Statistical accuracy of an extraction algorithm for linear image objects

by

Renate Nicole Thiede

Submitted in partial fulfillment of the requirements for the degree

Magister Scientiae

In the Department of Statistics
In the Faculty of Natural and Agricultural Sciences
Univeristy of Pretoria

October 2019

I, *Renate Nicole Thiede,* declare that this mini-dissertation (100 credits), which I hereby submit for the degree Magister Scientiae in Mathematical Statistics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: 

Date: 11 February 2020

Contents

Abstract

Informal unpaved roads in developing countries arise naturally through human movement and informal housing setups. These roads are not authorised nor maintained by council, nor recorded in official databases or online maps. Mapping such roads from satellite images is a common problem, as information on these roads is critical for sustainable city growth. Information on their location and extent may be gleaned from spatial big data, however, no automatic or semi-automatic approach is freely available. This research develops a novel algorithm for extracting informal roads from multispectral satellite images, using physical road characteristics. These include near-infrared reflectance, addressed via the NDVI index, shape, addressed via measures of compactness and elongation, and grey-value intensity. The crux of the algorithm is the Discrete Pulse Transform, implemented via the Roadmaker's Pavage. The algorithm provides a classification of road objects, along with an associated uncertainty measure for each road object. Accuracy is assessed using per-pixel assessment metrics and metrics based on road characteristics, including completeness, correctness, and Pratt's Figure of Merit, which is applied to road extraction accuracy for the first time. The algorithm is applied to areas in Gauteng and North West Provinces, South Africa. Sources of uncertainty and error are discussed, such as indefinite boundaries, surface type heterogeneity, trees and shadows.

Chapter 1

Introduction

1.1 Background

The world is currently experiencing unprecedented levels of urbanisation. Most of the world's population lives in urban areas, and city growth and sprawl is expected to continue [128]. According to estimates by the United Nations, 66% of the world's population will reside in urban areas by 2050 [95]. This global trend is most present in the developing world [6], where the rapid growth of cities often places a requirement for infrastructure that local governments are unable to provide [62]. Where the pressing need for infrastructure is not met by government, it may be fulfilled ad-hoc by residents, leading to the appearance of settlements and roads that have been constructed by the inhabitants of an area, without the knowledge or approval of government. In this dissertation, such infrastructure will be referred to as informal, while the term formal infrastructure is that constructed or recognised by government.

Two-thirds of the population of the global south lives in such informally developed settlements [112]. In South Africa, the informally settled population is growing at a faster rate than the population living in government-registered settlements [112]. This presents a national and international problem, as informally developed infrastructure may not comply with government regulations in terms of safety and sustainability. Services such as running water and electricity may not be available in these areas [140], and there may be limited to no access to job opportunities and facilities such as schools and clinics. Addressing the problem of informal development is critical for ensuring sustainable and equitable city growth.

The first problem to be addressed is determining the location and extent of informal infrastructure. This information is not usually recorded in government databases [63]. Many studies have focused on analysing informal settlements and housing setups [99, 1], including obtaining information on their spatial and socio-economic characteristics from remote sensing data [62, 97, 52]. However, less attention has been given to

the problem of informal roads. Informal roads occur inside informal settlements in order to link points of interest within these settlements, as well as to connect the settlement to the outside world. Informal settlements are mostly residential in nature, and do not generally contain establishments such as schools or clinics. Job opportunities within informal settlements are very limited. Informal roads may therefore be crucial for providing access to facilities and opportunities that are critical to residents. Informal roads may also occur in formally settled areas, where the settlements have grown to exceed the capacity of formal roads, or where no formal roads exist to connect these settlements with other areas.

In essence, informal roads indicate a real, crucial need experienced by citizens. However, the use of informal infrastructure is not sustainable or equitable in the long run. Informal roads may not be safely navigable. Their surfaces are not maintained officially, road shoulders may not be present, roads may not have constant widths or clear edges, turns may be too sharp, or gradients may be too steep. Additionally, roads may be constructed close to houses, leading to danger for pedestrians, as well as air and noise pollution. Despite this, these roads demonstrate a real navigational need within the community, and should therefore not simply be removed or ignored by government. Instead, informal infrastructure may be upgraded and registered as formal infrastructure. This is often done in the developing world to ensuring sufficient infrastructure to meet the real needs of the growing population [10].

The process of converting infrastructure from informal to formal is known as formalisation. For roads, this process should entail improving them to ensure that they comply with regulation and are safe to use, paving them if necessary. Formalisation improves conditions on navigable routes that residents have identified as desirable and convenient for travel. Thereby, access to schools, job opportunities, clinics, and other critical facilities and services is improved. During formalisation, roads should also be registered to government databases, and may be made available on GPS. Knowledge of the locations and capacity of these roads will enable local government to better gauge the need of its citizens. Services may be provided along these roads, including water pipes and power lines. Improving navigation in these areas will lead to better access for emergency services, and ensure safe and efficient navigation for non-residents who may not be familiar with the area and rely on GPS [154]. Formalising informal roads is therefore a critical step in ensuring sustainable and inclusive city growth. This is in line with Target 9.1 of Sustainable Development Goal (SDG) 9 on Industry, Innovation and Infrastructure, which specifies the need for sustainable infrastructure for affordable and equitable access for all.

1.2 Problem Statement

Information on informal roads may be gleaned from satellite data. Digitised road networks have recently been used as inputs to calculate access indicators in line with SDG 9 [56]. Given an efficient method for

extracting relevant information, this provides a non-intrusive, time- and cost-effective way of obtaining such information, as satellite images provide a comprehensive overview of settlements and infrastructure from the air. This reduces the need for ground-based surveying, which may be expensive, include security risks in areas of high crime, or disrupt the activities of residents. A semi-automatic or automatic technique for obtaining this information from satellite data is preferred. While roads may be manually digitised from satellite imagery with high accuracy, this is a highly time-consuming process, which is not suitable for informal roads, given the rapidity with which informal infrastructure may form.

Figure 1.1 illustrates the rapid development of an informal settlement, with associated roads. The images are of an area in Khayelitsha, Western Cape, South Africa, and were taken in 2018. On 21 April (a), there is a large unoccupied open area beside the formal settlement. On 16 May (b), less than a month later, sparse houses are visible. By 18 July (c), the area is well-populated, and various informal roads have arisen in that area which are visibly delineated by houses, and clearly used for navigation. Thus an informal settlement has come into existence, along within well-used roads, within a space of three months. Note how the S-shaped bare soil area in the south-western part of the image is immediately utilised as a road, with houses built around it. This demonstrates how existing natural features may encourage the formation of informal roads.

a)

b)

c)

Figure 1.1: Images showing the rapid development of an informal settlement, including informal roads. The area is in Khayelitsha, Western Cape, South Africa. All images are from 2018. a) 21 April. b) 16 May. c) 18 July. (Imagery ©2019 Maxar Technologies. Coordinates: 18°40′51.37”E, 34°02′02.54”S.)

Information on informal roads is not readily available in online maps, and the existing information in such maps is often outdated or incorrect. Figure 1.2 illustrates that reliable information on informal roads may not be available in online maps. A large part of the image in (a) has not been mapped, as shown in circle B, while circles A, C and D show that captured roads may not be at all aligned with the real-world roads on the ground. In Figure 1.2 (b), the road in circle A is not aligned with its corresponding real-world road. The roads in circle B are disconnected, and a piece of road is missing. In circle C, some footpaths were captured as roads, while others were not. The roads captured herein were compared to those mapped in other online maps, including Open Street Map[1], Sygic[2], coPilot[3] and Waze[4]. None of these alternatives provided comprehensive up-to-date coverage of either area. This demonstrates that the need for extracting information about informal roads from satellite images has not been satisfied.

A semi-automatic or automatic road extraction method is required to efficiently obtain up-to-date information on informal roads from satellite data in a time- and cost-effective manner. Such a method is yet to be developed, and is the central problem addressed in this research.

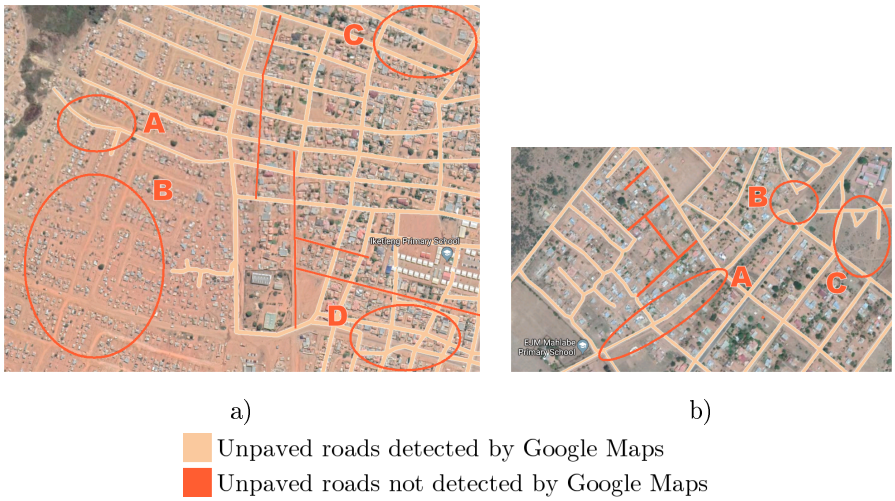


Figure 1.2: Images from Google Maps demonstrating that informal roads are not generally well-captured in online maps. a) An area of Hammanskraal, Gauteng Province. Areas A, C and D demonstrate misaligned roads, where the captured roads do not match the real-world roads. Area B shows roads not captured by Google Maps. b) A settlement in Klipgat, North West Province. Area A shows misaligned roads, area B shows disconnected roads, and area C demonstrates that some footpaths are captured, while others are not. (Imagery ©2019 CNES / Airbus Maxar Technologies. Map data ©AfriGIS (Pty) Ltd. Coordinates of (a): 28°14′56”E, 25°24′44”S. Coordinates of (b): 27°58′09”E, 25°28′33”S.)

[1] https://www.openstreetmap.org/
[2] https://www.sygic.com/gps-navigation
[3] https://copilotgps.com/en-us/
[4] https://www.waze.com/en-GB/

1.3 Literature

1.3.1 Overview of Road Extraction

Formal road extraction has been well-developed in the literature. Mena [90] provides a comprehensive overview of road extraction techniques, classifying them according to their objective, methodology, and the type of sensor used to collect the data. Objectives include general road extraction and road network reconstruction [50, 49], segmentation [92], vectorisation [35, 88], optimisation, which includes deep learning techniques such as neural networks [13], and finally evaluation [51, 139]. Methodologies are classified as low-level and high-level. Low-level methods including among others segmentation and classification [7, 89], morphological methods [144], dynamic programming, and multi-scale and multi-resolution analyses [47]. High-level methods include knowledge- and logic-based techniques, fuzzy modeling and fuzzy logic, and spatial reasoning. Various sources of data mentioned are monochromatic imagery, colour and infrared information, multi- and hyper-spectral imagery, RADAR and LiDAR. Although the work of Mena [90] is older, it nonetheless provides an encompassing classification of the types of road extraction methods currently used. This is confirmed by Wang et al. [136], who provide the most recent review of road extraction approaches, classifying them by methodology. Therein, techniques are grouped into knowledge-based methods [53], active contour models [5, 8], dynamic programming and grouping [94], mathematical morphology methods [131, 155], which are often used in combination with other techniques [136], and classification methods. Classification methods may be supervised, such as the artificial neural network in [59] and the Markov Random Field approach of [143], or unsupervised, such as the probabilistic and graph theory method of [130]. Deep learning approaches have become popular for road extraction [153, 135, 142], including convolutional neural networks [23, 137], generative adversarial networks [150], and other techniques [73]. However, deep learning methods typically require vast datasets for model or network training. This makes such methods infeasible for the financially constrained situations of the developing world. Very high resolution satellite imagery requires the use of advanced technology, including satellite launching and monitoring capabilities, coordinated by a national space agency. According to the United Nations list of national space agencies, most governments in the developing world do not possess such infrastructure[5]. Satellite data must be bought from more developed countries. This is very costly. An algorithm aimed at solving problems in the developing world must also be implementable within the constraints of the developing world.

The above methods were all proposed to detect formal roads. The only approach in the literature developed for informal roads is that of [98], for the problem of detecting roads in informal settlements in São Paolo, Brazil. The lack of informal road extraction methods indicates a need for further research and improvements in this area. Several challenges were experienced in [98], such as misdetecting buildings as

[5] http://www.unoosa.org/oosa/en/ourwork/space-agencies.html

roads, difficulty in detecting roads with heterogeneous surface colour, and misclassification caused by the presence of cars and other objects on roads. There is therefore a need for further research in this area. In addition, the method in [98] requires the use of proprietary software, namely eCognition. It can therefore not be implemented freely or in financially constrained situations, which may occur in the developing world. In general, formal road extraction methods are not geared towards addressing the particular problems experienced by informal roads, which will be discussed below. The characteristics of paved urban roads are very different from those of informal roads. Rural roads may experience some similar challenges to informal roads in terms of their appearance [81], but occur solely in rural environments. Informal roads may occur in urban as well as rural environments. In [129], it was argued that only a flexible, robust formal road detection method could be capable of detecting informal roads. It was shown that the formal road extraction method of [71] was able to detect some informal roads, however, this method also relied on proprietary eCognition software. There currently exists no road extraction method that is capable of addressing the unique circumstances of informal roads, and is implementable using freely available software. The road extraction algorithm developed in this dissertation works towards satisfying this need.

1.3.2 Informal Road Characteristics

As highlighted in [98], informal roads possess unique challenges due to their informal nature. They also differ from formal roads in some key characteristics. Wang et al. [136] classify road properties into geometric, topological, functional, photometric and texture properties. Anything that describes road shape is a geometric property, such as length and width. Formal roads tend to be long linear objects of constant width, either straight lines or regular curved shapes. Informal roads, on the other hand, are of varying lengths and widths [98], often exhibiting sharp changes in road width. Topological features are concerned with the way roads are organised and connected. Formal roads that have been officially designed are generally laid out in regular patterns, and connect to other roads at planned junctions. Informal roads may exhibit irregular patterns [98], as their locations are influenced by convenience and environmental factors, rather than pre-planned design. In addition, they may exhibit discontinuities. Functional properties relate to road usage. Formal roads are in general designed to be suitable for automobile navigation, and must adhere to specific standards, such as those specified by the South African National Road Traffic Act[6]. Informal roads are created ad-hoc. Their structure changes with their function, e.g. narrow roads for small vehicles and foot traffic or broad roads for larger vehicles, and the properties may change abruptly along the same road, if different stretches of it are more or less frequently travelled. Photometric properties are concerned with road colour and reflectance, and are also

[6]South African Legal Information Institute, National Road Traffic Act, 1996 [No. 93 of 1996] - G 17603. Available online: http://www.saflii.org/za/legis/num act/nrta1996189/

relevant for the areas around roads. Formal roads are generally paved or clearly delineated, separated from their surroundings by boundaries or sidewalks. Informal roads on the other hand often do not have clearly visible boundaries, and may have the same colour as their surroundings, including bare soil areas and yards, or buildings with roofs made from local soil. In addition, informal roads may exhibit heterogeneous colouring, such as brighter, harder soil on parts of the road that are better-travelled, or darker soil on wet or muddy patches. Roads with heterogeneous surface colour were difficult to identify in [98]. Texture features are concerned with the homogeneity of regions in the image. In the context of informal settlements, land cover may be highly heterogeneous at the scale of roads, which [98] mentions as a particular challenge for informal road extraction.

Any approach for detecting informal roads must be able to rely on the characteristics possessed by informal roads and take their circumstances into account. Spectral information, such as colour and near infra-red reflectance, can be used to remove areas that are not bare soil, e.g. vegetation, buildings and shadows [71]. Building removal is critical since long, linear buildings may be geometrically similar to roads [81], and shadow removal has been specifically mentioned to be advantageous [70]. These may be better detected by incorporating near-infrared information, which provides an advantage over many recent techniques that have made use of optical data only [91, 141, 146]. In particular, the Normalised Difference Vegetation Index (NDVI) is useful in this regard, which is calculated using information from the red and near-infrared wavelengths. NDVI has been widely applied in studies related to land cover [34, 82, 125], and is capable of detecting bare soil [28]. NDVI will be used to isolate bare soil areas, a step that is also present in the informal road extraction method of [98], as well as the state-of-the-art multi-step formal road extraction method of [71].

Due to the colour confusion between informal roads and their surroundings, as well as heterogeneous colour of the road surface, spectral information alone is not sufficient. Uniform colour cannot be assumed, as in some formal road extraction methods [3]. Additionally incorporating geometric and structural properties decreases the risk of confusing roads with spectrally similar areas and objects [81, 74]. When considering geometric features, the unique challenges of informal road shapes must be allowed for, such as sudden changes in road widths, which places a constraint on the type of method used. For instance, centreline extraction has been used in several recent road detection methods [42, 80]. This fails when road widths change suddenly [74]. Since the boundaries of informal roads are often not clearly defined, methods relying on clearly visible road boundaries will fail, such as the multiscale method in [32], or approaches incorporating edge detection such as that of [81]. Indistinct road edges have been mentioned as a significant challenge [136]. Rather than attempting to consider the exact widths or edges of roads, a more flexible understanding of road shape is required. Since roads are linear structures, detecting and measuring linearity will provide a way of detecting the presence of roads, even where their boundaries are not clearly visible. Mathematical morphology considers the shape of objects [116], and methods relying

on mathematical morphology have been successfully used for road extraction [131, 144]. Traditional mathematical morphology makes use of user-defined shapes called structuring elements (SE). Specifying linear SEs will allow the algorithm to identify straight linear objects in the image. Multiple road extraction methods have focused on detecting straight-line road segments [7, 120]. However, informal roads are not necessarily straight, but may be visually erratic or winding, as determined by the navigational needs and environmental constraints from which the roads have arisen. Detecting the linearity of roads that are not straight has proved a challenge in the past [7]. Measuring the compactness and elongation of objects [19] allows for detecting curvilinear objects, that are linear but not straight. Li et al. [71] use compactness and elongation to analyse the shape of image objects, to determine if they are road segments.

Image objects appear differently at different scales [16]. The scale at which an image is analysed determines what image objects will be identified [132]. At a large scale, small objects such as cars or individual trees may be identified, while at a very small scale, larger regions will be identified, such as built-up or open areas. Since informal roads differ in length and width, and roads of varying sizes may appear in the same area, searching for informal road segments at a single scale is not suitable. Considering various scales for road detection has been done in the case of formal roads, as in [71], [132] and [131], which consider objects of various sizes by representing image segments in a tree-like structure. Herein, the Discrete Pulse Transform (DPT) is used to obtain image objects at various scales.

1.3.3 Discrete Pulse Transform

The DPT provides a multi-scale decomposition of a greyscale image by recursive application of the LULU operators [109]. It identifies local maximum and minimum sets, known as pulses, at various scales. The DPT can be implemented using the Roadmaker's Pavage (RP). The RP represents the DPT pulses in a tree-like structure that can be efficiently traversed to find and analyse image objects of any scale or size.

Figures 1.3-1.7 provide an overview of the working of the DPT and illustrate the way in which pulses are stored in a tree-like structure. This example is provided in complete detail in Chapter 2. Consider the greyscale image in Figure 1.3(a). The numbers on the pixels are the pixel values, where lower numbers represent darker colours. The red arrows in (b) illustrate which pixels are considered to be adjacent to each other. The image is represented in graph form in (c). The left hand graph is called the work graph, which contains the structure of the image. Each node of the graph represents a pixel in the image. The value on each node is called its height, and is equal to the value of the corresponding pixel. The superscript above each node is called its size, and is equal to the number of connected equal-valued pixels corresponding to that node. In the initialisation step of the DPT, when the original image is initially represented as a graph, each node corresponds to exactly one pixel in the image, therefore the size of every node is one. The extra zero-valued node with an infinite size represents the image edge, and is called the

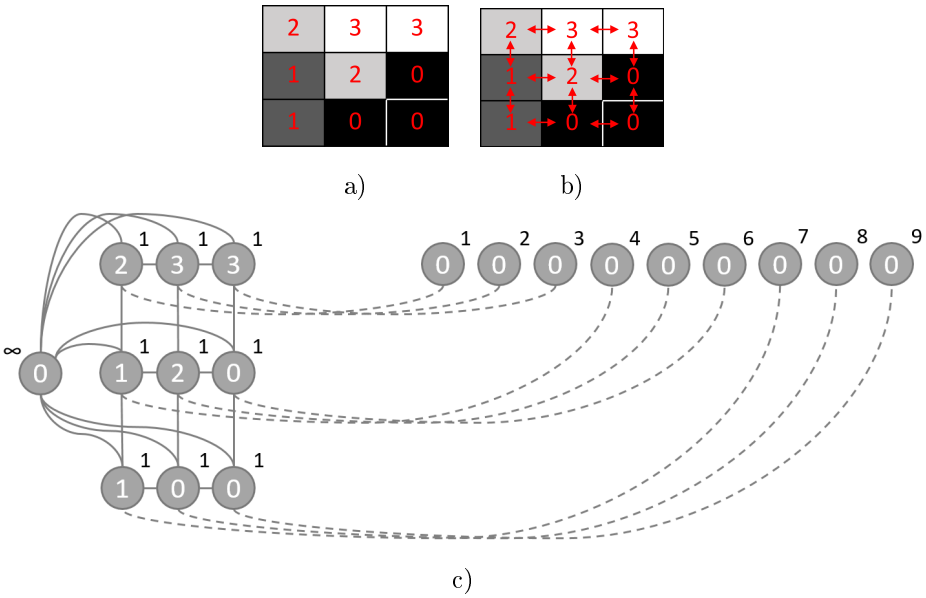


Figure 1.3: Example greyscale image with pixel values (a), connectivity (b) and work and pulse graphs (c). The graph on the left is the work graph, while the graph on the right is the pulse graph.

edge node. The nodes are connected via edges (solid lines), in accordance with the adjacency of pixels in the image. Any pixels on the edge of the image are additionally connected to the edge node.

The graph on the right hand side is called the pulse graph. This graph is modified during the working of the DPT, and will store the extracted image pulses at various scales. In the initialisation step, each node in the work graph has a corresponding node in the pulse graph. The height of each initial pulse node is zero, as no pulses have yet been extracted. The size of each initial pulse node is equal to the position of the corresponding node in the work graph, according to a one-dimensional vectorisation of the image. Each node in the work graph is connected to its corresponding node in the pulse graph via a virtual edge (dotted line).

Figure 1.4 shows the first step of the Roadmaker's Pavage, which merges adjacent pixels of the same value into coherent image objects. The changes are reflected in the work graph on the left hand side in (b), where each node now refers to an image object. The number of pixels in that object are represented by the node superscript. The connections to the pulse graph on the right hand side have also been updated.

Figures 1.5 and 1.6 show how the DPT is applied via the Roadmaker's Pavage. The DPT identifies and removes local maximum and minimum sets of increasing size. The smallest local maximum or minimum set is identified in Figure 1.5. In this case it is the local maximum set (bump) of white pixels (pixel value 3), shown on the image in (a) and highlighted in the work graph in (b). Figure 1.6 illustrates the process of the removal of this bump. It is smoothed on the image, as shown in (a), where those pixels take on

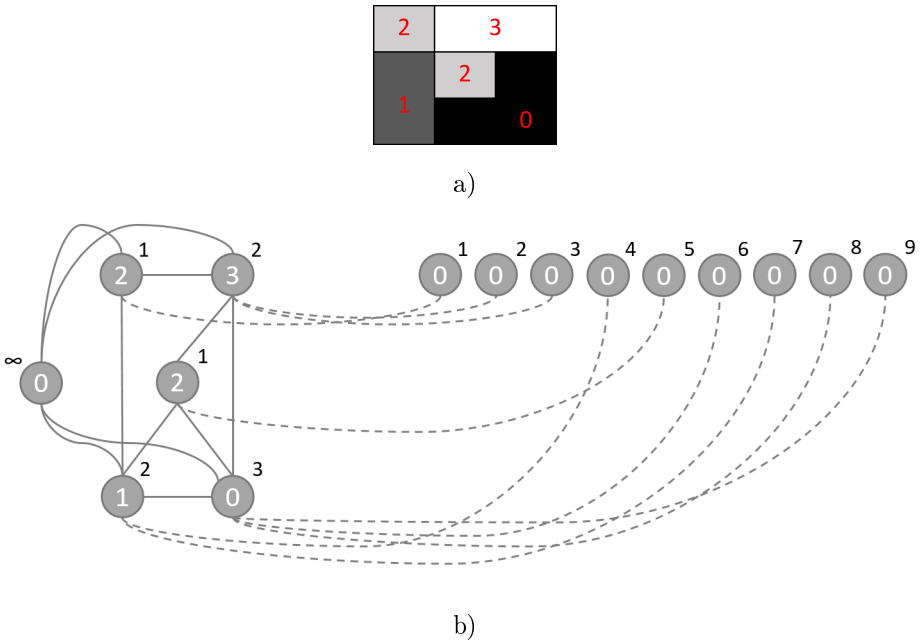


Figure 1.4: The initial merging step of the Roadmaker's Pavage, demonstrated on the image in Figure 1.3. a) Adjacent pixels of the same value are grouped in the image. b) The resulting changes in the work and pulse graph, where the work graph is on the left and the pulse graph is on the right.

the value of the adjacent pixels with the nearest value, namely the light grey pixels (pixel value 2). Since the value difference between the white pixels and the light grey pixels is 1, a pulse of size 2 and value 1 is extracted from the image, as shown in (b). The changes to the image are reflected in the work and pulse graphs in (c). The new light grey image object now includes 4 pixels, since all adjacent light grey pixels are considered as one image object. Therefore the superscript of the node with value 2 in the work graph increases to 4. The connections to the pulse graph are also updated. In the pulse graph, the new red node with a value of 1 represents the local maximum set that was extracted. The superscript of 2 indicates that two pixels are contained in this pulse.

The DPT progressively removes local maximum and minimum sets of increasing size, and pulses are extracted accordingly. Each time a pulse is extracted, the changes are reflected in the work graph, and a node is added to the pulse graph to store this pulse. Figure 1.7 (a)-(d) show the stepwise changes in the image, with the corresponding pulses extracted in (e)-(g). The final work and pulse graphs are shown in (h). The final work graph only contains the single resulting completely smoothed image, where all pixels have a value of 0. The pulse graph records the full process followed by the DPT to identify local maxima and minima, and stores sizes and values of the pulses extracted at all scales. A multi-scale decomposition of the image is obtained and stored in a tree-like structure. The pulse graph can be traversed after completion of the DPT to find pulses of a certain scale.

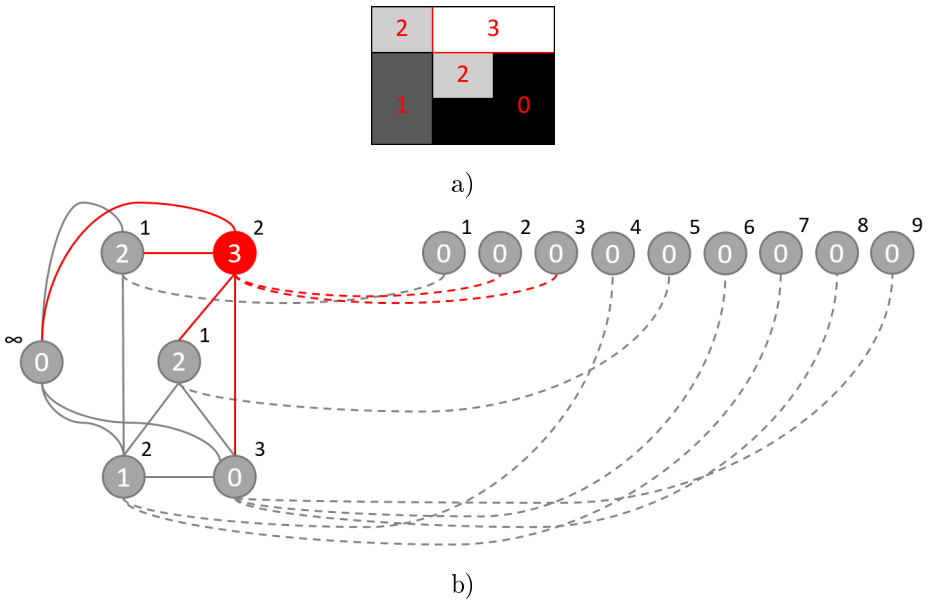


Figure 1.5: Identification of a bump of height 1 and size 2. a) The image with the local maximum set highlighted in red. b) The work and pulse graphs. The node corresponding to the bump is highlighted in red, along with the edges connecting to it.

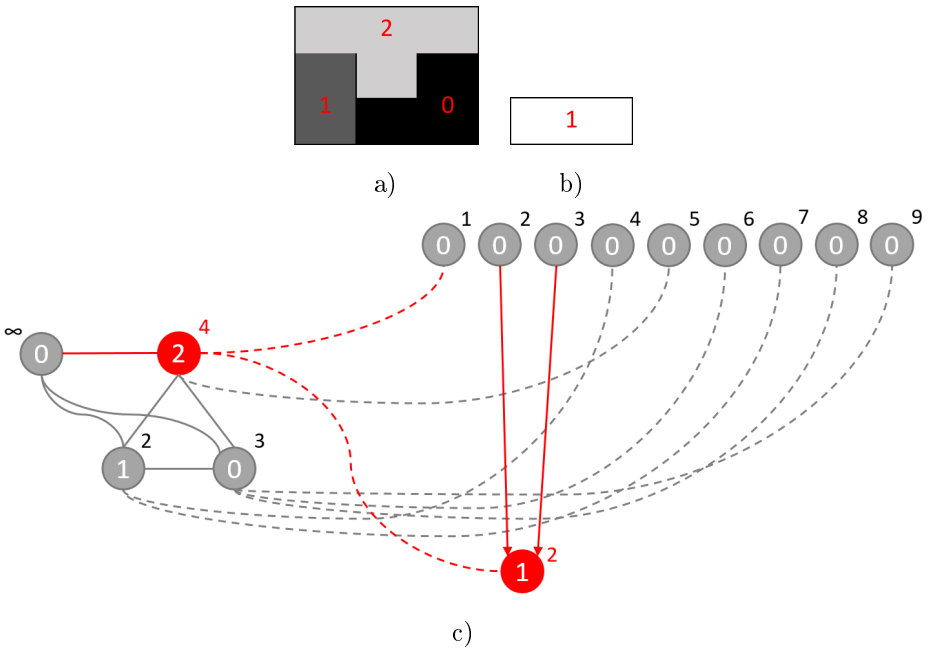


Figure 1.6: Continuation of Figure 1.5. Removal of a bump and extraction of a pulse of height 1 and size 2. a) The image after bump removal. b) The extracted pulse. c) The work and pulse graphs after bump removal.

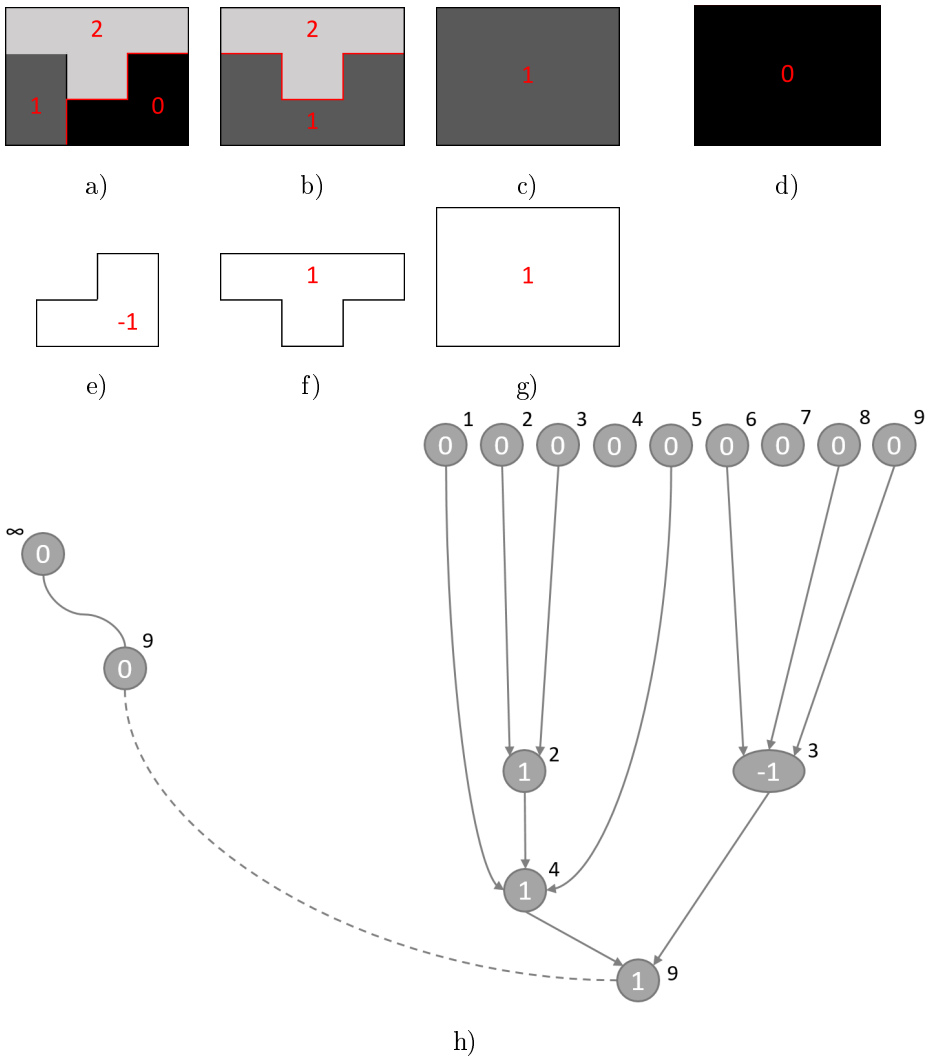


Figure 1.7: a)-d) The changes in the image as local maxima and minima are smoothed. e)-g) The pulses extracted corresponding to the local maximum/minimum removed in a)-c) respectively. h) The final work and pulse graphs. The pulse graph provides a complete representation of the image pulses at various scales.

It is straightforward to extend such a search to find pulses of a certain shape, in particular as measured by compactness and elongation. The extracted pulses stored in the pulse graph can be analysed to determine their compactness and elongation after the DPT has been executed. This is efficient in terms of both computation time and simplicity. Since the DPT operates on greyscale images, colour and reflectance information can be utilised in a separate step to isolate bare soil areas, which may then be converted to greyscale for further analysis by the DPT. In this way, colour, near-infrared, size and shape information available in satellite imagery can be combined to effectively detect unpaved informal roads of various sizes, exhibiting heterogeneous colour and irregular shapes. The only requirement is that the roads should be linear, which is implied by definition.

1.3.4 Uncertainty Quantification

Compactness and elongation may additionally be used to provide a measure of certainty related to the extracted objects [71]. The more linear the object, as determined by these measures, the more certain it is to truly be a road. In this way, an algorithm measuring compactness and elongation may easily be extended to provide an internal measure of reliability of the results. Many studies have focused on assessing the accuracy of the results of road extraction algorithms [26, 139, 48], and consider comprehensive sources of uncertainty, including those inherent to satellite imagery. This includes mixed pixels, which occur when more than one land cover type is present in the same pixel as represented by the satellite image. However, such uncertainty analyses are usually conducted in separate studies and not included as part of the road extraction process. Methods including uncertainty measures as part of the extraction algorithm output are rare in the literature [51].

1.3.5 Accuracy Assessment

In order to quantify the degree of accuracy of the algorithm, extracted roads are compared to reference data. Since informal roads are not captured in official databases, these are obtained via manual digitisation. The accuracy evaluation step is not a part of the extraction process, but is done to assess algorithm performance. There are various ways in which road extraction accuracy can be evaluated. Per-pixel accuracy metrics compare every pixel of the extracted roads to the corresponding pixel in the reference dataset. The higher the percentage of matching pixels, the higher the accuracy. The number of true positive pixels, false positives and false negatives are considered. True positives are pixels that were identified as belonging to roads in the extraction, and also corresponded to roads in the reference. False positives were extracted as forming part of roads, but did not match roads in the reference. False negatives are pixels that were not identified as roads in the extraction, but corresponded to road pixels in the reference. This is a conceptually and computationally simple approach. However, it is sensitive to errors, especially

on the edges of roads. These measures are included herein, but they may not provide comprehensive information of how well the presence of roads has been detected. The work of [48] is seminal in assessing the accuracy of automatic road extraction. In this approach, the length of roads in the extraction is compared with the length of roads in the reference. This approach has been widely used throughout the literature, either directly or as the basis of other evaluation methods [25, 121, 119]. For example, the above-mentioned method of [26] was adapted from [48], considering edges instead of centrelines. However, since informal roads may not have clearly visible boundaries, an edge-based approach is not suitable. In order to minimise the effects of variation at the edges and determine whether or not the presence of roads is detected accurately, a comparison is done herein between the extracted road centrelines and the full reference dataset, as well as a comparison between the full extracted dataset and the reference road centrelines. Metrics including completeness and correctness are calculated based on the metrics suggested by [48]. In order to determine how closely the centrelines of the extraction and reference are positionally aligned, Pratt's Figure of Merit (PFOM) is calculated on the road centrelines. This is a versatile and well-used accuracy measure [46, 45, 60]. Although it was originally proposed to assess edge detection, it has been applied for other purposes, including video segmentation [45]. PFOM is based on the distances between edges in an image. Applying PFOM to road centrelines provides a measure based on the distance between the extracted and reference centres.

1.4 Methodology

The multi-step semi-automatic algorithm proceeds as follows. The input is a satellite image with colour information in red, green and blue, as well as near-infrared information. Vegetation, buildings, shadows and paved surfaces are removed based on the NDVI. Default parameters are provided for this step which specify the NDVI values to identify bare soil. This leaves a binary (black and white) image containing white pixels corresponding to bare soil regions that are potential roads, and black pixels corresponding to filtered-out non-road areas. Since the DPT operates on greyscale information, the original image is converted to greyscale. Noise is removed via adaptive median smoothing, with a default maximum window size of 9×9 pixels. The greyscale image is then converted to binary to remove dark areas that are not likely to be roads. Default parameters for the binarisation are provided. The two binary images are multiplied, such that white pixels indicate potential roads, and black pixels indicate areas that are not considered further in the analysis. This binary image is multiplied with the smoothed greyscale image, which results in a greyscale image containing only those areas which are potential roads.

The DPT pulses are extracted from this image. This identifies image objects at various scales. These objects are filtered by size, elongation and compactness to identify linear pulses with a road-like shape. Parameters related to the size, elongation and compactness must be specified. User input is required for

the size parameters, as these will vary depending on image characteristics, such as spatial image resolution and the sizes of objects and roads in the image. Recommended defaults are provided for the compactness and elongation parameters, obtained from the literature [71].

Small holes in the objects are removed, and jagged edges are smoothed via adaptive morphological dilation, which further enhances the linear structure of the objects. Recommended structuring elements are provided for this step. The final road objects are once more filtered by elongation and compactness to ensure their linearity, using the same parameters as before, as these parameter values are specified in the literature. A certainty measure is calculated for each extracted road object, based on its elongation and compactness. The ultimate output of the algorithm is the final road objects along with their associated certainty measures. To assess the extraction accuracy, the extracted roads are compared with reference data. The reference data is obtained via manual digitisation, since informal roads are not generally available in official records.

A graphical representation of the process is provided by the flowcharts in Figures 3.1 and 3.2. These explain the process in detail, and are therefore best understood in the context of Chapter 3, which develops the algorithm in a stepwise manner.

1.5 Research Objectives

This dissertation provides a spatial statistical solution to the problem of detecting informal roads from remote sensing images, using the digital pulse transform. Applying such a solution is a critical step for ensuring sustainable city growth, an issue that is becoming increasingly important in the rapidly developing world. The solution developed herein is implementable using free open-source software, and can therefore be used in financially challenged situations. Additionally, the solution provides statistical measures of certainty related to the identified road objects, rather than hard binary classification. Methods for evaluating the accuracy of the extraction are also provided. The objectives and contributions are summarised below:

1. To develop a spatial statistical solution to detect the presence of informal roads in remote sensing images;
2. To provide associated uncertainty measures to give an indication of the reliability of results;
3. To achieve the above objectives using open-source software that is freely available online;
4. To provide assessment metrics to evaluate various aspects of the accuracy of the extraction;
5. To demonstrate the working of the solution as well as accuracy assessment at the hand of real-world

examples in South Africa.

6. To provide a thorough discussion of obtained results and propose avenues for future research.

1.6 Conclusion

This research provides a Discrete Pulse Transform solution to the problem of informal road extraction in remote sensing images. The dissertation proceeds as follows. Chapter 2 explains and discusses the theoretical background of image processing, as well as theory required for each step of the proposed solution, including the road extraction algorithm. Extensive literature related to the theory is provided as required. Chapter 3 develops the spatial statistical solution, describing each step in detail, and explains the methods of accuracy assessment employed. In Chapter 4, the solution is applied to four study areas in Gauteng and North West Province, South Africa. The accuracy is evaluated for each area. A thorough discussion is provided on sources of uncertainty for the images and algorithm used, and general challenges for informal road extraction are highlighted. Chapter 5 concludes the dissertation, highlighting the contributions made therein.

Chapter 2

Background Theory

This chapter develops the theoretical background for each of the elements of the road extraction algorithm. A variety of mathematical and statistical techniques are incorporated into the algorithm, leading to the necessity of discussing theory from various backgrounds. This chapter is divided into sections, each presenting theory on an aspect of the algorithm. Sections 2.1 and 2.2 present fundamental ideas regarding the mathematical representation of images. Sections 2.3 to 2.10 discusses the relevant theory corresponding to individual steps of the extraction algorithm, while Section 2.11 gives the theoretical background of the accuracy assessment techniques.

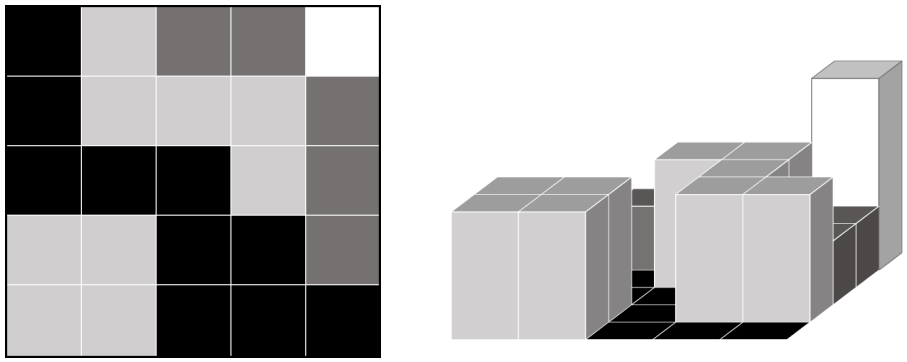
2.1 Image Representation

This section presents a mathematical definition of an image. An image is a grid in two dimensions where each cell of the grid is called a pixel [122]. Various values can be associated with each pixel, e.g. hue, intensity, saturation, or the value of a wavelength on the electromagnetic spectrum, such as red, green, blue, near infrared, or other values. In this way, an image can be represented as an object in $\mathbb{Z}^{n+2}$, where n is the dimensionality of the values associated with each pixel. In the case of binary and greyscale images, a scalar value is associated with each pixel. Therefore, binary and greyscale images are objects in $\mathbb{Z}^3$. Formally, a binary or greyscale image f may be defined as follows [116, 122]:

Definition 1. *An image f is a positive upper semi-continuous function bounded by a value $t \in \mathbb{R}$ which maps some subset $\mathcal{D}_f$ of $\mathbb{Z}^2$ onto a subset of consecutive nonnegative integers starting at zero, namely:*

$$f : \mathcal{D}_f \subset \mathbb{Z}^2 \rightarrow \{0, 1, ..., t\} \subset \mathbb{N}_0,$$

where $\mathcal{D}_f$ is a rectangular frame known as the definition domain of f.



a) b)

Figure 2.1: $3D$ representation of a greyscale image. a) The original image. b) The greyscale representation.

Figure 2.1 demonstrates how binary and greyscale images may be represented in three dimensions, where the height is given by the value associated with the pixel, e.g. grey intensity.

In a colour image, the value associated with each pixel is a 3-dimensional vector. This vector can contain various sets of values, depending on the image representation, e.g. hue-saturation-value (HSI) or red-green-blue (RGB). Figure 2.2(a) shows a colour image with its decomposition into channels. Its RGB decomposition is shown in (c)-(e), where (c) presents the red channel, (d) the blue and (e) the green. Each colour is referred to as a band or channel. A colour image can be defined as follows:

Definition 2. *A colour image g is a positive upper semi-continuous function bounded by a value $t \in \mathbb{R}$ which maps some subset $\mathcal{D}_g$ of $\mathbb{Z}^2$ onto a subset of $\mathbb{N}_0^d$, a subset of consecutive nonnegative integers starting at zero, namely:*

$$g : \mathcal{D}_g \subset \mathbb{Z}^2 \to \{0, 1, ..., t\}^d \subset \mathbb{N}_0,$$

where $\mathcal{D}_f$ is a rectangular frame known as the definition domain of f.

In the case of RGB images, $t = 255$ and $d = 3$. Images may also store other electromagnetic information. Multispectral images have multiple bands, which may include colour, as well as the values of other wavelengths, such as infrared or near-infrared. Figure 2.2(b) shows the near-infrared band captured by the satellite for the multispectral image in (a). Lighter pixel values indicate a higher amount of near-infrared output by the real-world object corresponding to the pixel, e.g. the white pixels are roofs that reflect direct sunlight.

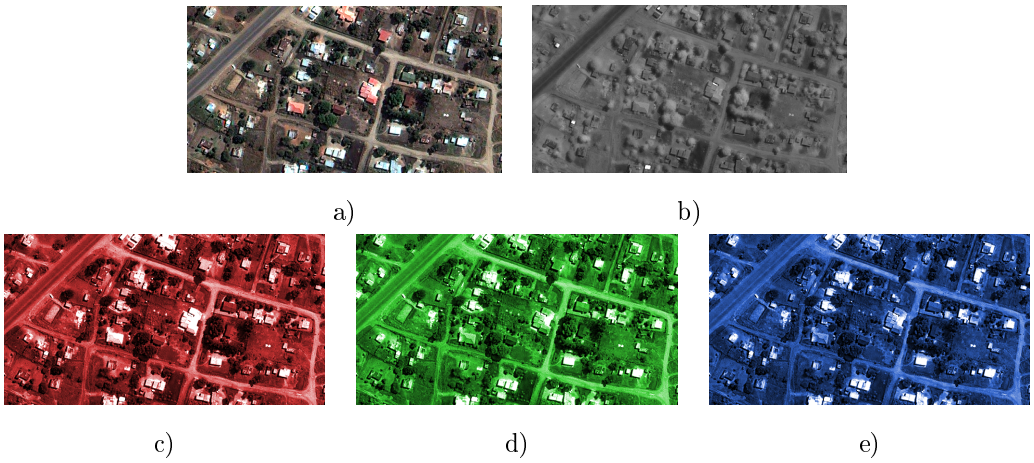


Figure 2.2: Decomposition of a multispectral image. a) The visible components of the multispectral image. b) The near-infrared band. c) The red band. d) The blue band. e) The green band.

2.2 Pixel Connectivity

Much of image analysis theory relies on pixels being ordered and connected in some way [31, 58, 75, 96]. The LULU operators discussed in Section 2.8.1 operate on connected sets [36]. It is therefore necessary to define the order and connectivity of pixels in a two-dimensional grid. Various orderings exist which specify which pixel is before or after another [12]. Pixel connectivity specifies which pixels can be said to be adjacent (neighbours) or *connected* to each other.

Figure 2.3 demonstrates a 2-dimensional connectivity. This specifies which pixel is a neighbour of another pixel. This connectivity is typically chosen for the context of the LULU operators and the Discrete Pulse Transform [36], discussed in Section 2.8. The figure also shows two popular ways of defining connectivity. Typically, the origin $(1, 1)$ is the top left hand pixel, with the order ascending to the right and downwards. The $(i, j)^{th}$ pixel is therefore found at i pixels down and j pixels to the right of the origin, $i, j \in \mathbb{N}_0$. This follows the traditional matrix notation.

Figure 2.3 a) shows 4-connectivity, where the pixels directly above, below and beside a pixel $\mathbf{x}$ are considered its neighbours. If $\mathbf{x}$ is the $(i, j)^{th}$ pixel in the image, the set of its neighbours is

$$\{(i, j-1), (i, j+1), (i-1, j), (i+1, j)\}.$$

Figure 2.3 b) shows 8-connectivity, where the pixels diagonally adjacent to $\mathbf{x}$ are additionally considered neighbours. If $\mathbf{x}$ is the $(i, j)^{th}$ pixel in the image, the set of its neighbours is

$$\{(i-1, j-1), (i-1, j), (i-1, j+1)(i, j-1), (i, j+1), (i+1, j-1), (i+1, j), (i+1, j+1)\}.$$

Figure 2.4 shows the graph representation of the image in Figure 2.3, with 4-connectivity in (a) and 8-

connectivity in (b). The pixels are represented as nodes of the graph, and are connected via graph edges. The red edges show how the red pixel is connected to its neighbours. The graph representation is relevant for the Roadmaker's Pavage implementation of the DPT, presented in Section 2.8.3.

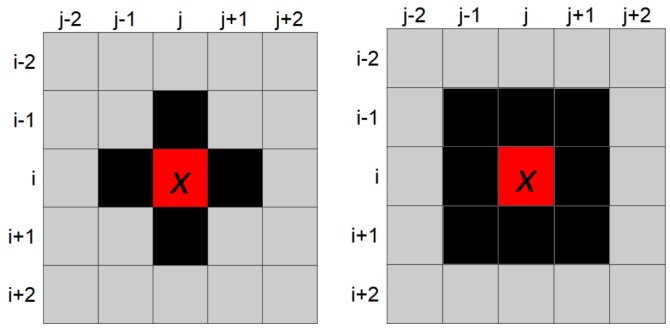


Figure 2.3: Well-known types of pixel connectivity. Black pixels are considered neighbours of the red pixel **x**. a) 4-Connectivity. b) 8-Connectivity.

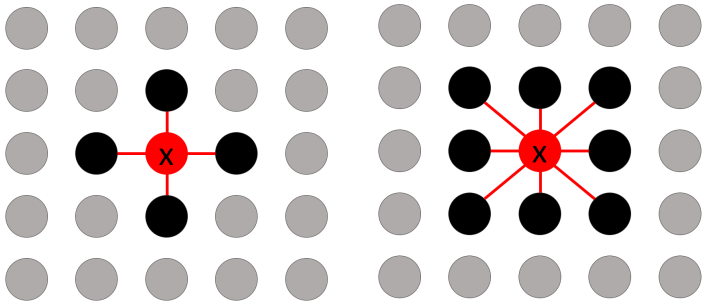


Figure 2.4: Well-known types of pixel connectivity. Black pixels are considered neighbours of the red pixel **x**, connected by edges. a) 4-Connectivity. b) 8-Connectivity.

2.3 Image Filtering

Images are often corrupted by image noise [78]. This can be defined as an undesired component of the image, and different types of noise can occur, e.g. Gaussian noise, speckle noise and salt-and-pepper noise [17]. Noise degrades image quality [54] and can adversely affect the results of image processing techniques, including feature recognition [79].

Various image filtering algorithms have been developed to reduce noise, including mean and median filters [54]. Median filters work by moving a window of a specified size across the image, and replacing each pixel

value with the median value of the window. These filters tend to smoothe noise effectively [79]. However, they may also blur features when the window size is too large, or smoothe inaccurately when the window size is too small [78]. Another disadvantage of median filters is that the filtering operation is applied indiscriminately to all pixels. They do not provide any way of distinguishing between noise and non-noise [114]. Many adaptive median filters have been proposed to mitigate or circumvent these disadvantages [78, 79, 114].

Herein we utilise an adaptive median filter. The algorithm specified is detailed in [101], and is also provided below. A maximum window size of 9 pixels is used. The filter adapts the window size and compares each pixel with its neighbours to determine whether or not the pixel constitutes noise. This filter preserves more information than the standard median filter [101]. The idea of the algorithm is as follows. Each pixel is considered along with its neighbours in a 3×3 window. If the median grey value within this window is strictly greater than the minimum and strictly less than the maximum, the grey value of the centre pixel is considered for smoothing. If the grey value of the centre pixel is strictly greater than the minimum and strictly less than the maximum, its value is not changed by the filter. Otherwise, it assumes the median grey value. In this way it is smoothed. If the median grey value is not strictly between the minimum and maximum, the window size is increased as necessary, either until the median is strictly between the minimum and maximum, or until the maximum window size (9×9 in this case) is reached. If this happens, the grey value of the centre pixel is not changed. In short, the centre value is only smoothed if both this value and the median grey value in the window are strictly between the minimum and maximum grey values in the window.

Peng's Adaptive Median Algorithm

```
 Level I
1.    If Wmin < Wmed < Wmax:  Go to Level II
2.    Else:  Increase Winsize
3.    If the Winsize < MaxWinsize Repeat Level I
4.    Else:  Wcenter(new) = Wcenter(old)
Level II
1.    If Wmin < Wcenter(old) < Wmax:  Wcenter(new) = Wcenter(old)
2.    Else:  Wcenter(new) = Wmed

Where:
Winsize is the window size,
MaxWinsize is the maximum window size,
Wmin is the minimum grey value in the window,
```

```
Wmax is the maximum grey value in the window,
Wmed is the median grey value in the window,
Wcenter(old) is the original grey value in the center of the window,
Wcenter(new) is the new grey value in the center of the window.
```

Figure 2.5 gives an example of the standard median filter and Peng's adaptive median filter applied to a greyscale image (a). It is clear that the adaptive median filter (b) preserves far more information than the standard median filter (c). The image in (c) appears blurred, especially around the edges of objects such as roads and objects. The image in (b) appears sharp and clear. Only noise has been reduced. A more in-depth comparison between the original image and the adaptive median filtered image is provided in Chapter 3, Section 3.2.2.

a)

b)

c)

Figure 2.5: A comparison of the results of the adaptive median filter and a non-adaptive median filter. a) Original image. b) Adaptive median filter (9×9 maximum window size). c) Median filter (5×5 window size).

2.4 The Normalised Difference Vegetation Index (NDVI)

The Normalised Difference Vegetation Index (NDVI) is a reflectance index used to study vegetation, based on the red and near-infrared channels of an image [40]. It is calculated as follows:

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}},$$

where NIR and Red refer to the reflectance values of the near-infrared and red channels, respectively. Red light refers to light with a wavelength in the range [0.625μm, 0.74μ] [22]. The precise values for near-infrared wavelengths depends on the classification scheme used to subdivide infrared into near, midrange, and far. Possible near-infrared ranges include [0.75μm, 1.4μm] [24], [0.7μm, 5μm] [124] or [0.78μm, 3μm] [38]. Regardless of the exact limits of red and NIR, NDVI ranges between -1 and 1, where higher values mean more dense, greener vegetation [28]. NDVI has been widely used to analyse vegetation [40, 57], and is also capable of detecting other types of land cover, such as water and bare soil. NDVI is used in the study of environmental quality [39], drought monitoring [103], crop yield assessment [107], and change detection, including vegetation and land cover change [41, 82]. Herein it is used to differentiate between land cover types. In particular, impervious surfaces such as rock and paving, as well as water give low NDVI values, while green vegetation gives high NDVI values. The focus will be on identifying bare soil and dusty areas that might contain roads. Relevant NDVI values are further discussed in Section 3.2.1.

2.5 Image Thresholding

Image thresholding or binarisation is the process of converting a greyscale image to a binary or black-and-white image, by separating an image into dark and light pixels [20]. Sezgin and Sankur [118] provide an overview of a wide array of techniques. Image thresholding works by applying some kind of threshold to the image, either locally or globally. Pixels with a grey value below this threshold are classified as dark, while those above the threshold are classified as light. The effectiveness of the thresholding results have a significant impact on any subsequent image processing and analysis [118].

In global thresholding, a single threshold is applied to the entire image. Otsu's binarisation is an example of a well-known global thresholding method [100], and is popular in remote sensing and satellite data applications [72, 106, 87]. However, many thresholding algorithms, including Otsu's binarisation, assume bi- or multi-modal intensity histograms, and have difficulty processing images with unimodal intensity histograms [110]. Furthermore, global thresholding methods assume that a single threshold level is appropriate for the whole image, which may not be the case.

Adaptive thresholding accounts for varying local contrasts between dark and light in different parts of the image [20], for instance as caused by shadows. In adaptive thresholding, a threshold is calculated for each pixel based on local statistics, such as the mean and variance of the grey values in a specified neighbourhood of the pixel [118].

Two simple adaptive thresholding techniques in OpenCV are available for implementation in Python, namely mean thresholding and Gaussian thresholding. In mean thresholding, the threshold value is the mean of the grey values of the neighbourhood of the pixel, where the neighbourhood is of a specified

a)

c)

d)

Figure 2.6: Similarity of the results of adaptive Gaussian and adaptive mean thresholding. a) The greyscale image to be thresholded. b) Mean thresholding with block size 71. b) Gaussian thresholding with block size 151.

size. In Gaussian thresholding, the threshold value for a pixel is a weighted sum of the grey values of the neighbouring pixels, in a neighbourhood of the specified size. In both methods, the threshold may be adapted by subtracting a constant value C from the mean (or weighted mean).

Depending on the neighbourhood size and constant value selected, the mean and Gaussian adaptive thresholding methods give very similar results. Figure 2.6 gives an example. The greyscale image in (a) is thresholded via mean thresholding with a block size of 71 in (b). These results are similar to those in (b), obtained via Gaussian thresholding with a block size of 151.

For theoretical transparency and simplicity, the mean adaptive threshold is used. Suitable values for the neighbourhood size and the constant value are determined in Section 3.2.3 for the application at hand.

2.6 Mathematical Morphology

Mathematical morphology is a rigorous mathematical theory for the analysis of the shape of image objects, first proposed by Matheron [86]. It is an extension of Minkowski's set theory [116, 102] and is based on set theory and lattice algebra [117]. It has been used in a wide range of applications, including

edge detection, noise suppression, image segmentation, texture analysis and shape smoothing [9, 84]. In particular, it is used for detecting, extracting and manipulating shapes in images [61]. Mathematical morphology has become popular in image analysis due both to its mathematical foundation, and its exploitation of the spatial relationships of pixels [9]. Serra [116] and Soille [122] are two of the fundamental developmental works on mathematical morphology. Therein, the theory for binary images was developed, and extended for greyscale images. Extending the theory to colour images is not straightforward, however, since mathematical morphology relies on the connectivity of pixel values for obtaining extrema. The mathematical operations used in this project will be applied to binary images. Therefore, only the relevant theory for binary images will be discussed.

2.6.1 Fundamental Concepts and Operators

Mathematical morphology relies on the concept of a structuring element (SE). An SE is a discrete pixel set by which the image is transformed via a morphological operator [102]. The two fundamental morphological operators are erosion and dilation, on which all others are based [123]. The definition of Minkowski addition is required to define erosion and dilation.

Definition 3. *The Minkowski addition of a set A by a set $\{b\}$ is denoted by A_b and is defined as follows:*

$$A_b = A \oplus \{b\} = \{a + b | A \in A\}.$$

Definition 4. *The dilation of a set X by a set B is given by the Minkowski addition of X and $\check{B}$ where $\check{B} = \bigcup \lim_{b \in B} \{-b\}$ is called the transpose of set B, namely:*

$$\delta_B(X) = X \oplus \check{B} = \{x | B_x \cap X \neq \emptyset\}, \text{ where } \emptyset \text{ denotes the empty set.}$$

Definition 5. *The erosion of a set X by a set B is given by the Minkowski subtraction of $\check{B}$ from X, namely:*

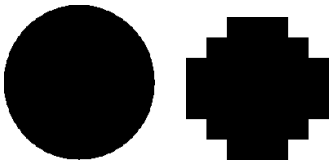
$$\varepsilon_B(X) = X \ominus \check{B} = \{x | B_x \subset X\}.$$

Definition 6. *The set B is called the structuring element (SE).*

These operators are defined within the context of set theory for continuous spaces. The definition of the erosion and dilation operators must be adapted for application to images on discrete domains. Figure 2.7 illustrates the difference between a circle represented on a continuous space in (a), versus the representation of a circle on a discrete (image) space in (b). The continuous representation has a smooth edge, while the discrete representation exhibits jagged edges. The definition below adapts dilation and erosion for images.

Definition 7. *The dilation of an image f with domain $X \subset \mathbb{N}$ by a set $B \subset \mathbb{N}$ is given by:*

$$\delta_B(f) = \bigvee_{b \in B} f_{-b} = \max_{b \in B} f(x + b).$$



a) b)

Figure 2.7: Continuous and discrete representations of a circle. a) A circle defined on a continuous space. b) A representation on a discrete space.

The erosion of f by B is given by:

$$\varepsilon_B(f) = \bigwedge_{b \in B} f_{-b} = \min_{b \in B} f(x + b).$$

Figure 2.8 shows the effects of erosion using various linear SEs of size 4, at different orientations. From the figure, it is clear how the SEs in b)-e) influence the erosion results, shown in f)-i). Structures in the image that are of a similar orientation to the SEs are preserved, while others are partially destroyed or completely removed. Note in particular the difference between the long southwest-northeast lines in h) and i), as eroded by the SEs in d) and e). The area circled in red shows how the choice of SE will determine whether image objects are connected or not. In f)-h), the thin line (road) is connected to the object to the larger object (yard), whereas in i), they are not connected. Such a connection between two separate image objects is known as leakage [126].

In general, dilation does not undo the effects of erosion, nor vice versa, so that $[\delta \circ \varepsilon](X) \neq [\varepsilon \circ \delta](X)$. In fact, the composition of dilation and erosion leads to the existence of new operators. Two such operators are the opening and the closing.

Definition 8. *The opening of a set X by an SE B defined as an erosion by B followed by a dilation by $\check{B}$, namely:*

$$\gamma_B(X) = [\delta_{\check{B}} \circ \varepsilon_B](X).$$

Definition 9. *The closing of a set X by an SE B defined as a dilation by B followed by an erosion by $\check{B}$, namely:*

$$\phi_B(X) = [\varepsilon_{\check{B}} \circ \delta_B](X).$$

For images, the definitions are as follows:

Definition 10. *The opening of an image f by an SE B is defined as an erosion by B followed by a dilation by $\check{B}$, namely:*

$$\gamma_B(f) = [\delta_{\check{f}} \circ \varepsilon_B](f).$$

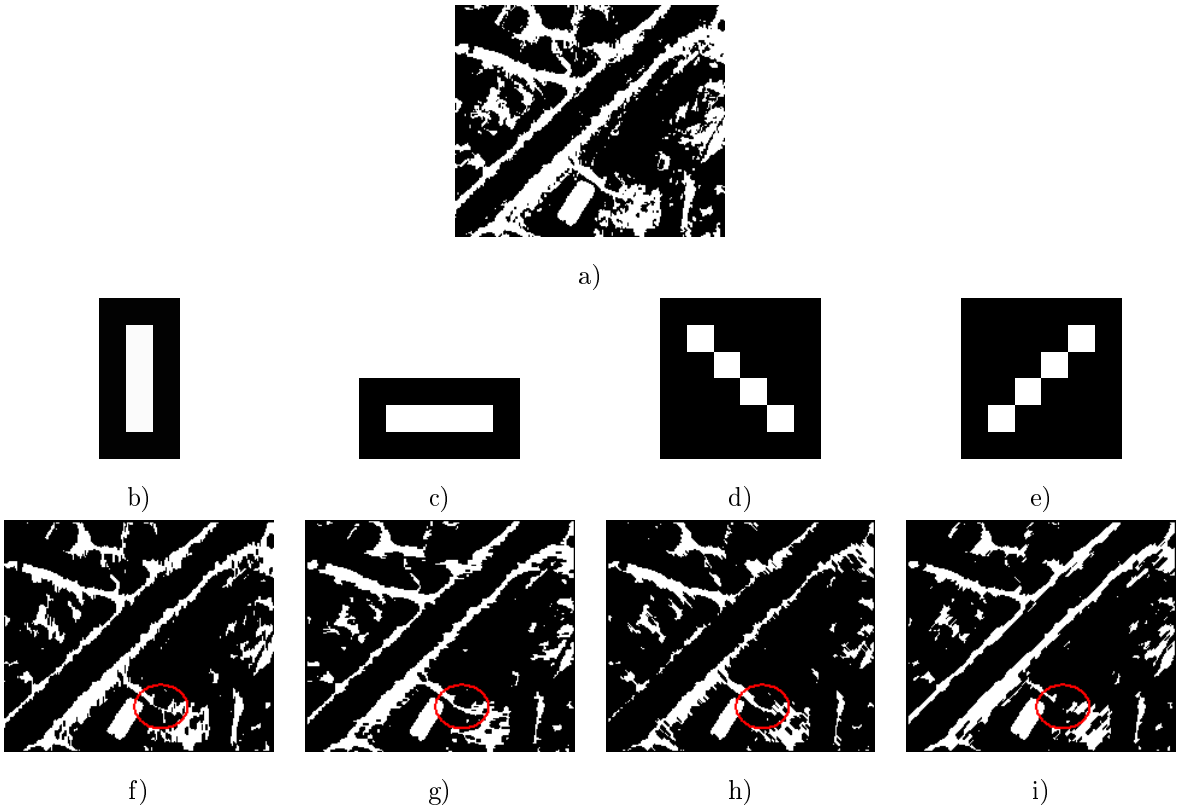


Figure 2.8: Erosions of a binary image shown in a), by linear SEs of different orientations. The results of erosion by the SEs in b)-e) are shown in f)-i). The white pixels are the foreground while the black pixels are the background. All SEs are of size 4 and the pixels are connected via 8-connectivity.

Figure 2.9 gives an example of binary mathematical operators. The original image shown in (a) is eroded, dilated, opened and closed by a 4×4 square SE, shown in (b). The opening in (e) removes small protrusions of the foreground (white pixels) and opens up cavities in image objects. The closing in (f) does the opposite, by filling up cavities within image objects, and bridges gaps between image objects. Note the image objects in the circled area. The opening accurately distinguishes these separate objects. Morphological opening is therefore appropriate to deal with leakage between image objects.

The concept of area operators now follows. An area opening and an area closing are independent of a particular SE, and are defined by the size or area of the SE, rather than its particular shape [134]. The definitions refer to [36]. We first define the operators $\bigvee$ and $\bigwedge$.

Definition 11. *The operator $\bigwedge$ and operator $\bigvee$ are given by:*

$$\bigwedge \boldsymbol{x} = \min \boldsymbol{x},$$
$$\bigvee \boldsymbol{x} = \max \boldsymbol{x},$$

where x is a set of pixels.

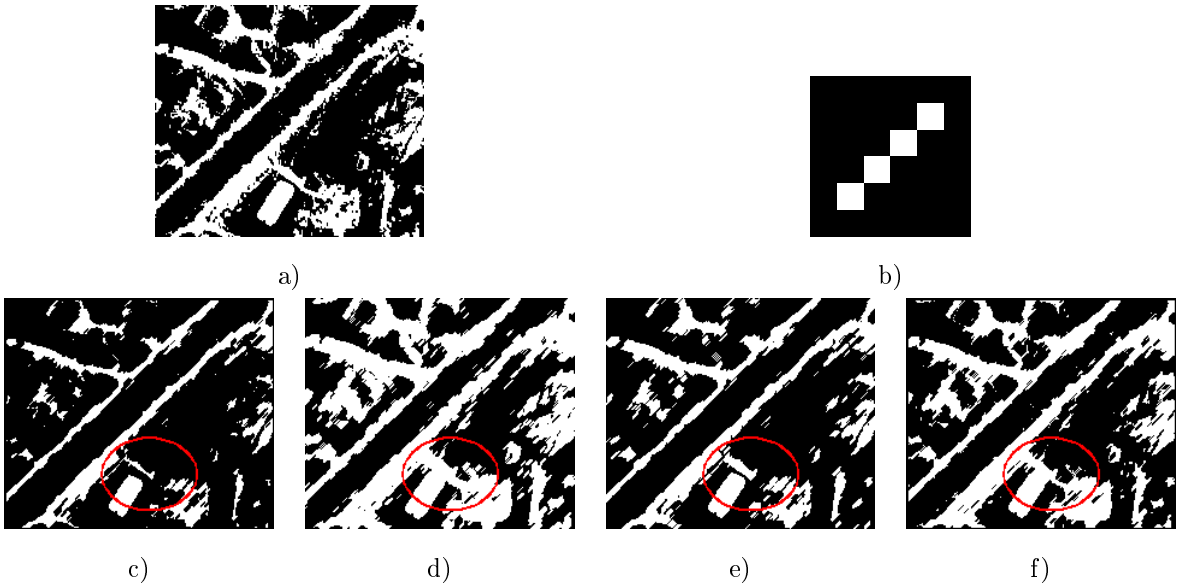


Figure 2.9: Morphological operations on a binary image shown in a), by the 4×4 square SE shown in b). c) Erosion. d) Dilation. e) Opening. f) Closing.

Definition 12. *An area opening is the maximum of all openings with connected structuring elements whose area is λ pixels, namely*

$$\gamma_\lambda = \bigvee_i \{\gamma_{B_i} | B_i \text{ is connected and } area(B_i) = \lambda\}.$$

Definition 13. *An area closing is given by the following:*

$$\phi_\lambda = \bigwedge_i \{\rho_{B_i} | B_i \text{ is connected and } area(B_i) = \lambda\}.$$

2.6.2 Adaptive Morphology

Classical morphology relies on the use of a single SE which is applied to the entire image. Such an SE is called rigid [33], and does not take into account the varying shapes that may be present in an image, in terms of characteristics such as size, orientation, and linearity. Figure 2.9 shows how an SE may be appropriate for some objects in an image, but not others. Various adaptive strategies exist to overcome this difficulty. The adaptivity may enter the method via adapting the SE to account for varying structures in the image, spatial distances between pixels, edges, the image gradient, or noise in the image. [33] provides a comprehensive overview of the various adaptive morphological methods.

Since the overall aim of the algorithm presented herein is to identify and evaluate image objects to determine whether or not they are roads, adaptivity enters our proposed algorithm on a per-object level. A set of linear SEs is considered, representing a line of length 4 at all its possible orientations, and for

each image object it is determined which SE best fits the object. This is done to enhance and evaluate the linearity of image objects. The SEs are presented in Figure 2.10.

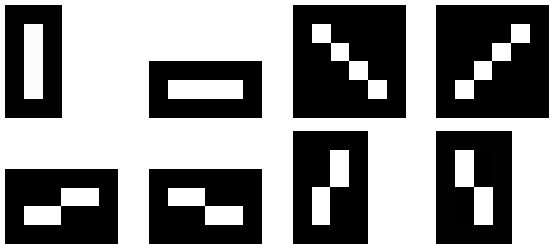


Figure 2.10: Structuring elements used for adaptive morphology of linear structures.

The adaptive algorithm herein is presented below, and illustrated with an example in Figure 2.11. The image in (a) is dilated adaptively in (b), and rigidly in (c), using the SE in (d). It is clear that the dilation in (b) generally results in thinner, more elongated shapes than those in (b), thereby better enhancing linearity. The algorithm determines which SE best fits the object by determining which SE least changes the object. This is measured by dilating the object with each SE under consideration, and counting the amount of pixels in the resulting dilation. The better the SE fits the object, the less dilation by that SE will add pixels to the object. The SE which best fits is therefore the one resulting in the dilation with the smallest amount of pixels.

Adaptive Dilation Algorithm

```
Define the vector of SEs.
1.  For each image object:
2.      For each SE:
3.          1.  Dilate the object by the SE
4.          2.  Determine the number of pixels in the dilation
5.      Determine which SE resulted in the smallest dilation.
This is the SE which best fits the image object.
```

2.7 Image Thinning and Skeletonisation

Many image analysis applications require the thinning or thin-line representation of an image [104, 66]. Thinning obtains a single-pixel width representation that summarise its shape, and contains information regarding its pixel connectivity, orientation and size [83, 116]. Saha, Borgefors and Sanniti di Baja [113] provide a survey of various skeletonisation techniques. Three prominent methods available for efficient

a) b) c) d)

Figure 2.11: Adaptive dilation example. a) The original image. b) The adaptive dilation. c) Dilation performed using only the SE shown in d).

implementation in Python are morphological thinning, the medial axis skeletonisation method of Peters [104], and the morphological method of Zhang [149]. Morphological thinning is an iterative process. At each iteration, the border pixels of the image are removed. Execution stops when no more edges can be removed, i.e. when pixels cannot be removed without changing the pixel connectivity. The method of Peters is a morphological image cleaning technique. It removes noise from images and preserves linear structures by considering the difference between an original image and smoothed versions of the image. Zhang's method executes effectively, resulting in a skeleton that preserves pixel connectivity. Similar to morphological thinning, it executes by removing the contours or edges of the image successively, until no more edges can be removed. Each iteration of Zhang's method consists of two parallel sub-iterations, where the one removes the south-east boundary points and the other removes the north-west boundary points. The parallel nature of the method leads to computational efficiency [148].

Figure 2.12 shows an example of thinning. The image in (a) is skeletonised using Zhang's method (b), Peters' medial axis technique (c) and morphological thinning (d). Zhang's method preserves connectivity and does not tend to result in spurious short branches.

Spurious lines obtained by the skeleton may be pruned [11], where these lines do not accurately represent the centres of the image objects. This is done by detecting end points of lines, and branch points, where the lines branch off from other lines in the skeleton. All pixels are removed that are nearer to an end point than to a branch point. Figure 2.13 illustrates the pruning process. The original skeleton is shown in (a), with end points marked in green and branch points in red, while the pruned skeleton is shown in (b), with the identified branch points circled in red, and all pixels closer to the end points than the nearest branch points removed.

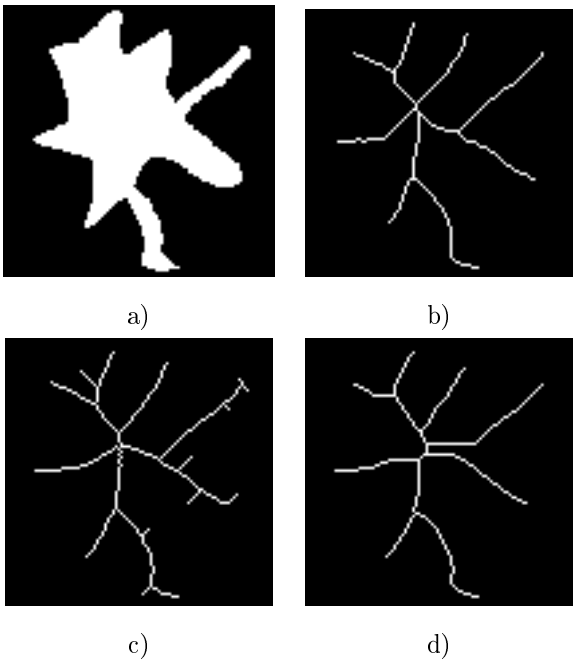


Figure 2.12: Thinning by various methods. a) Original image. b) Result of Zhang's method. c) Result of medial axis skeletonisation. d) The morphological skeleton.

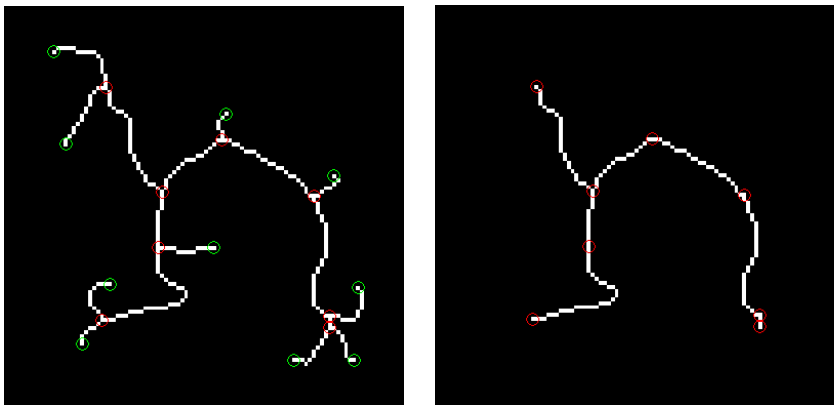
2.8 The Discrete Pulse Transform

2.8.1 The LULU Operators

The DPT works by recursively implementing the LULU operators [36], which were first developed by [109]. These operators act as smoothers on signal data. A signal may be in one or more dimensions, and includes image data [76]. In one dimension, LULU operators work on a sequence, e.g. $\{x_1, ..., x_{j-1}, x_j, x_{j+1}, ..., x_m\}, 1 < j-1, j+1 < m$ where the subset $\{x_i, ..., x_{i+n+1}\}$ is monotone $\forall i$. The LULU operators are the results of combining minimum and maximum operators. These operators were defined for individual elements in Section 2.6 and can be extended to multiple elements as follows [127]:

$$\bigwedge^{n} x = \min\{x_i, x_{i+1}, ..., x_{i+n}\},$$
$$\bigvee^{n} x = \max\{x_i, x_{i+1}, ..., x_{i+n}\}.$$

The definition of the LULU operators follows by applying these operators in composition.



a) b)

Figure 2.13: Skeleton pruning. a) The original skeleton. b) The pruned skeleton, with end points shown in red.

Definition 14. *Any finite composition of the operators* $\bigwedge$ *and* $\bigvee$ *is called a LULU operator:*

$$L_n(x) = \bigvee^{n}\bigwedge^{n} x = \max\{\ \min\{x_j; j \in \{i-n, i\}\}, ..., \min\{x_j; j \in \{i, i+n\}\}\},$$
$$U_n(x) = \bigwedge^{n}\bigvee^{n} x = \min\{\ \max\{x_j; j \in \{i-n, i\}\}, ..., \max\{x_j; j \in \{i, i+n\}\}\},$$

The following theorem states important mathematical properties of the LULU operators.

Theorem 1. *Properties of LULU operators:*

1. $L_nL_m = L_m$ *and* $U_nU_m = U_m$ *for* $m \geq n$,
2. L_n *and* U_n *are idempotent:* $(L_n)^2 = L_n$ *and* $(U_n)^2 = U_n$,
3. L_nU_n *and* U_nL_n *are idempotent operator combinations:*
   $(L_nU_n)^2 = L_nU_nL_nU_n = L_nU_n$ *and* $(U_nL_n)^2 = U_nL_nU_nL_n = U_nL_n$,
4. $U_nL_nU_n$ *and* $L_nU_nL_n$ *are idempotent operator combinations:*
   $(U_nL_nU_n)^2 = UnL_nU_nU_nL_nU_n = U_nL_nU_n$ *and* $(L_nU_nL_n)^2 = L_nU_nL_nL_nU_nL_n = L_nU_nL_n$,
5. L_n *and* U_n *are syntone:* $L_{n+1} \leq L_n \leq L_0 = I = U_0 \leq U_n \leq U_{n+1}$,
6. $L_n \leq L_nU_nL_n \leq L_nU_n$ *and* $U_nL_n \leq U_nL_nU_n \leq U_n$,
7. $L_nU_nL_n \leq U_nL_nU_n$.

In order to describe the working of L_n and U_n, the concepts of local maximum and minimum sets in one dimension are now defined.

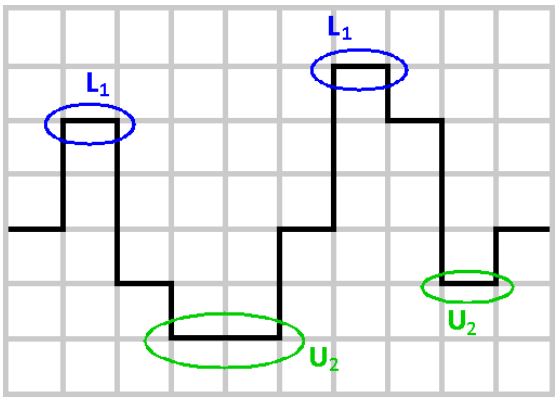


Figure 2.14: Example of LULU operators in one dimension. L_i removes local maxima of size less than or equal to i, while U_i removes local minima of size less than or equal to i. In this signal, i takes on the values 1 and 2.

Definition 15. *Let X be a sequence. Any subset of this sequence with values greater (lower) than the values directly to the left or right of this subset is a local maximum (minimum) set, i.e. the subset $\{x_i, ..., x_{i+j}\}$ of X is a local maximum (minimum) set if $x_k > (<) \quad x_{i-1}$ and $x_k > (<) \quad x_{i+j+1}, k = i, ..., i+j$.*

L_n smooths from above by taking a maximum of minima, and removes local maximum sets of size less than or equal to n. U_n smooths from below by taking a minimum of maxima, and removes local minimum sets of size less than or equal to n. Figure 2.14 gives a visual demonstration of the L_n and U_n operators in one dimension. The line heights represent the values in the sequence. Those line segments that are higher than the line segments around them represent local maximum sets, while those line segments that are lower than the line segments around them, represent local minimum sets. L_1 identifies and removes local maximum sets of size 1, as demonstrated by the blue circles. U_2 identifies and removes local minimum sets of sizes less than or equal to 2, i.e. sets of size 1 as well as sets of size 2. This is illustrated by the green circles.

Result 1. *L_n acts as an area opening and U_n acts as an area closing [36].*

For use in image analysis, the LULU operators are applied to two-dimensional signals. The LULU operators were extended to two dimensions in [36]. In this context, they work on a two-dimensional connected set, with connectivity defined as in Section 2.2.

Definition 16. *The 2-dimensional LULU operators acting on an image f are defined as follows:*

$$L_n(f)(x) = \max_{V \in \mathcal{N}_n(x)} \min_{y \in V} f(y), x \in \mathbb{Z}^2,$$

$$U_n(f)(x) = \min_{V \in \mathcal{N}_n(x)} \max_{y \in V} f(y), x \in \mathbb{Z}^2,$$

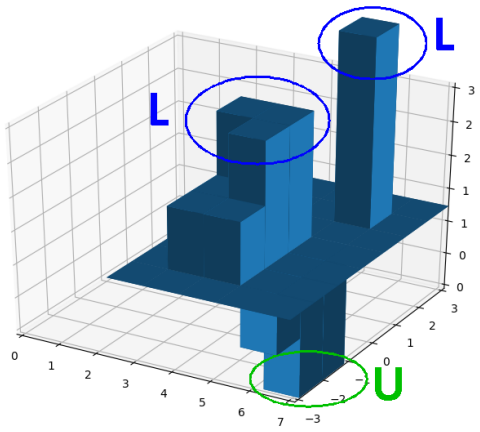


Figure 2.15: Example of LULU operators in two dimensions. L removes local maximum sets, while U removes local minimum sets.

where V is a connected set, $\mathcal{N}_n(x) = \{V \in \mathcal{C} : x \in V, card(V) = n+1\}$ is the set of all connected sets of size $n+1$ that contain the point x, and $\mathcal{C}$ is a connection, e.g. 4-connectivity or 8-connectivity.

Result 2. *The multi-dimensional LULU operators are morphological operators. In particular, L_n is an area opening while U_n is an area closing [36].*

The concepts of local maximum and minimum sets in two dimensions are now defined.

Definition 17. *Let f be an image with domain $X \in \mathbb{Z}^2$ and let C be a connection. Let S be a subset of X and let the pixels adjacent to S be denoted by $adj(S)$, where adjacency is defined by the connection. Then S is a local maximum (minimum) set if the values in S are greater (less) than the values of the adjacent pixels, i.e. if $x > (<)\ y\ \forall x \in S, y \in adj(S)$.*

Figure 2.15 gives a 2-dimensional example, which demonstrates that the L operator identifies maximum sets while the U operator identifies minimum sets. Figures 2.16 and 2.17 illustrates how sets are identified prior to applying L and U in two dimensions, based on a graph representation of an image with pixels connected via 4-connectivity. Figure 2.16 shows connected sets of size $n = 2$, where the red node with black text represents the pixel to which the operators are applied. In particular, the set considered in (a) is a constant set of value 7. Figure 2.17 illustrates sets of size $n = 3$. The set in (a) is a constant set of value 7.

2.8.2 The Discrete Pulse Transform

The Discrete Pulse Transform (DPT) provides a multi-scale decomposition of a signal. It is obtained by recursively applying the LULU operators until the signal is completely smoothed [36]. In the context of

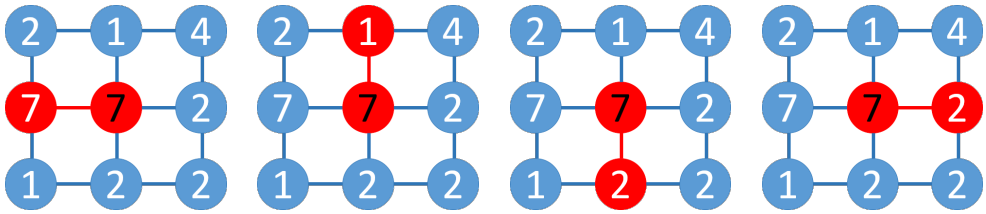


Figure 2.16: Sets of pixels to which the LULU operators are applied. A graph representation of an image is provided, where the number on each node is the pixel value. The red node with a pixel value of 7, in black text, is connected to four neighbours via 4-connectivity. a) A constant set of size $n = 2$ containing the node of interest. b)-d) Other possible sets of size $n = 2$ containing the red node.

image analysis, the DPT decomposes the image, and identifies image objects at various scales.

Definition 18. *The DPT decomposition of an image f is a mapping of an image f onto a vector of matrices:*

$$DPT(f) = [D_1(f), D_2(f), ..., D_N(f)],$$

where N is the number of pixels in the image. Each D_n is calculated as follows:

$$D_1(f) = (I - P_1)(f),$$
$$D_n(f) = (I - P_n) \circ Q_{n-1}(f), n = 2, ..., N - 1,$$

where $P_n = L_n \circ U_n$ or $P_n = U_n \circ L_n$, $Q_n = P_n \circ ... \circ P_1, n \in \mathbb{N}$, and I is the identity operator.

In the above definition, $I(f) = f$ is the original image, $P(f)$ is the smoothed image, and $(I - P)(f)$ is the removed noise component [127]. The original image can be reconstructed from the DPT in the following way:

$$f = \sum_{n=1}^{N} D_n(f).$$

Definition 19. *A pulse ψ_{ns} is a base function that describes D_n in the DPT as follows:*

$$D_n(f) = \sum_{s=1}^{\gamma(n)} \psi_{ns},$$

where each pulse ψ_{ns} consists of n non-zero values, and zero elsewhere, and where $\gamma(n)$ is the number of local minimum and maximum sets of size n.

Two important properties of a pulse ψ_{ns} are its height, given by α_{ns}, which may be positive or negative, and its cardinality, which is equal to n [127].

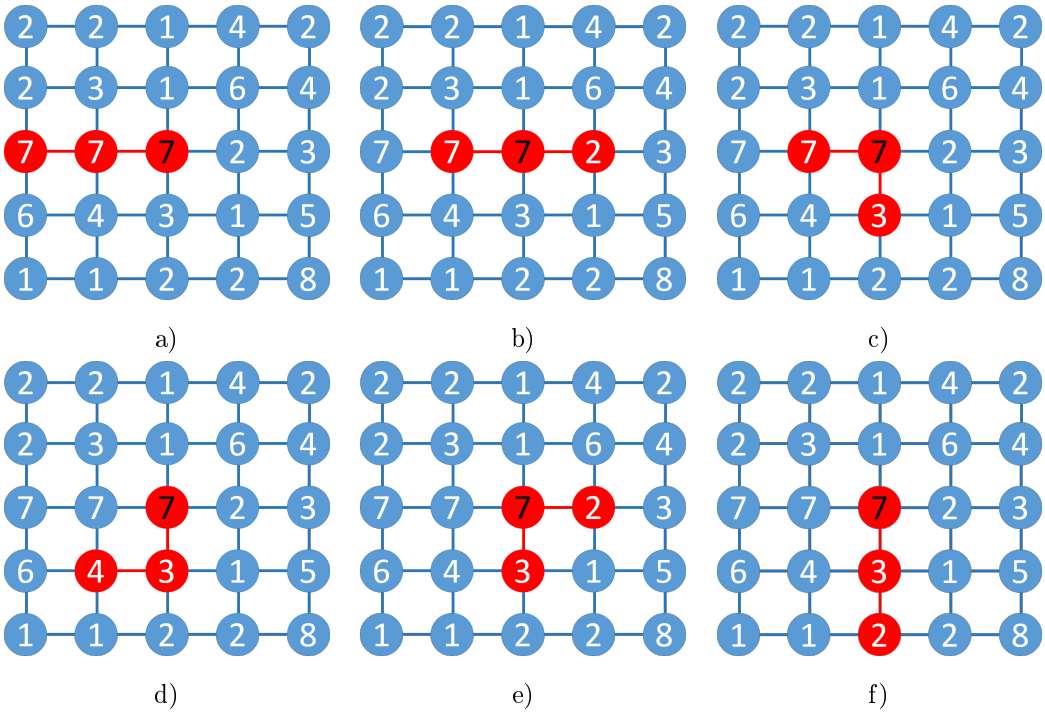


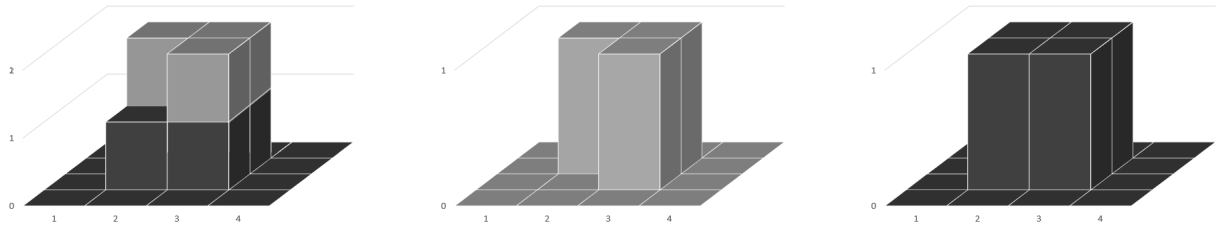
Figure 2.17: Sets of pixels to which the LULU operators are applied. A graph representation of an image is provided, where the number on each node is the pixel value. The red node with a pixel value of 7, in black text, is connected to four neighbours via 4-connectivity. a) A constant set of size $n = 3$ containing the node of interest. b)-f) Other possible sets of size $n = 3$ containing the node of interest.

An image f can now be written in terms of the pulses extracted by the DPT:

$$f = \sum_{n=1}^{N} \sum_{s=1}^{\gamma(n)} \psi_{ns}.$$

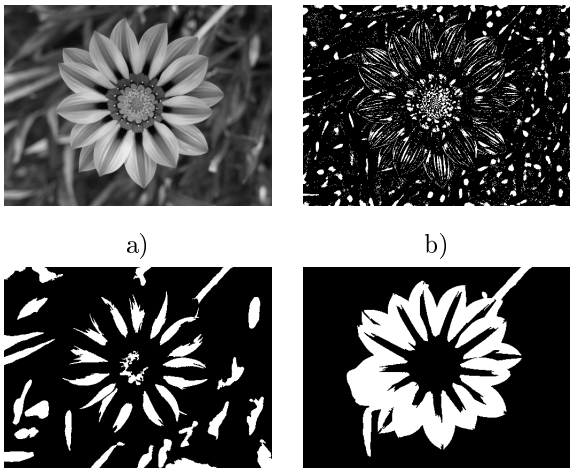
Greyscale images may be represented in three dimensions, with the grey value interpreted as the height, as discussed in Section 2.1 and illustrated in Figure 2.1. The DPT extracts pulses from such images. Figure 2.18 illustrates how pulse extraction is performed via the stepwise removal of local minimum and maximum sets, using the LULU operators. This process is repeated until the image is completely smoothed. The extracted pulses ψ_{ns} provide a multi-scale decomposition of the image. This provides information at each discrete scale, yielding richer data of the image content.

Figure 2.19 shows a greyscale image in (a). Small pulses (size $n = 0$ to 100) are shown in (b), while larger pulses (size $n = 500$ to 1000) are shown in (c), while (d) shows large pulses (size $n = 10000$ to the maximum pulse size). This illustrates how the structure of the image is represented at different scales.



a) b) c)

Figure 2.18: Pulse extraction using the DPT. Local maximum and minimum sets are removed and stored as separate pulses. a) Original 3D signal. b) Local maximum set. c) Remaining image after removal of local maximum set.



a) b)

c) d)

Figure 2.19: Pulse extraction using the DPT. a) A greyscale image. b) Pulses of size $n = 0$ to 100. c) Pulses of size $n = 500$ to 1000. d) Pulses of size $n = 10000$ to the maximum.

2.8.3 The Roadmaker's Pavage

The Roadmaker's Pavage is an efficient graph-based implementation of the DPT [127]. Implementing the DPT is not straightforward due to computational complexity. When implementing the LULU operators directly, the time required for the DPT to execute on a series of length n is of order $O(n^3)$ [68]. Figures 2.16 and 2.17 illustrate this complexity. For a pixel x, the number of sets of size n containing x increases drastically with n. There are 4 sets of size 2 containing x, but 14 sets of size $n = 3$. Attempting to find all sets of size n, $n = 1, 2, ..., N$ where N is the number of pixels in the image, therefore becomes a momentous task. The Roadmaker's Algorithm was developed by Laurie and Rohwer [69] in order to efficiently implement the DPT. This implements the DPT in $O(n)$ time. The idea behind this algorithm is to successively remove all local maximum and minimum sets of increasing scale until the signal is completely flattened, i.e. of a single constant value [69]. At a scale of $n = 1$, all local maximum and minimum sets of size 1 are identified. They are flattened by having their value set equal to the value of their nearest neighbour. The process is repeated for $n = 2, 3, ..., N$, where N is the total length of the signal. The Roadmaker's Algorithm efficiently implements the process of the DPT. It was implemented using graphs for the DPT decomposition in [68]. Python implementations were developed for two dimensions [37] and for d dimensions [68]. However, the Roadmaker's Algorithm results in a dense representation of the extracted pulses [127]. Once the DPT has been executed, it is computationally intensive to retrieve pulses of a given scale. A search must be performed on all extracted pulses in order to find those of the desired scale. This is a concern in the context of the approach developed herein, which requires that pulses of a given scale or scales be processed after completion of the DPT. Stolz [127] developed the Roadmaker's Pavage, which builds on the Roadmaker's Algorithm, but provides a more useful representation of the extracted pulses. These are extracted and stored in a graph in order to be accessible by other operations after the completion of the DPT. This leads to an increase in computational efficiency, and makes the Roadmaker's Pavage desirable for implementation herein. A brief description of its working will be provided below, after which a detailed example will be given, to contextualise the discussion of the algorithm.

In the Roadmaker's Pavage, the image is converted to a graph, called a *work graph*, where each node represents a pixel, and neighbouring pixels are connected via edges. Pixel connectivity was discussed in Section 2.2, and 4-connectivity and 8-connectivity were illustrated. In this research, we consider 4-connectivity when implementing the DPT for computational simplicity. The work graph updates to keep track of the changes in the image as the DPT is executed. A corresponding *pulse graph* is created, where each node in the work graph is connected to a node in the pulse graph via a virtual edge. The pulse graph keeps track of the pulses extracted during DPT execution, allowing for convenient access to the pulses after the completion of the DPT.

The Roadmaker's Pavage firstly merges all connected pixels of the same pixel value, which results in a

simplification of the work graph, and no change in the pulse graph. The DPT is then executed, during which local maximum and minimum sets are flattened at each scale, and these sets are extracted as pulses. Each extracted pulse is represented as a new node in the pulse graph. The local minimum or maximum set is called a *feature*, and a reference is made to it in the *feature* table, which keeps a record of all features in the image, at all considered scales.

The example in Section 2.8.3.1 illustrates the Roadmaker's Pavage on a greyscale image. Section 2.8.3.2 presents the relevant theory.

2.8.3.1 Roadmaker's Pavage Example

Consider the image in Figure 2.20(a). The numbers are hypothetical grey values. The connectivity is shown in (b). Figure 2.20(c) displays the corresponding work graph (left) and pulse graph (right). In the work graph, the values inside the circles (nodes) are the pixel values or height, while the superscripts indicate the size of the set, i.e. the number of pixels associated with the node. In the initial steps, each node corresponds to a pixel, therefore they are all of size 1. The nodes are connected by edges, represented in the figure as solid lines. Each edge between nodes of finite size, corresponds to an arrow in (b). The extra node of height 0 and infinite size is the edge node. In this way, a discrete image is manipulated to be on an infinite domain by padding it with zeroes. This is required since the theory of the LULU operators requires an infinite domain. Every pixel on the edge of the image is connected to this node. The only node that is not connected to the node of infinite size is the node that corresponds to the center pixel in the image, since this pixel is not on the edge. The pulse graph consists of nodes of size 0, one corresponding to each pixel or initial finite-size node in the work graph, and height corresponding to the pixel index. Each node in the pulse graph is connected to a (finite-sized) node in the work graph via a virtual edge, shown in the figure as dashed lines. The nodes that are in the pulse graph at initialisation are called the base nodes.

The first step in the Roadmaker's Pavage is to merge connected pixels with the same values. Figure 2.21 shows how merging is performed in the work graph. The merging step does not change the pulse graph, except that the virtual edges connect to different nodes in the work graph. Figure 2.21(g) corresponds to the merging of neighbouring pixels with the value of 3, as demonstrated in (a). One of the nodes of height 3 and size 1 is removed, and the remaining node of size 3 now receives a height of 2. In addition, the neighbours associated with the deleted node now become neighbours of the remaining node. This is done by connecting these neighbours to the node of height 2 via edges in the graph. These edges are shown as solid red lines. The virtual edge connecting the deleted node to the pulse graph, is transferred to the node of height 2. This virtual edge is shown as the dashed red line.

The process is repeated for the next set of neighbouring pixels of the same colour, namely the pixels of

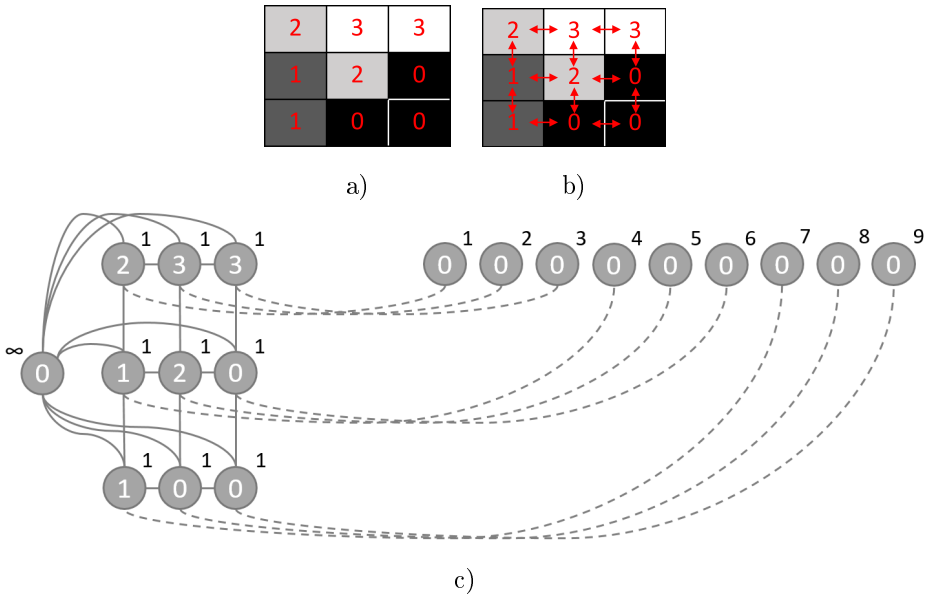


Figure 2.20: Example greyscale image with pixel values (a), connectivity (b) and work and pulse graphs (c). The graph on the left is the work graph, while the graph on the right is the pulse graph.

value 1, represented in the work graph as nodes of height 1 and size 1. These are merged to one node of height 1 and size 2, as shown in Figure 2.21(h). The merge is shown in terms of image pixels in (c) and (d). The last merge is between the black pixels of value 0, i.e. nodes of 0. The resulting node in the work graph is of height 0 and size 3, as shown in Figure 2.21(i). The final merged image is represented in (f).

The DPT is performed on this merged graph. The process is demonstrated in Figure 2.22. The decomposition searches for local minimum sets (dips) and maximum sets (bumps), and removes them sequentially in increasing size, starting at size 1. These bumps and dips are called *features*. In this case, there are no bumps or dips of size 1, since the only local minimum, of height 0, is of size 3. The only local maximum, of height 3, is of size 2. There are no dips of size 2. However, there is a bump of size 2, namely the node of height 3, as shown in Figure 2.22(a) and (b). Since its neighbour is of height 2, the bump is of height 1. This node is merged to its nearest neighbour, with a height of 2. Since there are two neighbours of height 2, all four nodes are merged, with the resulting image in (c). A pulse of height 1 and size 2 is extracted, shown in (d) and (e). The resulting graph in (f) shows the merged node of height 2 and size 4, with its updated connection in the work graph shown as a solid red line, and its updated connections to the pulse graph shown as red dashed lines. The pulse graph now changes. A new node of height 1 and size 2 is created, corresponding to the extracted pulse shown in (e). The parents of this node are connected to it via directional connections, shown as red arrows. It is connected to the node of height 2 and size 4 in the work graph via a virtual edge.

Next, a dip of $n = 3$ is found, namely the node of height 0 and size 3, shown in Figure 2.23(a). The dip

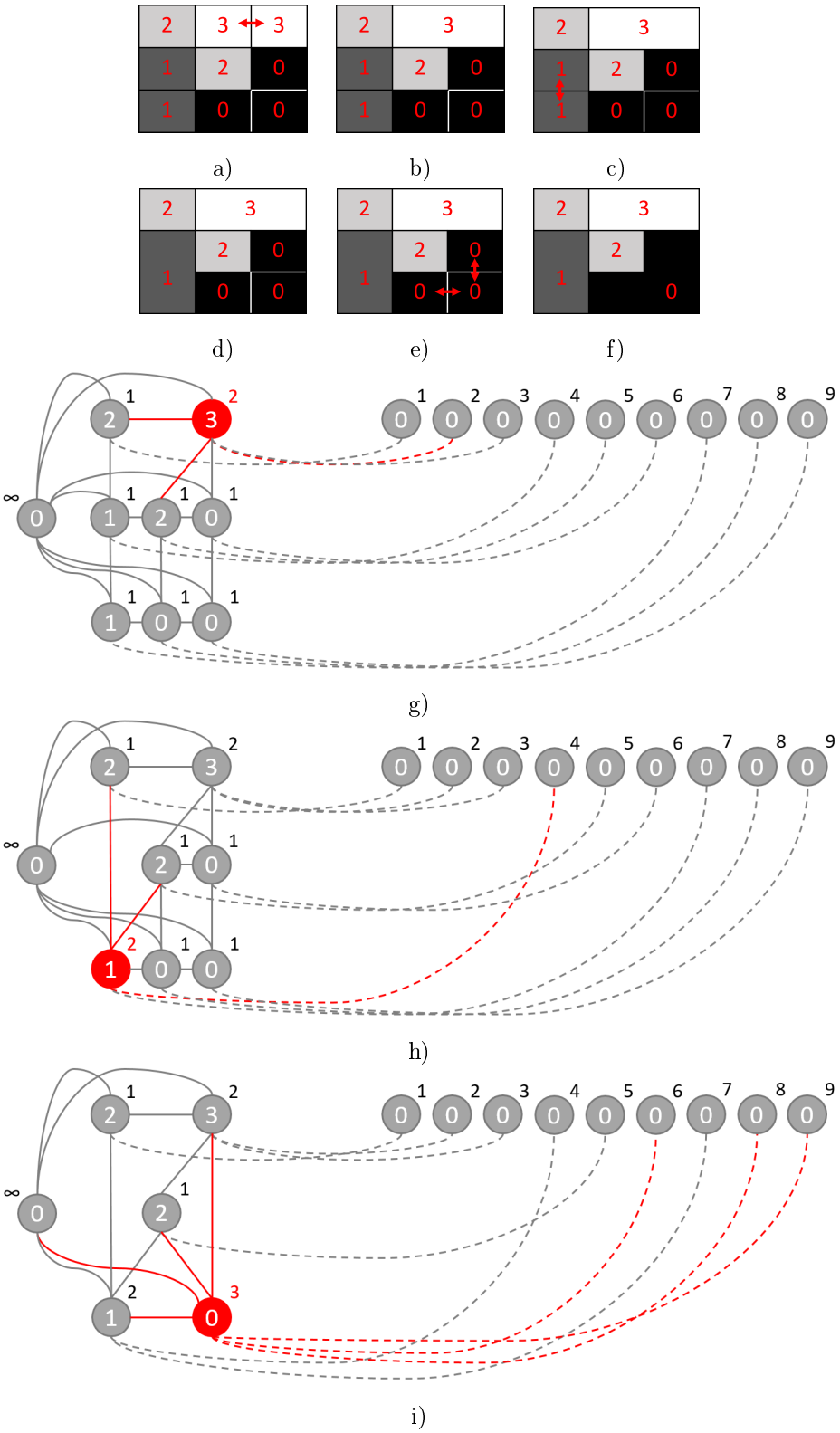


Figure 2.21: The initial merging step of the Roadmaker's Pavage, demonstrated on the image in Figure 2.20(a). The way in which connected pixels of the same colour are grouped in the image, is visualised in a)-f). The changes in the work and pulse graph are shown in g)-i), where the work graph is on the left and the pulse graph is on the right.

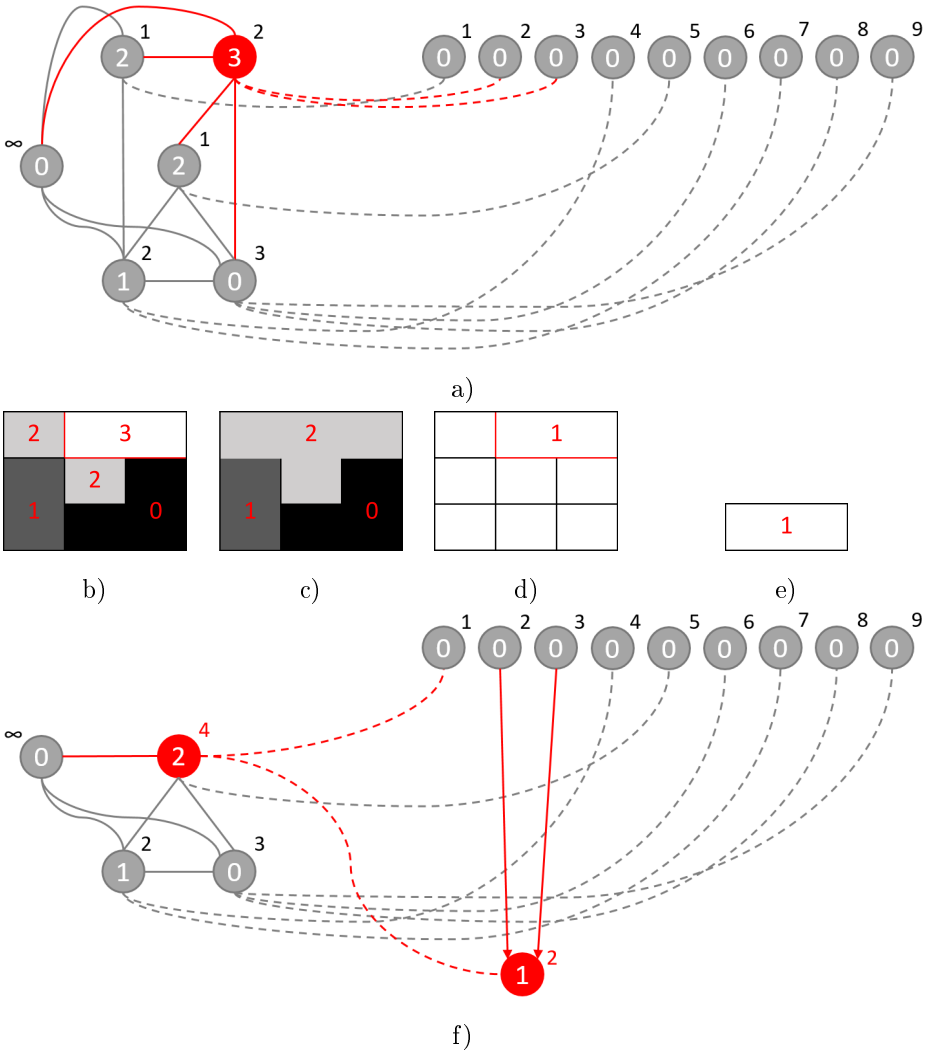


Figure 2.22: Removal of a bump and extraction of a pulse of height 1 and size 2. a) The work and pulse graphs prior to bump removal. b) The image prior to bump removal. c) The image after bump removal. d) The location of the pulse in the image. e) The extracted pulse.

is of height -1, since the node is of height 0 and its nearest neighbour is of height 1 and size 2. These are merged, with the resulting image in (b). A pulse of size 3 and height -1 is extracted, shown in (c) and (d). The graph in (e) shows that the new node of height 1 in the work graph now has a size of 5, and has inherited the connections associated with the node of height 0. Also, the pulse graph has been updated with a new child node of height -1 and size 3, corresponding to the extracted pulse (shown graphically in (d).

At $n = 4$, a dip of height 1 and size 4 is found, corresponding to the group of pixels of value 2, shown in Figure 2.24(a) and (b). The bump is flattened by merging the node with its nearest neighbour of size 4 and height 1. A new node of height 1 and size 4 is added to the pulse graph, as in (f).

The final pulse extracted is of height 1 and size 9, corresponding to a cluster of pixels the size of the entire image, as shown in Figure 2.25(a). In this step, the image is smoothed completely, shown in (b), with the extracted bump shown in (c). The bump is of height 1. The final work and pulse graphs are shown in (d). Figure 2.26 compares the work and pulse graphs prior to (a) and after (b) the execution of the DPT via the Roadmaker's Pavage implementation. The final node at the bottom of the pulse graph is called the head node.

2.8.3.2 Roadmaker's Pavage Theory

The Roadmaker's Pavage creates the graphs as shown in the example in Section 2.8.3.1 and updates them during the execution of the DPT, to result in a final pulse graph, and a feature table that stores all the extracted local maximum and mininum sets at each scale. An overview of the methodology of the Roadmaker's Pavage algorithm is provided below.

In the first step, the image is converted into a one-dimensional vectorised dataset. For the image in the above example, the conversion occurs as follows:

$$vec \begin{bmatrix} 2 & 3 & 3 \\ 1 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix} = [2, 3, 3, 1, 2, 0, 1, 0, 0].$$

The second step defines connectivity, specifying which nodes in the graph are neighbours. A set of connectivity functions $C_k(i)$ is defined, where i refers to the index of the node in the vector, and $k = 1, ..., p$ is the number of neighbours. In 4-connectivity, there can be up to 4 neighbours, and the connectivity

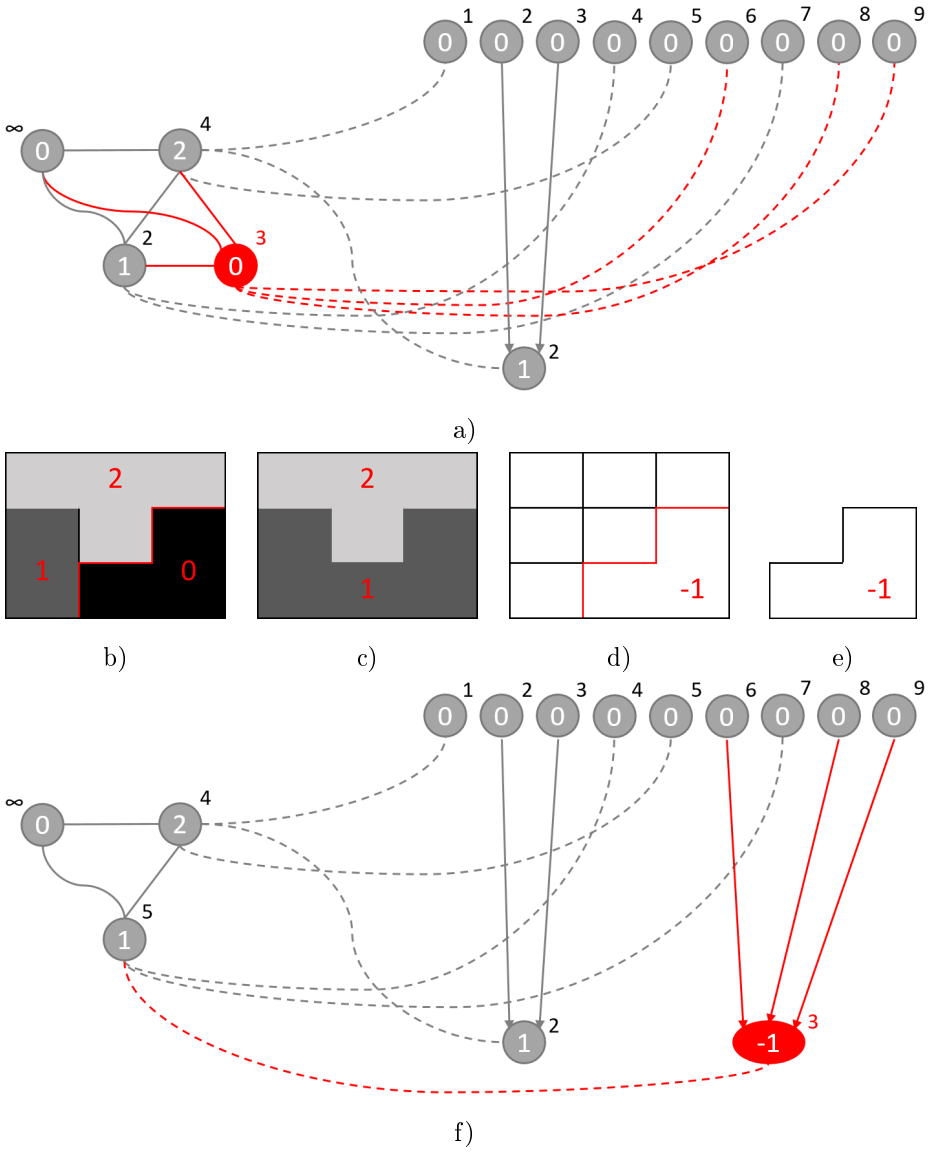


Figure 2.23: Removal of a dip and extraction of a pulse of height -1 and size 3. a) The work and pulse graphs prior to dip removal. b) The image prior to dip removal. c) The image after dip removal. d) The location of the pulse in the image. e) The extracted pulse.

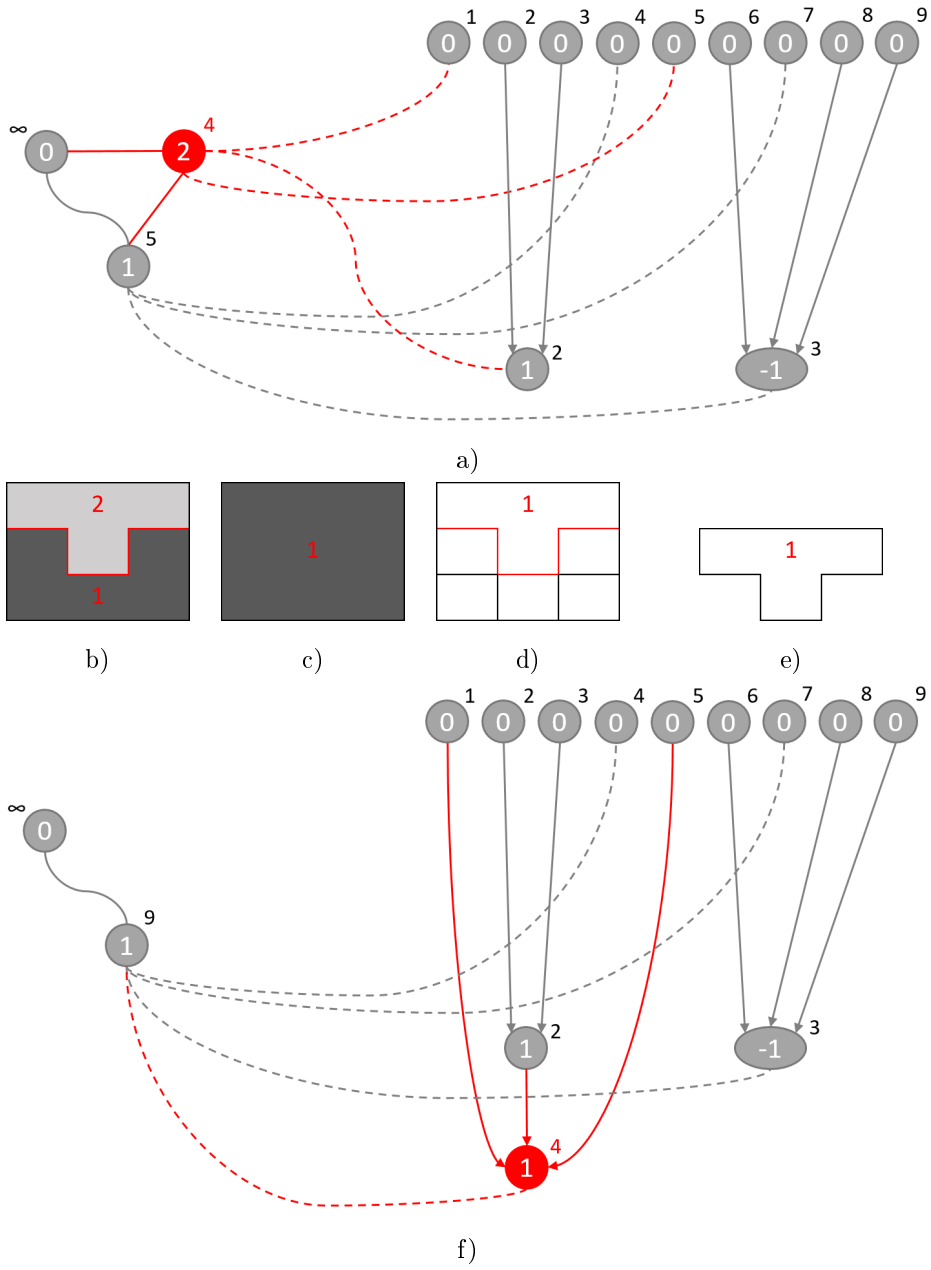


Figure 2.24: Removal of a bump and extraction of a pulse of height 1 and size 4. a) The work and pulse graphs prior to bump removal. b) The image prior to bump removal. c) The image after bump removal. d) The location of the pulse in the image. e) The extracted pulse.

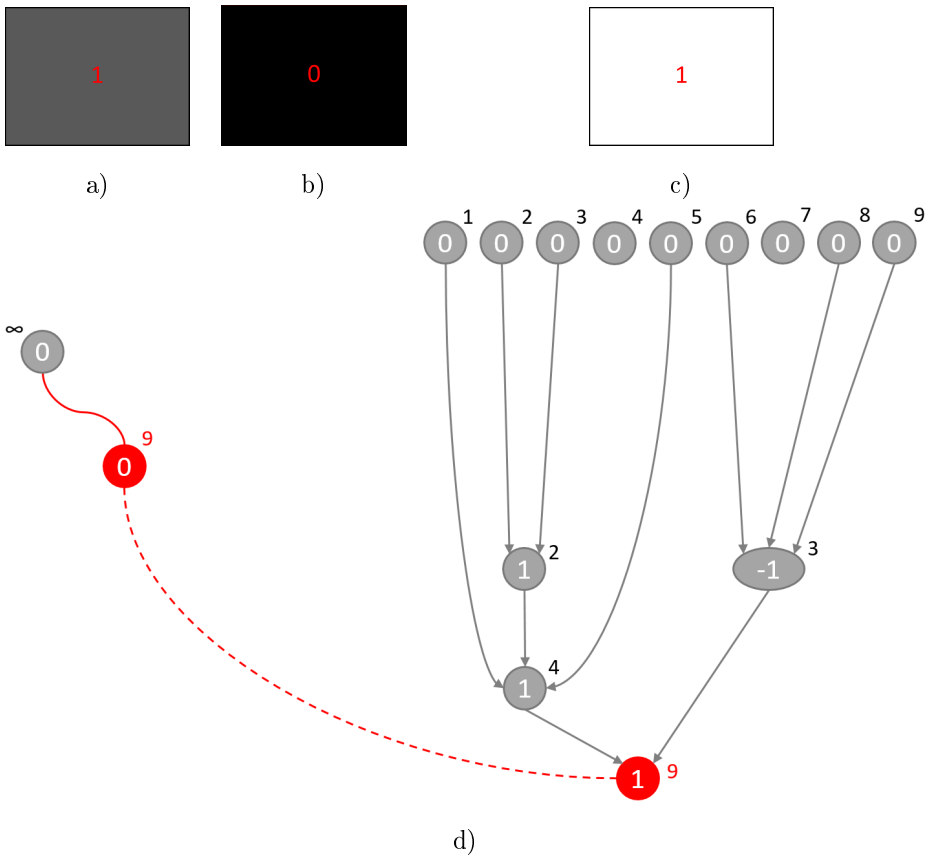


Figure 2.25: Removal of a bump and extraction of a pulse of height 1 and size 9.

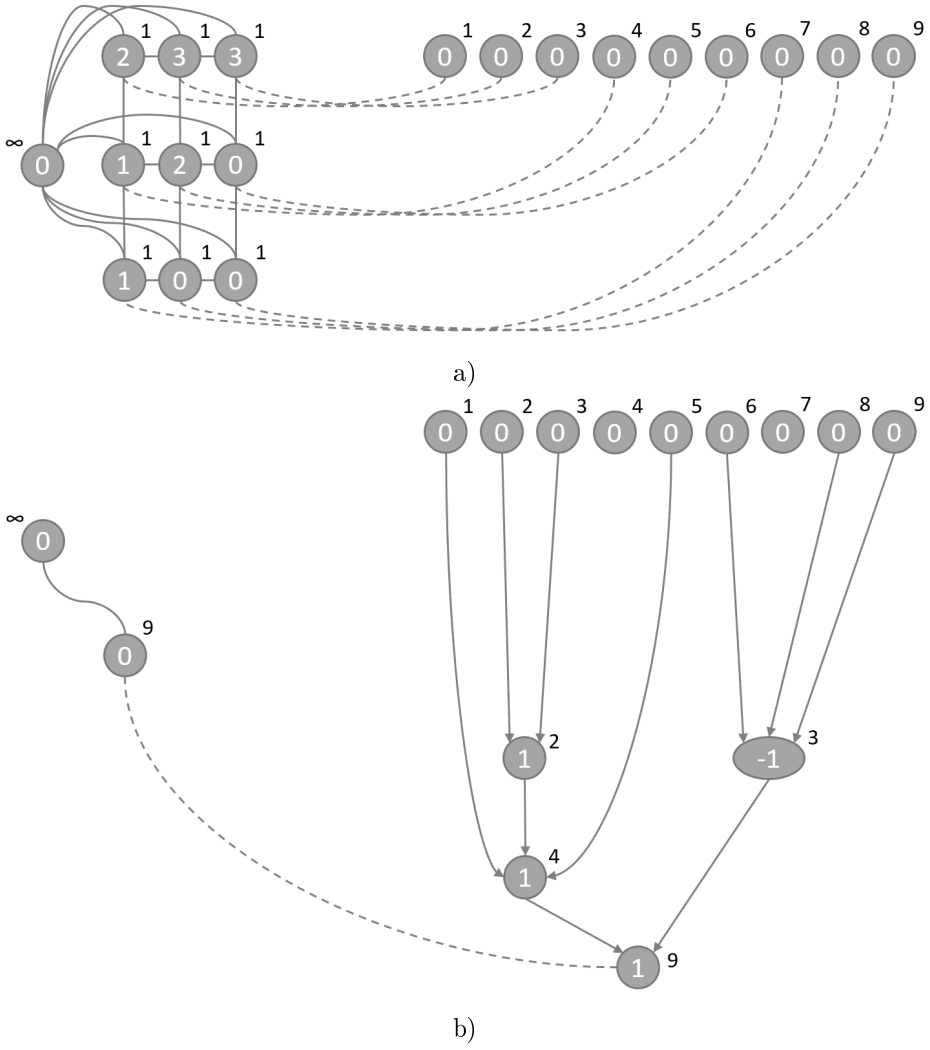


Figure 2.26: a) The work and pulse graphs prior to the DPT execution. b) The work and pulse graphs after DPT execution, as implemented via the Roadmaker's Pavage.

functions are defined as follows:

$$\begin{aligned}
\text{Top neighbour: } & C_1(i) = i - \max_1 \\
\text{Bottom neighbour: } & C_2(i) = i + \max_1 \\
\text{Left neighbour: } & C_3(i) = i - 1 \\
\text{Right neighbour: } & C_4(i) = i + 1,
\end{aligned}$$

where $\max_1$ is the maximum of the first dimension of the image. In the above example, $\max_1 = 3$ and the indices of the neighbours of the central pixel (value 2, $i = 5$) are given by:

$$\begin{aligned}
\text{Top neighbour: } & C_1(5) = 5 - 3 = 2(\text{pixel value} = 3) \\
\text{Bottom neighbour: } & C_2(5) = 5 + 3 = 8(\text{pixel value} = 0) \\
\text{Left neighbour: } & C_3(5) = 5 - 1 = 4(\text{pixel value} = 1) \\
\text{Right neighbour: } & C_4(5) = 5 + 1 = 6(\text{pixel value} = 0).
\end{aligned}$$

Next, the work and pulse graphs are initialised. The work graph is created with a node for each pixel, where the node height is equal to the pixel value, and the node size is equal to one. An additional node is created with an infinite height and a value of zero, representing the edge. The nodes are connected via the specified connectivity (e.g. 4-connectivity), and all the nodes corresponding to border pixels are connected to the 0 edge node. The nodes in the work graph are connected by virtual edges to their corresponding nodes in the pulse graph. The pulse graph is created with a node for each pixel, with a height of zero and a size equal to the pixel value. These are called the base nodes. The edge node in the work graph does not have a corresponding node in the pulse graph.

The feature table is now created. This is an important step of the Roadmaker's Pavage, and the main reason for its efficiency as an algorithm. The feature table stores a list of all the local maximum and minimum sets in the work graph, and reveals the maximum number of features that can be in the table at any one time, at the beginning of the algorithm. The application of the DPT does not lead to the creation of new features, only the extraction of existing features [127]. It is therefore not necessary to search for features during a later stage of the algorithm, reducing the need to traverse the work graph to identify features. This improves algorithmic efficiency and reduces computation time. The feature table is created by merging all equal-valued connected nodes (neighbouring pixels) into one node. The merged node has height equal to the pixel value, and size equal to the number of pixels in the merged nodes. The connections are also updated so that the merged node inherits the connections of all the nodes that were merged, both to nodes in the work graph and virtual connections to nodes in the pulse graph.

After the work and pulse graph initialisation and the creation of the feature table, the DPT is applied to the work graph. During this procedure, local maximum and minimum sets are removed from the work

graph. The final output of the DPT is the modified pulse graph, showing the extracted pulses at all considered scales. The DPT is applied by searching for local maximum and minimum sets at each scale, starting at a scale of one. In this implementation bumps are first searched for, and then dips, although the order may be altered and results in only minor differences in the images.

In the first step, bumps of size 1 are searched for in the feature table. If one is found, it is removed. A node is added to the pulse graph to represent the removed feature. This node has a size equal to the current scale, i.e. 1 at the first step. Its height is equal to the difference between the feature node in the work graph, and its nearest-valued neighbour. In the work graph, the height of the node which was a feature is changed to the height of its nearest-valued neighbour, and these nodes are merged. It is determined whether this resulting node, representing a constant set, is a feature (i.e. a local minimum or maximum set). If it is a feature, the old feature is updated in the feature table. If not, the old feature is removed from the feature table. Once the bumps of size 1 have been removed, the process is repeated for dips of size 1.

Due to the constant alteration to the feature table, it cannot be assumed that every node corresponding to a constant set is a feature. The value of an adjacent node may change during the execution of the DPT, and be made equal to the node in question, in which case the nodes must be merged. Prior to extracting a feature, therefore, the feature must be reaffirmed. This is done by revisiting its neighbours to determine if it is still a local minimum or maximum. If it is not, it must be merged with its equal-valued neighbour(s) and the feature table must be updated.

Once all local maximum or minimum sets of size 1 have been removed, and there are no more features in the feature table of size 1, the scale is increased to 2, where the above process is once more repeated. When there are no more features of the current scale in the feature table, the scale is incremented by one, and local maximum and minimum sets of the new scale are removed. This is done until no more features exist in the feature table. The final node added to the pulse graph consists of all pixels in the image, and has height equal to the number of pixels in the image. This node is called the head node. This completes the execution of the DPT via the Roadmaker's Pavage. The algorithm is provided below:

Roadmaker's Pavage Algorithm

1. `Convert the 2D image to a 1D vectorised dataset.`
2. `Define connectivity.`
3. `Construct Work Graph (WG) and Pulse Graph (PG).`
4. `Create Feature Table (FT).`
5. `Apply the DPT using WG, PG and FT as inputs.`

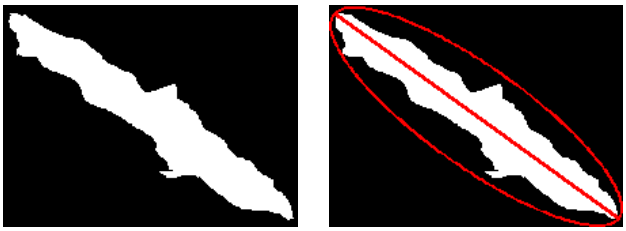
The advantage of applying the DPT using the Roadmaker's Pavage is that the original image can be easily reconstructed from the extracted pulses. In particular, specific pulses are reconstructed, namely those adhering to specific size criteria. This is done via conditional pulse reconstruction. This reconstruction is executed beginning from the initial pulses in the pulse graph, i.e. each data point is reconstructed individually. The conditional pulse reconstruction algorithm is given below. The reconstruction executes by traversing the pulse graph after application of the DPT. For each base node, the graph is traversed in the direction of the head node, and the size of each node is considered. If the height of the node is equal to the specified size, its corresponding pulse is reconstructed.

Conditional Pulse Reconstruction Algorithm

```
 Let Data be the image array.  Denote the ith base node by Bi, the required size by
RequiredSize.
1.  For every Bi:
2.  Let CNode = Bi, height = 0
3.  while there is a connection between CNode and Bj:
4.    if size(CNode) == RequiredSize:
5.        height = height + height(CNode)
6.    else if size(CNode) > RequiredSize:
7.        exit while loop
8.    else:
9.        CNode = Bi
10.  Data[height(Bi)] = height
11.  Return Data
```

2.9 Measuring Linearity

Geometric features can be used to differentiate between object types. Roads are linear features, and therefore linearity can be used to distinguish roads from other bare soil areas, and buildings with a similar colour. The algorithm herein measures linearity via compactness and elongation. The use of these shape measures for segmentation and feature classification is well-documented [19, 27], including for road extraction [71]. The calculation of these measures requires the area, length and perimeter of the image



a) b)

Figure 2.27: Illustration of the calculation of length. a) An image object. b) The major axis of the smallest ellipse containing the image object is the length.

region,

$$\text{Compactness} = \frac{2(\sqrt{\pi a})}{p},$$
$$\text{Elongation} = \frac{a}{\text{len}^2},$$

where a is the area of the region, p is the perimeter of the region, and len is the length of the region. Length is measured as the major axis of the smallest ellipse enclosing the region, as demonstrated in Figure 2.27. Compactness measures the relation between area and perimeter, where less compact areas have a larger area compared to the perimeter. Elongation considers the relationship between the area and the length, where areas with lower values have greater length compared to area. Note that elongation is an inverse measure, i.e. areas with a lower elongation value are more elongated, and vice versa. Section 3.5 provides examples of the compactness and elongation of shapes being considered. According to [71], built-up regions are found to generally be highly compact and not elongated, while roads are not compact, but are elongated.

The compactness and elongation per pulse are obtained following the execution of the DPT. The DPT obtains pulses based on size, or area. In addition, the length and perimeter must be calculated in order to calculate compactness and elongation. The availability of pulses in the work graph after the completion of the DPT allows for easy measurement of these characteristics. Calculating the elongation and compactness for each pulse is therefore a straightforward matter, providing a computationally efficient way to measure linearity.

2.10 Certainty

Certainty is quantified using the compactness and elongation measures, as in the method of [115], also used in [71]. For each image object, the necessity of the object being a road is calculated. The necessity specifies the degree to which it is certain that an object is a road [115], and may be used to evaluate

certainty [18]. This measure is based on compactness and elongation, where these are used as inputs to a Z-shaped fuzzy membership function f, given below and illustrated in Figure 2.28.

$$f(x;a,b) = \begin{cases} 1 & x \leq a \\ 1 - 2(\frac{x-a}{b-a})^2 & a \leq x \leq \frac{a+b}{2} \\ 2(\frac{x-b}{b-a})^2 & \frac{a+b}{2} \leq x \leq b \\ 0 & x \geq b \end{cases} \tag{2.1}$$

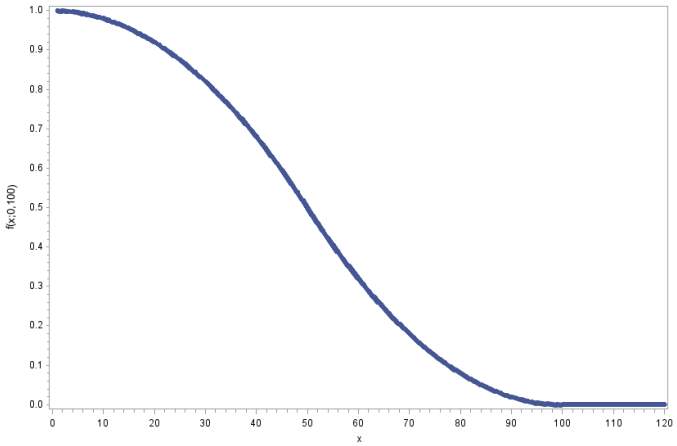


Figure 2.28: Example of the Z-shaped fuzzy membership function f with lower bound parameter 0 and upper bound parameter 100.

This results in the following measures:

$$\mu_c = f(c_x, a_c, b_c) \text{ based on compactness, and } \mu_e = f(e_x, a_e, b_e) \text{ based on elongation,}$$

where c_x and e_x are the compactness and elongation, respectively, of an image object x. The parameters a_c and a_e are lower bounds for the compactness and elongation respectively, while b_c and b_e are upper bounds. The values chosen for these parameters depend on the application. Section 3.7 specifies the choice for the algorithm herein. The necessity, or certainty that the object is a road, is quantified as follows:

$$\text{Certainty} = 1 - \max(1 - \mu_c, 1 - \mu_e).$$

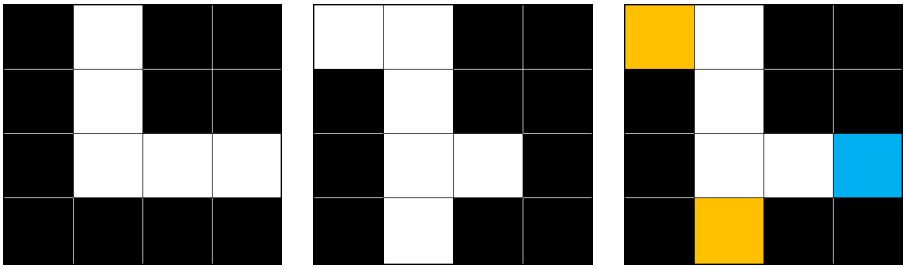
Image objects with lower compactness and elongation are more certain to be roads.

2.11 Accuracy Measures

Accuracy assessment is performed by comparing the extracted roads with a reference, which is assumed to represent the truth. The comparison may be done at various levels, such as by comparing individual pixels or considering roads as objects, and may consider different characteristics, such as the location and extent of roads. The output of accuracy assessment is a measure (measures) that quantifies (quantify) the accuracy of the extraction results.

2.11.1 Per-pixel Measures

Per-pixel measures are computed by comparing each pixel of the extraction, with each pixel of the reference. True positives and true negatives occur when the corresponding pixel in the extraction and the reference coincide in terms of type. A true positive occurs when the extraction classifies a pixel as a road, and the corresponding pixel in the reference is also a road. A true negative occurs when the pixel is non-road in both the extraction and the reference. If the extraction classifies a pixel as road, which is non-road in the reference, this pixel is a false positive. If a pixel is detected as non-road by the extraction, but is a road pixel in the reference, this is a false negative. Figure 2.29 illustrates these concepts. The extracted image in (b) is compared to the reference image in (a). The comparison in (c) shows false positives in yellow, where the extraction detected a pixel that was not present in the reference, and false negatives in blue, where the extraction did not detect pixels that were present in the reference.



a) b) c)

Figure 2.29: An example of accuracy assessment. The reference image in a) is assumed to be true. The extracted image in b) deviates from the reference image in some places. The comparison in c) shows false positives and false negatives. The white pixels are true positives, the yellow pixels are false positives and the blue pixel is a false negative.

Based on the true positive, false positive and false negative counts, the following per-pixel accuracy

measures are calculated:

$$\begin{aligned}\text{Completeness} &= \frac{TP}{TP+FN}\\ \text{Correctness} &= \frac{TP}{TP+FP}\\ \text{Quality} &= \frac{TP}{TP+FP+FN}\\ &= \frac{\text{Compl.} \times \text{Corr.}}{\text{Compl.} + \text{Corr.} - \text{Compl.} \times \text{Corr.}},\end{aligned}$$

where TP is the count of True Positives, FP of False Positives and FN of False Negatives.
These measures all range between 0 and 1, where higher values indicate higher accuracy. Completeness expresses the true positives as a proportion of the total pixels in the reference. The higher the completeness, therefore, the fewer false negatives were observed, and hence the more *complete* the extraction results. Correctness expresses the true positives as a proportion of the total pixels in the extraction. Higher correctness values indicate that fewer false positives were observed. The extraction is therefore more *correct*, i.e. the chances are higher that a detected road pixel truly being part of a road. Quality does not include extra information that the above present in completeness and correctness, but is useful for summarising the accuracy with a single measure for the sake of making comparisons.

The completeness in the case of Figure 2.29 is equal to the number of white pixels (true positives), divided by the number of white pixels plus the number of blue pixels:

$$\begin{aligned}\text{Completeness} &= \frac{4}{4+1} = 0.8,\\ \text{Correctness} &= \frac{4}{4+2} = 0.6667\\ \text{Quality} &= \frac{4}{4+2+1} = 0.5714\end{aligned}$$

2.11.2 Inclusion Completeness and Correctness

Inclusion completeness and correctness measures are desired that assess whether or not roads in the reference are included in the extraction. These should provide a measure of how well the presence of image objects is detected, without being sensitive to the exact boundaries. Two measures are considered, namely inclusion completeness and inclusion correctness.

Inclusion completeness sets out to measure the percentage of the roads in the reference dataset that are not detected, i.e, the false negatives. It is calculated by comparing the full extracted dataset with the centrelines of the reference, obtained by skeletonisation. These are intersected, resulting in the reference image comparison (RIC).

Inclusion correctness indicates the percentage of the roads in the extracted dataset that are falsely detected, not corresponding to objects in the reference. These are the false positives. It is calculated from

the extracted image comparison (EIC), namely the intersection of the extracted centrelines and the full reference dataset.

The formulae are given below:

$$\text{Overall completeness} = \frac{\text{number of white pixels in RIC}}{\text{number of white pixels in the reference}}$$
$$\text{Overall correctness} = \frac{\text{number of white pixels in EIC}}{\text{number of white pixels in the extraction}}.$$

Figure 2.30 provides an example. The extraction in (a) is compared with the reference skeleton in (b), resulting in the RIC in (c), which is evaluated to obtain an indication of the false negatives. The reference in (d) is compared to the extracted skeleton in (e), obtaining the EIC in (f). This is used to calculate the false positives.

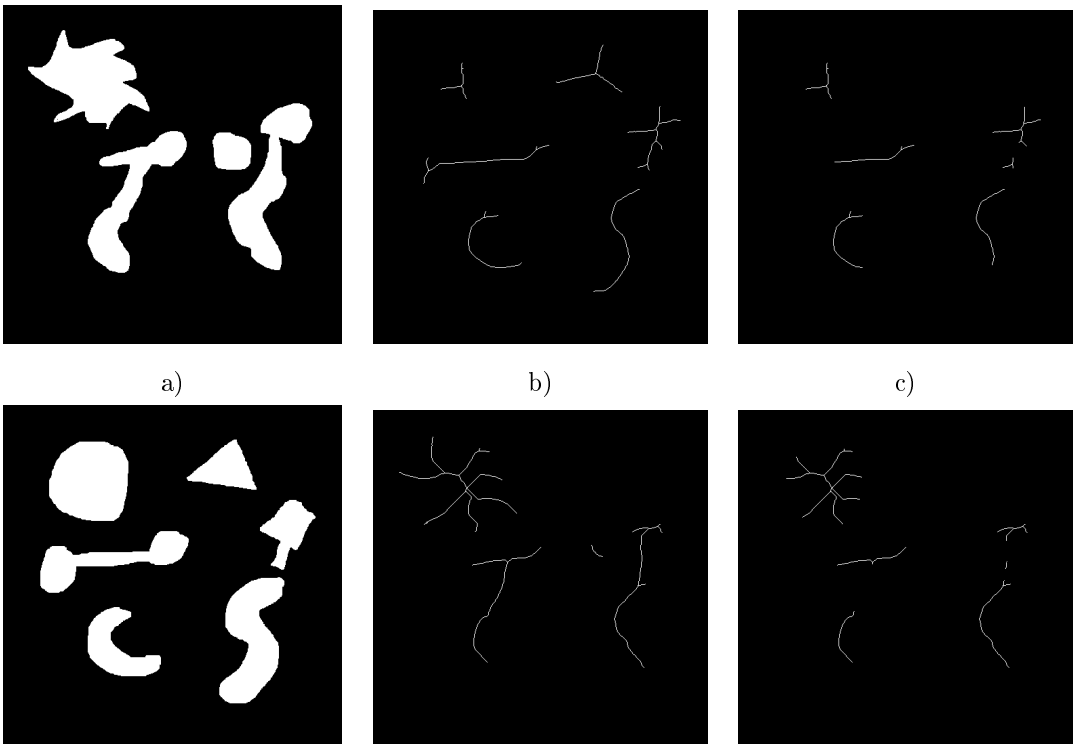


Figure 2.30: An example of accuracy assessment measuring road inclusion. a) The extraction. b) The skeleton of the reference. c) The RIC is the intersection of a) and b). c) The reference. d) The skeleton of the extraction. e) The EIC is the intersection of d) and e).

2.11.3 Pratt's Figure of Merit

Pratt's Figure of Merit (PFOM) is a versatile accuracy measure developed to assess edge detection [4]. It has been widely used for this purpose [46, 45, 60], for both binary and colour images [138]. It has been utilised as a basis for other edge detection measures [105], and has also been employed in other applications, including video segmentation in [45]. PFOM is based on the distances between the extracted and reference edges. The formula is given below:

$$PFOM = \frac{1}{\max\{N_R, N_E\}} \sum_{i=1}^{N_E} \frac{1}{1 + \alpha + \delta_i^2},$$

where N_R is the number of pixels of the reference edges, N_E is the number of pixels of the extracted edges, α is a scaling constant usually set equal to $\frac{1}{9}$, δ_i is the Euclidean distance between the i^{th} detected pixel and the nearest reference pixel. For a perfect match, $PFOM = 100\%$.

Herein, PFOM is calculated in two different ways. In keeping with the history of its use for edge detection, it is firstly calculated on the edges of the reference and extraction. Herein, the edges are calculated using the Laplace gradient, which has been used previously in road extraction applications to detect noise that may influence extraction quality [26]. The Laplacian gradient of an image f is given by the following formula:

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2},$$

where x and y are the horizontal and vertical axes, respectively, of the image. Since images are on a discrete space and not continuous, the derivatives must be approximated. Herein, the partial derivatives are approximated by filtering the image with the following kernel [21]

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

This calculation of PFOM assesses how precisely the boundaries of objects are captured. Secondly, PFOM is calculated on the reference and extraction skeletons, once the skeletons have been pruned to remove spurious lines. The formula above is not altered, but the reference and extracted skeletons are used as the reference and extracted image, respectively. This therefore measures the distance between the centrelines. The centrelines are obtained via skeletonisation of the full road objects (Section 2.7). The calculation of centrelines depends on the full extent of the objects, including their edges. Therefore, the edge-based and skeleton-based PFOM measures are not independent, but may be interpreted to measure different aspects of the extraction.

2.12 Conclusion

This chapter provides a theoretical foundation for the remainder of this dissertation. Image analysis requires an image to be rigorously mathematically defined. This chapter defined binary, greyscale, colour and multispectral images mathematically, and defined fundamental concepts related to images, such as pixel connectivity. The theoretical concepts required for each step of the road extraction algorithm were also provided, namely image filtering, the Normalised Difference Vegetation Index, image thresholding, mathematical morphology, image thinning, measuring linearity of image objects and certainty using compactness and elongation. The crux of this algorithm is the Discrete Pulse Transform applied via the Roadmaker's Pavage, which was discussed at the hand of a detailed example in the context of images. Finally, relevant theory for accuracy assessment was discussed, including per-pixel measures and measures based on road characteristics, namely inclusion completeness and correctness, and Pratt's Figure of Merit.

The algorithm is developed step-by-step in Chapter 3. Reasons are given for the chosen steps and the sequence of these steps, and parameter choices are discussed. The techniques for accuracy assessment are discussed in an applied setting. Chapter 3 practically implements the theory given in this chapter.

Chapter 3

Road Extraction Algorithm

This chapter describes the linear object extraction algorithm developed for the detection of informal roads in multispectral satellite images. The algorithm is implementable using free open source software. Various techniques exist that are capable of extracting formal roads [90, 136]. However, no informal road extraction technique which can be implemented freely currently exists. In light of the particular application, the extraction method should rely on the characteristics possessed by informal roads and take the circumstances of these roads into account. Incorporating geometric and structural properties instead of relying on spectral information decreases the risk of confusion with spectrally similar features [81, 74]. Other informal road characteristics also need to be considered. For instance, centreline extraction has been used in several recent methods. This fails when the road widths change suddenly [70]. Also, methods relying on clearly defined road boundaries will fail, such as the multiscale method in [32] or approaches incorporating edge detection such as that of [81]. Uniform colour of roads is another assumption that is used by recent methods [3], but is not met by informal roads. Li et al. [71] developed a method with adaptable parameters which can be modified depending on the road types and environments in the data. It also removes buildings, shadows and vegetation. However, in its current state this algorithm requires the use of proprietary software for implementation, and cannot be used under financially constrained circumstances.

An approach which overcomes the above road extraction challenges is desired, and should additionally be implementable via open source software. The technique developed herein is a flexible method that can be parametrised to suit various road and environmental characteristics. Vegetation and paved surfaces are reliably removed using near-infrared (NIR) data, an advantage over many recent techniques that have made use of optical VHR data only [91, 141, 146]. The Discrete Pulse Transform (DPT) [108] and shape measures are used to identify road objects in the remaining bare soil areas based on their geometry.

The chapter proceeds as follows. Section 3.1 provides an overview of the algorithm. Sections 3.2 to 3.5 describe the extraction technique, resulting in identified road objects. Uncertainty quantification is discussed in Section 3.7, and Section 3.8 explains the methods used to assess the accuracy of the results, given reference data.

3.1 Algorithm Overview

Figure 3.1 gives an overview of the process. The original multispectral image is subjected to a preprocessing step, during which vegetation and paved areas are removed. This results in a masked greyscale image containing regions of possible roads. The DPT is applied to the masked greyscale image, identifying image objects at various scales. The pulses are filtered by size to remove leakage, and then by elongation and compactness to identify linear pulses with a road-like shape. These road-like pulses are then dilated adaptively. Finally, the dilated pulses are once more filtered by compactness and elongation to ensure their linearity. This results in the final roads. Measures based on the elongation and compactness of the road objects are calculated to quantify the uncertainty of the extraction. The final roads are compared with reference data to evaluate the extraction accuracy. When implementing the algorithm to detect such roads, reference data will usually not be available. The evaluation step is not part of the algorithm to obtain the roads, but is used to assess the algorithm accuracy. The following sections develop the algorithm.

3.2 Preprocessing Algorithm

Preprocessing is the first step of the algorithm. During this step, the image is processed to isolate bare soil areas that potentially contain roads. The input to this step is a multispectral image with colour and near-infrared layers, and the output is a masked greyscale image, displaying the regions of interest that may contain roads. Figure 3.2 gives an overview of the preprocessing step. The original image is thresholded using the Normalised Difference Vegetation Index (NDVI), calculated from the red and near-infrared channels. The original colour image is also converted to greyscale, smoothed, and thresholded. The union of these thresholded binary images is taken, removing vegetation and dark areas. Morphological opening by various linear structuring elements (SEs) is then applied to the resulting binary image, and the opened images are filtered for linear shape via compactness and elongation. The results of these images are intersected, resulting in the final binary mask, which is then used to mask the smoothed greyscale image. This final preprocessed greyscale image is the output of the preprocessing step.

Preprocessing Algorithm

I. NDVI step:

Calculate the NDVI from the red and infra-red bands
of the original image.
Result: A greyscale image with values in the range [-1,1].

II. Smoothing step:

1. Convert the original image to greyscale.
2. Perform adaptive median smoothing on the greyscale image.

Result: A smoothed greyscale image.

III. Thresholding step:

1. Threshold the median smoothed image
   using adaptive mean thresholding.
2. Threshold the NDVI image using thresholds from literature,
   e.g. 0.1 and 0.3 to remove buildings and vegetation.

Result: A thresholded binary image.

III. Shape step:

1. Intersect the thresholded NDVI image
   and the thresholded smoothed image.
2. Filter the connected sets within this image
   by compactness and elongation.

Result: A binary image.

IV. Masking step:

Intersect the binary image from the shape step,
with the smoothed greyscale image from the smoothing step.
Result: The final preprocessed image. A greyscale image with
irrelevant areas masked out, i.e. set to zero.

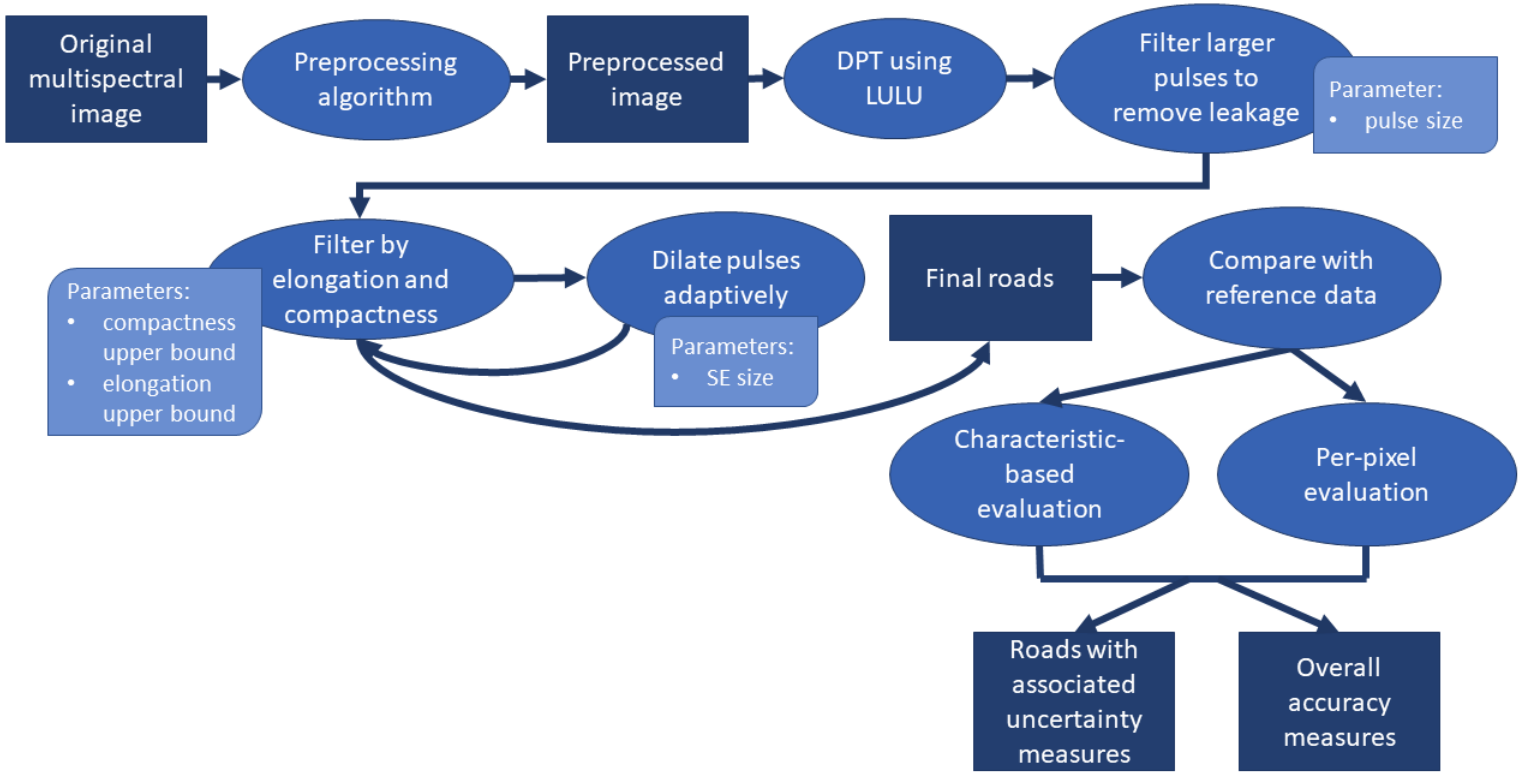


Figure 3.1: An overview of the road extraction algorithm.

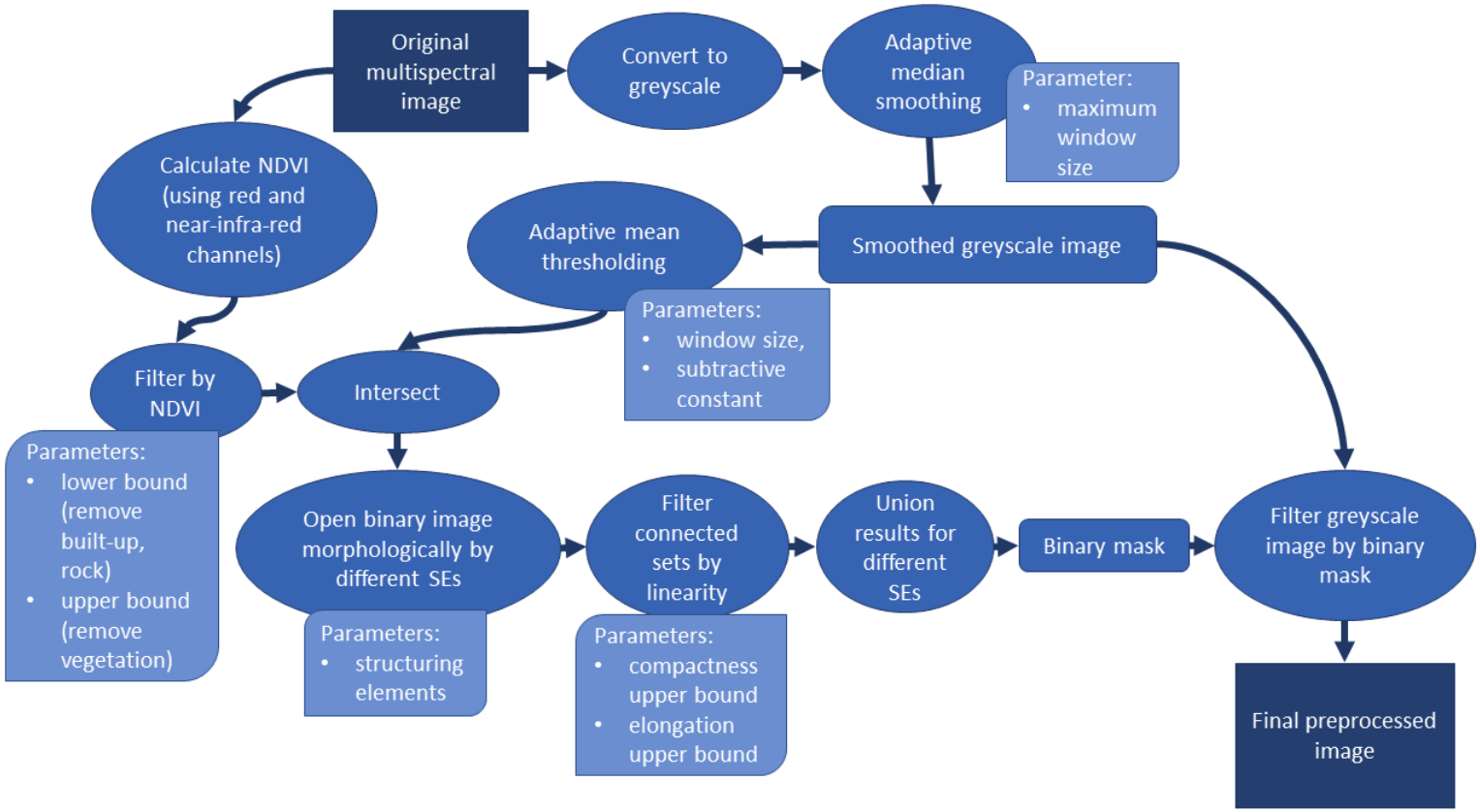


Figure 3.2: An overview of the preprocessing step of the algorithm.

3.2.1 NDVI Thresholding Parameters

The red and near-infrared channels of the original image are used to calculate the NDVI. This is then thresholded, with a lower bound that removes low NDVI pixels, such as rock and paving and an upper bound that removes high NDVI-valued pixels, namely vegetation. The greener the vegetation, the higher the NDVI value will be. In general, bare soil is in the range [-0.1, 0.2] [28]. Grasslands and shrubs are obtained between [0.2, 0.3] [15]. Dense, verdant vegetation is above 0.3. NDVI does vary based on temperature and other local conditions on the ground, as well as on the red and near-infrared wavelengths captured by the satellite. In general, thresholding with a lower bound around 0.1 and an upper bound around 0.3 gave good results, removing rocks, paved surfaces, as well as trees and green grass. The upper bound may be made higher in some cases, to account for grassy areas, especially where narrow roads are bordered by grass.

Figure 3.3 shows different NDVI values for one of the study areas used in the application (Chapter 4). The original image is shown in (a) with the final NDVI thresholded image in (b), which was obtained using an upper bound of 0.3 and a lower bound of 0.1. Images (c)-(d) show those pixels corresponding to an NDVI lower than 0.05, 0.1 and 0.15 respectively. While a threshold of 0.05 in (c) did not identify the paved road, a threshold of 0.15 in (e) identified many of the bare soil areas. The threshold of 0.1 (d) is desirable in this case, since it identifies paved areas and most house roofs, while not identifying bare soil areas. Experiments for the upper bound are shown in (f)-(h). A threshold of 0.25 in (f) identifies too much bare soil, while a threshold of 0.35 in (h) identifies only the most verdant vegetation, and not grassy patches. For this area, an upper bound of 0.3 is suitable.

Similar experiments were done for the other areas in the application. In general, the range 0.1-0.3 was suitable, requiring only minor modification depending on local conditions. Since area 2 was generally dusty, its lower bound was increased to 0.11 to remove more of the dusty formal road. For area 4, the lower bound was decreased to 0.09, since the area exhibited more formal characteristics, as evidenced by the housing in the area. This suggested that the roads would be better travelled and hence harder and more impervious.

3.2.2 Median Smoothing

An adaptive median filter is used to smooth the image and remove spurious noise. Section 2.3 discusses the theory of adaptive median filtering, and provides the algorithm used herein. Pixels are compared with their neighbours to determine whether or not they are noise. The neighbourhood is determined based on an adaptive window size. The relevant parameter is the maximum window size. A window size that is too large over-smooths the image, while a window size that is too small will preserve noise. In general,

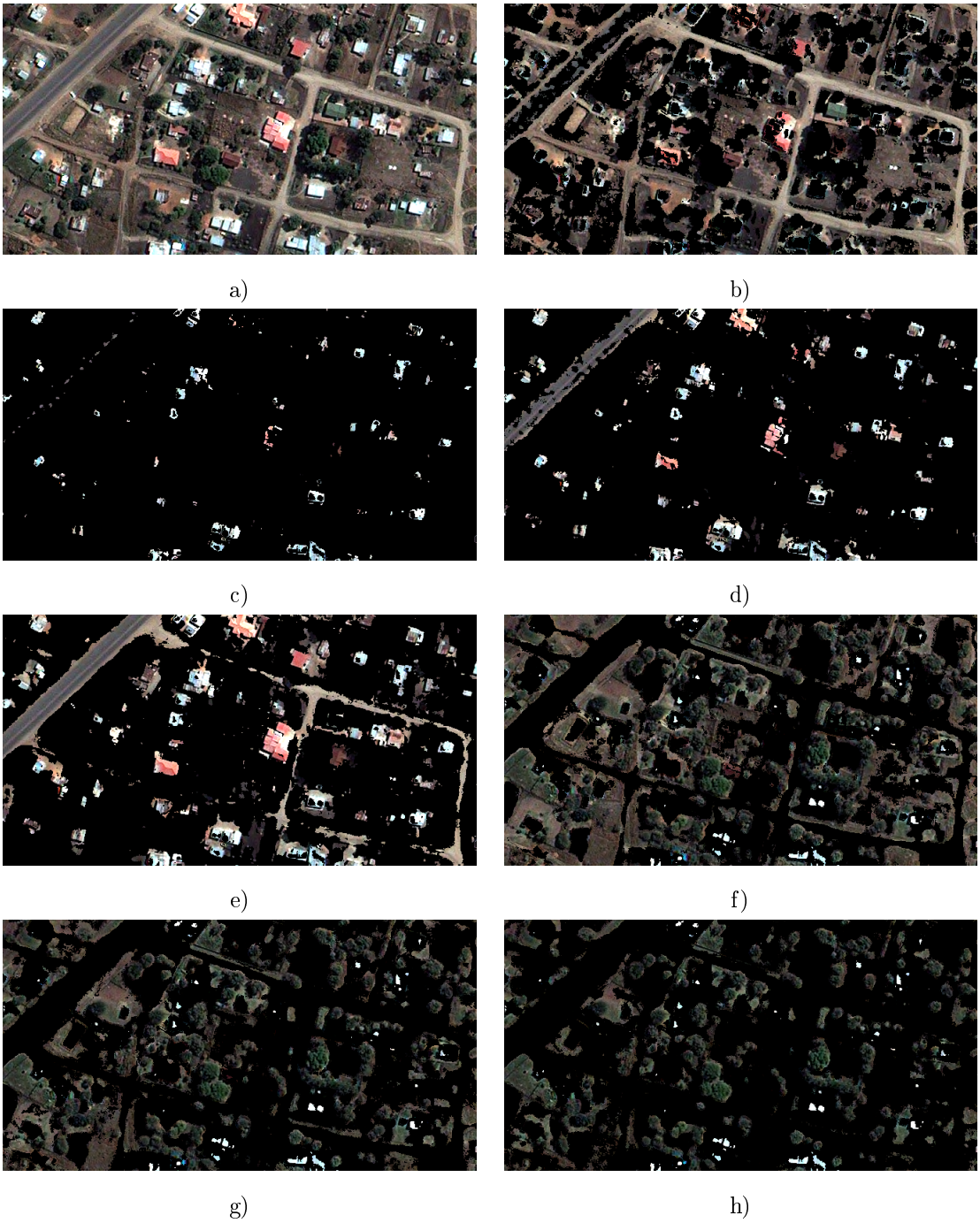
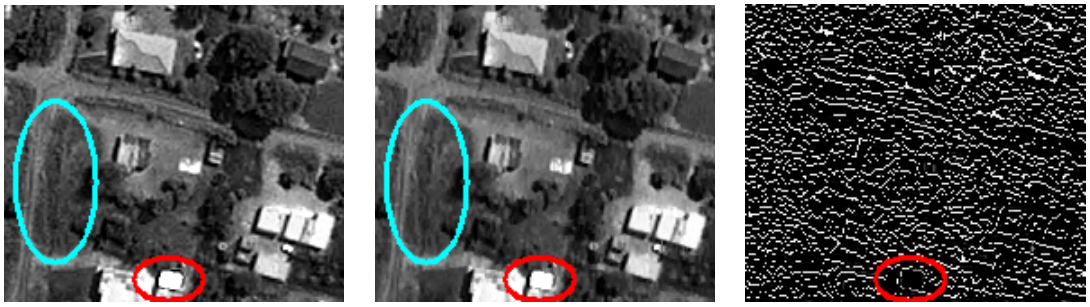


Figure 3.3: Selecting appropriate lower and upper bounds for the NDVI threshold. a) The original image. b) The final thresholded image, lower bound 0.1 and upper bound 0.3. c) All pixels with NDVI < 0.05. d) NDVI< 0.1. e) NDVI< 0.15. f) NDVI> 0.25. g) NDVI> 0.3. h) NDVI> 0.35.

smaller window sizes are preferred, e.g. 5×5 [55] and 7×7 [85, 78]. Herein, a maximum window size of 9×9 pixels was chosen. This was large enough to reduce noise effectively, but small enough to preserve essential detail.

Figure 3.4 illustrates the results of median filtering. Note especially how the irregularities in the field and road (circled in blue) are smoothed, as well as the dark shadows around the bright roof (circled in red). Figure 3.4 (c) shows the difference in the images, showing where smoothing has occurred. Note the lines in the red-circled area, which highlight the smoothing of the dark shadows around the roof in (a) and (b).



a) b) c)

Figure 3.4: Adaptive median filtering applied to a greyscale image. a) The original image. b) The result of median filtering. c) The difference between the filtered and original images.

3.2.3 Parameters for Mean Thresholding of the Greyscale Image

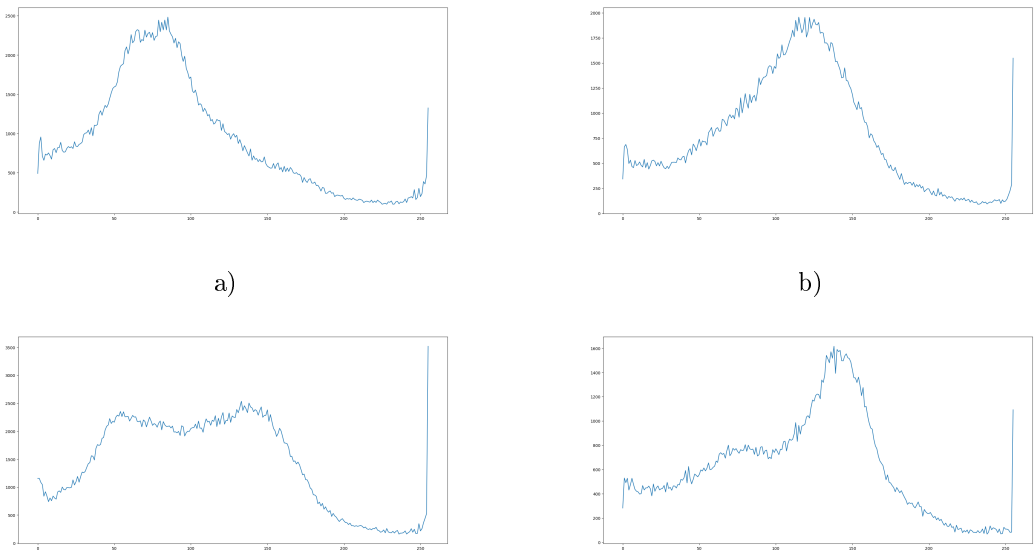
Adaptive mean thresholding is applied to the median smoothed greyscale image to remove dark pixels and preserve possible areas containing roads. This is a local technique, meaning that it compares a pixel with its neighbours, and calculates the threshold based on a specified neighbourhood size. This provides an advantage over global thresholding methods, including the popular Otsu's threshold [100], where a single threshold value is calculated for the whole image, regardless of local conditions. Additionally, Otsu's thresholding works best on bimodal images. Figure 3.5 shows some typical informal road setups, with their corresponding intensity histograms shown in Figure 3.6. The histograms in Figure 3.6(a) and (b) are clearly unimodal. The histograms in (c) and (d) display some bimodality. A threshold that does not rely on bimodal images is preferred, as images containing informal roads are not necessarily bimodal.

The thresholding technique applied herein requires two parameters, namely block size and a subtractive constant, which may be used to reduce the overall intensity of the image. Suitable parameter values were determined experimentally. Figure 3.7 illustrates the effect of the block size parameter. Suitable values depend on image resolution, as well as the result desired. For the images in this application, all of the same resolution, a block size of 71 was preferred. This preserved larger contiguous areas, including roads.

a) b)

c) d)

Figure 3.5: Greyscale images of informal roads. a)-d) show the four study areas (1-4 respectively) used in the application. Figure 3.6 shows the associated image histograms.



a) b)

c) d)

Figure 3.6: Histograms of greyscale images. The histograms in a)-d) correspond to the greyscale images in Figure 3.5(a)-(d) respectively.

a) b) c)

Figure 3.7: The results of various block size parameters for mean thresholding. Block sizes of 11, 71 and 501 are demonstrated in a), b) and c) respectively.

The subtractive constant was chosen to be 0, since any positive value only served to increase the number of white pixels in the image, thereby preserving undesired areas.

3.2.4 Structuring Elements for Morphological Opening

Mean thresholding of the greyscale image results in a binary image, which may be noisy, and contain structures of non-linear shapes. In order to remove noise, morphological opening is applied to the binary image. The choice of SE should accentuate linearity, without being so large that too much of the image is removed during preprocessing. For this reason, eight linear SEs of size 4 are chosen, representing linear structures of varying orientation. Figure 3.8 shows the chosen SEs. Figures 3.9 and 3.10 show other possible SEs of size 6 and 8 respectively. Note the increase in the number of SEs required to account for the various orientations of linear structures: for size 4, 8 SEs are needed. This number increases to 12 for SEs of size 6, and jumps to 20 for SEs of size 8. Figure 3.11 shows the results of morphological opening by these SEs, as well as subsequent filtering by linearity. The result of opening by SEs of size 4 is shown in (a), while (b) uses SEs of size 6, and (c) uses SEs of size 8. Note that much of the image is removed when using larger SEs, since all image objects that do not fit the SE are removed. This leads to information loss in the preprocessing step, excluding areas from further processing. For the sake of retaining information for further processing, SEs of size 4 are used. This has the further advantage of requiring the least amount of processing, due to only needing 8 SEs.

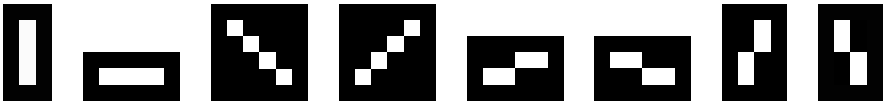


Figure 3.8: Structuring elements used for the morphological opening of the binary thresholded image.

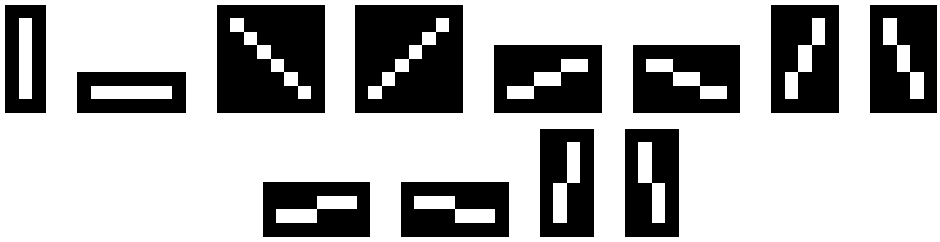


Figure 3.9: Selected structuring elements of size 6. Note that 12 SEs must be used to account for all relevant orientations.

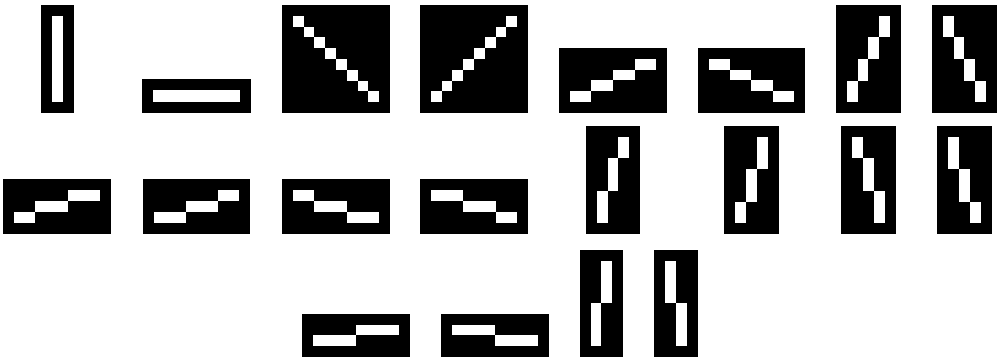
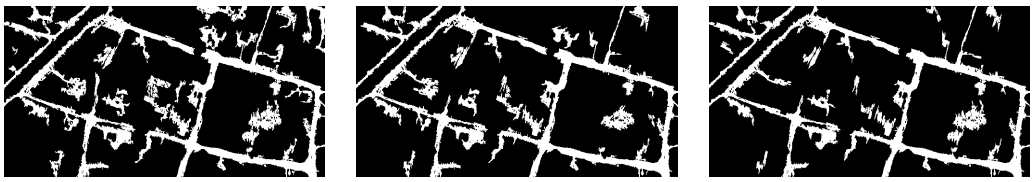


Figure 3.10: Selected structuring elements of size 8. Note that 20 SEs must be used to account for all relevant orientations.



a) b) c)

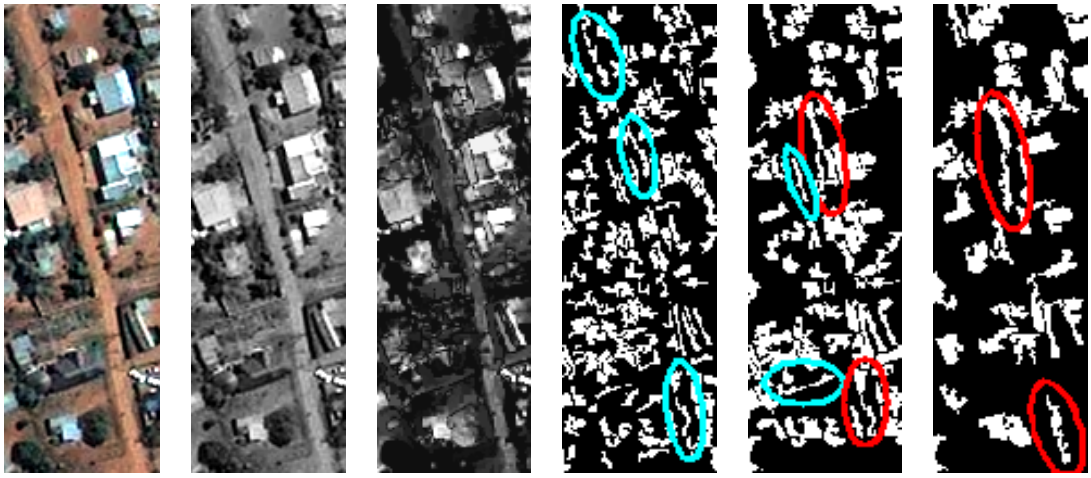
Figure 3.11: Unions of openings by the structuring elements. The results in a), b) and c) were obtained using SEs of size 4, 6 and 8 respectively.

3.3 DPT using LULU

The masked image resulting from the preprocessing step is a greyscale image with the undesirable non-road areas masked out. The remaining desirable areas are considered likely to contain roads, and must be further processed to identify roads. Image objects must be identified, and their size and shape characteristics must be measured to determine if they comply with the geometric characteristics of roads. In particular, image objects must be identified at various scales. A large image object may consist of part of a road network, as well as adjacent yards and non-road areas. Identifying smaller objects (e.g. houses or

individual roads) is therefore necessary. However, small objects that are not roads in and of themselves, may form part of a larger road. Therefore, identifying larger objects is necessary. The DPT provides a multiscale decomposition of the image. It identifies pulses at each scale by removing local maximum and minimum sets of that scale. After the completion of the DPT as implemented via the Roadmaker's Pavage, these pulses are available for fast processing. Their compactness and elongation can therefore easily be measured after the completion of the DPT, and can be used to determine whether or not they are road objects.

Figure 3.12 demonstrates the necessity and usefulness of the DPT for identifying road objects. The colour image in (a) is converted to greyscale in (b) and presented with the DPT decomposition in (c). Pulses of sizes 25 to 50, 50 to 100 and 100 to 150 are shown in (d), (e) and (f) respectively. The red circles in (e) and (f) highlight elongated pulses that may be road objects. The blue circles in (d) and (e) show smaller pulses that would not individually be classified as roads, but which form part of potential road objects. Note that the pulse circled in blue in (e) is not present in (f). This is a potential road object that may not be detected at some larger scales, and illustrates the importance of considering smaller scales.



a) b) c) d) e) f)

Figure 3.12: Result of filtering by various DPT pulse sizes. a) A multispectral image. b) The image converted to greyscale. c) The image reconstructed after the DPT. d) Pulses of size 25 to 50. e) Pulses of size 50 to 100. f) Pulses of size 100 to 150.

3.4 Filtering by Pulse Size

Once the DPT pulses have been extracted from the image, it must be decided which scales are to be analysed. Large pulses correspond to multiple real-world objects, e.g. roads as well as adjacent open

areas. It is desirable to consider smaller pulses. However, small pulses may correspond to real-world objects that are not connected to any roads at all, and very small pulses (e.g. size 1 or 2) may be noise. This problem is solved by intersecting the smaller pulses with the larger pulses, as defined by a threshold size. In this way, only small pulses that are contained within large pulses are considered. This eliminates noise and pulses that correspond to objects not connected to the roads.

The relevant parameter in this step is the threshold size, defining which pulses are considered small and which are considered large. Values for the threshold size depend on the image resolution, as well as the size of roads and other objects in the image. The threshold size was determined experimentally for the areas in the application. Figure 3.13 shows some possible threshold sizes for one of the areas.

The larger the threshold size, the more small objects will be filtered out, as demonstrated in Figure 3.13(a). A smaller threshold size will filter out large pulses, preserving smaller pulses, including those contained within the larger pulses. This is apparent in (c). In general, a threshold size must be chosen that is large enough to remove small objects not connected to roads, but small enough to preserve coherent large image pulses that correspond to real-world objects. Figure 3.13(b) shows the middle ground for this area, obtained using a threshold of 3000.

Due to the varying circumstances in images, including road types and environmental characteristics, suitable threshold size will vary greatly between application areas. There is no single rule for selecting an optimal size. In general, the threshold size must be selected according to the size of objects in the image, as determined both by their real-world area and the spatial resolution of the image. Thresholds of sizes in the range [1000, 3000] were generally found suitable for the areas herein, with a spatial resolution of 0.5m, and exhibiting narrow informal roads. For the case of broad roads ($5 - 6$m on the ground) in an urban setting, thresholding was not required, as the road and environmental characteristics did not lead to excessive small objects.

a) b) c)

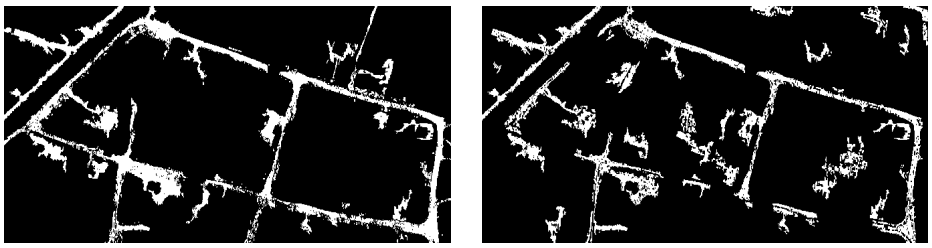
Figure 3.13: Result of filtering by various DPT pulse sizes. The results of using thresholds of size 1000, 3000 and 5000 are shown in a), b) and c) respectively.

3.5 Compactness and Elongation

After the pulses have been intersected based on a certain size, the remaining potential road objects are evaluated based on shape. The shape measures considered herein are compactness and elongation. These measures have the advantage of being able to detect linearity, even when a linear object is curved or twisting, as illustrated in the example at the end of the section.

The reason for first choosing the evaluation scales before considering shape is that small pulses may not adhere to the required compactness and elongation rules, even if they are part of larger road structures. The smallest possible pulse with satisfactory compactness and elongation is of size 32, namely a straight line. In the current step, a second threshold is introduced to ensure that such small pulses are not wrongfully removed. The pulses above this second threshold size are filtered based on shape. Their compactness and elongation is measured, and pulses with high compactness and elongation are removed, as high compactness and elongation indicate that a pulse is not linear.

Figure 3.14 illustrates the effects of first considering shape before considering size, i.e. swapping the current step and the previous step. This demonstrates the importance of first determining suitable evaluation scales before measuring compactness and elongation, as in (a). Considering shape with no regard for pulse size, in (b), may preserve unwanted non-road objects, while leading to the loss of pulses that correspond to road objects.



a) b)

Figure 3.14: Difference in the algorithm when first considering size versus first considering shape. a) Results of determining suitable evaluation scales before filtering by shape. b) Results of filtering by shape before considering size.

According to Li et al. [71], it is true for most roads in the literature that the compactness is lower than 0.3 and the elongation is lower than 0.2. The equations below describe what this means in terms of the area, length and perimeter of the pulse.

$$\text{Compactness} = \frac{2(\sqrt{\pi a})}{p} < 0.3 \quad \Rightarrow a < \frac{0.045}{\pi} p^2 \tag{3.1}$$

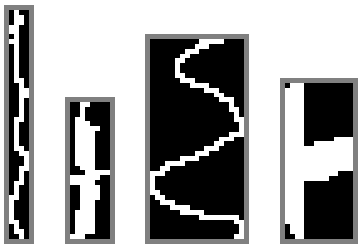
$$\text{Elongation} = \frac{a}{\text{len}^2} < 0.2 \quad \Rightarrow a < 0.2\text{len}^2 \tag{3.2}$$

$$\text{i.e. } a < \min\left\{\frac{0.045}{\pi}p^2, 0.2\text{len}^2\right\},$$

where a = area of the region, p = perimeter of the region, len = length of the region. The area is the number of pixels in the area, the perimeter is the number of pixels on the edge, and the length is the length of the major axis of the smallest ellipse fully containing the area. This is discussed more fully in Section 2.9.

An object with an area that is large relative to the squared perimeter will have high compactness, while an object with a large area relative to the squared length will have high elongation. Given an object has an area that is larger than the specified threshold size, objects with compactness above 0.3 or elongation above 0.2 are not considered road objects by this algorithm. Such objects will consist of smaller pulses which may or may not satisfy these criteria. The next step (Section 3.6) dilates and enlarges small pulses in a way that enhances their linearity. In the final step of the algorithm, the dilated pulses are again filtered by compactness and elongation, and pulses of all sizes are considered.

Figure 3.15 shows some example shapes, with their corresponding compactness and elongation measures given in Table 3.1. According to these measures, the images in (b) and (d) are elongated enough, but possess too high compactness. This means that their areas are too large relative to their squared perimeters, although they is not too large relative to their squared length. The elongated shapes in (a) and (c) may be classified as roads. This demonstrates that these shape measures allow for curvilinear and twisting roads, which provides an advantage over extraction methods that only consider straight lines.



a) b) c) d)

Figure 3.15: Example shapes used to demonstrate the calculation of compactness and elongation.

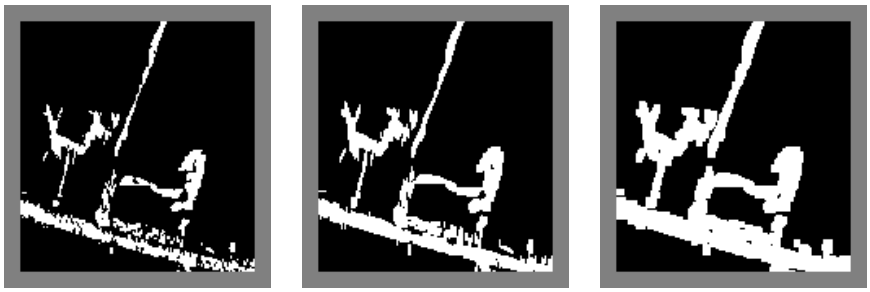
Shape	Compactness	Elongation
a)	0.2306	0.02852
b)	0.4163	0.1301
c)	0.1772	0.0607
d)	0.5183	0.1924

Table 3.1: Compactness and elongation values for the images in Figure ??.

3.6 Structuring Elements for Adaptive Morphological Dilation

After filtering image objects by shape and size, the final step in the algorithm is to dilate the remaining pulses to obtain the road objects. In order to accentuate the linearity of the road objects, as well as to remove specks, linear SEs are required. The same eight SEs of size 4 are used as in the preprocessing step, namely those shown in Figure 3.8. Similar reasoning is applied as in the preprocessing step. Firstly, SEs that are too large may incorrectly remove smaller pulses that may form part of a road. Secondly, increasing the size of the SE also leads to a large jump in the number of SEs required to take account of the various possible road orientations.

This step makes use of adaptive dilation. Instead of applying one SE to the full image, each contiguous image object is investigated to determine which single SE 'fits' best on that image pulse. The process is explained in Section 2.6.2. This is done to dilate the pulses in such a way that their inherent linear structure is enhanced. Figure 3.16 illustrates the difference between using adaptive dilation at this point, versus taking a union of the eight various SEs. The original image in (a) is adaptively dilated in (b), versus having all SEs applied to the full image, and taking the union, as shown in (c). This demonstrates that overall dilations are not sensitive to the shape of the individual pulse, and leads to an unwanted increase in size of the pulse. Adaptive dilation on the other hand takes the shape of the individual image pulse into account, and enhances its linearity in the direction of its orientation. Adaptive dilation is therefore preferred in this step.



a) b) c)

Figure 3.16: Illustration of alternative dilation techniques. a) The original image. b) Adaptive dilation. c) The union of all SEs applied to the full image.

3.7 Uncertainty

In order to quantify the certainty of the final extraction, a certainty measure is calculated for each extracted object. This measure was proposed by [71] and is based on compactness and elongation. The formula is

given below:

$$\text{Certainty} = 1 - \max(1 - \mu_c, 1 - \mu_e),$$

where μ_c and μ_e are calculated by inputting the compactness and elongation, respectively, of each image object into a fuzzy membership function f. The details of this function are given in Section 2.10. The compactness has associated lower and upper bounds a_c and b_c, while the lower and upper bounds of the elongation are denoted by a_e and b_e, respectively. The upper and lower bounds used herein are the same as in [71]. A lower bound of 0 is used for both compactness and elongation. Since the compactness of a road object must be less than 0.3, an upper bound of $b_c = 0.6 = 2 \times 0.3$ is used. The elongation of a road object is less than 0.2, therefore an upper bound of $b_e = 0.4 = 2 \times 0.2$ is used. Image objects with lower compactness and elongation are more likely to be roads, therefore they will have a higher associated certainty, and vice versa.

3.8 Accuracy Assessment

The accuracy of the final extraction is also computed. This is done by comparing the extracted roads with a manually digitised reference, which is assumed to represent the truth. Manual digitisation is the only way to obtain a reference dataset for informal roads, since such roads arise without government knowledge, and are hence not registered in official databases. Per-pixel metrics and two types of accuracy measures based on specific road characteristics are calculated. The inclusion completeness and correctness measures are obtained based on road skeletons (Section 2.7). Pratt's Figure of Merit is calculated based on the distances between the road edges in the extraction and the reference, as well as on the distances between the extracted and reference road skeletons.

3.8.1 Per-pixel Accuracy Assessment

Per-pixel measures are calculated by comparing each pixel in the extraction with the corresponding pixel in the reference image [139]. These measures provide assessment on a per-pixel level by direct comparison of the full extraction and the full reference datasets. The measures are based on true positives, which occur when a pixel is classified as a road in both the extraction and the reference, false positives, when a non-road pixel in the reference is detected as a road in the extraction, and false negatives, when a road pixel in the reference is classified as non-road in the extraction. The pixel-based measures computed in this case are completeness, correctness and quality, as described in Section 2.11.1. The formulae are

restated below:

$$\begin{aligned}\text{Completeness} &= \frac{TP}{TP+FN}\\ \text{Correctness} &= \frac{TP}{TP+FP}\\ \text{Quality} &= \frac{TP}{TP+FP+FN}\\ &= \frac{\text{Compl.} \times \text{Corr.}}{\text{Compl.} + \text{Corr.} - \text{Compl.} \times \text{Corr.}},\end{aligned}$$

where TP is the number of true positive pixels, FP is the number of false positives and FN is the number of false negatives.
Higher values of completeness indicate fewer false negatives, while higher values of correctness indicate fewer false positives. Quality combines the information of both to provide an overall single measure, useful for the sake of making comparisons between datasets and algorithms.

The reasoning for using pixel-based measures in the evaluation is that they provide a conceptually and computationally simple first-glance measure. However, these measures are strict regarding slight deviations. In particular, they are sensitive to small variations and errors occurring at the edges of objects, which is where most of the errors occur [26]. The disadvantage of such strictness is that variations at the road boundaries may lead to a low accuracy, even where the presence of objects was determined accurately. This is a concern in the case of informal road extraction, as some informal roads do not possess visually clear boundaries or edges, but fade into their surroundings. The need for assessing the extraction accuracy in terms of detecting the presence of roads, and allowing for variation at the boundaries, introduces the next section.

3.8.2 Inclusion Completeness and Correctness Based on Full Images and Skeletons

In order to gain an understanding of how well the algorithm captures the presence of objects, with or without accurately determining the boundaries, overall accuracy measures are introduced, which consider the centrelines of roads as indicators of road location. These measures are based on both the full extent of objects as well as their centrelines, which are obtained via skeletonisation according to the method of Zhang [149].

The inclusion completeness metric gives an indication of the false negatives, i.e. how many roads were not detected, while the overall correctness metric gives an indication of the false positives, i.e. how many spurious stretches were detected that are not truly roads. In order to obtain such metrics, the full extraction and reference datasets are used, as well as the skeletonised extraction and the skeletonised reference.

The inclusion completeness is obtained by comparing the full extraction with the skeletonised reference. The intersection shows those sections of the reference that were accurately detected by the extraction. The number of pixels in the intersection is divided by the number of pixels in the reference skeleton. Higher completeness indicates few false negatives. The overall correctness is obtained by comparing the skeletonised extraction with the full reference dataset. The intersection shows those sections of the extraction that accurately coincide with the reference. The number of pixels in the intersection is divided by the number of pixels in the extracted skeleton. High correctness means that few false positives were obtained.

3.8.3 Accuracy Assessment using Pratt's Figure of Merit

Pratt's Figure of Merit (PFOM) is a measure developed for assessing the accuracy of edge detection methods, based on the distance of the edges. Herein it is implemented as a road extraction evaluation metric. It is calculated from the extraction results in two ways, namely based on the road edges and on the road skeletons.

The edge-based PFOM is calculated by applying the formula in Section 2.11.3 to the extracted and reference edges, measuring the distance between these edges. The formula is given below:

$$PFOM = \frac{1}{\max(N_R, N_E)} \sum_{N_E}^{i=1} \frac{1}{1 + \alpha + \delta_i^2},$$

where N_R is the number of pixels of the reference edges, N_E is the number of pixels of the extracted edges, α is a scaling constant usually set equal to $\frac{1}{9}$, and δ_i is the Euclidean distance between the i^{th} detected pixel and the nearest reference pixel. A high edge-based PFOM means that the extraction is accurate in terms of the exact locations of the *road edges*. The edges of the extraction and reference are determined using the Laplacian gradient, which corresponds to the second derivative. The Laplacian is used because it detects noise in the image, which is expected to decrease the extraction accuracy, and must therefore be taken into account [26].

The skeleton-based PFOM is calculated using the pruned road skeletons of the extraction and the reference as inputs to the formula. This measures the distance between the road centrelines. A high skeleton-based PFOM means that the extraction is accurate in terms of the exact locations of the *road centrelines*. Note that a low PFOM does not mean that the road centrelines were not detected by the extraction, but that their exact locations do not overlay with the reference centrelines. Since the calculation of the skeleton involves the locations of the edges, we expect a high centreline-based PFOM if the edge-based PFOM is high, and vice versa. The skeletons are obtained via the efficient thinning method of [149], and pruned to remove spurious lines.

Figure 3.17(a), (b) and (c) show a reference dataset with its edges and skeleton, respectively. Figure

3.17(d) shows a corresponding extraction, with its edges in (e), its skeleton in (f) and its pruned skeleton in (g). Comparing (f) and (g) shows that many spurious lines were removed via pruning, including short lines that branched off from the main road skeletons. While pruning also removed some lines that could be considered true road skeletons, the pruned skeletons still provided an overall more accurate representation of the locations of extracted roads. Using the pruned extraction in the evaluation led to higher accuracy scores than comparing the reference with the extraction prior to pruning.

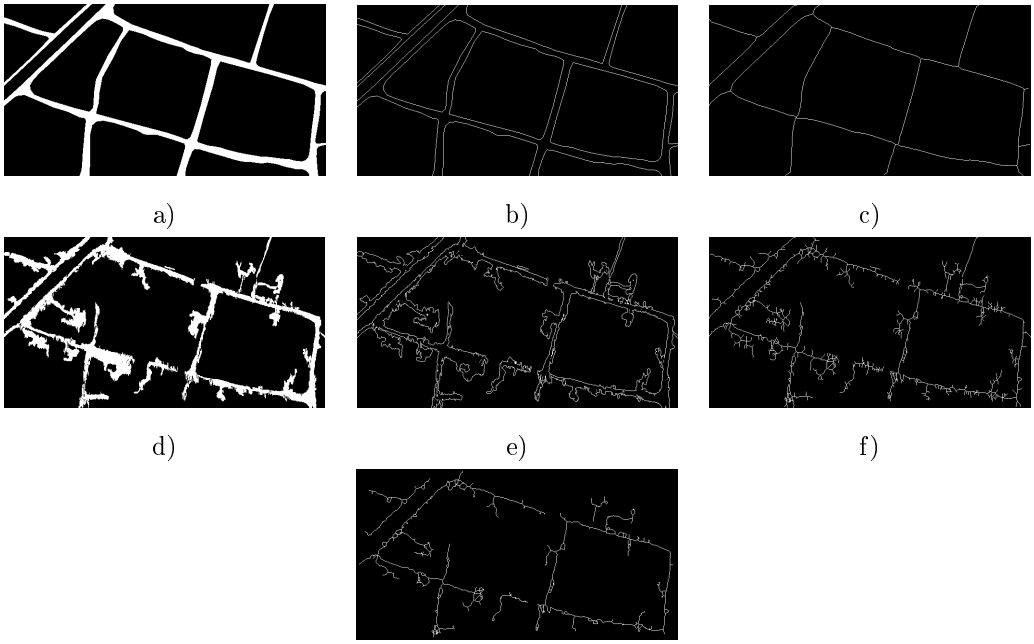


Figure 3.17: Images computed from the results and reference that are used in the PFOM calculations. a) Reference roads. b) Edges of the reference. c) Skeleton of the reference. d) Reference road edges. e) Skeleton of the extracted roads. f) Skeleton of the reference. g) Pruned skeleton of the extracted roads, removing branches.

3.9 Conclusion

This chapter developed the extraction algorithm used for detecting linear objects. The theory discussed in Chapter 2 was placed into the context of specific steps of the algorithm. Sections 3.1-3.6 developed the extraction technique, and the uncertainty measures were discussed in Section 3.7. The accuracy assessment metrics were laid out in Section 3.8. The next chapter will apply the algorithm to real-world images of areas in Gauteng and Northwest Province, South Africa.

Chapter 4

Application to Informal Road Extraction

The road extraction method developed in Chapter 3 is applied to the problem of informal road extraction. Figure 4.1 demonstrates informal roads that are available on Google Maps, highlighting the need for an informal road extraction method.

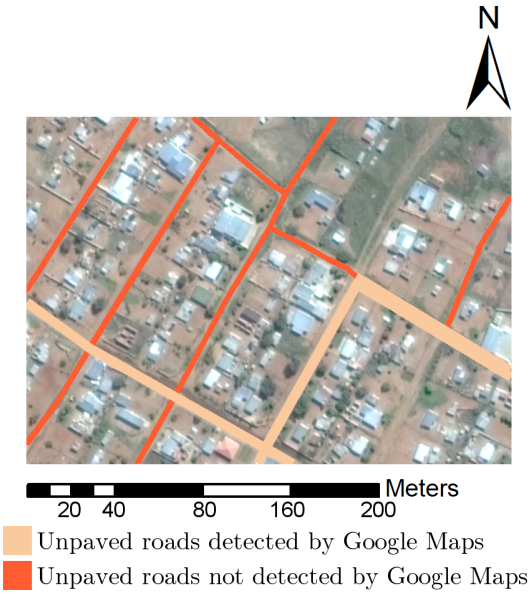


Figure 4.1: Formal and informal roads in Makanyaneng. The image illustrates that informal roads may not be captured by official mapping organisations such as Google Maps. (Imagery ©2017 DigitalGlobe, Map data ©2017 AfriGIS (Pty) Ltd.)

The motivation of this research was to develop a spatial statistical solution for the problem of informal road extraction in remote sensing images. In particular, the solution must be implementable in the developing world at any level of government, using only freely available open source software in the implementation. The only existing informal road extraction technique was developed in [98]. However, this technique makes use of proprietary Definiens eCognition software, which requires a paid license. It is therefore not available to all levels of government in the developing world, due to financial reasons. The state-of-the-art formal road extraction method of [71] has been shown to be effective for detecting informal roads, and was argued to be superior to other formal road extraction methods for this purpose [129]. However, it requires the use of eCognition as well as MATLAB. Therefore, the solution developed herein is the only existing informal road extraction method that is freely implementable at all levels of government in the developing world.

Informal roads occur in much of the developing world, including South Africa. These roads arise to satisfy a need identified on grass-roots level by inhabitants, but not provided by government. Informal roads occur in both formal and informal settlements. Informal settlements may arise without the knowledge of the government, and create a need for infrastructure of which local authorities are unaware. In particular, job opportunities inside informal settlements are limited, and in this context informal roads are critical for providing inhabitants with access to sources of income. Formal settlements may also require additional roads between points of interest. Informal roads connect points of interest within settlements, and connect these settlements with other points of interest. Informal roads may form when residents regularly travel across set routes on previously undeveloped land. The location of these roads is determined by necessity as well as convenience, which is influenced by the distance between points of interest, distance to other means of travel, e.g. formal roads, environmental factors such topography and the presence of vegetation, and safety concerns. In this way, infrastructure is created and used that may not comply with government regulations in terms of safety and capacity.

The formalisation process ensures that these roads comply with regulation, are registered by government, and are safe to use. They may also be paved, if this is deemed necessary by council. Formalising these roads will therefore improve travelling conditions along routes that have already been identified by citizens as being convenient. This will improve access to schools, job opportunities, and other facilities such as clinics and police stations. Additionally, local government will be able to provide services along these roads, as well as infrastructure such as water pipes and electrical lines. Emergency services will also have improved access to these areas. Formal roads are mapped by government and are hence available on GPS, leading to safe navigation for non-residents, who may not be familiar with the area. Sustainable and equitable city growth requires safe and inclusive access for all, as specified by the United Nations Sustainable Development Goal (SDG) 9 on Industry, Innovation and Infrastructure. Target 9.1 specifies the need for affordable and equitable access for all citizens. The goal of Target 9.1 is to "[d]evelop quality,

reliable, sustainable and resilient infrastructure, including regional and transborder infrastructure, to support economic development and human well-being, with a focus on affordable and equitable access for all"[1]. Formalising informal roads is critical for achieving this goal.

Given the availability of satellite imagery, an efficient road extraction method could provide a time- and cost-effective way to obtain information on informal roads. Ideally, the presence, location and extent of informal roads should be provided, reducing the need for physical surveying. This would be especially desirable in areas that pose security risks. While ground-based surveying is unavoidable prior to improving or constructing roads, automatic road detection methods could aid in decision-making, thereby limiting the surveying activity to the last step of the formalisation process, when it has already been decided that the road must be formalised. A method providing accurate information on the presence of informal roads, even if road boundaries are not captured with precision, is then a first step to automating the process of detecting informal roads. It is also a first step in a formalisation process involving more extensive ground-based surveying, since surveying is capable of capturing the exact boundaries of the road to be formalised.

Despite the benefits to obtaining informal road information from satellite images, little work has been done to address this problem. In particular, there is no extant algorithm that provides this information based on free open-source software, and could be available for use in the developing world. A road extraction solution was developed in Chapter 3, which included such an algorithm. Particular challenges of unpaved informal roads have to be considered. Colour and near-infrared information is invaluable for removing vegetation, built-up and paved areas, and identifying bare soil. However, there is colour confusion between roads and non-road areas, since those all consist of bare soil. Roofs made from local soil may have similar colour and reflectance, further leading to confusion. Informal roads may also have heterogeneous colour, e.g. light and dark patches. It is therefore critical to use more than colour information. Geometric information suggests itself, since roads are linear structures.

In order to measure the shape of image objects, it is first necessary to decompose the image into separate objects. The Discrete Pulse Transform (DPT) provides a decomposition of a greyscale image at various scales, by identifying local maximum and minimum sets. The shapes of these sets, or pulses, is measured via the shape measures compactness and elongation. Image objects are filtered accordingly, being excluded from the results if they do not comply with the compactness and elongation requirements of a road. Compactness and elongation are also used to provide a measure of uncertainty associated with the extracted road objects, thereby informing the user of the reliability of the results. The algorithm therefore uses colour, near-infrared and geometric information to detect informal roads and measure the reliability of the identified roads. It is implementable using free open-source software, namely Python[2] and QGIS[3],

[1] https://sdg-tracker.org/infrastructure-industrialization
[2] https://www.python.org/
[3] https://www.qgis.org/

and can therefore be applied by any institution, without the need for costly software licenses. This solution is tailored for the developing world, both in terms of the road extraction and the implementation. The following sections demonstrate how it is applied to extract informal roads in South Africa.

The chapter proceeds as follows. The study areas and data are described in Section 4.1, including expectations of the results. Section 4.2 demonstrates the application step-by-step for area 1. Results for all the areas are given in Section 4.3, and Section 4.4 provides a thorough discussion.

4.1 Study Area and Data

Five study areas are considered and shown in Figure 4.2. These exhibit roads that are good candidates for formalisation. Areas 1-4 (Figure 4.2 (a)-(d)) demonstrate roads that are clearly visible by eye, both in terms of their location and their extent or width. The datasets are all taken from a multispectral VHR Pléiades-1B image with an estimated azimuth angle of 65° and a spatial resolution of $0.5\,m$[4]. The study areas are situated in the north-western parts of Gauteng Province and the south-eastern parts of North West Province, South Africa. The informal settlements in these areas are in many cases starting to formalise and the informal roads are beginning to take on a grid-like structure.

Figure 4.3 shows possible challenges for road extraction. The roads in area 1 are of a light colour set in dark surroundings. Challenges for this area include roads occluded by trees, circled in green, narrow and grassy roads, circled in blue, as well as light-coloured patches of sand, circled in yellow. Areas 2 and 3 exhibit similar scenarios, namely narrow roads in a semi-rural setting, lined with open yards of similar colour to the roads. Both areas also contain some paved roads, which should not be detected by the algorithm. Some of the roads in area 3 are partly vegetated, which may lead to false negatives, while footpaths (circled in grey) may be falsely detected as roads. Similar results are expected for areas 2 and 3. Area 4 shows unpaved roads in an urban setting. A higher extraction accuracy is expected for this area, since there are few open areas that might be misdetected as road, and the roads are relatively broad. The trees occluding parts of the roads may lead to false negatives, while sandy patches may cause false positives.

The next section demonstrates the working of the algorithm with a comprehensive example.

[4]Data provided by and used with permission from the CSIR. Ethics approval number: NAS365/2019.

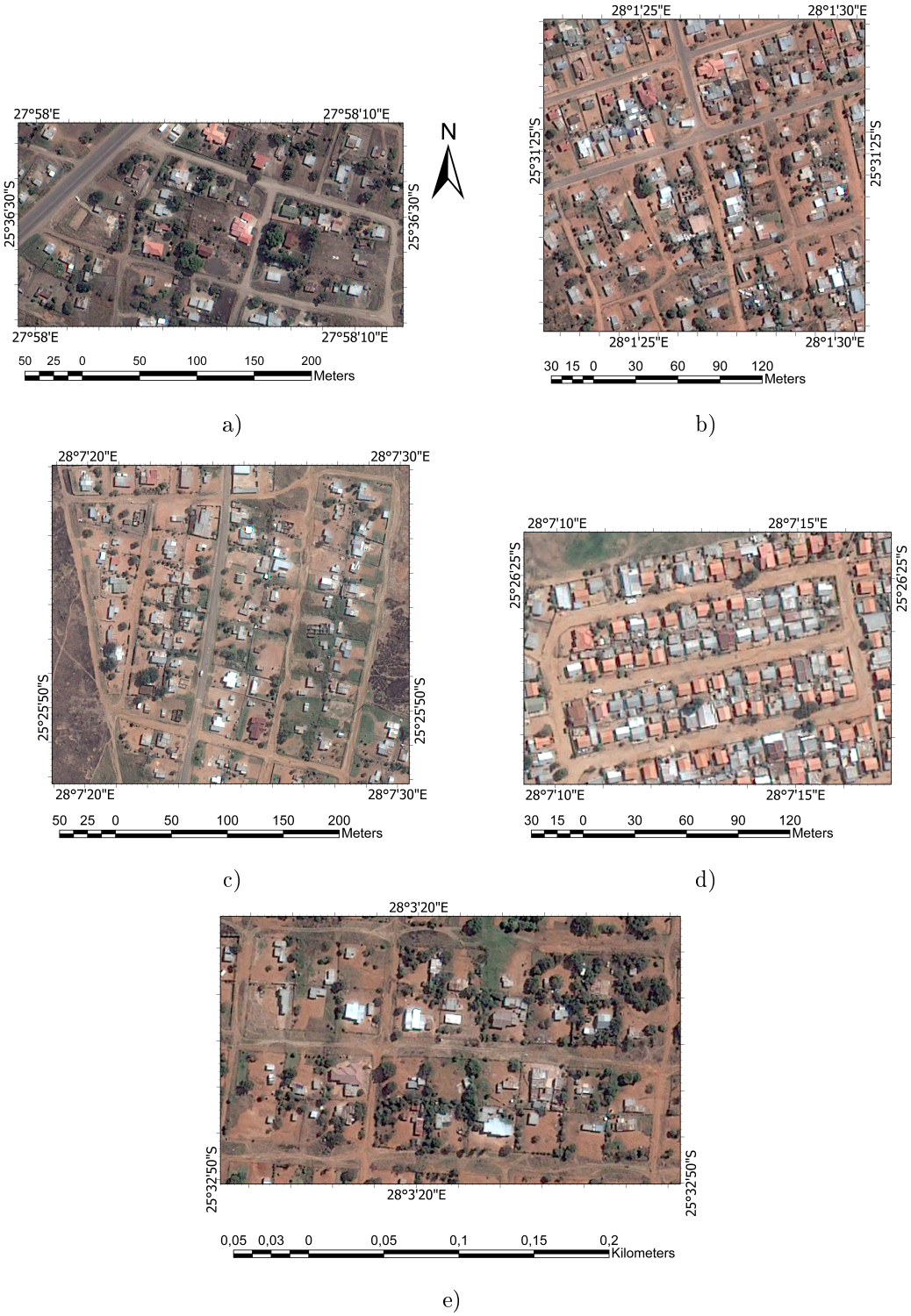


Figure 4.2: Areas to which the solution was applied. a) Hoekfontein, North West Province; b) Mabopane, North West Province; c) Soshanguve, Gauteng Province; d) Soshanguve, Gauteng Province; e) Mabopane, Gauteng Province

Figure 4.3: Possible sources of error for areas 1-4 shown in (a)-(d) respectively.

4.2 Demonstration for Area 1

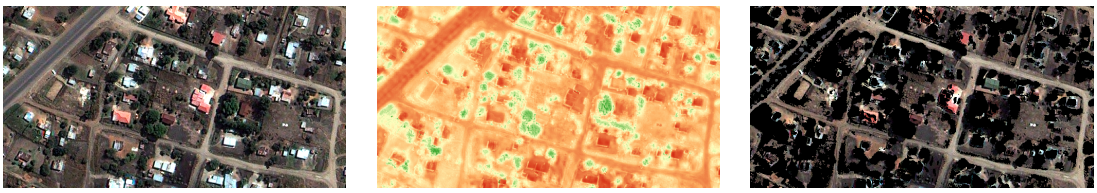
The full process is demonstrated for area 1 (Figure 4.2 (a)). This area is characterised by narrow roads of a bright colour relative to their surroundings. There is also a paved road present in the image. Figure 4.4 shows the first preprocessing steps.

The original image is converted to greyscale and noise is removed via median smoothing as discussed in Section 3.2.2. A maximum window size of 9 pixels was used. This reduced noise without smoothing important details. Next, the image was thresholded via mean thresholding (Section 3.2.3). Suitable values for the parameters were determined experimentally. The neighbourhood size was 71 pixels for this image, while the constant subtractive value was zero.

a) b) c)

Figure 4.4: First preprocessing steps for area 1. The original image in a) is converted to greyscale and median smoothed in b). Mean thresholding is applied in c).

Figure 4.5 shows the preprocessing steps associated with the NDVI. The red and NIR channels of the original image in (a) are used to calculate the NDVI image in (b). It is then thresholded. A lower bound of 0.1 was used to eliminate built-up and paved areas, while an upper bound of 0.3 was used to remove vegetated areas. The result of this threshold is shown in (c). Note that trees, many of the buildings and much of the paved road has been removed.



a) b) c)

Figure 4.5: NDVI preprocessing steps for area 1. The red and near-infrared (NIR) channels of the original image in a) are used to calculate the NDVI image in b). This image is thresholded to remove non-bare soil areas as shown in c).

The results of thresholding on the NDVI image and the mean thresholding applied to the median smoothed image are now intersected, shown in Figure 4.6(a). Morphological opening is applied to the binary image

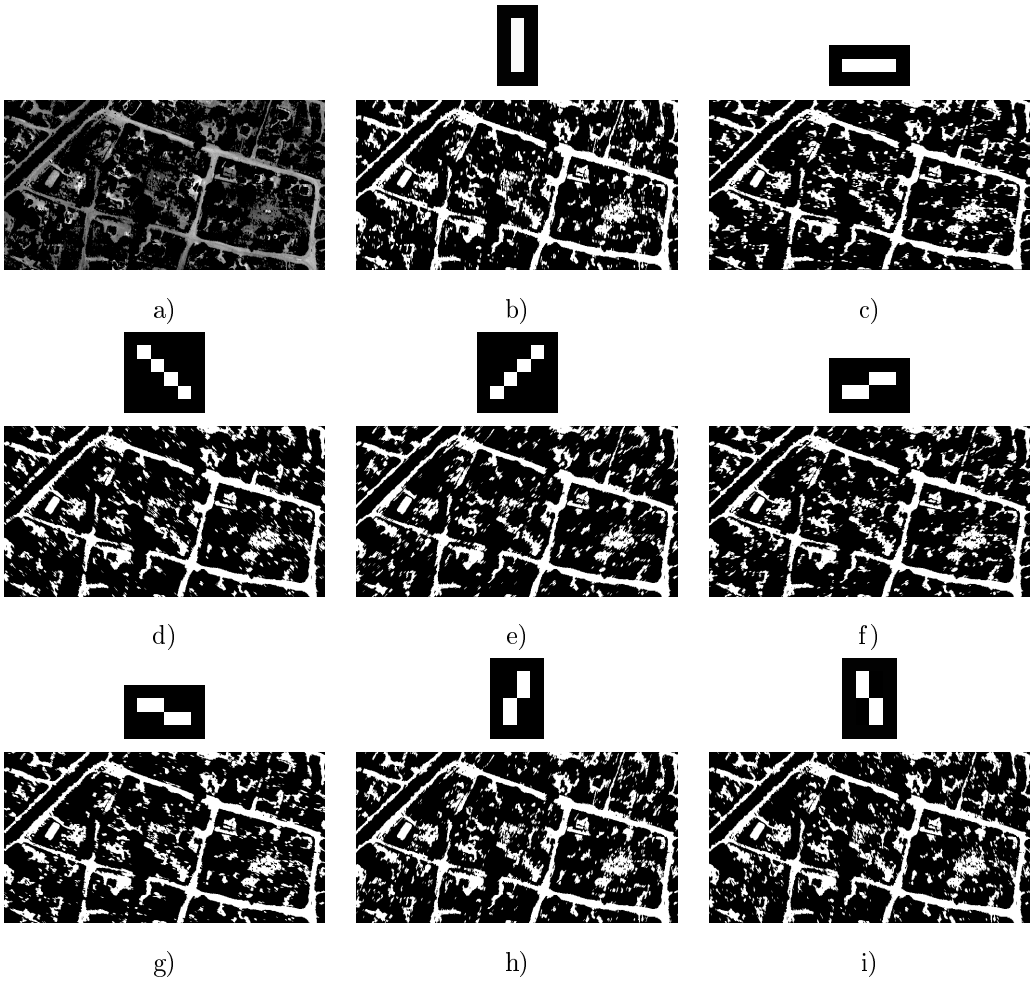


Figure 4.6: Binary openings by linear SEs of different orientations.

using eight structuring elements (SEs) as shown in (b)-(i). These opened images are then filtered by compactness and elongation, removing all contiguous image objects with compactness greater than 0.3 and elongation greater than 0.2. The results are shown in Figure 4.7. The union of these shape-filtered images is then taken, and filtered by compactness, elongation and size. Pulses above a given size threshold (500 for this image) are removed if they have compactness greater than 0.3 or elongation greater than 0.2. This result is shown in Figure 4.8(a). It is overlaid with the median smoothed image in (b), which is used as the input for the next processing step.

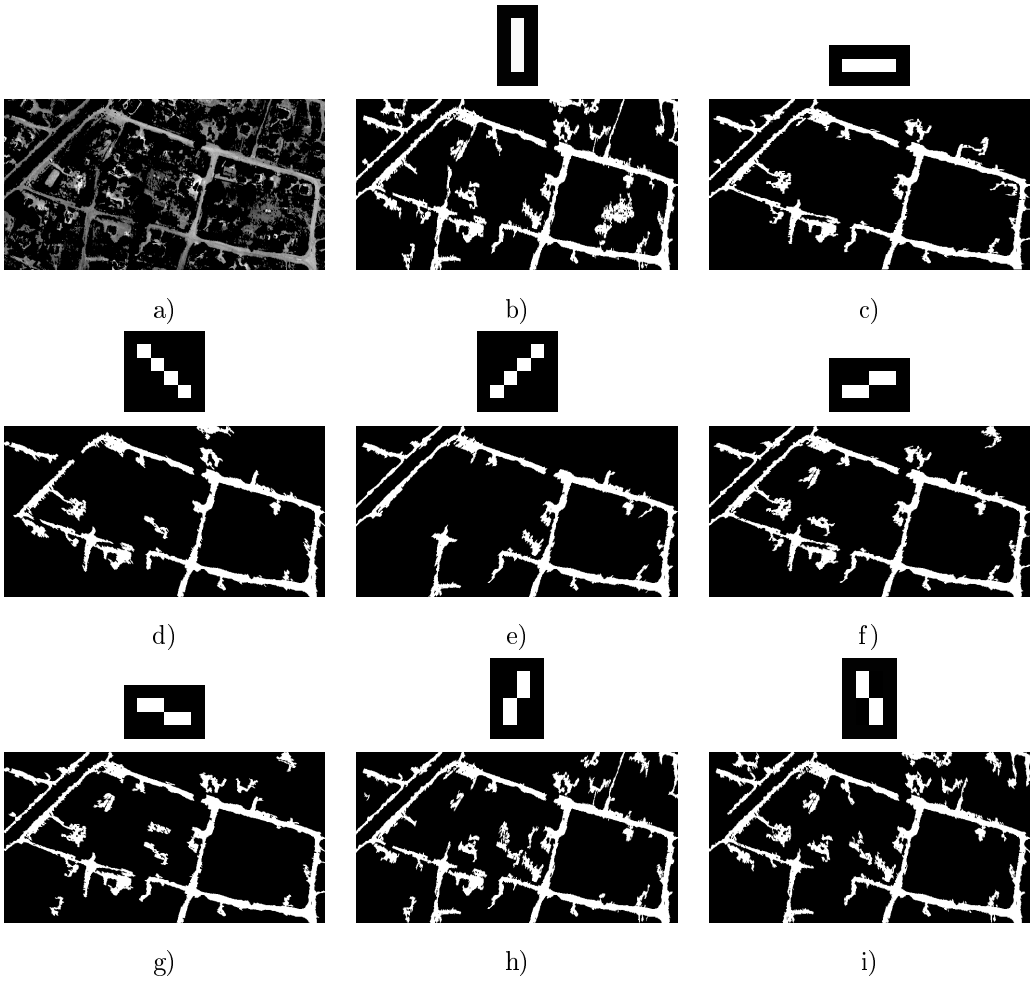
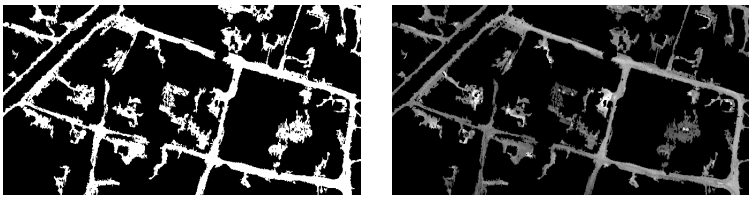


Figure 4.7: Binary opened images with objects of compactness > 0.3 and elongation > 0.2 removed.

The DPT is now applied to the preprocessed image, obtaining image pulses at various scales. Figure 4.9 shows some of these pulses, and illustrates that different image structures are detected at various scales. At small scales, such as in (a), pulses involving all the larger image objects are detected. However, this includes unwanted non-road objects. At large scales (e.g. Figure 4.9(d)), those parts of the road network that are joined, are detected. However, roads smaller than that scale are not detected. This suggests the next step in the algorithm, which consists of intersecting those pulses smaller than a given threshold, with

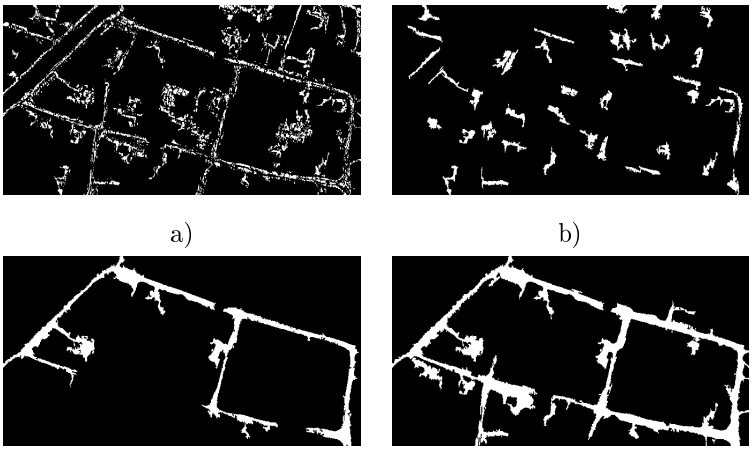


a) b)

Figure 4.8: Final result of the preprocessing step. a) The intersection of the opened, shape-filtered images, filtered by size, elongation and compactness. b) The mask overlaid with the median smoothed image is the input into the next processing step.

the pulses larger than that threshold. The goal of this thresholding is to remove noise and small pulses that are not part of road objects, while preserving image pulses that form part of larger objects, which may potentially be roads.

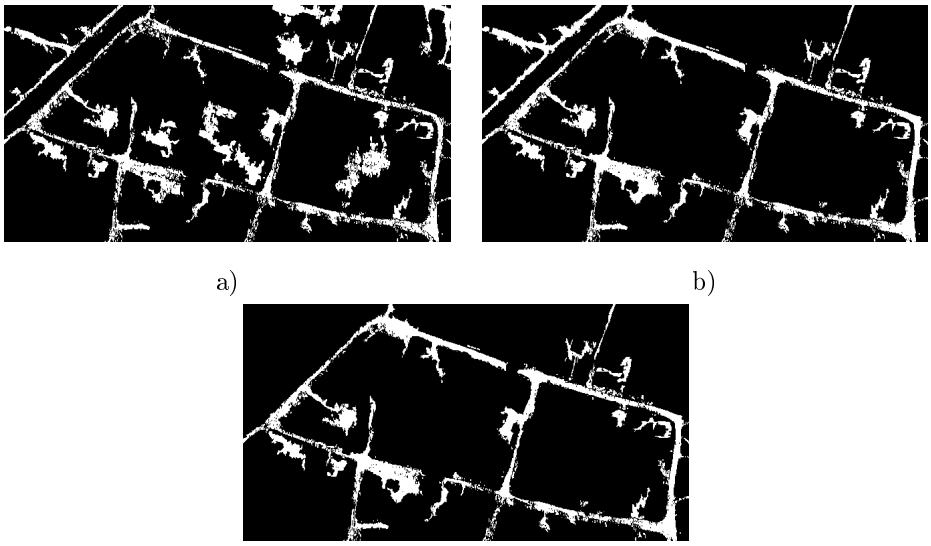
The effects of various threshold sizes for area 1 are shown in Figure 4.10. A smaller threshold size, as in (a) removes large contiguous objects, while preserving the small pulses that form part of those objects. A larger threshold size, as in (c) filters out small image pulses and noise, but may also remove potential road objects if it is too large. A desirable threshold size must be large enough to filter out unwanted noise and small objects, but small enough to ensure the preservation of large coherent image pulses that represent objects in the real world. A threshold size of 3000 was determined experimentally for area 1. The results of this intersecting by this threshold size are shown in Figure 4.10(b). This removed noise and undesirable smaller objects, while preserving larger areas that might contain roads.



a) b)

c) d)

Figure 4.9: Pulses of various sizes extracted using the DPT. a) Pulse sizes 0-100. b) 300-400. c) 8 000-9 000. d) 15 000-16 000.



a) b)

c)

Figure 4.10: Pulses intersected at various possible thresholds. a) Threshold size 1 000. b) Threshold size 3 000. c) Threshold size 5 000.

The image containing these pulses is now dilated adaptively. This is done by dilating the contiguous image objects with one of eight linear SEs of size 4, namely that SE which fits best to the image object (dilates it the least). The SEs are the same as those used during processing, as shown in Figure 4.6. The resulting dilated image is shown in Figure 4.11.

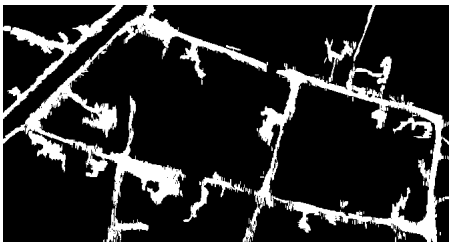
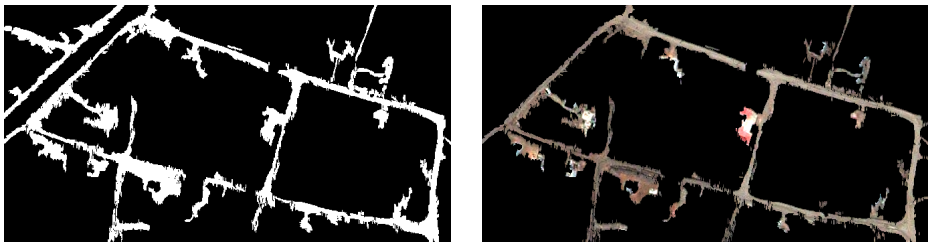


Figure 4.11: Results of adaptive dilation.

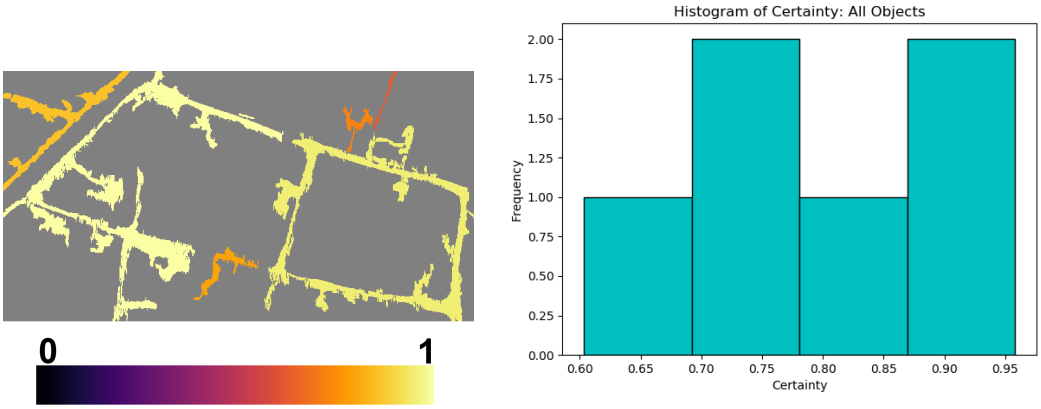
The dilated pulses are filtered once again for linearity, using the same compactness and elongation parameters as before. Pulses of all sizes are filtered in this step. This is because the adaptive dilation enhances the linearity of pulses, increases their size, and may connect previously separate road objects. Therefore, any object that does not satisfy the compactness and elongation requirements after dilation is removed. The final results are given in Figure 4.12(a), and overlaid with the original image in (b) for illustrative purposes.



a) b)

Figure 4.12: Results of the process. a) The binary image is the final output. b) The resulting image overlaid with the original image for illustrative purposes.

The certainty associated with the extracted road objects is quantified via the certainty measure defined in Section 2.10. This measure is based on the compactness and elongation of the road objects. Figure 4.13(a) illustrates the certainty. Lighter road objects are more certain to be roads based on their compactness and elongation, while darker objects are less certain. Figure 4.13(b) gives a histogram of the certainty distribution. This demonstrates a fairly even distribution of certainty between 60% and 96%. The certainty results are given in Table 4.1. The average certainty per image object was 79%. All road objects were more than 50% certain to be roads. The minimum certainty observed was 60%, while the maximum was 96%.



a) b)

Figure 4.13: Certainty associated with extracted road objects. a) The extracted road objects colour-coded according to certainty. The lighter the colour of the image object, the more certain it is to be a road, as measured its by compactness and elongation. b) Histogram of extraction certainty, showing the observed frequencies of the certainty of extracted road objects.

To quantify the extraction accuracy, extracted roads are compared to reference data, which was obtained via manual digitisation. Figure 4.14(a) and (b) show the extracted results and the reference roads,

	Minimum	Mean	Maximum
Certainty	60.35%	79.36%	95.80%

Table 4.1: Certainty measures for area 1.

respectively, overlaid with the original image for illustration. The binary images are compared at pixel-level, at edge level and at skeleton level to obtain various accuracy metrics as discussed fully in Section 3.8. The results are discussed in Section 4.3.

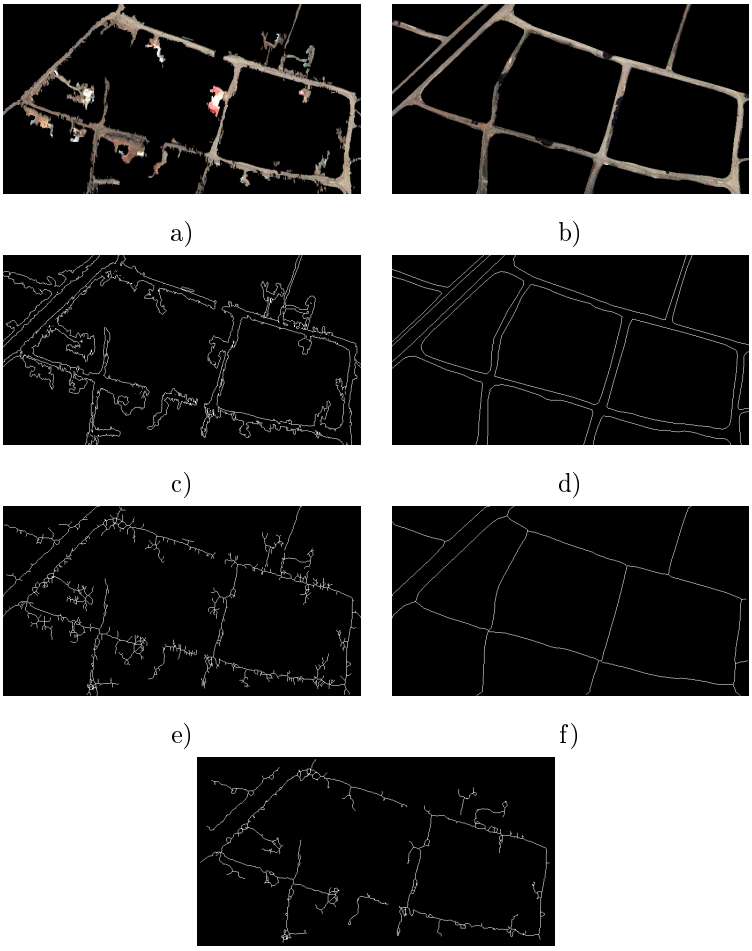


Figure 4.14: Comparison of results and reference. a) Extracted road objects (overlaid with the original image for illustrative purposes). b) Reference roads. c) Extracted road edges. d) Reference road edges. e) Skeleton of the extracted roads. f) Skeleton of the reference. g) Pruned skeleton of the extracted roads, removing branches.

4.3 Results

The results of the extraction are displayed in Figure 4.15, showing those parts of the roads that were detected correctly, as well as false positives and false negatives. The results for inclusion completeness, pixel-based accuracy and Pratt's Figure of Merit (PFOM) are given in Tables 4.2, 4.3 and 4.4 respectively. Results were calculated based on reference data. Due to the lack of any other implementable informal road extraction methods, no direct comparison could be made on the same datasets. However, the results compare favourably with those obtained for similar areas in [98], as well as those obtained in [71], using a similar multi-step algorithm. These will be further discussed in Section 4.4.

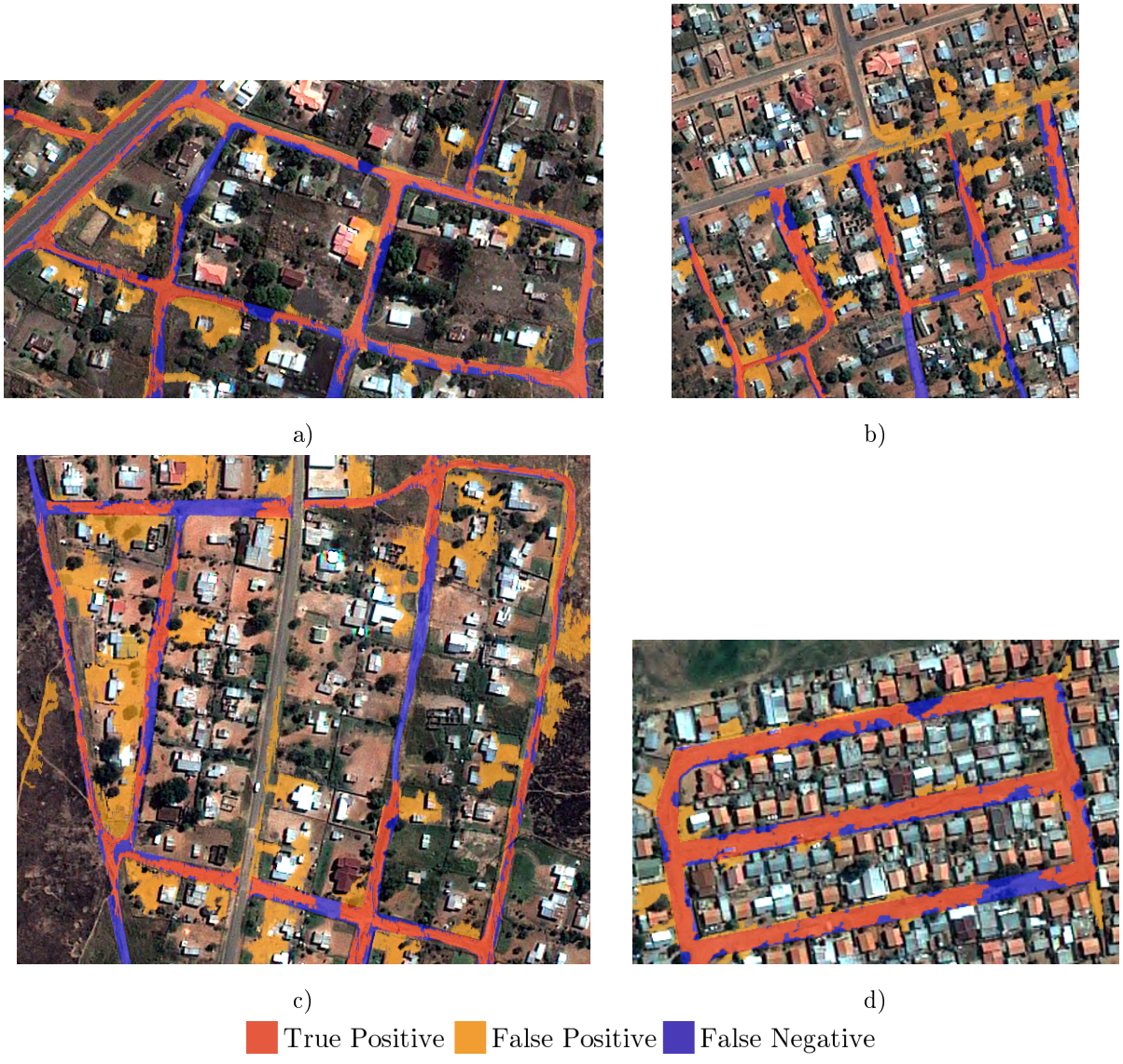


Figure 4.15: Results of the extraction for areas 1-4 shown in a)-d) respectively.

The inclusion completeness and correctness are given in Table 4.2. Comparing the road centrelines as determined via skeletonisation with the full extraction, indicates that 80.75% of the reference road centrelines are contained within the boundaries of the extracted roads for area 1. This means that most of the roads were detected successfully in terms of location. The correctness metric of 66.10% gives the percentage of extracted centrelines within the reference boundaries. This indicates that false positives were observed. The per-pixel measures quantify accuracy in terms of overall extent. The completeness figure of 66.93% indicates that over two-thirds of the road pixels were accurately captured by the extraction. The per-pixel quality corresponds closely to the PFOM metric based on road edges, indicating that the extent was captured with 45.33% accuracy. The PFOM metrics based on road skeletons measure the alignment of extracted and reference centrelines, namely 47.60%. While the presence of roads was therefore generally detected correctly, challenges were experienced at the edges. This is to be expected given the imprecise boundaries of unpaved roads, as well as bare soil areas adjacent to roads.

For area 2, the inclusion completeness signifies that 75.39% of road centrelines were detected successfully. However, the correctness score of 46.96% is considerably lower than for area 1, and indicates that a larger number of false positives was observed. Figure 4.15(b) shows that this is the case. In particular, parts of the paved roads were falsely identified as informal roads. The per-pixel completeness of 61.56% is relatively high, supporting the evidence that the presence of informal roads was mostly detected correctly. However, the correctness score is less than 50%, and the per-pixel quality is 34.11%. As in the case of area 1, the edge-based PFOM is very close to per-pixel quality. These measures indicate that the road edges were not detected with a high accuracy. This is largely due to the imprecise nature of informal road boundaries, as well as the presence of spectrally similar areas such as yards adjacent to roads. The skeleton-based PFOM of 33.50% further indicates that the centrelines of the extracted and reference roads were not aligned well. Since the process of skeletonisation considers the whole image, including the edges, this result is expected.

Although the circumstances of area 3 were similar to those of area 2, the particular challenges of area 3 led to worse accuracy experienced overall. Area 3 exhibited dusty areas along roads, and contained roads that did not have clear edges, but faded into their surroundings. The inclusion completeness was nearly 69.45%, indicating that in spite of these challenges, the presence of most roads was correctly identified. However, the inclusion correctness (44.81%), per-pixel quality (30.46%) and skeleton-based PFOM (32%) were the lowest of all areas. A great deal of false positives and negatives was experienced for this area. Road edges were also not detected accurately.

Area 4 presented the best results overall, as predicted a priori. Visual results (Figure 4.15) show that all roads were detected, including roads exhibiting a curvature. This is corroborated by the very high inclusion completeness (91.61%). The correctness (also 91%) confirms the visual result that no non-road stretches were detected as roads. The per-pixel completeness shows that over 75% of road pixels

were correctly identified, while the per-pixel correctness of 80.90% again confirms that few false positives were observed. The per-pixel quality (65.50%) was the highest observed for all areas by over 20%. The difference between this metric and the edge-based PFOM was the highest of all areas, with an edge-based PFOM of only 53.82%, signifying that road boundaries were not detected with complete accuracy. As consequently expected, the skeleton-based PFOM is not high (57.88%), though it is higher by over 10% than for any of the other areas.

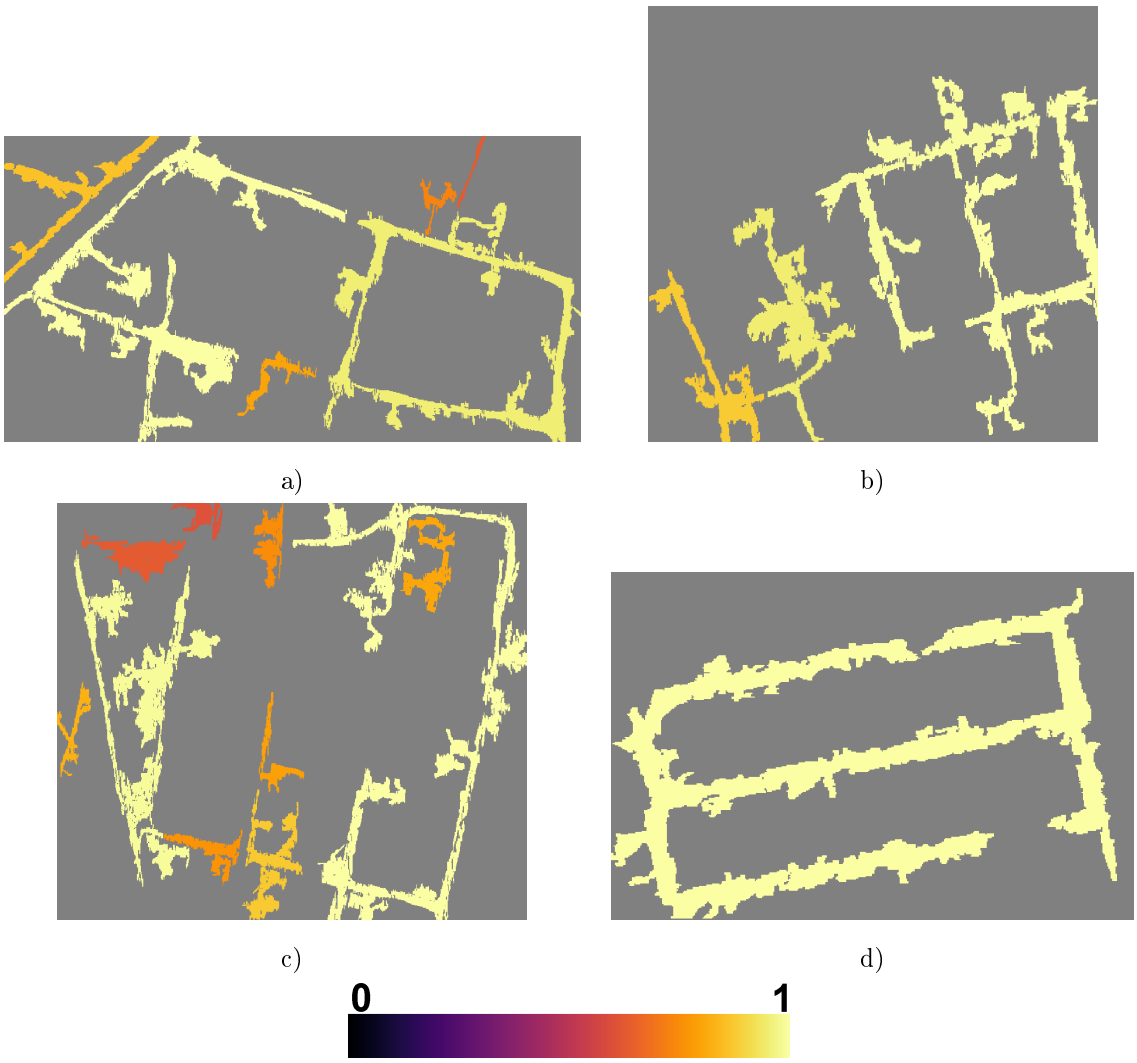


Figure 4.16: Certainty associated with extracted road objects. The lighter the colour of the image object, the more certain it is to be a road, as measured its by compactness and elongation. a)-d) show results for areas 1-4 respectively.

Figure 4.16 shows the certainty associated with the extracted road segments, where certainty was calculated based on compactness and elongation as in Sections 2.10 and 3.7. Table 4.5 gives relevant statistics. Area 3 generally had the lowest certainty scores, except for the maximum, where the lowest was ex-

Area	1	2	3	4
Completeness	80.75%	75.39%	69.45%	91.61%
Correctness	66.10%	46.96%	44.81%	89.45%

Table 4.2: Completeness based on the full extraction and the reference skeleton, and correctness based on the full reference and the extracted skeleton.

Area	1	2	3	4
Completeness	66.93%	61.56%	60.56%	77.49%
Correctness	57.00%	43.33%	38.00%	80.90%
Quality	44.48%	34.11%	30.46%	65.50%

Table 4.3: Per-pixel accuracy measures.

perienced by area 2. All areas had maxima of over 87%, and all means were over 75%. The highest variance was only 1.47% (area 1). For area 4, the entire road network was extracted as one object, with an associated certainty of 87.68%.

4.4 Discussion

This section discusses specific results obtained using this approach, offers a comparison with two similar approaches, provides a general discussion, and suggests possible avenues for future research.

4.4.1 Discussion of Results

The results indicate that the extraction method successfully detects the presence of informal roads in a variety of circumstances, exhibiting various road characteristics. Significantly, the presence of curved roads was detected accurately. However, improvement is needed in picking up the exact boundaries and in reducing false positives.

	Area			
PFOM based on:	1	2	3	4
Edges	45.33%	33.86%	34.72%	53.82%
Pruned skeletons	47.60%	33.50%	32.00%	57.88%

Table 4.4: Pratt's Figure of Merit.

Area	1	2	3	4
Mean	79.36%	89.22%	76.58%	87.68%
Minimum	60.35%	82.35%	57.65%	87.68%
Maximum	95.80%	93.46%	95.63%	87.68%
Percent $> 50\%$	100.00%	100.00%	100.00%	100.00%
Variance	1.47%	0.19%	1.40%	0%

Table 4.5: Certainty statistics for all areas, calculated on all extracted road objects.

Area 4 experienced the most accurate extraction overall, as well as the highest certainty. It is characterised by broad roads in an urban setting. As expected, trees and shadows occluding part of the roads were the main reasons for false negatives. These also led to imprecise boundaries. Contradicting the a priori expectations, the false positive detection was minor, and few sandy areas were misdetected as roads.

The next area in terms of accuracy is area 1, which displays bright roads against a dark background. This area also experienced generally high certainty, although area 2 had higher certainty in some regards. Sources of false positives include sandy areas that were inseparable from road segments, as well as sandy areas that have a road-like shape. It can be argued that the shoulders of the paved road should be considered false positives. However, these were captured in the reference due to their width and visual clarity, implying that they are used for navigation, and require government monitoring. Two notable non-road areas with road-like characteristics are circled in yellow (Figure 4.17(a)). These areas are clearly navigable and used for transport. They were not digitised as roads in the reference due to their short length and the fact that they terminate in yards. However, this demonstrates that the extraction algorithm is capable of picking up any reasonable navigable stretch. It also raises questions regarding the definition of a road, and suggests the need for the incorporation of further characteristics, including connectivity with other roads.

Area 2 was the next most accurately extracted area according to all metrics except the edge-based PFOM, which was higher for area 3. The completeness measures were not much lower than those of area 1, and most roads in the reference were detected. However, the correctness scores were much lower, implying a large number of false positives. This is confirmed by Figure 4.17(b). The high number of false positives was caused by the large section of formal road that was falsely detected as road. The road was not removed during pre-processing due to the layer of dust over the road, as is clear in Figure 4.2(b).

Area 3 experienced the least accurate results. It was the only area with an inclusion completeness below 70% and had the lowest per-pixel completeness at 60.56%, indicating that fewer of the roads were detected than for the other areas. It also experienced the lowest correctness, implying the presence of false positives. Figure 4.17 confirms this. False negatives were largely due to objects and shadows on the roads, as well as

narrow and vegetated roads. As with the other areas, false positives were caused by open areas attached to roads, and open areas with road-like shapes. Due to the NDVI thresholds chosen, some houses were not eliminated during preprocessing, which lead to their being falsely detected as road along with their surrounding open areas. Two notable road-like objects were in fact unpaved road shoulders, which shows similarity with area 1, where the road shoulders were also detected as road. This again demonstrates the need for the incorporation of information regarding the vicinity and orientation of nearby paved roads. An interesting phenomenon in area 3 is the presence of footpaths. These were detected as roads, but were not captured in the reference, since they are not candidates for formalisation. However, it may be argued that this is a desirable result, since the presence of footpaths demonstrates a navigational need experienced by the residents of this area.

4.4.2 Comparison to Existing Methods

The results of this method applied to these areas compares favourably to other road extraction techniques. Nobrega et al. [98] extracted informal roads in similarly challenging images of Sao Paolo, Brazil. Using per-pixel measures, a false positive rate of 76.5% and a false negative rate of 35.5% were obtained, which result in a completeness of 64.5%. Comparing road centrelines within a 9m buffer, the obtained completeness and correctness were 67.86% and 64.48% respectively. Errors and sources of uncertainty were also experienced at the boundaries of these informal roads. Li et al. [71] developed a state-of-the-art adaptable multi-step extraction algorithm, which was applied to two study areas. These contained formal roads rather than informal roads, which led to less uncertainty regarding road boundaries. However, some similar challenges were experienced, including trees and shadows covering roads. For these areas, respective completeness scores of 86% and 58% were obtained, as well as correctness scores of 59% and 45%. This gave overall quality scores of 54% and 35% respectively. Recall that the lowest completeness and correctness scores for the approach herein were 60.56% and 38% respectively, while the highest were 77% and 81%. A simulation study should be done in future to gain a more accurate understanding of how the proposed solution compares to other methods. However, this preliminary comparison suggests that the approach is suitable for detecting informal roads, and competes favourably with existing methods in terms of extraction accuracy.

4.4.3 General Discussion

This extraction approach was mostly successful in identifying informal roads for study areas 1-4, which all contained visually clear roads. However, many informal roads are visually unclear in satellite images. Figure 4.18(a) shows an area containing such unclear roads. Their exact extent and borders are difficult to determine by eye. No reference data was digitised for this area, and the extraction performance is

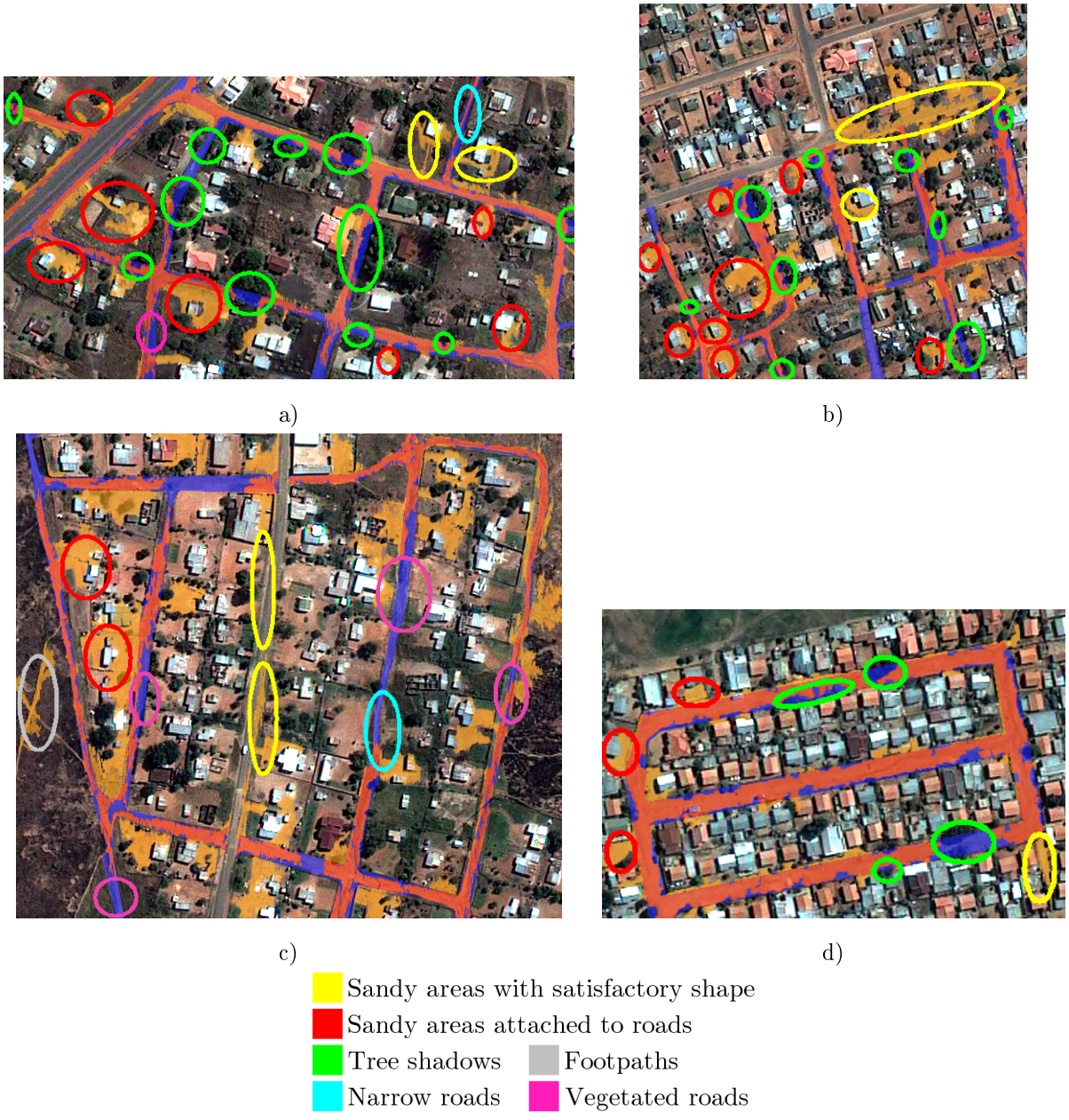


Figure 4.17: Results of the extraction for areas 1-4 shown in a)-d) respectively.

not evaluated. This image is used as an example to demonstrate general challenges for informal road extraction. As shown in (b), this method detects the presence of the roads, even for this challenging area. However, the sandy surroundings make distinguishing road from non-road very difficult. The image contains a high number of open areas that are detected as road. Low certainty is associated with most of the extraction, as shown in (c). This type of area is common in arid parts of South Africa, and generally presents a challenge for informal road extraction from satellite images. In these cases, it is advisable to use road extraction as a first step, and to investigate areas containing roads further by eye, as well as on the ground, to determine their extent.

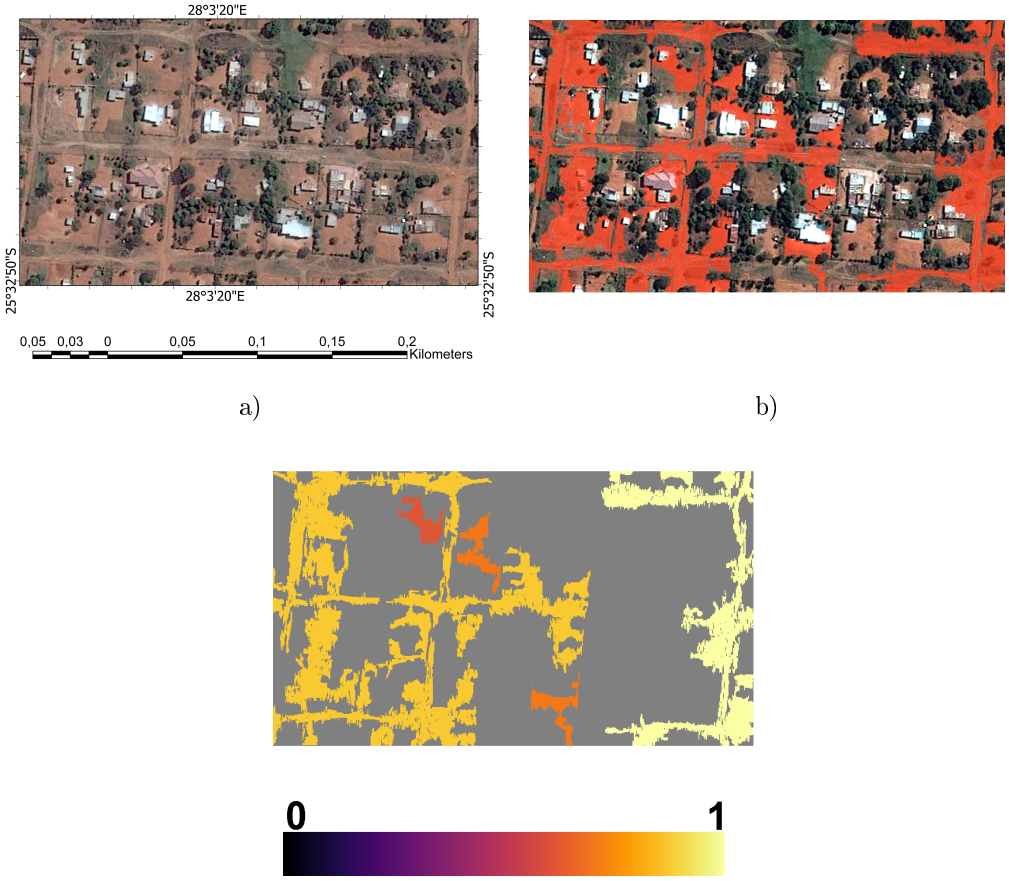


Figure 4.18: Results of the extraction for a challenging area. a) The original image. b) Extraction results. c) Uncertainty.

This technique generally works well for detecting the presence of informal roads. It additionally detects road shoulders, footpaths, and elongated navigable stretches, that can all be used for transport. Extracting footpaths is desirable, as these may grow into more substantial roads over time. In the context of informal settlements, these may also indicate the directions in which the settlement may expand in the future. Detecting footpaths therefore pre-emptively identifies a need for improved infrastructure.

False positives were caused either by non-road areas that are connected to roads, or non-road objects that have a road-like shape in terms of compactness and elongation. False negatives were caused by trees, shadows, and narrow and vegetated roads. The detection of precise extents and edges of roads presented a significant challenge. At this stage, this method can be used to identify the location of informal roads. Further refinement is needed to detect narrow roads, reduce false positives and refine road edges.

4.4.4 Future Research

Avenues of future research are suggested to improve quality assessment, as well as to address the sources of error described in Sections 4.4.1 and 4.4.3, in particular the problems of false positives and imprecise road boundaries. Alternative methods and additional data sources are also investigated.

The solution currently does not address the quality of the input image. It is assumed that preprocessing has been done by the data provider to reduce the effects of atmospheric disturbance, and other sources of error. However, a data validation step could be added to the algorithm prior to preprocessing, in which image quality is assessed. This step could also include a statistical evaluation of background noise.

The solution results included some false positives. Many of these were caused by non-road objects such as yards, that were adjacent to roads on the ground, and consequently visually attached to road objects in the image via leakage. Distinguishing between these non-road-like objects and the roads as separate objects requires leakage handling methods. A possibility is the Pulse Reformation Framework [126], which addresses leakage in the case of shapes of the same type, e.g. circles or rectangles of similar orientation. In order to address leakage in a road extraction context, this framework could be extended to consider linear shapes of various orientation.

False positives were also caused by image objects that had road-like shapes and linearity. These objects were not in general connected to any roads in the network. Determining the connectivity of extracted road objects will help to distinguish true roads from road-shaped objects. This is because roads generally form as part of a network, rather than in complete isolation from other roads or means of transport [93, 14]. Therefore, road-like objects that connect to roads or other road-like objects are more likely to be roads themselves. Road-like objects that do not connect to any other roads are less likely to be roads. Additionally, the connection between true roads could be incorporated into the certainty measurement, as road-shaped objects connected to roads are more certain to be roads. Incorporating the connections between roads into the conceptualisation of a road used for extraction could therefore lead to improved extraction quality. Road network topology has been considered in formal road detection [152, 151, 133], however, many of these methods rely on deep learning. Future research is required to incorporate road topology and connections into informal road extraction.

The challenge of road edges is due largely to the fact that informal unpaved roads do not possess very clear boundaries, but instead fade into their environment. Since these roads are not officially maintained, they may also become overgrown nearer to the edges, or on any other part of their surface that is not travelled on as regularly as the rest of the road surface. In addition, the presence of mixed pixels in the satellite image leads to confusion. Mixed pixels occur when more than one type of land cover, such as vegetation and bare soil, are present in the same pixel as captured by the satellite. This leads to a lack of clarity regarding whether or not the pixel is part of a road. Due to these factors, many of the extracted roads exhibited jagged edges that did not line up precisely with reference roads. While formal roads also experience more errors at the edges [26], formal roads are designed to be of certain width, constrained by boundaries, which are more clearly visible in high-resolution satellite images. Paved roads in particular exhibit a clear land cover difference from their surroundings. The challenge of road boundaries is particularly aggravated for informal roads due to the reasons mentioned above. A possible future solution is to determine the centrelines from an extraction via a thinning process such as skeletonisation, and then expand the roads along the thinned centrelines using a predefined width. The challenge of this method would be to obtain suitable widths, and may require expert knowledge.

The solution quantified the uncertainty of extracted roads. Uncertainty was assessed at the level of the road objects, providing an uncertainty score per road object. Future work could investigate the spatial distribution of uncertainty across an image, as well as providing an overall measure of uncertainty per image.

This solution considers only physical aspects of the roads themselves. Other environmental characteristics, both physical and social, could be considered additionally. These could give an indication of the reason that a road may or may not arise in a given area. The locations of points of interest and existing transport opportunities could be identified. Furthermore, the convenience of travel between such points could be investigated, using covariates such as distance between points, and environmental information such as topography and vegetation, that may influence travel convenience.

Additional data could be used to improve the extraction. For example, elevation data could aid in determining travel convenience of various possible routes. The areas could also be investigated at various points in time. Informal roads may arise quickly to serve an ad-hoc need, and may fall into disuse when alternative routes come into existence. Those informal roads that remain constant over time are more certain to be true roads, that may be good candidates for formalisation. Data from other sources could also be included, such as crime data to determine the safety of certain areas. Height data from LiDAR sensors could eliminate trees and improve building removal. LiDAR technology captures terrain data in a three-dimensional point cloud, providing a model of the surveyed area. It has been used extensively to map terrain models [44, 77], and to differentiate between trees and ground data [147]. It has proved useful for building and road extraction [111, 30]. Radio-wave data, in particular RADAR data has also been

widely used for land cover classification [29, 145], especially where optical conditions were not favourable, e.g. in the presence of clouds [2]. Kulkarni et al. [64] fuse RADAR imagery with multi-spectral data for the purposes of land cover classification. In the context of road extraction, RADAR data could be used to more effectively isolate potential bare soil road areas, by the removal of other land cover types. However, obtaining such data using specialised technology is costly, and therefore presents a financial constraint.

Additional multi-spectral satellite data could be used to train deep learning algorithms, which have become popular in road extraction [153, 135, 142]. However, deep learning methods typically involve large amounts of training data. When increasing the data requirement, it must be kept in mind that all such data must be available and affordable for all levels of government in the developing world. Capturing very high resolution satellite imagery requires sophisticated technology, including satellite launching and monitoring capabilities, that do not exist in much of the developing world. This data must therefore be bought from more developed countries, and is usually expensive. Methods requiring large amounts of such training data are therefore not feasible in the financially constrained conditions that dominate much of the developing world. In any future informal road extraction, it should constantly be considered what the trade-off is between the accuracy and completeness of the extraction afforded by the method, and its cost effectiveness and availability. Ultimately, the aim of the solution developed herein, and any further work building upon it, is to provide an accessible technique to determine the realities of access in the developing world.

In order to address the limited availability of high resolution satellite imagery in the presence of financial limitations, methods for enhancing the resolution of available satellite data should be investigated. Data from e.g. the SPOT satellites from the space agency of France[5] is available free of charge. However, the highest spatial resolution for the multi-spectral channels available from SPOT is 6m, meaning that the smallest roads visually identifiable in these images are 6m wide. Since many informal roads are considerably narrower, higher resolution imagery is required. Techniques such as data fusion [67], super-resolution [65] and side-scaling [43] may be employed to improve the spatial resolution of such data. This is a significant avenue for future research.

[5] https://cnes.fr/fr/

Chapter 5

Conclusion

This dissertation presents a spatial statistical solution for the problem of extracting informal roads in multispectral satellite images. The solution consists of a road extraction technique and an assessment of the accuracy of the results. The extraction technique provides an associated measure of certainty with each extracted object based on its linearity. The accuracy is assessed by comparing extracted results to manually digitised reference datasets. Various aspects of the extraction accuracy are evaluated by considering a variety of assessment metrics, either by evaluating the images at a per-pixel level, or considering specific road characteristics. The solution is implementable using free open-source software and requires colour and near-infrared satellite imagery from one point in time. This is suitable for the context of the developing world, where access to proprietary software and data availability may both be limited.

The research makes the following contributions:

1. The need is identified and addressed for a freely implementable method for informal road extraction from satellite images. A spatial statistical solution is developed for this purpose.

2. The road extraction technique provides associated uncertainty measures rather than hard classification. These give an indication of the reliability of results.

3. These objectives are achieved using freely available software.

4. Extraction accuracy is evaluated using various methods, both by considering images on a per-pixel level and by considering road characteristics. This leads to a richer understanding of the causes for errors, as discussed in Section 4.4.1. In particular, Pratt's Figure of Merit (PFOM) is proposed for the first time as a metric for road extraction accuracy evaluation, based on road edges as well as on road skeletons. The measure is based on the distance between edges that were found in an image

via edge detection, and the corresponding reference edges, that are assumed to represent the truth. Herein, the edges of roads found by the extraction technique were compared with the edges of the roads in the reference datasets. Additionally, the centrelines of the extracted roads were compared with the centrelines of the roads in the reference.

5. The solution is applied to images of rural, semi-rural and urban areas in South Africa. These include roads that are narrow or broad, straight or curved, partly overgrown or occluded by trees and shadows, and roads that exhibit heterogeneous surface colour. The results are favourable and demonstrate that the extraction technique generally detects the presence of informal roads exhibiting various characteristics in differing environmental contexts. The results also compare well with those obtained in similar circumstances by existing methods, namely the informal road extraction technique of Nobrega et al. [98] applied to informal settlements, and the binary partition tree-based method of Li et al. [71], applied to formal roads in urban settings.

6. Sources of uncertainty for informal road extraction are discussed. These include challenges afforded by the characteristics of informal roads, as well as their environments, as represented in satellite images. Future research options are proposed to reduce false positives and to detect road boundaries more accurately. Alternative techniques and additional data are investigated.

In light of the rapid urban expansion in the developing world, the need for information on informal roads can only increase. Solutions are required that address the problem in a time- and cost-effective way, and are implementable in the developing world. This research highlights that the analysis of informal roads in satellite images is a vastly undersaturated research area. This research provides a first step towards obtaining a comprehensive, accurate solution to this real and pressing problem. Such a solution will empower local governments to obtain critical information about informal infrastructure, thereby facilitating sustainable urban growth in the developing world.

Acknowledgements

I would like to thank my supervisor Dr. IN Fabris-Rotelli for her continuous guidance, support, and understanding. Her encouragement helped me to face my fears, conquer self-doubt and ultimately complete this dissertation! I am indebted to my co-supervisors Prof A Stein and Prof P Debba for their valuable advice, in particular Prof Stein's push for sharper, clearer writing. Also, Dr. M Li from the University of Twente is commended for his patient help in previous work that led to the formulation of the problem statement.

I acknowledge the National Research Foundation for the funding provided through the NRF-SASA Crisis in Academic Statistics grant, and STATOMET for the internship opportunity in the Department of Statistics. Without their funding, this research would not have been possible.

Finally, I would like to thank the University of Pretoria for this wonderful opportunity. It is truly a privilege to be a part of its vibrant academic life.

Bibliography

[1] John Abbott. An analysis of informal settlement upgrading and critique of existing methodological approaches. *Habitat International*, 26(3):303–315, 2002.

[2] Saygin Abdikan, Fusun Balik Sanli, M Ustuner, and F Calò. Land cover mapping using Sentinel-1 SAR data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41:757, 2016.

[3] Abolfazl Abdollahi, Hamid Reza Riyahi Bakhtiari, and Mojgan Pashaei Nejad. Investigation of SVM and level set interactive methods for road extraction from Google Earth images. *Journal of the Indian Society of Remote Sensing*, 46(3):423–430, 2017. doi: 10.1007/s12524-017-0702-x.

[4] Ikram E Abdou and William K Pratt. Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proceedings of the IEEE*, 67(5):753–763, 1979.

[5] Lizy Abraham and M Sasikumar. A fuzzy based road network extraction from degraded satellite images. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2032–2036. IEEE, 2013.

[6] Alireza Ajami, Monika Kuffer, Claudio Persello, and Karin Pfeffer. Identifying a slums' degree of deprivation from VHR images using convolutional neural networks. *Remote sensing*, 11(11):1282, 2019.

[7] J Amini, MR Saradjian, JAR Blais, C Lucas, and A Azizi. Automatic road-side extraction from large scale imagemaps. *International Journal of Applied Earth Observation and Geoinformation*, 4 (2):95–107, 2002. doi: 10.1016/s0303-2434(02)00004-1.

[8] PN Anil and S Natarajan. A novel approach using active contour model for semi-automatic road extraction from high resolution satellite imagery. In *2010 Second International Conference on Machine Learning and Computing*, pages 263–266. IEEE, 2010.

[9] Erchan Aptoula and Sébastien Lefèvre. A comparative study on multivariate mathematical morphology. *Pattern Recognition*, 40(11):2914–2929, 2007.

[10] Ben C Arimah. *The face of urban poverty: Explaining the prevalence of slums in developing countries*. Number 2010, 30. Working paper//World Institute for Development Economics Research, 2010.

[11] Xiang Bai, Longin Jan Latecki, and Wen-Yu Liu. Skeleton pruning by contour partitioning with discrete curve evolution. *IEEE transactions on pattern analysis and machine intelligence*, 29(3): 449–462, 2007.

[12] Ziv Bar-Joseph and Daniel Cohen-Or. Hierarchical context-based pixel ordering. In *Computer Graphics Forum*, volume 22, pages 349–358. Wiley Online Library, 2003.

[13] Arpad Barsi and Christian Heipke. Artificial neural networks for the detection of road junctions in aerial images. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(3/W8):113–118, 2003.

[14] Favyen Bastani, Songtao He, Sofiane Abbar, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, Sam Madden, and David DeWitt. Roadtracer: Automatic extraction of road networks from aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4720–4728, 2018.

[15] AK Bhandari, A Kumar, and GK Singh. Feature extraction using Normalized Difference Vegetation Index (NDVI): A case study of Jabalpur city. *Procedia Technology*, 6:612–621, 2012.

[16] Michael Bock, Panteleimon Xofis, Jonathan Mitchley, Godela Rossner, and Michael Wissen. Object-oriented methods for habitat mapping at multiple scales-Case studies from Northern Germany and Wye Downs, UK. *Journal for nature Conservation*, 13(2-3):75–89, 2005.

[17] Charles Boncelet. Image noise models. In *The Essential Guide to Image Processing*, pages 143–167. Elsevier, 2009.

[18] Bernadette Bouchon-Meunier, Radko Mesiar, Christophe Marsala, and Maria Rifqi. Compositional rule of inference as an analogical scheme. *Fuzzy Sets and Systems*, 138(1):53–65, 2003.

[19] Mourad Bouziani, Kalifa Goita, and Dong-Chen He. Rule-based classification of a very high resolution image in an urban environment using multispectral segmentation guided by cartographic data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(8):3198–3211, 2010.

[20] Derek Bradley and Gerhard Roth. Adaptive thresholding using the integral image. *Journal of Graphics Tools*, 12(2):13–21, 2007.

[21] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[22] Thomas J Bruno and Paris DN Svoronos. *CRC handbook of fundamental spectroscopic correlation charts*. CRC Press, 2005.

[23] Alexander Buslaev, Selim S Seferbekov, Vladimir Iglovikov, and Alexey Shvets. Fully convolutional network for automatic road extraction from satellite imagery. In *CVPR Workshops*, pages 207–210, 2018.

[24] James Byrnes. *Unexploded ordnance detection and mitigation*. Springer Science & Business Media, 2008.

[25] Guilherme Cardim, Erivaldo Silva, Mauricio Dias, Ignácio Bravo, and Alfredo Gardel. Statistical evaluation and analysis of road extraction methodologies using a unique dataset from remote sensing. *Remote Sensing*, 10(4):620, 2018.

[26] Guilherme Pina Cardim, EA Silva, and Maurício Araújo Dias. Algorithm development for analysis of statistical accuracy of the extraction of cartographic features in digital images. *Transactions on Machine Learning and Artificial Intelligence*, 2(2):32–47, 2014.

[27] AP Carleer and Eléonore Wolff. Urban land cover multi-level region-based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing*, 27(6):1035–1051, 2006.

[28] Toby N Carlson and David A Ripley. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, 62(3):241–252, 1997.

[29] João MB Carreiras, Joshua Jones, Richard M Lucas, and Yosio E Shimabukuro. Mapping major land cover types and retrieving the age of secondary forests in the Brazilian Amazon by combining single-date optical and RADAR remote sensing data. *Remote sensing of environment*, 194:16–32, 2017.

[30] Minjuan Cheng and Qihao Weng. Urban road extraction from combined data sets of high-resolution satellite imagery and LiDAR data using GEOBIA. In *Integrating scale in remote sensing and GIS*, pages 321–340. CRC Press, 2017.

[31] Dinu Coltuc and Philippe Bolon. Strict ordering on discrete images and applications. In *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, volume 3, pages 150–153. IEEE, 1999.

[32] Idrissa Coulibaly, Nicolas Spiric, Richard Lepage, and Michele St-Jacques. Semiautomatic road extraction from VHR images based on multiscale and spectral angle in case of earthquake. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(1):238–248, 2018. doi: 10.1109/jstars.2017.2760282.

[33] Vladimir Ćurić, Anders Landström, Matthew J Thurley, and Cris L Luengo Hendriks. Adaptive mathematical morphology–a survey of the field. *Pattern Recognition Letters*, 47:18–28, 2014.

[34] RS DeFries and JRG Townshend. NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing*, 15(17):3567–3586, 1994.

[35] Pete Doucette, Peggy Agouris, and Anthony Stefanidis. Automated road extraction from high resolution multispectral imagery. *Photogrammetric Engineering & Remote Sensing*, 70(12):1405–1416, 2004.

[36] IN Fabris-Rotelli. LULU operators on multidimensional arrays and applications. Master's thesis, University of Pretoria, 2009.

[37] Inger Fabris-Rotelli and Stefan J Van der Walt. The Discrete Pulse Transform in two dimensions. 2009.

[38] International Organization for Standardization. Iso 20473:2007 Optons and Photonics - Spectral bands, 2007.

[39] T Fung and W Siu. Environmental quality and its changes, an analysis using NDVI. *International Journal of Remote Sensing*, 21(5):1011–1024, 2000.

[40] John A Gamon, Christopher B Field, Michael L Goulden, Kevin L Griffin, Anne E Hartley, Geeske Joel, Josep Penuelas, and Riccardo Valentini. Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological Applications*, 5(1):28–41, 1995.

[41] G Meera Gandhi, S Parthiban, Nagaraj Thummalu, and A Christy. NDVI: vegetation change detection using remote sensing and GIS–a case study of Vellore District. *Procedia Computer Science*, 57:1199–1210, 2015.

[42] Lipeng Gao, Wenzhong Shi, Zelang Miao, and Zhiyong Lv. Method based on edge constraint and fast marching for road centerline extraction from very high-resolution remote sensing images. *Remote Sensing*, 10(6):900, 2018.

[43] Yong Ge, Yan Jin, Alfred Stein, Yuehong Chen, Jianghao Wang, Jinfeng Wang, Qiuming Cheng, Hexiang Bai, Mengxiao Liu, and Peter M Atkinson. Principles and methods of scaling geospatial Earth science data. *Earth-Science Reviews*, page 102897, 2019.

[44] Cody P Gillin, Scott W Bailey, Kevin J McGuire, and Stephen P Prisley. Evaluation of LiDAR-derived DEMs through terrain analysis and field comparison. *Photogrammetric Engineering & Remote Sensing*, 81(5):387–396, 2015.

[45] Miroslav Hagara and Adam Hlavatovic. Video segmentation based on Pratt's Figure of Merit. In *2009 19th International Conference Radioelektronika*, pages 91–94. IEEE, 2009.

[46] Russell C Hardie and Charles G Boncelet. Gradient-based edge detection using nonlinear edge enhancing prefilters. *IEEE Transactions on Image Processing*, 4(11):1572–1577, 1995.

[47] Christian Heipke, Carsten T Steger, and R Multhammer. Hierarchical approach to automatic road extraction from aerial imagery. In *Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II*, volume 2486, pages 222–231. International Society for Optics and Photonics, 1995.

[48] Christian Heipke, H Mayer, C Wiedemann, and O Jamet. Evaluation of automatic road extraction. *International Archives of Photogrammetry and Remote Sensing*, 32(3 SECT 4W2):151–160, 1997.

[49] Olaf Hellwich and Christian Wiedemann. Object extraction from high-resolution multisensor image data. In *Third International Conference Fusion of Earth Data, Sophia Antipolis*, volume 115, 2000.

[50] Stefan Hinz and Albert Baumgartner. Urban road net extraction integrating internal evaluation models. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY REMOTE SENSING AND SPATIAL INFORMATION SCIENCES*, 34(3/A):163–168, 2002.

[51] Stefan Hinz, Christian Wiedemann, and Heinrich Ebner. Self-diagnosis within automatic road network extraction. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY REMOTE SENSING AND SPATIAL INFORMATION SCIENCES*, 34(2):185–192, 2002.

[52] Peter Hofmann, Josef Strobl, Thomas Blaschke, and Hermann Kux. Detecting informal settlements from QuickBird data in Rio de Janeiro using an object based approach. In *Object-based image analysis*, pages 531–553. Springer, 2008.

[53] Jiuxiang Hu, Anshuman Razdan, John C Femiani, Ming Cui, and Peter Wonka. Road network extraction and intersection detection from aerial images by tracking road footprints. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4144–4157, 2007.

[54] Humor Hwang and Richard A Haddad. Adaptive median filters: new algorithms and results. *IEEE Transactions on Image Processing*, 4(4):499–502, 1995.

[55] Haidi Ibrahim, Nicholas Sia Pik Kong, and Theam Foo Ng. Simple adaptive median filter for the removal of impulse noise from highly corrupted images. *IEEE Transactions on Consumer Electronics*, 54(4):1920–1927, 2008.

[56] Atsushi Iimi, Farhad Ahmed, Edward Charles Anderson, Adam Stone Diehl, Laban Maiyo, Tatiana Peralta-Quirós, and Kulwinder Singh Rao. *New rural access index: main determinants and correlation to poverty*. The World Bank, 2016.

[57] Zhangyan Jiang, Alfredo R Huete, Jin Chen, Yunhao Chen, Jing Li, Guangjian Yan, and Xiaoyu Zhang. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote sensing of environment*, 101(3):366–378, 2006.

[58] Daniel A Keim, Ming C Hao, and Umeshwar Dayal. Hierarchical pixel bar charts. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):255–269, 2002.

[59] A Kirthika and A Mookambiga. Automated road network extraction using artificial neural network. In *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, pages 1061–1065. IEEE, 2011.

[60] Les Kitchen and Azriel Rosenfeld. Edge evaluation using local edge coherence. *IEEE Transactions on Systems, Man, and Cybernetics*, 11(9):597–605, 1981.

[61] Andreas Kleefeld and Bernhard Burgeth. Processing multispectral images via mathematical morphology. In *Visualization and Processing of Higher Order Descriptors for Multi-Valued Data*, pages 129–148. Springer, 2015.

[62] Divyani Kohli, Richard Sliuzas, Norman Kerle, and Alfred Stein. An ontology of slums for image-based classification. *Computers, Environment and Urban Systems*, 36(2):154–163, 2012.

[63] Monika Kuffer, Karin Pfeffer, Richard Sliuzas, Isa Baud, and Martin Maarseveen. Capturing the diversity of deprived areas with image-based features: nThe case of Mumbai. *Remote sensing*, 9(4): 384, 2017.

[64] Samadhan C Kulkarni, Priti P Rege, and Omkar Parishwad. Hybrid fusion approach for Synthetic Aperture RADAR and multispectral imagery for improvement in land use land cover classification. *Journal of Applied Remote Sensing*, 13(3):034516, 2019.

[65] Fayez Lahoud, Ruofan Zhou, and Sabine Susstrunk. Multi-modal spectral image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

[66] Louisa Lam, Seong-Whan Lee, and Ching Y Suen. Thinning methodologies - a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):869–885, 1992.

[67] Charis Lanaras, Emmanuel Baltsavias, and Konrad Schindler. Hyperspectral super-resolution by coupled spectral unmixing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3586–3594, 2015.

[68] Dirk P Laurie. The Roadmaker's Algorithm for the Discrete Pulse Transform. *IEEE Transactions on Image Processing*, 20(2):361–371, 2010.

[69] Dirk P Laurie and Carl H Rohwer. Fast implementation of the discrete pulse transform. In *Proc. Int. Conf. Numer. Anal. Appl. Math*, pages 15–19. Weinheim, Germany, 2006.

[70] Jiayuan Li, Qingwu Hu, and Mingyao Ai. Unsupervised road extraction via a Gaussian mixture model with object-based features. *International Journal of Remote Sensing*, 39(8):2421–2440, 2018. doi: 10.1080/01431161.2018.1425563.

[71] Mengmeng Li, Alfred Stein, Wietske Bijker, and Qingming Zhan. Region-based urban road extraction from VHR satellite images using binary partition tree. *International Journal of Applied Earth Observation and Geoinformation*, 44:217–225, 2016. doi: 10.1016/j.jag.2015.09.005.

[72] Mengmeng Li, Alfred Stein, Wietske Bijker, and Qingming Zhan. Urban land use extraction from Very High Resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122:192–205, 2016. doi: 10.1016/j.isprsjprs.2016.10.007.

[73] Peikang Li, Yu Zang, Cheng Wang, Jonathan Li, Ming Cheng, Lun Luo, and Yao Yu. Road network extraction via deep learning and line integral convolution. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1599–1602. IEEE, 2016.

[74] Runsheng Li and Fanzhi Cao. Road network extraction from high-resolution remote sensing image using homogenous property and shape feature. *Journal of the Indian Society of Remote Sensing*, 46 (1):51–58, 2018. doi: 10.1007/s12524-017-0678-6.

[75] Xiaolong Li, Jian Li, Bin Li, and Bin Yang. High-fidelity reversible data hiding scheme based on pixel-value-ordering and prediction-error expansion. *Signal Processing*, 93(1):198–205, 2013.

[76] Jae S Lim. Two-dimensional signal and image processing. *Englewood Cliffs, NJ, Prentice Hall, 1990, 710 p.*, 1990.

[77] Kevin Lim, Paul Treitz, Michael Wulder, Benoît St-Onge, and Martin Flood. LiDAR remote sensing of forest structure. *Progress in physical geography*, 27(1):88–106, 2003.

[78] H-M Lin and AN Willson. Adaptive-length median filters for image processing. In *1988., IEEE International Symposium on Circuits and Systems*, pages 2557–2560. IEEE, 1988.

[79] Tzu-Chao Lin. A new adaptive center weighted median filter for suppressing impulsive noise in images. *Information Sciences*, 177(4):1073–1087, 2007.

[80] Ruyi Liu, Qiguang Miao, Jianfeng Song, Yining Quan, Yunan Li, Pengfei Xu, and Jing Dai. Multiscale road centerlines extraction from high-resolution aerial imagery. *Neurocomputing*, 329:384–396, 2019.

[81] Weifeng Liu, Zhenqing Zhang, Shuying Li, and Dapeng Tao. Road detection by using a generalized Hough transform. *Remote Sensing*, 9(6):590, 2017. doi: 10.3390/rs9060590.

[82] Ross S Lunetta, Joseph F Knight, Jayantha Ediriwickrema, John G Lyon, and L Dorsey Worthy. Land-cover change detection using multi-temporal MODIS NDVI data. *Remote sensing of environment*, 105(2):142–154, 2006.

[83] Petros Maragos and R Schafer. Morphological skeleton representation and coding of binary images. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1228–1244, 1986.

[84] Petros Maragos and Ronald W Schafer. Morphological filters-part 1: Their set-theoretic analysis and relations to linear shift-invariant filters. Technical report, 1987.

[85] Gary A Mastin. Adaptive filters for digital image noise smoothing: An evaluation. *Computer Vision, Graphics, and Image Processing*, 31(1):103–121, 1985.

[86] Georges Matheron and Jean Serra. The birth of mathematical morphology. In *Proc. 6th Intl. Symp. Mathematical Morphology*, pages 1–16. Sydney, Australia, 2002.

[87] Farid Melgani, Gabriele Moser, and Sebastiano Bruno Serpico. Unsupervised change-detection methods for remote-sensing images. *Optical Engineering*, 41(12):3288–3298, 2002.

[88] JB Mena. Vectorización automática de una imagen binaria mediante k-means y degeneración de la triangulación de delaunay. *Revista de la Asociación Espanola de Teledetección*, 7:21–29, 2002.

[89] JB Mena and JA Malpica. Color image segmentation using the Dempster-Shafer theory of evidence for the fusion of texture. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY REMOTE SENSING AND SPATIAL INFORMATION SCIENCES*, 34(3/W8):139–144, 2003.

[90] Juan B Mena. State of the art on automatic road extraction for GIS update: a novel classification. *Pattern Recognition Letters*, 24(16):3037–3058, 2003.

[91] Zelang Miao, Wenzhong Shi, Alim Samat, Gianni Lisini, and Paolo Gamba. Information fusion for urban road extraction from VHR optical satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5):1817–1829, 2016. doi: 10.1109/jstars.2015.2498663.

[92] Amir Monadjemi, BT Thomas, and Majid Mirmehdi. Experiments on high resolution images towards outdoor scene classification. Technical report, tech. rep., University of Bristol, Department of Computer Science, 2002.

[93] Agata Mosinska, Pablo Marquez-Neila, Mateusz Koziński, and Pascal Fua. Beyond the pixel-wise loss for topology-aware delineation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3136–3145, 2018.

[94] Sahar Movaghati, Alireza Moghaddamjoo, and Ahad Tavakoli. Road extraction from satellite images using particle filtering and extended Kalman filtering. *IEEE Transactions on geoscience and remote sensing*, 48(7):2807–2817, 2010.

[95] United Nations. *World Urbanization Prospects, The 2014 Revision*. United Nations: New York, NY, USA, 2014.

[96] Shing-Chung Ngan and Xiaoping Hu. Analysis of functional magnetic resonance imaging data using self-organizing mapping with spatial connectivity. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 41(5):939–946, 1999.

[97] Susan Niebergall, Alexander Loew, and Wolfram Mauser. Integrative assessment of informal settlements using VHR remote sensing data-The Delhi case study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1(3):193–205, 2008.

[98] RAA. Nobrega, CG. O'Hara, and JA. Quintanilha. Detecting road in informal settlements surrounding Sao Paulo City by using object-based classification. In *1st International Conference on Object-based Image Analysis*, 2006.

[99] Wilna H Oldewage-Theron, Emsie G Dicks, Carin E Napier, and Rajab Rutengwe. Situation analysis of an informal settlement in the Vaal Triangle. *Development Southern Africa*, 22(1):13–26, 2005.

[100] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. doi: 10.1109/tsmc.1979.4310076.

[101] Lei Peng. Adaptive median filtering. In *Seminar Report, Machine Vision*, volume 140, 2004.

[102] Martino Pesaresi and Jon Atli Benediktsson. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2): 309–320, 2001.

[103] Albert J Peters, Elizabeth A Walter-Shea, Lei Ji, Andres Vina, Mlchael Hayes, and Mark D Svoboda. Drought monitoring with NDVI-based standardized vegetation index. *Photogrammetric engineering and remote sensing*, 68(1):71–75, 2002.

[104] Richard Alan Peters. A new algorithm for image noise reduction using mathematical morphology. *IEEE transactions on Image Processing*, 4(5):554–568, 1995.

[105] Armando J Pinho and Luis B Almeida. Edge detection filters based on artificial neural networks. In *International Conference on Image Analysis and Processing*, pages 159–164. Springer, 1995.

[106] Moacir P Ponti. Segmentation of low-cost remote sensing images combining vegetation indices and mean shift. *IEEE Geoscience and Remote Sensing Letters*, 10(1):67–70, 2012.

[107] Michael S Rasmussen. Developing simple, operational, consistent NDVI-vegetation models by applying environmental and climatic information. part II: Crop yield assessment. *International Journal of Remote Sensing*, 19(1):119–139, 1998.

[108] CH Rohwer and DP Laurie. The discrete pulse transform. *SIAM journal on mathematical analysis*, 38(3):1012–1034, 2006.

[109] CH Rohwer and LM Toerien. Locally monotone robust approximation of sequences. *Journal of Computational and Applied Mathematics*, 36(3):399–408, 1991.

[110] Paul L Rosin. Unimodal thresholding. *Pattern Recognition*, 34(11):2083–2096, 2001.

[111] Franz Rottensteiner and Simon Clode. Building and road extraction from LiDAR data. In *Topographic Laser Ranging and Scanning*, pages 485–522. CRC Press, 2018.

[112] Simon Runsten, Francesco Fuso Nerini, and Louise Tait. Energy provision in south african informal urban settlements-a multi-criteria sustainability analysis. *Energy Strategy Reviews*, 19:76–84, 2018.

[113] Punam K Saha, Gunilla Borgefors, and Gabriella Sanniti di Baja. A survey on skeletonization algorithms and their applications. *Pattern Recognition Letters*, 76:3–12, 2016.

[114] Aswini Kumar Samantaray and Priyanka Mallick. Decision based adaptive neighborhood median filter. *Procedia Computer Science*, 48:222–227, 2015.

[115] Imane Sebari and Dong-Chen He. Automatic fuzzy object-based analysis of VHSR images for urban objects extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 79:171–184, 2013. doi: 10.1016/j.isprsjprs.2013.02.006.

[116] Jean Serra. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1982.

[117] Jean Serra and Pierre Soille. *Mathematical Morphology and its Applications to Image Processing*, volume 2. Springer Science & Business Media, 2012.

[118] Mehmet Sezgin and Bülent Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–166, 2004.

[119] Moslem Ouled Sghaier and Richard Lepage. Road extraction from very high resolution remote sensing optical images based on texture analysis and beamlet transform. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5):1946–1958, 2015.

[120] Wenzhong Shi and Changqing Zhu. The line segment match method for extracting road network from high-resolution satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 40 (2):511–514, 2002.

[121] KS Sim, YY Kho, Chih Ping Tso, ME Nia, and HY Ting. A contrast stretching bilateral closing top-hat Otsu threshold technique for crack detection in images. *Scanning*, 35(2):75–87, 2013.

[122] Pierre Soille. *Morphological Image Analysis: Principles and Applications*. Springer, 1999.

[123] Pierre Soille, Edmond J Breen, and Ronald Jones. Recursive implementation of erosions and dilations along discrete lines at arbitrary angles. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (5):562–567, 1996.

[124] IPAC Staff. Near, mid and far-infrared. *NASA ipac. Retrieved on*, pages 04–04, 2007.

[125] William L Stefanov and Maik Netzband. Assessment of ASTER land cover and MODIS NDVI data at multiple scales for ecological characterization of an arid urban center. *Remote sensing of Environment*, 99(1-2):31–43, 2005.

[126] Gene Stoltz and Inger Fabris-Rotelli. Pulse reformation algorithm for leakage of connected operators. In *VISAPP (1)*, pages 583–590, 2015.

[127] George Gene Stolz. Roadmakers Pavage, Pulse Reformation Framework and image segmentation in the Discrete Pulse Transform. Master's thesis, University of Pretoria, 2014.

[128] Sylvia Szabo, Sabu Padmadas, and Jane Falkingham. Is rapid urbanisation exacerbating wealth-related urban inequalities in child nutritional status? evidence from least developed countries. *The European Journal of Development Research*, 30(4):630–651, 2018.

[129] RN Thiede, IN Fabris-Rotelli, A Stein, P Debba, and M Li. Road extraction in remote sensing images of South African informal settlements, 2017.

[130] Cem Unsalan and Beril Sirmacek. Road network detection using probabilistic and graph theoretical methods. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4441–4453, 2012.

[131] Silvia Valero, Jocelyn Chanussot, Jon Atli Benediktsson, Hugues Talbot, and Björn Waske. Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images. *Pattern Recognition Letters*, 31(10):1120–1127, 2010. doi: 10.1016/j.patrec.2009.12.018.

[132] Silvia Valero, Philippe Salembier, Jocelyn Chanussot, and Carles M Cuadras. Improved binary partition tree construction for hyperspectral images: application to object detection. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 2515–2518. IEEE, 2011. doi: 10.1109/igarss.2011.6049723.

[133] Carles Ventura, Jordi Pont-Tuset, Sergi Caelles, Kevis-Kokitsi Maninis, and Luc Van Gool. Iterative deep learning for road topology extraction. *arXiv preprint arXiv:1808.09814*, 2018.

[134] Luc Vincent. Morphological area openings and closings for grey-scale images. In *Shape in Picture*, pages 197–208. Springer, 1994.

[135] Jun Wang, Jingwei Song, Mingquan Chen, and Zhi Yang. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. *International Journal of Remote Sensing*, 36(12):3144–3169, 2015.

[136] Weixing Wang, Nan Yang, Yi Zhang, Fengping Wang, Ting Cao, and Patrik Eklund. A review of road extraction from remote sensing images. *Journal of Traffic and Transportation Engineering*, 3 (3):271–282, 2016. doi: 10.1016/j.jtte.2016.05.005.

[137] Yanan Wei, Zulin Wang, and Mai Xu. Road structure refined CNN for road extraction in aerial image. *IEEE Geoscience and Remote Sensing Letters*, 14(5):709–713, 2017.

[138] Slawo Wesolkowski, ME Jernigan, and Robert D Dony. Comparison of color image edge detectors in multiple color spaces. In *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)*, volume 2, pages 796–799. IEEE, 2000.

[139] Christian Wiedemann. External evaluation of road networks. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(3/W8):93–98, 2003.

[140] Peta Wolpe and Yachika Reddy. Alleviating urban energy poverty in the informal sector: The role for local government. 2010.

[141] Rui Xu. A multistage method for road extraction from optical remotely sensed imagery. *Journal of Information Hiding and Multimedia Signal Processing*, 7(2):438–447, 2016.

[142] Yongyang Xu, Zhong Xie, Yaxing Feng, and Zhanlong Chen. Road extraction from high-resolution remote sensing imagery using deep learning. *Remote Sensing*, 10(9):1461, 2018.

[143] Osama Yousif and Yifang Ban. Improving SAR-based urban change detection by combining MAP-MRF classifier and nonlocal means similarity weights. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(10):4288–4300, 2014.

[144] Chunsun Zhang, Shunji Murai, and Emmanuel Baltsavias. Road network detection by mathematical morphology. In *ISPRS Workshop on 3DGeospatial Data Production: Meeting Application Requirements*, pages 185–200, 1999. doi: 10.3929/ethz-a-004334280.

[145] Hongsheng Zhang, Jiang Li, Ting Wang, Hui Lin, Zezhong Zheng, Yu Li, and Yufeng Lu. A manifold learning approach to urban land cover classification with optical and RADAR data. *Landscape and Urban Planning*, 172:11–24, 2018.

[146] Jing Zhang, Lu Chen, Chao Wang, Li Zhuo, Qi Tian, and Xi Liang. Road recognition from remote sensing imagery using incremental learning. *IEEE Transactions on Intelligent Transportation Systems*, 18(11):2993–3005, 2017. doi: 10.1109/TITS.2017.2665658.

[147] Keqi Zhang, Shu-Ching Chen, Dean Whitman, Mei-Ling Shyu, Jianhua Yan, and Chengcui Zhang. A progressive morphological filter for removing nonground measurements from airborne LiDAR data. *IEEE transactions on geoscience and remote sensing*, 41(4):872–882, 2003.

[148] TY Zhang and Ching Y Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3):236–239, 1984.

[149] TY Zhang and Ching Y Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3):236–239, 1984.

[150] Xiangrong Zhang, Xiao Han, Chen Li, Xu Tang, Huiyu Zhou, and Licheng Jiao. Aerial image road extraction based on an improved generative adversarial network. *Remote Sensing*, 11(8):930, 2019.

[151] Yang Zhang, Xiang Li, and Qianyu Zhang. Road topology refinement via a multi-conditional generative adversarial network. *Sensors*, 19(5):1162, 2019.

[152] Yang Zhang, Zhangyue Xiong, Yu Zang, Cheng Wang, Jonathan Li, and Xiang Li. Topology-aware road network extraction via multi-supervised generative adversarial networks. *Remote Sensing*, 11 (9):1017, 2019.

[153] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual U-net. *IEEE Geoscience and Remote Sensing Letters*, 2018. doi: 10.1109/lgrs.2018.2802944.

[154] Tingting Zhou, Chenglin Sun, and Haoyang Fu. Road information extraction from high-resolution remote sensing images based on road reconstruction. *Remote Sensing*, 11(1):79, 2019.

[155] C Zhu, W Shi, M Pesaresi, L Liu, X Chen, and B King. The recognition of road network from high-resolution satellite remotely sensed data using image morphological characteristics. *International Journal of Remote Sensing*, 26(24):5493–5508, 2005.