Information extraction from paper maps using object oriented analysis (OOA)

Mahmoud Ibrahim Mahmoud
19865-AES
Enschede, The Netherlands,
March, 2012.

Information extraction from paper maps using object oriented analysis (OOA)

Mahmoud Ibrahim Mahmoud
19865-AES
Enschede, The Netherlands,
March, 2012

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Disaster Geo-Information Management

Thesis Assessment Board:

Chairman:	Professor Dr. F. D. van der Meer
External Examiner:	Dr. R. van Beek - University of Utrecht
First supervisor:	Dr. N. Kerle
Second supervisor:	Dr. M. Gerke

DISCLAIMER

ABSTRACT

Historical topographic maps are distinct sources of spatial information for hind-cast studies. They are acclaimed to be one of the most reliable legacy archives representing and describing geographic features prior to aerial photography and the present day satellite imagery. However, two major challenges are encountered in extracting information from these sources. These challenges are conceptual and technical emanating from scanning artefacts, inherent map complexity and analogousness, although information extraction has been manually done through digitizing, pixel-based methods and visual map analysis, which are time consuming and tedious. Hence, there is urgent need to explore robust and reliable methods such as the object-oriented analysis (OOA) to efficiently develop new information extraction techniques for scanned topographic paper maps. Therefore, this research investigated and answered questions about conceptualization, development, implementation and transferability of an OOA-based information extraction method for complex papers maps and potential applications.

This study demonstrates the OOA-based information extraction technique on a 1967 topographic map of Nigeria. The work was structured broadly into three major parts. The first part investigated underpinning theoretical concepts of saliency and semantics to conceptualize a generic OOA-based information extraction framework. The second part consequently, translated the conceptualized procedure to develop multi-step object-centered rules and implemented the developed algorithm on input maps to robustly and accurately extract tangible thematic information, despite typical complexities known with paper maps. Thirdly, the research further probed how far OOA rules-sets can be transferred to comparable data sets of same series with slightly changed imaging conditions to see how robust and reliable will the once developed rule-sets perform. Results show the suitability of saliency and semantic to conceptualize and develop an OOA-based information extraction formalism for complex paper maps. Similarly, the created OOA rule-sets robustly and reliably extracted thematic information that corresponds to targeted map objects with accuracies of 95% for the hydrographic layer, 97% and 92% for correctness and completeness of symbols respectively, 70% for texts and 55 % for contour lines. Interestingly, transferability of the once created rule-set proved realistic after testing it on different map section of same map series with slight modifications. Empirical observation of the developed method reveals that OOA-based information procedure was swifter than manual method and is thus useful for environmental modelling/monitoring programs. More importantly the method is suited for applied earth's sciences especially disaster risk management programs where rapid understanding and mapping of multi-temporal hazards, element at risk and vulnerability assessment are increasingly demanded. Therefore, the increasing demand for quick insights on hazards and risk assessment over time are critical milestones that OOA-based method and extracted information can achieve since we can rapidly unlock such relevant information from historical paper maps using this approach.

Index Terms—Information extraction, paper maps, object oriented analysis, saliency, semantics, transferability, robustness and reliability

PREPOSITION

"If we have a promising method but merely talked about it and do not explore how far we can operationalize its utility, then such method is, however good, with no significance"

DEDICATION

To Allah for his grace and mercies and my parents Professor Ibrahim S. Mahmoud and Late Mrs. Maimunat I. Mahmoud

ACKNOWLEDGEMENTS

My genuine gratitude and appreciation first goes to Almighty Allah who has giving me life from cradle till this moment that I am able to do an MSc to develop my career in GIScience (GI) and Earth Observation (EO) at faculty ITC, University of Twente, Enschede Netherlands. Secondly, a grateful to NUFFIC for providing me with a fully funded fellowship to further my career dreams to obtain an MSc degree in GI & EO. Thirdly, I express my indebtedness to four seasoned academicians and mentors who supported my dream immeasurably 1). Professor Harold Annegarn of University of Johannesburg, South Africa with strong recommendation letters and his willingness to consider /(sponsor) me to actively participate in advanced learning opportunities as a volunteer student researcher under the university network for disaster risk reduction for Africa (UNEDRA) platform as well as taught me about scientific thinking despite my undergraduate level. 2). Professor Mohammed T. Usman of Federal University of Technology, (FUT) Minna, Nigeria with strong recommendation letters and his inspiring words of motivation to sail through though hurdles of being a good scientist on his notice of my dogged interest in research. 3). Dr. Halilu Shaba Ahmed of the National Space Research and Development Agency, (NARSDA), Nigeria with strong recommendation letters, providing me with the dataset used for this research and his immeasurable support in developing my critical scientific attitude since my undergraduate research endeavours while in (FUT) Minna, and 4). Dr. Tsehai Woldai of the ITC, University of Twente, for his confidence in supporting my candidature for the NUFFIC fellowship to do my MSc in GIScience towards building requisite skills young African scientists and UNEDRA.

At ITC, I like to express my appreciation to my supervisor (s) Dr. Norman Kerle of the Applied Earth Science department (Geohazards) stream and Dr. Markus Gerke of the Earth Observation Science department. Their willingness to accept me as an MSc student researcher despite their busy schedules but also on critical and thorough scrutiny of research ideas and finished products by working through each chapter with me on how to think scientifically. In spite of their tight schedule, they were able to provide guidance, advice and help in this research project. I also really appreciate the confidence they showed in me for launching into a research area that was completely new to me. I am indeed indebted to your understanding and kindness for helping me throughout this study, from beginning to the submission of the thesis. Similarly, I am thankful to Shruthi Shruthi a Ph.D student who has listened to my research challenges and argument and offered useful suggestions in her capacity. Many thanks to the experts who responded to the manual boundary delineation questionnaire I served them to complete my thesis.

I would also like to extend my gratitude to the National Oil Spill Detection and Response Agency (NOSDRA) management team [Late Dr. Bamidele Ajakaiye] for his willingness to endorse my study application. Mrs. U. H Okwchime, Mr. O. A. Abe, Mr. I. O. Musa, Mr. D. A. Gidado, Mr. Rotimi, Mal. Aliyu Ringim and Mr. Okon] for granting me study leave to this MSc degree. In addition, my heartfelt gratitude goes to my colleagues Mal. Hamza Jakada, Hajiya. Rahmat Akerele & family - friendly support, Ms. Utuk Ekaete, Mr. Josiah and Mrs. Bilkisu Bello & family, of the Department of Planning Policy Analysis and Research for their support. Similarly, Mrs. Uju for her kind support while I was preparing to embark on study travel to Netherlands.

My deepest thanks go to my dad, Professor Ibrahim. S. Mahmoud, for his great comprehension, love, and prayers during this study. To my siblings Muinat, Mubarak, Muhammad-l Kabir, Ahmad, Abdul kadir, Habibullah, Amina, Safinatul-Najat, Kamalludeen and Adam thank you so much for enduring my absence. And my friends Mr. Usman Tosin Edun, Uztaz Hassan Hambal Babatinde & family, Mr. Kamar Edun & family, Mal. Yusuf Isah Ibrahim & family, Mal. Lukman Enilolobo and Mal. Abdullahi Abdul Majeed for their thoughtfulness during my study.

My special thanks go to Ms. Esther Shupel Ibrahim, Mr. John Essien, Ms. Mosunmola Adeniya, Mal. Ahmed Abubakar Sarki, Mr. Akinlapa Femi, Mr. Adegbe Momodu etc.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AA	-	Accuracy Assessment
DRM	-	Disaster Risk Management
DEM	-	Digital Elevation Model
ESP	-	Estimation of Scale Parameter
FSO	-	Feature Space Optimization
NN	-	Nearest Neighbourhood
OOA	-	Object Oriented Analysis
RGB	-	Red Green Blue
USGS	-	United States Geological Survey

CHAPTER 1

1 INTRODUCTION

This chapter describes the general overview of the study. Background concepts, motivation and challenges are presented. The rationale of information extraction from paper map to improve multi-temporal hazard and risk assessment using Object-Oriented Analysis (OOA) as well as software to be used is given. Overall and specific objectives are outlined and are further broken down into relevant research questions. Relevance and structure of the study is described and put in perspective of contemporary information extraction framework.

1.1 Background

Lately enormous demands for spatio-temporal information aimed at good governance and development planning such as environmental monitoring and disaster risk management (DRM) programs is on the increase. Consequently, geospatial information technologies have become integral to any of these comprehensive good governance plans.

Many datasets for this information needs exist and their availability is becoming quite easy and overwhelming in the recent past. Links to these pools of data sources is guaranteed through governmental agencies, census data archives, earth observation and monitoring missions like remote sensing of the environment etc.

Historically speaking, phenomenal changes due to strong human and earth system dynamics have been persistent and there is need to understand these changes in proper historic context. Although getting historic information is difficult, especially in countries that were colonized and where no good archiving was done, or no functional government after independence.

One of the most reliable historic sources of spatial information are paper maps such as the historic topographic or thematic maps that allow insight into the historic location, extent of built-up areas, elements at risk and hazard processes are analogue allowing visual information analysis such as multi-temporal land cover changes (McChesney & McSweeney, 2005). Nonetheless, the geographic information content portrayed in these maps is complex and crowded with high information density consisting of concatenated lines, characters, symbols and area objects making it not easily accessible.

Therefore there is urgent need to explore robust and reliable, state of the art methods such as the object oriented analysis (OOA) to accurately extract information from topographic paper map. This is because the realization of their potentials for detailed and comprehensive understanding of array of applications at global, regional, national and local scales. Robust here means the accuracy and completeness of OOA method in extracting same information type when transferred on a different map sheet of same map series without failure. While reliability refers to how users can depend on the OOA based information extraction method for various applications compared to existing methods.

1.2 Motivation

Human use of maps has a long and remarkable history and today they are used to visualize and satisfy a variety of information needs. While human cognition is achieved on this historic information source, digital information extraction remains a great challenge through the time consuming and error-prone traditional digitizing method (Dillabaugh et al. 2002; Ali & Algarni 2004). Consequently, Graeff (2006) reported that manual information extraction of map layers suffers from inconsistency and biased interpretation coupled with the labour intensiveness often requiring a lot of workforce for quick interpretation. Although concerted mapping efforts as those done for past hazard event are good options, but faster automated mapping method are now needed. Hence a comprehensive conceptual formalism, for developing robust and reliable algorithm for information extraction using OOA on complex topographic maps is the original motivation of doing this thesis. This knowledge driven information extraction approach was found to be an appealing and promising technique for automatic information extraction from paper maps. However, OOA methods have only been recently explored, and only on simple maps. Here OOA technique will be applied on complex topographic paper maps. The method works with images broken into homogeneous and corresponding segments that, in turn, are analyzed and classified, using not only the multispectral or colour information but also attributes such as size, shape, or texture inherent to these segments, e.g. Kerle & de Leeuw (2009) and (Leyk et al. 2006). The method is efficient in extracting information from topographic and thematic maps considering its advantages over traditional digitizing and pixel-based methods. With extraction, we mean here the object modelling, starting from spectral characteristics (colour), spatial attributes (geometric description, additional attributes, etc.) and semantic/functional properties or topologic information. The thesis focus is on topographic map and will consider structural analysis for knowledge extraction of various layers of spatially embedded objects such as vegetation symbols, extent of objects, associated texts and line objects detection. Key terminologies used will be intermittently defined where necessary and discussed perceptively in proceeding sections, sub-sections and chapters.

1.3 The concepts of sailency and semantics in information extraction challenges

The works of Kerle & de Leeuw (2009), Khotanzad & Zink (2003) as well as Leyk et al. (2006) have highlighted two notable broad challenges (technical and conceptual). These were the main focus this research intends to systematically tackle. Technically, challenges emanate from the scanning resolution of these maps leading to 1) aliasing, i.e. an effect induced by the scanner's point spread function that blurs map objects such as lines, dots, symbols and edges, leading a "bleeding" effect; 2) spacing of close objects, whereby higher spatial resolution lends partial solution, but at the expense of large file size thus increasing computer processing time; and 3) the manifestation of false colour generated in the digital image due optical misalignment typical in scanner head. Additionally, due to file-format issues specific artefacts can be introduced during document conversion, such as the well-known "jpeg effect" with the exception of lossless conversion techniques.

Conceptually, challenges are rooted to saliency and semantics concerns affecting feature extractability as a result of strong interaction of symbols, text and objects with each other originating from lower graphical quality of map production and printing methods. Saliency here is the principal concept which signifies the existence of non-ambiguous color or shape features attributed to a prototype map object class, and do not occur in all other classes. Such features allows for explicit detection of the complete set of objects of the considered map category. While semantics aids in distinguishing map objects with identical color or shape characteristics using their location or proximity to other features. For example, this allows the distinction of mangrove prototype symbols from marsh, areas liable to flood or river layer as well as the symbols in green color representing six different classes.

Before carrying out reliable geo-information based OOA information extraction on topographic paper maps for further analysis (e.g. environmental monitoring or DRM programs), the conceptual extraction formalism of the data themselves as well as those which relate to the intended method, have to be investigated. This means suitable theoretical concepts for information extraction on topographic paper maps have to be identified, conceptually understood and developed with regards to the map in question. This aspect will be extensively addressed in chapter two and three. Basically the first step will involve logical visual map interpretation, breaking the map content into information layers similar to the way paper map were traditionally prepared (e.g. water, symbol, text and lines layer). Then the second step is followed by explicitly defining the various information layer contained in each of the deconstructed map layers. Due to complexity inherent in the map it is essential to associate map elements meaningfully using semantics to establish logical map units. Also the different types of uncertainties have to be identified, conceptually understood and, if possible, quantitatively assessed to determine where automation is possible and restricted. Additionally, ways of using this knowledge of uncertainty for the improvement of the planned target application are needed. Finally, this thesis investigates such ways of conceptual and analytical treatment of topographic maps to extract information using saliency and semantic in an OOA context. Doing this research is an effort to further develop OOA method to investigate how far OOA based automatic information extraction capabilities can be applied on complex topographic maps. Thereby conserving man hours for complex tasks, minimizing bias information extraction problems faced using manual extraction methods and traditional classification approaches. This thesis desires to develop an OOA based method which is able to recognize and extract different information (symbols or pattern) represented on topographic maps within eCognition software environment.

1.4 Relevance of study

The importance of time in extracting relevant information needs for environmental resource management is indisputable. But inadequacies of technological support and available traditional methods are chiefly responsible for the monotonous challenges of extracting desired information (Worboys 1995). Extraction of map features e.g. area, symbols, text and line is sometimes performed using manual digitizing or pixel based methods for several reasons. As such, this has remained one of the fundamental impediments to meeting the geo-information science information extraction demands particularly in topographic maps where various types of map features often appear within a single color plane, and they touch, cross, and overlap. Considering the complexity of topographic map if features are represented as objects entities that correspond to the logical components of real objects, the information can be extracted and manipulated in a highly productive manner for applications and can be used for activities such as multi-temporal hazard and risk assessment. While on the other hand advancing information extraction for paper map in the OOA context which will reduce man hours invested in extracting relevant GIS information layer for intended applications. Developing such generic method will be more efficient in unlocking encoded information on complex topographic maps using OOA method. In the end the method could provide a launching pad for many other applications thereby establishing the capabilities of OOA and determining its strength and limitation in information extraction.

1.5 Objectives

According to the challenges described above, three specific objectives and nine research questions are considered to achieve the overall objective of this thesis. Below, specific objectives relevant research questions were briefly outlined. Questions I-II considers the theoretical basis of knowledge driven information extraction from paper maps. Question III-VI addresses how to translate human cognitive interpretation of paper maps into an OOA-based robust, standardized, less laborious, cost effective information extraction procedure. While questions VIII-IX are set to tackle accuracy assessment concerns and misclassification cost and penalty issues.

1.5.1 Overall objective

To further develop information extraction procedure on complex topographic paper maps by object oriented analysis (OOA) in eCognition software environment.

1.5.2 Specific objectives and research questions

1. Develop knowledge driven procedure for systematic information extraction of topographic paper maps.

I. *What are the suitable theoretical concepts for investigating, understanding and extracting thematic information contents depicted on historical topographic paper maps?*

II. *In what order should the information extraction procedure be applied considering the complexity of individual and composite map layer content and associations?*

2. To provide a conceptual frame work and formalism of an OOA-based approach to object information extraction for historical topographic paper maps

III. *How can we develop an OOA-based information extraction procedure for topographic paper maps?*

IV. *How can human cognitive interpretations be translated to develop reliable and robust OOA information methods and what modeling techniques can be applied in extracting geospatial objects from historical topographic paper maps?*

V. *What kind of measurement can be used for delineating geographic features? i.e. how to go from symbol to boundary.*

VI. *How important is transferability of the once created rule-set and method applicability to other maps from the same series, with the same symbology and imaging conditions but showing another region?*

VII. *How robust, consistent, unbiased will such rule sets be if transferred and compared to ambiguity of manual extraction?*

3 To compare the accuracy and completeness of the digital information extraction procedure with a manual extraction method.

VIII. *To what extent can OOA approach be applied to develop reliable and robust information extraction method and how accurate is the OOA method compared to manual information extraction?*

IX. *What is the cost and penalty of misclassifying objects wrongly?*

1.6 Research approach and structure of thesis

This thesis was conducted with the belief that employing state-of-the-art OOA (tools) to develop a robust and reliable procedure to extract relevant information from paper map using the concepts of saliency introduced by Rosin (1997), Smeulders et al. (2000) and demonstrated on paper maps (Leyk et al. 2006). Saliency and semantic modelling can lead to developing a formalism of extracting information behind spatially embedded objects on complex topographic maps. In this respect known complexities of information encoded on topographic maps are systematically unlocked which could be utilized to answer questions of significant societal or environmental concern, such as for multi-temporal hazards and risk assessment. The research will not test the utility of the extracted GIS ready layers but will rather focus on the core problem in the development of OOA-based information extraction method for unlocking thematic information behind spatially embedded objects due to their complexities. Essentially, the approach here was to use the concepts of saliency and semantics to initially identify the classes (object) at various stages of the modeling base on attributes and topological relation of spatially embedded object. Then use same concept to develop relevant rules for unlocking object information considering geometric, thematic and semantic attributes. For completeness the developed rules for object information extraction will be validated by transferring the same rules on separate map section to test for robustness and reliability of the proposed method. The thesis was organized into five chapters, introduction (Chapter-1), literature review (Chapter-2), method development and testing (Chapter-3), result presentation and discussion Chapter-4), conclusion recommendations and limitations (Chapter-5). However, it can be technically divided into three major parts Figure 1 firstly conceptualization of information extraction procedure is considered, secondly implementation of the developed method and thirdly the transferability of the method and algorithm on separate image of the same and similar image.

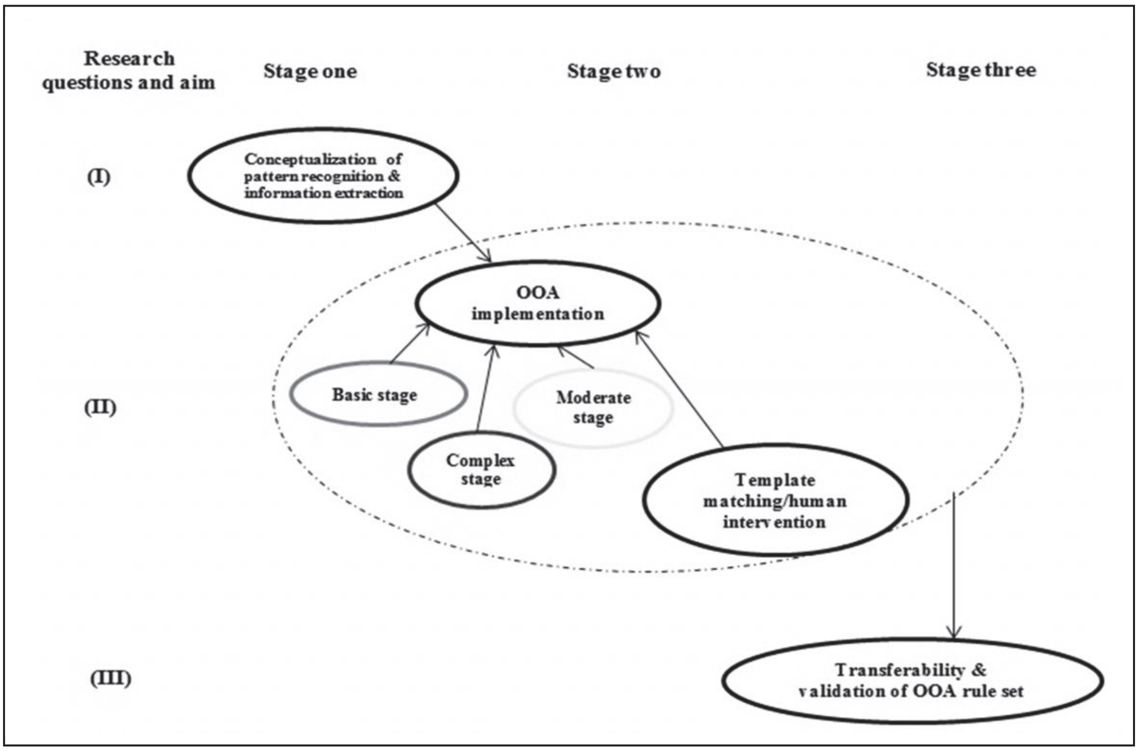


Figure 1: Overview of individual stages of this thesis.
Stage one is related to the conceptualization of pattern recognition and information extraction framework for paper maps in OOA in response to research questions I to II. Stage two encompasses the implementation of OOA using conceptual and analytical framework developed using saliency and semantic modelling approach (simple, moderate and complex) objects and focuses on research questions III to VI. Stage three is dedicated to testing the transferability of OOA method for validation of the rule set.

CHAPTER 2

2 LITERATURE REVIEW

In this chapter prevailing theoretical issues and recurrent trends found in the state of the art scientific literature aiming at semi-automatic and automatic information extraction are reviewed. In a broader sense, underpinning theoretical concept for developing OOA based information extraction framework for topographic paper maps are highlighted cascading into specific problem and related works in pattern analysis and cartographic recognition. Central problems, key methods and concepts relevant to the contributions of this thesis are outlined briefly to position them within the corresponding research areas.

2.1 Theoretical background and key issues

A paper map is a representation, usually on a flat paper surface, of a whole or part of an area. Paper maps are interesting real world artefacts to augment in part because they are themselves an information resource (Reilly et al. 2006). Similar to the uses of books, information presented on a paper map can be used and, as a launching point for queries by scanning to convert the analogue paper map to digital information (Koike et al. 2000). Many information extraction methods have been developed and are found to be appealing in remote sensing and GIS fields. Baltsavias (2004) presented a list of such and their area of applications. Simplest amongst these methods is the visual assessment of paper maps to extract relevant information, although their utility depends on a wide range of map qualities (scale, accuracy, unambiguous legend, content description and the map's georeferencing quality). Consequently, development and increased demand for advanced information extraction methods paved way for commercial digitizing tablets, introduced in 1979. This technique lent a more considerable and precise information extraction of point, line, or polygon information in geo-information science and was later replaced by on-screen digitizing of scanned maps to achieve greater accuracy. Nonetheless the extraction of information from scanned paper maps and remotely sensed imagery using on-screen digitization method is presently intensive, laborious, time-consuming and error-prone process requiring highly qualified remote sensing experts, making this information source expensive and slow. By 1980s these challenges stimulated works to further explore automatic procedures thus focused on pattern recognition, text separation and line tracing using different methods e.g. Goodson & Lewis (1988) extracted line features from line drawing images. Similarly, Kasturi & Alemany (1988) proposed automatic information extraction from paper-based maps. While, Chang et al. (1985) proposed generic image rule based information systems and Watson et al. (1984) traced lines from grey level digitizations of industrial drawings. Afterwards from 1990s until now information extraction methods have undergone refinement over time but the overall objectives never changed. However, Madry (2006) highlighted salient challenges of integrating extracted information from paper maps to contemporary GIS layers. Kerle & de Leeuw (2009) suggested that it is more useful to process the actual information represented in the map, as opposed to using visual interpretation or integrating a scanned paper map into raster format. Remarkably, automatic information extraction methods have remained an active field of research, and in the recent past more procedures have been developed for extraction of information from paper maps. However, these researches can be categorized in milestones [e.g. applied computing, machine/artificial intelligence for pattern recognition such as area recognition (polygon), structural pattern analysis (symbols), text separation (alphanumeric character), and line extraction (contours)]. Map interpretation and GIS analysis, as well as text / graphics separation and processing of map labels. In the following sub-section a transitory appraisal of related works is given and the used methods.

2.2 Pattern analysis and cartographic recognition of salient geographic objects

Interestingly significant works have been published globally aiming at semi-automatic and automatic information extraction from historic paper based maps. Some of these contributions have been noted here and are referred to owing to the fact that these contemporary works have also demonstrated automatic data capture and recognition methods on a variety of maps across the globe (e.g. Kerle & de Leeuw (2009) unlocked population signature from the Kenyan legacy paper map in African, Leyk et al. (2006) extracted forest cover from historical Swiss national topographic map series in Europe. Similarly, Chen, et al. (1999) extracted information from the Chinese land register maps, Deseilligny et al. (1998) recognized and extracted information from the French national topographic map and Luyang et al. (1999) used the United States Geological Surveys (USGS) map for his work and many others etc.) Another reason for referring to these contributions is that they are useful in building a logically sound conceptualization base for this research. For example Guerin & Deseilligny (1995) proposed automatic interpretation and digitization of scanned maps using spatial relationships to exhaustively extract an inventory of meaningful geographical objects to improve GIS analysis. Lichtner (1985) proposed pattern recognition by raster-vector-conversion where skeletonization of objects is realized using topologic method combined with the information from a distance matrix. Also Kasturi & Alemany (1988) developed an efficient algorithm to automatically detect, extract symbol information, identify and track various types of lines, follow closed contours, compute distances and find shortest paths, etc. from paper-based maps and answer queries related to spatial features and structure of geographic data. While Samet & Soffer (1998) demonstrated information extraction on paper maps using a system named MAGELLAN (denoting Map Acquisition of GEographic Labels by Legend ANalysis) the method utilizes the symbolic knowledge found in the legend of the map to drive geographic symbol (or label) recognition. They first scan the geographic symbol layer(s) of the map, locate the legend of the map and then segment the map portion. The geographic symbols (i.e., labels) are identified, and their semantic meaning is attached. An initial training set library was constructed based on available information and the training set library was subsequently used to classify geographic symbols on input maps using statistical pattern recognition. The following subsection presents detailed technical review of related works that fit into the aforementioned research categories and are evaluated for their applicability to the identified problems.

2.3 Overview of specific problems and related works

From the perspective of general pattern analysis and cartographic symbol recognition, one or more patterns are of interest in map object recognition (e.g. spatial extent mapping (area coverage), structural pattern recognition, and text/graphic extraction are frequent problems in information extraction techniques). On this note information extraction in this thesis adopts a stepwise recognition procedures similar to that implemented by Kerle & de Leeuw (2009), Leyk et al. (2006) and Khotanzad & Zink (2003) due to topographic map complexity. Stepwise here means identifying contents of individual map layer and essentially preparing the extracted layer into a new GIS usable layer for further analysis such as e-map proposed by (Dhar & Chanda 2006). Detailed description of the stepwise strategy is presented in the next chapter.

2.3.1 Area recognition and mapping

For most interpreters working on two-dimensional data such as paper map, extracting relevant information layers for further GIS analysis is very important. For example the boundaries of map objects are useful information for evaluation and monitoring of water resources, flood prediction, GIS database updating and water pollution detection. In this thesis the hydrographic layer is one of the objects that will be extracted (e.g. extent of rivers, lakes and creek). Over the years extracting such features is commonly done by manually digitizing or photogrammetric methods in creating polygons to determine the geometry, area or extent of a particular landscape feature. This procedure can be a time consuming and costly process (Dillabaugh et al. 2002). Other traditional approaches use the spectral signature to distinguish between water bodies and other objects through density slicing and the multispectral maximum-likelihood image classification. However these methods have their shortcomings since they depend on spectral distinguishing factor and can lead misclassification problems due to the complex interaction often represented in a single colour plane as well as touching and overlapping issues. Although Shane & Page (2000) proposed the rule-based classification of water in Landsat MSS image using the variance filter and the algorithm is used as a non-linear non-directional edge detector. The filter emphasizes sudden changes in image brightness without any directional bias and is successful at identifying shorelines. Similarly, Zhu & Joao (2003) developed lake extraction algorithm using attribute information. Firstly the method is based on a segmentation module in order to accommodate different types of images (e.g. gray scale and RGB composite images) such that images with digital values falls in a certain threshold values is classified as a water bodies. Secondly, they implemented a filter module in order to remove sharp spikes and small regions from the binary image. These methods are similar to the Definiens object extraction parameter which can be used in the case of paper maps as well to extract polygonal features.

2.3.2 Structural pattern recognition in maps

Structural pattern here is used as a compound term that describes a variety of complex object/features used to portray information about the location and extent of landscape features. For clarity structural pattern recognition can be considered as a symbol recognition application that can be used to extract targeted information. Assuming this analysis based on the description of spatial relations between graphical objects using "contextual" information such as distances and directions allows object recognition (Pavlidis 1977; Delalandre et al. 2004). Similarly Myers et al. (1996) and Llados et al. (2002) proposed set of rules to find symbols in the images. To efficiently extract structural pattern e.g. symbols on topographic maps an ideal method development requires topological relationship of spatially embedded object for extraction. Den Hartog et al. (1996) and Ogier et al. (2006) demonstrated the use of contextual reasoning to find certain types of objects in the vicinity of an already recognized objects. Considering that vegetation cover types depicted on the topographic map are represented by various complex patterns, structural analysis is an appropriate method to define relevant rules for identification of prototype symbols base on geometric and thematic attributes. Hence after successful recognition semantic processing can be applied to aggregate similar symbol for classification into appropriate class unit to determine the spatial extent of each vegetation cover type depicted on the map (Cordella & Vento 2001).

2.3.3 Text recognition on maps

Recurrent concerns in information extraction procedure applied to paper maps are character recognition, text/graphic separation and Roy et al. (2007) suggested a method to improve text/graphics separation suitable for maps. The approach works with colour separation using clustering method for the purpose of text/graphics separation. Similarly, Kasturi & Alemany (1988) developed an algorithm that mimics and perform the same logical operations humans do to extract information from maps. The method was demonstrated using the fundamental procedures human beings use to locate, extract spatial feature in maps which is a brilliant idea this research intends to work out in an OOA context. Cao & Tan (2002) proposed a separation of overlapping text and graphics method, the method works with the assumption that the constituent strokes of characters are usually short segments in comparison with those of graphics. It combines line continuation with the feature line width to decompose and reconstruct segments underlying the region of intersection. On maps text show salient shapes which distinct them from other features. Their typical characteristic aids extraction as Velázquez & Levachkine (2004) proposed a method to separate and recognize the touching/overlapping alphanumeric characters. While Zhong (2002) isolated other character-like features from lines and other graphic for complete text detection. To obtain multiple information and structured objects from digital topographic maps in raster format Frischknecht & Kanani (1998) used knowledge-based template matching approach which generated high recognition rates (about 95%). Additionally for complex conditions Casey & Lecolinet (1996) proposed template matching method, a well-known method for character recognition and can be applied where necessary for overlapping or un-isolated texts extraction.

2.3.4 Contour line extraction

Nowadays, many GIS applications require topographic information and have made line extraction a topical subject of interest in information extraction. Especially contour line extraction and recognition which is greatly significant in digital elevation modeling (DEM) data (Arrighi & Soilees 1999). Nonetheless contour line extraction and recognition remains the most difficult and time requiring process using manual extraction procedures. Research on automated extraction of contour lines has been demonstrated by some authors using different extraction and recognition steps. Some of the reviewed methods are presented as follows. Samet & Namazov (2008) demonstrated the use of fuzzy set for filtering contour lines in topographic map image. Firstly, the method broadly constitutes a filtering approach to distinguish between local variations due to noise and due to image structure based on Van de Ville et al. (2003) approach by looking for edges which helps in developing a robust estimate by applying fuzzy rule. Secondly a fuzzy smoothing method which filters the image to reduce noise component of pixels by means of correction of pixel values was applied. The method achieves good noise discrimination and also preserves and enhances the contour lines. Similarly Samet et al. (2010) proposed an automatic contour line recognition method. They firstly applied an initial colour image segmentation to recognize the features on digital topographic maps, followed by morphological and filtering operations to eliminate all unwanted information from digital topographic map except from contour lines. Then, resolved the terminal and crossed points and finally matched and reconnected broken contour lines to automatically extract contour lines form paper map. Furthermore Cao & Tan (2002) tested the concept of line continuation based on similar slopes, adjacency and size measures while Chen et al. (1999) applied graphical line tracing based on tests for slope equality and offsets. Since features on the topographic map lack discrete boundary thereby making the delineation of feature fuzzy, thus a reliable fuzzy set and fuzzy logic method will be appropriate in semantically defining boundaries of the geo-spatial objects due their semantic fuzziness. Also the use of line detection can be applied in order to preserve some feature structures using priori thinning (Nowak et al. 1992). Although degree of conjoining varies between maps sheets, objects, group of vegetation in the topographic maps such that features touch at some point and are vaguely delineated in other spaces. On the Nigeria topographic map it is obvious that line features are not uniformly represented especially where numeric characters appear on contour. Here the Hough transformations approach by Yamada (1993) is not applicable everywhere on the map except for unambiguous contour line on map. However, extraction of contour line here will begin with optimal segmentation and might require relying on well-defined line segments as proposed by Samet, Askerbeyli, & Varol (2010), Samet & Namazov (2008), Khotanzad & Zink (2003), Ghircoias & Brad (2010) and (San et al. 2004). Consequently, the segmentation quality can be managed using colour and the most common is the RGB colour format which is used in digital images. The primary reason for this is because it possesses compatibility with computer displays which aid unambiguity and it was used to advance line extraction. Also steps and significance of each step are described considering their circuital importance in contour line extraction.

2.4 OOA-based information extraction from paper maps

Recent advances in computer vision and machine intelligence have led to the development of new techniques, such as object-oriented analysis (OOA) for automatic information content extraction for both natural and man-made geospatial objects from remote sensing images(Akcay & Aksoy 2008; Blaschke 2010). The method is a concatenation of two broad iterative procedures, namely segmentation and classification. Segmentation is when image is explicitly broken down into meaningful features corresponding homogenous regions or segments with a process controlled by scale, colour, shape, smoothness and compactness criteria. Many segmentation techniques can be found in literature such as thresholding, edge based and region based segmentation (Miliaresis & Kokkas 2007; Rekik et al. 2007). Further discussion on segmentation is limited to segmentation methods in Definiens and will be used in this research. Classification is when the segments are in turn analyzed and labelled, using not only the multispectral or colour information but also attributes such as size, shape, or texture inherent to these segment (Akcay & Aksoy 2008; Kerle & de Leeuw 2009; Blaschke 2010). After converting analogue maps into digital images for viusalization, information extraction on topgraphic maps can also be achieved using semi-automatic or automatic algorithms in OOA context. In OOA expert knowledge is incoporated to develop rule sets based spectral, topological and morphologcal attributes of spatially embedded map objects and their false positives. This is also refered to as OOA classification. OOA uses salient objects on images with respect to their geometric, thematic and semantic attributes.
Research on information extraction from paper maps using the OOA approach only began to receive attention fairly in the recent past and is still in its developmental phase particularly for applications such as multi-temporal hazards and risk assessment. Remarkably, OOA based information extraction procedures from paper maps have been demonstrated in some application areas, such as forest cover extraction Leyk, et al. (2006) and population dynamics (Kerle & de Leeuw 2009). Primarily this research considers the above as a major inspiration for this research and the overarching extraction procedure hinges on the concepts saliency and semantic. Basically, saliency of such features would allow for explicit detection of the complete set of objects of the considered map category. While semantics aids in distinguishing map objects with identical color or shape characteristics using their location or proximity to other features. For example, this allows the distinction of mangrove prototype symbols from marsh, areas liable to flood or river layer as well as the symbols in green color representing six different classes. Thus using spatial and structural information can be complementary to pixel-based processing as established by Akcay & Aksoy (2008) on high-resolution images and further demonstrated by Kerle & de Leeuw (2009) on legacy paper maps. These concepts are used in this thesis for designing a systematic multi-step recognition framework for a robust and reliable information extraction procedure for paper maps using OOA approach in the Definiens software environment.

2.4.1 Segmentation and optimization procedures

Fundamentally, a requisite in OOA method was and is still built on image segmentation (Lu et al. 2011). This is the very first step required for information extraction. Here image is broken into objects or pixel clustersfor further analysis in OOA after which classification is done. Segmemnation of image into objects is formed based on certain image charateristics such as of heterogeniety or homgeniety. In Definiens software, several sgementation algorithms are provided like the multiresolution, quadree and chessboard

(Definiens 2010). Despite the variety of option for segementation algorithms in Definiens software, the choice of one most suitable algorithm for a good segmentation remains a challenge due to object variabilty in color, size and structure as well as the spatial distribition of map objects. Therefore, these segementation algorithms can be combined to achive complete and realistic object identtifcation with respect to spectral, shape, and hierarchical characteristics of image objects. Since the quality of segmentation phase determines the completeness of map object identification. Specifically, OOA image analysis is often associated with a number challenges since actual analysis relies on accurate image segmentation (e.g. *under or over estimation* of object boundaries). This has made segmentation in OOA to be thought of a shighly subjective making it a trial-and error approach previously. But recently OOA based information extraction method has been developed taking advantage of hierarchy of features, spatial dimensions at which objects (features), patterns, and processes can be observed and characterized (Marceau 1999). Thus making scale a tactical parameter for accurate object detection in OOA and requires optimization. Optimizating segmentation in OOA to achieve proper information extraction on paper maps can be carried out in Definiens using the vast number of parameters for each objects, such as mean layer, shape, texture and topological relationships based on the available data layer, to be used as class discriminators on our topographic maps in OOA. Recent advances in OOA researches have improved on the shortcoming of existing segmentation approaches and it is no more a trial and error approah as it use to be. Most recent of segment optimizartion work is that of Martha et al. (2011) they created a plateau objective function using Moran's I index and intrasegment variance that allowed an objective selection of the optimal scales required for identification of false positives. Estimation of dynamic parameters wad achieved using the K-means cluster analysis for several classification steps in OOA. Result of their work showed that multi-scale-based identification of false positives helped in achieving a higher overall recognition accuracy (76.9%) of targeted objects compared to single-scale identification and significantly reduced the error of commission which affected their earlier results. Similarly Dragut et al. (2010) developed a procedure called estimation of scale parameter (ESP) tool for the optmization of scale parameter and it works iteratively using the buttom top approach to generate image objects at multiscale levels as well as calculate the local variance for each scale. But, before now Esch et al. (2008) proposed an optimazation technique that iteratively combines an array of multiscale segmentation, feature based classification and classfication based object refinement by merging or cliping of segments. The technique was tested and experimental result suggeted the method to be an adaptive procedure that can facilitate more accurate and rubust image segmentation. Result showed improved segmentation process by 20 and 40 percent. Although, this method is said to increase processing time. Since each image analysis problem requires algorithms that deals with structures of a certain spatial scale, and the average image objects size must be freely adaptable to the scale of interest. This was achieved by a general segmentation algorithm based on homogeneity definitions in combination with local and global optimization techniques particularly asscale parameter can be used to control the average image object size. Different homogeneity criteria for image objects based on spectral and/or spatial information can be applied for paper map information extraction.Additionally, while doing this state-of-the-art review a plethora of image segmentation and clustering algorithms also exists in the literature such as the works of Bhattacharyya et al. (2011) who proposed a collection of adaptive thresholding mechanisms that incorporate the image context information in the thresholding process. Likewise, Kaftan et al. (2008) used mean shift algorithm which is based on an iterative scheme to detect modes in a probability density function. Furthermore, Hurtut et al. (2008) used spatial organization of colors between images through a global optimization procedure relying on the Earth Mover's Distance in image segmentation). Therefore, to achieve good OOA segmentation the Definiens scale parameter which is a value that determines the maximum allowed heterogeneity for the formation of image objects and will be tested to determine the optimal segmentation threshold most suitable for information extraction on our complex topgraphic map.

2.4.2 Classification

In OOA based information extraction, after segmentation and desired object primitives is identifed based on saliency of objects and semantic modeling, classification of aggregate feature layer into corresponding landcover class category is the proceeding step. However, during this stage false positives are identified and eliminated concurrently and true postives are accepted and classified accordingly. Various approaches are available to classify the features, such as the hard, soft and knowledge based approaches. The hard (crisp boundary) classification method label features based on some threshold values or classification algorithms while fuzzy classification uses the knowledge of fuzzy set to label and image using membership function. In Definiens, different classification algorithms are provided like the simple nearest neighbor classifier, membership classification method and the rule-based classification. Here we want to explore the rule based classification on complex topographic maps and at some point feature space optimation will be applied where necessary.

2.5 Accuracy assessment

Accuracy assessment (AA) is critical in digital image analysis. Therefore this thesis will demonstrate this critical attitude to validate the robustness and completeness of the OOA-based information extraction method developed for complex topographic map with a specified accuracy. To independently and comprehensively assess the overall performance of the entire information extraction procedure, a rule set transferability approach demonstrated by Tumuhairwe (2011) and comparison between the automatic and manual approach Lee et al. (2003) is appropriate and was established in this thesis. Cost and penalty of misclassification was considered and evaluated in developing the OOA-based information extraction method for topographic maps for better accuracy assessment. Cost here refers to classifier learning problems i.e. misclassification costs between targeted classes often present when classifying real world objects which exists on the topographic maps used in this thesis. In a map object/symbol/layer-classification analysis, for example classifying a valuable land cover class of mangrove vegetation into a marshland vegetation area (perhaps an area with least potential to the mapping interest-*non* protected vegetation) can be considered to cause higher costs than a misclassification of same land cover class into heavy forest area (perhaps a protected land cover class similar to mangrove). Additionally, misclassifying mangrove (implies not identifying a land cover class with high protection requirement) is in general more costly than misclassifying scrub vegetation. As classifier learning can hardly ever achieve a perfect classification, due to issues corresponding kinds of biases that machine classifier usually experience. Hence, quality control measures should be capable of considering a cost matrix in order to evaluate bias misclassifications in such a way, that costly misclassifications are managed. This method lessens costly misclassifications and therefore reduces error rates. Since search aims and classifiers in Definiens can be optimized using expressive powers that best guess the target function for learning and classification.

2.6 Multi-temporal hazards and risk assessment

Natural and anthropogenic environmental disasters have escalated in the past decades and have been identified and reported to pose serious threat to life and property all over the world which has prompted a reorientation of emergency management systems away from simple post event response (Cutter et al. 2000; Bouwer 2011). Though it is not possible to prevent natural disasters, but it is certainly possible to model their occurrence and reduce their impact by evolving appropriate preparedness plans and mitigation measures. Thus fairly recent guiding principle behind these mitigation strategies is the paradigm amongst disaster experts that disaster risk reduction (DRR) campaigns should be approached from a multi-temporal hazards perspective given that some hazards occur in chains and or sequential orders. Such a compound approach not only allows for proper segmentation and classification of hazard processes as well as specific hazard modeling but improves hazards understanding and risk management and heterogeneity in disasters. Since multi-temporal hazards and risks are dynamic processes and understanding these temporal changes can lead to improved risk and vulnerability management. Therefore, relevant multi-temporal GIS-based spatio-temporal hazards and risk modelling assessment suffering from lack of rapid information extraction techniques using available methods with errors and making input data unreliable especially from the aforementioned data sources in (section 1.1) can be improved. Hence OOA-based method is suited for rapid information extraction expertise needed within applied earth science especially disaster geoinformation management programs owing to the increasing demand for swift and reliable information needed for multi-temporal hazards and risk analysis.

2.7 Chapter summary

In this review in-depth theoretical appraisal of information extraction procedures, from traditional methods to current state-of-the art methods was done. Persistent challenges identified by previous researchers were highlighted in this work as recurring issues that might impede on achieving the objective of this research. Pattern analysis, cartographic recognition and its applicability to information extraction on paper maps were also reviewed tactically by discussing specific challenges and related work from the global perspectives. Fundamental requisites in OOA-based method segmentation and classification inclusive of optimization procedures were thoroughly reviewed to address information extraction on complex topographic maps. Accuracy assessment issues were also reviewed for quality assessment. In addition the relevance of developing this method and its suitability for applied earth science expert especially disaster geoinformation management experts was contextualized. In the next chapter, the study data, tools and method applied in this research are thoroughly described. Furthermore, implementation of the conceptualized information extraction method developed was put to test as proposed. This study aimed to use the concept of saliency and semantics to interpret and recognize patterns on paper maps, and consequently translate them into rule set which can be applied for feature or objects extraction in OOA context for suited applications e.g. disaster risk management programs.

CHAPTER 3

3 DATA AND METHOD IMPLEMENTATION

This chapter begins by describing the Nigerian topographic map of 1968 used for visual interpretation, manual feature extraction, development of OOA-based semi-automatic information extraction method and OOA rule set transferability analysis. Section 3.1 presents the data used, elucidating on the specific technical feature representations on the topographic map and a thorough description of salient objects with logical semantic meaning of object classes. Keeping the focus on method development, section 3.2 illustrate with good example of map sections and layer transparencies to write a clear and logical map treatment steps based on saliency and semantics present on the dataset to conceptualize the information extraction procedure developed for complex paper maps. In the next sub-section, an outline of the data pre-processing and OOA implementation steps applied to the map and derived information is provided. The segmentation and optimization techniques applied in Definiens 8 and subsequent analyses are also described briefly. Finally, a synopsis of the method developed to extract information on paper map was transferred to validate for robustness and reliability of the method.

3.1 Data

3.1.1 The Nigerian topographic map and graphical representation

The purpose of this research was to test the capability of OOA to extract information from two Nigerian topographic paper maps of (1968 and 1974) at 1:50,000 and 100,000 scales respectively (Figure 2 and Figure 3). The maps were compiled and drawn by the Federal Survey of Nigeria from a 1:50,000 graphic plots based on field survey carried out in 1968. The topographic map consists of approximately 1300 map sheets at 1:50,000 and 676 sheets at 1:100,000 published over 40 years. This motivated the efforts of developing an automated information extraction procedure for this historic dataset. Basically, the map in Figure 2 contains seven broad land cover layers or types resulting in a high density of information. Those are hydrographic, vegetation, other landscape features, built-up, utilities/public infrastructure, associated text/label and miscellaneous layer. The map background carries a whitish or beige like colour. While the hydrographic layer comprises of four water features types (Ocean, lakes, creeks and rivers) and are represented in blue. Nine vegetation types are signified differently with small graphical symbols unevenly and relatively spaced to each other using the plotter, yielding a tree like/lambda pattern rendition distributed all over the map. Some of these symbols are standalone objects (e.g. the mangrove and palm tree symbols) while others have separate parts and when combined forms a complete object (e.g. the heavy forest and marsh land symbols). Largely, these symbols were printed in blue and green colour (e.g. the mangrove vegetation in blue, heavy and light forest, savannah, park country, scrub, plantation and palm trees in green and the scattered plantation is in black with a zigzag like symbol etc). Three other landscape features closely related to water features printed in blue are the (i) areas likely to flood in bluish dash lines; (ii) the marshland symbols and; (iii) the sand or mud areas depicted in brown dots. Additionally, the map caries approximately six built-up layers comprising settlements layer, depicted in small black rectangular dots and associated texts in black describing the settlement. Present on the map are two distinct texts categories printed in blue and black colour and can semantically be associated to some of the generic layers on the map. For instance the blue texts can linked to the hydrographic layer, while the black text can be linked to settlement class and other non-hydrographic map features present on the map.

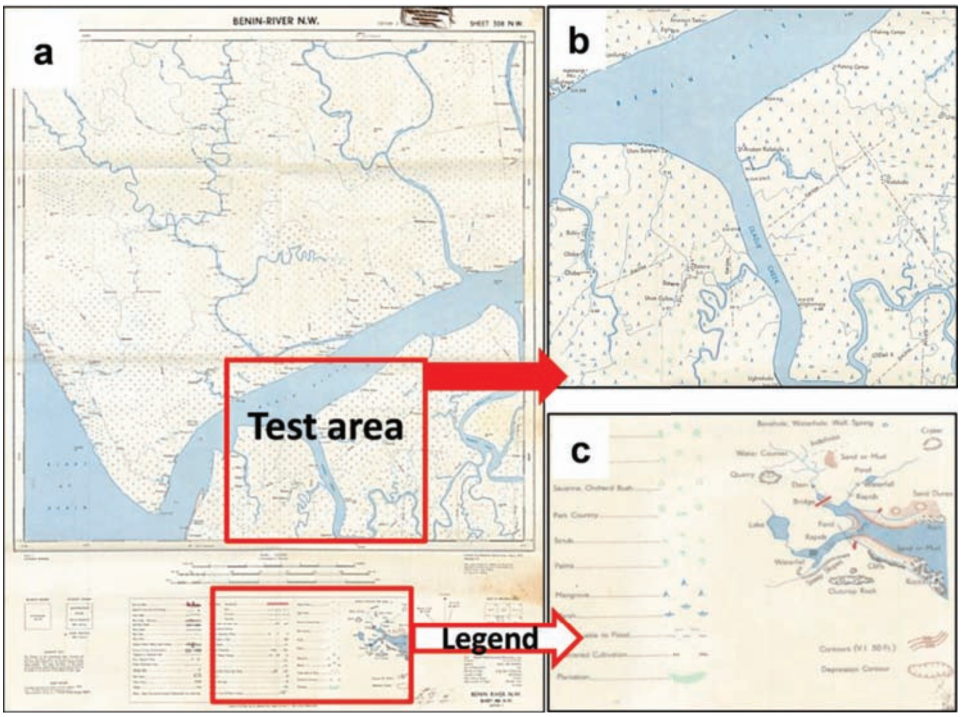


Figure 2: Data used in this thesis to extract hydrographic, symbol and text information.
(a) Shows 1968 topographic map of south western Nigeria with highlighted sections in red box. (b) Top right is the enlargement of highlighted section with red in (a) subsetted as test area, illustrating the relevant information layers considered in this thesis (e.g. hydrographic layer in blue, vegetation symbols in blue and green and, text in blue and black) and (c) Bottom right zoom of the map legend used to support visual interpretation of salient objects focused on in this research.

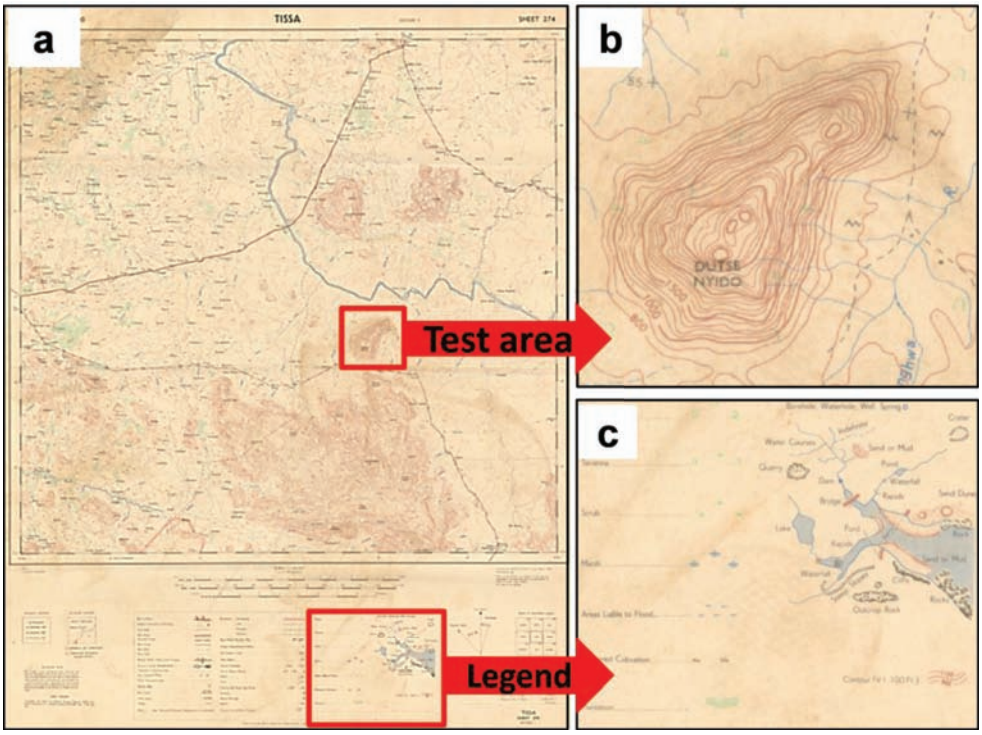


Figure 3: Data used for contour line extraction.
(a) 1974 topographic map of south eastern Nigeria with highlighted sections in red and yellow squares. (b) Zoom of highlighted section with red in (a) subsetted as test area showing contour lines represented in brown colour with varying length, shape, sizes and width and (c) Zoom of the map legend overview for interpretation.

Additionally, the map carries dotted lines with texts such as “cut line”, plus like symbol of coordinate line and other bleeding artefacts but will not be considered for extraction. Another layer is the utilities/public infrastructure layer (e.g. roads, telecommunication etc.) which was not considered for extraction. The second map Figure 3 contains contour line which is a typical example of line feature representation on topographic maps and is so far the most complex object considered for extraction in this research and requires the most complex extraction procedure. The map background is light brownish similar to the colour which the contour lines was printed. Contour lines are represented in dark brown with varying colour tone, length, shape, sizes and width. The lines are intercepted by height values in the same colour plane. Similarly the contours overlap, cross and touch other geographic features such as road, text vegetation symbols and hydrographic layer.

3.1.2 The conceptualized OOA information extraction order

Apparently, objects embedded on these maps can be characterized from simple to more complex map signatures. Similarly, these objects can be categorized into either of the standard object representation in GIS (e.g. polygon, point and lines). Thus extraction model conceptualized for objects extraction was a simple to complex procedures alongside object representation category they belong (Figure 4). For example the hydrographic layer was about the simplest extractable layer and reveals both polygon and line object class property due to its double and single object boundary characteristics (Figure 4). Standalone objects without broken parts and are unambiguous were categorized as moderate objects requiring moderate extraction procedures (Figure 4). Features overlapping, touching and have multi-parts (e.g. contour lines and ambiguous other objects) were categorized as complex representations and require complex object extraction procedures. Accordingly, depicted objects on the scanned map with no physical boundary but rather having ill-defined spaces representing object boundaries were labelled as zone of uncertainties.

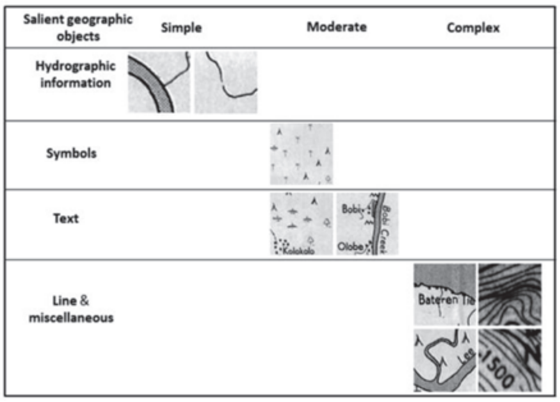


Figure 4: OOA-based information extraction order. This figure show the proceeding order developed for extracting targeted information from paper maps. The designed order logically facilitated the extraction of relevant information starting from most salient and unambiguous to moderate and complex features (e.g water, symbols, text and contour lines).

3.1.3 Software used

A number of software was utilized for this research. Table 1 presents the list and the purposes they were used for in this work.

Table 1: List of software used

Software	Purpose
Microsoft word	Used for report preparation and graphics representation
ArcGIS10	ArcGIS10 was used to georeferenced the entire input maps used in this study
eCognition	eCognition for all OOA information recognition and extraction tasks from paper
Photoshop	Photoshop Was used to select colour object in specific colour range to be extracted by OOA
GIMPs	GIMPs Was used to select colour object in specific colour range to be extracted by OOA
ERDAS	Erdas was mainly used for image to image georeferencing of the pre-treated maps.
Endnote	Endnote was used to prepare the list of references used in this study
Microsoft Visual	This was used for preparation of visuals

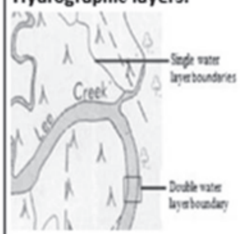
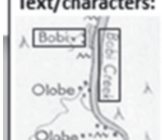
3.1.4 Homogeneity of saliency and semantic description of map layers

The homogeneity criterion was discussed here because it is a requisite property that supports the complete object creation process with other relevant properties. Its presence within object is determined by color and shape. Especially for region growing segmentation algorithms since they work with similarity of pixel to form objects. Critically speaking the topographic map used in this thesis contains both homogeneous and heterogeneous sections which in turn affects object creation in the segmentation process. This phenomenon can be managed through optimized segmentation algorithms. On the test map a substantial part of the hydrographic layer can be described as homogeneous, while hydrographic feature with single boundary mainly river channels possess some challenge due to their shape. The vegetation layer is made up of a variety of symbols; hence the layer is a combination of homogeneous and heterogeneous sections. Homogeneity here refers to section having similar symbols in a group while heterogeneity is the situation where neighbouring symbols differ from one another. For the texts layer, association can be done using semantics since texts belonging to the hydrographic layer are recognizable because of their colour, proximity and orientation while the black texts are settlement names. Also for line extraction contour line is the focus and has both homogeneous and heterogeneous sections. These aspects of the map are systematically tackled using the homogeneity parameter provided in Definiens software.

3.1.5 Visual interpretation of topographic map

Visual interpretation was essential for recognizing relevant information contained on paper maps and was logically done (Table 2). This was a fundamental step implemented for overall object identification using the concepts of saliency and semantics to develop knowledge base for automatic information extraction. The interpretation began from hydrographic features, followed by vegetation symbols and then associated texts belonging to various layers and lines (contour) and was described (Table 2).

Table 2: Object type and their logical visual interpretation

Object Class	Map feature interpretation
Hydrographic layers: 	Basically, the map legend was very useful for visual interpretation of salient features (e.g. water body). Spectrally, water is represented in blue with single and double boundaries (See left caption). Inland water consist of a major river, creeks, tributaries and channels while towards the coast is the open sea which is the international boundary of Nigeria and the Atlantic ocean. Obviously, on the map there are other features/objects with blue colour not belonging to the hydrographic layer (e.g. vegetation symbols and marshland). Thus they are regarded as false positive to the hydrographic layer. Also we have text printed in blue, semantically these texts can be associated to the water layer, and since they essentially carry names of the water feature they are close to, or printed on water feature. Hence, they can be extracted separately as text objects.
Vegetation symbol: 	Clearly, on the map numerous symbols (objects) are visible representing an array of vegetation types. Saliently, these objects are in green while some are in blue. Although some of these symbols are disjointed (made of two parts), while others overlap, touch and cross other objects and will pose extraction challenges. For discrimination and classification of these symbols the map legend was useful in interpreting the meaning of each symbol prototype. Example of symbol prototypes representing (mangrove, heavy forest, palm, light forest and marsh land) are identified on map section and zoomed out in red boxes beside the map.
Text/characters: 	Specifically, the map contains two main text types highlighted in red boxes belonging to two different layers. The blue text can be associated to the hydrographic layer with varying orientations. Some are inscribed around or within the hydrographic object. Semantically, these bleu texts describe the object e.g. names of river and creeks. While majority of the black text are name of town or settlement belongs to the built-up areas . Some other black text exist on the map but they will not be considered.
Line features (Contour): 	The map carries a series of linear objects on it representing roads, water contour lines etc. All of these linear features are saliently printed using different colours and this spectral characteristic was useful for identifying objects and logically assign semantic meaning to them by visual interpretation. Here, contour line is the focus and is depicted with brown colour. Apparently, some of these contour lines are enclosed while other are intermittently broken, crossed with numeric characters or overlapping other map objects as we have in the map.

3.2 Conceptualization of OOA method implementation

Conceptualizing a robust and reliable information extraction procedure for topographic paper maps containing layered information with high degree of complexity requires a map deconstruction approach (Figure 5). The deconstruction approach referred to here is operating a strategic reversal concept which is opposite of the map making process to retrieve targeted cartographic details (e.g. polygon, point, text and line features). This strategy was useful in the reinterpretation of relevant cartographic information on the topographic map into synthetic map layers. For instance the hydrographic layer was the most salient feature requiring the most trivial extraction procedure on the map and thus was initially extracted based on interpretation of salient features. To illustrate, this was done through the manual digitizing method using ArcGIS10 software to generate a synthetic map that illustrates the target layer transparencies. Same approach was used for the second relevant layer based on the visual interpretation (vegetation symbols).

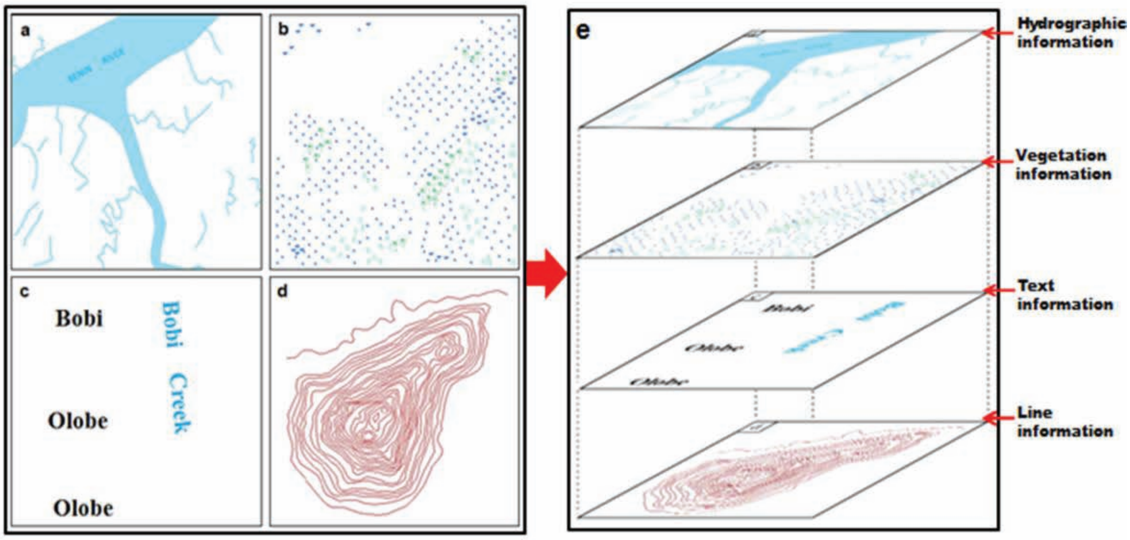


Figure 5: Synthetic map illustrating the transparency concept introduced in this thesis.
Maps (a-d) describe four deconstructed relevant information layer present on the topographic map. The deconstructed map layers are considered as transparencies (see e) and in turn are useful in the overall information extraction process. For instance transparency (a) blue map object/layer, it represents the hydrographic information i.e. water. Transparency (b), different green and blue symbol/object represents an array of vegetation types. Map (c) depicts texts belonging to different map layers and transparency (d) represents the line transparency in brown colour used for depicting contour on the map.

This same concept was applied throughout the map to extract relevant targeted information (text and line) and worked quite well in for developing a robust and reliable information extraction method for paper maps (Figure 5). Although, the procedure can be very complex if interactions between different objects have to be considered which is present on the test map and thus multi-step recognition procedure similar to that proposed by Dupont et al. (1998) and Leyk et al. (2006) was conceptualized for this research (see section 3.3 for details).

A Step-wise approach becomes inevitable if the object of interest cannot be explicitly detected by unambiguous characteristics directly. Cordella &Vento (2000), Llados et al. (2002) described how complicated it could be to extract information from maps compared to other documents. The difficulty is linked to the strong interaction of crowded spatially embedded thematic layers of same colours or shape depicting different objects in the landscape. Complexity is defined here based on the structural patterns of prototype symbology on the map, the correspondence of different map features having same colour, texts and linear features from different segments running into one another (e.g. text breaking continuous entity or a roadway/river layer crossing contours and the background effect) making them touch, cross, and overlap. All of these problems are present in the map used for this research. However, for segments of linear features (continuous, with no branches or crossings) automatic extraction can be easily achieved. Watanabe (2000) and Llados et al. (2002) further established that in maps, topographic maps are labelled the highest complex spatial information source compared cadastral or utility maps due to the high density of information contained in them, generalization i.e. displacing features and labels, smoothing features, and aggregating features based on scale and symbolization. Thus map information is highly dependent on the specific feature-to-feature relationships, scale, and use of the map. To overcome the complexity challenge associated to information extraction from paper maps, applying a map deconstruction procedure was found appealing to use in conceptualizing a systematic and multi-step information extraction. The procedure also supports stepwise targeted map content extraction from an OOA perspective which is glaring on the topographic map especially the density of information layer embedded on it. But this research considered the most relevant layers which in turn can be reconstructed into GIS ready map layers useful for diverse applications e.g. multi-temporal hazard and risk analysis (Figure 6).

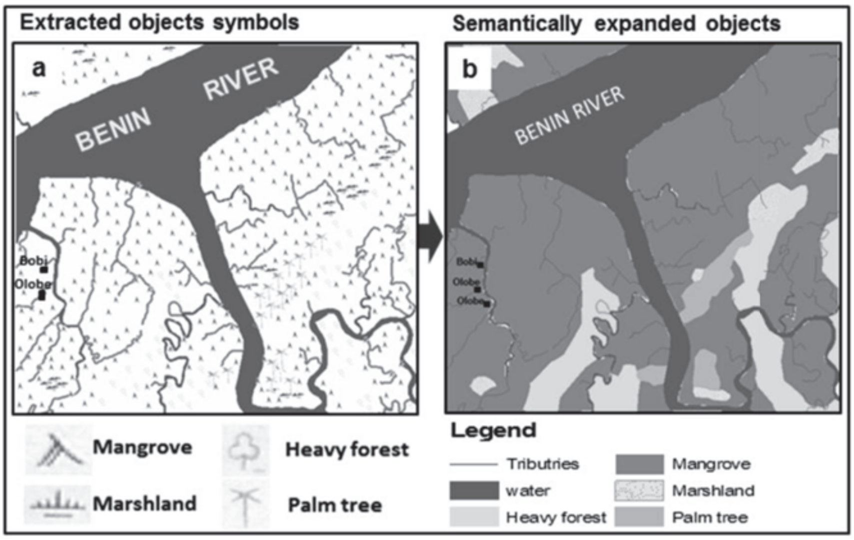
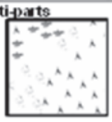
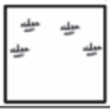
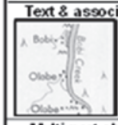
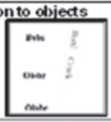


Figure 6: Exemplified overlay of the extracted GIS ready layers.
(a) Overlay of map transparencies (b) Strategic map deconstruction approach conceptualized into a new digital e-map (e.g. hydrographic layer in blue, vegetation layer in various green; mangrove, heavy forest, palm and other landscape features. Associated texts belonging to the water layer in blue and settlement in black.

3.3 Multi-step approach for OOA information extraction method

The "stepwise strategy" ensures that each step of human cognitive interpretation of a particular layer is developed into relevant OOA rule sets based on topological relationship of map objects for information extraction. The logical process begins with the "easier" layer; at the end of each step, the recognized objects are assigned to semantic class in the map, to facilitate the recognition of the remaining more difficult layers. To overcome the sequential aspect of this method, it was necessary to apply an iterative multi-step OOA based segmentation and classification with a strong presence of saliency and semantic concepts (Table 3). To position the “stepwise strategy" into the entire information extraction process a general overview was designed and is presented here. This simply is the pipeline of ordered list of activities leading to the information extraction processes on the map image (Figure 7). This must be empirically tested by observing the performance of the different sub-processes to arrive at a rule of thumb for subsequent application.

Table 3: Outline of multi-step recognition of salient object extraction methods developed

Salient object layers	Object transparencies	Derived Knowledge	Steps	Results
Water 		Large bluish object, homogeneous with single/double boundaries and has irregular shape, sizes & length signifying water on map.	1) Optimized multiresolution segmentation, 2) Region merging -blue segment 3) Rule based classification (1) 4) Region merging -back ground segments. 5) Rule based classification (2)	Recognized & extracted hydrographic layer
Vegetation & multi-parts 		Color & shape of vegetation symbols was helpful. Semantic meaning derived from legend along side associated map symbols.	1) Optimized multiresolution segmentation 2) Region merge-blue /green segment 3) Structural analysis & rule based classification /miscellaneous 5) Region merge background segment. 6) Rule based classification/ semantics expansion (convex hull)	Recognized & extracted vegetation types with different (symbols)
Non-vegetated symbols & multi-parts 		Color, semantic meaning derived from map legend and shape differentiating vegetation and non-vegetation symbols	1) Optimized multiresolution segmentation 2) Region merge (blue objects) 3) Structural analysis & rule based classification. 4) Region merge & rule based classification – (background segments) 5) Semantics expansion (convex hull)	Recognized & extracted non- vegetation information (symbols)
Text & association to objects 		Text colour, shape & orientation and regions. Association of semantic meaning derived from visual map interpretation	1) Optimized multiresolution segmentation 2) Region merge (black text) 3) Region merge (blue text) 4) Rule based classification 5) Template matching (optional)	Recognized & extracted texts belonging to water and place names
Multi-part objects (miscellaneous) 		Bluish and greenish line part of marshland & heavy forest class	1) Optimized multiresolution segmentation 2) Region merge 3) Structural analysis 4) Rule based classification	Recognized & extracted symbol multi-part belonging to vegetation and non-vegetation objects.
Line extraction and numeric characters 		Contour line colour & spatial description optimal gaps between objects	1) Optimized multi-resolution segmentation 2) Region merge 3) Complete line detection using geometric properties (line connection) 4) Character detection 5) Rule based classification	Recognized and extracted contour lines

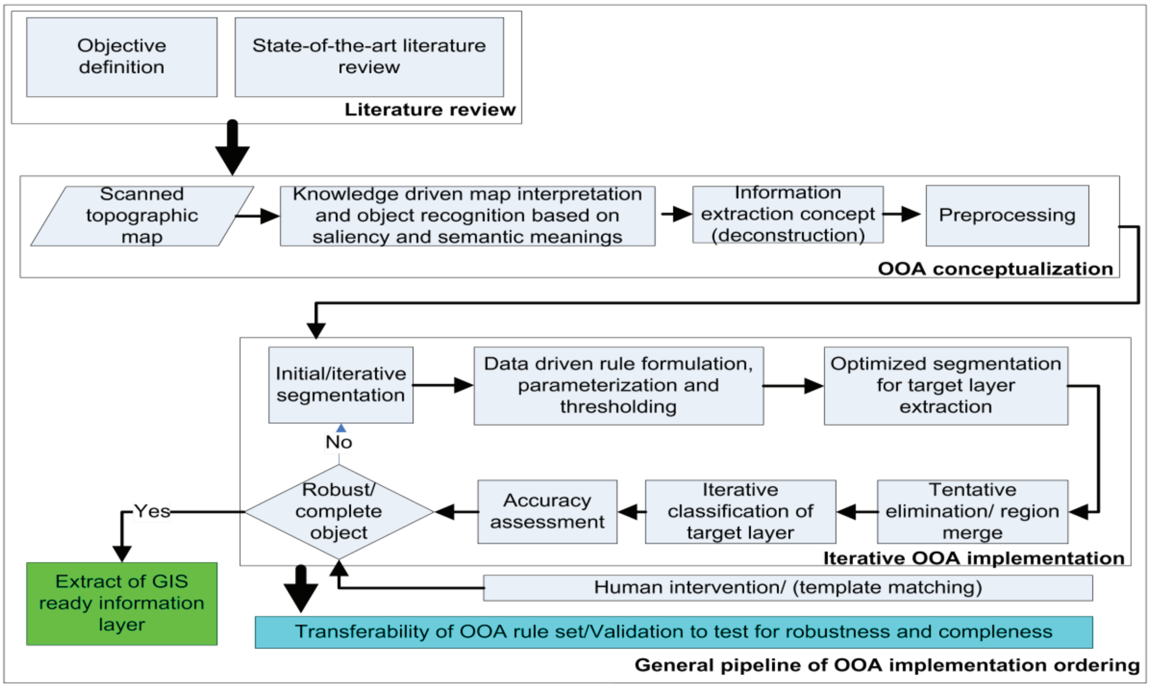


Figure 7: General methodology followed in this thesis.

3.4 Map pre-processing

Figure 8: Overall map section.
Scanned map separated from map legend, border and other unwanted section using the subset tool in Definiens, inset of test area highlighted in red square used for the method development.

The maps were scanned at 200dpi resolution and the choice of scanning was motivated by using a modest resolution that preserves accuracy of the original map, but certainly do not improve it and keeping data size within a manageable for fast computing hours. Also scanning was done with the TIFF file format using a color bit-depth of 8 bits per Red Green Blue (RGB) color channel being a suitable visualization spectrum for humans in the computer screen. The original maps were georeferenced in ArcGIS10 using the coordinate grid of the map and the pre-treated maps were subsequently georeferenced in ERDAS by image to image procedure. To avoid unnecessary processing time it was essential to fulfil certain logical preconditions such as the (removal of unnecessary parts as well as map treatment) before commencing to implement the requisite OOA based information extraction procedures. For replicability important OOA implementation steps are successively described below:

Step 1- Removal of unwanted areas: The large legend area in the bottom of the topographic map inclusive of scale bars, the map boarder housing coordinate of the map, as well as the outer space after the boarders were eliminated as absolute exclusion areas. This was done via subset tool provided in Definiens while loading the map to block out unwanted areas (Figure 8). This same step can be achieved using the coordinates of a georeferenced image as well as specifying rules the minimum/maximum *x* and *y* values.

Step 2- Subset of test area: To create a representative (test area) or subset sufficient to implement an OOA based procedure to extract targeted information layers, rule specifying the minimum/maximum (*x* and *y*) values was demonstrated to temporarily block out relevant map section used for the research (Table 4). Figure 8 illustrates the OOA rule based test area creation process by blocking out map section for subsequent analysis.

Table 4: Minimum/maximum x and y rules for sub setting the test area

S/No	Logical steps	Parameter	Remarks
1.	Chessboard segmentation	Scale factor: 250	Initial uniform fast segment
2.	Update region by specifying minimum/maximum (x and y) values	Origin (min) x: 1750 Origin (min) y: 0 Extent (max) x: 1800 Extent (max) y: 1550	} min/max (x and y) values of test area
3.	Region merge	Test area (defined)	Image object merging

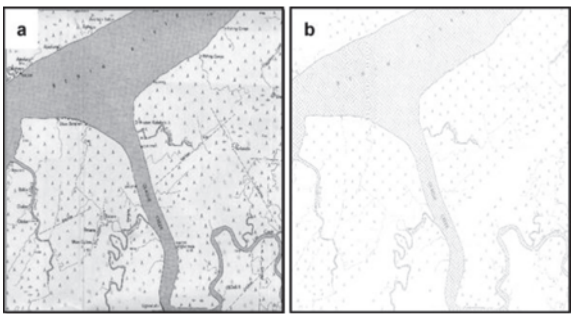


Figure 9: Subset test area.
(a) Original map showing the test area. (b) Pre-treated map of test area.

To overcome lossless compression effects emanating from the map dirtiness and scanning artefact such as aliasing and bleeding on the topographic map Figure 9(a-b), a map pre-treatment approach was considered using the colour range tool in Photoshop graphic software (Figure 9b). This was to allow experimentation of information extractability between the original map and the proposed map pre-treatment approach for comparison. Interestingly, this experimentation facilitated the removal/reduction of background artefacts emanating from map conditions (e.g. bleeding, map dirtiness, crossing, touching and overlap issues present on the map). Thus the map pre-treatment approach helped to establish the need for paper map pre-processing step before implementing an OOA based information extraction of targeted layers on scanned complex paper map as was demonstrated on the blue information layer using the colour range selection tool of Photoshop graphics software (Figure 9).

3.5 Map processing

Step 1- Segmentation and optimization:

A proceeding step was to further segment the subsetted test area using initial chessboard segmentation at a scale factor of 1 so as to get back to the pixel level from earlier object level on the image in question. Next was to estimate the optimal scale parameters suitable to further segment the test area image using the estimation of scale parameter (ESP) tool developed by Dragut et al. (2010) instead of the trial and error method. The tool produced an array of possible optimal scales suitable (Figure 10) and was used to implement the multiresolution segmentation algorithm that facilitated the sequential extraction of targeted object of interest based on saliency and sematic as proposed in the multi-step recognition outline (Table 3 above).

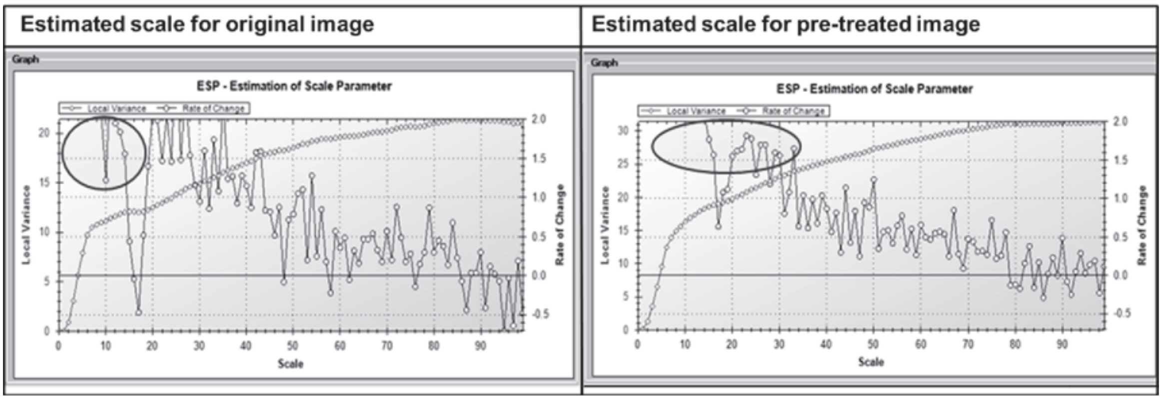


Figure 10: Estimation of scale Parameters.
(Left) Possible scales suitable for segmenting the original image circled in red. (Right) Possible scales suitable for segmenting the pre-treated map circled in red.

Recall from (chapter 3.3) that a "stepwise strategy" was prescribed which is simply a pipeline of ordered sequence of activities starting from (trivial to complex) aspects leading to the overall information extraction processes on paper map. Therefore, information extraction began with the "easiest" being the hydrographic layer and was used to illustrate the segmentation optimization on both the original test area and the pre-treated maps. This was demonstrated considering the possible effects of map complexity and conditions on the intended information extraction development procedure. To robustly and reliably extract the targeted information layers due to the complexity of cartographic features embedded on the topographic map the choice of a multiresolution segmentation algorithm was most preferred bearing in mind the dynamic computational capability of the algorithm in efficiently recognizing parallel and distributed objects. The multiresolution algorithm segmentation was demonstrated with the lowest ESP tool derived scale factor and other necessary parameters (Table 5).

Table 5: Optimized segmentation parameter

Image types	Scale	Shape	Compactness
Original map	07	0.1	0.1
Pre-treated maps	07	0.1	0.1

3.6 Logical treatment of topographic map for information extraction

3.6.1 False positive treatment in information extraction procedure

Generally because targeted object layers belonging to different object classes on the test map carry similar colours, this increased the challenges posed in the information extraction process. A fundamental challenge was the manifestation of false positives while extracting specific information layers on topographic map due to aliasing bleeding and convolution etc. False positive is defined as unwanted objects recognized along with the targeted information layer which poses object extraction challenges. Hence, such false positives needed to be separated from the true positives (i.e. the target layer). Therefore, they were eliminated where possible or classified to a separate class (e.g. temporary class) using relevant spatial or contextual rules to enable specific object extraction. For instance due to the appearance of objects on the same colour plane like the hydrographic layer and the vegetation (mangrove), blue texts and landscape feature (marshland) on the map to extract the hydrographic layer either of these object might appear as false positive and had to be treated accordingly.

A. Hydrographic candidate (tranparency-1):

The hydrographic layer was the most trivial object to be considered for OOA information extraction on the paper map with respect to the proposed map deconstruction approach described in (section 3.2) in form of transparencies. Saliently, the hydrographic layer is represented with blue colour on the map and was made up of a broadly homogeneous main river and small meandering tributaries etc. Additionally, the hydrographic layer had two distinct boundary types (single boundary linear features and double boundary dynamic features (Figure 11).

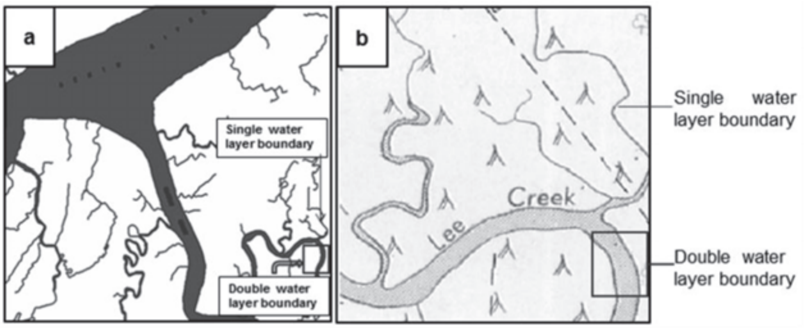


Figure 11: Steps of OOA for hydrographic layer extraction.
(a) Conceptualized transparency of the hydrographic layer with a section highlighted in red box. (b) Zoom of representative section highlighted with red box and line arrows of the test area transparency in (a) from the scanned map carrying hydrographic layer as a salient feature with single and double boundaries.

Step 1: Recognition of hydrographic information began with optimized multi-resolution segmentation algorithm which was implemented on the original topographic map and pre-treated map (Figure 12). Here the optimized scale parameter derived using the (ESP) tool with a rate of change (ROC) above the local change peaks was used. Also shape and compactness parameters were kept at minimum considering smaller object of interest (see Table 5). Clearly, Figure 12 shows how homogeneous map sections form segments that correspond to the targeted information on the map (e.g. hydrographic layer in blue with larger segment having double boundaries as well as meandering river channels having single boundary).

Step 2: Ideally, when segments are created as in Figure 12 inspection of the segment attributes is next line of action to facilitate object extraction. The mean layer values of image object information was used.

Here the *Blue Mean layer* clearly highlighted the corresponding hydrographic objects robustly and was used for object recognition. The Blue layer information is a mean layer value which reveals an array of image object information threshold across image segments. For complete object identification the Blue layer value derived was used as the threshold value to perform the merge operation. *Blue layer threshold of <= 227 and <= 254* respectively was used to implement the merge rule for all blue objects corresponding to hydrographic primitives, vegetation symbols and blue text thereby making all related segments to agree as single objects ready for further analysis (e.g. classification).

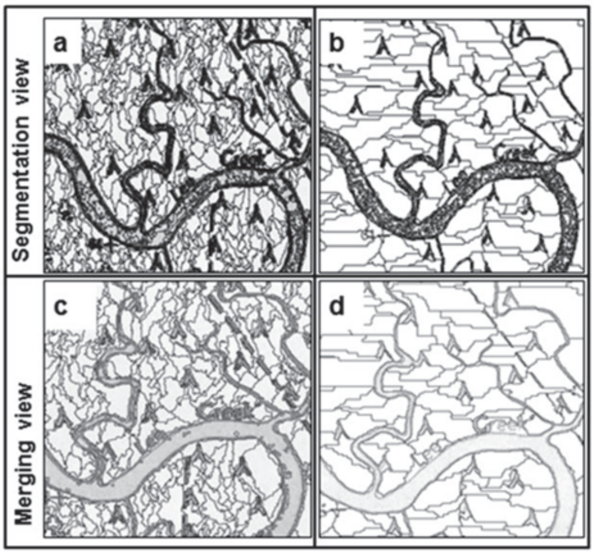


Figure 12 (a-d): OOA multiresolution segmentation and merging views.
(a-b) Shows segmented image section of corresponding hydrographic segment on original and pre-treated maps. (c-d) Shows the merge operation performed on the original and pre-treated image to recognize hydrographic objects.

Step 3: Next was to create a hydrographic layer class and assigned the merged object to the created class using size criterion of the object geometry (Table 6). Here the area of the merged hydrographic layer derived from geometry/extent tool of the image object information was used to assign the hydrographic layer to the created class. Accordingly, all other segments such as the map background segments as well as the vegetation and texts were logically assigned to their respective classes.

For instance a background class was created in the class hierarchy; a region merge rule was implemented on the background segments using *Brightness parameter of >= 227* and was logically classified with the same parameter on the original map. Similarly, the *Red layer mean parameter of >= 254* was used to merge background objects on the pre-treated image and the same layer threshold applied for classification. Same procedure was also applied to create and assign other objects to their logical classes (e.g. segments to be further analyzed such as vegetation objects from the unclassified layer).

Table 6: Parameters used for merging and classifying the hydrographic segments

Image types	Marge parameter	Classification parameter
Original map	Blue <=225	Size (area = 640133)
Pre-treated map	Blue <=254	Size (area = 642076)

Although, hydrographic object recognition was achieved, but not entirely recognized and this challenge was addressed by further inspecting the unclassified class to identify linear objects corresponding to the hydrographic layer. Here object length parameter was calculated which proved useful. Linear objects with *Length threshold >= 35 within the Blue layer mean of <=227* were further assigned to the water layer class.

Critically, the extracted hydrographic object/layer was accompanied with false positives due to spectral resemblace of some objects (e.g. mangrove, blue texts and marshland symbols carrying same colour with the hydrographic layer). These false positive were tactically evaluated based on cost and penalty of misclassification. Considering the degree of object proximity to the hydrographic feature, fuzzy membership can be introduced in assigning the cost of miclassified objects. For instance in the case of riverine settlement and vegetation misclassfied to the river boundary which attarcts a moderate penality due to its proximity. While blue text was assigned a low penalty since it is a discriptor to the hydrographic layer.

B. Symbol recognition, extraction and aggregation (transparency-2):

The vegetation information behind these symbols remains the sole motivation for extracting the second transparency. On the topographic map vegetation was represented with different symbols in blue and green colours dipicting mangrove, palm tree, heavy forest vegetation types and lanscape features (Figure 13).

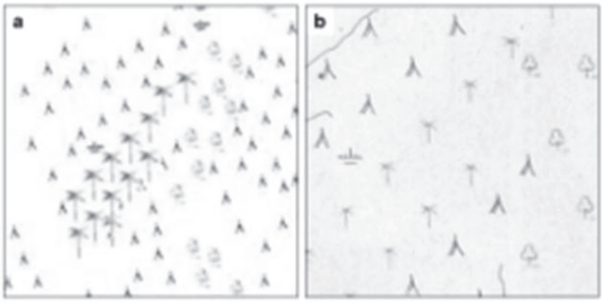


Figure 13: Image showing vegetation symbols.
(a) Vegetation layer transparency. (b) Topographic map showing various vegetation types to be extracted using OOA method.

Step 1: To begin vegetation symbol extraction the multiresolution segmentation alogithm was applied on the subset of the original and pre-treated images using scale parameter of *(07),* shape and compactness parameter of *(0.1)* respectively (as in Table 5 above). Here to sucessfully extract vegetation objects two pretreated images were considered for extracting the vegetation types based on colour variation. The initial pretrated map contains all blue symbol capturing mangrove vegetation and landscape features. The later pretreated map carries all green symbol capturing palm tree and heavy forest vegetation. The idea was to break the images into meaningful segments corresponding to vegetation and non vegetation symbols on the map irrespective of the colour and shape variation.

Step 2: To facilitate complete vegetation layer recognition the image object information was applied to inspect the most useful object property to efficiently recognize symbols. Here the mean layer attributes proved useful in vegetation symbol recognition. The preceeding step was then to implement the region merge rule on the images to merge the targeted segments such that meaningful objects are formed for extraction (e.g. mangrove, heavy forest, palm tree vegetation types and marshland). This was achieved using the *Mean Blue threshold of <=227* for the original topographic image. While for the the pretreated images with blue symbols at *Blue Mean threshold of <=254* and for the pretreated image carrying green symbols at *Mean Brightness threshold of <=250* (Table 7).

Table 7: Parameters used to merge and classify the recognized vegetation objects

Image types	Marge parameter	Classification parameter
Original map	Blue layer mean <= 227	Feature space optimization Nearest Neighborhood classification
Pre-treated map (Blue symbols)	Blue layer mean <= 254	Width <= 13.5 for marshland Width >= 13.6 for mangrove
Pre-treated map (Green symbols)	Brightness layer mean <= 250	Green layer mean <= 246 *(Heavy Forest with Area > 100 & Width >=1 pixels for the multi part).* *(Palm Tree with Width >=9 pixels).*

Step 3: Consequently, logical classes were created for the target information layers considered for extraction such as mangrove, heavy forest, palm tree vegetation and marshland. After these classes were created the recognized objects were then assigned to their respective classes using relevant rules (Table 7).

On the original topographic map extraction of vegetation and non-vegetation symbols was attempted with the Feature Space Optimization (FSO) function considering challenges faced using rule-based approach and classification was done using nearest neighbourhood classification (NN) method. The FSO function is a technique that offers a method to mathematically calculate the best combination of features in the feature space dimentions. This was done by selecting representative samples of the target symbols using the sample selection tool with brush (e.g. marshland and mangrove symbol). Then NN was implemented using the FSO tool by selecting the sampled classes, features used and subsequently calculated. While on the pretreated images the rule based classification approach was used. For the mangrove vegetation symbol and marshland the pre-treated image carrying blue objects was used. Here *Width <= 13.5* was specified to assign the marshland object to the initially created class and this was robust to capture the multi-part of the marshland objects while *Width >= 13.6* was specified to assign the mangrove objects accordingly (Table 7). Similarly, heavy forest and palm tree vegetation types were tactically extracted and to assigned to their appropriate classes from the pretreated image carrying green coloured vegetation types. Extraction of the heavy forest was achieved with *Mean Green threshold of <=246 and Area >= 100 pixels* and the multipart belonging to the heavy forest was extracted using *Width parameter of >= 1* from the unclassified segments since other feature have already been extracted. While the extraction and classification the palm tree vegetation type was achieved by specifying the rule *Green layer mean of <=246 and Width >= 9 pixels* (Table 7).

Table 8: Parameters used to merge/ classify background segments of vegetation extraction images

Image types	Marge parameter	Classification parameter
Original map	Brightness layer mean *of >= 227*	Brightness layer mean *of <= 227*
Pre-treated map *(Blue symbols)*	Red layer mean *of <= 246*	Red layer mean *of <= 246*
Pre-treated map *(Green symbols)*	Red layer mean *of <= 239*	Red layer mean *of <= 239*

Step 4: To completely recognize targeted objects the image segments corresponding to the background class were merged by specifying the region merge and classification rules. Here *Mean Brightness parameter >=246* was used to merge the background segments of the original image. The threshold was subsequently used to assign the merged background segments to the created background class (Table 8). Accordingly, the background segments of the pretreated images for both the blue and green objects were merged and assigned to their logically created classes. This time the background segments of the pretreated image carrying blue objects was merged with *Mean Red parameter >=246* and was assigned to the created class using same *Mean Red parameter of >=246* threshold. Furthermore, background segments of the pretreated image carrying green objects was merged and assigned to the apprpriate class with *Mean Red parameter of >=239* threshold. Although, vegetation and non-vegetation object recognition was achieved, but objects were accompanied with false positives. For example the misclassification of vegetation as water is costly misclassification and thus attracted a high penalty. Similarly, the misclassification of mangrove as marshland attracted high penalty and was assigned a high misclassification cost. But in the case of misclassified vegetation type of palm tree to the heavy forest here the misclassification cost is low since misclassified objects belong to the same category. Hence penalty of misclassifying objects from one category to another e.g. vegetation into water or vegetation into landscape feature was evaluated with severe costs and high penalties. While the misclassification of object within same object category attracted low to moderate penalty. Misclassification matrix is presented in the result section.

Step 5: To complete extraction of symbols it was necessary to aggregate the various symbols cartgories to overcome the concavity of symbol polygon in order to to determine the entire boundary of each symbol types. Here the extracted layers were exported as shape files using the Definiens export routine. Afterwards, aggregation of the individual symbols representing specific thematic information was logically done according to the cartegory such that missaggregation is avoided. This was aciheved using the ArcGIS geoprocessing tool to create a convex hull generalization sequentially (Figure 14). The convex hull is a computational approach of creating a feature class containing polygons which represent a specified

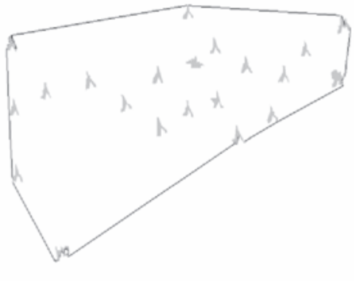


Figure 14: Convex hull of extracted objects

minimum bounding geometry enclosing each input feature or group of features. The feature outline masking operation of the cartography tool was used for feature aggreation. Then the output was dissolved. Merging was considered where necessary. Here two measurement thresholds were investigated for extracted objects. The first measurement was base on the object angular unit representation on the map at 0.0174meters and the second was base on object extent which was determined by the software measurement tool at 0.5meters. This was done with the assumption that each symbol represents certain amount of landscape feature within the spatial context of its location. It is important to note that defining decrete boundaries for two distinct objectes was concepualized as fuzzy boundaries and thus areas with these semantic and imprecision issues were identified as zone of uncertainties which exist through out the vegetation and non-vegetation symbols. Hence, expert opinion for boundary modelling was explored by giving about 10 experts the test area image to manually digitize and delineate boundaries of the symbol units which was compared with convex hull approach. The image was accompanied with an ArcGIS project shapefile produced by the author for consistency and quality control. Overlay of the results from responses gotten are presented in result section.

C. Text recognition (transparency-3): Recall from (Chapter 3.3) that embedded texts on the Nigerian topographic map was interpreted and associated to different map classes. Saliently, texts are in blue and black colours and were semantically interpreted base on their colour association, adjacency and orientation. Although there are a variety of black text but in this thesis focus will be on the black text describing only geographic features (e.g. settlement/town). Figure 15 captures the map section and the corresponding transparency prescribed while conceptualizing the OOA method development.

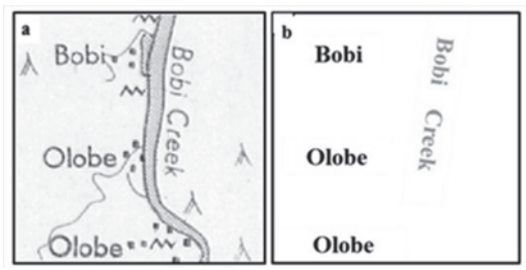


Figure 15: Map (a) shows texts in two different colours belonging different layers.
Map (b) is the deconstructed transparency of the text layer.

Step 1: To efficiently extract text information belonging to various class layers on the images, the first step here was to implement the optimized multiresolution segmentation algorithm at a scale of 7 keeping shape and compactness parameters at 0.1 respectively (same as Table 5). This is because texts/characters are similar to symbol and this algorithm works well in recognizing object morphology satisfactorily. Here colour, adjacency and orientation was used as the first basis to identify, discriminate and associate text features to specific semantic meanings. However, the aim of this investigation was to explore how far OOA method can be used for text recognition and extraction rather than text understanding.

Step 2: After segmentation the variety of Definiens object/segment feature properties such as the mean layer values were used to inspect the most useful object property that best recognize targeted text layers. Here the *Blue layer Mean* was useful in identifying blue associated to the hydrographic layer and black texts linked to the settlement layer on the original map image. To overcome extraction challenges the pre-treatment approach was applied here by using colour range tool to separate the blue and black texts associated to the two distinct layers to aid extraction. Figure 16 illustrates

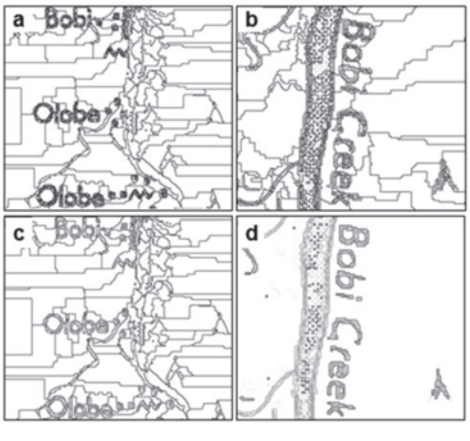


Figure 16: OOA region merge on the segmented images highlighted in red.
(a) Merged text objects on the pre-treated topographic image. (b) Merged black text on the pre-treated image describing names of settlement, schools and other geographic features (c) Shows merged background and blue text segments demonstrated on the pre-treated image of blue texts belonging to the hydrographic layer.

Table 9: Parameters used for merging text segments and blue objects pre-treated image

Image types	Marge parameter
Original map	Blue layer mean <= *227*
Pre-treated map (Black texts)	Brightness layer mean <= *250*
Pre-treated map (Blue texts)	Red layer mean >= *246* (Background segments) Blue layer mean <= *254*

the segmented views of two the pre-treated images the initial capturing black texts and the later showing blue text black texts. For the black text *Brightness layer Mean* was found useful in identifying black texts. While *Blue layer Mean* was useful in identifying blue texts (Figure 16). Hence, the next activity was to implement the region merge rule for complete text recognition. To advance text extraction the region merge rule was implemented using relevant *Mean layer thresholds* for grouping target objects (Table 9). The region merge rule was implemented using relevant parameters to facilitate strategic text extraction belonging to specific geographic objects present on the images.

Step 3: To efficiently proceed with text extraction it was necessary to merge all unwanted segments such that they do not impede on text information extraction (e.g. background segments). As this was the case while implementing the initial region merge rule for targeted texts. Therefore, to overcome text extraction impediment, background segment was firstly merged for the pre-treated image carrying blue text before merging the text objects. But for other images the background and other unwanted segments, the region merge and classification was done after initially merging the text objects. Then the background was subsequently merged and classified using relevant rules (Table 10). Next was to create model classes (e.g. hydrographic texts and settlement texts). This was done in the class hierarchy section of Definiens such that each of the extracted text layers were assigned to the logically created class since texts saliently carry different colours and semantically describes different geographic features on the topographic map.

Table 10: Parameters used to merge/classify segmented image background

Image types	Marge parameter	Classification parameter
Original map	Red layer mean *of >= 219*	Red layer mean *of >= 219*
Pre-treated map *(Black texts)*	Red layer mean *of >= 225*	Red layer mean *of >= 225*
Pre-treated map *(Blue texts)*	Red layer mean *of >= 254*	Red layer mean *of >= 254*

Step 4: The OOA based recognized texts were systematically extracted by assigning them to the logically created class types in the preceding step. On the original topographic image both blue and black texts were extracted using colour information.

Thus Max. Diff. which is calculated by subtracting the minimum mean value of an object in the spectral bands (here RGB) form maximum value, divided by brightness of the segments. *Max. Diff. threshold of >=3.0* was used to extract the blue text belonging to the hydrographic layer. While *Brightness threshold of >=134* was used to extract the black text belonging to the settlement layer (Figure 17a-d). Black text extraction was also experimented on the pre-treated image. But black rectangle signifying settlements such as Bobi and Olobe and wavy lines signifying scattered cultivation were initially extracted although were not in the initial target objects to be extracted (Figure 17b).

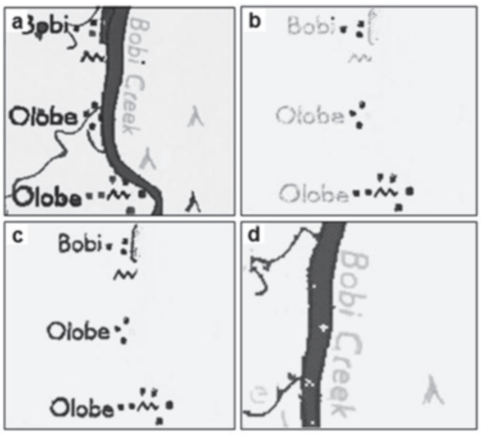


Figure 17: OOA based text extraction.
(a) Extracted texts from original image in brown and blue. (b) Extracted settlement dots in red (c) Extracted Black texts belonging to settlement class in brown colour and (d) Extracted blue text in brown and green. All background is in yellow.

This was achieved using the *Length threshold of* <=10 to extract the black dots depicting settlement and length/width = 2.100 for the black wavy line. Then black texts belonging to the settlement was finally extracted using the unclassified segments and was further assigned to the created black text class (Figure 17c). Additionally, blue text extraction was also demonstrated on a separate pre-treated image. This was achieved by initially classifying other object after which the unclassified segment was used assigned to the blue text class created (Figure 17d). Generally, the extracted texts were accompanied by false positives as illustrated in and were evaluated using the misclassification cost/penalty matrix table in the result chapter. The cost/penalty of misclassification was used to appraise the false positives generated from the OOA based text extraction in order to determine the severity of object misclassifications. On the original image in (Figure 17a-d) misclassification was as a result of bleeding between black text and hydrographic layer as well as black dot/wavy line with the hydrographic layer. This attracts high penalty/cost since they are from the same category. Similarly Figure 17 illustrates the misclassification of wavy line to the settlement dot this also attracts high penalty. Clearly the misclassification of dot above the letter (i) of Bobi describing the name of a settlement was evaluated as moderate considering its association to the text describing the extracted object. The extraction of text on the entire test area images was not robust due to inconsistency in text object characteristics and this necessitated a theoretical investigation.

Experimental text extraction by OOA: This theoretical experiment was done using synthetic image containing series text and numbers on a full document image with different font sizes and the same color and orientation. This experimentation aimed at investigating and identifying most useful text object properties such that they could be used as dictionary to robustly guide successful text extraction on maps. Here the image was subjected to multiresolution segmentation at the scale of 7, shape and compactness of 1. The text segments were analyzed and *Max.diff of >=3.9* was useful for merging corresponding text segments. Similarly, *Max diff. threshold of 0* facilitated the merging of the background segments and *Max. Diff. of <=0.003* aided classification accordingly. Finally, *Asymmetry threshold of <=0.9* robustly extracted text on the synthetic image used as input data.

D. Line recognition (transparency-4): Generally contour lines are represented as brown linear features on the topographic maps same as the one used in this research (Figure 18a-b). Line extraction was the fourth layer transparency proposed for OOA extraction (Figure 18). Basically, contour line extraction can be manageable if they are presented in uniformly controlled colour. But can become complex due to change in colour, more or less in contrast to background due to aliasing, bleeding and colour falsity. On the input map used for this research all these challenges were present alongside other challenges such as crossing of multiple features belonging to different layers (e.g. road, texts, vegetation symbols and rivers). Also contours lines had similar colour with image background as well as the evident map conditions. Altogether these factors tend to impede on information extraction processes. Hence this thesis demonstrated two OOA techniques in a multi-step fashion to recognize and extract contour lines. The first technique entails exploring the simple segmentation, region merge and classification in a multistep approach. While the second technique investigates the outputs of applying more advance algorithms on the recognized contour lines. Here morphological operation, convolution filtering and line extraction algorithms of Definiens were used to experiment the possibilities of overcoming these challenges highlighted above on both images. The approach is similar to the method proposed by Samet et al. (2010) where they initially segmented the input data to recognize the topographic maps

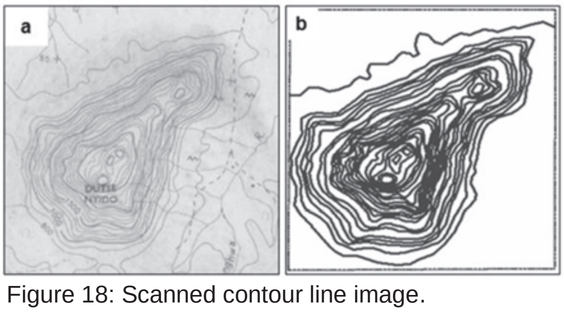


Figure 18: Scanned contour line image.
(a) Subset of contour line image used as test area in brown colour as well as character showing height value. (b) Transparency of deconstructed contour line layer.

features, then executed morphological and filtering operations to eliminate all unwanted information from topographic map but retained contour lines. They further resolved and processed terminal and crossed points and finally, matched as well as reconnected broken contour lines.

Step 1: To commence contour line extraction two OOA based segmentation algorithms were used to attempt contour line recognition on paper maps. Firstly the fast quadtree segmentation algorithm was experimented using a scale factor of 10. The choice of the scale was arbitrary and was applied on two image type mainly the original and the pre-treated image as was done for other transparencies. This was to experiment on how to overcome specific information extraction challenges encountered due to image artefacts when extracting information on paper map. Also an optimized multi-resolution segmentation algorithm was implemented on the original colour and pre-treated images as the very basic OOA step at specified parameters *(scale of 05, shape 0.1 and compactness 0.1)* respectively using the ESP derived scale.

Step 2: Next was to inspect the segmented contour lines using the array of Definiens object properties thoroughly in order to robustly recognize contour line segments and other objects. Here *Brightness threshold value of >=186* and *Brightness mean parameter of >=209* were found useful in recognizing background segments on the original and pre-treated images respectively. Accordingly, these *Mean layer values* were used to merge and classify the corresponding background segments for both images**.** This was demonstrated with the aim of also improving the chances of recognizing the contour lines for further analysis. To recognize the contour lines on the original image, the *Max. Diff. Mean value* was found useful for contour line segment identification. Similarly, *Max. Diff. Mean value* identified contour line segment on the pre-treated image.

Step 3: The proceeding step was to experiment on merging the fragmented contour line segments into complete isolines such that they can be extracted and used for further GIS applications. The region merge was done with *Max. Diff. thresholds of >= 0.40 as well 0.21* respectively. Practically, *Max. Diff.* alone was not sufficient for complete contour line recognition and extraction due to strong interferences emanating from background colour, bleeding and aliasing effects. Consequently, the second approach earlier mention was explored on segmented images and is described in step 4.

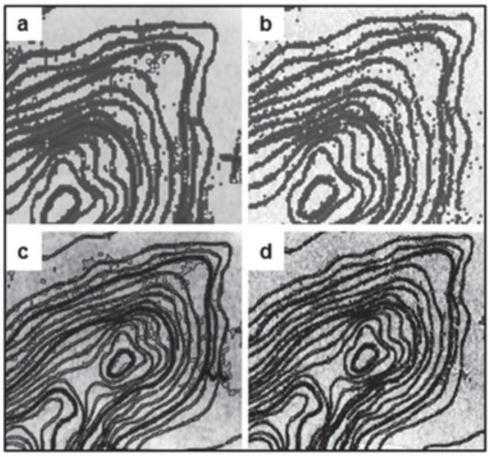


Figure 19: OOA-based contour line region merge.
(a-d) Merged background segments of the original maps. Merged background segments of the pre-treated images to better visualize the contour lines.

Step 4: Apparently, contour lines were at some point disjointed due to either character crossing, colour spread emanating from background which in turn impeded on complete line segments creation in the previous step (Figure 19). Also false colour emanating from scanning effects such as bleeding and aliasing resulted to convoluted segments and posed contour line extraction challenges. Hence morphological operation, convolution filtering and line extraction algorithms of Definiens were used to experiment the possibilities of overcoming these challenges highlighted above on both images. Here, morphological algorithm in Definiens which is an image processing techniques based on mathematical morphology was executed on the segmented image, and then convolution filter was applied on the morphologically treated contour line segments yet contour lines were still disjointed. Additionally region growing algorithm was applied to connect the filtered contour lines with disjointed segments and assign the merged contour line to a logical class with specified rule.

Contour line extraction performance analysis: Contour line was the most complex object considered for recognition and extraction in this research. This necessitated the author investigate the performance of the method and result obtained but output were not satisfactory. Therefore, to compensate for limitations encountered the potential utility of partial result was motivated and demonstrated for contour lines since point height were recognized and extracted. Hence, ancillary knowledge of the height intervals were also investigated from the map interpretation earlier done in (section 3.1.5) which was complementary to the partial results obtained to construct terrain profiles similar to cross section used for terrain assessment.

3.7 Transferability of slightly altered rule-set on separate map section

Recall from (chapter 3.4) that the dataset used in this thesis was initially subsetted and the subset was used as the test area for the OOA based information extraction method development. Here the rule sets developed to extract information from the representative test area used as input data was transferred on the entire map to test for robustness and reliability of the OOA information extraction procedure developed for paper maps. The transferability exercise was a very important step this research demonstrated as a means of ascertaining the feasible information extraction milestones achievable using the OOA based method. To proof the transferability of the OOA algorithm developed, a separate section of the original image was used to demonstrate the performance of adapted rule set with slight modifications. Rule set transferability test was investigated on extracting only two features type chiefly the hydrographic layer and background segments. Results of the transferability investigation are presented in the next chapter.

3.8 Accuracy assessment

Accuracy assessment (AA) is a critical element of any information extraction procedure executed on remotely sensed data to assure its fitness for use for further GIS or remote sensing analysis (Congalton, 1991). In that sense this rule applies to scanned topographic maps since extracted landscape information from scanned paper maps are comparable to remotely sensed imagery. In this thesis AA implies subjecting the recognized and extracted objects from the reference image in comparison to the reference objects base on geometry (i.e. extent), thematic category (e.g. water and vegetation types) and label describing geographic objects (Figure 20). Here extracting the entire features was not the most critical milestone rather the ultimate goal was to correctly determine the entire object extent inclusive of associated text (e.g. recognizing settlement features and associated text in black colour). The accuracy of the extracted hydrographic layer was assessed by boundary correctness and label beside it (Figure 21). For vegetation extent and pattern similarity of the symbols was used including the multi part apparent for the heavy forest and marshland symbol (Figure 21). The accuracy of text was evaluated using by comparing with manually extracted as well as completeness and semantic meaning it describes (e.g. Olobi Creek and Benin River). In the case of contour line AA was done in an objective way by also evaluating how well the extracted contour line correctly corresponds to the reference contour line (Figure 21). Therefore, the OOA based extracted layers were analyzed for both producer and consumer accuracies in percentages. Recall from sections 2.5 that cost and penalty of misclassification was proposed to evaluate the OOA based information extraction method applied on topographic maps for better AA. Also by comparing the OOA based extraction with the manually extracted object using ocular agreement as the criteria (e.g. comparing the hydrographic layers extracted from both the OOA based and the manual extraction method). However, the manual extraction avoided misclassification better than the OOA algorithm due to human semantic learning such that the hydrographic text “Lee Creek” was not misclassified as a water object contrary to that of the OOA.

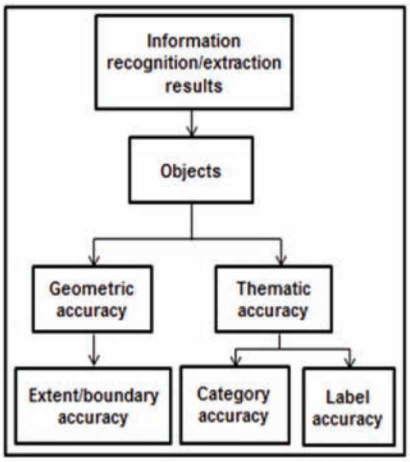


Figure 20: OOA based information extraction accuracy assessment based on objects.

Hence, cost penalty of misclassification was assigned base on logical and semantic decision allocated to the misclassified object to another class using the OOA algorithm. For instance the cost of misclassifying blue text to the hydrographic layer was low compared to misclassifying a black text, settlement or vegetation to the hydrographic layer. Same goes for the vegetation and marshland objects/layer because the mangrove symbol has same colour with marshland objects. Thus, the cost/penalty of misclassifying either mangrove to marshlands or vice versa was high compared to misclassifying mangrove to heavy forest and other vegetation types. Nevertheless, misclassifications were managed by optimizing spatial and contextual attributes to avoid false positives where necessary but for unavoidable cases penalty was assigned logically. False positives were much in the texts recognition and extraction step and this emanated from colour similarity with different features on the map. Cost penalty of misclassification of contour line to other liner features or viz-versa was evaluated based on the importance of contour line for generating relevant information for many applications (e.g. slope derivation etc). The cost penalty of misclassifying contour line to any linear feature was weighted higher.

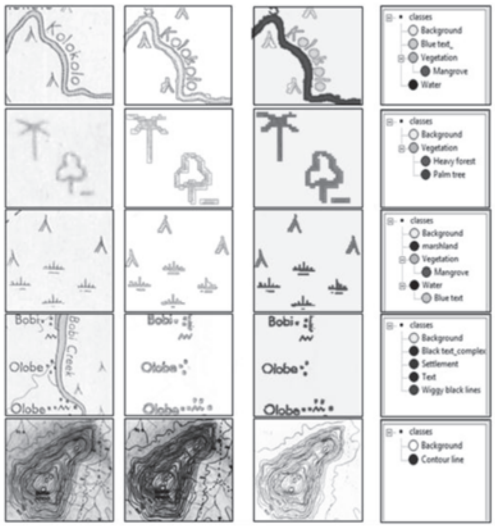


Figure 21: Method applied to assess accuracy of extracted objects.
Extracted objects were assessed based on area of object on the original image in comparison with OOA as well the similar manually extracted object. This was method was applied to all target objects.

3.9 Chapter summary

This chapter described the method used in this research. It has explained the data used, the procedure used to convert the data and the scanning resolution used. It gave a brief account of software's used to prepare the data at the pre-processing and processing levels. The chapter further described the methodology by proceeding to the OOA based logical steps used to treat the topographic map to extract relevant information such as the map pre-treatment concept. The author conceptualized an OOA-based information extraction procedure using the map deconstruction approach such that topographic map was treated as transparencies. Hence for every map layer (transparency) topological and spatial attributes were used to develop relevant rules for automatic recognition and extraction using a multi-step approach. This stepwise strategy suggested a logical process that begins with the easier to complex layers considering the homogeneity of salient objects and semantic expansion (Table3.3) in a logical order empirically verifiable. Aspects and applications yet to be explored are outlined as future research outlook for this area (see Recommendations). A multi-step implementation of the conceptualized OOA method was further demonstrated and the milestones achievable in OOA method to extract information from paper maps were established. OOA-based segmentation algorithms and optimization steps were implemented. Next was to implement a data driven thresholding to robustly extract relevant information from paper maps. Then a transferability test of the once created rule set was investigated to validate the robustness of the OOA method. Accuracy assessment of the method was demonstrated taking into account misclassification costs/penalty to evaluate the method as a quality control measure. In the next chapter, results of the developed and implemented method applied in this research were presented and thoroughly discussed. This was done in a synchronized fashion by presenting and discussing results in the order of research objectives investigated. Accordingly, method transferability test, result accuracy assessment and misclassification penalty were presented and discussed respectively.

CHAPTER 4

4 RESULTS

The previous chapter conceptualized, developed and implemented OOA-based information extraction method for paper maps. The objective of this chapter is to present experimental results documented from chapter 3 in details. The experiment was based on real and simulated data and results are synchronized based on the specific objectives and research questions presented in chapter 1. In this chapter the practicalities of suitable theoretical concepts to investigate and understand the order of extracting layered information contents on paper maps were researched. This exercise facilitated the conceptualization of the OOA based information extraction method for paper maps (section 3.2). Afterwards experimental results of the demonstrated OOA based information extraction method and its transferability on a separate map section was proofed and assessed to determine accuracy, robustness and reliability of the method.

4.1 Research objective 1- Develop knowledge driven information extraction framework for paper maps

4.1.1 Suitable theoretical concepts for investigating, understanding and extracting thematic information contents depicted on historical topographic paper maps.

In order to develop OOA-based knowledge driven information extraction formalism questions pertaining suitable theoretical concepts useful for extracting thematic information from paper maps were answered. Remarkably, the concepts of saliency and semantics were found appealing for investigating; understanding and establishing the order of extracting targeted layered information from paper maps (see section 1.3) and serves as a theoretical contribution to researchers within the GI Science community. Primarily, the framework started with visual image interpretation as a prerequisite to investigate paper maps and was done using the concept of saliency on the input maps shown in (Figure 2 and Figure 3). This detailed visual interpretation facilitated a scientific and thorough development of a knowledge base about the information depicted on the map (see Table 2). The concept enabled visual recognition of objects depicted on the maps. The saliency characteristic as a precondition for OOA information extraction procedure present on the map lead to the identification unambiguous of objects of interest for information extraction. The concept proved very useful for recognizing and systematically description of objects such as the isolated hydrographic information. Indeed the suitability of the concept for analysing and understanding recognized object enabled systematic design of unlocking complex geospatial features using object properties. Since OOA technique has the advantage of using contextual information for feature extraction which includes object morphology, thankfully relevant contextual properties complementary to colour were used where relevant. Theoretically, this notable concept lent a central base in developing a robust OOA based information extraction method. Saliently, colour, shape, and size attributed to target objects of one particular category which do not occur the same in all other categories supported explicit detection of complete set of objects of the targeted information. This analogy applies particularly in the cases of hydrographic, vegetation, text and line objects on the map. Therefore the concept was suitable for developing the OOA-based information extraction method. Complimentarily, another important concept found suitable for establishing a knowledge base for developing robust OOA method to facilitate information extraction from paper maps was semantics. In objects recognition and visual image interpretation, human cognitive capability is swift in semantically assigning logical meaning to what is seen but the computer cannot do that on its own but can be instructed by the human expert. A typical example is the texts types carrying different colour and orientation. Semantically human interpretation associates the blue text to water and the black text to settlement but the computer cannot do this without being instructed accordingly. OOA being a digital pattern recognition system that can extract relevant information on maps and relies on human definition to form meaningful geographic object was systematically instructed to extract such features accordingly. Therefore, to robustly extract objects such as the vegetation types with high variability in size, colour and shape of the graphic symbol as a result of the manual production of the map semantics modelling and expansion was necessary. Although these symbol/objects classes extracted have no physical boundaries and this can be called fuzziness of conceptualized boundaries. Meanwhile efforts to consult metadata about the map production were abortive as such the boundaries to these identified symbols were logically defined in order to reconstruct useful information. Consequently, to delineate boundaries for extracted object, semantic expansion was useful in achieving this decision making milestone and was objectively done and compared with some expert judgement to conclude on a median tolerable reinterpretation.

4.1.2 Order of OOA-based information extraction procedure for paper maps

As no formal OOA procedure for extracting information from paper maps was available to the researcher, a logical approach was established for OOA-based information extraction from paper maps. The author conceptualized a simple to complex information extraction procedure (section 3.1.2, Figure 4). Here hydrographic information was firstly conceptualized for extraction being the most trivial object on the paper map. Next were the moderate objects such as unambiguous symbols chiefly depicting vegetation information and marshland features. However, when geographic information is extracted it is logical such information is annotated for completeness. Thus in OOA text recognition and extraction can be achieved similar to objects as well. Interestingly, text extraction has been recurrent concern in information extraction procedures. Therefore text could be recognized as objects in OOA method and was demonstrated after geographic objects were recognized and extracted such that geographic objects are properly described. Lastly, is contour line and its extraction has been the most complex in geo-information feature extraction tasks. In establishing a logical order for OOA-based information extraction contour line was appraised the most complex too due to its characteristics and tediousness of its extraction process. Altogether, the designed order of extraction proved logical and robust for complex paper map extraction by OOA method.

4.2 Research objective 2- Conceptual framework and formalism of OOA method

4.2.1 How can we develop OOA-based information extraction procedure for topographic paper maps?

Recall that the entire chapter 2 dwelled on documenting state of-the-art underpinning literatures to develop OOA based information extraction framework for topographic paper maps. The chapter also cascaded central problems, key methods and concepts from relevant works were highlighted and contextualized from cartographic recognition and pattern analysis for developing OOA-based information extraction method. These contributions were positioned in this thesis within the corresponding research areas to formalize the OOA method. Thus, the over lift concept proposed and action demonstrated in section (section 3.1.5) resulted in a formalized conceptualized framework in (sections 3.2) which was a map deconstruction approach developed and was useful in the reinterpretation of relevant cartographic information on the topographic map into synthetic map transparencies (Figure 5). Also recall that the map deconstruction approach simply refers to operating a strategic reversal idea which is opposite of the map making process to retrieve targeted cartographic details. Figure 5 shows the targeted map information comprised of (a) hydrographic information (b) vegetation layer (c) Text information associated to different objects and (d) contour lines. These four major classes were conceptualized into object transparencies as shown in (e) of this figure. As mentioned in the previous section also information extraction order preceded from simple to complex shown in (e) of same figure. The conceptualization further established the procedure such that every target layer was mentally visualized as transparencies and works similar to pealing of individual information contained on the map. In the end when information is extracted into GIS ready e-maps layer, they can be reconstructed into new map with updating possibilities (Figure 6). Although this approach seems appealing but can be hindered by series of challenges inherent on complex topographic maps. Hence, some of these complex challenges can be overcome by considering a multi-step information extraction alongside the map deconstruction procedure conceptualized to systematically extract target layer and it is presented and discussed in the next subsection.

4.2.2 How can human cognitive interpretations be translated to develop reliable and robust OOA information methods and what semantic modeling techniques can be applied in extracting geospatial objects from historical topographic paper maps?

This research question shows the result of steps involved to develop knowledge base for OOA information extraction. As previously mentioned challenges present on the map are impeding factors hindering a straight forward information extraction procedure for paper maps. Therefore, derived knowledge from visual interpretation of map was translated into a multi-step OOA approach such that bottlenecks in information extraction are managed for OOA method in a stepwise fashion (Table 3).

4.2.2.1 Experimental results of object oriented analysis

Scale parameter selection:- Firstly, the scale at which the image was segmented in the very basic step of OOA was estimated using the ESP tool developed by Dragut et al. (2010) instead of the trial and error method (see Figure 10). The ESP method was preferred since the procedure is aimed at automatically deriving optimal scale in order overcome under and over segmentation problems from the tool although observation showed additional time is required in the scale estimation process. Hence the image was segmented with scale factor of 7 which was about the smallest though the ESP result was zoomed out to visualize the choice of scale. However, shape and compactness parameter were kept at minimum in order to recognize smaller objects such as the symbols and single boundary hydrographic features. It is important to mention that the estimated scale performed remarkable and the investigated the transferability of the initial optimal scale by estimating the scale parameter for the separate image and 7 emerged to also be the most optimal scale and justifies the scale automation when transferred on separate map section with same imaging condition but contrasted on map with different imaging condition

Segmentation and classification

Multiresolution segmentation was mostly executed on all images (i.e. the original and pre-treated images). Each image was initially, broken down into sub-objects from previous level and were subsequently merged into super objects after target information have been satisfactorily recognized. The result of segmentation, merge operation and classification for the four transparencies (hydrographic, vegetation symbols, text and contour line) are shown sequentially.

A Hydrographic information recognition and extraction

Results of the recognized and extracted hydrographic layer are presented in the subsection (Figure 22). The results show that the choice of smallest scale applied keeping shape and compactness at minimal mostly fits well considering small objects such as hydrographic tributaries, vegetation and non-vegetation symbols and texts. As earlier mentioned proceeding steps varies according to target object/transparencies and complexity. Table 11 illustrates areas where false positives resulted to misclassification and logical penalties were assigned based on cost of such misclassifications.

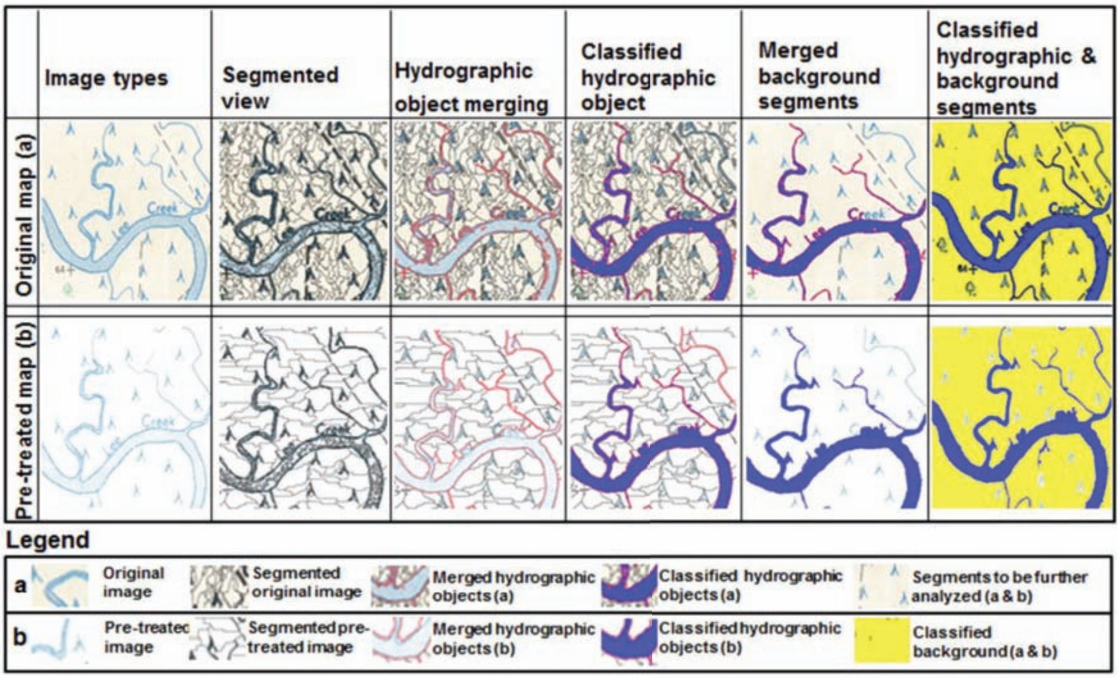
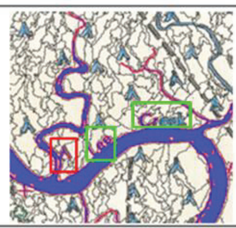


Figure 22: OOA-based extracted information from paper maps.
The first row shows the original image and the transitional steps from segmentation, merging of hydrographic primitives, classified hydrographic segments in blue and the merged and classified background in yellow. The second row shows the pre-treated image and the transitional steps from segmentation, merging of hydrographic primitives, classified hydrographic segments in blue and the merged and classified background in yellow. Legend is included for guidance as well.

Table 11: Treatment of recognized and extracted hydrographic information with false positives

Image type	Challenges	Misclassified objects	Penalty/cost of misclassification	Screen capture
Original map	Color bleeding	(a)Mangrove vegetation to water in red box.	High (refer to section 3.7)	
		(b)Blue texts to water in green boxes.	Low (refer to section 3.7)	
Pre-treated map	Color bleeding	(a)Mangrove vegetation to water in red box.	High (refer to section 3.7)	
		(b)Blue texts to water in green boxes.	Low (refer to section 3.7)	

B Vegetation and non-vegetation symbols/information recognition and extraction

OOA-based method ensured the robust symbol recognition of all the vegetation and non-vegetation symbol prototypes on both images. At the segmentation phase object recognition rate was visually appraised and scored approximately 50% recognition (Figure 23). To robustly and completely recognize target objects a merge operation was executed to improve the symbol recognition and was scored a satisfactory recognition performance as shown in the symbol region merge view in (Figure 23). Recognition performance was carefully assessed by comparing symbol structure on the original image in terms of area, shape and category as shown in (Figure 23). The result presented in this figure confirms the robustness of the OOA-based approach in symbol recognition.

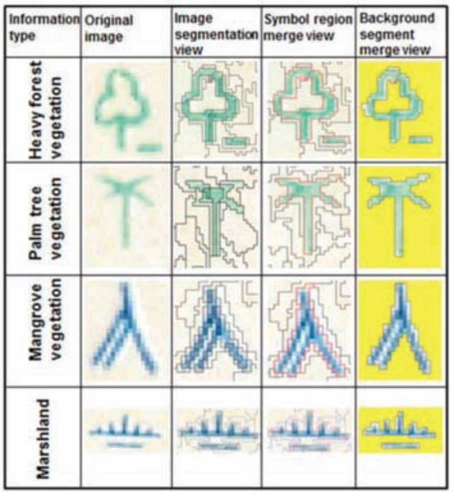


Figure 23: OOA based extracted symbols. This figure shows OOA-based recognized vegetation and non-vegetation symbols. Each row shows specific information prototype and the columns show logical step followed to recognize symbols.

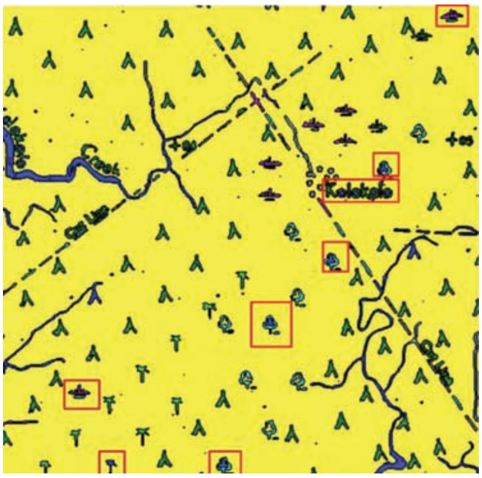


Figure 24: OOA-based symbol recognition and extraction.
This figure show the result of feature space optimization classification using nearest neighbourhood method for symbol extraction with costly misclassification highlighted in red boxes. (i.e. high costs).

Therefore, robust extraction of recognized vegetation information was based on object category. However, extraction was not achieved from the original image using rule-based classification. Hence feature space optimization (FSO) was applied in order to achieve robust extraction. Yet results were not encouraging as only mangrove vegetation extraction was performed compared to other symbols which recorded large amount of false positives from various object due to similarity in colour plane (Figure 24). To overcome the impeding challenges experienced the map pre-treatment approach was adopted here and successful symbol extraction was achieved. Following the result achieved by OOA-based symbol recognition and extraction of vegetation and non-vegetation symbols, extracted it was necessary to implement structural analysis and semantic expansion on these objects. Since information behind these symbols is most paramount and this is the central milestone to complete the process. Therefore a number of techniques were investigated for semantic expansion and spatial partitioning in order to spatially cluster the group of symbol/object to determine their unit boundaries. The convex hull technique was explored as a means of aggregating symbols to establish boundaries for the extracted

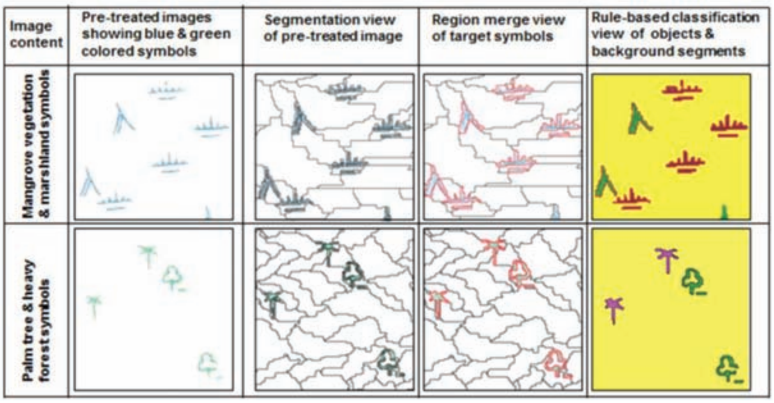
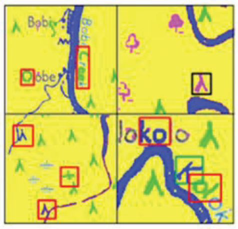


Figure 25: OOA-based symbol extraction from pre-treated paper maps. The first row shows the result of extracted mangrove vegetation and marshland symbol from pre-treated maps carrying blue objects. While the second row shows the result of extracted palm tree and heavy forest symbols from pre-treated maps carrying green objects.

objects and the result of this investigation is presented in proceeding section. Apparently, misclassifications of objects was noticed due to false positive and these errors were appraised and assigned logical penalties based on the cost/magnitude of the misclassification observed Table 12.

Table 12: Misclassification penalties for wrongly recognized and extracted vegetation information

Image type	Challenges	Misclassified objects	Penalty/cost of misclassification	Screen captures
Original map	Colour bleeding & convolution	(a) Black/Blue text to vegetation; mangrove to heavy forest in black box signifying moderate cost; mangrove to water in red box. (i.e. high penalty)	Red boxes high cost (refer to section 3.7). While black box implies moderate cost.	
		(b)Blue texts to water in green boxes.	Green box means low cost. (refer to section 3.7)	
Pre-treated maps	Colour bleeding & convolution	(a)Heavy forest to palm tree.	Green box means low cost. (refer to section 3.7)	
		(b) Palm tree to heavy forest.	Green box means low cost. (refer to section 3.7)	

C Text recognition and extraction

Results of OOA-based text recognition and extraction are presented here. The performance of the multiresolution segmentation algorithm at appropriate parameters alone was not robust for texts recognition initially on the image as a first step (Figure 26a-d). However, to robustly recognize text as logical objects with semantic meaning, it was imperative to apply region merge operation for precision text prototype recognition (Figure 26b).

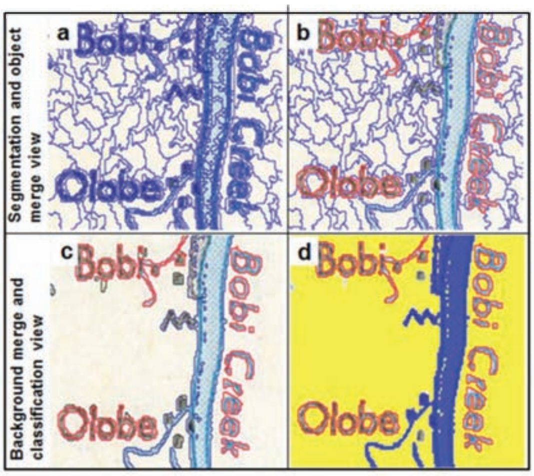


Figure 26 (a-d): OOA segmentation on images. First row shows segmented and merged text on the images. While the second row shows merged and classified background and water segments on the images.

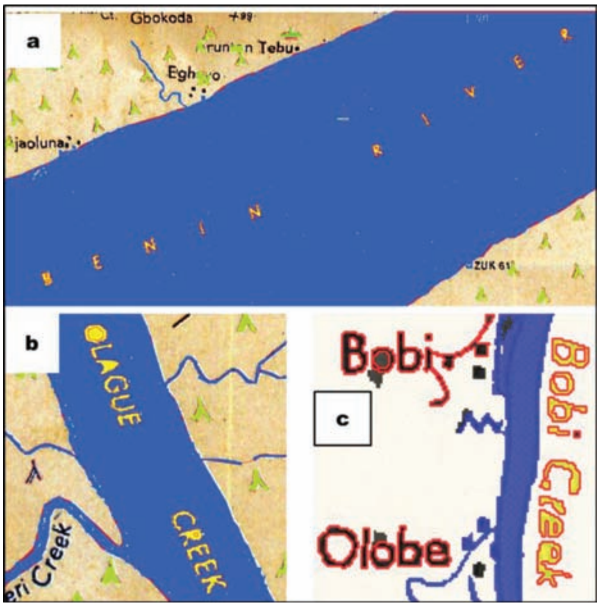


Figure 27: OOA extracted hydrographic texts.
(a and b) extracted blue text inside the river in yellow.
(c) Shows extracted water and settlement texts yellow and black highlighted with red.

Clearly results of the merge algorithm showed high text recognition rate thereby making recognized objects ready for extraction. Accordingly, other segments such as background and semantically important objects (e.g. water) were already merged and classified towards improved visualization text extraction Figure 26(c-d) background merge and classification view. Semantically blue text was associated with the hydrograph information and black linked to settlements, town etc.

On the original image extraction was challenging with high proportion of false positives due to the fact that other semantically important objects have same color with text objects and that remained the chief hindering factor coupled with bleeding, touching, crossing and convolution of these objects. Notwithstanding these challenges extraction possibilities were demonstrated with the river name, major creek and captioned from small image portions of the original image (Figure 27a-c). Similarly, precision text prototype extraction was demonstrated on the pre-treated images as mean of separating the blue and black texts to overcome the recurring extraction challenges (Figure 28a-b). Further investigative steps taken by the author to explain the challenges encountered in OOA and push text extraction beyond limiting factors resulted into creating synthetic text/character image for theoretical experiment. The investigation was targeted toward finding relevant object properties for robust extraction using both FSO and rule based classification. The FSO performed satisfactorily with text but misclassified small text for numbers (Figure 29a-b). However, results of the rule based classification showed that asymmetry and roundness could be potential text extraction parameter if used to extract recognized texts (Figure 29b). Importantly, text extraction showed strong dependency on font type, size and orientation. This was found out after the image was subjected to segmentation and logically merged to form complete text objects and a dictionary of these text properties were investigated (Table 13).

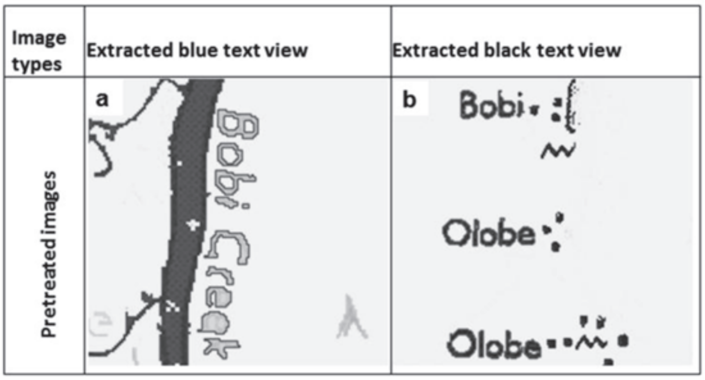


Figure 28: (a-b): OOA extracted text from pre-treated images.
(a) Extracted hydrographic text in cyan with red outline. (b) Extracted place name text for settlement and towns.

Figure 29: OOA based text recognition and extraction.
(a) Shows text recognition and extraction based on multiresolution segmentation and FSO extraction. (b) Shows text recognition and extraction based on multiresolution segmentation and rule extraction information extraction.

Table 13: A dictionary of two potential text extraction properties in OOA

Texts	Asymmetry	Roundness	Texts	Asymmetry	Roundness
A	0.09	1.5	N	0.10	1.17
B	0.12	1.1	O	0.02	1.17
C	0.18	1.38	P	0.3	1.51
D	0.08	1.34	Q	0.1	1.50
E	0.2	1.38	R	0.08	1.41
F	0.4	1.55	S	0.2	1.05
G	0.06	1.3	T	0.4	1.19
H	0.1	1.25	U	0.07	1.52
I	0.8	0.1	V	0.08	1.67
J	0.5	1.55	W	0.3	1.49
K	0.06	1.43	X	0.1	1.21
L	0.5	1.16	Y	0.2	1.52
M	0.2	1.44	Z	0.3	1.33

D Contour line information recognition and extraction

OOA-based line recognition and extraction results are presented in this section. Here results of the investigated techniques are presented as a group and discussed appropriately. Firstly, results of the two segmentation algorithms used on the original and pre-treated images are shown in the segmentation views (Figure 30 a1-d3). From this figure results shows the performance of both segmentation algorithm in line recognition. To visualize contour line segments clearly segments corresponding to the contour line and background layer were merged accordingly as shown in the background merge view of (Figure 30a2-d2).

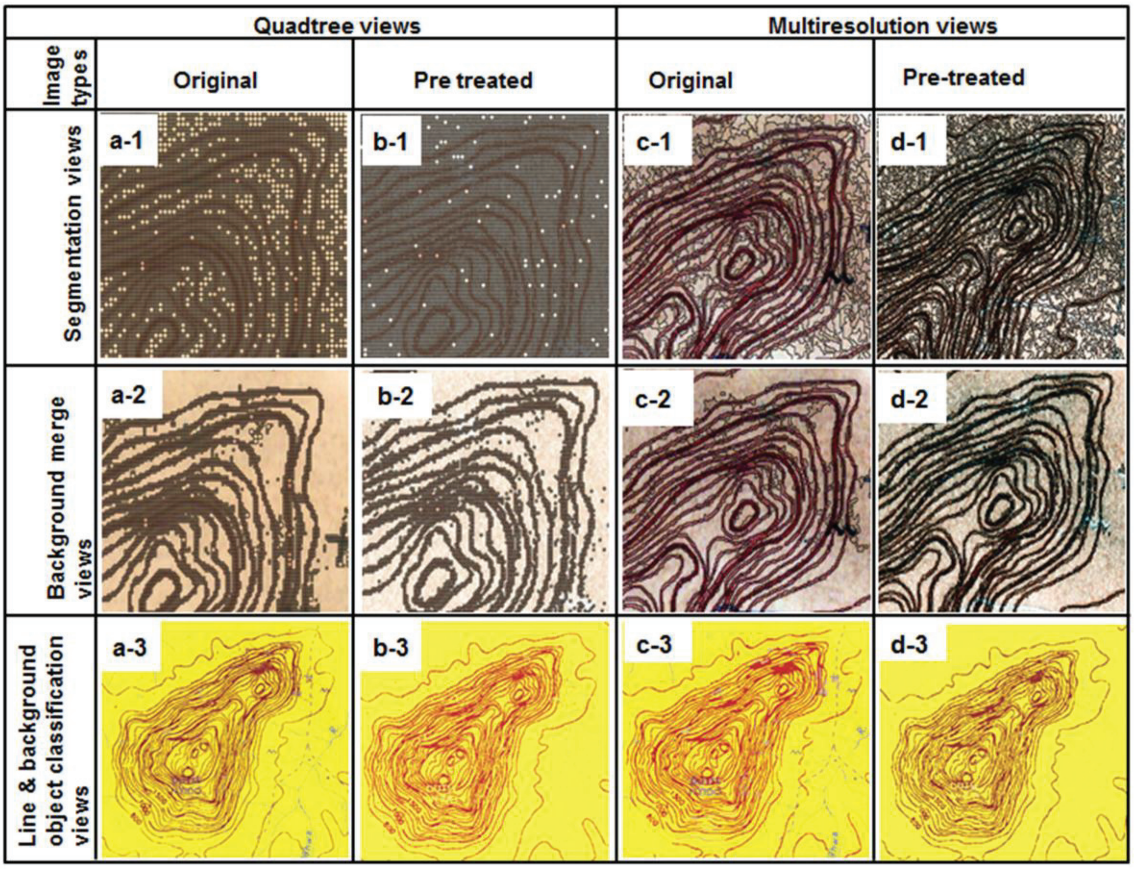


Figure 30: OOA segmentation algorithms on paper map.
(a and b) Left view columns show quadtree segmentation, merging and classification of contour line on original pre-treated images respectively. While (a and b) on the right side view/columns show optimized multiresolution segmentation, merging and classification of contour line respectively.

Consequently, the merged background and contour line segments were assigned to their appropriate classes created. Here the footprints of the contour lines became visible although the merged contour lines did not form complete isolines. As mentioned in (section 3.6D) two methods were proposed. Here results obtained from the second investigated techniques are presented on the original and pre-treated image respectively. Figure 31 and Figure 32 shows results of the convolution filtering, morphologic algorithm, line extraction and merge region analyses to establish the capabilities OOA method in contour line extraction.

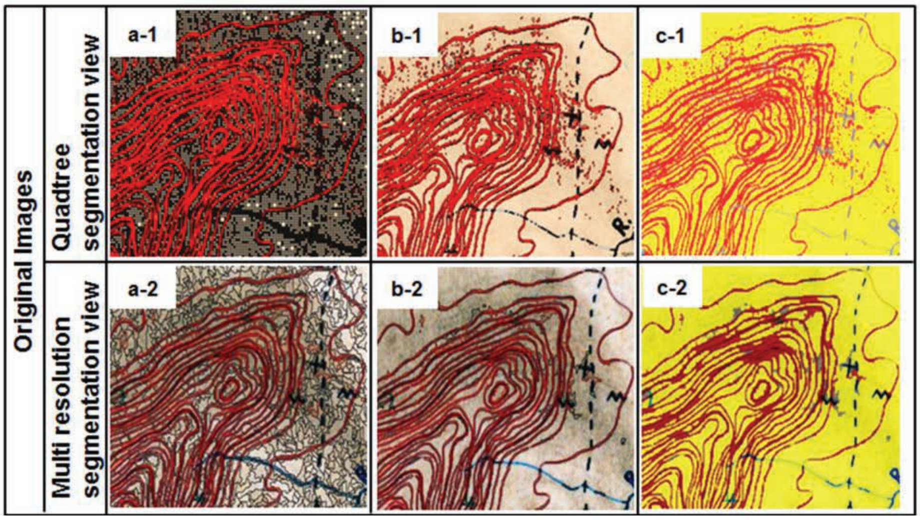


Figure 31: Convolution filtering, morphological algorithm and line extraction on original images.
(a) Shows results of contour line morphology, convolution filtering and line extraction on original images. (b) Background segment and contour line merging and (c) Classified contour line & background segments.

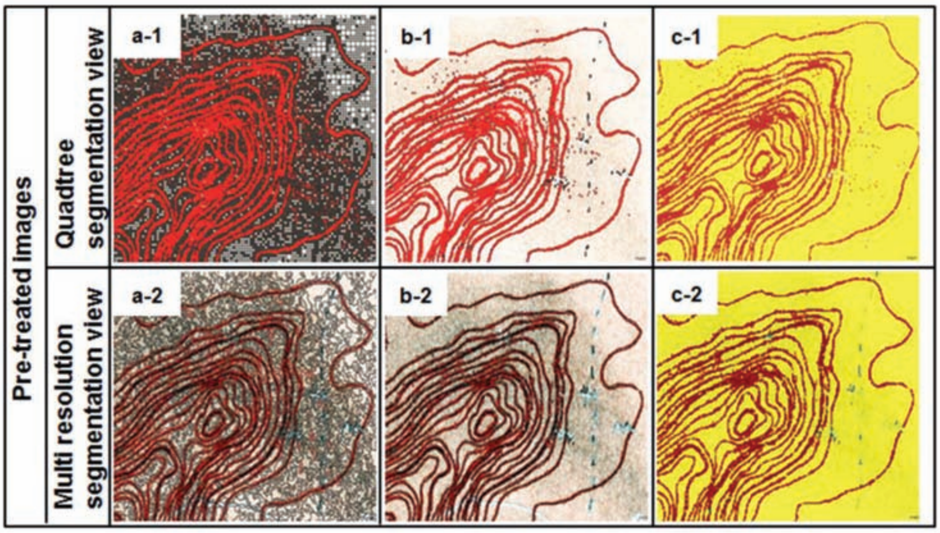


Figure 32: Convolution filtering, morphological algorithm and line extraction on pre-treated images.
(a) Shows results of contour line morphology, convolution filtering and line extraction on original images. (b) Background segment and contour line merging and (c) Classified contour line & background segments.

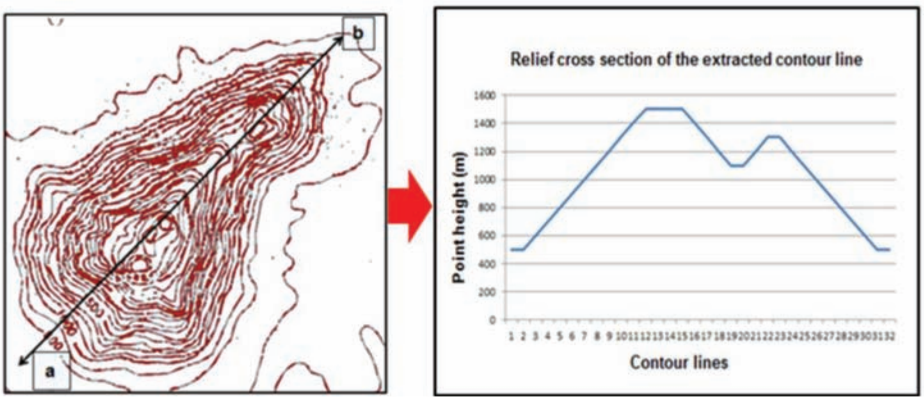


Figure 33: Extracted contour line by OOA.
Left image shows the cross section of the contour line point (a to b). Right shows relief information plotted from the OOA extracted contour line.

4.3 What kind of measurement can be used for delineating geographic features? i.e. how to go from symbol to boundary.

One simple geographic object clustering example was shown in the previous chapter. It illustrated the approach of clustering spatially extracted objects by convex hull. Here the results of convex hull of each cluster were extracted and delineated to indicate objects that belong to a specific thematic landscape category. Example of such results, using the measurement of 0.0174 and 0.5 meter respectively were investigated respectively. These measurements were derived from two sources the initial measurement was from the angular unit of the image and the later was gotten from direct object measurement on the map by ruler. Result in Figure 34 shows zone of uncertainties generated from the convex hull approach used to semantically expand extracted symbol and the zones are highlighted in red outlines. Figure 35 presents the result obtained from the measurement of individual objects derived from ruler on screen. Here individual object were measured at 0.5 meters and was used to implement the convex hull based on object category extracted by OOA. Apparently, performance of the measurement for convex hull did not aggregate objects into a single unit with generalized boundary which can be said to be an underestimation.

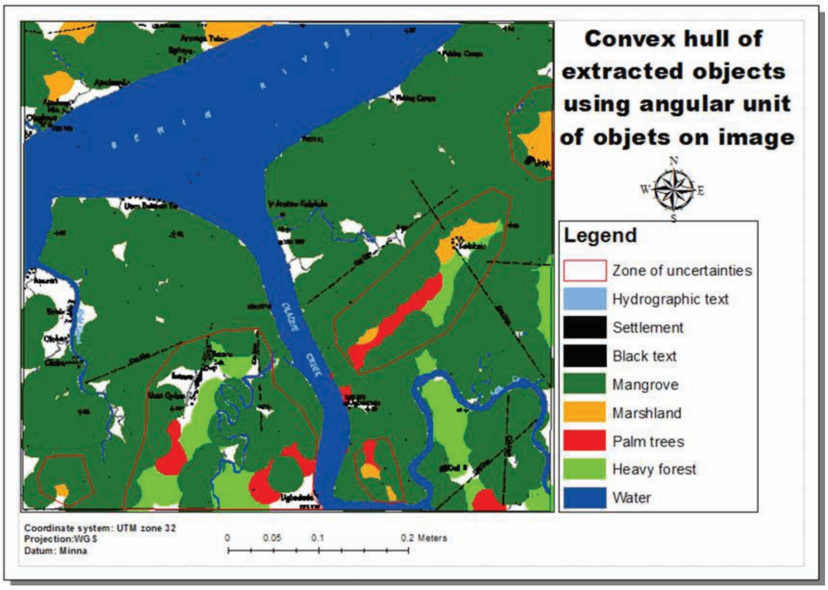


Figure 34: Reconstructed layers of overestimated convex hull of extracted objects into e-map.

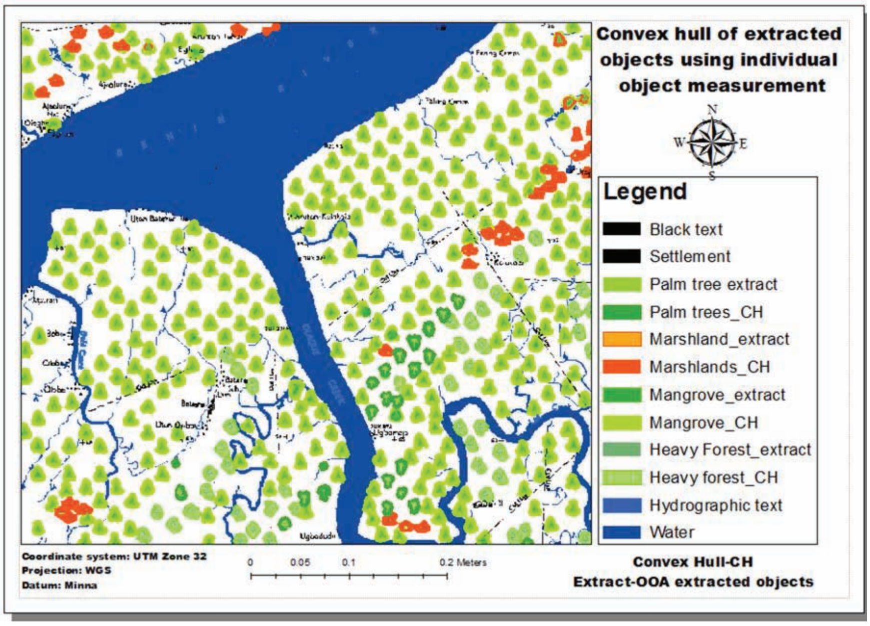


Figure 35: Reconstructed layers of underestimated convex hull of extracted objects.

4.4 *How important is transferability of the once created rule-set and method applicability to other maps from the same series, with the same symbology and imaging conditions but showing another region?*

As proposed in the research objectives of this thesis the importance of transferability will be investigated (section 1.5.2 (VI). Here the results of the transferred once created OOA rule-set are presented (Figure 36). The transferability test was demonstrated on two image types the initial being a separate section of the original image used as test area and later is a separate map sheet with different imaging condition but from same series. It worked efficiently on the map with similar imaging conditions and slight modification. The modification was on the geometry of the merged hydrographic layer since size criterion was used on the region merge and classification parameter. Hence the performance of rule set was subjective to the image characteristics when transferred. But rule set transferability experiment on separate map with different imaging conditions did not work unaltered.

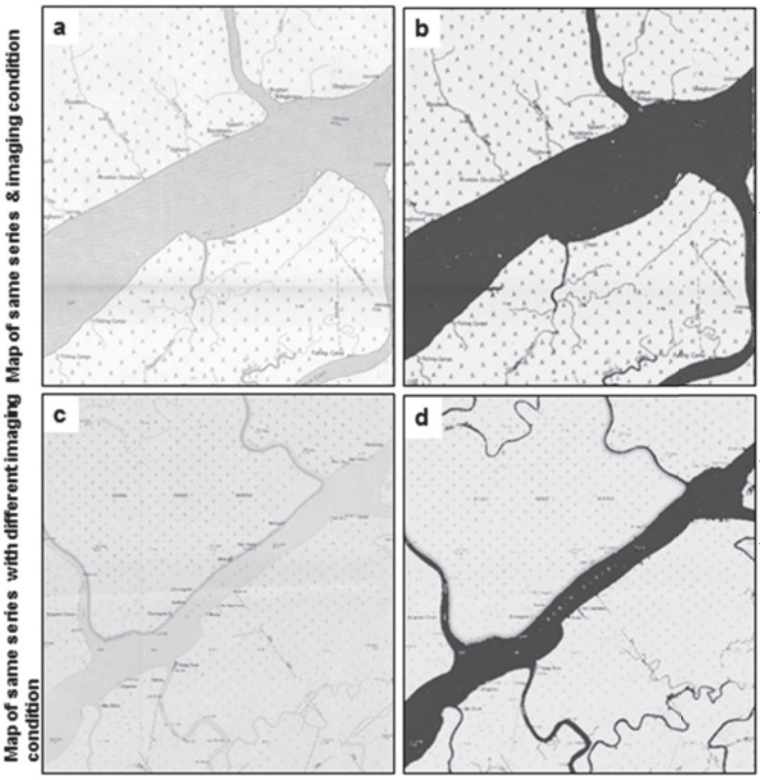


Figure 36: Demonstration of rule-set transferability on different images. (a) Shows separate section of input image with same imaging condition. (b) Subset of separate image with different imaging condition. (c) Transferred rule set on map with same imaging condition to extract only the hydrographic information with slight modification on object geometry and (d) Transferred rule set on map with different imaging condition to extract only the hydrographic information with considerable modification on extraction parameter.

4.5 *How robust, consistent, unbiased will such rule sets be if transferred and compared to ambiguity of manual extraction?*

Altogether, the OOA based information extraction showed varying significant robustness for the target transparencies from the results shown in (Figure 22, 23-25, 26, 28, 29, 30, 31 and 32). Interestingly, the hydrographic layer showed remarkable robustness and consistency in performance compared to the manual extraction procedure used to extract same separately. However, a good agreement was observed between the manually extracted and OOA extracted hydrographic feature although the manual method avoided false positive compared to the OOA method Figure 37. Clearly, most hydrographic objects and symbols were recognized and extracted from original input map by OOA and corresponds correctly with the reference map as well as the manually extracted feature with minimum ambiguity.

However, results of individually derived symbols by OOA-based method were also robust but inconsistency in object shape and sizes were observed similar to that of result obtained from the manual extraction procedures and lacked consistency robustness. Apparently, the manual method was not transferrable for symbols extraction and boundary delineation and this was further established from the result obtained from the manual extraction experiment given to ten experts. However, the OOA method was repeated several times and it showed robustness, consistency and was un-biased while the manual method was biased, not robust and inconsistent. These problems are obvious as shown in Figure 37 with rough, biased, and inconsistent boundaries shown in red and green. This investigation experimented by comparing result of the OOA extraction alongside a series of manually extracted vegetation objects from the result 10 out of 20 experts who responded validates the proof of rule-set transferability in OOA (Figure 37). Although results of some of the manually extracted boundaries were similar with the convex hull boundaries mapped especially the mangrove layer boundary. However, the OOA method showed transferability, robustness and consistency advantage above the manual extraction procedure for information extraction from paper maps.

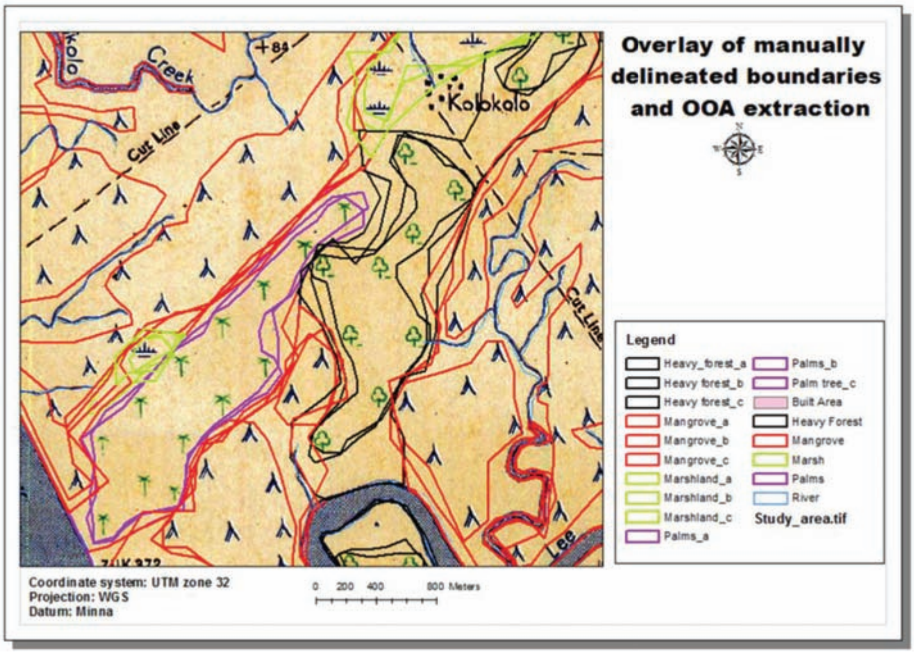


Figure 37 Comparison of OOA based and manually extracted features.
Result shows delineated symbol boundaries manually determined by experts. OOA based extraction in red. Water is shown in different shades of blue, heavy forest and palm in different shade of green, marshland in cyan and mangrove in red colours.

4.6 Research objective 3- To compare the accuracy and completeness of the digital information extraction procedure with a manual extraction method.

4.6.1 To what extent can OOA approach be applied to develop reliable and robust information extraction method how accurate is the OOA method compared to manual information extraction?

The experience obtained from this thesis asserts the reliability and robustness of OOA approach to develop consistent and accurate information extraction method. Since the ultimate measure in any feature information extraction approach is the similarity and matching objective between the recognized, classified/extracted and the reference data hypothetically representing reality. When this is met the higher the assurance of OOA capabilities which establishes how far the method is applicable. Although uncertainties inherent in input data provided for OOA cannot be avoided due to scale dependency of an object based delineation which contains a broad range of ontology making OOA generated products are hard to evaluate in binary mode. However, OOA method demonstrated in this thesis and results obtained by this technique performed objectively in extracting information from paper maps and quality of results are assessed accordingly. Based on the proposed framework for AA in chapter 3, AA of objects extracted from the images was obtained. Since AA of a map object being produced by OOA requires evaluation, the units to be tested are the extracted image objects. The overall accuracy was assessed for extracted objects from the images according to the figure shown in Figure 38. Apparently, the result of the hydrographic layer shows satisfactory geometric and thematic boundary agreements based on overlay inspection in term of boundary similarity and extent matching in ArcGIS. Furthermore, the spatial extent in square meters computed for both the manual delineation and automatic extraction had comparable area measurements of 0.00207cm for manual delineation and automatic 0.00216cm. The observed overshoot can be accounted for as the unavoidable false positives such as texts, and symbols on same colour plane as well as, slight shift between the extraction methods. Hence an overall accuracy of 95% was assigned to the OOA hydrographic object extraction base on the comparison of reference object and OOA extracted object the (Figure 38). Objectively, quality measures are needed in objects perspective and can be done by counting the number of objects correctly detected and those wrongly detected as shown in (Table 14).

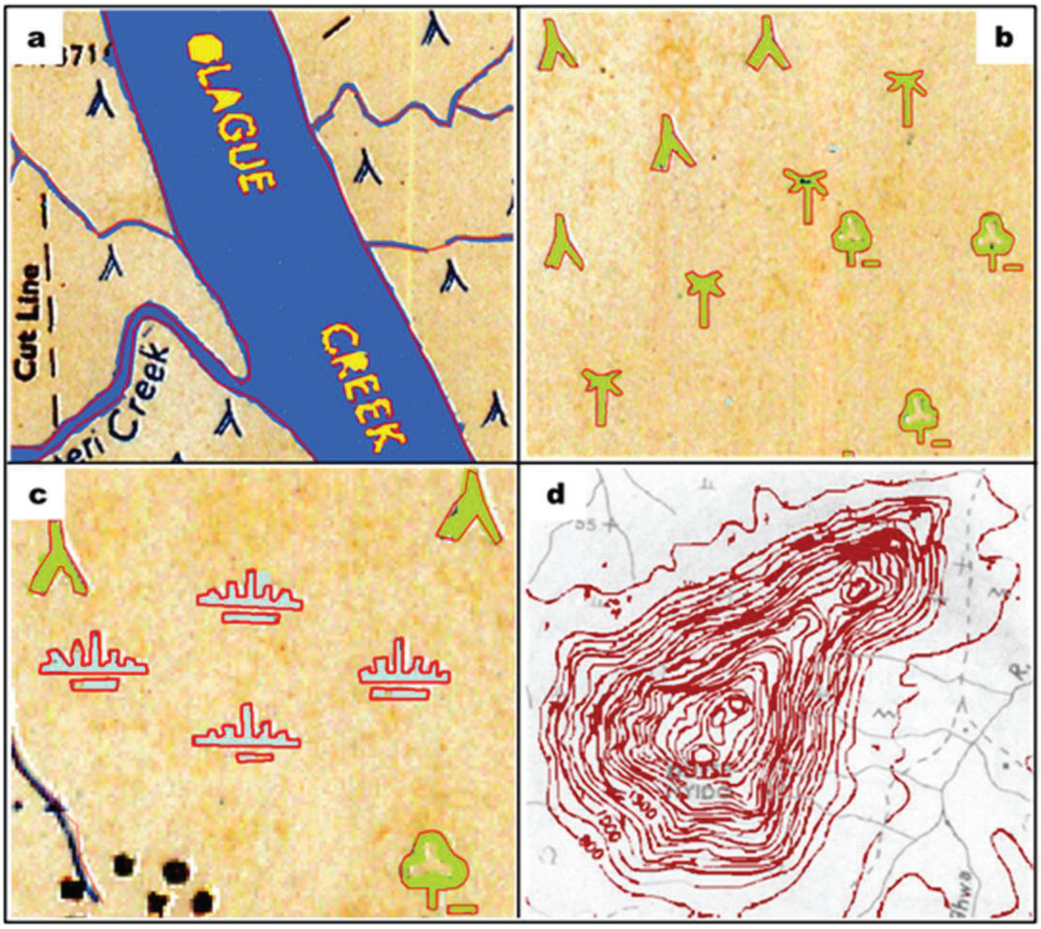


Figure 38: Accuracy assessment of OOA extracted objects.
(a) Shows boundary overlay of extracted hydrographic layers (blue automatic extraction; red manual delineation. Similarly, yellow texts signifies automatic extraction and red outline around the text shows manual extraction (b and c) Shows boundary overlay of extracted symbols (green automatic extraction; red manual delineation and marshland in cyan) and (d) Shows contour line in dark brown overlaid on filtered image.

Additionally, undetected objects can also provide quality information about extracted objects too. However, evaluation here was done for the extracted symbols to establish overall geometric and thematic accuracies for the vegetation and non-vegetation symbols by overlaying both manually extracted and OOA-based extraction to observe extent matching and boundary similarity and this assessment yielded good ocular agreement. Also counting object category for correctness and completeness were estimated as 92% and 97% respectively. Similarly, accuracies estimated for recognition and extraction of text was assessed at 65% and 70% respectively Figure 27 and Figure 38(a) and lastly line extraction was assessed at 55% since only partial information was extracted. Although errors were recorded for every extracted object but varied according to objects complexity which resulted as misclassification and cost of such misclassifications were evaluated.

Table 14: Error matrix for AA by counting numbers of extracted symbols on the test area

Classified data	Method				
	Automatic	Misclassification	Miss	Manual	Correctness %
Mangrove	357	-	12	369	97
Heavy forest	45	-	5	50	90
Palm	24	4	-	20	83
Marshland	25	-	5	27	93
Total	451	-	22	466	97
Completeness %	97	-			92

4.6.2 *What is the cost and penalty of misclassifying objects?*

It is a known fact that there is no perfect classifier. However, results obtained from classifiers must be close to the truth. Here misclassified objects were evaluated using cost/ penalty matrix ranging from low to high. The severity of misclassification determines the penalty assigned (Table 15).

Table 15: Cost /Penalty matrix for misclassifications

Object transparencies	Misclassification cost/ penalty matrix			
Hydrographic layer	**Low**	**Moderate**	**High**	
Blue text	•			Descriptor of extracted object
Black text			•	
Symbols			•	
Settlement			•	
Background			•	
Symbols layer				
Blue text			•	
Black text			•	
Hydrographic object			•	
Settlement			•	
Background			•	
Vegetation	•			Similar object category
marshland			•	
Blue text				
Black text			•	
Symbols			•	
Settlement			•	
Background			•	
Hydrographic object	•			Mother object
Black text				
Blue text			•	
Symbols			•	
Settlement	•			Mother object
Background			•	
Hydrographic object			•	

4.7 Need for OOA-based information extraction method in disaster risk management

The results obtained from the OOA-based method developed for paper maps is argued to be plausible for hazard analysis because of the robustness and reliability potential it lends for rapid information extraction which is pivotal to DRM programs. Furthermore, the utility of extracted information such as hydrographic layer presented in Figure 22 can be used for assessing coastal hazards e.g. integrating the information with satellite imagery can help risk managers to understand potential effects of erosion and accretion in the coastline. It can also be used to investigate the spatiotemporal variability of changes at various time scales in hind-cast modes and can lead to insight about questions that interest's disaster risk management programs and other environmental applications. Similarly, vulnerability and risk assessment studies can be done for the extracted marshland landscape shown in Figure 34 which is known to be a scarce habitat for endangered species and are at the verge of extinction. Hence such studies can be used to investigate how much fragmentation damage has been done to these precious habitat that is useful to man and his environment and subsequently protected and monitored. The extracted vegetation information in Figure 34 can be utilized for assessing the vulnerability pattern of mangrove dating back to the age of the map which also known to be seriously threatened world over. Result of such analysis can be used to enlighten the locals about how unsustainable human actions have affected the vegetal cover which are shelter belts to coastal communities against damaging storm threatening coastal inhabitants. Extent of settlements growth can be assessed by also integrating the few extracted settlement with satellite data to identify and categorize settlements alongside potential risk exposures (see Figure 28). Although complete extraction was not obtained from the contour line extraction shown in (Figure 30 -Figure 32). However, the potential utility of the partially extracted information was also explored to gain insights about the terrain pattern through cross section analysis and the result clearly proved useful since a mountainous impression of the area was established with two peaks Figure 33. Thus, these arguments substantiates the relevance of the OOA-based method and its fitness for use in DRM theme especially for its swiftness in extracting tangible information needed for analysis such as multi-temporal hazards and risk assessment.

4.8 Chapter summary

This chapter presented the results and discussion of this research. It was organized by presenting the result in a synchronized order with respect to the thesis objectives itemized in chapter one. It began with presenting the findings about suitable theoretical concepts for extracting relevant information on paper maps by OOA and was used in this research. Then the outcome of the developed and implemented OOA-based method was logically presented and thoroughly discussed respectively. The experimental result included extracted hydrographic layer, vegetation and non-vegetation symbols which were semantically expanded using convex approach. Comparisons of the manually extracted objects were presented from 10 out 20 of respondents. For each information layer extracted result was shown, discussed and false positive present were assigned to misclassification cost/ penalties depending on the severity. Also the chapter presented the result of transferability test investigated and emphasized the importance of the step in OOA. Additionally, accuracy assessment of extracted objects by OOA was compared with manually extracted objects and the relevance and applicability of the method and extracted objects was emphasized in DRM context. The proceeding chapter presents conclusions; outlook and limitations.

CHAPTER 5

5 DISCUSSIONS

The main objective of this study was to develop a robust and reliable OOA-based information extraction procedure for complex paper maps. Hence the research was structured into three major parts which began with the conceptualization of the entire procedure. However, an in-depth investigation of theoretical issues and recurrent trends proceeded in chapter 2 of this thesis. Here state of the art scientific literatures aiming at semi-automatic and automatic information extraction were thoroughly reviewed. This major chapter investigated underpinning theoretical concepts suitable for developing OOA-based information extraction framework for topographic paper maps. Highlighted works were cascaded further to gain insight about specific problem and related works in pattern analysis and cartographic recognition suitable for developing this method. The second part focused on method implementation and the third aspect investigated the transferability of the developed method on a separate map section to validate the applicability of the method. In this chapter presented results in the previous chapter are thoroughly discussed in a synchronized order in relation to the research objectives.

5.1 Methodological approach

5.1.1 Method conceptualization

Objectively, findings of this research revealed the suitability of two underpinning concepts for investigating, understanding and conceptualizing a robust and reliable OOA-based information extraction procedure for paper maps. The concepts of saliency and semantics facilitated the building of fundamental knowledge that was utilized in visual interpretation of the maps used in this research. Table 3 in (section 3.1.5) was the product of knowledge derived from the visual map interpretation done for the topographic maps. From the visual interpretation each of the relevant salient information were identified, investigated, understood and modelled as a transparencies similar to onion pealing approach and this systematically facilitated the successful extraction of thematic information encoded on the input maps (see Figure 5). In addition the concepts helped in establishing the logical order for extracting information taking a simple to complex configuration as illustrated in Figure 4. Consequently, for each the modelled target layer considered, relevant logical steps required to robustly and reliably extract such features in an OOA context were sequentially outlined in a multi-step fashion in order to robustly recognize and extract these tangible information. Fundamentally, relevant object characteristics such as colour, shape, size orientation formed the basis of object knowledge and was translated into OOA rules for guidance. Apparently, the multi-steps information extraction procedure for paper maps shown in Table 3 established the fact that OOA based-approach is not a straight forward procedure as would be thought of like the segmentation and classification process. Thus at specific object recognition and extraction level the multi-step processes demonstrated for developing the OOA-based information extraction from paper maps was logical and pragmatic to objectively extract target information on input maps in a stepwise approach considering inherent map complexities. As noticed from the performance of these concepts for developing robust and reliable information for paper maps, the suitability of the concepts is in agreement with recent researches that applied same concepts for unlocking useful information from paper maps. Hence the works of Kerle & de Leeuw (2009) and Leyk et al. (2006) were resourceful in establishing the suitability of saliency and semantics for information retrieval from simple paper maps which this research further demonstrated on complex paper maps using OOA-based method.

5.1.2 Method implementation

5.1.2.1 Data set-up scale parameter selection

This research used a representative image section as a test area and this approach was adopted to manage time and overcome computing time and memory challenges which the software could not process. Similarly, the test area approach was conceived and motivated in order to investigate transferability of the developed method on separate map section.

5.1.2.2 Scale parameter selection

To begin information extraction using OOA-based methods segmentation techniques forms the very basic and fundamental step that facilitates extraction of spatial object from the image scene. Image segmentation is primarily used to break down image into meaningful primitives. However, deciding what parameter to use has always been the challenge in image segmentation which if not optimized results to under or over segmentation issues. Hence to avoid this problem recently developed estimation of scale parameter (ESP) tool by Dragut et al. (2010) was used to segment the input image for this research. The estimated scales were randomly tested and they generally produced good segmentation result that correspond to target object but considering the target object this research aimed to extract the smallest scale was adopted in order to identify small object dependent to scale parameter. To identify the estimated scale it was necessary to zoom out the graph produced by the tool. Thus the ESP tool enabled an objective estimation of most optimal scale parameterization automatically to perform image segmentation being the first step necessary in OOA-based information extraction approach. Similarly the estimated scale was validated for transferability and good segmentation was also achieved and this investigation further confirmed the transferability of the estimated scale for maps of same series with same imaging conditions.

5.1.2.3 Results discussion

Segmentation was the basis of recognizing and extracting targeted map information. Thus suitable segmentation should robustly and completely recognize objects of interest. However, a one-step best fit segmentation was not achieved for the entire target information as objects were mostly over segmented and negligible under segmentation cases as shown in the segmented view (Figure 22) despite the optimized scale parameter used. Therefore, a region merge operation rule was necessary in order to accurately recognize the hydrographic information robustly. The region merges rule cuts across all created object primitives with relevant parameters specific to target objects for complete object recognition. Results of region merge for the targeted transparency shown in column 3 of (Figure 22) performed satisfactorily by forming an ideal hydrographic object. Consequently, a rule based classification was applied using the size criterion to logically assign the recognized object as shown in blue in the classified hydrographic object column. The two next columns in Figure 22 further illustrate the merging of the background segment and classification accordingly. Challenges encountered in the hydrographic layer recognition and extraction processes such as the prevalence of false positives were finally evaluated with misclassification penalty matrix introduced in chapter 3. The resulting misclassifications were due bleeding and convolution of the hydrographic objects with other blue coloured objects and are shown in Table 11. Despite the map pre-treatment technique applied and manual intervention the bleeding, convolution and aliasing challenges were not entirely supressed. Hence the false positives were managed by evaluating cost of misclassifications and assigning logical penalties with regards to misclassification severity. Table 11 shows some areas where false positives were present and are highlighted in colour in boxes signifying misclassification severity. Here the green implies low misclassification cost and red implies high misclassification penalty.

Similarly, to recognize and extract symbols the multi-resolution segmentation algorithm did not completely recognize objects as a first step in OOA procedure. However, region merge algorithm was useful to robustly and accurately recognize symbols (Figure 23 and Figure 25). Apparently, the result shown in Figure 24 confirms the poor symbol extraction performance on the original image despite the FSO which was due colour impediments as well as the difficulty of identifying robust object properties that allow explicit extraction especially on the original image. The result showed abundance in object misclassification highlighted in red boxes signifying severe misclassifications in Figure 24 (e.g. misclassifying background segments as water within the heavy forest enclosures etc). Also aside from the samples taken all other FSO classification were misclassified excluding the mangrove vegetation symbol that performed slightly better. Although, useful object properties that could best discriminate symbol was suggested by the FSO result and if investigated with time can lead to better extraction rates *(Roundness, Asymmetry, Max. Diff., Mean Green and Length and Width).* Efforts made to further investigate these properties using rule-based extraction did not yield interesting results due to inherent complexities and data quality. However, results of vegetation and non-vegetation symbols extracted in Figure 25 showed remarkable performance using the image pre-treatment technique which substantially overcame the extraction impediments encountered directly using the original image for OOA extraction. Subsequently, where false positives resulted in misclassifications logical penalties were assigned accordingly (Table 12).

Interestingly, OOA segmentation and merging algorithms were plausible for text recognition. However, extraction proved challenging especially from the original image. This was largely due to text orientation, color bleeding, convolution and crossing of features inherent on the map and impeded on robust text extraction from paper maps. This necessitated the use of the earlier devised map pre-treatment procedure as a means of overcoming prevailing challenges in OOA-based information extraction for paper maps. This approach yielded promising results especially with the black texts associated with the settlement layer. Although high amount of false positives were present but can be hypothetical eliminated when exported as shapefile and edited carefully and leaving relevant black texts. However, performance of blue text extraction was tasking due to map complexity and false positive emanated from vegetation symbols such as mangrove and marshland. Despite the map pre-treatment procedure substantial proportion of false positive recurred and robust OOA text extraction was not achieved. The misclassification of text to other classes was largely experienced as such cost of misclassifications was generally high. However, results shown Figure 29 (a-b) gotten from the theoretical experiment clearly established the fact that if asymmetry and roundness parameters are explored robust text extraction could achieved. Recall that the experiment was performed on synthetic image to investigate suitable parameters that could facilitate robust text extraction and findings from experimental result confirm the dependency of precision and robust text extraction on font type and size as well as orientation. Therefore, to achieve robust OOA-based text extraction advanced text matching algorithms will be required taking into account sensitivity of text to orientation, font size and types.

Performance of OOA segmentation algorithms was robust to recognize well-defined contour lines but showed some limitation in recognizing convoluted section of the contour line sections. For contour lines close to each other specific contour line recognition was not achieved but rather a group of convoluted contour (Figure 30 -Figure 32). Similarly, situations where crossing between two or more objects intersect, continuation conditions were not met and OOA-based line object recognition was not achieved.

Occasionally this was linked to the interception by characters signifying height points. This was actually the case on the original map making recognition rate low using both segmentation algorithms. This necessitated the introduction of the earlier devised map pre-treatment approach but recognition rate only improved slightly above average on pre-treated images (Figure 32 a1-c3). Apparently, the pre-treated images slightly performed better than the original image as recognition was more prominent on the pre-treated images using both segmentation algorithms. Contour line recognition and extraction was objectively established using the simple OOA approach but results were obviously not impressive. Accordingly, OOA capabilities were further investigated using other algorithms to advance contour line extraction. Here morphologic algorithm was initially applied on both segmentation algorithms and images (Figure 32a1-c3). This algorithm was used to investigate how well the recognition rate will improve compared to the initial method demonstrated. Especially on the pre-treated images since the procedure was a plausible approach in reducing the effects of bleeding and crossing of objects on paper maps. Hence expectation was that line morphology will be well established to unambiguously recognize and extract contour lines but that was partially achieved. Secondly, convolution filter operation used to investigate how well the algorithm will further reduce the effect of convolution on closely spaced contour lines since this was a hindrance in the first approach. However, results showed no improvement after applying these algorithms and this was partly due to the strong colour bleeding and convolution in places with closely rendered contours. Apparently, Figure 30Figure 32 clearly shows how convoluted the problematic zones look after assigning the contour lines to appropriate classes and this areas were confirmed convoluted on the original map. Altogether, both OOA methods investigated could not recognize contour lines completely despite the logical OOA steps implemented on the images. However, this thesis fundamentally investigated the applicability of OOA to empirically extract contour lines. Vital findings established from experience gained in this research revealed the limitation of OOA method in contour line extraction on low quality complex maps considering inherent challenges on paper maps. There was no accuracy measures provided separately for line extraction since its performance was not encouraging for subsequent GIS applications although visual appraisal can be made from (Figure 30 - Figure 32). Nonetheless, the partially extracted contour lines were explored to derive useful information and this was unlocked using cross section profile approach. Here density of contour line which conveys some relief description about the terrain can be obtained from the partial information extracted as well the height points which are very useful for elevation determination (Figure 33).

Recall that symbols were extracted but it was important to move from symbol to area. Hence aggregation was done using convex hull in ArcGIS. Objectively, the idea behind the convex hull approach was to semantically expand the extracted object to useful information such as extent of coverage for specific vegetation unit. However, the convex hull implemented using the angular unit measurement of 0.0174m over performed in delineating boundaries of extracted symbols by crossing outside vegetation area or into water layer and other objects. Consequently, this challenge was managed by performing some logical GIS spatial analysis such as clipping and erasing operations on the over fitting area on the water since the water boundary has been initially defined (Figure 34). Although over estimations across other symbols in the same category were not modified and can be considered as areas of uncertainty which exist all over the map. These areas are delineated in red and with such existential uncertainties and where the semantically expanded objects overlapped can be validated with field data to classify such zones as either areas thematic or spatial uncertainty. Similarly, extent of individual object on the original map was extracted and this onscreen measurement of 0.5m was derived by measuring object extent as it is done in conventional cartography using software ruler in ArcGIS. Here the convex hull performed only around individual objects leaving lot of empty spaces indicating fuzzy boundary as it was on the original map although at some points the convex hull over-estimated object boundaries by crossing outside vegetation area or into water layer Figure 34. Despite the under and overestimations challenges observed in the convex hull approach to semantically expand extracted objects in order to determine extent of coverage estimation with varying uncertainties prevalent the procedure can be said to be a logical approach for clustering extracted OOA geographic objects to semantically expand information behind extracted objects.

Investigation of transferability is an important aspect of this study and from the results shown in (Figure 36) it is logically arguable that the transferability possibilities OOA-based information extraction procedure offers is very crucial in the semi-automatic method compared to manual driven information extraction methods. Especially, considering the numerous challenges encountered in manual method and non-transferability constraint which leads to inconsistencies when repeated either by same expert of different expert. Transferability of the once created OOA rule-set and method applicability to same or similar images is ultimately desirable. The demonstrated rule-set transferability proved logically significant and pragmatic in OOA. The experiment was a rapid procedure used to extract target information from paper maps and it was found to be a consistent information extraction method which is unrealistic in manual based methods. The rule-set transferability performed satisfactorily with reduced time investment compared to the required time when developing the rule-set. The transferred rule-set was faster than other methods especially on maps with same imaging condition requiring slight modification. However, the performance of the transferability test on a separate map sheet of the same series with different imaging condition was in-efficient. Hence substantial modification of rule-set was done to efficiently and robustly extract target information which was also faster and consistent than the manual method.

Visually, good extent agreement was established between the manually extracted water feature and the OOA-based method and this was validated by overlay analysis in ArcGIS. However, the OOA-based extraction was accompanied with false positives mainly small bright pixels with colours close to the background segments as well as text and symbols while the manual method avoided such errors. Similarly, a number of extent generalizations, inconsistencies, and biasness were observed within and between the OOA method and manual extraction technique on the extracted symbols. These irregularities were detected with the individual symbol extraction and aggregate boundary delineation from the manual delineation result obtained from experts. Clearly Figure 37 illustrates the variation of inconsistencies and biasness present on the manually derived boundary delineation prominent all over the map coupled with the fact that it is time consuming. Theses inconsistencies make the OOA method consistent compared to the manual method. Hence, OOA extraction was confirmed to be robust, consistent and unbiased and there was remarkable time lag reduction between the rule-set development phase and transferability by more than half of the time invested in extraction. In that sense OOA-based approach is fundamentally a faster information extraction procedure since manual method is known to be time consuming coupled with biasedness and inconsistency from the derived information. Errors encountered in OOA method were compensated by the applicability of the conceptualized method, fast and semi-automatic process and the gains achieved with the map treatment approach. Objectively, since the ultimate goal of this research was to develop and investigate how far OOA can perform in extracting target information from paper maps and this was realistically achieved with verifiable explanations about consistency, un-biasness, robustness and satisfactory accuracy. Thus OOA-based approach is indeed a promising method and can be further developed to complement existing methods as well as improve information extraction in aspect where manual methods are inefficient.

Apparently, AA of extracted object can be evaluated in two ways. This is because this work began with object recognition after which recognized objects are subsequently extracted. Recognition rate was generally remarkable for all target objects especially for the hydrographic information. Similarly, other objects such as the vegetation symbols had remarkable recognition rate with impressive visuals despite colour match with the hydrographic layer. Recognition rate in texts showed promising result but requires further in-depth investigation for robust extraction being a topical subject in recognition science. Lastly, accuracy of contour line recognition was low especially in problematic regions while in areas with well-defined contours recognition rate was fair. Apparently result showed that contour line recognition did not perform well and indicate the limitation of OOA for information extraction method. On the other hand accuracies of extraction rate measured differently. Here extraction rate of the hydrographic layer also performed remarkably with an overall accuracy of 95%. Considering the success rate of the extracted hydrographic and associated text qualitatively extraction is complete here since accuracy criterion of geometry, category and label is met. The accuracy of symbols was assessed at 97% for correctness and 92% completeness. Hence, the accuracy of extracted symbols is considered complete. The result shown in Figure 38 and Table 14 indicates that the extracted symbol conform to reference data and also when compared with the manually extracted. Similarly result of the semantic expansion done by convex hull aggregation is in agreement with the area each symbol represent and existential thematic and spatial uncertainties are present and were defined as zone of uncertainties. The accuracy of text extraction was generally impressive with accuracy of 70% and this was achieved by manually extracting relevant text object and superimposed on the OOA-based extracted text. This was done specifically for the blue text belonging to the hydrographic layer. Although, text extraction in OOA-based approach was not robust due to difficulty in identifying relevant object properties suitable for robust extraction but OOA-based text extraction capability was investigated and established. Line extraction was the most challenging, both recognition and extraction results were not encouraging due to occasional interference. However, relevant information such as contour line density was examined across the image. Notwithstanding the limiting factor that hindered complete contour line information extraction, the partial information extracted was evaluated with cross section charts for relief understanding and height points Figure 33. This was achieved from the extracted contour line and point heights and was complimented with derived visual map interpretation to unlock the contour line intervals which was used for the profile. Generally, AA done for the extracted object was largely by visual evaluation and overlay of both manually and automatically extracted object the reader of this work is also welcome to judge the accuracies obtained for comparison. However, the author opines that improved accuracy assessment can be a forward to push this research beyond current limitation.

This research did not explore the extracted information for multi-temporal hazards and risk assessment as well as other potential uses of the result due to time limitation. However, the method developed was efficient in extracting useful information that can be utilized by government agencies and disaster managers since good accuracy was achieved some objects except contour lines. The applicability of the developed OOA-based method for paper maps is briefly presented in disaster risk management perspective. Especially since the data used in this research was a coastal community map of the Niger-delta area of Nigeria. Hence, this substantiates the development of this method within the disaster Geoinformation management theme considering hazards in coastal communities. It is obvious that numerous multi-temporal hazards and risk assaults from nature over the century, including flood, sedimentation, formation of new deltas and erosion can be reduced with GIS proffered solution. The constantly wreaking damage caused to such areas are process based by man or nature and the utility of OOA-based derived information especially from paper maps having long history about formation and development pattern of such locations. Such information can be very useful in modelling potential multi-temporal hazard and risk assessment in such areas. For instance the extracted hydrographic layer from the input map can be used to monitor and understand the changing river course when integrated with more recent satellite data of same location. Such analysis can reveal potential damaging flood and erosion hazards to inhabitants who have perhaps encroached into flood plains, newly formed delta/back swamps and riverine settlements along the river banks. Similarly, this research extracted marshland landscape feature from the test map and the importance of this scarce landscape as well as the impact risk of marshland habitat fragmentation can be studied to assess whether this geographic feature exist anymore. If they do the spatial network of marshlands can be mapped and protected against biodiversity vulnerability since they are nature conservation zones etc. Also vegetation objects extracted such as mangrove, heavy forest, and palm trees are important landscape features that need to be mapped, quantified and monitored since they serve various purposes to the coastal communities e.g. Shelter belts against heavy storm and direct or indirect effect of tsunamis peculiar to coastal communities. Thoughtfully, the Niger-delta is located where extreme rains are potential hazard and can result to inundation and massive erosion and accretion in rainy periods. Hence the extracted features can be integrated with remote sensing data, statistical and hydro meteorological as well as local information to objectively monitor the trends of such disasters. This timely information can educate such communities about hind-cast unsustainable loss of vegetal cover that should have serves as shield against deadly storms and if their actions are not controlled might pave way for extreme natural condition likely to happen in the near future in coastal zones. Additionally, dynamics about expansions from village to town and cities can be assess and potential risks in such coastal communities can be studied with extracted towns in case of for resettlement purposes though that was not the research focus but is argued for the purpose of buttressing the relevance and need of this robust method for DRM programs. Thus many other thinkable applications includes assessing potential hazards from historic mines sites which have been overthrown by settlement and such information are only available on paper maps which this developed method can lend itself to robustly extract when needed.

5.2 Chapter summary

This chapter discussed the results obtained in this research. The discussion was synchronized systematically in the order of the thesis objectives and structure itemized in chapter one. It began with presenting a brief introduction of the research aim. Then data setup and scale parameter selection was discussed being fundamental to the very basic OOA method. Consequently experimental result was discussed in the order of method implementation, transferability, accuracy and relevance/applicability issues. The proceeding chapter presents conclusions, recommendations, outlook and limitations.

CHAPTER 6

6 CONCLUSIONS, RECOMMENDATIONS AND LIMITATIONS

This final chapter summarises the main research findings and recommends future investigations to advance the field of geo-information science and earth observation techniques for extracting information from paper maps using advanced techniques such as OOA. Some limitations in the developed method are also communicated.

6.1 Conclusions

The purpose of this study was to conceptualize, develop and implement OOA-based information extraction procedure for complex paper maps and test the transferability of the method on the same and similar map series. The idea was conceived due to the growing demand for advanced information extraction techniques in geo-information science to complement conventional methods in semi-automated fashion. Such that time invested in manual extraction, inconsistencies and bias can be reduced to the optimum. With the potentials of OOA it was necessary to investigate capabilities of OOA to robustly, completely and reliably extract relevant information to undermine the challenges experienced from conventional methods. Generically, result of this study can be used to robustly and swiftly extract information from paper maps for further GIS applications e.g. DRM programs. Based on the results obtained in chapter 4, the following conclusions were reached.

- The concepts of saliency and semantics identified as suitable theoretical basis to investigate understand and develop OOA method worked efficiently for extracting information from paper map using eCognition software. Accordingly, the knowledge developed allowed the author established the logical order of extracting relevant information encoded in paper maps for further GIS analysis and is argued to be suited for other applications in the DRM theme.
- Similarly, these concepts were useful to systematically conceptualize the OOA-based information extraction procedure and formed the umbrella to determine the step-wise sequence of the method. The procedure being a map deconstruction approach visualized thematic contents embedded on maps as transparencies and entails a map reversal method similar to the onion pealing process and reconstructed back into GIS ready e-maps with updating advantages.
- The map pre-treatment step established was effective in reducing the subjective impediments encountered using the original image. Hence the map pre-treatment approach objectively facilitated OOA-based method in extracting information from paper maps by overcoming some of the aliasing, colour bleeding and object crossing impediment inherent on the map.
- The multi-step approach introduced overcomes some of the inherent map complexity known to hinder information extraction from paper maps. The procedure showed robustness in the result of the method implementation to unlock targeted information.
- The OOA algorithm developed and implemented performed remarkably in extracting the target information depicted on the input map. This was due the data driven thresholding nature of the OOA method. The performance of the algorithm varied across the target information transparencies. Generally, recognition rate recorded using OOA-based method was realistic on paper maps. The performance evaluation showed that the algorithm was robust in recognizing and extracting the hydrographic information with 95% although few challenges were persistent such as false positive resulting misclassification and over or under fitting issues despite the map

pre-treatment approach. Similarly, symbol recognition and extraction using the algorithms recorded high performance with 97% completeness and 92% correctness. Challenges experienced were linked to the representation of different thematic information and category on same colour plane and unique object properties to robustly extract the various symbols was difficult to find. However, the map pre-treatment method was used to overcome this major impediment to an objective level. Furthermore, performance of text recognition by OOA-based algorithms showed impressive results on the input image with and overall accuracy of 70%. Both blue and black texts were averagely recognized. However, extraction challenges recurred and obstructed the extraction performance especially on the original image. Objectively, few relevant texts were extracted on the original image to demonstrate capabilities of the OOA-based method and ensure complete object recognition and extraction. Similarly, the map pre-treatment measure was implored to robustly extract text, yet robustness was impeded. The results obtained showed presence of enormous false positives for both black and blue texts which was not satisfactory. Hence, theoretical experiment was demonstrated to hypothetically establish OOA capabilities for text extraction if probed further. Here recognition and extraction was encouraging and was visually assessed with accuracy of 90% based on comparison with input image and very minimal errors. Additionally, the investigative effort demonstrated for contour line recognition and extraction by OOA-based method was not satisfactory due to inherent convolution and challenges which were hindering factors to contour line extraction. Here accuracy of 55% was assigned from visual assessment and in-depth scrutiny of isolines completeness, plausibility and fitness for future use. Although partial information derived is argued to be useful to extract certain information when probed (e.g. terrain profiling and relief information through cross sectioning of the extracted contour line).

- Results of sematic expansion demonstrated on the OOA-based symbol extraction symbols was useful in improving the quality of extracted symbols after the map was reconstructed into an e-map. Here two approaches used produced varying results and the out derived from the computed convex hull using the angular unit measurement produced a realistic out. Although, at some point over estimations were observed but were managed using logical GIS analysis. Hence, the concept of semantically expanding objects unlocked from historical map and reconstructed into e-maps by convex hull was rational to derive an improved meaning of extracted objects.
- The performance of the transferability exercise demonstrated on a separate section of the map substantiated the possibilities of transferable OOA algorithms. Additionally, the significance of the once created OOA rule-set was established and can drastically reduce time consuming challenges faced using the conventional method to extract information from historical maps. However, for an OOA algorithm developed for a specific historical map to be effective when transferred to a separate map with slight modifications, the maps should be saliently and semantically comparable. This is necessary because the result obtained from the transferability experiment demonstrated for the once created OOA rule-set on images with slightly different imaging conditions did not work efficiently as in the original image it was initially developed for. This also confirms the data dependency of OOA-based method should be considered if robust method is to be developed. Therefore, transferability can be challenging if imaging conditions of data exists between the data used to initially create rule-set and the data the rule-sets are transferred on due to the inherent variability in data thresholding. As was illustrated in the transferability section where the separate map of the same series and symbology had a different imaging condition such as the background colour and different features not present in the initial map used for creating the rule-set. This made the map incomparable and was a limiting factor to the transferability test and many other factors which can be better understood by supplementary research.
- Generally, accuracy assessment results showed the capabilities of OOA to robustly and reliably extract relevant information from paper maps rapidly. The accuracies of 95% recorded from the

OOA for hydrographic object, 97 and 92% for symbol correctness and completeness respectively, 70% in text and 55% was recorded in line extraction confirm the reliability suitability of the method. It is important to keep in mind that false positives recorded which resulted to misclassification were logically assigned cost/penalty of such misclassification and where necessary manually treated after exporting the extracted object as shapefiles.

- Outputs of the questions investigated in this research elucidated the capabilities of OOA to extract timely, relevant, reliable and robust information from paper maps. This advancement is valuable to geoinformation science, pattern recognition, information extraction realm of science and DRM programs.
- Ambiguity of visual map interpretation was a major challenge as to decide where to define object boundaries especially in the case of symbol boundary delineation. Hence the result obtained from the 10 experts to manually delineate object boundary on the test area image further confirm the ambiguity on deciding where precisely to delineate objects manually. However the OOA-based method was objectively unbiased and showed consistency in boundary delineation compared to the manual approach.
- Finally it is arguable that the conceptual schema devised, the product derived after implementing the OOA-based method, transferability of the created rule-set and the applicability of the method in applied earth sciences e.g. DRM analysis is a remarkable milestone this thesis achieved.

6.2 Research contributions

- This study identified suitable theoretical concepts for investigating, undemanding and developing a knowledge base for extracting relevant information from complex paper maps.
- The concepts of saliency and semantics were found to be useful to conceptualize, develop and implement an OOA based information extraction procedure from paper complex maps.
- Result of the research has added a body of knowledge to geoinformation science by bridging a gap between pattern recognition, information extraction, OOA and cartographic refinement in a multi-disciplinary fashion. Hence the method can be used to rapidly, robustly and reliably extract relevant information from historic topographic maps. Especially in developing countries where historic topographic maps are out dated and require update within short periods. Also that the maps are enormous and require concerted, rapid information extraction technique in a consistent manner which is lacking in conventional methods.
- Results of this study also proved the possibility of transferring a once created OOA algorithm to extract relevant information from complex paper map on same map series with same imaging conditions. Hence result of this experiment was important in understanding conceptual and technical issues to be considered for developing transferable OOA rule-sets.
- This research also found out how far OOA method can go in extracting information from complex historical paper maps and attempted to explain insights of the capabilities and in capabilities of the developed OOA method.
- Result of this thesis can be a useful piece of knowledge to build exercise for academic training and incorporations such as governments where there is urgent need to extract relevant information from historical maps.
- Extracted objects can be used for environmental management program e.g. DRM by incorporating this information with satellite data for further analysis.

6.3 Recommendations and futher research prospects

- Prospect for further research will include the utility of extracted information for further landuse change analysis (e.g. DRM programme change in river course, settlement expansion, vegetation change/landscape degradation analysis especially the vulnerable marshlands, mangrove vegetation and palm tree plantations etc).
- Effect of scanning resolution on information extraction enhancement.
- It important to note that though this study attempted a map pre-treatment approach to improve information extraction on paper map which performed remarkably in facilitating symbol extraction. This approach need to be investigated further since information extraction in satellite data also requires pre-processing before getting qualitative result.
- Further investigation on actual text recognition is needed to automatically derive text meaning and linking such information on atlas with name of places for location base analysis.
- Another research outlook is how to deal with symbol areas that overlap/intersect with one another.
- Similarly, further research in contour line extraction is needed on quality topographic maps to investigate extractability of contour line by OOA-based method and overcome current limiting factors in low quality maps.
- Improved accuracy assessment of OOA-based extracted objects from paper maps.

6.4 Research limitations

6.4.1 Data limitation

- The data used in this research was low quality and over time the map condition in terms of dirtiness had effect on information extraction.
- The effect of scanning such aliasing, convolution and bleeding hindered a better contour line extraction OOA investigation.
- Map complexity due to printing of different thematic information on the same colour plane.
- Thematic and spatial uncertainties present all over the map which resulted in objects boundary ambiguity.

6.5 Chapter summary

This chapter highlighted salient observation and conclusion derived from the result obtained in developing an OOA based information extraction procedure for paper maps. Highlight of the major contributions by this study are also presented in realm of geoinformation science, a number of recommendations and research outlooks have been pointed out. Limiting factors that significantly impeded on the success of the research objectives were outlined.

LIST OF REFERENCES

Akcay, H. G., & Aksoy, S. (2008). Automatic detection of geospatial objects using multiple hierarchical segmentations. *Ieee Transactions on geoscience and remote sensing, 46*(7), 2097-2111. doi: 10.1109/tgrs.2008.916644

Ali, E. A., & Algarni, D. A. (2004). Evaluating cartographic accuracy of some digitizing tables. *Journal of king Saud university, engineering sciences, Riyadh, 17* (1), 15-24.

Arrighi, P., & Soilees, P. (1999). *From scanned topographic maps to digital elevation models*. Paper presented at the international symposium on imaging applications in geology geovision'99, University of Liege, Belgium.

Baltsavias, E. P. (2004). Object extraction and revision by image analysis using existing geodata and knowledge: Current status and steps towards operational systems. *ISPRS Journal of photogrammetry and remote sensing, 58*(3-4), 129-151. doi: 10.1016/j.isprsjprs.2003.09.002

Bhattacharyya, S., Maulik, U., & Dutta, P. (2011). Multilevel image segmentation with adaptive image context based thresholding. *Applied soft computing, 11*(1), 946-962. doi: 10.1016/j.asoc.2010.01.015

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of photogrammetry and remote sensing, 65*(1), 2-16. doi: DOI: 10.1016/j.isprsjprs.2009.06.004

Bouwer, L. M. (2011). Have disaster losses increased due to anthropogenic climate change? *Bulletin of the American meteorological society, 92*(1), 39. doi: 10.1175/2010bams3092.1

Cao, R., & Tan, C. (Eds.). (2002). *Text/graphics separation in maps* (Vol. 2390). Berlin: Springer.

Casey, R. G., & Lecolinet, E. (1996). A survey of methods and strategies in character segmentation. *Ieee Transactions on pattern analysis and machine intelligence, 18*(7), 690-706. doi: 10.1109/34.506792

Chen, L. H., Liao, H. Y., Wang, J. Y., & Fan, K. C. (1999). Automatic data capture for geographic information systems. *Ieee, Transactions on systems man and cybernetics part c-applications and reviews, 29*(2), 205-215.

Cordella, L. P., & Vento, M. (2000). Symbol recognition in documents: A collection of techniques? *International journal on document analysis and recognition*, 73-88.

Cordella, L. P., & Vento, M. (2001). Symbol and shape recognition. In A. K. Chhabra & D. Dori (Eds.), *Graphics recognition, recent advances* (Vol. 1941, pp. 167-182).

Cutter, S. L., Mitchell, J. T., & Scott, M. S. (2000). Revealing the vulnerability of people and places: A case study of georgetown county, south carolina. *Annals of the association of American geographers, 90*(4), 713-737.

Definiens. (2010). Developer 8 *Reference book*. Germany: Timble.

Delalandre, M., Trupin, E., & Ogier, J.-M. (2004). Local structural analysis: A primer. In J. Lladós & Y.-B. Kwon (Eds.), *Graphics recognition. Recent advances and perspectives* (Vol. 3088, pp. 223-234): Springer Berlin / Heidelberg.

den Hartog, J., ten Kate, T., & Gerbrands, J. (1996). Knowledge-based segmentation for automatic map interpretation. In R. Kasturi & K. Tombre (Eds.), *Graphics recognition methods and applications* (Vol. 1072, pp. 159-178): Springer Berlin / Heidelberg.

Deseilligny, M. P., Mariani, R., Labiche, J., & Mullot, R. (1998). Topographic maps automatic interpretation: Some proposed strategies. In K. Tombre & A. K. Chhabra (Eds.), *Graphics recognition: Algorithms and systems* (Vol. 1389, pp. 175-193).

Dhar, D. B., & Chanda, B. (2006). Extraction and recognition of geographical features from paper maps. *International journal on document analysis and recognition, 8*(4), 232-245. doi: 10.1007/s10032-005-0010-9

Dillabaugh, C. R., Niemann, K. O., & Richardson, D. E. (2002). Semi-automated extraction of rivers from digital imagery. *Geoinformatica, 6*(3), 263-284. doi: 10.1023/a:1019718019825

Dragut, L., Tiede, D., & Levick, S. R. (2010). Esp: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International journal of geographical information science, 24*(6), 859-871. doi: 10.1080/13658810903174803

Dupont, F., Deseilligny, M. P., & Gondran, M. (1998). Automatic interpretation of scanned maps: Reconstruction of contour lines. In K. Tombre & A. K. Chhabra (Eds.), *Graphics recognition: Algorithms and systems* (Vol. 1389, pp. 194-206).

Esch, T., Thiel, M., Bock, M., Roth, A., & Dech, S. (2008). Improvement of image segmentation accuracy based on multiscale optimization procedure. *Ieee Geoscience and remote sensing letters, 5*(3), 463-467.

Frischknecht, S., & Kanani, E. (1998). Automatic interpretation of scanned topographic maps: A raster-based approach. In K. Tombre & A. Chhabra (Eds.), *Graphics recognition algorithms and systems* (Vol. 1389, pp. 207-220): Springer Berlin / Heidelberg.

Ghircoias, T., & Brad, R. (2010). A new framework for the extraction of contour lines in scanned topographic maps. In M. Essaaidi, M. Malgeri & C. Badica (Eds.), *Intelligent distributed computing iv* (Vol. 315, pp. 47-52).

Goodson, K., & Lewis, P. (1988). Feature extraction from line drawing images. In J. Kittler (Ed.), *Pattern recognition* (Vol. 301, pp. 216-221): Springer Berlin / Heidelberg.

Graeff, C. (2006). *Ethical implications of biases and errors in geographic information systems.* Paper presented at the Ieee International symposium on technology and society, ISTAS 2006 (8-10 June).

Guerin, P., & Deseilligny, M. P. (1995, 4-6 Jul 1995). *Global analysis of scanned maps for an automatic interpretation.* Paper presented at the Fifth international conference on image processing and its applications, 1995.

Hurtut, T., Gousseau, Y., & Schmitt, F. (2008). Adaptive image retrieval based on the spatial organization of colors. *Computer vision and image understanding, 112*(2), 101-113. doi: 10.1016/j.cviu.2007.12.006

Kaftan, J. N., Bell, A. A., & Aach, T. (2008). *Mean shift segmentation - evaluation of optimization techniques.*

Kasturi, R., & Alemany, J. (1988). Information extraction from images of paper-based maps. *Ieee Transactions on software engineering, 14*(5), 671-675. doi: 10.1109/32.6145

Kerle, N., & de Leeuw, J. (2009). Reviving legacy population maps with object-oriented image processing techniques. *Ieee Transactions on geoscience and remote sensing, 47*(7), 2392-2402. doi: 10.1109/tgrs.2008.2010853

Khotanzad, A., & Zink, E. (2003). Contour line and geographic feature extraction from usgs color topographical paper maps. *Ieee. Transactions on pattern analysis and machine intelligence, 25*(1), 18-31. doi: 10.1109/tpami.2003.1159943

Koike, H., Sato, Y., Kobayashi, Y., Tobita, H., & Kobayashi, M. (2000). *Interactive textbook and interactive venn diagram: Natural and intuitive interfaces on augmented desk system.* Paper presented at the SIGCHI conference on human factors in computing systems, The Hague, Netherlands, April 2000.

Lee, D. S., Shan, J., & Bethel, J. S. (2003). Class-guided building extraction from ikonos imagery. *Photogrammetric engineering and remote sensing, 69*(2), 143-150.

Leyk, S., Boesch, R., & Weibel, R. (2006). Saliency and semantic processing: Extracting forest cover from historical topographic maps. *Pattern recognition, 39*(5), 953-968. doi: 10.1016/j.patcog.2005.10.018

Lichtner, W. P. o. (1985). *Pattern recognition procedures for automatic digitizing of cadastral maps.* Paper presented at the AutoCarto 7, digital representations of spatial knowledge, Washington D.C.

Llados, J., Valveny, E., Sanchez, G., & Marti, E. (2002). Symbol recognition: Current advances and perspectives. In D. Blostein & Y. B. Kwon (Eds.), *Graphics recognition: Algorithms and applications* (Vol. 2390, pp. 104-127). Kingston, Canada.

Lu, P., Stumpf, A., Kerle, N., & Casagli, N. (2011). Object-oriented change detection for landslide rapid mapping. *Ieee, Geoscience and remote sensing letters, PP*(99), 701-705.

Luyang, L., Nagy, G., Samal, A., Seth, S., & Yihong, X. (1999, 20-22 Sep 1999). *Cooperative text and line-art extraction from a topographic map.* Paper presented at the ICDAR Proceedings of the fifth international conference on document analysis and recognition

Madry, S. (2006). *The integration of historical cartographic data within the gis environment.*

Marceau, D. J. (1999). The scale issue in social and natural sciences. *The Canadian journal of remote sensing, 25*(4), 347-356.

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, PP*(99), 1-16.

Miliaresis, G., & Kokkas, N. (2007). Segmentation and object-based classification for the extraction of the building class from lidar dems. *Computers & geosciences, 33*(8), 1076-1087. doi: 10.1016/j.cageo.2006.11.012

Myers, G., Mulgaonkar, P., Chen, C., DeCurtins, J., & Chen, E. (1996). Verification-based approach for automated text and feature extraction from raster-scanned maps. In R. Kasturi & K. Tombre (Eds.), *Graphics recognition methods and applications* (Vol. 1072, pp. 190-203): Springer Berlin / Heidelberg.

Nowak, A., Florek, A., & Piascik, T. (1992). Edge tracing in a priori known direction
computer vision — eccv'92. In G. Sandini (Ed.), (Vol. 588, pp. 38-42): Springer Berlin / Heidelberg.

Ogier, J. M., Bessaid, A., & Bechar, H. (2006). Automatic topographic color map analysis system. *Association for the advancement of modelling & simulation echniques in entreprises*, 229–244.

Pavlidis, T. (1977). *Structural pattern recognition*. NewYork: Springer.

Reilly, D., Rodgers, M., Argue, R., Nunes, M., & Inkpen, K. (2006). Marked-up maps: Combining paper maps and electronic information resources. *Personal and ubiquitous computing, 10*(4), 215-226. doi: 10.1007/s00779-005-0043-6

Rekik, A., Zribi, M., Hamida, A. B., & Benjelloun, M. (2007). Review of satellite image segmentation for an optimal fusion system based on the edge and region approaches. *IJCSNS International journal of computer science and network security, 7*(10), 242-250.

Roy, P. P., Llados, J., & Pal, U. (2007). *Text/graphics separation in color maps.* Paper presented at the International conference on computing: Theory and applications (ICCTA'07).

Samet, H., & Soffer, A. (1998). Magellan: Map acquisition of geographic labels by legend analysis. *International journal on document analysis and recognition, 1*(2), 89-101. doi: 10.1007/s100320050009

Samet, R., Askerbeyli, I. N. A., & Varol, C. (2010). An implementation of automatic contour line extraction from scanned digital topographic maps. *Applied and computational mathematics, 9*(1), 116-127.

Samet, R., & Namazov, M. (2008). Using fuzzy sets for filtering topographic map images. *Applied and computational mathematics, 7*(2), 242-254.

San, L. M., Yatim, S. M., Sheriff, N. A. M., & Ismail, N. (2004). Extracting contour lines from scanned topographic maps *Proceedings of international conference on computer graphics, imaging and visualization* (pp. 187-192).

Shane, F. P., & Page, K. J. (2000). Water body detection and delineation with landsat tm data. *Photogrammetric engineering & remote sensing letters, 66*(12), 1461 - 1467.

Tumuhairwe, S. (2011). *Evaluation of the transferability of a generic algorithm for object oriented landslide mapping and pattern analysis for the 2010 haiti earthquake.* Master of science, University of Twente, Enschede. Retrieved from http://www.itc.nl/Pub/Home/library/Academic_output/AcademicOutput.html ITC academic output database.

Van de Ville, D., Nachtegael, M., Van der Weken, D., Kerre, E. E., Philips, W., & Lemahieu, I. (2003). Noise reduction by fuzzy image filtering. *Ieee Transactions on fuzzy systems, 11*(4), 429-436. doi: 10.1109/tfuzz.2003.814830

Velázquez, A., & Levachkine, S. (2004). Text/graphics separation and recognition in raster-scanned color cartographic maps. In J. Lladós & Y.-B. Kwon (Eds.), *Graphics recognition recent advances and perspectives* (Vol. 3088, pp. 63-74): Springer Berlin / Heidelberg.

Watanabe, T. (2000). *Recognition in maps and geographic documents: Features and approach.* Paper presented at the Selected papers from the third international workshop on graphics recognition, recent advances.

Watson, L. T., Arvind, K., Ehrich, R. W., & Haralick, R. M. (1984). Extraction of lines and regions from grey tone line drawing images. *Pattern recognition, 17*(5), 493-507. doi: 10.1016/0031-3203(84)90047-5

Worboys, M. (1995). *Gis, a computing perspective*: Taylor & Francis.

Yamada, H., Yamamoto, K., & Hosokawa, K. (1993). Directional mathematical morphology and reformalized hough transformation for the analysis of topographic maps. *Ieee Transactions on pattern analysis and machine intelligence, 15*(4), 380-387. doi: 10.1109/34.206957

Zhong, D. X. (2002). Extraction of embedded and/or line-touching character-like objects. *Pattern recognition, 35*(11), 2453-2466. doi: 10.1016/s0031-3203(01)00225-4

Zhu, Y., & Joao, N. C. (2003). Recognition of lakes from remotely sensed imagery. *University of Nebraska journal on computational aspects of GIS, 896.*