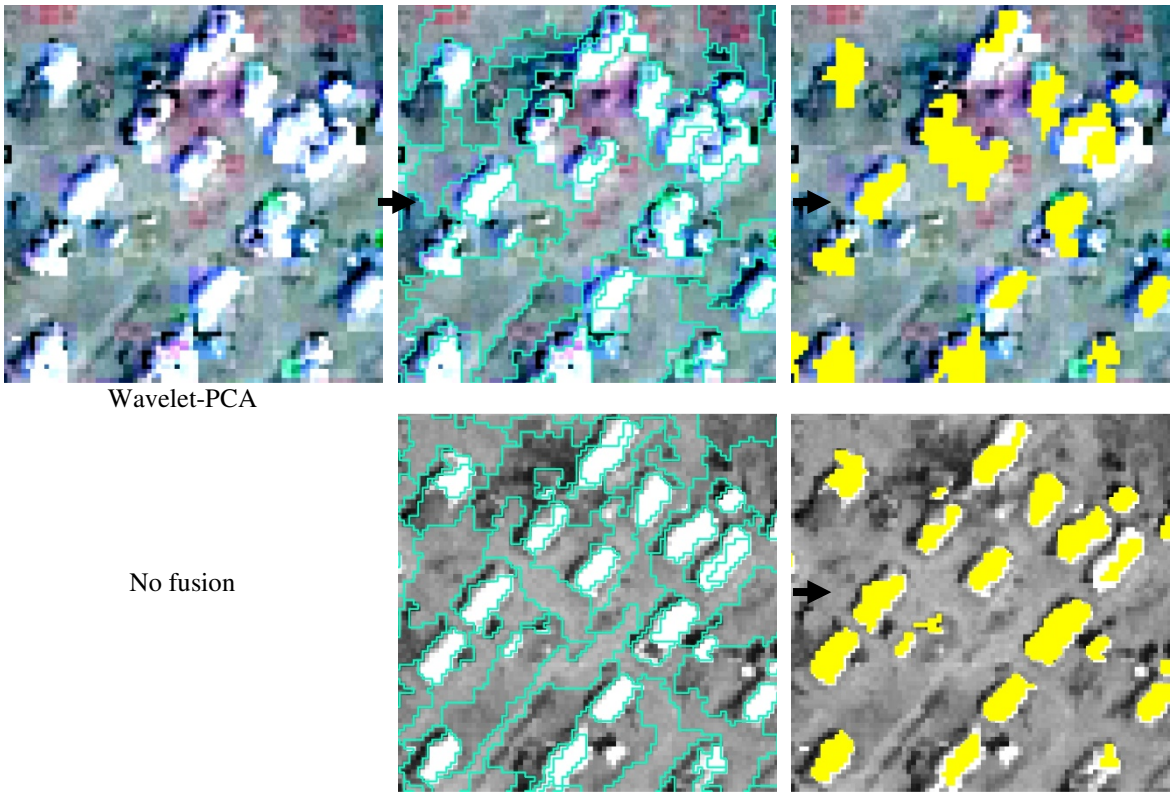


Figure 28. A representative location from the MF2 study site that encompasses IDP tents. Fusion, segmentation, and classification results for each fused product are given for visual comparison

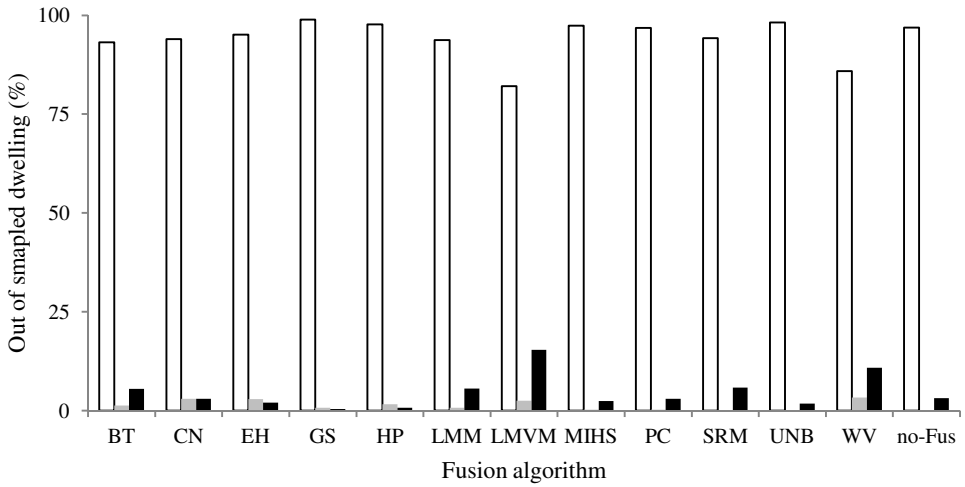
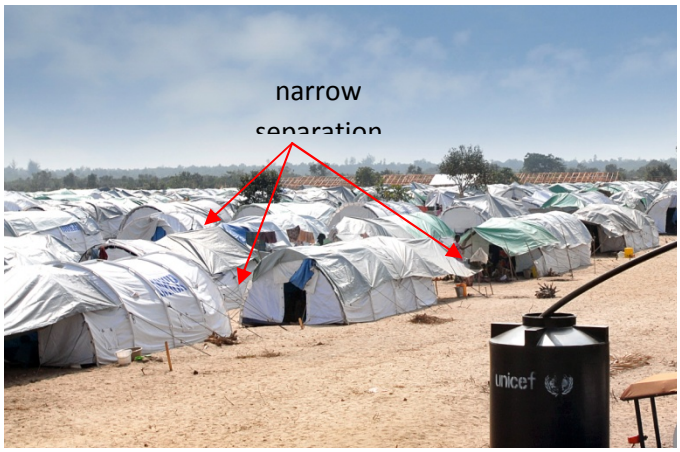
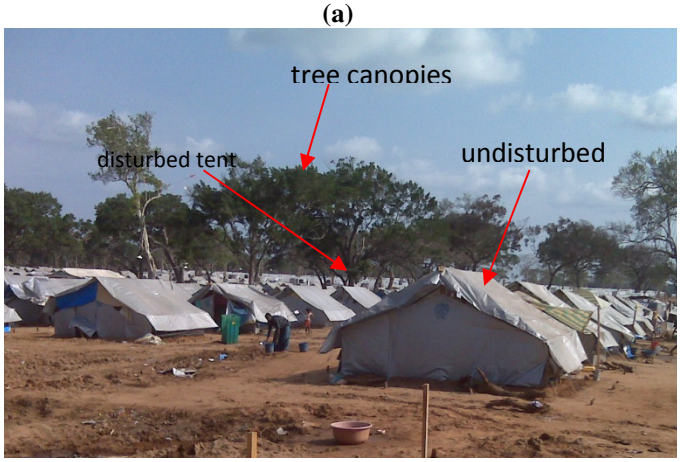


Figure 29. Summarizes the number of reported cases for each fused products (detected, merged, and lost) as a percentage of the total reference dwellings in the MF2 study site. The percentages are calculated based on a detailed visual inspection carried by two remote sensing experts with respect to a thousand randomly extracted IDP shelters

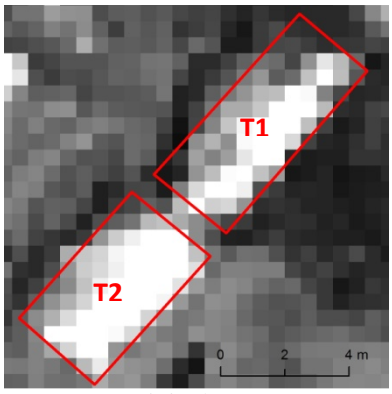


(a)

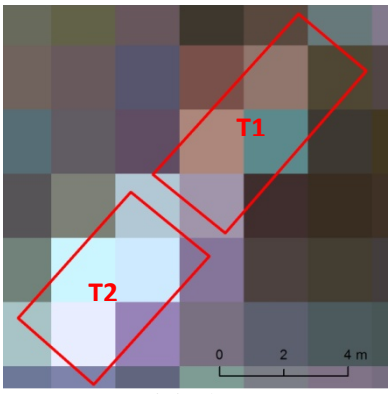


(a)

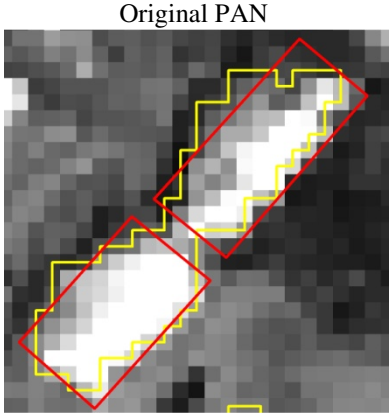
Figure 30. A Ground-based photographs from the MF2 study site showing how IDP tents are closely packed leaving very narrow spaces between adjacent structures (a) and disturbed by tree canopies (b).

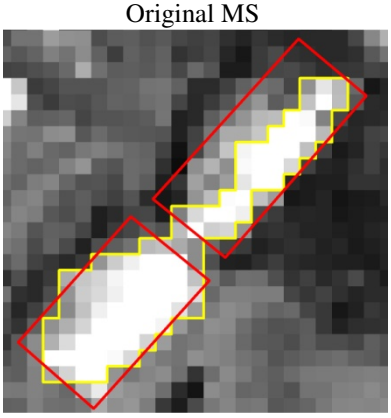


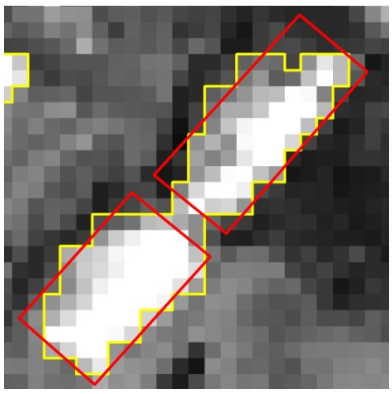
Original PAN

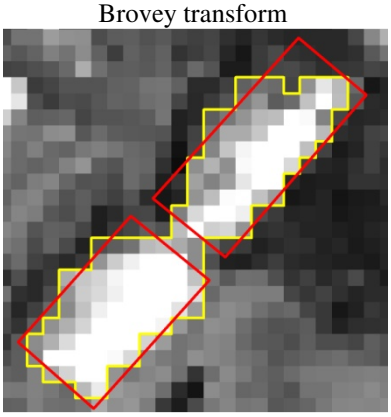


Original MS


No fusion


Brovey transform


Gram-Schmidt fusion


UNB fusion

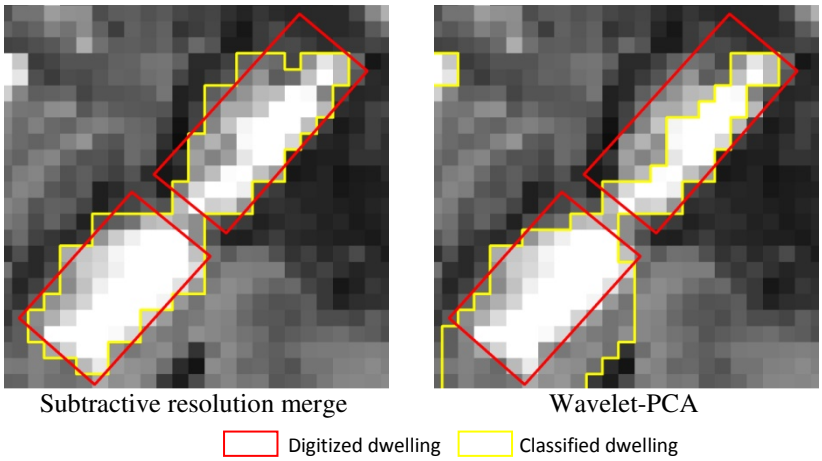


Figure 31. An ideal exemplar scenario showing a closely-packed and canopy-disturbed IDP tents as seen on VHSR images and possible errors incur during segmentation and classification steps.

3.4 Discussion

We envision this as an exploratory study pillared on multiscale internally displaced persons' camp information extraction. In ongoing- and post-crisis situations, it is critical to monitor constantly and map affected populations living in large-scale ephemeral settlements, such as IDP and refugee camps and transitional shelters sites. We used two subsets (MF1, MF2) of a GeoEye-1 image scene that encompassed a major war-induced ephemeral settlement in Sri Lanka. Our approach is centered on the geographic objects-based image analysis framework.

We applied twelve data fusion algorithms to the candidate subsets and tested their performances using a series of quality metrics. From the point of visual inspections, no single algorithm was able to produce superior results by simultaneously preserving spectral and spatial properties of the original MS and PAN images. Algorithms like HPF, Gram-Schmidt, and UNB exhibited mediocre fusion results with respect to color similarity and spatial improvement. Taking objective spectral quality evaluation (Table 9) into account, it is important to examine how spectrally-superior algorithms (Ehlers fusion, LMVM, and Wavelet-PCA algorithms) behave across eleven quality metrics. With respect to two global indicators (ERGAS and SAM) and other band-wise metrics in our quality budget, the Ehlers fusion produced better-quality results. Our results support the repeated claims of Ehlers and his colleagues (Klonus and Ehlers 2007; Ehlers et al. 2010; Ling et al. 2007) on the spectral quality preservation character of the Ehlers fusion algorithm. Regarding spatial quality assessment, despite the superior performances with respect to spectral similarity, the Ehlers, wavelet-PCA, and LMVM algorithms exhibited poor spatial improvement. Among this subset, the Wavelet-PCA algorithm showed the

worst scores consistently for all spatial indices. The CN sharpening algorithm, Brovey transform algorithm, and subtractive resolution merge algorithm achieved notably better scores for spatial indices than for spectral indices. These observations emphasize the necessity of a combined approach (i.e., spectral and spatial fidelity) for benchmarking fusion results because the best color preservation of an algorithm can be observed even if no pansharpening is performed; on the other hand, a fusion algorithm can achieve the best spatial improvement while producing results with poor color preservation. The UNB algorithm showed superior results consistently for all spatial metrics across all study sites while Gram-Schmidt produced mediocre spatial improvement.

The intriguing question is "do fused images maintain their benchmarks achieved with respect to fusion quality in the segmentation results?". In general, one could argue in three different ways that the fused product with the highest 1) spectral fidelity, 2) spatial fidelity, 3) and spectral and spatial fidelity lead to better segmentation results. For example, if the first case is considered, the Ehlers, the Wavelet-PCA, and the LMVM should produce good image object candidates. Segmentation algorithms are scene-dependent. Thus, when investigating the synergies of data fusion and segmentation, it is necessary to select a benchmarking segmentation algorithm to minimize the plurality of solutions. Although, many image segmentation algorithms have been built into to commercial software packages, based on the highlighted success rate, we utilized eCognition developer's multiresolution segmentation algorithm to segment twelve pansharpened images of the MF1 study site at varying scale, shape, and compactness setting. Like segmentation itself there is no universal way of assessing segmentation quality. Quality metrics have different design goals addressing different aspects of

segmentation, thus, we used three empirical discrepancy measures to assess the segmentation quality of fused products. By doing this we aimed to achieve multi-dimensional assessment on the quality of image object candidates. Although we employed a single segmentation algorithm, there is no guarantee that all pansharpened images (even though they represent the same real-world scene) yield best segmentation results at the same parameter setting of the multiresolution segmentation algorithm. There is a plethora of literature on image segmentation and quality assessment; however, the importance of the statistical analysis on the quality of image objects has been overlooked. We used non-parametric statistical techniques because, in general, area-based metrics are not normally distributed. Therefore, comparing sample medians of quality metric values are more meaningful than sample means. When comparing the segmentation quality of individual fused images with respect to QR, ED1, and ED2, only the scale and shape parameters significantly affected on the quality of image objects. This reveals that the compactness parameter has a less influence on the quality of image objects. With this observation we would like to make a note on the usage of the MRS algorithm and perhaps it might be valid for other segmentation algorithm as well. Due to the success of the MRS algorithm and its limited control available to user, several studies (e.g, Dragut et al. 2010; Liu et al. 2012; Tong et al. 2012) have aimed to develop methods for finding optimum segmentation parameter combinations. We undoubtedly accept the importance of the scale parameter in achieving good segmentation results. However, we question the real value of fine tuning the shape and compactness parameters. Unlike the scale setting, with a parameter space of zero to infinity (theoretically), the shape and compactness occupy a small parameter space (0 -1). The most important question is that at the

optimum scale setting does the MRS algorithm produce significantly different image objects, if the shape or the compactness parameter is changed slightly. For example, at the optimum scale (s =50), is the quality of image objects derived at shape = 0.7 and compactness = 0.3 significantly different from those achieved at shape = 0.7 and compactness = 0.4?. A detailed discussion on finding MRS algorithm's optimum parameter settings is beyond the scope of this study; however, it would be interesting to perform a sensitivity analysis to unravel the significance of finding optimum parameter settings.

With respect to each quality metric (QR, ED1, and ED2), we compared the segmentation quality (quality of image object candidates) among twelve fused images. Based on fusion evaluation results, we would expect a significant difference in the segmentation results among fused images. In visual inspections, we observed a notable segmentation quality variation among fused images. For example, spectrally superior algorithms like the Ehlers fusion and Wavelet-PCA fusion produced poor segmentation results compared to spectrally inferior algorithms like the Brovey transform and CN pansharpening. Interestingly, our multidimensional assessment (i.e., QR, ED1, and ED2) reported that image objects of twelve pansharpened products are not statistically different. In most image segmentation studies, conclusions are drawn based only the quality of raw image object candidates and overlook the quality of classification (classified image objects).This is mainly due the common understanding that good segmentation leads to better classification results. However; our contention is that even a satisfactory segmentation could lead to an excellent classification because we have many opportunities to refine initial image objects in an evolutionary manner. On the other hand,

the definition of a good segmentation largely depends on the design goals of the quality metric. For example, when finding optimum segmentation settings for the pansharpened product of the BT fusion algorithm, the QR metric proposed a scale of 10 as the optimum scale for obtaining a good segmentation, whereas the ED2 proposed a scale of 40.

Although the analysis of the discriminative capacity of empirical discrepancy measures is beyond the main scope of this study, our results provide useful information on the behavior of QR, ED1, and ED2 quality metrics. A given segmentation quality measure is required to be sensitive the quality change of image object candidates, i.e., the metric should yield increase or decrease assessment scores with respect to increasing or decreasing quality of image object candidates. In this respect, if a metric is highly sensitive (according to statistical assessment results) to quality change, it can be realized as strong metric with high discriminative capacity. When considering segmentation quality variation (QR, ED1, and ED2) relative to increasing scale (10-100 at 10 unit steps) for all fused images (see Figures 18,19, and 20), visually, three metrics seemed to be sensitive to the changing quality of image objects. However, statistical results (see Tables 15.1-15.12) revealed that QR and ED1 consistently produced significantly different assessment scores (i.e., high sensitivity for quality change) when the scale parameter is changed by 10 units. In contrast, the ED2 metric clearly showed a poor sensitivity for the segmentation quality change. Our understanding is that despite of less sensitivity, the ED2 metric is more meaningful than the other two metric because it accounts for arithmetic discrepancy between image object candidates and reference objects. Many quality metrics have been developed and applied for assessing segmentation quality, however, the sensitivity of these metrics has been poorly addressed.

Thus, we emphasis the necessity of comparing widely-used segmentation quality indices and identify the most sensitive ones.

Because we aimed to compare the classification accuracy among twelve fused products, we relied on a classifier-based (SVM classifier) approach other than a knowledge-based approach (rule-based classification) to maintain the consistency of classification. It is interesting to see the connection between the classification accuracy and quantitative fusion evaluation results. When taking into account the quantitative fusion evaluation results (Table 9), we would expect the Ehlers fusion algorithm (also Wavelet-PCA and LMVM) to exhibit notably high values for the classification accuracy indicators. However, as seen in Table 19, the GS fusion algorithm reported the highest overall accuracy and Kappa. Further, with respect to the IDP dwelling class, the GS fusion, the UNB fusion, the PCA fusion, and the MIHS algorithms achieved the best values for user's and producer's accuracy and kappa. Standard accuracy assessment techniques are pixel-based, in fact, the validity of these quality indicators are frequently challenged in the GEOBIA framework. Thus, we aimed to support further the quantitative classification accuracy assessment results by conducting a detailed visual inspection. Our contention is that human interpreters are more capable of (i.e., superior to quantitative metrics) benchmarking image objects in multidimensional aspects (e.g., detection, geometric fit, arithmetic fit, shape, and thematic accuracy). Similar to previous case, if fusion quality is only considered (see Table 9), the Ehlers fusion algorithm (also Wavelet-PCA and LMVM) should lead to good classified image objects and it should be manifested in our visual assessment (see Figures 25 and 26). It should also be noted that exhibiting thebest spectral fidelity (e.g., Wavelet-PCA, Table 5) does not necessarily

mean that the fusion algorithm has made spatial improvements, i.e., a fusion algorithm can exhibit a high degree of spectral preservation even if no pansharpening is performed.

In the MF1 study site we followed a systematic approach to experiment the effect of fusion in segmentation and classification steps. In contrast, the MF2 site simulates a more realistic crisis information extraction scenario where the domain expertise is highlighted in segmentation (i.e., trial-and-error approach for finding optimum setting of MRS algorithm and classification (i.e., rule-based classification supplemented with the class modeling approach (Lang et al. 2010) steps. In light of GEOBIA framework, there is a growing tendency of introducing original both the PAN and MS images (with fusion) in to the MRS segmentation algorithm. For example, Tiede et al., 2011 bypassed pre-processing steps (e.g., image fusion and geometric correction) in the context of rapid-mapping workflows. We do not know of any other study that has compared the segmentation and classification results of non-fused images against a large sample of fused products. It is interesting to compare fusion evaluation results (Table 9) with segmentation and classification results (Figure 28 and 29). Algorithms like Wavelet-PCA and LMVM reported poor detection compared other fused products and non-fused image. The most highlighted feature is that the non-fused image high produced quality classified image objects comparable to those of GS fusion and the UNB fusion algorithms. The MF2 study site has closely-packed small tents and frequently disturbed with tree canopies (see Figures 30 and 31). When estimating IDP population based indirect measures like shelter count, undetected shelters could lead to under estimation of IDP families and individuals. In this respect it is necessary to take the full advantage of the PAN image's high frequency information. The stressing question is that is it necessary to invest time

for pansharpening when the non-fused images provide better segmentation results than fused products. For example, the Wavelet-PCA's segmentation and classification were clearly inferior to those of non-fused image.

3.5 Conclusion

In humanitarian emergencies the timeliness of data provision and the short time window available for dispatching value-added information pose major challenges to the mapping community. To satisfy the humanitarian information demand in ongoing- and post-crisis situations, earth observation data must be streamed through time-critical workflows. We aimed to unravel the synergies of data fusion and image segmentation in the context of EO-based rapid mapping workflows. In light of geographic object-based image analysis framework, our findings have challenged the well established view of data fusion community that "any error in the synthesis of the spectral signatures at the highest spatial resolution incurs an error in the decision". We have shown that the GEOBIA framework has the ability to create meaningful image objects during the segmentation process by compensating the fused image's spectral distortions with the high-frequency information content that has been injected during fusion. We further questioned the necessity of the data fusion step in rapid mapping context. Bypassing time-intense data fusion steps helps to intensify EO-based rapid mapping workflows. However, we emphasize the fact that data fusion is not limited to VHSR image data but expands over many different combinations of multi-date, multi-sensor EO-data. There might be certain fusion scenarios where pansharpening is really necessary and high-quality fused products are expected. In future research we will investigate further the synergies of data fusion and image segmentation with respect to multi-date, multi-sensor fusion scenarios and extrapolate our findings to other remote sensing application domains beyond EO-based crisis information retrieval.

CHAPTER 4

Optimizing multi-resolution segmentation scale using empirical methods: Exploring the sensitivity of a supervised discrepancy measure

Abstract

Multiresolution segmentation (MRS) has proven to be one of the most successful image segmentation algorithms in the geographic object-based image analysis (GEOBIA) framework. This algorithm is relatively complex and user-dependent; scale, shape, and compactness are the main parameters available to users for controlling the algorithm. Plurality of segmentation results is common because each parameter may take a range of values within its parameter space or different combinations of values among parameters. Finding optimal parameter values through a trial-and-error process is commonly practiced at the expense of time and labor, thus, several alternative supervised and unsupervised methods for supervised automatic parameter setting have been proposed and tested. In the case of supervised empirical assessments, discrepancy measures are employed for computing measures of dissimilarity between a reference polygon and an image object candidate. Evidently the reliability of the optimal-parameter prediction heavily relies on the sensitivity of the segmentation quality metric. The idea behind pursuing optimal parameter setting is that, for instance, a given scale setting provides image object candidates different from the other scale setting; thus, by design the supervised quality metric should capture this difference. In this exploratory study, we selected the Euclidean distance 2 (ED2) metric, a recently proposed supervised metric, whose main design goal is to optimize the geometrical discrepancy (potential segmentation error (PSE)) and arithmetic discrepancy between image objects and reference polygons (number-of segmentation ratio (NSR)) in two dimensional Euclidean space, as a candidate to investigate the validity and efficacy of empirical discrepancy measures for finding the optimal scale parameter setting of the MRS algorithm. We chose test

image scenes from three different space-borne sensors with varying spatial resolutions and scene contents and systematically segmented them using the MRS algorithm at a series of parameter settings. The discriminative capacity of the ED2 metric across different scales groups was tested using non-parametric statistical methods. Our results showed that the ED2 metric significantly discriminates the quality of image object candidates at smaller scale values but it loses the sensitivity at larger scale values. This questions the meaningfulness of the ED2 metric in the MRS algorithm's parameter optimization. Our contention is that the ED2 metric provides some notion of the optimal scale parameter at the expense of time. In this respect, especially in operational-level image processing, it is worth to re-think the trade-off between execution time of the processor-intensive MRS algorithm at series of parameter settings targeting a less-sensitive quality metric and an expert-lead trial-and-error approach.

4.1 Introduction

Modern satellite sensor technology provides remote sensing imagery whose spatial resolution rivals aerial images (Blaschke, 2010; Dey et al., 2010). Satellite sensors like IKONOS, QuickBird, GeoEye-1, and WorldView-2 provide very high spatial resolution (VHSR) multispectral (MS) imagery at sub-meter level that can capture the fine details needed (e.g. city-block to individual house, forest stand to single tree crown or an internally displaced persons (IDP) camp to an individual shelter) for a spectrum of application domains ranging from land use/cover mapping (Song et al. 2010; Kim et al. 2011; Lu et al., 2011; Pu et al., 2011; Ardila et al., 2012; Huang and Jia 2012; Li Saho, 2012; Pinho et al., 2012) to humanitarian crisis information retrieval (Giada et al., 2003; Al-Khudairy et al., 2005; 2007; Vu et al., 2009; Lang et al., 2010; Tiede et al, 2011; Voigt et al., 2011; Hangenlocher, et al. 2012; Kim et al . 2012, Witharana et al. 2013).

Despite the prolific advantageous features of commercial satellite sensors, dispatching time-critical conditioned geoinformation (Lang, 2008) from VHSR satellite image data to meet ever-increasing user demand still poses a main challenge to the remote sensing community.

During the past decade, the geographic object-based image analysis (GEOBIA) framework, a novel conceptualization of image understating that attempts to mimics innate cognition abilities of humans (Hay et al. 2005; Hay et al 2008; Blaschke, 2008; Marcal et al. 2008; Blaschke 2010; Marpu et al. 2010), has widely been acknowledged in multi-faceted remote sensing applications owing to the conspicuous shortfalls of pixel-based spectral data alone model when applied to the VHSR imagery (Jyothi et al. 2008; Zhou and Troy, 2008; Blaschke 2010; Smith and Morton, 2010; Kim et al., 2011). GEOBIA means more than sequential feature extraction (Lang et al. 2008). By design, it is a cyclic, adaptive, and evolutionary expert-lead workflow (Tiede et al., 2010). It provides a cohesive methodological framework for machine-based characterization and classification of spatially-relevant, real-world entities by using multiscale regionalization techniques augmented with nested representations and rule-based classifiers (Hay and Castilla 2008; Lang et al. 2008; Hangenlocher, et al. 2012). Among other reasons, the paradigm shift from the conventional pixel-based model to the GEOBIA framework has further been catalyzed by the emergence of the first commercially-available, feature rich geographic object-based classification software package - eCognition® (now known as eCognition Developer ®(Munich, Germany)). Several other GEOBIA software contenders (e.g., Feature Analyst, FeatureObjex, ERDAS Objective) have also been introduced into the market owing to the proven success of image segmenation and object-

ased classificastion; however, detailed studies (Neubert et al., 2008; Blaschke, 2010, Tong et al., 2012) report that its dominance in remote sensing applications.

Image segmentation (regionalization), a process of partitioning of a complex image-scene into non-overlapping homogeneous regions with nested and scaled representations in scene space (Lang, 2008), is the fundamental step within the geographic object-based information retrieval chain (Pham, 2001; Schiewe 2002; Blaschke and Strobl 2001; Costa et al. 2008; Dey et al. 2010; Kim et al., 2011; Liu et al. 2012; Tong et al. 2012; Duro et al., 2012). This step is decisive because the resulting image segments - image object candidates (Blaschke 2010) - form the basis for the subsequent classification, which is based on spectral, form, topological, and semantic features (Burnett and Blaschke 2003; Hay et al. 2003; Benz et al. 2004, Baatz et al., 2008; Lang 2008; Marcal et al. 2008; Neubert et al. 2008; Marcal and Rodrigues 2009; Sturm and Weindner 2009; Smith and Morton 2010; Tong et al. 2012). Segmentation algorithms are broadly grouped into three taxa as ; point-based, edge-based, and region-based techniques (Gonzalez et al., 2002; Acharya and Ray, 2007; Tian and Chen, 2007; Sonka et al., 2008). Many segmentation algorithms targeting RS applications have been developed and tested, however, only a few lead to qualitatively-convincing results that are robust and applicable under operational settings (Baatz and Schape 2000; Marcal 2009; Tong et al ,2012). When thinking beyond perfect image object candidates (i.e., an ideal correspondence between an image object and a real-world entity), two main cases of anomalies are typically expected in segmentation, i.e. over-segmentation and under-segmentation (Clinton et al., 2010; Marpu et al. 2010; Kim et al., 2011). The former is generally acceptable, although it could be problematic if the geometric properties of image objects candidates are used in

the classification step.The latter is highly unfavorable because of the resulting mixed-segment problem and it has to be avoided (Sturm and Weidner 2009; Marpu et al. 2010; Liu et al., 2012). Thus, optimal segmentation results in which the average size of image objects is similar to that of the targeted real-world objects, are desired for achieving better classification accuracy (Dorren et al. 2003, Addink et al. 2007, Smith 2010; Kim et al., 2011; Myint et al., 2011).

Of the plethora of segmentation algorithms, eCognition Developer's proprietary multiresolution segmentation (MRS, Baatz and Schäpe, 2000), a region-based technique, has proven to be the most successful segmentation algorithm capable of producing meaningful image objects in many remote sensing application domains (Neubert et al. 2008). The MRS algorithm has also been implemented in similar form into other software packages, such as, example BerkeleyImageSeg (BIS; http://www.imageseg.com) or InterImage (Costa et al., 2008). This algorithm is relatively complex and image- and user-dependent (Hay et al., 2003; Marpu et al., 2010). Scale, shape, and compactness are the main parameters available to users for controlling the algorithm (Smith and Morton, 2010; Liu et al., 2012; Tong et al., 2012). Of the three parameters, the scale parameter is considered as most important as it controls the relative size of the image objects, which has a direct impact on the subsequent classification steps (Dorren et al., 2003; Smith, 2010; Kim et al., 2011; Myint et al., 2011).

Plurality of segmentation results is common because each parameter may take a range of values within its parameter space or different combinations of values among parameters. Finding optimal parameter values through a trial-and-error process (Myint et al., 2011; Duro et al. 2012), a subjective visual estimation of segmentation stability, is

commonly practiced at the expense of time and labor, thus, a significant number of alternative supervised and unsupervised methods (Zhang et al., 2008) have been proposed and tested. Supervised methods are designed to measure the dissimilarity between segmentation results and user-generated (e.g., hand digitized) reference objects. The parameter setting (especially the scale parameter) at which segmentation result is the most similar to the reference objects is considered as the optimal parameter combination of the MRS algorithm for the targeted land use/cover class. Examples for supervised optimal parameter selection methods include those by Carleer et al., (2005); Chabrier et al., (2006); Moller et al., (2007); Tian and Chen (2007); Costa et al., (2008); Neubert et al., (2008); Trias-Sanz et al. (2008); Chen et al., (2009); Smith and Morton (2010); Liu et al., (2012); Tong et al., (2012). Supervised methods provide a certain degree of freedom to integrate expert knowledge into optimal parameter selection procedures, however, these methods are typically time- and labor-intensive. Unsupervised methods, on the other hand, are robust, rigorous, self adaptable, transferable, and require minimum user intervention. In recent years, several unsupervised methods have been developed and tested (Espindola et al. 2006; Kim and Madden 2006; Dragut at al. 2010; Johnson and Xie, 2011; Martha et al., 2011).

In the case of supervised empirical assessments, discrepancy measures (Zhang, 1996) are employed for computing measures of dissimilarity between a reference polygon and an image object candidate. Based on different design goals, many segmentation quality metrics have been developed and tested. Typically, in the case of high-quality segmentation, the minimum geometric and arithmetic discrepancies between image objects and reference polygons are desired (Neubert et al., 2008; Sturm and Weidner,

2009). This state is considered as the optimal parameters setting (scale, shape, and compactness) of the MRS algorithm with respect to the quality metric used in the segmentation quality evaluation. Despite the plethora of segmentation quality metrics, their sensitivity to quality variation has rarely been investigated. Evidently the reliability of the optimal-parameter prediction heavily relies on the sensitivity of the segmentation quality metric. The idea behind pursuing optimal parameter setting is that, for instance, a given scale setting provides image object candidates different from the other scale setting; thus, by design the supervised quality metric should capture this difference. In this exploratory study, we aimed to unravel the discriminative capacity of a recently-proposed supervised quality metric (Euclidean Distance-2 (ED2), Liu et al. (2012) for selecting optimal parameter values. The ED2 metric's main design goal is to optimize the geometrical discrepancy (potential segmentation error (PSE)) and arithmetic discrepancy between image objects and reference polygons (number-of segmentation ratio (NSR)) in the two dimensional Euclidean space. Liu et al. (2012) tested this metric to estimate the optimal parameter setting of the MRS algorithm when applied to moderate (Landsat 5 TM) and high spatial resolution (ALOS) images. The ED2 metric was tested against four exiting indices (quality rate (QR), over-segmentation rate (OR), under-segmentation rate (UR), and Euclidean distance 1 (ED1)). Liu et al. (2012) concluded that the ED2metric is superior to its contenders when optimizing the MRS algorithm and it can also be used to optimize other segmentation algorithms (e.g., BIS, ENVI EX, ERDAS Imagine Objective). In this study, we systematically segmented images from three sensors (GeoEye-1, WorldView-2, EO-1 ALI) representing different geographies and target land used/cover classes using the MRS algorithm at a spectrum of scale, shape,

and compactness settings. We then statistically analyzed the sensitivity of ED2 metric to the segmentation-quality variation.

The rest of the paper is structured as follows. Section 2 describes study area, image data, image segmentation and quality assessment, and design goals of ED2 metric. Section 3 reports the quality of image object candidates and statistical significance and discusses the results based the discriminative capacity of the ED2 metric and its validity in the MRS algorithm's optimal parameter estimation. Finally, conclusions are drawn in Section 4.

4.2. Methodology

4.2.1 Study area and image data

We selected four study areas from the state of Connecticut and Sri Lanka (Figures 32 and 33). The QuickBird and WorldView-2 image scenes of Connecticut capture the University of Connecticut (CT-1) and a part of the Town Mansfield (CT-2), respectively. The GeoEye-1 and a Earth Observation (EO)-1 Advanced Land Imager (ALI) scenes of Sri Lanka comprise a former war-induced ephemeral settlement (the Menik Farm internally displaced persons (IDP) camp)(SL-1) and a part of the Eastern Province (SL-2), respectively. The images are spatially registered to the Universal Transverse Mercator (UTM) coordinate system on the WGS 84 datum. General characteristics of the study areas and candidate image scenes are summarized in Table 20.

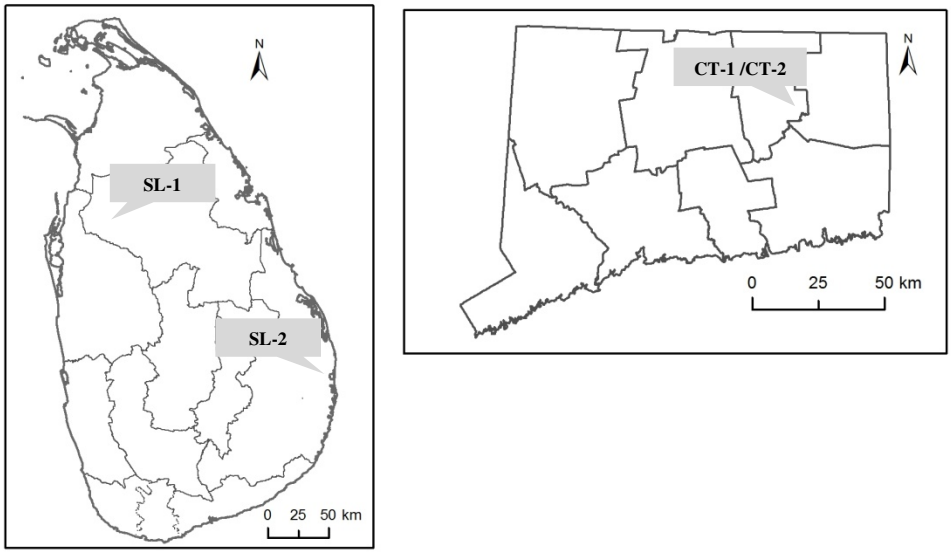


Figure 32. Index maps of the study areas

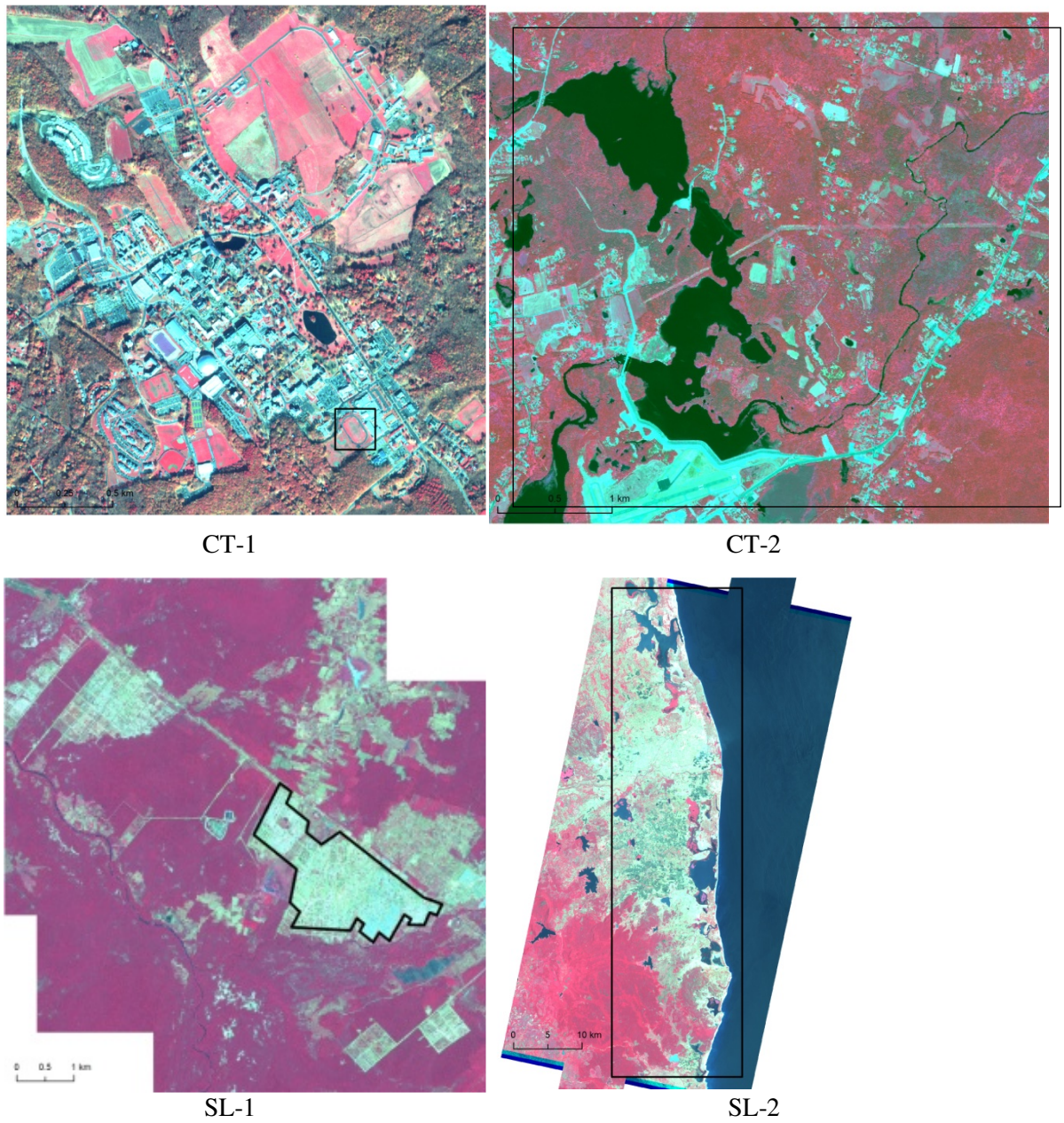


Figure 33. Image subsets used for segmentation. a) QuickBird MS image (CT-1), b) WorldView-2 MS image (CT-2) , c) GeoEye-1 resolution-enhanced MS image (SL-1), and d) EO-1-ALI MS image (SL-2) . Black-outlined polygons cover area of the targeted land use/cover classes. All images are shown in false-color composites.

Table 20. General characteristics of the study sites and image scenes

Geographic setting	Study site	Target features	Sensor	Used product for segmentation
Connecticut	University of Connecticut	Sport field	QuickBird (PAN- 0.61m, MS - 2.44m)	MS bands only
	Town of Mansfield	Lakes, pastures, single-family houses	WorldView-2 (PAN-0.50m, MS-2m)	MS bands only
Sri Lanka	Menik farm IDP camp	IDP shelters	GeoEye-1 (PAN-0.50m, MS-2m)	Fused image
	Eastern Province	Water bodies, agricultural areas, residential areas	EO-1 ALI (PAN-10m, MS-30m)	MS bands only

4.2.2 Methods

4.2.2.1 Image segmentation - Multiresolution segmentation algorithm (MRS)

The MRS algorithm is a bottom-up region merging technique (Tian and Chen, 2007; Duro et al. 2012). It aggregates individual pixels into increasingly larger size segments at multiple levels in an iterative process with respect to three parameters: 1) scale, 2) shape, and 3) compactness. At each step, merging of two adjacent segments is decided by spectral heterogeneity and shape heterogeneity the weighted sum of which is used to calculate image-segment fusion values (Benz et al. 2004; Trimble GmbH, 2011). The amalgamation of two adjacent image segments occurs when the calculated fusion value (f) is less than the square of user-defined scale threshold (s) (i.e, $f < s^2$). The spectral heterogeneity determines the size of the segment. The shape heterogeneity is the weighted sum of two additional parameters; compactness and smoothness. Individual weight and weighted sums of the parameters are scaled between zero and one. A detailed review on the mathematical formulation of the MRS algorithm is beyond the scope of this paper, thus, we encourage readers to refer relevant literature (e.g., Baatz and Schäpe 2000; Benz et al. 2004; Tian and Chen, 2007; Trimble GmbH, 2011; Tong et al. 2012) for more details.

4.2.2.2 Segmentation quality assessment-Conceptual framework

After partitioning an image scene into non-overlapping homogenous areas, the resulting disjoint segments (image object candidates) serve as the basic building blocks for reconstructing reference objects (Figure 15 in Chapter 3). The quality of segmentation can be assessed with respect to three scenarios; optimal-, over-, and under-segmentation. The optimal segmentation serves as the best case scenario required to

achieve accurate classification results. We categorize image segments into three different object types (Figure 16) as; 1) image object candidate, 2) corresponding image object, and 3) satisfactory image object. An image object candidate is realized as a corresponding image object when it intersects with a reference object (geo-object, Lang et al., 2010) that needs to be reconstructed. Being a corresponding image object does not necessarily guarantee that it would be used in the reference object reconstruction process. Therefore, a corresponding image object is further realized as a satisfactory image object when it meets the user-defined criterion (equation 3) for being elected for the reconstruction process.

Let image scene be I and its segments as S.

$$I = \bigcup_{i}^{n} s_i \quad (1)$$

where, $S = \{s_i: i = 1, 2, ..., n\}$

for any s_i and s_j ;

$$s_i \cap s_j = \emptyset$$

Let R be reference objects of land cover class C. In order to reconstruct reference objects (R), corresponding image objects (S_c) should be identified from image object candidates (S).

□ = ∪□ □□□ (2)

Here s_{ck} image objects are assigned to the reference object r_p ;

$$s_{ck} \cap r_p = \emptyset$$

where, $R = \{r_p: p = 1, 2, ..., q\}$, $S_c = \{s_{ck}: k = 1,2,....,l\}$

We refine corresponding image objects (S_c) as satisfactory image objects (S_s) based on the areal overlap criteria (Clinton et al., 2010).

Here s_{su} image objects are assigned to the reference object r_p ;

$$\frac{|s_{su} \cap r_p|}{|s_{su}|} > 0.5 \quad or \quad \frac{|s_{su} \cap r_p|}{|r_p|} > 0.5 \tag{3}$$

where, $S_s = \{s_{su}: u = 1,2,....,v\}$

2.2.2 Segmentation quality metrics - *Euclidean Distance-2 (ED2)*

Basic quality indices, such as quality rate (QR), over-segmentation rate (OR) , under-segment rate (UR), and an optimization of UR and OR in the Euclidean space (ED1) estimate the geometrical congruence between the reference objects (r_p) and satisfactory image objects. In reality, the highest geometrical fit does not necessarily manifest a good segmentation because at the highest geometrical congruency the size of satisfactory image objects can be the size of individual pixels. Figure 17 (in Chapter 3) illustrates three possible arithmetic outcomes expected between reference and image object candidates: 1) one-to-one, 2) one-to-many, and 3) many-to-one. The first and second serves as the ideal and the most expected scenarios, respectively, while the last one is the most undesirable scenario. To address these critical scenarios, Liu et al. (2012) proposed the 1) potential-segmentation error (PSE)) and 2) the number-of-segmentation ratio (NSR), and optimized the PSE and NSR in Euclidean space. Mathematical formulation of these metrics is given in equations 4, 5, and 6. We employed ED2 for assessing segmentation quality and optimum parameter setting of the MRS algorithm. The highest segmentation quality is reported when the calculated ED2 metric value is close to zero.

$$PSE = 1 - \frac{\sum|s_{su} - r_p|}{\sum|r_p|} \quad (4)$$

$$NSR = \frac{abs(q - v)}{q} \quad (5)$$

$$ED2 = \sqrt{PSE^2 + NSR^2} \quad (6)$$

$QR, OR, UR, \text{ED1} \in [0,1]$ and $PSE, NSR, ED2 \in [0, \infty)$

4.2.2.3 Statistical analysis

As previously noted, our contention is that the segmentation quality metric of interest (in our case ED2) should be sensitive to the quality variation of image object candidates produced at different parameter settings of the MRS algorithm. In general, area-based quality metrics are not normally distributed (Marpu et al. 2010), thus, comparison of sample medians (non-parametric tests, e.g., Kruskal-Wallis) are more meaningful than sample means (parametric tests, e.g., ANOVA). Compared to parametric tests, non-parametric tests have generally low power but they are conservative; when the underlying distribution is uncertain it is more appropriate to use non-parametric tests (Siegel and Castellan 1998; Quinn and Keough 2002; Rogerson 2006). In order to verify the shape of the distribution we applied the Shapiro-Wilk test (Rogerson, 2006) to both raw data and model residuals. Based on the test results of Shapiro-Wilk test, we

employed two non-parametric statistical tests (Kruskal-Wallis test and Wilcoxon pairwise test) for comparing the quality of image objects. The Kruskal-Walli test (Kruskal and Wallis, 1952) performs the comparison of population medians among multiples groups (two or more). When the obtained value of Kruskal-Walli test is significant (p-value < α level (e.g., 0.05)), it indicates at least one of the groups is different from at least one of the others (Sigel and Castellan, 1988). This test does not indicate which groups are different or how many of the groups are different from each other. The Wilcoxon pairwise test, an alternative to two-sample t-test, entails a pair-wise comparison of groups (Gibbons and Chakraborti, 2011). This test can be performed as one-tailed (H_0: $\theta_u = \theta_v$ and H_A: $\theta_u > \theta_v$) or two-tailed (H_0: $\theta_u = \theta_v$ and H_A: $\theta_u \neq \theta_v$). These statistical tests should be carried out under a user-nominated significance level (e.g., α = 0.05, 0.10, and 0.15), which specifies the probability level to accept that an event did not occur by chance. At lower significance levels, stronger evidence is needed for rejecting null-hypothesis. In the case of multiple testing situations, special corrections, such as Bonferroni correction (Quinn and Keough, 2002) and Benjamini procedure (Benjamini and Yekutieli, 2001), are needed for adjusting significance levels to control Type I error rates (Makarau et al., 2012).

4.2.2.4 Analysis workflow

The key steps involved in the analysis workflow are shown in Figure 34. In order to maintain varying spatial resolutions among candidates images, we enhanced only the GeoEye-1 MS image to 0.5m resolution using the Gram-Schmidt data fusion algorithm, while the other MS images were kept at their original resolutions (EO-1 ALI :30m, QuickBird: 2.44m, and WorldView-2: 2m). These candidates image scenes were then

systematically segmented using the MRS algorithm found in the 64-bit version of eCognition Developer 8 (Trimble GmbH, 2011) at a series of parameter settings. The MRS algorithm's scale (*s*) parameter is unbounded while shape (*sh*) and compactness (*cp*) parameters are bounded between 0 and 1. Therefore, theoretically, it is possible to have an infinite number of segmentation cycles depending on the combination of values that feed into scale, shape, and compactness. The MRS algorithm is highly time- and processor intensive, thus, we gradually increased (based on few apriori trails) the scale parameter at 10 unit steps from 10 to 150 for GeoEye-1 and WorldView-2 scenes and 10 to 200 for QuickBird and EO-1 ALI scenes. Shape and compactness parameters were changed at 0.1 intervals from 0.1 to 0.9. For a given scale (e.g. $s = 10$), the other two variables (shape and compactness) can yield 81 numerical combinations or 81 different segmentation cycles. We hand-digitized the outlines of the target features. As seen in Table 20, we aimed on; 1) a single large artificial feature (a sports field) in the QuickBird MS image, 2) three land use/cover types from the WorldView-2 MS image, 3) four IDP shelter types from the GeoEye-1 fused image, and 4) four land cover/use types from the EO-1 ALI MS image. Except for the sports field, we manually extracted 30 reference samples for each target class from each image. The resulting image object candidates were exported as shapefiles and PSE, NSR, and ED2 metrics were computed. Subsequently, scatter plots (scale parameter vs ED2) were constructed (as proposed by Liu et al., 2012) for each class in each image to estimate the optimal scale parameter. For statistical analysis, from each scale group (e.g., scale 10 of IDP shelter type-1 in GeoEye-1 image), 40 shape and compactness settings of the 81 combinations were randomly selected. The ED2 values produced at those 40 parameter combinations were initially

tested for the sample independence and utilized in further statistical tests. The normality was tested for both raw data and model residuals using Shapiro Wilk test. We used Kruskal Wallis test to compare multiple scale groups within each target class of each image. If the test was significant, we further employed the Wilcoxon pairwise test to compare scale groups (i.e., scale 10-20, 20-30, 30-40, .etc). As previously noted, when a large number of groups are compared there is a higher possibility of rejecting the null hypothesis incorrectly. To alleviate the Type-I error, we adjusted *p* values using the Benjamini procedure. We tasked detailed visual inspections to corroborate the objective segmentation quality analysis.

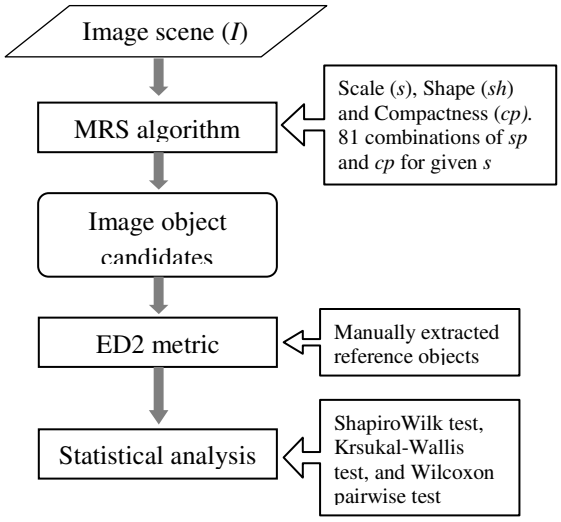


Figure 34. Flow chart showing key steps involved in the analysis workflow.

4.3 Results and discussion

We statistically analyzed the behavior of the ED2 metric as a supervised method for optimizing the multiresolution segmentation (MRS) algorithm. We selected multi-sensor candidate image scenes (QuickBird MS only, WorldView-2 MS only, resolution-enhanced GeoEye-1, and EO-1 ALI MS only) with varying spatial resolutions and scene contents.

Figures 35-38 depict the scatter plots constructed based on the image objects of each target class in each image across different scale settings of the MRS algorithm. Each scale group (we use 'scale group' because it pools a number of segmentation cycles) represents 81 numerical combinations (on the scatter plots, each dot indicates a parameter combination) of the shape and compactness parameters. With respect to the design goals of the ED2 metric, when the scale parameter (s) is increased across a bounded domain ($[s_i,s_k]$, let, $s_i < s_k$) in which the optimal scale (s_o) exists , the ED2 values monotonically decrease until the optimal scale group that exhibits low scores with the least scatter and monotonically increase after the optimal scale group. In general, the ED2 metric produces concave-shaped scatter plots along an increasing scale parameter. The number-of-segment ratio (NSR) dominates the left limbs of these scatter plots and the potential segmentation error (PSE) dominates the right limbs. In the case of the sports field (QuickBird MS image, Figure 35) scale group 50 can be elected as the optimal scale of the MRS algorithm as it exhibited the lowest ED2 values with the least scatter. For the three target classes of WorldView-2 image (Figure 36), the ED2 metric has suggested scale group 40 as the optimal for both single-family houses and small lakes. However, in the case of pastures the ED2 has indicated three possible optimal scale groups (60, 70,

and 80). A similar situation can be observed for the target classes of the GeoEye-1 image (Figure 8). While the ED2 metric suggested single optimal scales for the IDP shelter types 1 and 2 (s_o = 50, s_o=70, respectively), a window of optimal scales was indicated for the IDP shelter types 3 and 4. This visual inseparability of optimal scales on scatter plots is further highlighted in the target classes of EO-1 ALI image (Figure 9). Plurality of optimal scale groups is highly aggravated in the agricultural class where the ED2 metric exhibited almost similar results for s >30.

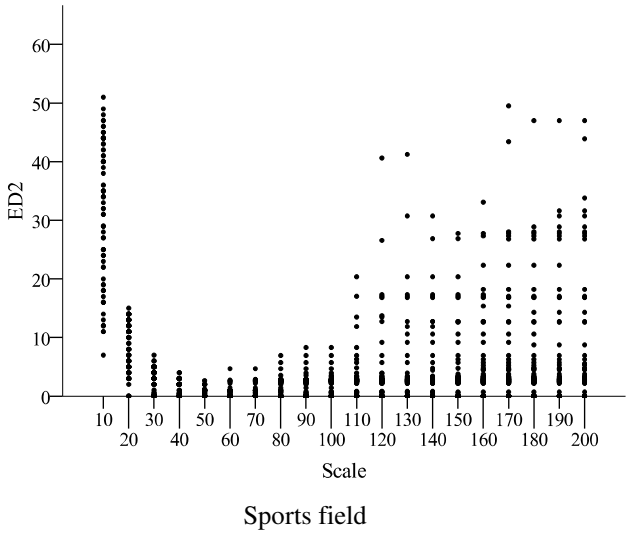


Figure 35. Scatter plot depicting the variation of the ED2 metric across different scale parameter setting of the multiresoultion segmentation algorithm for the sport field in the QuickBird MS image

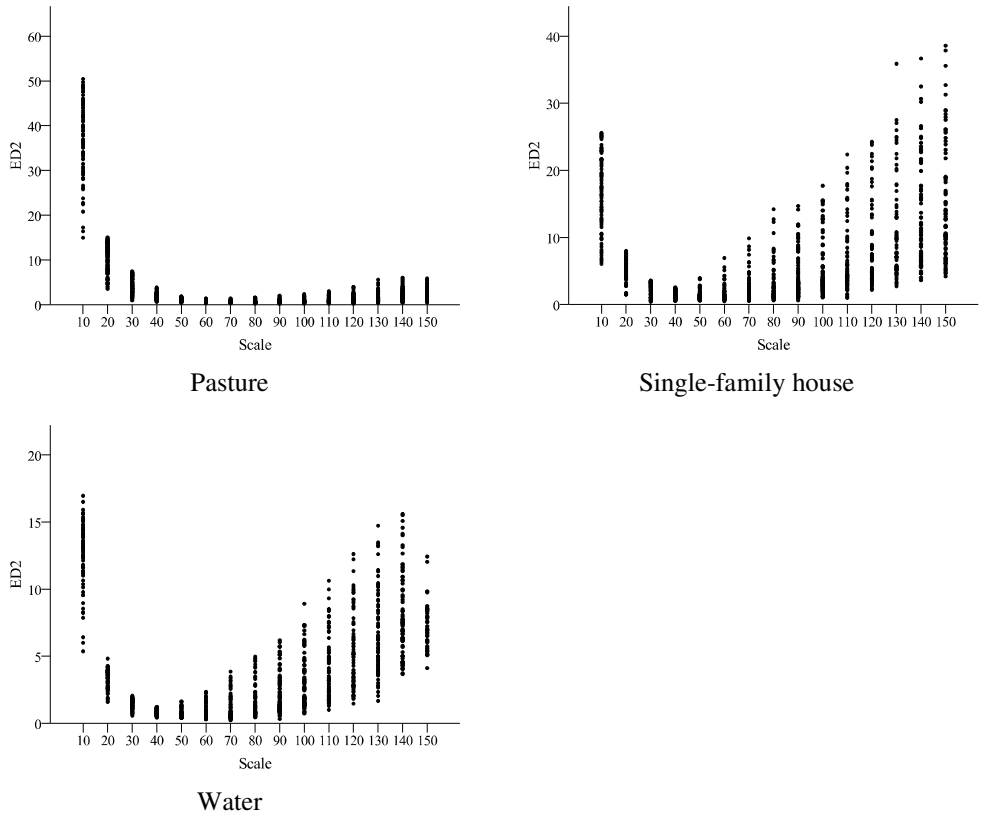


Figure 36. Scatter plot depicting the variation of the ED2 metric across different scale parameter setting of the multiresoultion segmentation algorithm for the land use/cover types in the WorldView-2 MS image.

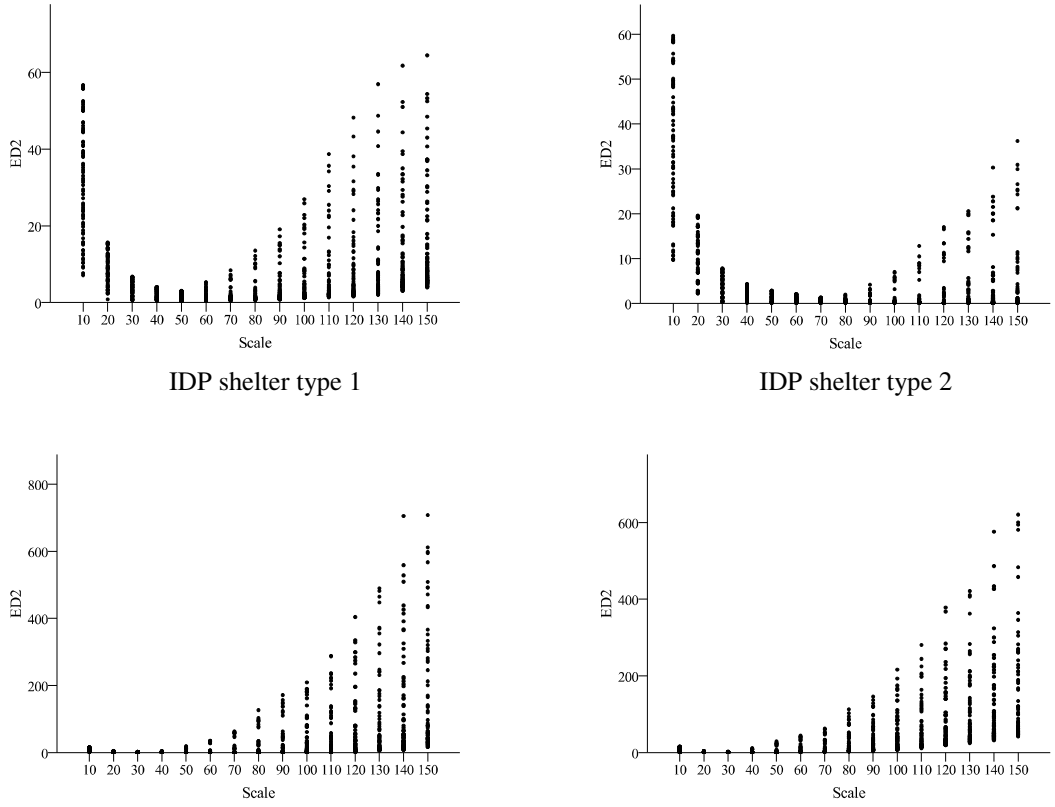


Figure 37. Scatter plot depicting the variation of the ED2 metric across different scale parameter setting of the multiresoultion segmentation algorithm for the IDP shelter types in the resolution-enhanced GeoEye-1 MS image.

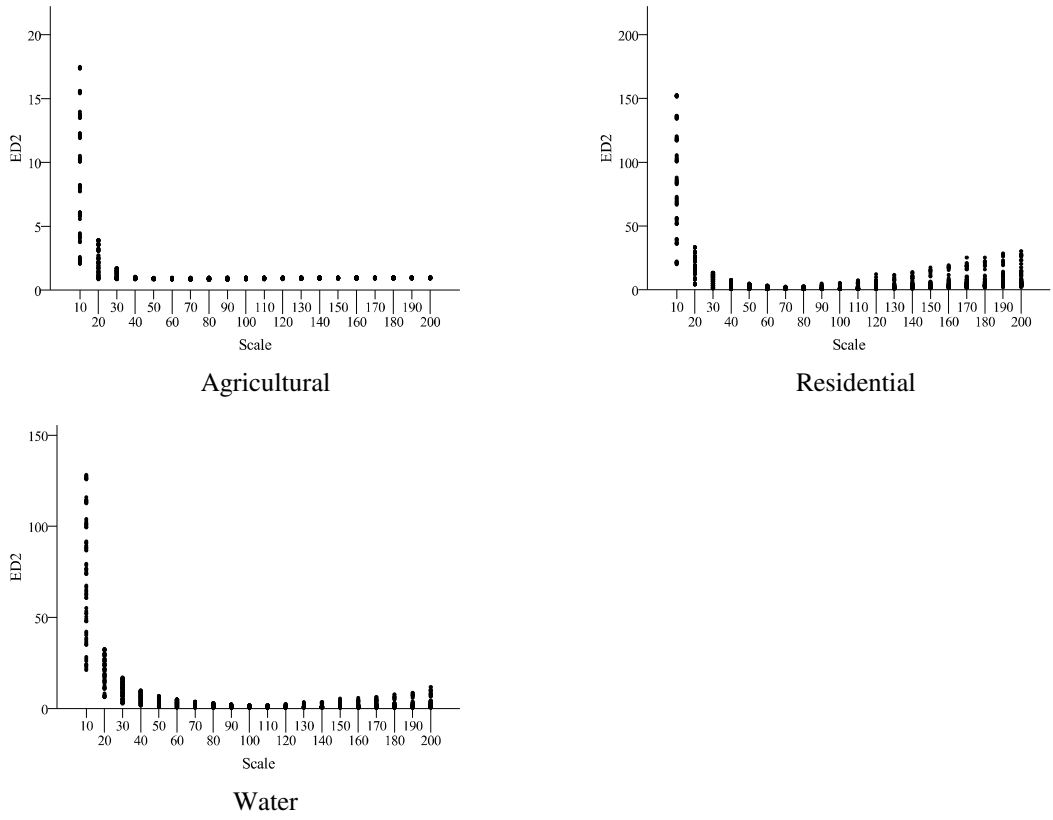


Figure 38. Scatter plot depicting the variation of the ED2 metric across different scale parameter setting of the multiresoultion segmentation algorithm for the land use/cover types in the EO-1 ALI MS image.

Generally speaking, it is true that the scatter plots of the ED2 metric provide some hints for selecting the MRS algorithm's scale parameter; however, the reliability of those estimations might be disputable. It is necessary to revisit the design goals of the ED2 metric and closely analyze whether it behaves as it was intended. Theoretically, a well-performing quality metric should be sensitive to the segmentation quality variation. In the case of the ED2 metric, it should significantly discriminate two adjacent scale groups (e.g., *s*=10 and *s*= 20). Our sensitivity analysis (Tables 21-24) is twofold: 1) the Kruskal Wallis test was employed to compare multiple scale groups (see the first column of each

table), 2) the Wilconxon pairwise comparison was employed to monitor ED2 metric's behavior relative to the increasing scale parameter.

Table 21. Reported Kruskal-Wallis and Wilcoxn pairwise test results for the QuickBird MS image

Land use/cover class	Kruskal-Wallis test. *p*-value	Wilcoxon pairwise tests, *p*-values																		
		Scale pairs																		
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	100-110	110-120	120-130	130-140	140-150	150-160	160-170	170-180	180-190	190-200
Sports field	<0.0001*	<0.0001*	<0.0001*	0.0045*	0.0264	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 22. Reported Kruskal-Wallis and Wilcoxn pairwise test results for the WorldView-2 MS image

Land use/cover class	Kruskal-Wallis test. *p*-value	Wilcoxon pairwise tests, *p*-values													
		Scale pairs													
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	100-110	110-120	120-130	130-140	140-150
Pasture	<0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*	0.0033*	0.0405*	1	1	1	1	1	1	1	1
Residential	<0.0001*	<0.0001*	<0.0001*	0.0061*	0.1245	1	1	1	1	1	1	1	1	1	1
Water	<0.0001*	<0.0001*	<0.0001*	<0.0001*	0.0088*	1	1	1	1	1	1	1	1	1	1

(Superscript * indicates significance levels (α) of 0.05)

Table 23. Reported Kruskal-Wallis and Wilcoxn pairwise test results for the GeoEye-1 resolution-enhanced MS image

Land use/cover class	Kruskal-Wallis test. *p*-value	Wilcoxon pairwise tests, *p*-values													
		Scale pairs													
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	100-110	110-120	120-130	130-140	140-150
Shelter 1	<0.0001*	<0.0001*	<0.0001*	0.006*	0.3213	1	1	1	1	1	1	1	1	1	1
Shelter 2	<0.0001*	<0.0001*	<0.0001*	0.0026*	0.0615	1	0.9093	0.0135*	1	0.9768	1	1	1	1	1
Shelter 3	<0.0001*	<0.0001*	<0.0001*	0.001*	0.448	0.0506	0.2922	1	0.9047	1	0.9106	1	0.8531	1	1
Shelter 4	<0.0001*	<0.0001*	<0.0001*	0.0997	0.0997	1	0.7037	0.7037	0.1318	1	0.0997	1	1	1	0.7037

Table 24. Reported Kruskal-Wallis and Wilcoxn pairwise test results for the EO-1 ALI MS image

Land use/cover class	Kruskal-Wallis test. *p*-value	Wilcoxon pairwise tests, *p*-values																	
		Scale pairs																	
		10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	100-110	110-120	120-130	130-140	140-150	150-160	160-170	170-180	180-190	190-200
Agri(forest)	<0.0001*	<0.0001*	<0.0001*	<0.0022*	0.0047*	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Residential	<0.0001*	<0.0001*	0.0008*	0.0003*	0.0004*	1	0.0046*	1	1	1	1	1	1	1	1	1	1	1	1	1
Water	<0.0001*	<0.0001*	<0.0001*	0.0002*	0.1784	0.0531	0.0094*	1	0.0531	1	0.6559	1	1	1	1	1	1	1	1	1

(Superscript * indicates significance levels (α) of 0.05)

As seen in the Tables 21-24, reported *p*-values for the Kruskal-Wallis test revealed that in each test site at least one scale group was significantly different (at $\alpha = 0.05$) from others. However, as we noted in the methodological section, the Kruskal Wallis test does not voice which group(s) is/are different. In the post-hoc analysis, we took measures to reduce the Type-I error (falsely rejecting the null hypothesis of having equal medians). We avoided all possible scale group comparisons and selected only meaningful ones (for example, comparing scale group 10 and 100 or 30 and 70 are less meaningful) as we have good apriori knowledge on the behavior of the ED2 metric. The *p*-values reported in summary tables for the Wilcoxon test have been adjusted using the Benjamini procedure. We corroborated numerical values shown in Tables 21-24 with box plots (see Figures 39-42), which indicate the variation of ED2 metric optimal and near-optimal scale groups. In the case of Sports field (Table 21), the ED2 metric was able to discriminate (see *p*-values) image object produced at scales 10, 20, 30, and 40; however, beyond that it failed to significantly capture the quality variation. It is important to revisit the corresponding scatter plot (Figure 39), which elected the scale group 50 as the optimal. However, the Wilcoxon test results revealed that scales groups 40-50 and 50-60 are not significantly different. The implication might be that the ED2 metric has lead to plurality of optimal scales. A similar trend can be observed in other test areas (see Tables 22,23, and 24) as well. For instance, in case of the four target classes of the GeoEye-1 image (Table 23), the ED metric exhibited a significant sensitivity for scale groups 10-20, 20-30, and 30-40. Our contention is that human interpreters are more capable of (i.e., superior to quantitative metrics) benchmarking image objects in multidimensional aspects (e.g., detection, geometric fit, arithmetic fit, shape, and thematic accuracy). Figures 44-48

depict the variation of the segmentation quality (arithmetic and geometric discrepancies) across scale settings with respect to a targeted reference object in the candidate images. In general, visually-detectable segmentation quality variation (subjective assessment) across scale settings should be captured by the objective quality metric. When comparing the test statistics (Table 21) and visual inspection (Figure 44), the ED2 metric agrees with the subjective segmentation quality interpretations from scales 10 to 40. However beyond that the ED2 exhibits disagreements. For example, when moving scale from 80 to 90, a striking under-segmentation can be visually detected; however, the ED2 was not significantly sensitive to that change. Overall, the ED2 metric discriminates image object candidates at smaller scale values (the left limb of scatter plots) but it loses the sensitivity at larger scale values. The reason is that at small scale values the NRS overpowers the PSE due to the high level of arithmetic discrepancy (one-to-many), however, when the scale parameter increases the PSE disproportionately overpowers the NSR.

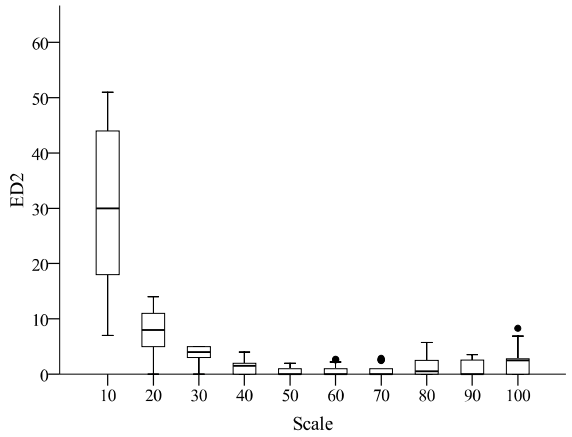


Figure 39. Box plots showing the variation of the ED2 metric across all scale settings for the sports field in the QuickBird MS image.

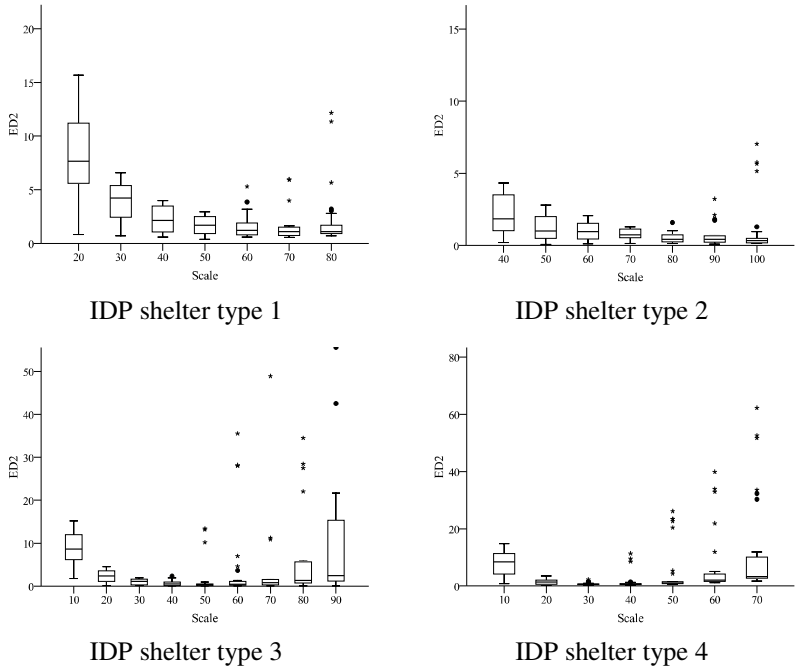


Figure 40. Box plots showing the variation of the ED2 metric across selected (the optimal and near-optimal scale) scale settings for the target land use/cover classes in the WorldView-2 MS image.

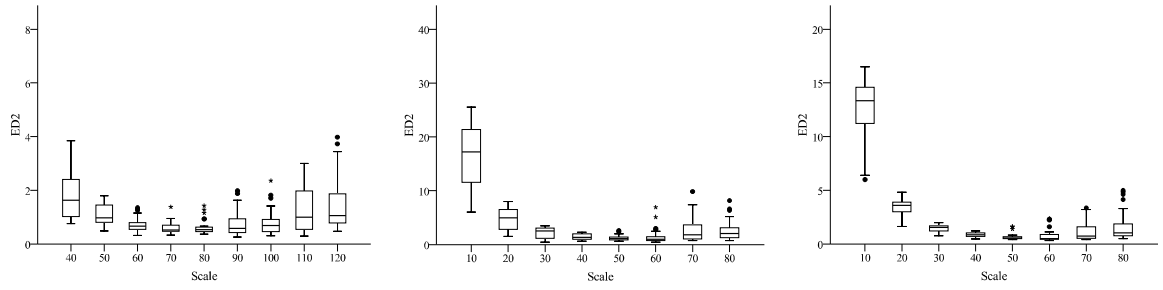


Figure 41. Box plots showing the variation of the ED2 metric across selected (the optimal and near-optimal scale) scale settings for the target IDP shelter types in the resolution-enhanced GeoEye-1 MS image.

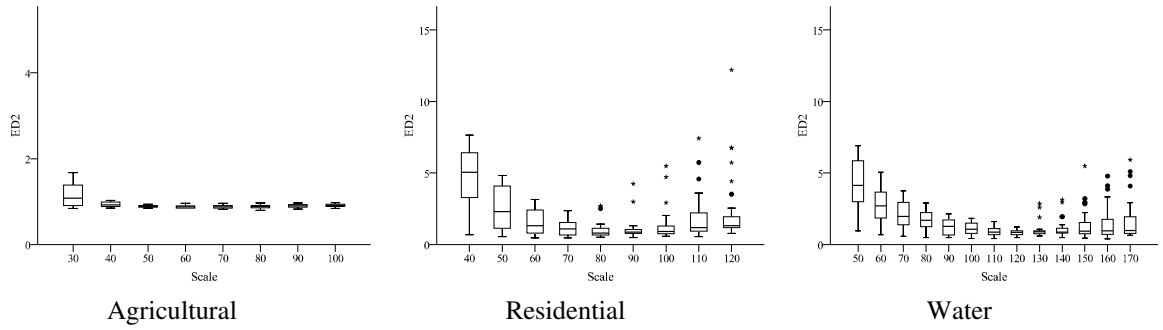


Figure 42. Box plots showing the variation of the ED2 metric across selected (the optimal and near-optimal scale) scale settings for the target land use/cover classes in the EO-1 ALI MS image.

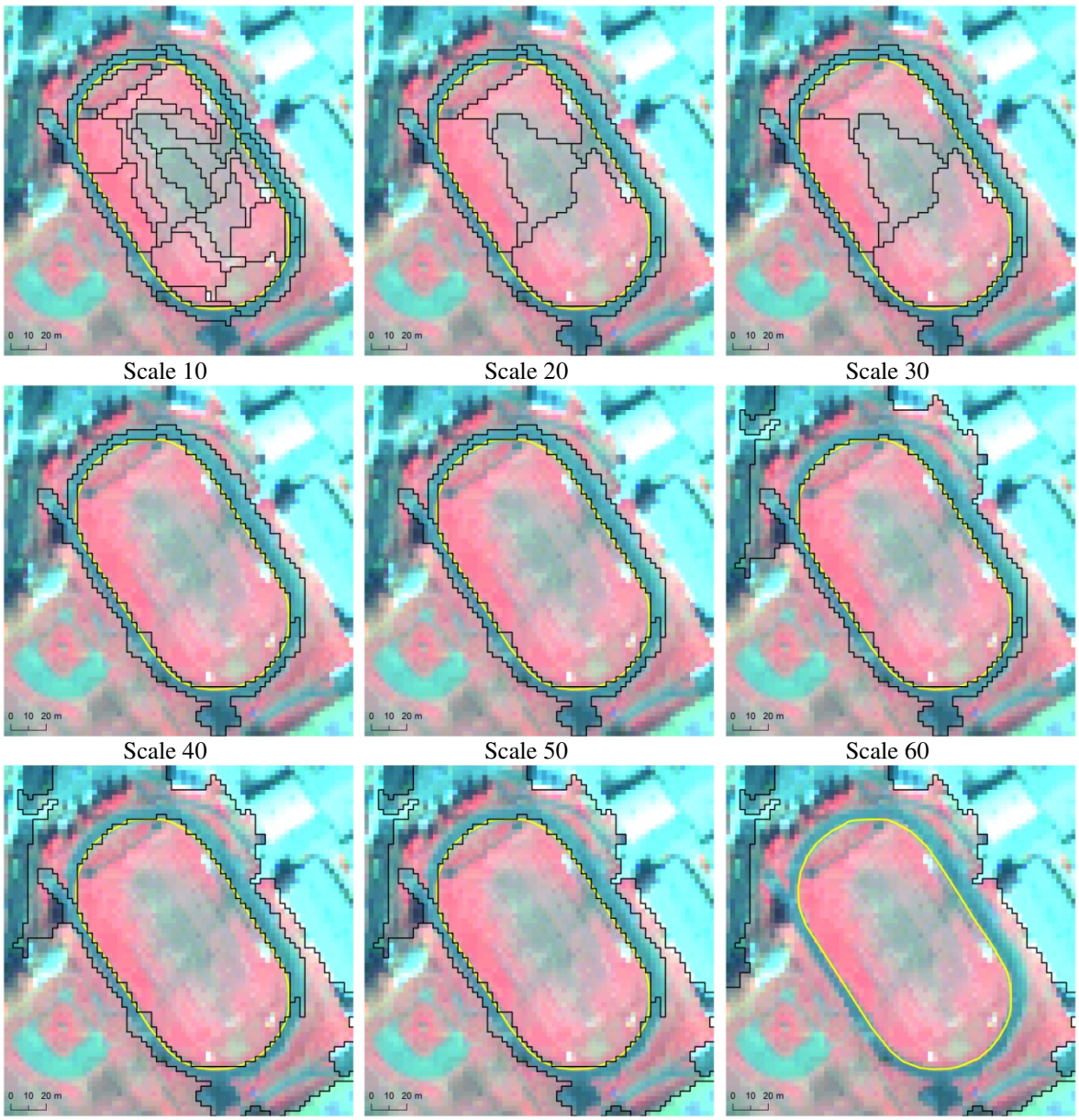
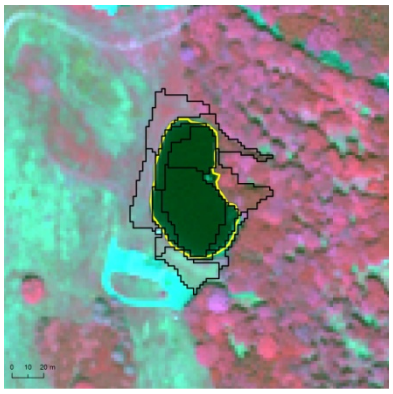
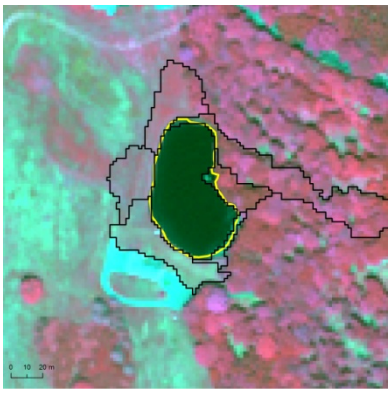


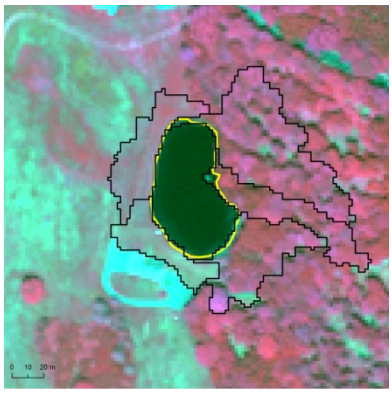
Figure 43. Variation of the segmentation quality (arithmetic and geometric discrepancies) across scale settings with respect to the target object (sports field - outlined yellow, corresponding image objects are in black) in the QuickBird MS image.



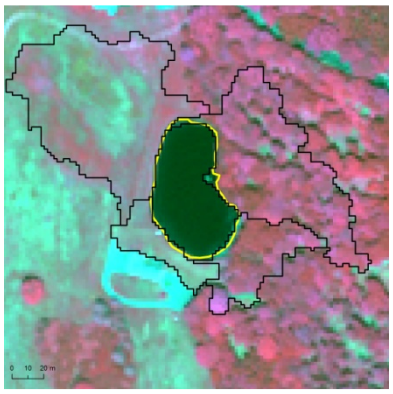
Scale 10



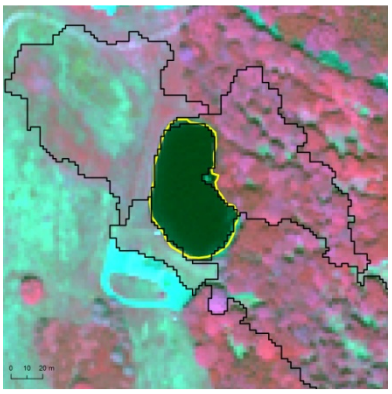
Scale 20



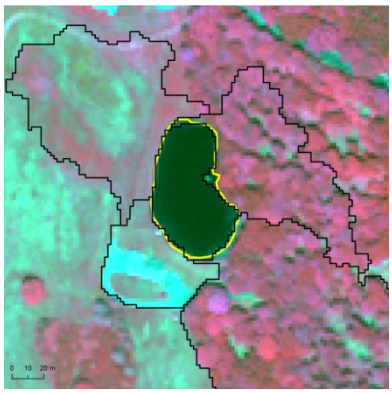
Scale 30



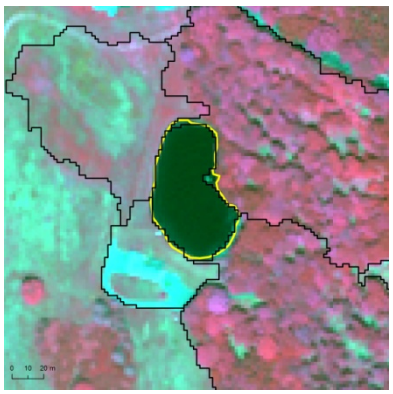
Scale 40



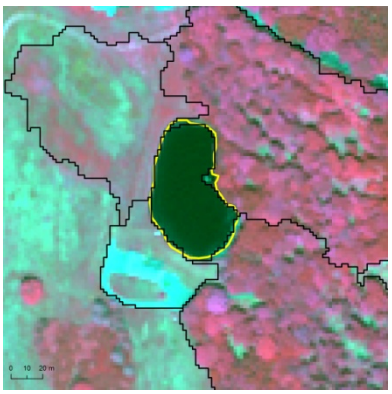
Scale 50



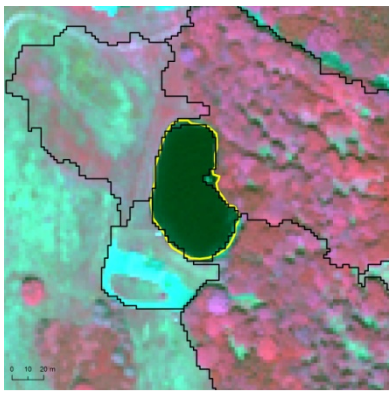
Scale 60



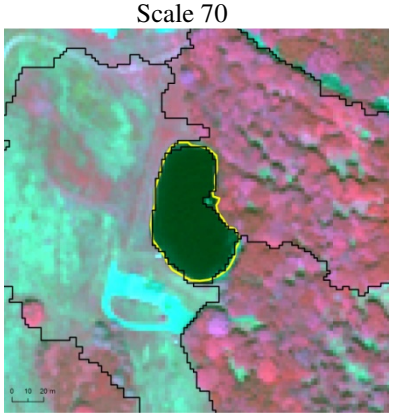
Scale 70



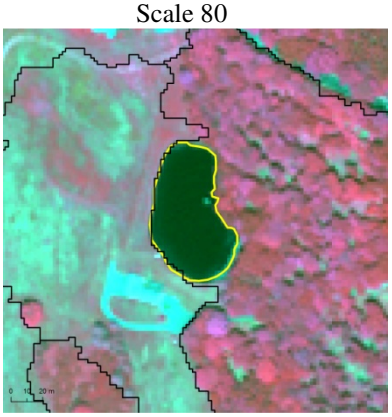
Scale 80



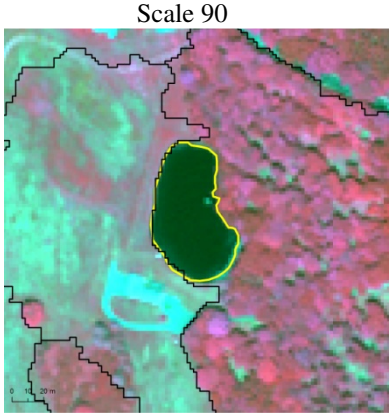
Scale 90



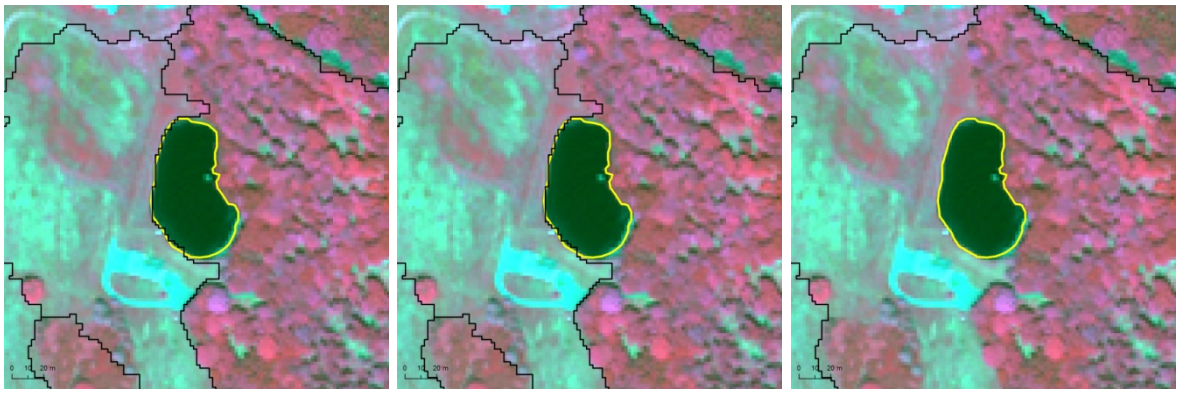
Scale 100



Scale 110

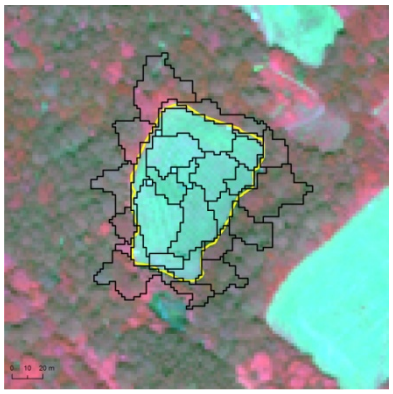


Scale 120

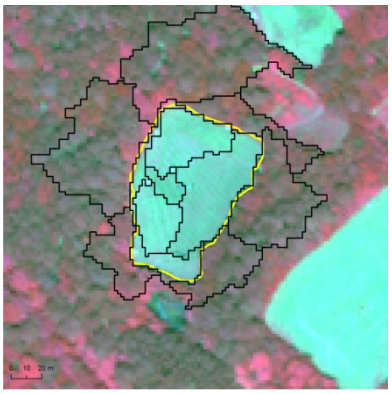


Scale 130 Scale 1140 Scale 150

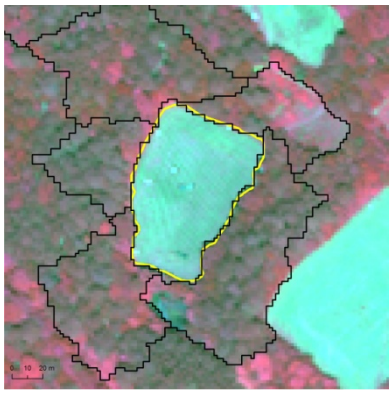
Figure 44. Variation of the segmentation quality (arithmetic and geometric discrepancies) across scale settings with respect to the target object (a water body - outlined yellow, corresponding image objects are in black) in the WorldView-2 MS image.



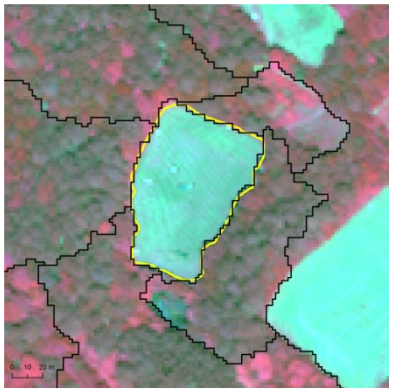
Scale 10



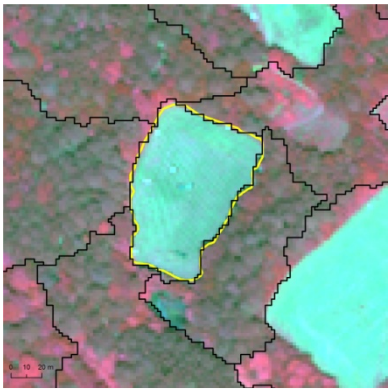
Scale 20



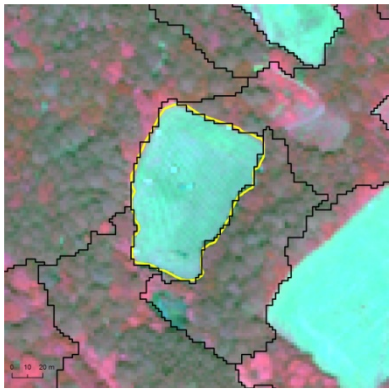
Scale 30



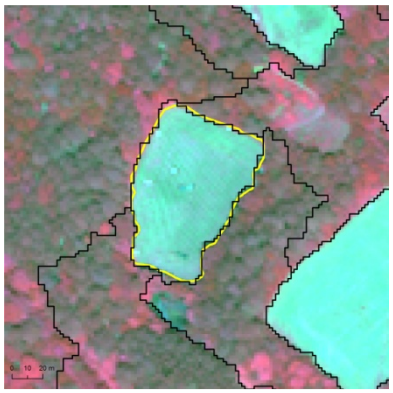
Scale 40



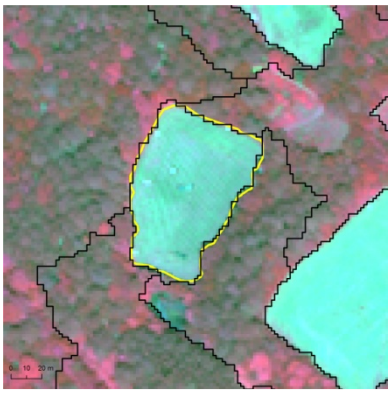
Scale 50



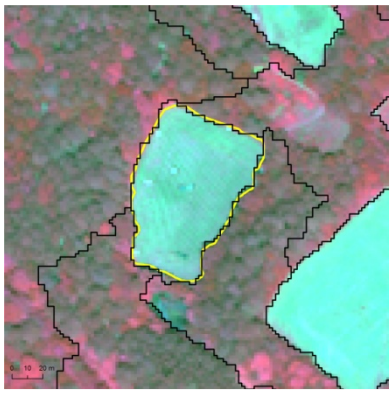
Scale 60



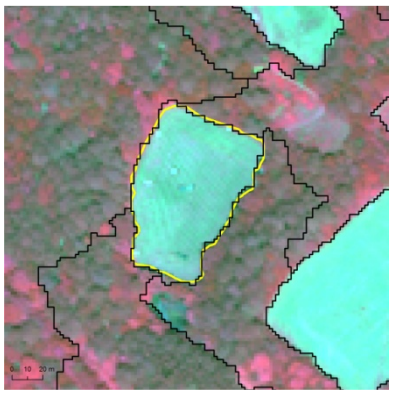
Scale 70



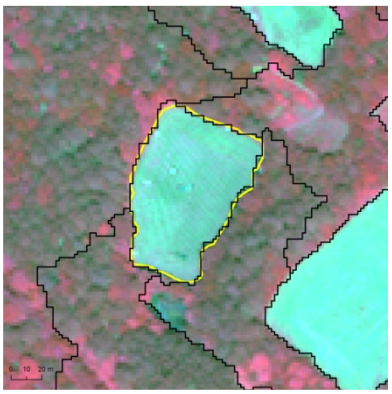
Scale 80



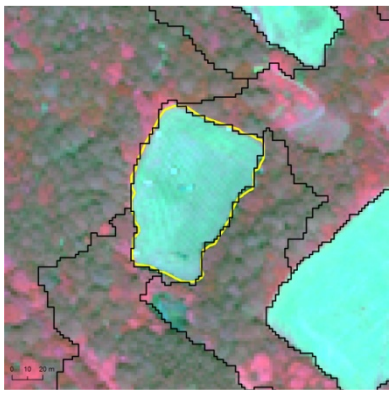
Scale 90



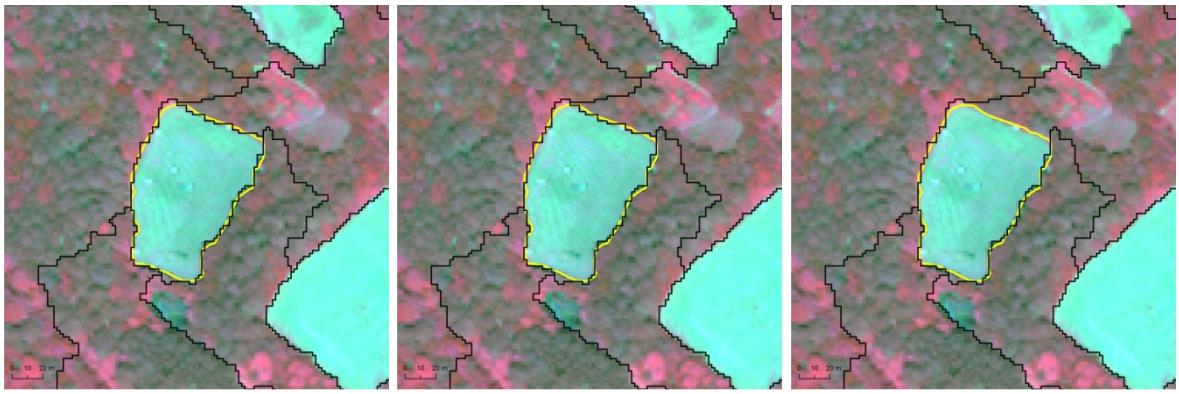
Scale 100



Scale 110

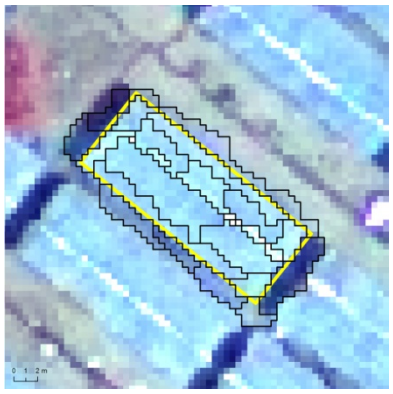


Scale 120

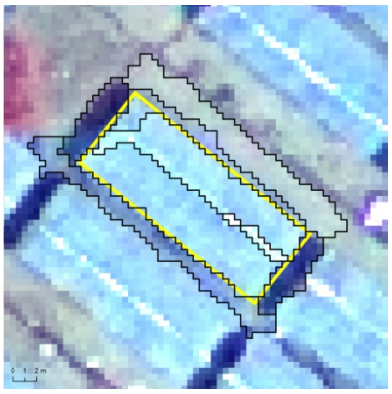


Scale 130 Scale 140 Scale 150

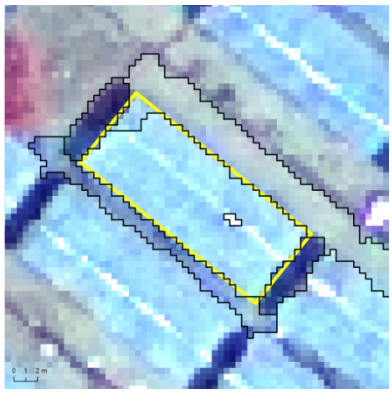
Figure 45. Variation of the segmentation quality (arithmetic and geometric discrepancies) across scale settings with respect to the target object (a pasture- outlined yellow; corresponding image objects are in black) in the WorldView-2 MS image.



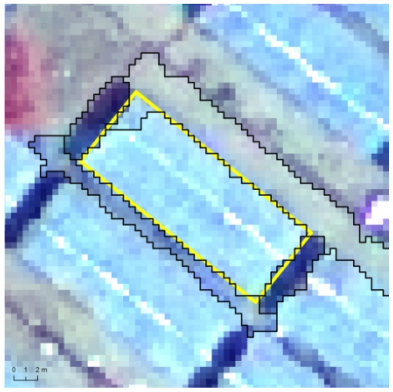
Scale 10



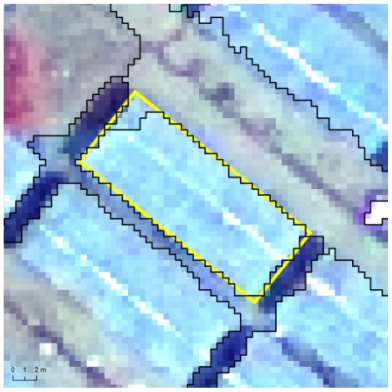
Scale 20



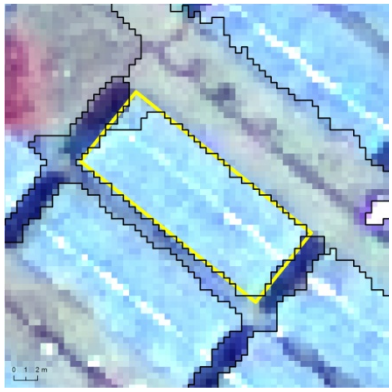
Scale 30



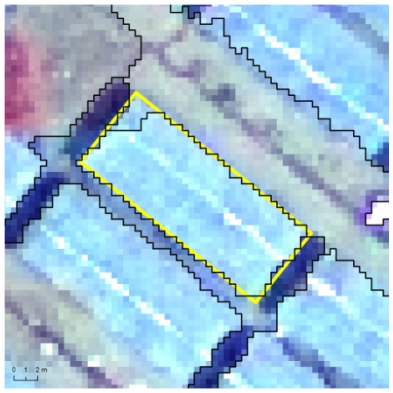
Scale 40



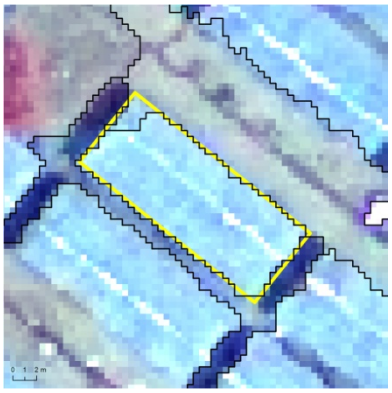
Scale 50



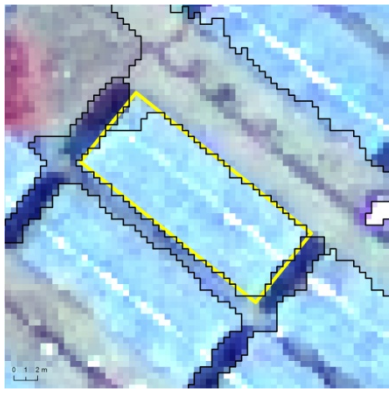
Scale 60



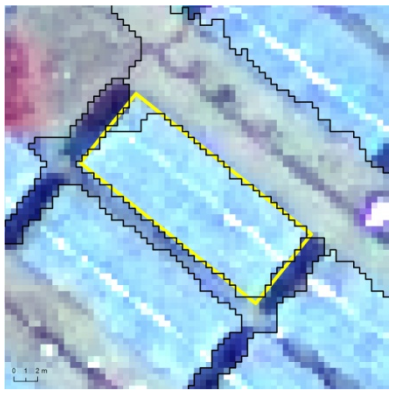
Scale 70



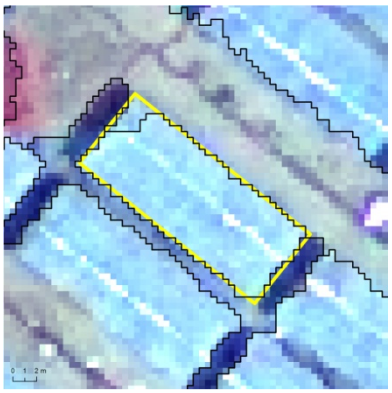
Scale 80



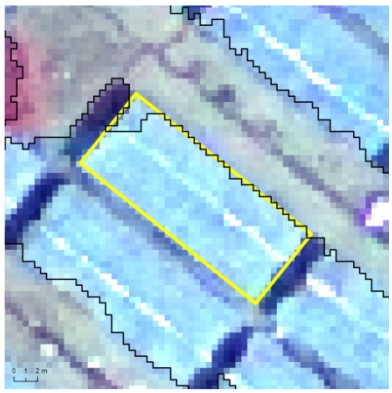
Scale 90



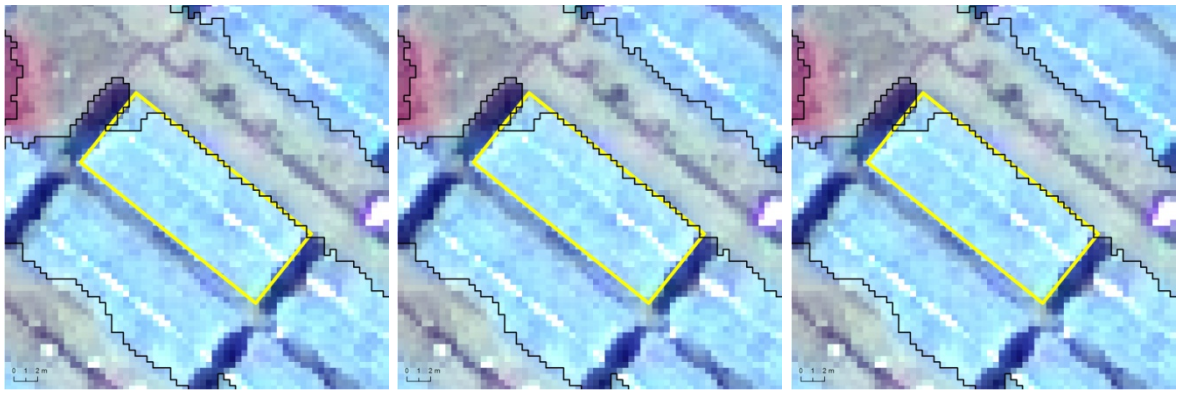
Scale 100



Scale 110

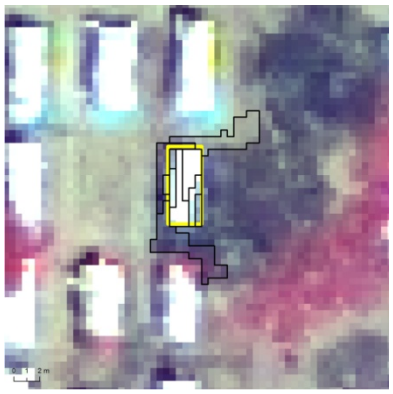


Scale 120

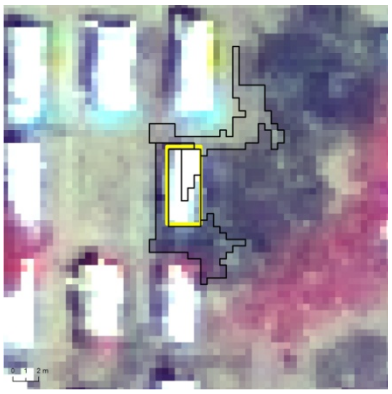


Scale 130 Scale 140 Scale 150

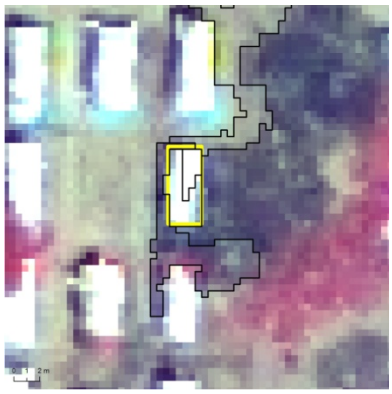
Figure 46. Variation of the segmentation quality (arithmetic and geometric discrepancies) across scale settings with respect to the target object (IDP shelter type 1 - outlined yellow, corresponding image objects are in black) in the resolution-enhanced GeoEye-1 MS image.



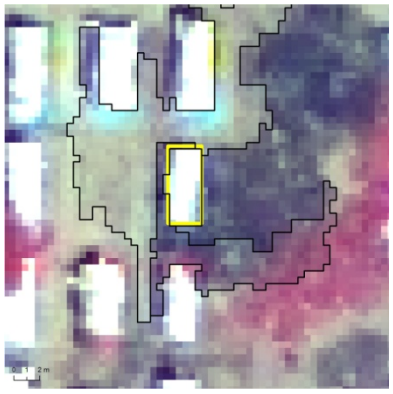
Scale 10



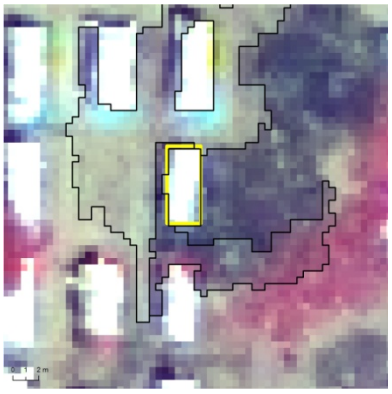
Scale 20



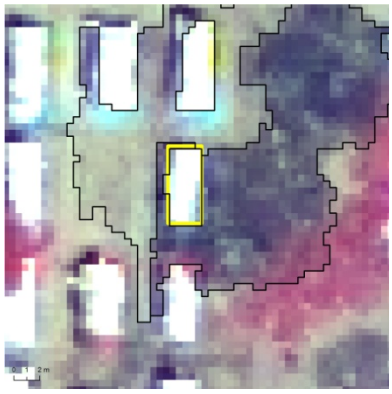
Scale 30



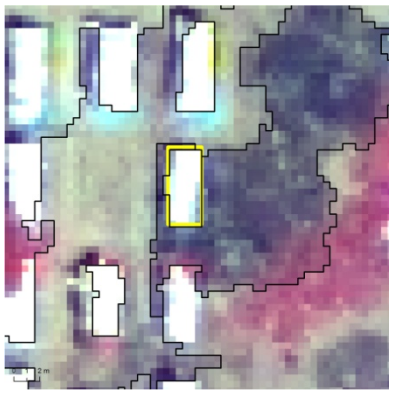
Scale 40



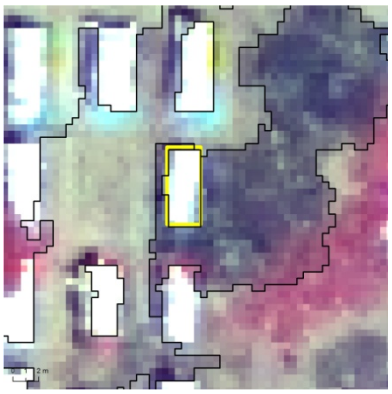
Scale 50



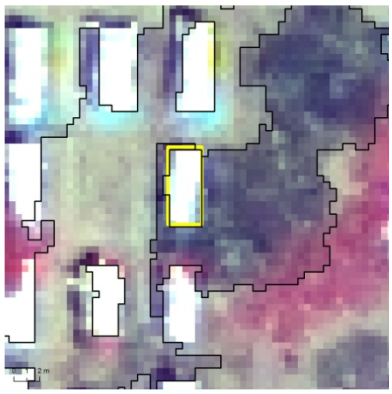
Scale 60



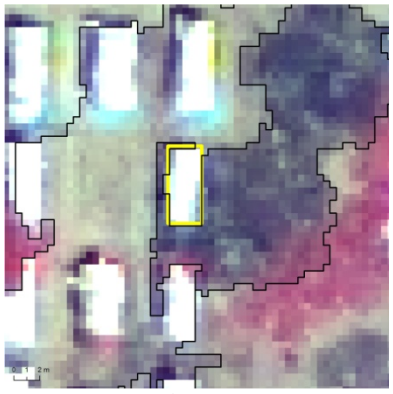
70



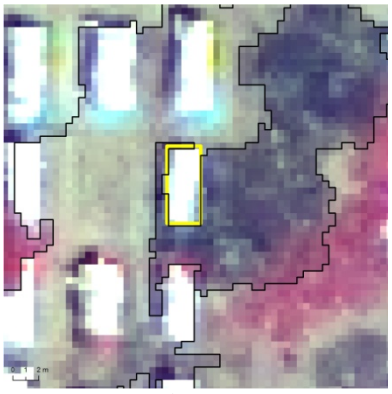
80



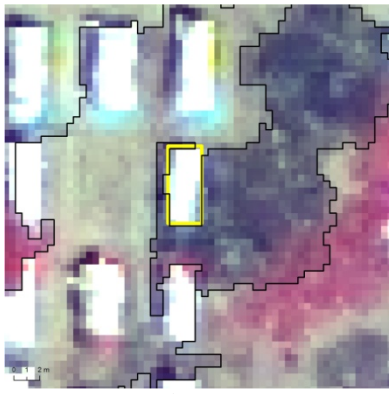
90



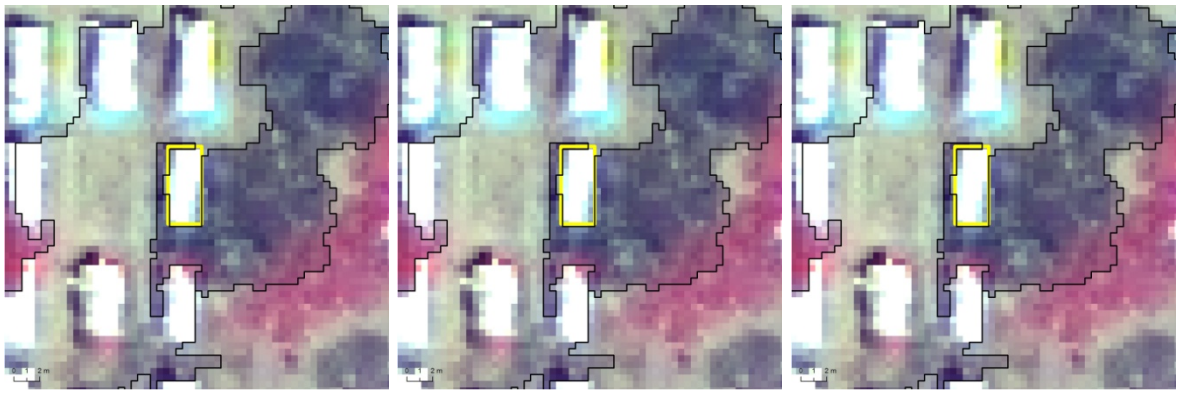
Scale 100



Scale 110

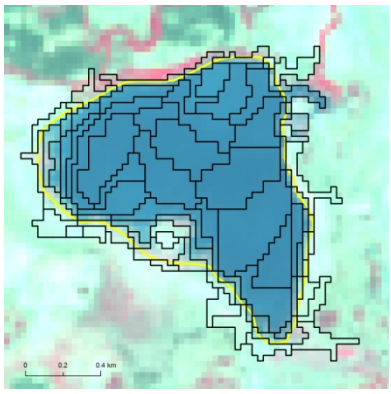


Scale 120

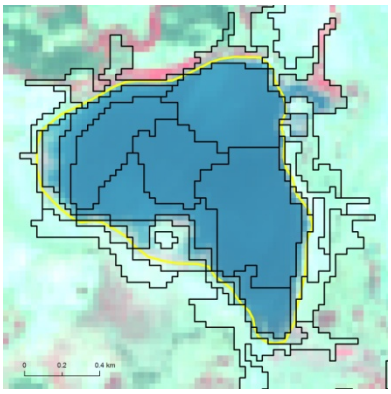


Scale 130 Scale 140 Scale 150

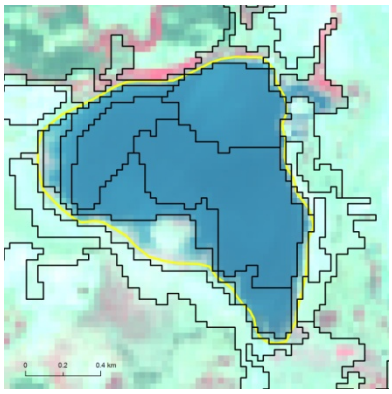
Figure 47. Variation of the segmentation quality (arithmetic and geometric discrepancies) across scale settings with respect to the target object (IDP shelter type 3 - outlined yellow, corresponding image objects are in black) in the resolution-enhanced GeoEye-1 MS image.



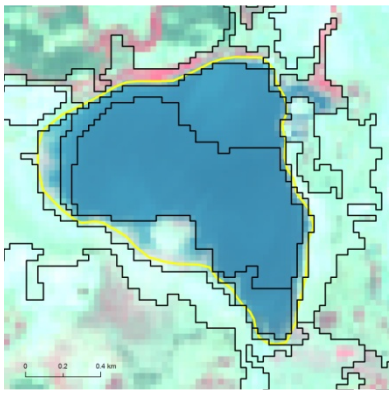
Scale 10



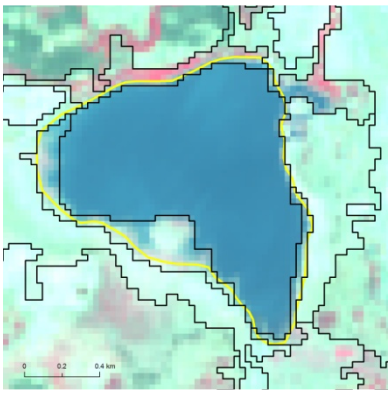
Scale 20



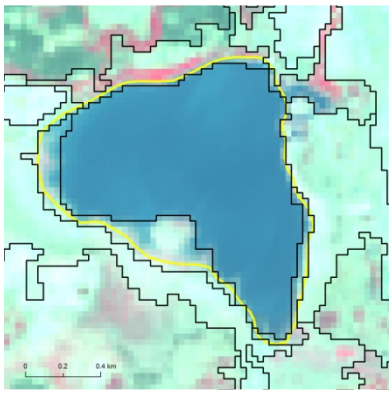
Scale 30



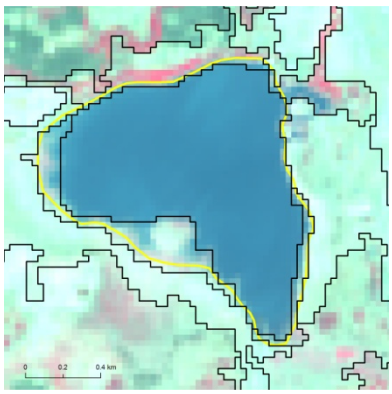
Scale 40



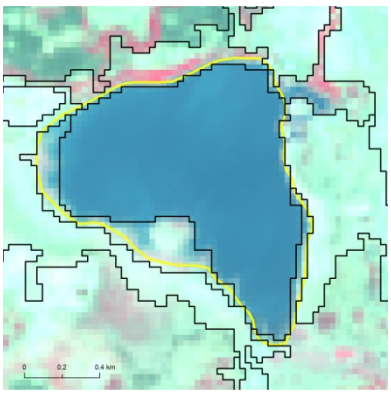
Scale 50



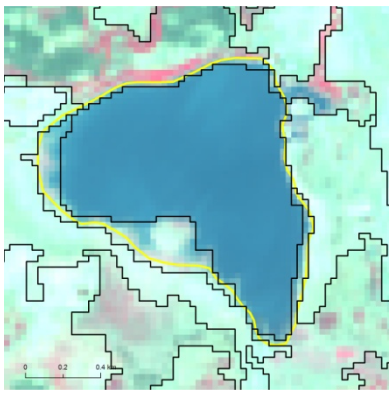
Scale 60



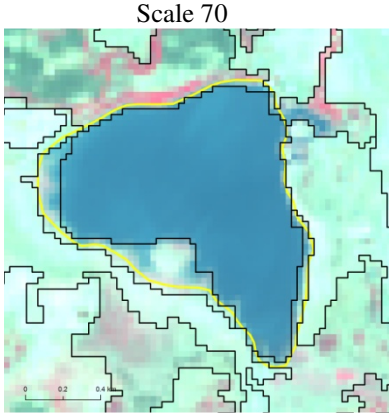
Scale 70



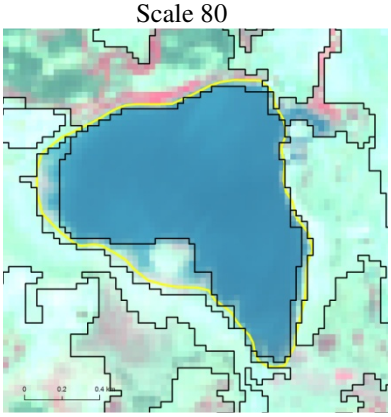
Scale 80



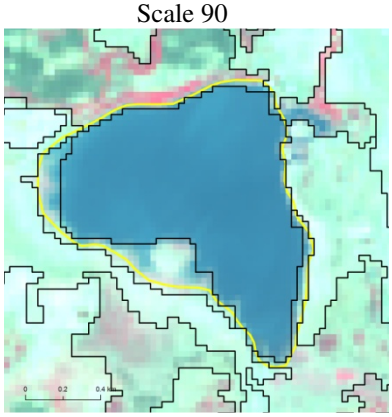
Scale 90



Scale 100



Scale 110



Scale 120

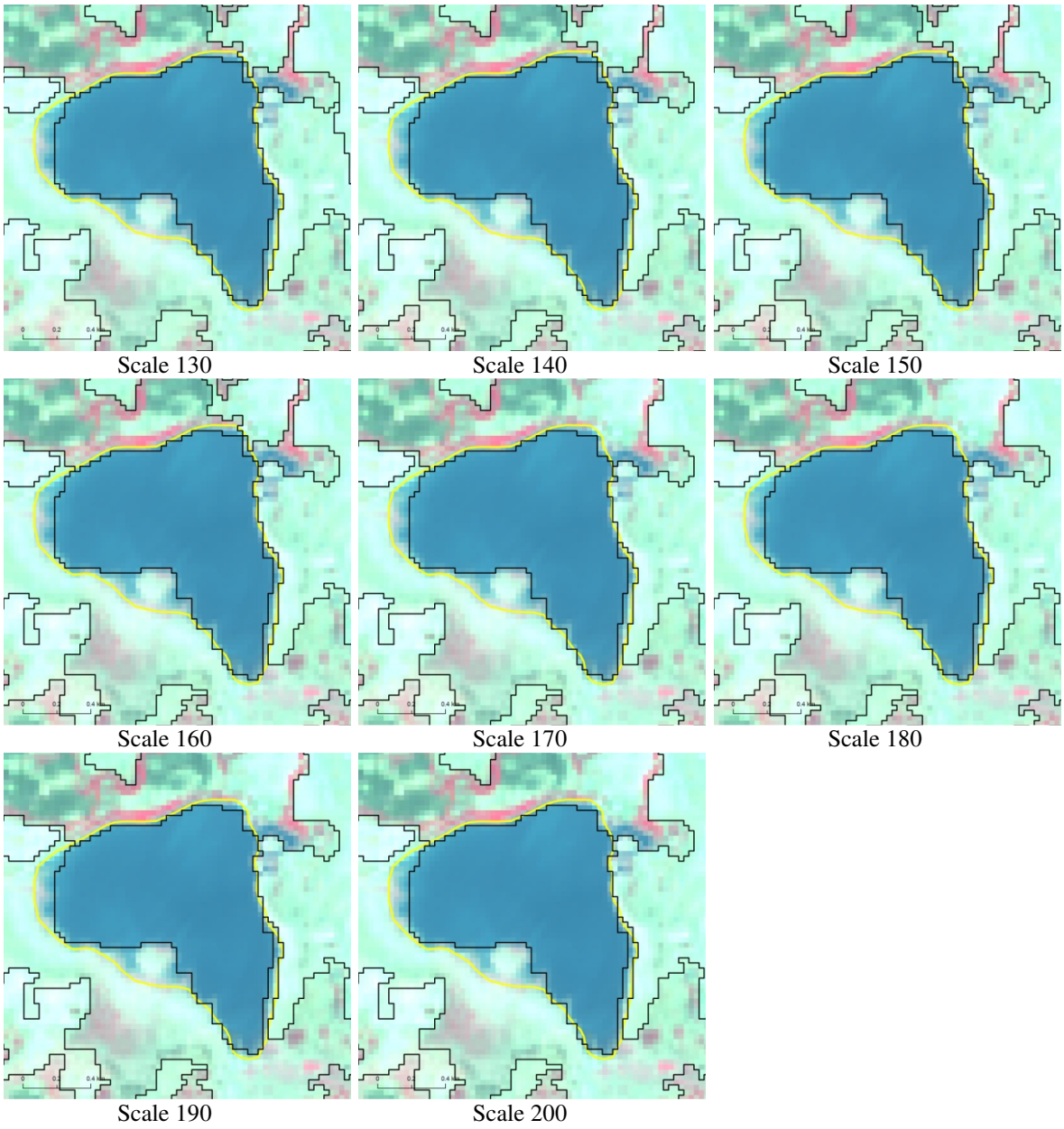


Figure 48. Variation of the segmentation quality (arithmetic and geometric discrepancies) across scale settings with respect to the target object (IDP shelter type 3 -outlined yellow, corresponding image objects are in black) in the EO-1 ALI MS image.

Optimizing the MRS algorithm using the ED2 metric is a time critical process. As discussed, the user needs to apply the time-intensive MRS algorithm to an image scene across a series of scale settings with some apriori knowledge on the lower and upper bounds of the scale parameter. The number of segmentation cycles seriously procreates depending on the parameter increment factors (i.e. scale at 1 or 10, ... unit intervals, shape and compactness at 0.1 or 0.2 ... intervals). To portray some sense of the time factor, the MRS algorithm consumed approximately 20 hours to segment the candidate GeoEye-1 image across the selected scale domain and the shape and compactness parameter combinations. The user expects either from a(n) supervised or unsupervised metric to efficiently (compared to trial-and-error parameter selection) provide a 'general notion' of the optimal scale parameter for a given target land use/cover class. The intriguing observation is that is it advantageous to invest an extensive amount of time to find the optimal parameter setting of the MRS algorithm. Our sensitivity analysis clearly exhibits that the ED2 metrics' discriminative capacity diminishes near or after the optimal scale setting. Thus, the question arises that is it necessary to go over all the segmentation cycles even after approaching the ED2-elected optimal scale setting. In other words, it is possible to manipulate intelligently the ED2 metric in which the segmentation cycles automatically terminates when the metric fails to distinguish significantly different scale groups. This is a way of adapting a supervised metric to a self-tuning state with respect to scene contents and user-defined reference objects.

4.4. Conclusion

There is a growing interest of optimizing eCognition Developer's multiresolution segmentation algorithm's parameters, especially the scale parameter, by means of supervised and unsupervised methods as opposed to the commonly-practiced trial-and-error approach. In the case of supervised methods, the reliability of the optimal-parameter prediction heavily relies on the sensitivity of the segmentation quality metric. In this study, we selected the Euclidean distance 2 (ED2) metric, a recently proposed supervised metric that measures dissimilarity between a reference polygon and an image object candidate, as a candidate to investigate the validity and efficacy of empirical discrepancy measures for finding the optimal scale parameter setting of the multiresolution segmentation algorithm. We chose test image scenes from three different sensors with varying spatial resolutions and scene contents and systematically segmented them using the multiresolution segmentation algorithm at a series of parameter settings (scale, shape, and compactness). The discriminative capacity of the ED2 metric across different scales groups was tested using non-parametric statistical methods. Our results showed that the ED2 metric significantly discriminates the quality of image object candidates at smaller scale values but it loses the sensitivity at larger scale values. This questions the meaningfulness of the ED2 metric in the MRS algorithms parameter optimization. Our contention is that the ED2 metric provides some notion of the optimal scale parameter at the expense of time. It is true that image objects serve as basic building blocks of subsequent classification steps, however, by design, the GEOBIA provides rich opportunities to refine image object candidates in a cyclic evolutionary manner until the user expectations are met. In this respect, there is always a compromise between

investing time for fine tuning the parameter setting of segmentation algorithm based on a supervised metric and for refining satisfactory image object candidates through approaches like class modeling. In future research, we aim to investigate the sensitivity of other supervised metrics that have been proposed for finding the optimal scale parameter of the multiresolution segmentation algorithm.

REFERENCES CITED

Acharya, T., Ray, A.K., 2007. Image Processing: Principles and Applications. Wiley Publishing Inc.

Addink, E.A., De Jong, S.M. and Pebesma, E.J., 2007. The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery. Photogrametric Engineering and Remote Sensing, 72(8): 905-912.

Aiazzi, B., Alparone, L., Baronti, S. and Garzelli, A., 2002. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. IEEE Transactions on Geoscience and Remote Sensing, 40(10): 2300-2312.

Aiazzi, B., Baronti, S. and Selva, M., 2007. Improving Component Substitution Pansharpening Through Multivariate Regression of MS +Pan Data. IEEE Transactions on Geoscience and Remote Sensing, 45(10): 3230-3239.

Al-Khudairy, D.H., Caravaggi, I. and Glada, S., 2005. Structural damage assessments from Ikonos data using change detection, object-oriented segmentation, and classification techniques. Photogrammetric Engineering & Remote Sensing, 71(7): 825-837.

Alparone, L., Aiazzi, B., Baronti, S., Garzelli, A. and Nencini, F., 2006. Information-theoretic image fusion assessment without reference, Image Information Mining for Security and Intelligence. ESA-EUSC 2006Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P. and Bruce, L., 2007. Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest. IEEE Transactions on Geoscience and Remote Sensing, 45(10): 3012-3021.

Alparone, L., Baronti, S., Garzelli, A. and Nencini, F., 2004. A global quality measurement of pan-sharpened multispectral imagery. IEEE Geoscience and Remote Sensing Letters, 1(4): 313-317.

Ardila, J.P., Bijker, W., Tolpekin, V.A., Stein, A., 2012. Context-sensitive extraction of tree crown objects in urban areas using VHR satellite images. International Journal of Applied Earth Observation and Geoinformation 15(2012), 57-69.

Ashraf, S., Brabyn, L. and Hicks, B.J., 2012. Image data fusion for the remote sensing of freshwater environments. Applied Geography, 32(2): 619-628.

Baatz, M., Schäpe, M., 2000. Multiresolution segmentation - An optimization approach for high quality multi-scale image segmentation. In: Strobl, J.,Blaschke, T., Griesebner, G. (Eds.), Angewandte Geographische Informations-Verarbeitung XII. Wichmann Verlag, Karlsruhe, pp. 12-23.

Benjamini, Y. and Yekutieli, D., 2001. The Control of the False Discovery Rate in Multiple Testing under Dependency. The Annals of Statistics, 29(4): 1165-1188.

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

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

Blaschke, T., Lang, S., Hay, G.J., Castilla, G. and Hay, G.J., 2008. Image objects and geographic objects. In: W. Cartwright, G. Gartner, L. Meng and M.P. Peterson (Editors), Object-Based Image Analysis. Lecture Notes in Geoinformation and Cartography. Springer Berlin Heidelberg, pp. 91-110.

Blaschke, T., Strobl, J., 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. GIS _ Zeitschrift für Geoinformationssysteme 14 (6), 12-17.

Burby, R.J. 1998. Natural Hazards and land use: An introduction. In: Cooperating with Nature: Confronting natural Hazards with Landuse Planning for Sustainable Communities. Washington DC. Joseph Press.

Burnett, C. and Blaschke, T., 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. Ecological Modelling, 168(3): 233-249.

Canny, J., 1986. A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8(6): 679-698.

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

Chabrier, S., Emile, B., Rosenberger, C., Laurent, H., 2006. Unsupervised performance evaluation of image segmentation. Journal on Applied Signal Processing 2006(1), 1-12.

Charles, J.J., Kuncheva, L.I., Wells, B. and Lim, I.S., 2006. An evaluation measure of image segmentation based on object centres. Image Analysis and Recognition, Pt(1).

Chavez, P.S., Sides, S.C. and Anderson, J.A., 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. Photogrammetric Engineering & Remote Sensing, 57(3): 295-303.

Cheema, U., 2007. Expert systems for earthquake damage assessment, IEEE Aerospace and Electronic Systems Magazine, 22(9):2403-2420

Chen, Y., Su, W., Li, J. and Sun, Z., 2009. Hierarchical object oriented classification using very high resolution imagery and LIDAR data over urban areas. Advances in Space Research, 43(7): 1101-1110.

Chiroiu, L., 2005. Damage Assessment of the 2003 Bam, Iran, Earthquake Using Ikonos Imagery. Earthquake Spectra, 21(S1): S219-S224.

Civco, D. and Witharana, C., 2012. Assessing the spatial fidelity of resolution-enhanced imagery using fourier analysis: a proof-of-concept study. SPIE Remote Sensing, Edinburg, United Kingdom Paper no. 8538-3.

Civco, D., Chabaeva, A. and Parent, J., 2009. KH-series satellite imagery and Landsat MSS data fusion in support of assessing urban land use growth. In: M. Ulrich and L.C. Daniel (Editors). SPIE, pp. 747-80I.

Clinton, N., Holt, A., Scarborough, J., Yan, L. and Gong, P., 2010. Accuracy assessment measures for object-based image segmenation goodness. Photogrametric Engineering and Remote Sensing, 76(3): 289-299.

Corban, C. et al., 2011. A Comprehensive Analysis of Building Damage in the January 12, 2010 M w 7 Haiti Earthquake using High-Resolution Satellite and Aerial Imagery. Photogrammetric Engineering & Remote Sensing, 77(9): 0997-1009.

Costa, G.A.O.P., Feitosa, R.Q., Cazes, T.B., and Feijo, B. 2008. Genetic adaption of segmentation parameters. In: Blaschke, T., Hay, G.,v& Lang, S., eds. Object-Based Imagev Analysis – Spatial Concepts for Knowledge-Driven Remote Sensing Applications. Lecture Notes in Geoinformation & Cartography, 18,v Berlin, Springer, pp. 679–695.

Costantini, M., Farina, A. and Zirilli, F., 1997. The fusion of different resolution SAR images. Proceedings of the IEEE, 85(1): 139-146.

Cutter, S.L., 2003. GI Science, Disasters, and Emergency Management. Transactions in GIS, 7(4): 439-446.

De Bethune, S., Muller, F. and Donnay, J.P., 1998. Fusion of multispectral and panchromatic images by local mean and variance matching filtering techniques, Fusion of Earth Data. IRS–1D Users Handbook Hyerabad: National Remote Sensing Agency, Sophia Antipolis, Nice, France, pp. 31-37.

Dell`Acqua, F. and Polli, D.A., 2011. Post-event Only VHSR Radar Satellite Data for Automated Damage Assessment: A Study pn COSMO/SkyMed and the 2010 Haiti Earthquake. Photogrametric Engineering and Remote Sensing, 77(10): 1037-1043.

Dey, V., Zhang, Y. and Zhong, M., 2010. A review on image segmentation techniques with Remote Sensing perspective. In: W. W. and B. Székely (Editors), ISPRS TC VII Symposium – 100 Years ISPRS. IAPRS, Vienna, Austria, pp. Part 7A.

Dorren, L.K.A., Maier, B. and Seijmonsbergen, A.C., 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. Forest Ecology and Management, 183(13): 31-46.

Dragut, L. and Blaschke, T., 2006. Automated classification of landform elements using object-based image analysis. Geomorphology, 81(3-4): 330-344.

Du, Q., Younan, N.H., King, R.L. and Shah, V.P., 2007. On the Performance Evlaution of Pan-Sharpening Techniques. IEEE Geoscience and Remote Sensing Letters, 4(4): 518-522.

Duro, D.C., Franklin, S.E. and Dub, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sensing of Environment, 118(2012): 259-272.

Ehlers, M., Gähler, M. and Janowsky, R., 2003. Automated analysis of ultra high resolution remote sensing data for biotope type mapping: new possibilities and challenges. ISPRS Journal of Photogrammetry and Remote Sensing, 57(5-6): 315-326.

Ehlers, M., Klonus, S., Johan Ã…strand, P.r. and Rosso, P., 2010. Multi-sensor image fusion for pansharpening in remote sensing. International Journal of Image and Data Fusion, 1(1): 25-45.

Ehrlich, D., Guo, H.D., Molch, K., Ma, J.W. and Pesaresi, M., 2009. Identifying damage caused by the 2008 Wenchuan earthquake from VHR remote sensing data. International Journal of Digital Earth, 2(4): 309-326.

ERDAS, 2011. ERDAS Imagine electronic Help document. The Earth to Business Company, Atlanta, Georgia, USA.

Espindola, G.M., Camara, G., Reis, I.A., Bins, L.S., Monteiro, A.M., 2006. Parameter selection for region growing image segmentation algorithms using spatial autocorrelation. International Journal of Remote Sensing 27(14), 3035-3040.

Gamba, P., Dell'Acqua, F. and Odasso, L., 2007. Object-oriented building damage analysis in VHR optical satellite images of the 2004 Tsunami over Kalutara, Sri Lanka, Urban Remote Sensing Joint Event, 2007, pp. 1-5.

Gangkofner, U.G., Pradhan, P., S. and Holcomb, D., W., 2008. Optimizing the High-Pass Filter Addition Technique for Image Fusion. Photogrammetric Engineering & Remote Sensing, 74(9): 1107-1118.

Garzelli, A. and Nencini, F., 2005. Interband structure modeling for Pan-sharpening of very high-resolution multispectral images. Information Fusion, 6: 213-224.

Garzelli, A. and Nencini, F., 2005. Interband structure modeling for Pan-sharpening of very high-resolution multispectral images. Information Fusion, 6: 213-224.

Giada, S., De Groeve, T., Ehrlich, D. and Soille, P., 2003. Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. International Journal of Remote Sensing, 24(22): 4251 - 4266.

Giada, S., De Groeve, T., Ehrlich, D. and Soille, P., 2003. Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. International Journal of Remote Sensing, 24(22): 4251 - 4266.

Gibbons, J. D., and Chakraborti, S. 2011. Nonparametric Statistical Inference, 5th Ed., FL: Chapman & Hall/CRC Press, Taylor & Francis Group.

Gibbons, J. D., and Chakraborti, S. 2011. Nonparametric Statistical Inference, 5th Ed., FL: Chapman & Hall/CRC Press, Taylor & Francis Group.

Gillespie, A.R., Kahle, A.B. and Walker, R.E., 1987. Color enhancement of highly correlated images – II. Channel ratio and 'chromaticity' transformation techniques. Remote Sensing of Environment, 22: 343-365.

Goetz, A.F.H., Boardman, J.W. and Yuhas, R.H., 1992. Discrimination Among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper(SAM) Algorithm, 3rd Annual JPL Airborne Geoscience Workshop, pp. 147-149.

Gonzalez, R.C., Woods, R.E., Eddins, S.L., 2002. Digital Image Processing, second ed. Pearson Education Inc., Published as Prentice Hall

Gonzalez-Audicana, M., Otazu, X., Fors, O. and Seco, A., 2005. Comparison between Mallat's and the 'a` trous' discrete wavelet transform based algorithms for the fusion of multispectral and panchromatic images. International Journal of Remote Sensing, 26(3): 595-614.

Goodchild, M.F., 2006. GIS and disasters: Planning for catastrophe. Computers, Environment and Urban Systems, 30(3): 227-229.

Gregg C. E., and Houghton B.F., 2006. Natural Hazards. In: Disaster Resilience: An Integrated Approach, Paton, D., and Johnson, D. (eds), pp 19-39, Chalres C. Thomas Publsihers, Springfiled.

Guo, Q., Chen, S., Leung, H. and Liu, S., 2010. Covariance intersection based image fusion technique with application to pansharpening in remote sensing. Information Sciences, 180(18): 3434-3443.

Gusella, L., Adams, B.J., Bitelli, G., Huyck, C.K. and Mognol, A., 2005. Object-Oriented Image Understanding and Post-Earthquake Damage Assessment for the 2003 Bam, Iran, Earthquake. Earthquake Spectra, 21(S1): S225-S238.

Hagenlocher, M., Lang, S. and Tiede, D., 2012. Integrated assessment of the environmental impact of an IDP camp in Sudan based on very high resolution multi-temporal satellite imagery. Remote Sensing of Environment, 126(0): 27-38.

Hagenlocher, M., Lang, S. and Tiede, D., 2012. Integrated assessment of the environmental impact of an IDP camp in Sudan based on very high resolution multi-temporal satellite imagery. Remote Sensing of Environment, 126(2012): 27-38.

Hallada, W.A. and Cox, S., 1983. Image sharpening for mixed spatial and spectral resolution satellite systems, 17th International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, pp. 1023-1032.

Hay, G.J., Blaschke, T., Marceau, D.J. and Bouchard, A., 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. ISPRS Journal of Photogrammetry and Remote Sensing, 57(5-6): 327-345.

Hay, G.J., Castilla, G., 2008. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. In: Blaschke, T., Lang, S., Hay, G. (Eds.), Object Based Image Analysis. Springer, Heidelberg, Berlin, New York, pp. 93_112.

Hay, G.J., Castilla, G., Wulder, M.A. and Ruiz, J.R., 2005. An automated object-based approach for the multiscale image segmentation of forest scenes. International Journal of Applied Earth Observation and Geoinformation, 7(4): 339-359.

Hill, R.A. and Broughton, R.K., 2009. Mapping the understorey of deciduous woodland from leaf-on and leaf-off airborne LiDAR data: A case study in lowland Britain ISPRS Journal of Photogrammetry and Remote Sensing, 64(2): 223-233.

Hill., A.A., Keys-Mathews, L.D., Adams, B.J., Podolsky, D. 2006. remote sensing and recovery: A case study on the Gulf Coasts of the United States. 4th International workshop on Remote Sensing for Post-disaster Response. Cambridge, UK.

Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P., Bunke, H.,Goldgof, D., Bowyer, K., Eggert, D., Fitzgibbon, A. and Fisher,R., 1996. An experimental comparison of range image segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(7), pp. 673–689.

Huang, Z. and Jia, X., 2012. Integrating remotely sensed data, GIS and expert knowledge to update object-based land use/land cover information. International Journal of Remote Sensing, 33(4): 905-921.

Hussain, E., Ural, S., Kim, K., Fu, C. and Shan, J., 2011. Building Extraction and Rubble Mapping for City Port-au-Prince Post-2010 Earthquake with GeoEye-1 Imagery and Lidar Data. Photogrammetric Engineering & Remote Sensing, 77(10): 1011-1023.

IPCC 2011.Summary for Policymakers. In: Intergovernmental Panel on Climate Change Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Ito, A., 2005. Issues in the implementation of the International Charter on Space and Major Disasters. Space Policy, 21(2): 141-149.

Jensen, J. R. , 2005. Introductory Digital Image Processing: A Remote Sensing Perspective (3rd edn).Prentice-Hall, Upper Saddle River, NJ.

Jing, L. and Cheng, Q., 2011. An image fusion method for misaligned panchromatic and multispectral data. International Journal of Remote Sensing, 32(4): 1125-1137.

Johnson, B., Xie, Z., 2011. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. ISPRS Journal of Photogrammetry and Remote Sensing 66(4), 473-483.

Joyce, K., Wright, K., Ambrosia, V. and Samsonov, S., 2010. Incorporating remote sensing into emergency management. The Australian Journal of Emergency Management, 25(4): 14-23.

Joyce, K.E., Wright, K.C., Samonsov, S.V. and Ambrosia, V.G., 2009. Remote sensing and the disaster management cycle, Advances in Geoscience and Remote Sensing. In: G. Jedlovec (Editor), Advances in Geoscience and Remote Sensing. pp. 317-346.

Jyothi, B.N., Babu, G.R. and Murali Krishna, L.V., 2008. Object Oriented and Multi-Scale Image Analysis: Strengths, Weaknesses, Opportunities and Threats-A Review. Journal of Computer Science, 4(9): 706-712.

Kalpoma, K.A. and Kudoh, J., 2007. Image fusion processing for IKONOS 1-m color imagery. IEEE Trans. Geosci. Remote Sens., vol. 45, no. 10, 3075-3086,

Karathanassi, V., Kolokousis, P. and Ioannidou, S., 2007. A comparison study on fusion methods using evaluation indicators. International Journal of Remote Sensing, 28(10): 2309-2341.

Kaya, G.T., Musaogglu, and Ersoy, O.K., 2011. Damage Assessment of 2010 Haiti Earthquake with Post-Earthquake Satellite Image by Support Vector Selection and Adaptation. Photogrametric Engineering and Remote Sensing, 77(10): 1025-1035.

Kim, M., Holt, J.B. and Madden, M., 2011. Comparison of Global- and Local-scale Pansharpening for Rapid Assessment of Humanitarian Emergencies. Photogrammetric Engineering & Remote Sensing, 77(1): 51-63.

Kim, M., Holt, J.B. and Madden, M., 2012. Comparison of Global- and Local-scale Pansharpening for Rapid Assessment of Humanitarian Emergencies. Photogrammetric Engineering & Remote Sensing, 77(1): 51-63.

Kim, M., J.B. Holt, R.J. Eisen, K. Padgett, W.K. Reisen and J.B. Croft. 2011. Detection of swimming pools by geographic object-based image analysis to support West Nile Virus control efforts, Photogrammetric , Engineering & Remote Sensing, 77(11):1169-1179.

Kim, M., Madden, M., 2006. Determination of optimal scale parameter for alliance-level forest classification of multispectral Ikonos images, Proceedings of the 1st International Conference on Object-based Image Analysis, Salzburg, Austria.

Kim, M., Warner, T.A., Madden, M., Atkinson, D.S., 2011. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. International Journal of Remote Sensing.

Klonus, S. and Ehlers, M., 2007. Image Fusion Using the Ehlers Spectral Characteristics Preservation Algorithm. GIScience & Remote Sensing, 44(2): 93-116.

Kouchi, K. and Yamazaki, F., 2005. Damage detection based on object-based segmentation and classification from high-resolution satellite images for the 2003 Boumerdes, Algeria earthquake, Proceeding ACRS 2005. CD-ROM, Hanoi, Vietnam, pp. 7-11.

Kranz, O. et al., 2010. GMES services for conflict prevention and mitigation: supporting mission planning of DG RELEX. In: S. Zlatanova (Editor), Geoinformation and Cartography for Early Warning and Crisis Response. NY Springer, New York, pp. 177-188.

Kruskal, W.H. and Wallis, W.A. 1952, Use of Ranks in One-Criterion Variance Analysis, Journal

Laben, C.A., Bernard, V. and Brower, W., 2000. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening., United States Patent Application 6011875.

Lang, S., Albrecht, F., Kienberger, S. and Tiede, D., 2008. Object validity for operational tasks in a policy context. Journal of Spatial Science, 55(1): 9-22.

Lang, S., Tiede, D., Holbling, D., Fureder, P. and Zeil, P., 2010. Earth observation (EO)-based ex post assessment of internally displaced person (IDP) camp evolution and population dynamics in Zam Zam, Darfur. International Journal of Remote Sensing, 31(21): 5709-5731.

Lang, S., Tiede, D., Holbling, D., Fureder, P. and Zeil, P., 2010. Earth observation (EO)-based ex post assessment of internally displaced person (IDP) camp evolution and population dynamics in Zam Zam, Darfur. International Journal of Remote Sensing, 31(21): 5709-5731.

Li, C. and Shao, G., 2012. Object-oriented classification of land use/cover using digital aerial orthophotography. International Journal of Remote Sensing, 33(4): 922-938.

Li, P., Xu, H. and Guo, J., 2010. Urban building damage detection from very high resolution imagery using OCSVM and spatial features. International Journal of Remote Sensing, 31(13): 3393 - 3409.

Li, Q. and Hu Q., 2004. 3D wavelet compression to multiple band remote sensing images based on edge reservation. ISPRS, Commission VII, Istanbul, CD-ROM, paper no.11.

Li, S., Kwok, J.T. and Wang, Y., 2002. Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images. Information Fusion, 3(1): 17-23.

Li, Z., Jing, Z., Yang, X. and Sun, S., 2005. Color transfer based remote sensing image fusion using non-separable wavelet frame transform. Pattern Recognition Letters, 26(13): 2006-2014.

Ling, Y., Ehlers, M., Usery, E.L. and Madden, M., 2007. FFT-enhanced IHS transform method for fusing high-resolution satellite images. ISPRS Journal of Photogrammetry and Remote Sensing, 61(6): 381-392.

Ling, Y., Ehlers, M., Usery, E.L. and Madden, M., 2008. Effects of spatial resolution ratio in image fusion. International Journal of Remote Sensing, 29(7): 2157-2167.

Liu, Y. et al., 2012. Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis. ISPRS Journal of Photogrammetry and Remote Sensing, 68(0): 144-156.

Lu, P., Stumpf, A., Kerle, N., Casagli, N., 2011. Object-oriented change detection for landslide rapid mapping. IEEE Geoscience and Remote Sensing Letters 8, 701-705.

Lucieer, A. and Stein, A., 2005. Texture-based landform segmentation of LiDAR imagery. International Journal of Applied Earth Observation and Geoinformation, 6(3-4): 261-270.

MacEachren, A.M. and Cai, G., 2006. Supporting group work in crisis management: visually mediated human ? GIS ? human dialogue. Environment and Planning B: Planning and Design, 33(3): 435-456.

MacEachren, A.M. and Cai, G., 2006. Supporting group work in crisis management: visually mediated human ? GIS ? human dialogue. Environment and Planning B: Planning and Design, 33(3): 435-456.

Makarau, A., Palubinskas, G. and Reinartz, P., 2012. Analysis and selection of pan-sharpening assessment measures. Journal of Applied Remote Sensing, 6.

Marçal, A.R.S. and Rodrigues, A.S., 2009. A method for multi-spectral image segmentation evaluation based on synthetic images. Computers & Geosciences, 35(8): 1574-1581.

Marpu, P.R., 2009. Geographic Object-based Image Analysis, technische Universitat Bergakademie Freiberg, Freiberg, 135 pp.

Marpu, P.R., Neubert, M., Herold, H. and Niemeyer, I., 2010. Enhanced evaluation of image segmentation results. Journal of Spatial Science, 55(1): 55 - 68.

Martha, T.R., Kerle, N., van Westen, C.J., Jetten, V., Kumar, K.V., 2011. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. IEEE Transactions on Geoscience and Remote Sensing 49, 4928-4943.

Meinel, G. and Neubert, M. 2004. A comparison of segmentation programs for high resolution remote sensing data. International Archives of the ISPRS 35 (Part B, Commission 4), pp. 1097–1105.

Möller, M., Lymburner, L. and Volk, M., 2007. The comparison index: A tool for assessing the accuracy of image segmentation. International Journal of Applied Earth Observation and Geoinformation, 9(3): 311-321.

Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S. and Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery, Remote Sensing of Environment. doi:10.1016/j.rse.2010.12.017.

Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S. and Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery, Remote Sensing of Environment. doi:10.1016/j.rse.2010.12.017.

Neubert, M., Herold, H. and Meinel, G. 2006. Evaluation of remote sensing image segmentation quality – further results and concepts. Int Arch Photogram Rem Sens Spatial Inform Sci XXXVI-4/C42, 6 pp.

Neubert, M., Herold, H., and Meinel, G. 2008. Assessing image segmentation quality - concepts, methods and application. In: Blaschke, T., Hay, G., & Lang, S., eds. Object-Based Image Analysis – Spatial Concepts for Knowledge-Driven Remote Sensing Applications. Lecture Notes in Geoinformation & Cartography 18, Berlin, Springer, pp. 769–784.

Nikolakopoulos, K.G., 2008. Comparison of nine fusion techniques for very high resolution data. Comparison of nine fusion techniques for very high resolution data, Photogrammetric Engineering & Remote Sensing ,74(5): 647-660.

of the American Statistical Association, Vol 47, No. 260, 583–621.

Pagot, E. and Martino, P., 2007. Post-conflict Urban Reconstruction Assessment Using Image Morphological Profile on 1-m-Resolution Satellite Data in a Fuzzy Possibilistic Approach Application: Koidu, Sierra Leone, Africa. European Commission.

Pal, N.R. and Pal, S.K., 1993. A review on image segmentation techniques. Pattern Recognition, 26(9): 1277-1294.

Pesaresi, M. and Pagot, E., 2007. Post-conflict reconstruction assessment using image morphological profile and fuzzy multicriteria approach on 1-m- resolution satellite data; Application test on the Koidu village in Sierra Leone, Africa, Urban Remote Sensing Joint Event, 2007, pp. 1-8.

Pham, D.L., Xu, C. and Prince, J.L., 2000. Current methods in medical image segmentation. Annual Review of Biomedical Engineering, 2(1): 315-337.

Pohl, C. and Van Genderen, J.L., 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. International Journal of Remote Sensing, 19(5): 823-854.

Pradhan, P.S., King, R.L., Younan, N.H. and Holcomb, D.W., 2006. Estimation of the Number of Decomposition Levels for a Wavelet-Based Multiresolution Multisensor Image Fusion. , IEEE Transactions on Geoscience and Remote Sensing, 44(12): 3674-3686.

Pu, R., Landry, S. and Yu, Q., 2011. Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery. International Journal of Remote Sensing, 32(12): 3285-3308.

Quinn, G. P., Keough, M.J., 2002. Experimental design and data analysis for biologists, Cambridge, United Kingdom, pp 32-54.

Radoux, J., Bogaert, P., Fasbender, D. and Defounrny, P. 2011. Thematic accuracy assessment of geographic object-based image classification. International Journal of Geographical Information Science 25(6): 895-911.

Ranchin, T. and Wald, L., 2000. Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation. Photogrammetric Engineering & Remote Sensing, 66(1): 49-61.

Ranchin, T., Aiazzi, B., Alparone, L., Baronti, S. and Wald, L., 2003. Image fusion--the ARSIS concept and some successful implementation schemes. ISPRS Journal of Photogrammetry and Remote Sensing, 58(1-2): 4-18.

Renyuan, Q., Xiaoling, Z. and Baoguo, W., 2009. An Improved Spatial/Spectral-Resolution Image Fusion Method Based on IHS-MC for IKONOS Imagery, Image and Signal Processing, 2009. CISP '09. 2nd International Congress on, pp. 1-4.

Rogerson, P.A. 2006 Statistical methods for Geography, pp-132-148, SAGE Publication, London

Roy, D. and Blaschke, T. 2010. International Charter and UNOSAT in Managing Disasters and Emergency Responses - An Initial Evaluation. Conference Proceedings of the third International Disaster and Risk Conference IDRC Davos 2010: Extended Abstract Collection, Global Risk Forum GRF Davos, Davos, Switzerland, pp 621-624.

Saito, K., Spence, R.J.S., Going, C. and Markus, M., 2004. Using High-Resolution Satellite Images for Post-Earthquake Building Damage Assessment: A Study Following the 26 January 2001 Gujarat Earthquake. Earthquake Spectra, 20(1): 145-169.

Schiewe, J., 2002. Segmentation of high-resolution remotely sensed data-concepts, applications and problems. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Ottawa, Canada, Vol. XXXIV, Part 4, pp. 358-363

Shettigara, V.K., 1992. A Generalized Component Substitution Technique for Spatial Enhancement of Multispectral Images Using a Higher Reolsution Data Set. Photogrammetric Engineering & Remote Sensing, 58(5): 561-567.

Siddiqui, Y., 2003. The modified IHS method for fusing satellite imagery, ASPRS 2003 Annual Conference. American Society for Photogrammetry and Remote Sensing, Anchorage, Alaska, CD-ROM.

Siegel, S., and Castellan, N. J., 1988. Non parametric statistics for the behavioral sciences. MacGraw Hill Int., New York. pp 213-214

Smith, G.M. and Morton, R.D., 2010. Real worl objects in GEOBIA through the exploitation of existing digital cartography and image segmenation. Photogrametric Engineering and Remote Sensing, 76(2): 163-171.

Smith, G.M. and Morton, R.D., 2010. Real world objects in GEOBIA through the exploitation of existing digital cartography and image segmenation. Photogrametric Engineering and Remote Sensing, 76(2): 163-171.

Song, C., Dickinson, M.B., Su, L., Zhang, S., Yaussey, D., 2010. Estimating average 548 tree crown size using spatial information from Ikonos and QuickBird images: 549 Across-sensor and across-site comparisons. Remote Sensing of Environment 114 (5), 1099-1107.

Sonka, M., Hlavac, V., Boyle, R., 2008. Image Processing Analysis and Machine Vision. Router Learning. Thomson Cooperation.

Stryker, Timothy and Jones, Brenda.2009. Disaster Response and the International Charter Program, Photogrammetric Engineering and Remote Sensing, 75 (12), 1342-1344.

Sturm, U., and Weidner, U. 2009. Further investigations on segmentation quality assessment for remote sensing applications. ISPRS Hannover Workshop 2009 High-Resolution Earth Imaging for Geospatial Information, Hannover, Germany, June 2–5, available at: http://www.ipi.uni-hannover.de/fileadmin/ institut/pdf/isprs-hannover2009/Sturm-144. pdf.

Taubenböck, H, Post, J, Roth, A, Zosseder, K, Strunz, G and Dech, S. 2008. A conceptual vulnerability and risk framework as outline to identify capabilities of remote sensing', Natural Hazards and Earth System Sciences, 8(3), 409-420.

Tian, J. and Chen, D.M., 2007. Optimization in mult-scale segmentation of high-resolution satellite images for artificial feature recognition. International Journal of Remote Sensing, 28(20): 4625-4644.

Tiede, D., Lang, S., Füreder, P., Hölbling, D., Hoffmann, C. and Zeil, P., 2011. Automated damage indication for rapid geospatial reporting. An operational object-based approach to damage density mapping following the 2010 Haiti earthquake. Photogrammetric Engineering & Remote Sensing, 77 (9), 933-942.

Tiede, D., Lang, S., Hölbling, D. and Füreder, P., 2010. Transferability of OBIA rulesets for IDP camp analysis in Darfur, GEOBIA 2010, Ghent, Belgium.

Tong, H., Maxwell, T., Zhang, Y. and Dey, V., 2012. A supervised and fuzzy-based approach to determine optimal multi-resolution image segmenation parameters. Photogrametric Engineering and Remote Sensing, 78(10): 1029-1043.

Trias-Sanz, R., 2005. A metric for evaluating and comparing hierarchical and multi-scale image segmentations, Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International, pp. 5647-5650.

Trimble Germany GmbH, T.G., 2011. eCognition Developer 8.7.2 Reference Book, Trimble Germany GmbH, Germany.

Tu, T.-M., Su, S.-C., Shyu, H.-C. and Huang, P.S., 2001. _A new look at IHS-like image fusion methods. Information Fusion, 2: 177-186.

UN. 2011.. Global Assessment Report on Disaster Risk Reduction. Revealing Risk, Redefining Development. United Nations Publications.

UNEP., 2011. Livelihood security: Climate change, migration and conflict in the Sahel. Nairobi, Kenya.

UNISDR, 2004.Terminology: Basic terms of disaster risk reduction 2004, UNISDR (United Nations International Stratergy for Disaster Risk Reduction), <http://www.unisdr.org/eng/terminology/terminology-2004-eng.html>, Accessed on January 2009.

Unnikrishnan, R., Pantofaru, C. and Hebert, M., 2007. Toward Objective Evaluation of Image Segmentation Algorithms. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(6): 929-944.

Vijayaraj, V., Nicolas, H.Y. and Charles, G.O.H., 2006. Quantitative analysis of pansharpened images. Optical Engineering, 45(4): 046202.

Voigt, S. et al., 2007. Satellite Image Analysis for Disaster and Crisis-Management Support. Geoscience and Remote Sensing, IEEE Transactions on, 45(6): 1520-1528.

Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K.. and Mehl, H., 2007. Satellite Image Analysis for Disaster and Crisis-Management Support. Geoscience and Remote Sensing, IEEE Transactions, 45(6): 1520-1528.

Voigt, S., Schneiderhan, T., Twele, A., Gahler, M., Stein, E. and Mehi, H., 2011. Rapid Damage Assessment and Situation Mapping: Learning from 2010 Haiti Earthquake. Photogrammetric Engineering & Remote Sensing, 77(9): 923-931.

Vrabel, J., 1996. Multispectral Imagery Band Sharpening Study. Photogrammetric Engineering & Remote Sensing, 62(9): 1075-1083.

Vrabel, J., Doraiswamy, P., McMurtrey, J. and Stern, A., 2002. Demonstration of the accuracy of improved resolution hyperspectral imagery SPIE, pp. 556-567.

Vu, T.T., Masashi, M. and Fumio, Y., 2005. Detection and Animation of Damage Using Very High-Resolution Satellite Data Following the 2003 Bam, Iran, Earthquake. Earthquake Spectra, 21(S1): S319-S327.

Vu, T.-T., Yamazaki, F. and Matsuoka, M., 2007. Context-based detection of post-disaster damaged buildings in urban areas from satellite images, Urban Remote Sensing Joint Event, pp. 1-5.

Vu, T.T., Yamazaki, F. and Matsuoka, M., 2009. Multi-scale solution for building extraction from LiDAR and image data. International Journal of Applied Earth Observation and Geoinformation, 11(4): 281-289.

Vu, T.T., Yamazaki, F. and Matsuoka, M., 2009. Multi-scale solution for building extraction from LiDAR and image data. International Journal of Applied Earth Observation and Geoinformation, 11(4): 281-289.

Wald, L. and Ranchin, T., 1997. Fusion of satellite images of different spatial resolutions: Asssessing the quality of resulting images. Photogrammetric Engineering & Remote Sensing, 63(6): 691-699.

Wald, L., 2000. Quality of high resolution synthesised images: Is there a simple criterion ?. In: T. Ranchin and L. Wald (Editors), Fusion of Earth data: merging point measurements, raster maps and remotely sensed images. SEE/URISCA,Nice, France, Sophia Antipolis, France, pp. 166.

Wald, L., 2002. Data Fusion: Definitions and Architectures - Fusion of Images of Different Spatial Resolutions. Les Presses, Ecole des Mines de Paris,, Paris, France, 200 pp.

Wang, Z. and VBovik, A.C., 2002. A universal image quality index. IEEE Signal Processing Letters, 9: 81-84.

Wang, Z. and VBovik, A.C., 2002. A universal image quality index. IEEE Signal Processing Letters, 9: 81-84.

Wang, Z., Bovik, A.C., Sheikh, H. and Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(1): 600-612.

Weidner, U., 2008. Contribution to the assessment of segmentation quality for remote sensing applications. International Archives of Photogrammetry and Remote Sensing. vXXXVII iPart B7. 479-484.

Weidner, U., 2010. Pansharpening – relating quantitative quality measures to impact on results of subsequent processing steps. In: W. W. and B. Székely (Editors), ISPRS TC VII Symposium – 100 Years ISPRS. IAPRS, Vienna, Austria.

Weidner, U., 2010. Pansharpening – relating quantitative quality measures to impact on results of subsequent processing steps. In: W. W. and B. Székely (Editors), ISPRS TC VII Symposium – 100 Years ISPRS. IAPRS, Vienna, Austria.

Witharana, C. 2012. Who does what where? Advanced earth observation for humanitarian crsisi management. 6th IEEE International Conference on Information and Automation for Sustainability. Bejing, China. .Paper no. ICIAfS'12 1569613211

Witharana, C. and Civco, D.L., 2012. Evaluating remote sensing image fusion algorithms for use in humanitarian crisis management. SPIE Remote Sensing, Edinburg, United Kingdom, Paper no. 8538-5.

Witharana, C. and Civco, D.L., 2012. Evaluating remote sensing image fusion algorithms for use in humanitarian crisis management. SPIE Remote Sensing, Edinburg, United Kingdom, Paper no. 8538-5.

Witharana, C. Meyer, T.W. Civco, D.L, Osleeb, J. 2010. Developing a new ArcGIS tool to quantify building-content vulnerability from storm-surge inundation. ASPRS Annual Conference 2010, California, USA.

Witharana, C., Civco, D.L. and Meyer, T.H., 2013. Evaluation of pansharpening algorithms in support of earth observation based rapid-mapping workflows. Applied Geography, 37(0): 63-87.

Witharana, C., Civco, D.L. and Meyer, T.H., 2013. Evaluation of pansharpening algorithms in support of earth observation based rapid-mapping workflows. Applied Geography, 37(2013): 63-87.

Witharana., C. A GIS Approach to Estimate Storm Surge Damage Risk to Buildings in Groton, Connecticut", M.S Thesis, University of Connecticut, 2009.

Witharana.. C. and Meyer. T.H. 2010. Developing customized ArcGIS tools for disaster management. In Proc. of the IEEE ICIAfS10, Colombo, Sri Lanka.

Wu, B., Xiong, Z.G., Chne, Y.Z. and Zhao, Y., 2009. Classification of quickbird image with maximal mutual information feature selection and support vector machine. Procedia Earth Planetary Science, 1: 1165-1172.

Xydeas, C. and Petrovic, V., 2000. Objective image fusion performance measure. Electronic Letters, 36(4): 308-309.

Yakhdani, M.F. and Azizi, A., 2010. Quality assessment of image fusion techniques for multisensor high resolution satellite images (case study: IRS-p5 and IRS-p6 satellite images). In: W. W. and B. Székely (Editors), ISPRS TC VII Symposium – 100 Years ISPRS. IAPRS, Vienna, Austria, pp. Part7B.

Yamazaki, F., Vu, T.T. and Matsuoka, M., 2007. Context-based detection of post-disaster damaged buildings in urban areas from satellite images, Urban Remote Sensing Joint Event, 2007, pp. 1-5.

Yang, B., Kim, M. and Madden, M., Assessing Optimal Image Fusion Methods for Very High Spatial Resolution Satellite Images to Support Coastal Monitoring. GIScience & Remote Sensing, 49(5): 687-710.

Zerger, A., and Smith, D. I. 2003. Impediments to using GIS for real-time disaster decision support, Computers, Environment and Urban Systems, 27, 123-141.

Zhang, H., Fritts, J., Goldman, S., 2008. Image segmentation evaluation: A survey of unsupervised methods. Computer Vision and Image Understanding 110(2), 260-280.

Zhang, Y., 2002. A new automatic approach for effectively fusing Landsat-& as well as IKONOS images. IEEE Transactions on Geoscience and Remote Sensing: 2429-2431.

Zhang, Y., 2002. Problems in the fusion of commercial high-resolution satellite as well as Landsat 7 images and initial solutions, Symposium on Geospatial Theory, Processing and Aplications, Ottawa, Canada.

Zhang, Y., 2004. System and method for image fusion, United States Patent Application No. 2004014659.

Zhang, Y.J., 1996. A survey on evaluation methods for image segmentation. Pattern Recognition, 29(8): 1335-1346.

Zhang, Y.J., 1996. A survey on evaluation methods for image segmentation. Pattern Recognition, 29(8): 1335-1346.

Zhou, D.L. Civco, and J.A. Silander. 1998. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. International Journal of Remote Sensing 19(4):743–757.

Zhou, W., Troy, A., 2008. An object object oriented approach for analysing and 574 characterizing urban landscape at the parcel level. International Journal of Remote Sensing 29 (11): 3119-3135.

ZKII-DLR, 2013. Center for Satellite based Crisis Information (ZKI) of the German Aerospace Center (DLR). http://www.zki.dlr.de/