Cover Page

Declaration

Abstract

Acknowledgement

Table of Contents

1. Chapter 1: Introduction

1.1 Chapter Overview

This chapter covers the overview of the project, it's background, the justification of the problem being addressed and the research question relevant to this research. The motivation behind attempting this research is also explained. Furthermore, the Aim and the scope that will be attempted in this research will also be defined. The Research objectives that will be fulfilled and also personal objectives of conducting this research will be mentioned along with an overview of the solution. The requirements for such proposed solution will also be defined.

1.2 Project Overview

1.2.1 Project Background

Images

Human beings can be considered predominantly visual creatures as we use our sense of vision when available to understand our surroundings. Due to this reason we have created images to capture moments of this ever constantly moving world. Images have since then evolved to not only capturing real world moments but to digital drawings and other representations of visual still media. An image can be a single picture which represents something. It may be a picture of a person, of people or animals, or of an outdoor scene, or a microphotograph of an electronic component, or the result of medical imaging[DL1]. When image files are attempted to be classified, two main classifications can be identified. Vector images and Raster images. There are different pros and cons of using either a Vector image or a Raster image.

Raster Images

A raster image is built up of color pixels which are arranged to form the necessary result image. Raster images are mostly suited for linear art images because they can better represent subtle chromatic gradients due to the fact that each pixel can change its value independently of other pixels to form the necessary image. These types of images are also called continuous tone images.

A raster image while being faster to process and display as there is no methodical processing or such involved will have a larger file size as information is electronically stored on a pixel-by-pixel basis (Winnemoeller *et al.*, 2018), the size of the image is directly proportional to the resolution of the image

Vector Images

Vector images on the other hand are created from points and lines and curves joining them. These are based off of mathematical formulas that create combinations of multiple true geometric primitives to create a final image (Seel-audom, Naiyapo and Chouvatut, 2017)..

Vector format based computer graphics tools have become very powerful tools allowing artists, designers etc. to mimic many artistic styles, exploit automated techniques, etc. and across different simulated physical media and digital media (SEVERENUK *et al.*, 2019). Similarly, in real world applications according to the factors that need to be considered, the image may be required in either Raster or in Vector format.

As stated above, as a vector file and raster file of the same image may have similar resulting image, when observed on a deeper level multiple differences can be identified between them.

1.2.2 Problem Domain

Geographic Information System (GIS) is the processes of managing, manipulating, analyzing, updating and presenting metadata according to its geographic location, to be effectively used in different aspects of life (Al-Bayari, 2018). The proliferation of GIS technology has greatly increased the access to and the usage of spatial data. Making maps is relatively easy even for those who do not have much cartographic training (Wong and Wu, 1996).

According to the analysis needed to be performed on a certain image obtained, the requirement for a Raster image or a Vector image may vary. There are several pros and cons when you consider each type of image. There is an old GIS adage stating that “Raster is Faster but Vector is Corrector”.(Berry, 1995).

Vector GIS results in the geometrization of the geographical world, and generalizing and reducing its theory into theories about relations between points, lines, polygons and areas. Such objects which are in a GIS can be counted, moved about, stacked, rotated labelled, cut etc. and be handled like a variety of other everyday solid objects that bear no particular relationship to geography (Couclelis, 1992).

Vector maps use simple geometric components such as points, lines and polygons in adjacencies, unions and inclusions to describe spatial information and Raster maps are based on pixel matrices and are richer and realistic than vector maps (Lin and Guo, 2011).

There are several advantages of using a Vector data format. These can be stated as the output being more aesthetically pleasing and zoom able to very close detail as it is made up of points and line segments connecting them and not using fixed number of pixels which might look pixelated and less clear when zoomed into more than its resolution allows. It also provides higher geographical accuracy due the same reason as it being not of a fixed resolution. There are other reasons as why vector images are used in GIS such as data integrity, and allowing network analysis and proximity operations as they both use vector data structures. As well as there are advantages are there are disadvantages to using vector images as well. As these images are a result of mathematical calculations it is often very processing intensive. Vector data structures are also poor performing when displaying continuous data, and needs to be generalized in some manner to display, which can result in loss of some information.

While vector data structures in GIS over determines the geographic world by forcing it into a geometric objects generalizing them, the raster data structure feigns maximal ignorance on the nature of things in the world. Yet Raster data structures provide an implicit view of the geographical world with measurable values discretized into pixel arrays (Couclelis, 1992).

Raster data can store unique values per each pixel without any generalization being required. Therefore, is a good option when continuous data is required to be displayed. Even though continuous data can be very accurately represented in a Raster image, because of the resolution. Raster graphics display devices are capable of reproducing very complex images (Sloan and Tanimoto, 1979). It struggles when representing linear features and can cause pixilation if the resolution of the image obtained is low or when zoomed to obtain a closer look. Raster datasets can also be very large file as when the resolution is increased to get a more accurate image with high detail the file size increases proportionally with it.

As we can see from the statements above, we can identify that both these formats are equally important when considering the use of imagery in GI Systems. Therefore there becomes necessary a method of conversion between these two data types.

1.2.3 Problem Justification

Automated conversion of engineering drawings and such similar content into Raster and vector data has been a very widely discussed topic. A critical step in this process can be considered as the conversion of these images into a vector format (Liu and Josep Lladós (Eds.), 2005). Many techniques for conversion of raster to vector have been proposed which has even led to development of commercial solutions to tackle this issue. The systems created all did provide quite acceptable results but each had their own drawbacks (Hilaire and Tombre, 2006). (Lacroix, 2009)

1.2.4 Research Question

How is the problem of a common platform that identifies the properties of a Raster image and converts it into a Vector file by using the best method of conversion using parameters which best fit the use case of the resulting vector image addressed currently in the research domain of GIS graphics processing?

1.2.5 Motivation

After researching on the basic concept of converting Raster images into Vector images for graphic design purposes, I have come to find the importance of it but in a different domain which is in the field of GIS. Raster and Vector data structures are widely used in analysis in GIS and as both of these type of images are needed according to different situations. It has motivated me to create this automated Raster to Vector conversion tool.

1.3 Project Aim and Scope

1.3.1 Project Aim

To investigate design and implement a Raster to Vector conversion platform that selects the best method of conversion using image processing techniques.

1.3.2 Project Scope

In-Scope

- Raster to Vector conversion tool is only developed geared towards GIS
- Training an image processing model to identify certain properties of images that affect conversion algorithms.
- Integrating Image processing for the identification special characteristics to identify the best conversion algorithm
- Considering of continuous tone images as well as line based images for the conversion process
- Set conversion method from either one of Accurate or Fast conversion

Out-Scope

- Conversion of Raster to Vector for other domains such as graphic design.
- OCR functionality out of image text is not considered, and will be represented in the converted images as graphical data and not textual data.

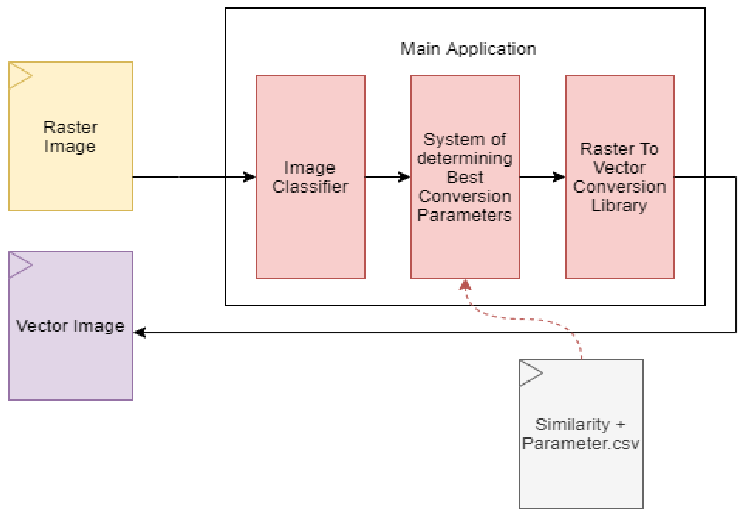
1.4 Objectives

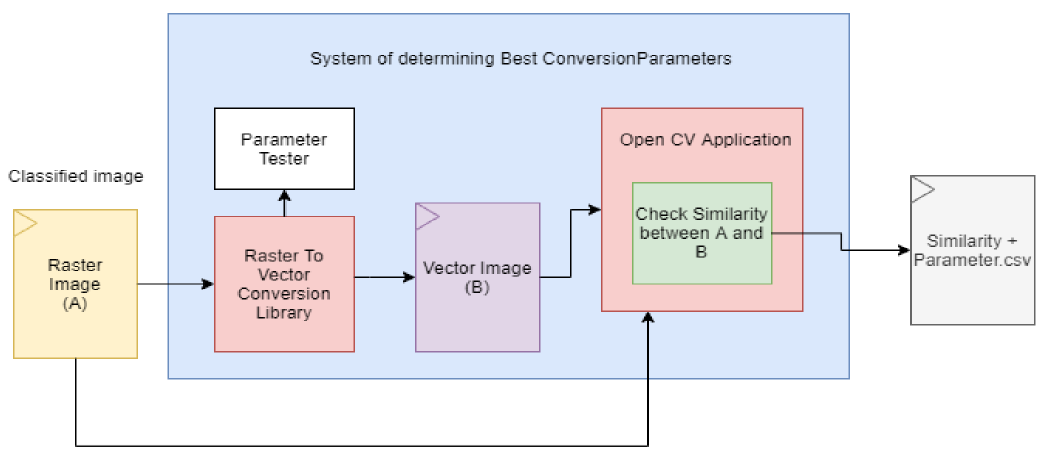
I have identified the following as the objectives to achieve to complete my Research successfully.

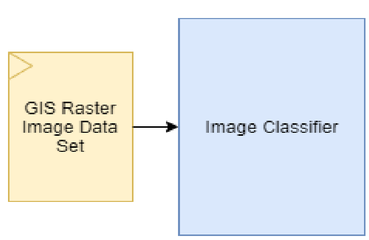
1.4.1 Research Objectives

- To conduct a thorough literature review on existing solutions and platforms
- To design an image processing model to identify properties of a Raster image.
- To implement functionality to determine the most favorable algorithm for the conversion through the parameters identified from the image processing process.
- To evaluate the converted Vector image in terms of accuracy and speed of conversion.

1.5 Overview of the Solution







As shown in the figure above the main system is made up of three components. The image classifier, system of determining best conversion parameters and the raster to vector conversion library.

When taking a closer look at each system; the system of determining best conversion parameters is a machine learning based system that finds the best conversion parameters by interpolating data from a set of previously identified conversion parameters for each image classification and the match rate of the image to its particular classification determined by the Image classifier.

The initial best fit parameters are identified by converting a single image into multiple vector images each time with different parameters within a certain range of each parameter and comparing its visual similarity with the original using an image processing library. This similarity is then recorded in a csv file for the system of determining the best conversion parameter to use.

A data set of several hundred labelled GIS image types are used to train the image classifier.

The raster to vector conversion library is used to convert the images from its raster to vector format in each step that its necessary to do so.

These three main components make up the main system and together convert any given GIS based image which can be classified into an accurate vector image.

1.6 Resource Requirements

*Note these requirements can be subjected to change

1.6.1.1 Software Requirements

- **Windows 10 (64-bit version):** To accommodate and run the software
- **Java, Python:** For the conversion algorithms to function
- **NodeJS:** For backend related scripting
- **MSWord:** For documentation Requirements
- **GraalVM:** For cross platform application support

1.6.1.2 Hardware Requirements

- Core i7 Processor
  - High processing power required for algorithms to be executed
- Minimum 4GB RAM
  - For application to run smoothly and not run out of system memory
- Disk Space: Up to 10GB
  - For storing of application and images and temporary files created while converting and running algorithms

1.7 Chapter Summary

Have to write

2. Literature Review

2.1 Overview

This chapter will provide a critical analysis on the domain of the problem based on literature. This chapter will analyze the current methods of raster to vector conversion why they are needed, the different variations of raster to vector conversion system. This chapter will also review similar work for the process of conversion of Raster to Vector.

2.2 Problem Domain

2.2.1 Geographic Information Systems

Geographic Information Systems (GIS) are computer based tool for process of manipulating, updating, managing, analyzing and presenting data of geographic locations and cartographic data which can be thereafter used in different fields of study or be applied in daily applications which require such data. (Al-Bayari, 2018) (Chrisman, 1999).

Due to proliferation of GI science and the usage of spatial data, map making has been simplified even for someone with little to no experience in cartographic training (Wong and Wu, 1996).

During the earlier stages of GI Science, the definition of the it simply putting maps into computers. The process of implementing this idea however was a tedious task and to complete satisfactorily was a major undertaking. Recently however the limitation of spatial data in 2 dimensional vertical view of the surface of the globe is not as sought after and scientist demand for 3 dimensional views and the ability to simulate various geographic processes though GIS data (Gold, 2006).

This GIS information can be stored and processed in image formats after being generated. There are two main types of image structures that can be used when storing this information. This can be classified as Raster data and vector data (Wade *et al.*, 2003).

2.2.2 Raster Image Data

Raster images, which are also known as bitmap images are categorized under digital images as being formed of tiny rectangular pixels which are arranged in a grid formation that combines together to represent an image. This format of image can support a wide range of colours and depict subtle gradients, it allows for a very accurate visual representation of continuous tone images such as shaded drawings, photographs and other highly detailed and complex imagery.

Raster graphics initially originated in television technology with images constructed much like the pictures on a television screen. Raster graphics are made of small uniform sized pixels which are arranged in a two dimensional grid which is made up of vertical and horizontal columns. A single pixel contains information of a single or multiple bits depending on the degree of information in the image. For example, black and white images contain only one bit per pixel in the image this can either be a true black bit or a true white bit, an image with shading or colour commonly contains 24 bits of information per pixel, this allow more than 16 million possible states of colour value for the pixel. Images with 24 bits of information per pixel are known as 'truecolor' images with 24-bit colour and can realistically depict colour images. The amount of detail stored into a single bit is represented by the colour depth and the number of pixels that form together to form the image can be represented as its resolution and affects how much detail is depicted in an image (Britannica, 2014).

2.2.3 Vector Image Data

Vector graphics are digital images that are constructed through a sequence of commands or mathematical formulas that place lines points to create polygons in a two dimensional or three dimensional space. In the field of physics, a vector is the representation of a quantity that contains both a direction and a quantity or value. For example, instead of storing the data as an array of pixels as when a raster images or bitmaps, a vector file creates its complex images by mathematically aligning and stacking multiple polygons created by connecting a series of point with lines segments. This results in the creation of typically smaller files with extremely high fidelity, lossless compress and scalability of any kind without the distortion or loss of fidelity of it (He, Xu and Zheng, 2009).

2.2.4 Need for both Raster and Vector based data in GI systems

According to the type of processing and analysis to be performed on a certain data set obtained the requirement for the format of the data type, either Raster or Vector may be different. Each data type has its advantages and disadvantages, strengths and weaknesses and to further solidify the need for both data types an old GUS adage states that “Raster is Faster but Vector is Corrector” (Berry, 1995).

A well-known logical consequence of the difference of the data structures vector and raster is that as while vector data can record position to and degree of accuracy, raster data have a built in level of positional accuracy. Therefore we can also classify raster positional data as integer and vector positional data as real(Holroyd and Bell, 1992).

GIS data in vector format is produced by geometricizing the real geographical world, generalizing it and reducing it into theories about points, lines polygons and areas. Objects thus created using these theories can be counted translated, stacked, rotated, labeled, cut and etc. and can be handled like every day geometrical shapes that are not directly related to geographical data in any manner (Couclelis, 1992).

While vector maps use simple geometric structures as points, lines, polygons and relationships between them such as adjacencies, unions and inclusions to depict the geographical data, Raster based maps are stacked matrices of pixel based data which can be layered in a sandwich like structure and are more rich and realistic than vector data (Lin and Guo, 2011).

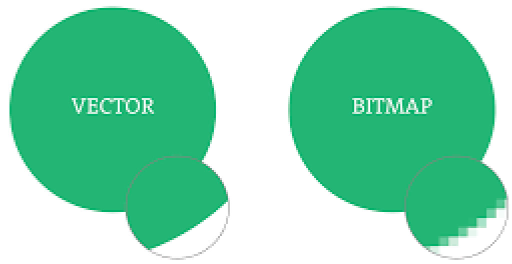


Figure 2.2-1 Visual comparison between vector and bitmap image

Advantages of using a vector data format can be stated as the output produced being more aesthetically pleasing, highly zoom able as it is made up points and line segments contented to each other to form multiple polygons and not a pixel array which when zoomed into more than its set resolution allowed may look pixelated. There are a number of other advantages of using a vector data structure rather than a raster data structure, and these can be listed down as follows.

1. Data integrity
2. Allowing network analysis
3. Proximity operations as they both use vector data structures

While there is strength in using vector data structures to represent and process GIS data, there are weakness of it as well. As vector data is produced by the result of geometrical mathematical calculations, this data produced can be very processor intensive when for example depicting continuous data regarding a point if stored in vector format is much more complex than a single pixel storing information which makes up a raster image. Therefore, due to the feasibility of the idea of storing data as a series of points made up of geometrical data being not practical these images must be generalized in some manner to ease storing and processing when needed to be displayed. This result in the loss of some portions of information.

Even though GIS data that is stored as vector data structures over determines the geographic world by generalizing them and forcing them into geometric objects, Raster data structures provides an implicit view of the world with measurable values discretized into pixel arrays but simulates ignorance on the nature of an actual object in a physical space (Couclelis, 1992).

A simple advantage of using raster data over vector is the ability to store continuous tone images as pixel array which stores unique values of data per pixel without performing any generalization which may lead to data loss in some manner. But while continuous data can be stored very accurately using raster images, as the resolution of a raster graphic is set and cannot be varied, it may limit the fidelity of final output if the process requires the image to be enlarged. Raster graphics display devices are capable of reproducing very complex images (Sloan and Tanimoto, 1979). Raster images also have the weakness of not being able to produce true linear features as pixels used to form as arranged horizontally and vertically, are limited when reproducing line data at any other angle. Raster data sets also have very big file sizes as the higher the resolution the more pixel data it contains. Therefore, the file size is directly proportional to the resolution of the image, and by resulting the file size of the image to be directly proportional to the visual fidelity of the image.

2.2.5 Raster to Vector Conversion in GIS

Spatial data obtained as aerial imageries, for example is very continuous in nature and must be represented as raster data. These types of maps are unable to be displayed by vector maps (Lin and Guo, 2011).

With development of high speed computing and its usage in a wide range of fields of study leading to improve process times and increasing the ability of highly computational tasks to be performed on smaller and more affordable system. Vector data can be used as a much more accurate data structure as the images even with complex processing requirements can be easily displayed and rendered faster than before (Winnemoeller *et al.,* 2018). Vector images are also known for its flexibility and compact nature etc.(Lin *et al.*, 2015). By going through the above use cases for each raster and vector images, we can come to the conclusion that both of these data structure is widely used in the Geographical Information Sciences. Therefore, there arises a need for conversion between these data structures. Even though many commercial solutions and researches have been conducted in achieving a satisfactory result, they each have their drawbacks and hence arises a need for a common solution to this problem.

2.2.6 Image Processing

Digital Image Processing is the process of manipulation of images using computers. Image processing use cases have been increased exponentially during recent times. Its application ranges from entertainment, passing by geological processing and remote sensing. Multimedia systems which play a huge role in modern society also heavily rely on image processing.

When studying the various disciplines of image processing we can identify that it is a vast topic with many techniques that are applied specific to the type and qualities of an image. An image can be regarded as a function $f(x, y)$ of two continuous variables x and y. *For images to be processed digitally the image has to be sampled and transformed into a matrix of number. And quantization is then required as a computer represents number using finite precision. Image processing then can be identified as the manipulation of those finite precision numbers.*

Image processing can be mainly categorized into (Eduardo A.B. da Silva, 2005).

- **image enhancement**
- **image restoration**
- **image analysis**
- **image compression**

2.3 Comparison of Similar Solutions

Name of Software	Maximum Zoom	Control point limit	Color management for print	Adjust Image	OCR	Scripting Support
Easy Trace	12800x	-	Yes	Yes	No	No
Free Hand	64x	-	No	No	No	Yes
Illustrator	640x	-	Yes	No	No	Yes
Potrace	N/A	-	No	No	No	No
Scan2Cad	64K	No Limit	Yes	No	Yes	Yes
WinTopoPro	512x	No Limit	Yes	Yes	No	Yes

Table 2.3-1 Comparison of Similar Solutions

2.3.1 Easy Trace

The easy Trace pro software package is a convenient and powerful tool for vector map generation with editing capabilities. The program also is widely used in GI systems and supports most of the import and export formats such as.

- ArcINFO
- ArcView
- AutoCAD
- Credo
- MapInfo
- MicroStation

This software application is only functional on a windows operating system. It also contains raster to vector conversion tools and is create by the Easy Trace group (Russia). Old versions are free versions but all newer updated version are paid versions of this software.

2.3.2 FreeHand

Adobe Freehand (a.k.a Macromedia Freehand) is a commercial computer application for creating vector graphic content. Created by the Altsys Coporation in 1988 and licensed to Aldus corporation this software released version 1 through4 in 1994, and was discontinued since 2003. This application also allowed a user to convert a placed raster image into a vector.

2.3.3 Illustrator

Adobe illustrator is an industry grade application for multimedia, online and print graphics creation. It provides tools to create technical illustrations or graphic related content for print publishing and also allows web related graphic content creation. It is a commercial tool. This software also allows a functionality known as image trace which allows the conversion of images from raster to vector.

2.3.4 Potrace

Portace(TM) is a bitmap tracing tool which allow the transforming of bitmaps into raster based scalable images. The input type of the bitmap can be of any of the following formats

- PBM
- PGM
- PPM
- BMP

And the output will be into a vector format file. A usecase of this software can be create SVG or PDF files from scanned raster images. The resulting vector image then can be scaled to any resolution without any distortion in the image quality.

Protrace uses the output formats

- SVG
- PDF
- EPS
- PostScript
- DXF
- GeoJson
- PGM

It is a software package that is currently in development and further image type support will be added in future times. Portrace does not preprocess the image before it is converted to a vector file format and this has to be performed beforehand by a different application.

2.4 Justification of the Selected Approach

As the literature review has been conducted, the various similar solutions available on a commercial level and on a research level are identified not address identifying the qualities of the raster image before conducting the process of conversion of the image from raster to vector.

Hence this project will be further looking into the usage of image processing to identify various image processing techniques for the preprocessing of the image before converting it from raster to vector and also identifying the most suitable parameters for the algorithm to be executed on for an accurate conversion.

2.5 Algorithmic Analysis

2.5.1 General Raster to Vector Conversion Algorithms

The task of converting a raster based data structure in to line data or into vector based data structures can be divided into three basic operations. These are as follows

- Skeletonization/Line Thinning
  - This is process of reducing or thinning the thickness of the lines found in the image to a unit thickness relative to the given resolution
- Line Extraction/Vectorization
  - The process of identifying a series of data entities or coordinates that are used o form a single line segment in the input document
- Topology Reconstruction
  - The process of determining the adjacency relationship among the lines identified during line extraction.

The individual line segments are joined into whole lines if desired and can also be combined as polygons for continuous representation. Two other additional post processing tasks that can applied onto the basic raster to vector conversion process can be stated as follows.

- Line smoothening
- Gap Removal

2.5.2 Potrace Raster to Vector Algorithm

The potrace algorithm converts bitmaps into a vector files by following though the following steps. Initially the bit map is segmented into a paths which form several boundaries between dark and light (black and white) areas.

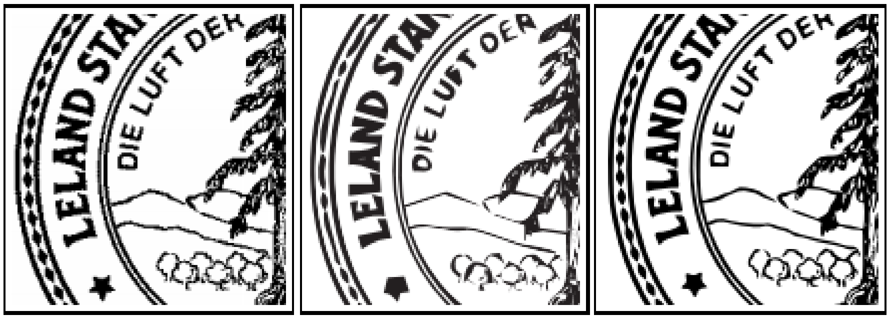


Figure 2.5-1A detail from the seal of Stanford University; the original scanned image,

Next, each of those identified paths are approximated by an optimal polygon, and each polygon is transformed into a smooth outline. An optional step after the previously mentioned steps are completed is the resulting curve being optimized by joining consecutive Bezier curve segments together where it is possible to do so. After all these steps are completed the output is generated in the required format (Selinger, 2003).

2.6 Chapter Summary

As outlined in this chapter, an image processing solution to identifying the best conversion method and parameter identification when raster to vector conversion is considered is discussed.

This is to combat the fact of there not being a standard when identifying the best method of conversion from raster to vector. A number of raster to vector conversion tools have been reviewed and researched on for this purpose. And a conclusion was reached upon after carefully analyzing their functionality and algorithmic features.

3. Chapter 3: Methodology

3.1 Chapter Overview

Write something

3.2 Research Methodology

This research is based on the conversion of the process of Raster to Vector in classified GIS imagery with the use of best-fit conversion parameters. The theory of the research is based on a hypothesis that has been established as follows. *"Images that have similar characteristics and can be classified as such using an image classifier can have a similar set of best-fit conversion parameters for Raster to Vector conversion to obtain optimal results of conversion."*

3.2.1 Research Paradigm

When considering the different approaches that can be considered when approaching this research topic. We can identify three clear approaches that have been established by researchers. These are the approaches of 1) Positivism 2) Interpretivism 3) Critical theory. This is essential because as consumers of research, we have to be able to look deeper into claims made by researchers who adhere to different research paradigms (Reference here). [DL2]

we are establishing a hypothesis and are testing to obtain empirical data to form a relation between the phenomena we can consider the research paradigm of this research to be of a **positivism approach**.

3.2.2 Research Methodology

When considering research methodologies, there are two main types of methods that researches generally can be categorized under. 1) Deductive approach 2) Inductive approach. The deductive approach starts with specifying the objective. The inductive approach emphasizes on observation and deriving conclusions through observation[DL3]. (Reference here).

And as we are establishing this hypothesis and directing our research to test it. We can consider this research to be of the **deductive approach**.

3.3 Design Methodology

Two main design methodologies can be considered to be used to design the proposed system. These can be identified as Object-Oriented Analysis and Design (OOAD) and Structured Systems Analysis and Design Method (SSADM). Each of these methods have their pros and cons according to what kind of system is being designed. In this section we will identify factors when considering each design methodology to pick the most suitable approach for designing the proposed system.

In the Object-Oriented analysis and design approach initially, requirements are identified and a software specification and documentation is developed in terms of an object model. These are objects that integrate both data and function ns and are modeled after real-world objects. After these objects are identified it is then mapped into classes and constraints and relationships are identified. This methodology gives access to certain OOD principles such as classes and objects, encapsulation, polymorphism and interfaces/abstract classes. OOD can be considered a good design methodology to manage complexity in applications and to enable the reusability of components.

In the structured system analysis and design methodology graphical tools are used in a systematic approach to refine objectives out of well-defined user requirements. In the case of these requirements not being clearly described initially, it can lead to problems in the process of the solution created. This approach also does not accommodate dynamic user requirements that might be subjected to change through the development life cycle of the solution.

When considering the above factors as the requirements of the proposed solution may be subject to change through the life cycle of this research. The **OOAD** methodology is better suited for the analysis and design methodology of this project.

3.4 Development Methodology

When starting a project that has as purpose the software development, it is very important to use a methodology that increases its success rate. The success rate of software development projects can be increased by using a methodology that is adequate for the specific characteristics of those projects.[DL4] Multiple software development methodologies that can be employed when determining the development life cycle of a software application. These can be identified as follows. Waterfall, Agile, Feature-Driven, Iterative, Spiral, Prototyping and RAD (Rapid

Application Development). The following features that might be considered as factors that affect which mentioned methodology is most suitable for this project can be identified due to their repeated occurrence during this project.

- The initial requirements are identified and recorded in the Project initiation document.
- These requirements are subjected to change over the development of the project due to feedback and development constraints
- Development is done by breaking down the project into components
- Development will be done component by component.
- Each component will have the ability to be tested independently.

The Agile development methodology can be considered an iterative and incremental process that focuses on rapid delivery of working prototypes that finally create the working product in incremental builds. When considering the above stated factors, it is evident that an **Agile** Software development methodology is better suited for the development methodology of this project.

3.5 Chapter Summary

4. Chapter 4: Project Management

4.1 Chapter Overview

Write something here

4.2 Project Management Methodology

For the Project management methodology of this project, the PRINCE 2 Project Management methodology has been selected. PRINCE 2 is a project management framework that is scalable and can be adaptable to any type of project with ease. PRINCE 2 focuses on dividing the work process of a project into manageable and controllable stages. PRINCE 2 is a widely used standard among many professionals in the industry and is a flexible project management methodology that can accommodate with changing requirements of a project. PRINCE 2 is based on seven principles. These can be stated as follows, ***Continued Business Justification, Learn from Experience, Defined Roles and Responsibilities, Manage by Stages, Manage by Exception, Focus on Products, Tailor to Suit Project Environment.***

When considering these principles, and as we can see that the project being considered in this research does have tendency for requirement alteration as progress is being made and there is much to learn and discover through the life cycle of the research. PRINCE 2 can be considered a suitable project management framework to be employed for the management methodology of this project.

4.3 Potential Risks and Mitigation Plans

The following tables define potential risks and its respective mitigation plan.

Risk 01	Lack of Domain Knowledge				
Risk Level	High	**Phase**	All	**Frequency**	Medium
Mitigation	• Perform an in-depth review and analysis of currently available literature regarding domain and relevant topics				

Risk 02	Inability to meet set deadlines				
Risk Level	High	**Phase**	All	**Frequency**	High
Mitigation	• Meet with mentor and keep track of progress made on a regular basis. • Keep personal deadlines for each task by breaking it down further within main timeline				

Risk 03	Changes in Project Requirements				
Risk Level	High	**Phase**	All	**Frequency**	Medium
Mitigation	• Maintain priorities and make changes accordingly so as not to affect project timeline				

Risk 04	Technical limitations				
Risk Level	High	**Phase**	Implementation	**Frequency**	Medium
Mitigation	• Alter method of development to find work around said limitation				

Risk 04	Progress loss due to loss of data				
Risk Level	High	**Phase**	All	**Frequency**	Low
Mitigation	• Regular data back-ups to cloud storage • Maintain Github repository for code management • Have local backup of project if drastic change is being made in case roll-back is necessary.				

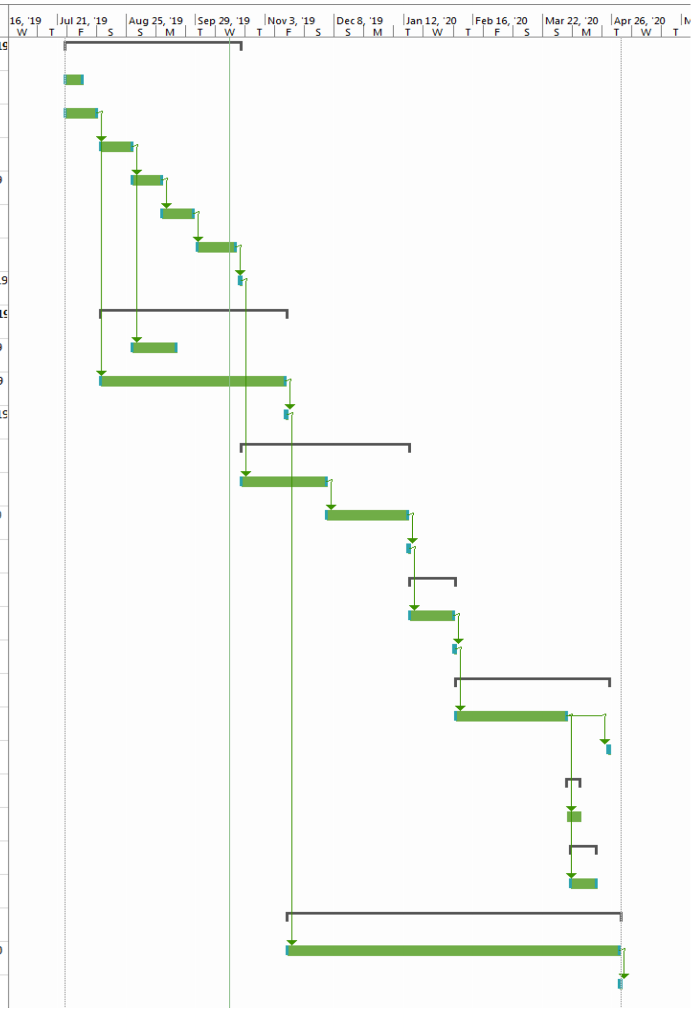
4.4 Work Breakdown Structure

Please refer on Appendix A

4.5 Gantt chart

Please refer on Appendix B.

ID		Task Mode	Task Name	Duration	Start	Finish
1			**Project Initiation**	**63 days**	**Thu 7/25/19**	**Mon 10/21/19**
2			Research on the project background	7 days	Thu 7/25/19	Fri 8/2/19
3			Research on the project domain	12 days	Thu 7/25/19	Fri 8/9/19
4			Research on the exsisting systems and libraries	12 days	Mon 8/12/19	Tue 8/27/19
5			Specify Project scope	11 days	Wed 8/28/19	Wed 9/11/19
6			Identifying resource requirements	12 days	Thu 9/12/19	Fri 9/27/19
7			Finalizing the PID	15 days	Mon 9/30/19	Fri 10/18/19
8			Submit PID	1 day	Mon 10/21/19	Mon 10/21/19
9			**Literature Review**	**68 days**	**Mon 8/12/19**	**Wed 11/13/19**
10			Research on existing systems	16 days	Wed 8/28/19	Wed 9/18/19
11			Prepare draft literature review	67 days	Mon 8/12/19	Tue 11/12/19
12			Submit draft literature review	1 day	Wed 11/13/19	Wed 11/13/19
13			**Requirement Gathering**	**61 days**	**Tue 10/22/19**	**Tue 1/14/20**
14			Requirement Specification	31 days	Tue 10/22/19	Tue 12/3/19
15			Create SRS document	29 days	Wed 12/4/19	Mon 1/13/20
16			Submit SRS document	1 day	Tue 1/14/20	Tue 1/14/20
17			**Design Prototype**	**17 days**	**Wed 1/15/20**	**Thu 2/6/20**
18			Create design specification	16 days	Wed 1/15/20	Wed 2/5/20
19			Submit design specification	1 day	Thu 2/6/20	Thu 2/6/20
20			**Implementation of prototype**	**56 days**	**Fri 2/7/20**	**Fri 4/24/20**
21			Build prototype	40 days	Fri 2/7/20	Thu 4/2/20
22			Prototype report	1 day	Fri 4/24/20	Fri 4/24/20
23			**Prototype Testing**	**5 days**	**Fri 4/3/20**	**Thu 4/9/20**
24			System testing	5 days	Fri 4/3/20	Thu 4/9/20
25			**Prototype evaluation**	**10 days**	**Sun 4/5/20**	**Fri 4/17/20**
26			Critical evaluation of the system	11 days	Sun 4/5/20	Fri 4/17/20
27			**Project Finalized**	**121 days**	**Thu 11/14/19**	**Thu 4/30/20**
28			Preperation of the finalized document	120 days	Thu 11/14/19	Wed 4/29/20
29			Submission for the final project document	1 day	Thu 4/30/20	Thu 4/30/20



4.6 Social, Legal, Ethical and Professional Aspects

Social	**Ethical**
• The platform will be available to anyone who wishes to use the software under a Standard End User License Agreement (EULA)	• Creating inaccurate result images - • In terms of an ethical standpoint, for data which might define county boundaries for example if produced inaccurately can cause a lot of issues if used for high influential statements.
Legal	**Professional**
• Computer Miss use Act Violation	• Violation of BCS Code of Conduct • Due regard for public privacy

4.7 Chapter Summary

5. Chapter 5: System Requirements Specification

5.1 Chapter Overview

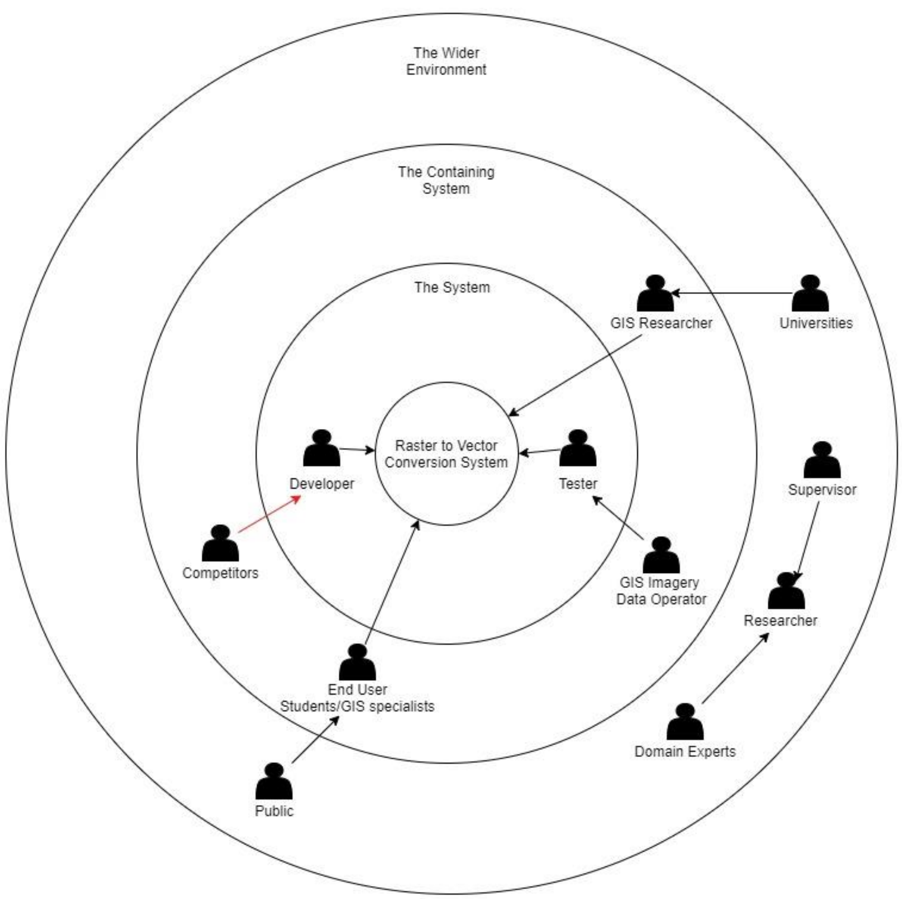
Write overview please

5.2 Stakeholder Analysis

The stakeholder analysis will be visually represented using an onion model diagram in this chapter and be further described using the roles and benefits table for each stakeholder.

5.2.1 Onion Model

A diagram used to visually represent the stakeholders' relationship among themselves and to the system proposed.



5.2.2 Stakeholders and Roles

The following table describes the stakeholders and their roles and benefits to the system.

Stakeholder	Roles	Benefits
Developer	Develop System	Develops the platform with less cost
Tester	Test System	Test and report the accuracy of the system
Domain Experts	Expert on the field of study (GIS)	Provide expertise on domain related matters to make system results accurate and to evaluate them
Universities	Functional Beneficiary	Is allowed to convert valuable geographical data from raster formats into vector formats and other educational uses
GIS Researches		Is allowed to convert valuable geographical data from raster formats into vector formats
End User		
Public		
GIS Imagery Data Operator	Gather Training and Testing Data	Make system more accurate if more data sets are found for training and testing
Supervisor	Assist in documentation and process of building system	Improve quality of system and its documentation by giving appropriate feedback

Researcher	Conduct Literature reviews and research to help implement system and identify requirements	Creates valuable documentation for ease of implementation of system
Competitor	Negative Stakeholder	Build similar solutions with better features

5.3 Requirement Elicitation Process

There are multiple techniques that can be used in order to validate and verify the requirements gathered. These can be states as Questionnaires, Observations, Literature Reviews etc. This section will briefly discuss the strengths and weaknesses of each such method and justify the method(s) of approach selected for this project.

Method 1: Questionnaire	
Questionnaires are carried out to gain knowledge of the developers who have a similar experience in the industry. This form is sent out to the target audience to get their feedback regarding the research.	
Advantages	• Larger research audience • Easy to analyses data obtained from questionnaire quantitatively
Disadvantages	• Feedback could contain facts that are untrue
Execution: A questionnaire is circulated among the target audience of the project which mostly comprises of developers working in the same field along with GIS Researchers who work on a daily basis with the cartographic and geographical image data that might require conversion between raster and vector. The questions were as follows. • To identify the use of raster to vector conversion software in GIS • To identify the level of the domain where GIS uses these conversion methods • To analyze the issues faced when developing such a system • To understand the domain understanding of developers when building a similar system.	

The information gathered from the questionnaire will be further discussed as this section continues.

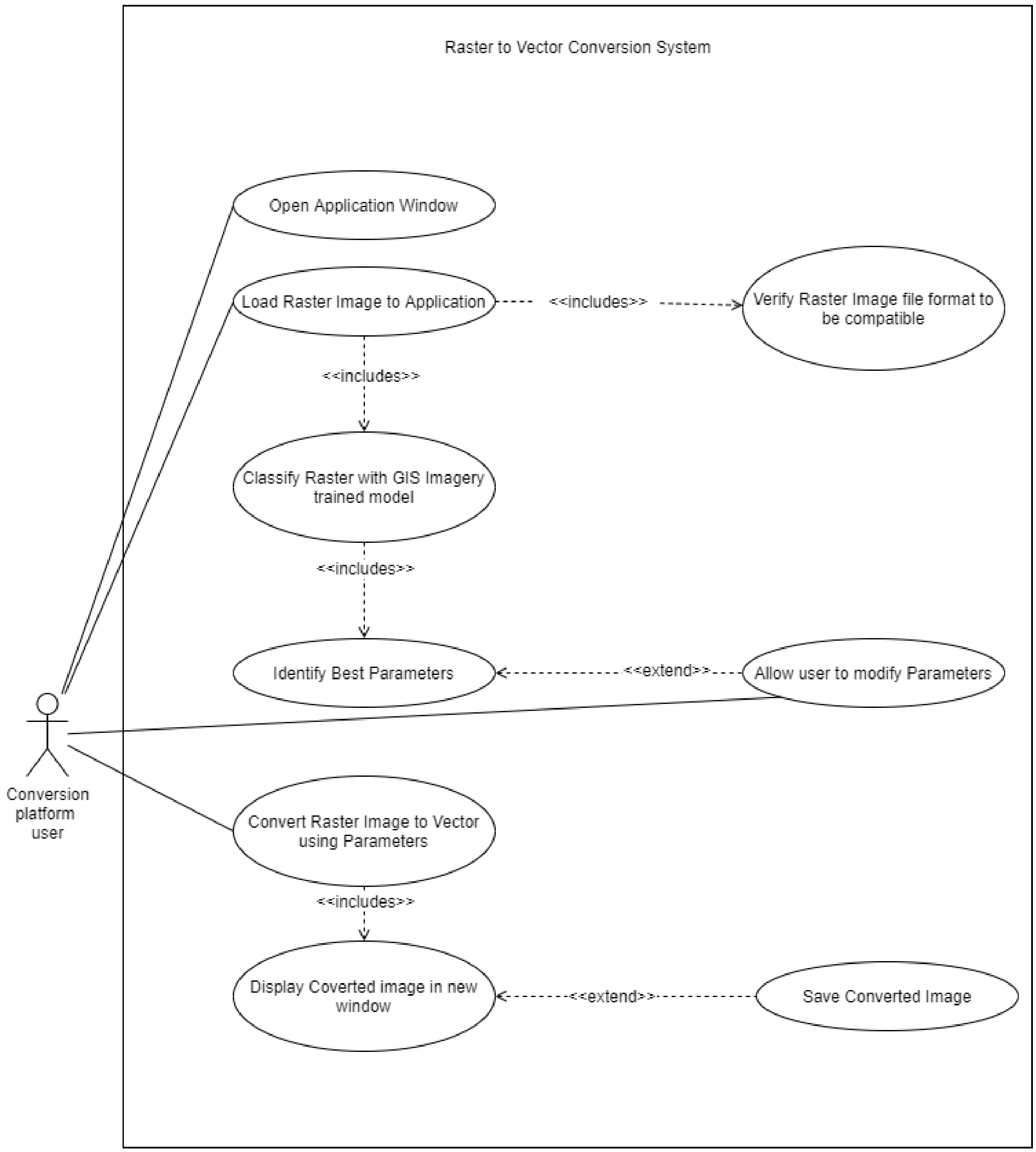
Method 2: Observations	
Observation of existing solutions proprietary and open source available in this domain, and to a find a unique gap that can be addressed and solved in this research.	
Advantages	• Identify short comings of current solutions
Disadvantages	• Observation varies according to observing individual
Execution: The author is reviewing existing systems and their functionalities, to identify their strength and weaknesses to address them in this research	

Method 3: Literature Review	
A widely used method in the research community when it is necessary to gain knowledge on a domain or various techniques and technologies currently being used is by conducting a Literature review on existing material. Research repositories such as IEEE, Science Direct and Google Scholar can be used for this purpose. Hence by studying this material a research gap can be identified in current system to be addressed.	
Advantages	• Identify short comings of current solutions
Disadvantages	• Observation varies according to observing individual
Execution: A literature review has been conducted using the reference of research papers and documents found from the research libraries mentioned. The literature found has been categorized as Image Processing, GI systems and Raster to Vector Conversion related topics. This section is addressed in the second chapter of this thesis document.	

5.4 Use Case Diagram and Description

5.4.1 Use Case Diagram

The following is a diagram that represents the use cases of the system visually.



5.4.2 Use Case Description

The following table will further elaborate the use cases shown in the above diagram.

Use Case ID	**UC-1**
Use Case Name	**Open Application Window**
Priority	High
Actors	Conversion Platform User
Description	Open the application and initialize the system
Pre-condition	Application package must be downloaded and extracted onto machine running a windows OS
Extending Use Cases	none
Including Use Cases	none
Triggering Event	Click on Application icon from windows file explorer
Main Flow	1. Open Application 2. Perform Initialization functions to prepare application
Alternative Flow	none
Exceptional Flow	none

Use Case ID	**UC-2**
Use Case Name	**Load Raster Image to Application**
Priority	High
Actors	Conversion Platform User
Description	Load a valid raster type image file onto application for processing and conversion
Pre-condition	Application must be opened and initialization functions should have been executed

Extending Use Cases	none
Including Use Cases	• Verify Raster Image file format to be compatible • Classify Raster with GIS imagery trained model
Triggering Event	Click on open file from application window
Main Flow	1. Open file explorer window to browse and locate file 2. Once file is selected load file into application as an array or matrices for processing and classification
Alternative Flow	none
Exceptional Flow	none

Use Case ID	**UC-3**
Use Case Name	**Verify Raster Image file format to be compatible**
Priority	High
Actors	
Description	Verify if the file format of the loaded file is one of the compatible file formats accepted by the system
Pre-condition	Application must be opened and initialization functions should have been executed and a file must be selected for the conversion process
Extending Use Cases	none
Including Use Cases	none
Triggering Event	Click on open file from file browser window

Main Flow	1. Load file onto the application 2. Check if the file belongs to one of the valid types 3. Display file loaded successfully message
Alternative Flow	3. If File is not one of the valid types show failed to load file message
Exceptional Flow	none

Use Case ID	**UC-4**
Use Case Name	**Classify Raster with GIS imagery trained model**
Priority	High
Actors	Conversion Platform User
Description	Classify image with trained classification model to aid automated identification of best fit conversion parameters for that particular loaded image
Pre-condition	Image must be loaded onto the application and be of a valid file format type
Extending Use Cases	none
Including Use Cases	Identify Best Parameters
Triggering Event	Click on Analyze button on application
Main Flow	1. Run classification model and classify image 2. Return classification accuracy data to find best match of GIS image classification
Alternative Flow	none
Exceptional Flow	none

Use Case ID	**UC-5**
Use Case Name	**Identify Best Parameters**
Priority	High
Actors	Conversion Platform User
Description	Use classification data to identify best fit conversion parameters for the image that was loaded
Pre-condition	Image classification must be performed on the loaded image
Extending Use Cases	Allow user to modify parameters
Including Use Cases	None
Triggering Event	System triggered after classification of image
Main Flow	1. Use parameter and classification data to identify best parameters for analyzed image 2. Return parameters for the conversion 3. Display parameters and allow editing to the user
Alternative Flow	none
Exceptional Flow	none

Use Case ID	**UC-6**
Use Case Name	**Convert Raster image to vector using parameters**
Priority	High
Actors	Conversion Platform User

Description	Convert Image using raster to vector conversion library using the parameters defined
Pre-condition	Conversion parameters for the particular image must be set
Extending Use Cases	none
Including Use Cases	Display converted image in new window
Triggering Event	Click on convert button on application
Main Flow	1. Initialize conversion package 2. Convert image using parameters set 3. Display success message for conversion
Alternative Flow	none
Exceptional Flow	3. If image conversion fails, due to not enough memory, catch exception and handle crashes that might occur

Use Case ID	**UC-7**
Use Case Name	**Allow User to Modify Parameters**
Priority	High
Actors	Conversion Platform User
Description	Allow the user to adjust parameters obtained for best conversion within a certain range at the users own discretion
Pre-condition	Conversion parameters for the particular image must be set
Extending Use Cases	none
Including Use Cases	none
Triggering Event	Edit parameter button clicked

Main Flow	1. Display editable fields for user to change parameters found before conversion 2. Edit parameters 3. Save new edited parameters

Use Case ID	**UC-8**
Use Case Name	**Display Converted Image in new Window**
Priority	High
Actors	Conversion Platform User
Description	Show preview of converted image to user
Pre-condition	Image conversion must be executed
Extending Use Cases	Save Converted Image
Including Use Cases	none
Triggering Event	System triggered after image conversion
Main Flow	1. Open new preview window 2. Show converted image in window
Alternative Flow	none
Exceptional Flow	none

Use Case ID	**UC-8**
Use Case Name	**Display Converted Image in new Window**
Priority	High
Actors	Conversion Platform User
Description	Save converted image
Pre-condition	Preview window of converted image must be open
Extending Use Cases	none
Including Use Cases	none
Triggering Event	System triggered after image conversion

Main Flow	1. Open file explorer window 2. Select save location 3. Press save button

5.5 Functional and Non Functional Requirements

5.5.1 Functional Requirements

FR No.	Requirement	Inputs	Process	Outputs	Priority	Use case
1	Convert Image from Raster to vector	Raster Image	Convert Raster to vector from identified parameters	Vector Image	Critical	UC-6
2	Classify GIS image with trained model	Raster Image	Use image classification model trained to identify types of GIS imagery and classify image	Image classification score	Critical	UC-4
3	Identify conversion parameters	Classification score	Use classification score to find best fit conversion parameters	Best fit parameters	Critical	UC-5
4	Review output vector image			Vector Image	Critical	Do tom.
5	Change auto selected parameters and redo		Convert Raster to vector from		Critical	Do tom

	conversion process		manual parameters			
6	Save converted file to local or cloud storage		Save file to storage		Important	Do tom

Table 5-1 Functional Requirements

5.5.2 Non-functional Requirements

#	Requirement title and description	Specification	Priority
1	Give user proper feedback on conversion process as it can be a long and time consuming process depending on the raster image and conversion parameters	Usability	Desirable
2	Result image should be visually similar and accurate to input raster image	Accuracy	Important
3	Develop API to allow conversions using online platforms	Usability	Desirable

Table 5-2 Non Functional Requirements

5.6 Chapter Summary

As chapter summary, first the stakeholders were identified and their roles were defined. After defining the roles, the requirement elicitation was carried out mainly with a questionnaire and a literature review. The outcomes of the questionnaire were discussed above with the statistics. Then the use case diagram of the system with the use case descriptions were discussed. After the use case diagram, the functional and non-functional requirements were identified.

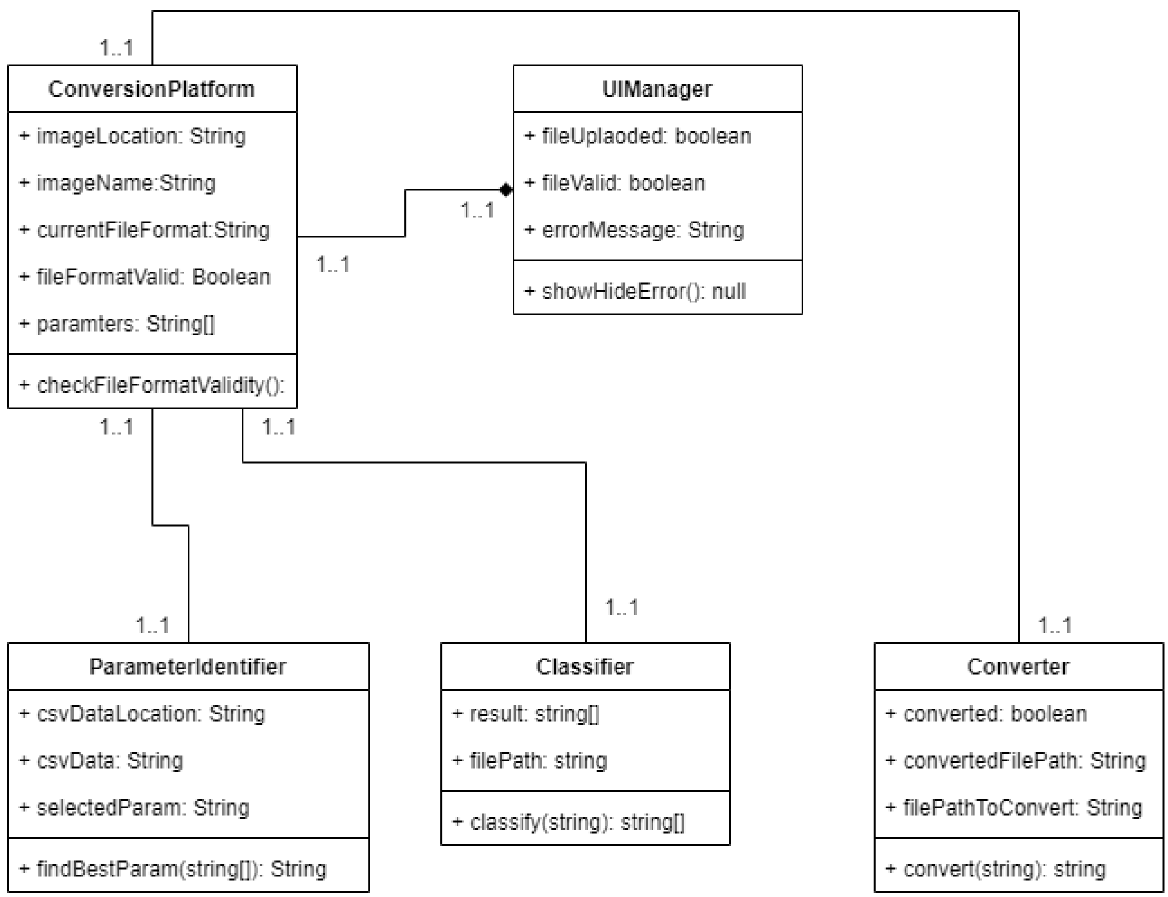
6. Chapter 6: Design

6.1 Chapter Overview

In this chapter, the class diagram, activity diagram and sequence diagram of the proposed solution will be elaborated and visually represented. Also an overview of all the diagrams will be elaborated at the end of this chapter.

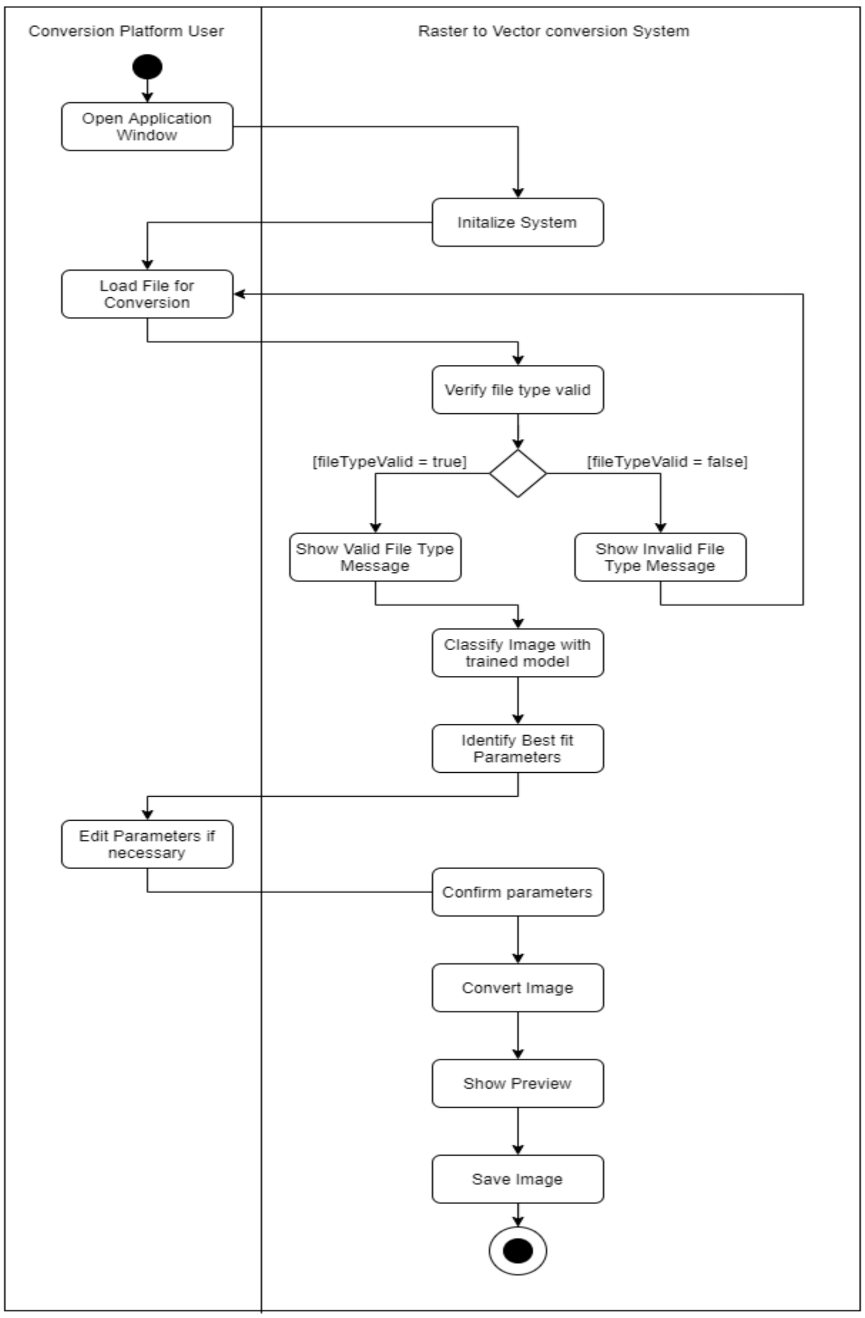
6.2 Class Diagram

Class diagram below represents the attributes and methods of each class in the proposed raster to vector conversion platform. Each class's relationship with each other is also shown and also aggregation/composition relationships.

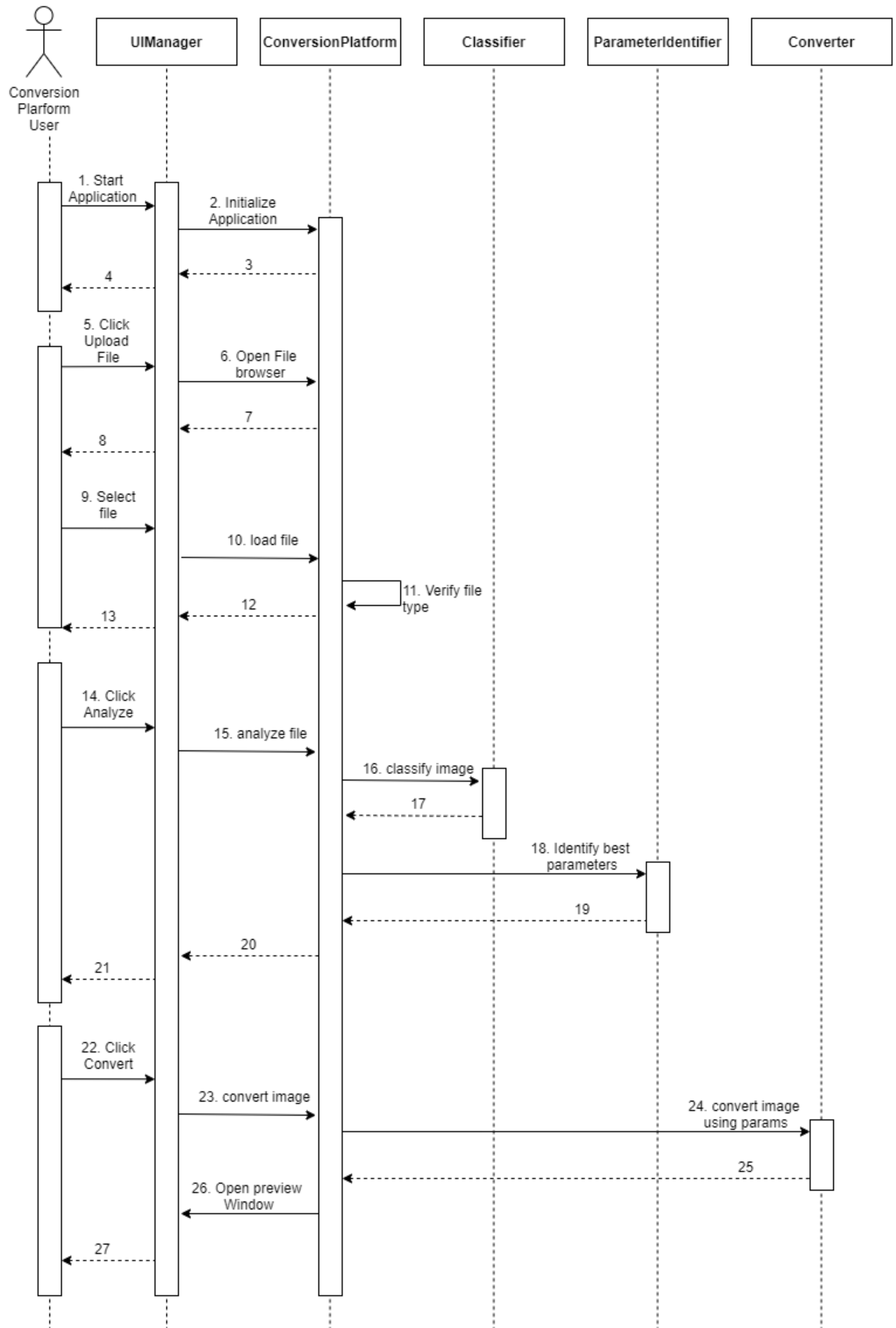


6.3 Activity Diagram

The flow of activities in the Raster to Vector conversion platform is shown by the activity diagram shown below. This diagram is created according to the flows that were described in the use case diagrams and descriptions. This can be identified as the successful flow of activities for converting an image from raster to vector.



6.4 Sequence Diagram



The diagram above represents how the user interacts with the raster to vector conversion platform. The user starts by opening up the application, which then initializes the conversion platform. Then the user can locate an image file that they wish to convert. Once the file is loaded. It will be verified if it is of the valid file types accepted by the system. If this is the case. The user will be allowed to analyze the file. The conversion platform then uses the image classification trained model to classify the image and get its accuracy as to which classification it belongs to. Using this information then the best fit parameters are identified. Then the user is allowed to click the convert button which then converts the image into a temporary location. The user is also showed a preview of the converted image. This can be considered the sequence followed for a successful image conversion.

6.5 Chapter Summary

Summary