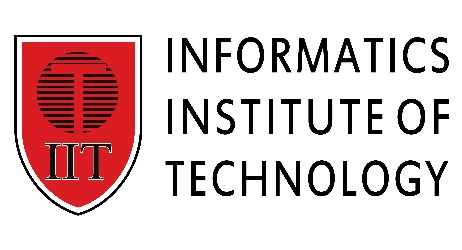
****

**Informatics Institute of Technology**

**In Collaboration With**

**University of Westminster (UoW)**

BSc. (Hons) in Computer Science

Final Year Project 2019/2020

**Interim Progress Report (IPR)**

**TransformVector**: Image processing based

Raster to Vector converter

By

Damodaran Dillon Thisaru Lakshman

2016086 | w1626703

Supervised By

Mr. Pumudu Fernando

**Contents**

[1. Project Initiation Document 3](#_Toc32220823)

[1.1 Introduction 3](#_Toc32220824)

[1.2 Background 3](#_Toc32220825)

[1.3 Problem Statement 5](#_Toc32220826)

[1.4 Research Question 5](#_Toc32220827)

[1.5 Draft Literature Review 6](#_Toc32220828)

[1.5.1 Problem Justification 6](#_Toc32220829)

[1.5.2 Requirement for Raster to Vector Conversion in GIS 6](#_Toc32220830)

[1.5.3 Similar Solutions and Differences between them 6](#_Toc32220831)

[1.5.4 Requirement for different types of conversion process depending on raster data structure properties 7](#_Toc32220832)

[1.6 Project Aim 8](#_Toc32220833)

[1.7 Motivation 8](#_Toc32220834)

[1.8 Objectives 8](#_Toc32220835)

[1.8.1 Research Objectives 8](#_Toc32220836)

[1.9 Proposed Solution 9](#_Toc32220837)

[1.9.1 Scope of the Project 9](#_Toc32220838)

[1.9.2 Anticipated Rich Picture Diagram 9](#_Toc32220839)

[1.9.3 Social, Ethical, Legal and Professional Aspects 10](#_Toc32220840)

[1.9.4 Functional and Non-Functional Requirements 10](#_Toc32220841)

[1.9.5 Resource Requirements 11](#_Toc32220842)

[1.10 Contribution 11](#_Toc32220843)

[1.11 Conclusion 12](#_Toc32220844)

[1.12 Timeline 13](#_Toc32220845)

[2. Literature Review 14](#_Toc32220846)

[2.1 Overview 14](#_Toc32220847)

[2.2 Problem Domain 14](#_Toc32220848)

[2.2.1 Geographic Information Systems 14](#_Toc32220849)

[2.2.2 Raster Image Data 15](#_Toc32220850)

[2.2.3 Vector Image Data 15](#_Toc32220851)

[2.2.4 Need for both Raster and Vector based data in GI systems 16](#_Toc32220852)

[2.2.5 Raster to Vector Conversion in GIS 18](#_Toc32220853)

[2.2.6 Image Processing 18](#_Toc32220854)

[2.3 Comparison of Similar Solutions 19](#_Toc32220855)

[2.3.1 Easy Trace 19](#_Toc32220856)

[2.3.2 FreeHand 20](#_Toc32220857)

[2.3.3 Illustrator 20](#_Toc32220858)

[2.3.4 Potrace 20](#_Toc32220859)

[2.4 Justification of the Selected Approach 21](#_Toc32220860)

[2.5 Algorithmic Analysis 21](#_Toc32220861)

[2.5.1 General Raster to Vector Conversion Algorithms 21](#_Toc32220862)

[2.5.2 Potrace Raster to Vector Algorithm 22](#_Toc32220863)

[2.6 Chapter Summary 22](#_Toc32220864)

[3. System Requirements Specification 24](#_Toc32220865)

[3.1 Chapter Overview 24](#_Toc32220866)

[3.2 Stake Holder Analysis 24](#_Toc32220867)

[3.2.1 Onion Model 24](#_Toc32220868)

[3.2.2 Stakeholders and Roles 25](#_Toc32220869)

[3.3 Requirement Elicitation Techniques and Execution 26](#_Toc32220870)

[3.4 Use Case Diagram 28](#_Toc32220871)

[3.5 Use Case Description 28](#_Toc32220872)

[3.5.1 Functional and Non Functional Requirements 29](#_Toc32220873)

[3.5.2 Functional Requirements 29](#_Toc32220874)

[3.5.3 Non-functional Requirements 30](#_Toc32220875)

[3.6 Chapter Summary 30](#_Toc32220876)

[4. References 31](#_Toc32220877)

**List of Figures**

[Figure 1Rich Picture Diagram 12](#_Toc32220878)

[Figure 2 Gantt Chart 15](file:///C:\Users\Dillon%20Lakshman\Documents\IPR.docx#_Toc32220879)

[Figure 2.4‑1 Visual comparison between vector and bitmap image 18](#_Toc32220880)

[Figure 5.2‑1A detail from the seal of Stanford University; the original scanned image, 24](#_Toc32220881)

[Figure 2‑1Onion Model 26](#_Toc32220882)

[Figure 4‑1Use case 1 30](#_Toc32220883)

**List of tables**

[Table 1 Comparison of Commercial Raster to Vector Software 8](#_Toc32220884)

[Table 2 In Scope and Out of Scope Analysis 11](#_Toc32220885)

[Table 3 SLEP Analysis 12](#_Toc32220886)

[Table 2.6‑1 Comparison of Similar Solutions 21](#_Toc32220887)

[Table 2‑1Stakeholders and Roles 27](#_Toc32220888)

[Table 3‑1 Requirement Elicitation - Questionnaires 28](#_Toc32220889)

[Table 3‑2 Requirement Elicitation - Observations 29](#_Toc32220890)

[Table 3‑3 Requirement Elicitation - LR 29](#_Toc32220891)

[Table 5‑1Use case 1 Description 31](#_Toc32220892)

[Table 6‑1 Functional Requirements 32](#_Toc32220893)

[Table 6‑2 Non Functional Requirements 32](#_Toc32220894)

# Project Initiation Document

## Introduction

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. 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. 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).

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. 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 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.

## Background

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.

## Problem Statement

Some spatial data, such as aerial imageries, must be represented by raster data. Vector maps are unable to describe and render them (Lin and Guo, 2011).

With the development of fast high performance computing, many fields are widely applying it to improve performance in processing large amounts of data and accomplishing high computational tasks in smaller time. As Vector data is much simpler and more precise in location this helps to take advantage of the high performing computational power in a much more efficient manner (Wei *et al.*, 2012).Vector base graphics also take advantage over raster images for its flexibility, compactness etc.(Lin *et al.*, 2015). By considering the following reasons, it can be understood that Raster as well as Vector images are equally needed when data sets are required for processing in Geographical Information Systems. Even though many commercial solutions have been developed to achieve this task, many of these software only work well under certain parameters and conditions. There arises a need for a common solution of conversion between these two formats. This research will be focused on the conversion of Raster data structures into Vector data structures by identifying different properties of the Raster data structure and using the best method of conversion to convert it into Vector data.

## 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?

## Draft Literature Review

### 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)

### Requirement for Raster to Vector Conversion in GIS

Vector map is widely used in geographic information system (Wu, Yang and Wu, 2013).We generally use images (satellite or airborne) in GIS by stacking vector information to visualize thematic maps(Grandchamp, 2010). Current graphic devices which are used for high speed computer input and output of cartographic data are mostly raster oriented. Some of these can be identified as the rotating drum scanner and the color raster display. Even though the solutions are mostly based on Raster techniques, most of the widely used manipulative techniques in computer assisted cartography and automated spatial data handling require the data to be in Vector data structure. This is one of the main reasons for the need of techniques for converting data structures from Raster to Vector (Peuquet, 1981).

### Similar Solutions and Differences between them

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 1 Comparison of Commercial Raster to Vector Software

### Requirement for different types of conversion process depending on raster data structure properties

By going through existing literature regarding Raster to Vector conversion methods, it can be identified that as the output produced by each conversion method is unique. There are quantitative results as well as use case based situations that show us that the conversion method to be used in a certain situation can depend heavily on the Raster data structure that the conversion has to be performed on.

#### Raster to vector algorithms where properties of the raster data structure has to be considered

There are many optical scanners which scan cartographic data such as maps, engineering data or text. These scanners produce a matrix of binary pixels (zero or one) which convert into black and white pixels. These scanners are scan at a resolution of 25 microns and at a rate on the order of 1 million bits per second. These scanners are also able of handling material as large as 40 x 48 inches. The binary matrix created from such a scanner could contain over 1.9 billion pixels. Due to the large size of the Raster data structure created by such a scanner high speed Raster to vector conversion algorithms have been created.

These types of algorithms involve in finding and operating on general representations of pixel data in a way in which where the actual pixels are not examined. These techniques can be applied for rapid feature from imagery, therefore making the process of Raster to conversion possible for such large images as mentioned above (L. Gibson and D. Lucas, 1982).

#### Qualitatively analyzing results of Raster to Vector conversion depending on conversion method

In a study done to quantitatively evaluated several raster to vector conversion methods such as   
(i) thinning, (ii)medial line and (iii)line fitting algorithms show data on properties such as End points, T-junctions, Crossing points and Corners in the final output line produced according to each algorithm.

This quantitative measurement was produced by measuring produced synthetic ideal CAD models and the difference between input and output lines. The number of missing and matching feature points and lines, and the accuracy of their locations result in quantitative measures (Kasturi and (eds.), 1996).

Even though this qualitative measurement has been done on CAD data sets, the result of these being performed on GIS data sets can also be considered similar as both these data sets are ultimately Raster images.

## Project Aim

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

## 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.

## Objectives

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

### 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.

## Proposed Solution

The proposed solution for the problem domain identified will be discussed in this section.

### Scope of the Project

|  |  |
| --- | --- |
| In Scope | Out of 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 | * 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. |

Table 2 In Scope and Out of Scope Analysis

### Anticipated Rich Picture Diagram

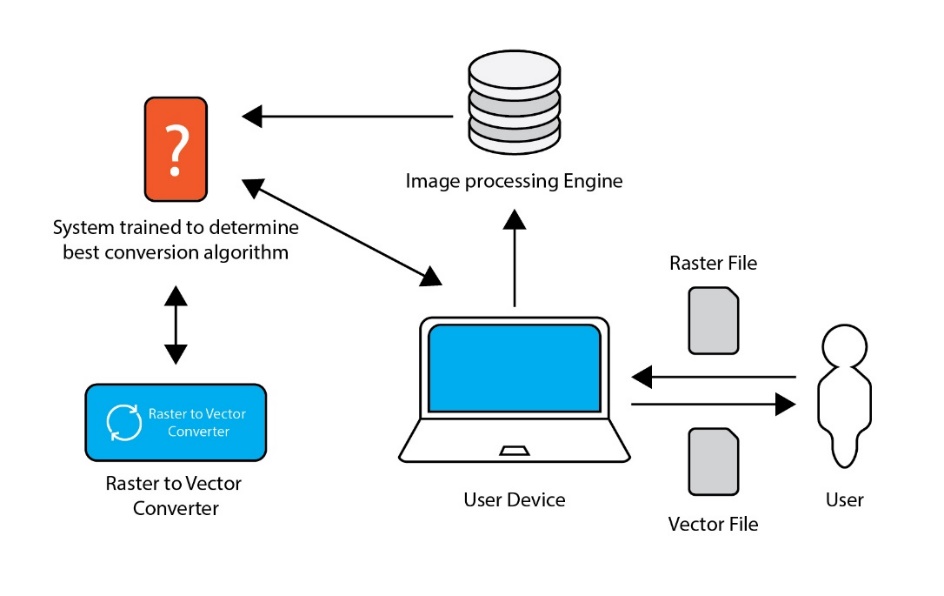


Figure 1Rich Picture Diagram

As shown in the figure 9.1 the user of the platform will be upload a file of a Raster format this file then will be fed into the image processing system which will determine the best conversion method for the image to be converted. Then using these conversion settings, the image will be converted to of a Vector format and returned to the user through the platform solution created.

### Social, Ethical, Legal and Professional Aspects

|  |  |
| --- | --- |
| Social:  The platform will be available to anyone who wishes to use the software under a Standard End User License Agreement (EULA) | Ethical:  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:  Computer Miss use Act Violation | Professional:  Violation of BCS Code of Conduct |

Table 3 SLEP Analysis

### Functional and Non-Functional Requirements

#### Functional Requirements

* Developing an image processing system that identifies certain properties of a raster image
* Developing a system to identify the most suitable conversion method out of given methods
* Developing a system to convert the Raster image to Vector using the best found method
* Allowing user to select between performance and accuracy of conversion

#### Non Functional Requirements

* Generating visual feedback on conversion progress
* Allow users to tweak settings using simple user friendly UI

### Resource Requirements

#### 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

#### 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

## Contribution

In this section, the individual contribution of this project which will be stated to the research domain will be discussed. The contributions are as follows.

* Proposing a new standard for Raster to Vector Conversion
* Create an image processing platform that identifies certain properties that aid in determining the best method of Raster to Vector conversion.
* Creating a model which determines best method of conversion according to properties of image.
* Creating a platform which converts the Raster to Vector with best conversion method after identifying it

## Conclusion

Concluding the Project Initiation document, this research mainly focuses on Raster to Vector conversions in GIS. The drawbacks faced by most commercial and widely used products are taken into consideration and a platform in which the strengths of those solutions are identified and then combined to form a better solution.

## TimelineC:\Users\Dillon Lakshman\Downloads\Dillon.gif

Figure 2 Gantt Chart

# Literature Review

## 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.

## Problem Domain

### 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).

### 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).

### 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).

### 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).

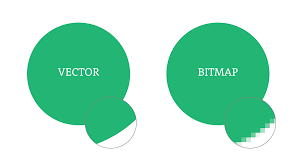


Figure 2.4‑1 Visual comparison between vector and bitmap image

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.

### 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.

### 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**

## **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 2.6‑1 Comparison of Similar Solutions

### 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.

### 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.

### 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.

### 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.

## 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.

## Algorithmic Analysis

### 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

### 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.

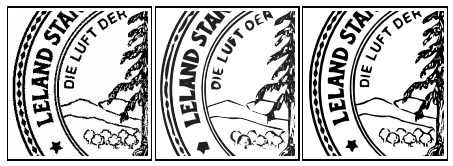


Figure 5.2‑1A detail from the seal of Stanford University; the original scanned image,

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).

## 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.

# System Requirements Specification

## Chapter Overview

The aim of this chapter is to gather requirements for the Transform Vector System and to identify functional and non-functional requirements from them. A stakeholder analysis is initially carried out and then requirement gathering techniques with each ones’ outcome respectively will be discussed. Use case diagrams and descriptions will also be discussed in order to further clarify the functional and non-functional requirements identified.

## Stake Holder Analysis

The stakeholder analysis is depicted with the help of the following onion model diagram. The diagram is further explained under the stakeholders and roles table after the diagram.

### Onion Model

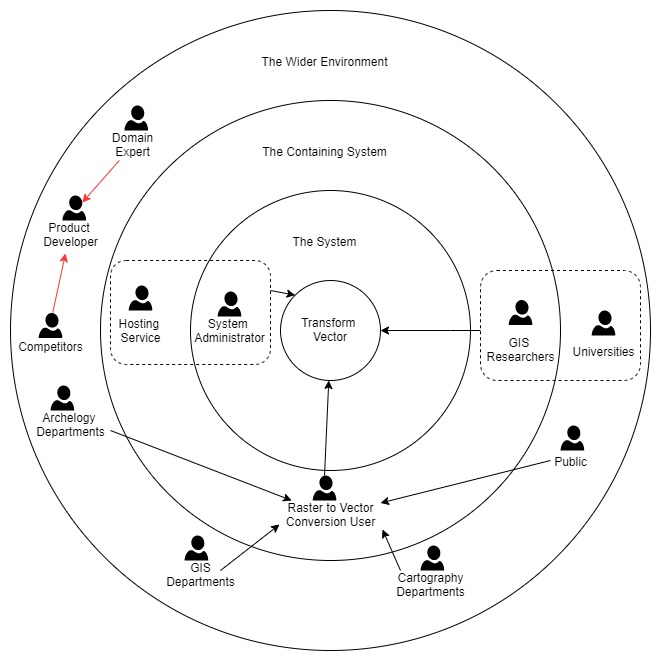


Figure 2‑1Onion Model

The stakeholder roles which the onion model in *Figure 2-1* are pointed out below the table.

### Stakeholders and Roles

|  |  |  |
| --- | --- | --- |
| **Stakeholder** | **Roles** | **Benefits** |
| Product Developer | Developer | Develops the platform with less cost |
| Domain Expert | Expert | Provide expertise on domain related matters to make system results accurate and to evaluate them |
| GIS Researches | Functional Beneficiary | Is allowed to convert valuable geographical data from raster formats into vector formats |
| Universities | Is allowed to convert valuable geographical data from raster formats into vector formats and other educational uses |
| Public | Is allowed to convert valuable geographical data from raster formats into vector formats |
| Cartography Department |
| GIS Departments |
| Archeology Departments |
| R2V Conversion User |
| Hosting Service | Host the servers for the application processing on the Cloud |
| System Administrator | Operational Administration | Deploys the system from the developing environment to production environment |
| Competitor | Negative Stakeholder | Build similar solutions with better features |

Table 2‑1Stakeholders and Roles

## Requirement Elicitation Techniques and Execution

There are multiple ways identified that will be used to identify and validate the requirements found. These can be stated as Questionnaires, Observations, Literature Reviews and etc. This section of this document will discuss the strength and weakness of each such method and how it’s execution is carried out.

|  |  |
| --- | --- |
| **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. | |

Table 3‑1 Requirement Elicitation - Questionnaires

|  |  |
| --- | --- |
| **Method 2: Observations** | |
| Observation of existing solutions for this domain and to a find a uniqueness 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 | |

Table 3‑2 Requirement Elicitation - Observations

|  |  |
| --- | --- |
| **Method 3: Literature Review** | |
| A widely used method to research on existing solutions and studies conducted on the field that this paper wishes to address is by reading through the highly esteemed research repositories such as IEEE and ScienceDirect. These contain the latest knowledge and the unique approaches which have been addressed by other researches to find a research gap. | |
| **Advantages** | * Identify short comings of current solutions |
| **Disadvantages** | * Observation varies according to observing individual |
| **Execution:** IEEE, ScienceDirect and Google Scholar research repositories have been used and the papers categorized under Image processing, Raster to vector conversion and GIS have been studied. This section is covered in the second chapter of the thesis. | |

Table 3‑3 Requirement Elicitation - LR

## Use Case Diagram

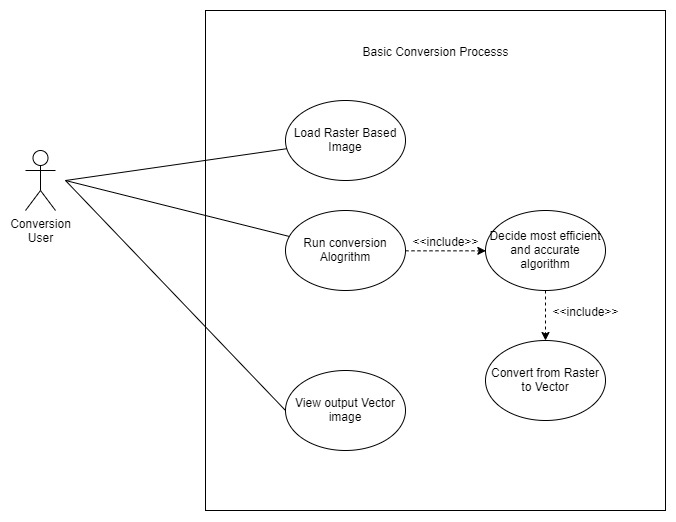


Figure 4‑1Use case 1

## Use Case Description

|  |  |
| --- | --- |
| **Use Case Name** | **Basic Conversion Process** |
| Description | The conversion process of a raster image to a vector format |
| Priority | High |
| Participating actors | Conversion User |
| Pre-conditions | Image should be of compatible raster types |
| Include Use Cases | * Decide most efficient and accurate algorithm for conversion * Convert from Raster to vector by running found method |
| Main Flow | * Load Raster image for conversion * Run conversion process * Decide most efficient algorithm * Convert image using that algorithm * Display converted vector image |
| Exceptional Flows | * Invalid raster image added * Conversion process ran out of hardware resources when running algorithm(Raster image dimension to large) |
| Post Conditions | * Actor can re-run conversion with custom properties if auto found properties not satisfactory |

Table 5‑1Use case 1 Description

### Functional and Non Functional Requirements

### Functional Requirements

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **FR No.** | **Requirement** | **Inputs** | **Process** | **Outputs** | **Priority** | **Usecase** |
| **1** | Convert Image from Raster to vector | Raster Image | Convert Raster to vector from identified parameters | Vector Image | Critical | 1 |
| **2** | Review output vector image |  |  | Vector Image | Critical | 1 |
| **3** | Change auto selected parameters and redo conversion process |  | Convert Raster to vector from manual parameters |  | Critical | 1 |
| **4** | Save converted file to local or cloud storage |  | Save file to storage |  | Important | 1 |

Table 6‑1 Functional Requirements

### 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 6‑2 Non Functional Requirements

## 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.

# References

Al-Bayari, O. (2018) ‘GIS cloud computing methodology’, *CITS 2018 - 2018 International Conference on Computer, Information and Telecommunication Systems*. IEEE, pp. 1–5. doi: 10.1109/CITS.2018.8440176.

Berry, J. K. (1995) ‘Raster is Faster, But Vector is Corrector’, *GIS World*, 8(6), p. 35. Available at: https://www.esri.com.

Britannica, T. E. of E. (2014) ‘No Title’, *Encyclopædia Britannica*. Encyclopædia Britannica, inc. Available at: https://www.britannica.com/technology/raster-graphics.

Chrisman, N. R. (1999) ‘Chrisman - What does GIS Mean’, 3(2), pp. 175–186.

Couclelis, H. (1992) ‘People Manipulate Objects but Cultivate Fields’, *Proceedings from International Conference on GIS*, pp. 65–76. Available at: http://clamsitel.pbworks.com/w/file/fetch/49252349/Couclelis 1992 Objects Fields.pdf.

Eduardo A.B. da Silva, G. V. M. (2005) *The Electrical Engineering Handbook*. Edited by W.-K. CHEN. Academic Press. Available at: https://doi.org/10.1016/B978-012170960-0/50001-3.%0A(http://www.sciencedirect.com/science/article/pii/B9780121709600500013).

Gold, C. M. (2006) ‘What is GIS and what is not?’, *Transactions in GIS*, 10(4), pp. 505–519. doi: 10.1111/j.1467-9671.2006.01009.x.

Grandchamp, E. (2010) ‘Raster-vector cooperation algorithm for GIS: Application to Ecological Units delineation’, *2nd International Conference on Advanced Geographic Information Systems, Applications, and Services, GEOProcessing 2010*, pp. 39–43. doi: 10.1109/GEOProcessing.2010.32.

He, B. S., Xu, X. L. and Zheng, T. (2009) ‘Vector graphics rendering on mobile device’, *Proceedings - 2009 WRI International Conference on Communications and Mobile Computing, CMC 2009*, 3, pp. 8–11. doi: 10.1109/CMC.2009.285.

Hilaire, X. and Tombre, K. (2006) ‘Robust and accurate vectorization of line drawings’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6), pp. 890–904. doi: 10.1109/TPAMI.2006.127.

Holroyd, F. and Bell, S. B. M. (1992) ‘Raster GIS: Models of raster encoding’, *Computers and Geosciences*, 18(4), pp. 419–426. doi: 10.1016/0098-3004(92)90071-X.

Kasturi, R. and (eds.), K. T. (1996) ‘Graphics Recognition: Methods and Applications’, *First International Workshop, University Park, PA, USA, August 1995, Selected papers published as Lecture Notes in Computer Science*, 1072(May). doi: 10.1007/3-540-61226-2.

L. Gibson and D. Lucas (1982) ‘Vectorization of Raster images Using Hierarchical Methods’, *COMPUTER GRAPHICS AND IMAGE PROCESSING*, 20, pp. 82–89.

Lacroix, V. (2009) ‘Raster-to-vector conversion: problems and tools towards a solution a map segmentation application’, *Proceedings of the 7th International Conference on Advances in Pattern Recognition, ICAPR 2009*, 1(c), pp. 318–321. doi: 10.1109/ICAPR.2009.96.

Lin, F. and Guo, C. (2011) ‘Raster-vector integration based on SVG on mobile GIS platform’, *Proceedings - 2011 6th International Conference on Pervasive Computing and Applications, ICPCA 2011*, pp. 378–383. doi: 10.1109/ICPCA.2011.6106534.

Lin, J. *et al.* (2015) ‘Diffusion Based Vector Graphics on Mobile Devices’, *Proceedings - 2015 9th International Conference on Frontier of Computer Science and Technology, FCST 2015*, pp. 153–156. doi: 10.1109/FCST.2015.12.

Liu, W. and Josep Lladós (Eds.) (2005) *Graphics Recognition: Ten Years Review and Future Perspectives*, *6th International Workshop, GREC 2005 Hong Kong, China, August 25-26, 2005 Revised Selected Papers*. doi: 10.1007/11767978\_5.

Peuquet, D. J. (1981) ‘An examination of techniques for reformatting digital cartographic data/part 1: the raster-to- vector process.’, *Cartographica*, 18(1), pp. 34–48. doi: 10.3138/VL12-8581-0733-0Q11.

Seel-audom, C., Naiyapo, W. and Chouvatut, V. (2017) ‘A search for Geometric shapes in a Vector Image’, *9th International Conference on Knowledge and Smart Technology (KST)*, pp. 305–310.

Selinger, P. (2003) *Potrace: a polygon-based tracing algorithm*.

SEVERENUK, T. *et al.* (2019) ‘VECTOR GRAPHICS BASED LIVE SKETCHING METHODS AND SYSTEMS’.

Sloan, K. R. and Tanimoto, S. L. (1979) ‘Progressive Refinement of Raster Images’, *IEEE Transactions on Computers*, C–28(11), pp. 871–874. doi: 10.1109/TC.1979.1675269.

Wade, T. G. *et al.* (2003) ‘Vector & Raster Methods’, 69(12), pp. 1399–1405.

Wei, J. *et al.* (2012) ‘Parallel algorithm designed for polygon vectorization’, *Proceedings - 2012 20th International Conference on Geoinformatics, Geoinformatics 2012*, pp. 0–3. doi: 10.1109/Geoinformatics.2012.6270305.

Winnemoeller, H. *et al.* (2018) ‘ENHANCED VECTORIZATION OF RASTER IMAGES’.

Wong, D. W. S. and Wu, C. V. (1996) ‘Spatial metadata and GIS for decision support’, *Proceedings of the Annual Hawaii International Conference on System Sciences*, 3, pp. 557–566. doi: 10.1109/HICSS.1996.493251.

Wu, J., Yang, F. and Wu, C. (2013) ‘Review of digital watermarking for 2D-vector map’, *Proceedings - 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, GreenCom-iThings-CPSCom 2013*, pp. 2098–2101. doi: 10.1109/GreenCom-iThings-CPSCom.2013.394.