

**Marcel Ziems**

**Automatic Verification of Road Databases  
using multiple Road Models**

**München 2014**

**Verlag der Bayerischen Akademie der Wissenschaften  
in Kommission beim Verlag C. H. Beck**

**ISSN 0065-5325**

**ISBN 978-3-7696-5136-2**

---

**Diese Arbeit ist gleichzeitig veröffentlicht in:  
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover  
ISSN 0174-1454, Nr. 313, Hannover 2014**





## Automatic Verification of Road Databases using multiple Road Models

Von der Fakultät für Bauingenieurwesen und Geodäsie  
der Gottfried Wilhelm Leibniz Universität Hannover  
zur Erlangung des Grades  
Doktor-Ingenieur (Dr.-Ing.)  
genehmigte Dissertation

von

Dipl.-Ing. Marcel Ziems

München 2014

Verlag der Bayerischen Akademie der Wissenschaften  
in Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

ISSN 0065-5325

ISBN 978-3-7696-5136-2

---

Diese Arbeit ist gleichzeitig veröffentlicht in:  
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover  
ISSN 0174-1454, Nr. 313, Hannover 2014

Adresse der Deutschen Geodätischen Kommission:



Deutsche Geodätische Kommission

Alfons-Goppel-Straße 11 • D – 80 539 München

Telefon +49 – 89 – 23 031 1113 • Telefax +49 – 89 – 23 031 -1283 / - 1100

e-mail hornik@dgfi.badw.de • <http://www.dgk.badw.de>

Diese Dissertation ist auf dem Server der Deutschen Geodätischen Kommission  
unter <<http://dgk.badw.de/>> elektronisch publiziert

Prüfungskommission

Vorsitzender: Prof. Dr.-Ing. Steffen Schön

Referent: Prof. Dr.-Ing. Christian Heipke

Korreferenten: Prof. Dr.-Ing. Monika Sester

Prof. Dr.-Ing. Jörn Ostermann

Prof. Dr. techn. Franz Rottensteiner

Tag der mündlichen Prüfung: 21.02.2014

---

© 2014 Deutsche Geodätische Kommission, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,  
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen

ISSN 0065-5325

ISBN 978-3-7696-5136-2



As far as the laws of mathematics refer to reality, they are not certain;  
and as far as they are certain, they do not refer to reality.  
– Albert Einstein

## Abstract

In this thesis a new approach for automatic road database verification based on remote sensing images is presented. In contrast to existing approaches, the applicability of the new approach is not restricted to specific road types, context areas or geographic regions. The general applicability of the approach is achieved by combining several state-of-the-art road detection and road verification approaches that can deal well with different road types, context areas and geographic regions. Each road detection or verification approach is realized as an independent module representing a unique road model and a specific processing strategy. All modules provide independent solutions for the verification problem of each road object stored in the database. Statistical reasoning is applied to the outputs of the individual modules in order to find the optimal combined solution.

Each module provides a solution consisting of two parts. The first part is a probability distribution with respect to the two possible states of a database object, *correct* and *incorrect*. This part is based on the developments of the original state-of-the-art approaches. The second part is a probability distribution with respect to the two states of a particular model, *applicable* and *not applicable*. This part is a specific contribution of this thesis. It is always based on an assessment of the assumptions implicitly made for the particular model, and only if all assumptions are fulfilled a model is considered as *applicable*. In accordance with the original literature, such model assumptions refer to the presence of specific road types, context areas or geographic regions, but also to the visibility of features and the availability of representative training samples in case of statistically defined models. Consequently, each module provides probabilities with respect to two state spaces (*correct*, *incorrect* and *applicable*, *not applicable*). In accordance with the Dempster-Shafer Theory both probability distributions are mapped to a new state space that includes *correct*, *incorrect* and *unknown* as the possible states of a database object. Combination and classification are realized in the new state space, whose outcome is considered to be the starting point for manual post-processing.

Experiments with real cartographic datasets provided by three national mapping agencies, showing 7742 correct and 669 incorrect road objects, are used to reveal the potential and the limitations of the proposed approach. If a human operator investigates all database objects assigned to the states *incorrect* or *unknown* and edits them if necessary, a database quality of 97.5–100% is achieved, while the human operator must look at 25–33% of the road objects in the database. If a human operator concentrates only on database objects assigned to the state *incorrect*, the database quality is reduced by approximately 5%, but the human operator must investigate not more than 9% of the road objects. Additionally, a comparison with other state-of-the-art road detection approaches using benchmark datasets shows that in general the proposed approach provides better results than other state-of-the-art approaches.

**Keywords:** geo-spatial databases, verification, image analysis, data fusion

## Zusammenfassung

Diese Arbeit stellt einen neuen Ansatz zur automatischen Verifikation von Straßendatenbanken auf Grundlage von aktuellen Luft- und Satellitenbildern vor. Im Gegensatz zu bereits vorliegenden Forschungsarbeiten ist der neue Ansatz nicht auf spezifische Straßentypen, Kontextbereiche oder geographische Regionen beschränkt. Dies wird primär durch eine Kombination unterschiedlicher Forschungsarbeiten, welche als unabhängige Verifikationsmodule im Gesamtsystem integriert sind, erreicht. Jedes der Module basiert auf einem spezifischen Straßenmodell, deren Anwendbarkeit auf bestimmte Straßentypen, Kontextbereiche oder geographische Regionen beschränkt bleibt, die in Kombination jedoch nahezu Allgemeingültigkeit erreichen. Die nicht modellierten Straßenobjekte werden automatisch erkannt und können somit manuell verifiziert werden.

Die Kombination der Modelle basiert auf einer statistischen Auswertung der Ergebnisse aller Module. Hierfür liefert jedes Modul zu jedem Straßenobjekt der zu verifizierenden Datenbank ein aus zwei Teilen bestehendes Ergebnis. Der erste Teil ist eine Wahrscheinlichkeitsverteilung, welche den Zustand eines Straßenobjektes (*richtig*, *falsch*) beschreibt. Die zugrunde liegenden Ansätze sind weitgehend aus der Literatur übernommen. Der zweite Teil ist eine Wahrscheinlichkeitsverteilung, welche den Zustand eines Modells (*anwendbar*, *nicht anwendbar*) beschreibt. Dieser Teilaspekt ist grundlegend neu und somit ein wichtiger Beitrag der vorliegenden Arbeit. Die Idee dabei ist, dass ein Modell nur genau dann anwendbar ist, wenn alle Modellannahmen bezüglich eines Straßenobjektes im Bild erfüllt sind. Ausgehend von den ursprünglichen Forschungsarbeiten beziehen sich die Annahmen auf das Vorhandensein bestimmter Straßentypen, Kontextbereiche oder geographischer Regionen, aber auch auf die Sichtbarkeit von Merkmalen oder, bei statistischen Ansätzen, auf die Verfügbarkeit von repräsentativen Trainingsdaten. Somit liefern die Module Wahrscheinlichkeiten bezüglich zweier Zustandsräume, welche auf Grundlage der Dempster-Shafer Theorie auf einen neuen Zustandsraum abgebildet werden. Dieser neue Zustandsraum unterscheidet zwischen *richtig*, *falsch* und *unbekannt*.

Experimente mit drei realen Straßendatenbanken, welche insgesamt 7742 richtige und 669 fehlerhafte Straßen enthalten, zeigen das Potential des neuen Ansatzes. Dabei stellt sich heraus, dass eine Datenbankqualität von 97.5–100% erreicht wird, wenn alle Straßenobjekte, die als *unbekannt* oder *falsch* klassifiziert sind, von einem menschlichen Operateur betrachtet und gegebenenfalls berichtigt werden. Hierbei müssen nur 25–33% aller Straßenobjekte vom Operateur betrachtet werden. Reduziert man den manuellen Aufwand weiter und berücksichtigt nur die maximal 9% der als *falsch* klassifizierten Straßenobjekte vermindert sich die Datenbankqualität im Durchschnitt nur um ca. 5%. Weitere Experimente auf Grundlage von Benchmark-Datensätzen ordnen den neuen Ansatz wissenschaftlich ein. Die Ergebnisse zeigen, dass der neue Ansatz in allen Testszenen mindestens gleich gute Ergebnisse liefert wie der jeweils beste Ansatz des entsprechenden Benchmark Tests. In zwei der sechs Szenen sind die Ergebnisse mit ca. 25% höherer Vollständigkeit sogar deutlich besser.

**Schlagworte:** räumliche Datenbank, Verifikation, Bildanalyse, Datenfusion



# Contents

<b>1. Introduction</b>	<b>1</b>
1.1. Objectives and focus . . . . .	2
1.2. Outline . . . . .	3
1.3. State of the art . . . . .	3
1.3.1. Road models . . . . .	4
1.3.2. Combination strategies for different road models . . . . .	8
1.3.3. Working with road databases . . . . .	11
1.3.4. Discussion . . . . .	13
<b>2. Basics</b>	<b>15</b>
2.1. Combination of uncertain information . . . . .	15
2.1.1. The reasoning problem . . . . .	15
2.1.2. The Dempster-Shafer Theory . . . . .	17
2.1.3. Decision making . . . . .	20
2.1.4. Derivation of probability masses . . . . .	22
2.2. An uncertainty metric for Support Vector Machines . . . . .	24
2.2.1. The problem of partly non-representative training data . . . . .	24
2.2.2. The Support Vector Data Description . . . . .	26
2.2.3. Hyperparameter adjustment . . . . .	29
2.2.4. Kernel space metric . . . . .	31
<b>3. New methodology</b>	<b>35</b>
3.1. Basic strategy . . . . .	35
3.2. The fusion framework . . . . .	41
3.2.1. Reasoning strategy . . . . .	41
3.2.2. Model-uncertainties . . . . .	43
3.3. Verification modules . . . . .	47
3.3.1. Colour classification . . . . .	47
3.3.2. Line detection . . . . .	56
3.3.3. Parallel edge detection . . . . .	65
3.3.4. The SSH method . . . . .	68
3.3.5. Acupuncture method . . . . .	72
3.3.6. Building and grassland detection . . . . .	78
3.3.7. Adjacency analysis . . . . .	84
3.4. Practical aspects of the approach . . . . .	87
3.4.1. System parameters . . . . .	87
3.4.2. Implementation issues . . . . .	90
3.5. Discussion . . . . .	90
<b>4. Experiments</b>	<b>93</b>
4.1. Datasets . . . . .	93
4.2. Evaluation of the proposed methodology . . . . .	98
4.2.1. The SVDD-based uncertainty metric . . . . .	98
4.2.2. The data fusion strategy . . . . .	103
4.2.3. The overall approach in a cartographic context . . . . .	128
<b>5. Conclusions and outlook</b>	<b>147</b>
<b>A. Appendix</b>	<b>151</b>
A.1. The cat in the box – Bayesian inference versus Dempster-Shafer Theory . . . . .	151
A.2. Gaussian Mixture Models for colour-based road verification . . . . .	158
A.3. Images used for the experiments described in the thesis . . . . .	163
<b>Bibliography</b>	<b>171</b>
<b>Curriculum vitae</b>	<b>177</b>
<b>Acknowledgements</b>	<b>179</b>



# 1. Introduction

Road databases have large economical relevance as the basis for navigation systems and web services like Google maps ([www.google.maps.com](http://www.google.maps.com)) or bing-maps ([www.bing.com/maps](http://www.bing.com/maps)), whereby they have become part of our everyday life, e.g. for finding the way to the office, booking a hotel or finding the nearest shop for a good coffee. Beyond that, road databases have been known all along to be an important part of the geodata infrastructure, e.g. as the basis for urban planning or organizing emergency services. For many of those applications frequent updates and quality control are of vital importance. Hence, the private sector, which is mainly represented by a relatively small number of mapping companies, and the public sector, which is represented by national mapping agencies, spend high efforts on their road databases. As an alternative to those professionally working institutions, recently also crowdsourcing projects such as the Humanitarian OpenStreetMap Team ([www.hot.openstreetmap.org](http://www.hot.openstreetmap.org)) are able to provide valuable road databases as they can activate thousands of volunteers worldwide.

The availability of remote sensing data such as aerial and satellite images with a geometric resolution better than 2.5 m has significantly increased over the last two decades. Hence, up-to-date imagery where roads can be observed is basically available (cf. Figure 1.1). Web mapping services use these circumstances to visualize road databases superimposed on imagery (and other map contents). Mapping agencies commonly use the available remote sensing imagery rather than in-situ observations with GPS devices to update or check their databases. In the meantime the open street map project that was initially based on in-situ GPS measurements also uses remote sensing imagery. This allows the volunteers to participate in mapping from remote locations, which is especially interesting for crisis events.



Figure 1.1.: Up-to-date imagery superimposed by an outdated road database.

Despite the advantages of imagery to update road databases, the bottleneck is the manpower required to search through the imagery. In this regard, editing efforts become less time consuming if the frequency of updates increases because only small changes can be expected. Therefore, approaches that automatically identify the database errors on the basis of the imagery are of interest, while the editing can still be carried out manually.

Figure 1.1 shows a possible scenario for the problem of automatically finding database errors. It turns out that detecting these errors is challenging as roads and even more database errors concerning roads are complicated to describe in general. This problem can partly be circumvented by introducing additional knowledge into the process. Therefore, the actual research field dealing with automatic updates basically distinguishes two tasks: First, the automated *verification*, i.e. the analysis of all existing road database entries, and second the automated *detection of missing* road database entries. Both subtasks induce specific strategies that exploit the knowledge given with the original database in different ways.

Recent research has shown that approaches to *verification* and *detection of missing roads* can provide promising contributions to the problem of updating road databases with a high time efficiency. However, neither of these approaches gained practical relevance. One reason for this is that such approaches are only capable to deal with relatively small areas, where the roads appear more or less similar, and thus can be explained by relatively simple models. For instance Youn et al. [2008] basically restrict their approach to straight roads in a dense urban area, while Gerke & Heipke [2008] basically restrict their approach to rural areas. Hence, the existing approaches cannot simply be scaled up to achieve the goal of automation because real datasets usually cover large areas with heterogeneous conditions for the underlying automatic image analysis processes.

## 1.1. Objectives and focus

The objective of the work described in this thesis is to develop a method that automatically verifies road database objects on the basis of remote sensing imagery. In this context, *verification* means a check whether a road database object can be found in the image and whether the required positional accuracy is met. Hence, the approach has to define the state of a database object that principally can be either *correct* or *incorrect*.

The approach presented in this thesis does not deal with the verification of attribute information that is stored in some road databases such as road name, road width or road type. The approach is not about automatically achieving corrected road geometries either, nor is it about finding completely new roads. These three tasks, while important for practical applications, are beyond the scope of this work.

In contrast to existing approaches for road database verification, the appearance of roads and their local surroundings is not restricted. The new method can deal with very different conditions, e.g. roads in urban, sub-urban and rural context areas, straight and winding roads. This is realized by a combination of different state-of-the-art road detection methods, each of which has been proven to operate successfully under specific circumstances, e.g. in urban, sub-urban and rural context areas. The main contribution of this thesis lies in the integration of the partially incomplete and



possibly contradicting solutions that the different methods provide. For that purpose, a statistical framework on the basis of the Dempster-Shafer Theory (DST) is developed, which requires an individual uncertainty measure for each solution provided by each method. It conceptually considers always two states of a method: either its underlying model is *applicable* or it is *not applicable* for a particular database object. The state is always determined by checking the correspondence of the actual situation in the imagery with the assumptions made by the underlying model of the method which represents the second important contribution of the thesis.

One method in the framework uses a support vector machine (SVM) classifier, and hence its model is implicitly defined on the basis of training samples. This model assumes the training data to be representative for the test data. In order to check this assumption, the feature space distance of a test sample from the training samples is evaluated. Therefore, a new kernel space metric is developed on the basis of a support vector data description (SVDD) which is the third important contribution of this thesis.

The verification results of each method (*correct*, *incorrect*) and the corresponding model-uncertainty (*applicable*, *not applicable*) are mapped to a new state space that includes three states: *correct*, *incorrect* and *unknown*. The experiments are carried out with real cartographic datasets which particularly demonstrates the practical relevance of the additional state *unknown*. Further experiments are carried out with benchmark datasets so that the new approach can also be put into the context of the scientific literature.

## 1.2. Outline

This thesis is structured as follows: First, state-of-the-art approaches dealing with road detection and verification of road databases are reviewed. In Chapter 2, the mathematical basics are described in the context of the new approach. The methodology of the new approach is explained in detail in Chapter 3. Seven different ways to verify road objects based on ten different road models are explained and the strategy to combine their outputs is presented. Chapter 3 concludes with a discussion of the presented approach. Chapter 4 reports on experiments, including an evaluation of the SVM-based uncertainty metric, an evaluation of the fusion concept based on benchmark datasets and an evaluation based on real cartographic datasets. Finally, in Chapter 5 conclusions and an outlook are given.

## 1.3. State of the art

Developments in the field of automatic road detection in remotely sensed imagery started in the 80's, reaching a climax in the later 90's; see [Auclair-Fortier et al., 1999; Mena, 2003] for extensive literature surveys. Recent research tries to combine the basic concepts developed in the 90's but also integrates successful concepts from neighbouring disciplines, such as machine learning. However, the scientific goal of providing techniques for practical applications has not yet been achieved.

The following review will primarily focus on approaches that use RGB aerial or satellite images with a ground sampling distance (GSD) of 0.2–1.0 m, because this is currently the most realistic input

available for potential applications of the proposed approach, e.g. at national mapping agencies or other web map providers. Furthermore, approaches based on infrared information and stereo images will also be considered because these kinds of data are expected to become more available in the near future. Approaches relying on very high resolution imagery (GSD<0.2 m), Synthetic Aperture Radar (SAR) and Airborne Laserscanner (ALS) data are not explicitly reviewed but included if their concepts seem to be transferable to the other data.

The objective of this thesis is *road verification* which in the first place requires the detection of roads in the imagery. Hence, the first part of the following review (Sections 1.3.1 and 1.3.2) is about road detection approaches, independently of what their motivation might be: *detection from scratch* or specific tasks related to road databases such as *database enhancement*, *verification* or *update*. In accordance with [Mayer, 1998] road detection approaches can be distinguished by their underlying model which always describes at least some properties of roads in the imagery. In Section 1.3.1 the most promising properties considered in state-of-the-art approaches will be investigated one after the other and independently of each other. In Section 1.3.2 strategies to combine different road properties will be presented. In Section 1.3.3 particularly road detection approaches that deal with road tasks such as *database enhancement*, *verification* or *update* will be investigated, before in Section 1.3.4 the deficits of the existing road verification approaches will be discussed.

### 1.3.1. Road models

**Line-based models:** A significant group of approaches models roads as lines of more or less constant brightness and width. The maximum curvature and the width of a line can be specified by introducing knowledge about the road type and the image resolution. Line-based models were defined for panchromatic imagery with a GSD of about 1–2 m. For instance, Wiedemann & Ebner [2000] initially apply the line detection algorithm described in [Steger, 1998], and then evaluate the resulting line segments according to their *homogeneity*, *width* and *curvature*. Higher resolution images are usually down-sampled to a resolution of 1–2 m in order to reduce disturbances with respect to the line model, e.g. [Bacher & Mayer, 2005]. Other authors apply the Steger line-detector but use different input data. For instance, Gerke & Busch [2005] use NDVI<sup>1</sup> images and Hinz & Baumgartner [2003] use nDSM<sup>2</sup> images. A deficit of the Steger line-detector is its restriction to homogeneous background regions. Alternative line detection algorithms that are slightly more effective for heterogeneous background regions rely on wavelet-transform [Gruen & Li, 1995], Radon-transform [Zhang & Couloigner, 2006] or Hough-transform [Hu et al., 2004]. Furthermore, specific line detectors have been developed for SAR data [Tupin et al., 1998] and for ALS data [Clode et al., 2007]. If an initialization is given, e.g. from a road database, the line property of roads can also be exploited by applying active contours [Kass et al., 1988], as Klang [1998] and Koutaki et al. [2006] have demonstrated.

In conclusion, numerous line-based models have been developed for different types of input data that are most effective for homogeneous background, as a benchmark test, organized by EuroSDR

<sup>1</sup>The normalized difference vegetation index (NDVI), depending on the red and the infrared bands of an image shows the vitality of the vegetation in form of grey values. Roads usually represent non-vegetated areas, and thus differ significantly from vegetated areas such as grassland.

<sup>2</sup>A normalized digital surface model (nDSM) shows the height above the ground. Roads usually represent ground surface areas, and thus differ from 3D objects such as buildings.

has shown [Mayer et al., 2006]. However, for a heterogeneous background as it typically appears in urban areas, such approaches suffer from high false alarm ratios due to confusion of roads with other objects, e.g. rails, rivers, elongated houses or gaps between two elongated houses, rows of trees and walls. As a consequence, some authors use stricter parameters but then achieve lower detection ratios, e.g. [Hu et al., 2004; Zhang & Couloigner, 2006].

**Models based on parallel edge pairs:** These approaches model road borders as pairs of anti-parallel<sup>3</sup> edges. The spacing between those edges is often specified according to the road type and the given image resolution. Such methods usually rely on aerial images with a GSD  $\leq 0.5$  m, where *parallelism* and *distance* between previously extracted edge segments are evaluated to determine road candidates, e.g. [Heipke et al., 1995]. Zhang [2004] searches for parallel edge pairs in stereo images, and thus considers three dimensions to determine road candidates. Given some initialization, Fua [1996] uses ribbon snakes, a specific form of active contours, to consider anti-parallel edges. Göpfert et al. [2011] present an extension of this method to the 3D case.

The benchmark test described in [Mayer et al., 2006] concludes that parallel edge pairs are mainly useful in rural context areas. As an advantage to line-based approaches, approaches relying on parallel edge models produce less false alarms even in heterogeneous image regions. However, they also provide lower detection ratios because road borders are often partially occluded by trees, vehicles and shadows.

**Models based on edge directions:** These approaches model roads as regions that are less frequently intersected by edges than the background. Furthermore, if an edge intersects a road, it is expected to have the same direction as the road. For instance, Youn et al. [2008] present an approach for satellite imagery of about 1 m GSD in an urban environment, called *acupuncture method*. At first, so-called *acupuncture nails* are defined by analysing the main directions of the road network. The *acupuncture nails* represent plausible positions of roads. Then, the intersections of *acupuncture nails* with image edges are counted: *nails* having low intersection counts are classified as roads. Gamba et al. [2006] present an approach for SAR images that applies direction-sensitive filters to determine the frequency and the directions of edges and then integrate the filter responses along plausible road paths that are known in advance from another processing step.

The results presented in [Gamba et al., 2006; Youn et al., 2008] show that these models can deal with urban areas even if the imagery has comparably low geometric resolution. However, such a model is restricted to a specific type of urban area, where roads appear in a regular grid between buildings aligned in a rectangular pattern. If these preconditions are not fulfilled, for instance in rural areas, the model is not applicable.

**Colour-based models:** Such approaches model roads as image regions with specific radiometric properties in each colour band. Such models require colour imagery, e.g. RGB or IRRG<sup>4</sup>, mostly with a GSD  $\leq 0.5$  m. As illumination conditions and the materials of the road surface and background objects can hardly be predicted and defined by humans, usually machine learning methods are applied to parametrise the model on the basis of training samples. The training samples are either provided

<sup>3</sup>In the related literature the term *anti* means that the gradients have opposite directions.

<sup>4</sup>Here IRRG denotes the colour bands infrared, red and green. Some authors use the term *false colour infrared image* (CIR).

by a human operator (e.g. Song & Civco, 2004; Poullis & You, 2010) or by a second road detection algorithm (e.g. Doucette et al., 2001; Bacher & Mayer, 2005). Nearly the full range of popular supervised classification methods have been applied, e.g. the maximum likelihood classifier [Doucette et al., 2001], classifiers based on Fuzzy-set theory [Bacher & Mayer, 2005], neural networks [Kuo-Tu, 1995], expectation maximization to determine the parameters of a Gaussian mixture model [Peng et al., 2010; Poullis & You, 2010; Butenuth et al., 2011] and support vector machines [Song & Civco, 2004; Ma et al., 2008; Fujimura et al., 2008]. In addition, unsupervised classification methods have been tested, e.g. ISODATA [Zhang, 2004] and k-means clustering [Zhang & Couloigner, 2006]. In that case, the class assignment to the clusters is determined by fixed rules, e.g. low brightness in the infrared and high brightness in the blue band.

Reviewing the experiments for colour-based models, their principal applicability to different context regions is an advantage. A deficit, pointed out by Das et al. [2011], is that such approaches usually suffer from many false alarms due to parking lots and roofs that have similar colour properties as roads. Thus, many authors using the colour model concentrate their experiments on scenes where the roofs have different colour properties to the roads and parking lots are less relevant, and thus achieve rather good results.

**Texture-based models:** Roads can also be modelled as image regions of specific texture. In this thesis *texture* is understood to describe relations of grey values of different pixels within a certain local neighbourhood, whereas *colour* is understood to describe the grey values of a pixel (or a set of pixels) in different colour bands. A rather typical texture-based model is described in Quam [1978], where cross correlation coefficients between a central image patch and several patches in a local surrounding are computed to determine the most probable road direction starting from an initial road pixel. Haverkamp [2002] defines a set of rotated rectangular patches around each image pixel, where the width of the rectangle corresponds to the expected road width and its length depends on how straight the roads are expected to be. For each pixel, the variance of the grey values of each rectangle is determined and compared to those of the other rotated rectangles with respect to the same pixel. The minima are assumed to indicate the road directions. The presence of one or two significant minima indicates a road pixel, while the presence of more than two significant minima indicates a road junction. Another strategy for exploiting texture properties is given in [Mena & Malpica, 2005], where a histogram defined from a quadratic test patch is compared to training histograms. The similarity of histograms is determined by the Bhattacharyya distance. Fujimura et al. [2008] use a road database to define plausible road regions. Then, they determine the similarity between the alternative solutions by comparing the respective histograms via the Bhattacharyya distance. As it is assumed that only one image region in a local neighbourhood can represent the true position of a road, the most dissimilar region from the rest of the regions is classified as a road, i.e. the one that has the largest integrated Bhattacharyya distance from the rest. Das et al. [2011] apply an SVM classifier using the features *mean*, *standard deviation*, *skew energy* and *entropy* of the grey value distribution extracted from a quadratic image patch around a pixel.

Usually, texture-based methods require a GSD  $\leq 0.5$  m, e.g. [Mena & Malpica, 2005; Das et al., 2011]. However, this demand can be reduced to a GSD  $\leq 2.5$  m if rectangular image patches are employed [Haverkamp, 2002; Fujimura et al., 2008]. As Haverkamp [2002] points out, adjusting the

lengths of the patches to values larger than parking lots or buildings makes the method robust against false alarms but also reduces the detection ratios for curved and short roads. Furthermore, errors in texture-based models are particularly caused by occlusions or shadow as mostly homogeneous texture properties along the road are expected.

**Segment-shape models:** A comparably small number of approaches models roads as elongated and more or less rectangular regions of homogeneous image properties. One example is given by Ruskoné & Airault [1997] who initially perform a *watershed segmentation* in a panchromatic image of about 0.5 m GSD, followed by a sub-selection of segments that have parallel borders. Jin & Davis [2005] initially apply a *k-means segmentation* on the basis of Ikonos images and then select the road segments based on the features *mean width*, *elongation* and *size*. Grote et al. [2012] first apply a *normalized cuts segmentation* on the basis of an IRRG image with 0.1 m GSD, considering features such as *brightness*, *hue*, NDVI and *local gradients*. The resulting segments are classified according to the shape descriptors: *elongation*, *convexity*, *width* and *width constancy*.

As Jin & Davis [2005] point out, approaches relying on segment shapes tend to produce low false alarm ratios even for challenging environments. Grote et al. [2012] show that the detection ratios of such approaches may decrease in situations where shadows or occlusions due to trees or vehicles occur frequently.

**Local context models:** Local context objects such as *road markings*, *vehicles*, *trees*, *buildings*, *grass-land* and *water* allow to indirectly infer the presence of roads. However, such models become complex with an increasing number of considered objects. Therefore, most authors restrict possible interactions. For instance, [Hinz & Baumgartner, 2003; Zhang, 2004] define semantic networks, i.e. a relatively small number of (directed) edges represent specific relations between roads and context objects. More recent approaches often concentrate on one type of context object only and model the inference statistically. For instance, Gerke & Heipke [2008] consider rows of trees to infer occluded roads. Poulain et al. [2010] use buildings to infer shadowed road regions and Butenuth et al. [2011] use initially known road centrelines and areas classified as water to infer flooded roads.

As several authors have demonstrated, local context models can be very effective to provide solutions for particular situations, e.g. for tree-lined roads [Gerke & Heipke, 2008] or for urban canyons [Poulain et al., 2010]. However, they are far away from being generally applicable as they concentrate on a few interactions with specific context objects only. Furthermore, it is worth noting that such approaches require the positions of the context objects in the imagery which increases the overall complexity of such approaches.

**Network topology models:** Roads are arranged in networks which implies that roads are never isolated, a property that is considered by many authors to remove false alarms of former processing steps, e.g. [Zhang & Couloigner, 2006]. Furthermore, it can be assumed that a road network connects places at relatively short distances. One possibility to exploit this property is to add road segments where a short connection seems to be missing, e.g. [Baumgartner et al., 1999]. Another possibility is to adjust the detection process to be less restrictive in situations where a short connection seems to be missing, e.g. [Fischler et al., 1981; Gerke & Heipke, 2008]. The approach presented in [Wiedemann & Ebner, 2000] determines the *shortest paths* based on road candidates detected by a pre-processing

step that itself achieves high detection ratios but also produces many false alarms. The basic idea is that the true road candidates correspond to the most effective connections, whereas the false alarms represent less effective connections. Another kind of approach that explicitly exploits the network characteristics of roads relies on active contours that are initialized as a network and where the internal energy forces the solution to preserve the character of that network, e.g. [Peng et al., 2010; Butenuth & Heipke, 2012].

It can be stated that approaches that exploit the network properties of roads are quite powerful if the initialization and hence the prior knowledge about the road network is nearly complete.

### 1.3.2. Combination strategies for different road models

The road models discussed in the previous section all have particular strengths and deficits. Many authors combine these basic model properties to define more comprehensive models. To do so, the combination of rather different extraction or classification methods is required in some way. The subsequent review will examine the different combination concepts.

**The either-or concept:** An obvious possibility to combine (partially) complementary road models is to select the respective method that seems to be the best for a given situation. For instance, McKeown & Denlinger [1988] combine a texture-based method [Quam, 1978] and a method based on parallel edge pairs [Nevatia & Babu, 1980] that both provide specific confidence measures depending on local image properties. If the selected method falls below a pre-defined confidence threshold, the system automatically switches to the alternative method. Bordes et al. [1997] enhance that strategy by adding a network-based method that connects dead end roads if both other methods are unconfident. Jin & Davis [2005] use prior knowledge from a database to select one of two possible methods. One method is better suited for dense urban areas while the other is more appropriate for sub-urban areas. The first method uses a Hough-transform to extract only straight lines that have a length larger than the expected size of a building block. The second method is particularly useful for sub-urban areas as it relies on a segment shape model that also considers curvilinear roads.

An advantage of the *either-or concept* is that complementary model properties can be combined without the requirement of handling correlations. A problem of the *either-or concept* is that the combined solution is never better than the most confident solution, and thus, a possible advantage of a fusion strategy is not exploited, namely the gain of knowledge by combining different knowledge sources. Another considerable problem is the definition of the criterion to choose the (one) optimal method.

**The evaluation concept:** This concept combines the outputs from one *basic method* with those of other independently applied *evaluation methods*. The main idea is that the *basic method* provides nearly complete results but it also produces many false alarms. The *evaluation methods* provide solutions with lower completeness but high correctness. Then, the outputs of the *basic method* are evaluated on the basis of the other outputs that potentially represent different road models. For instance, Baumgartner et al. [1999] apply a line-based method [Steger, 1998] that produces many false alarms in urban or sub-urban context regions. Then, the detected line primitives are evaluated by analysing the presence of parallel edge pairs and homogeneous image regions in the vicinity.

One difficulty of this concept seems to be the parametrisation of the *basic method* which is required to provide relatively stable solutions in terms of the completeness.

**The sequential concept:** A further concept to combine different road models is to put some methods into a sequence, where the output of the first method serves as the input for the second method and so on. In contrast to the *evaluation concept*, the initially applied method is not required to provide a complete solution. For instance, Bacher & Mayer [2005] start with a very restrictive edge-based method which only provides the training samples for a colour classification method. The output of the classification serves as the input for a line detection algorithm which provides the final result. Zhang [2004] starts with colour classification to restrict the search space, followed by a method relying on parallel edge pairs. The output of that method is used as the input for a context-based model that considers *road markings* and *shadow*. Grote et al. [2012] start with a segmentation method and then classify the segments by considering *colour* and *shape* features. The classification results in fuzzy membership values for each segment which serve as a basis for a context-based analysis considering *vehicles*, *trees*, *buildings* and *grassland*. Finally, the output is passed on to a global network analysis that delivers the final road network including road junctions.

The main advantage of the *sequential combination concept* is that very complex models with complementary road properties can be defined. All cited approaches [Zhang, 2004; Bacher & Mayer, 2005; Grote et al., 2012] use fuzzy membership values according to the *Fuzzy-set theory* [Zadeh, 1965] in order to specify the relative quality of the outputs. The fuzzy membership values basically describe the degree to which the outputs correspond to the underlying model and are mainly important to control the interactions between the different methods rather than serving as a quality metric for the final result. In theory, approaches like [Zhang, 2004] and [Grote et al., 2012] are able to provide good results for nearly every imaginable situation if the parametrisation of all methods is optimal. However, it turned out that the parametrisation becomes rather difficult, because the parameters interact in a complex way. Changing the parametrisation of the first method may have direct consequences for the optimal parametrisation of the second method and so on. Finally, the relatively complex parametrisation makes an adaptation to new scenes very time consuming which is a relevant problem when applying these approaches to real cartographic tasks.

**The concept of independent methods:** A further concept is to apply independent methods where each method focuses on a particular road property. The outputs are then considered in a fusion framework that has a relatively flat structure, i.e. the underlying methods are basically equally ranked. One example is given by Hinz & Baumgartner [2003] who apply a parallel edge extraction method, once on an image of fine scale and once on an image at a coarse scale. A third method extracts single road lanes based on road markings. A quality measure based on the detection process is defined in form of fuzzy membership values that control the impact of the method-specific outputs on the final result. Gerke & Heipke [2008] apply a line-based method and a colour-texture-based classification method that extracts rows of trees. Based on a statistical model, both methods provide conditional probabilities (likelihoods) that are combined in accordance with the Theory of Bayesian inference [Box & Tiao, 1992]. Gerke [2005] also presents an alternative approach considering the same inputs but relying on the DST [Dempster, 1967; Shafer, 1976] instead of Bayesian inference. Another approach using the DST is described by Mena & Malpica [2005] who basically apply the

same classification algorithm three times, each time based on different colour-texture features. The quality measures for classification outputs are defined in the form of probability masses that are defined on the basis of the ratio of a distance from a test sample to the closest training sample and the average distance between all training samples. A third approach using the DST is described in [Poulain et al., 2010], where five linear classifiers are combined. All classifiers rely on the same basic classification algorithm but the features always describe different road properties and the parameters differ. The probability masses depend on the feature-space-distance to the decision-threshold; large distances represent classification results of better quality than shorter distances.

Consequently, three significant fusion techniques *Fuzzy-set theory*, *Bayesian inference* and DST can be identified in the related work. According to [Klein, 2004] all of them represent scientifically founded techniques to combine confirming and conflicting information. The requirements for the solutions to be combined (*fuzzy membership values*, *probabilities* or *probability masses*) are always *independence* and *consistency*. The approaches described in [Hinz & Baumgartner, 2003; Mena & Malpica, 2005; Poulain et al., 2010] combine methods using different input data, and thus the authors assume *independence*. The strategies of the combined methods are similar which makes it relatively easy to define *consistent* quality measures by simply using the same strategy for their definitions. In contrast, Gerke & Heipke [2008] have to define a rather complex *geometric-topologic relationship model* to achieve the required *consistency* because the methods to be combined follow very different strategies.

Another fusion technique for independently operating detection methods incorporates machine learning methods. For instance, Porway et al. [2010] combine the outputs of image classifiers that deliver solutions for *roads*, *buildings*, *vehicles*, *trees* and *parking lots*. Potential conflicts concerning the outputs of those methods are solved by non-recursive grammars that are learned in advance. Das et al. [2011] apply two different methods, a texture-based SVM classifier and a parallel edge detection method. The outputs of both methods are passed on to a neural network classifier. The advantage of solving the fusion problem with supervised machine learning methods is that one does not have to care about the definition of *consistent* quality measures. However, the training data must represent all possible qualities to all possible outputs in all possible constellations of the underlying methods. Hence, the required amount of training data increases drastically with an increasing number of methods to be combined.

Due to the stated difficulties concerning *independence*, *consistency* and *amount of training data*, so far approaches that perform a combination of independent methods either consider just a small number of methods [Gerke & Heipke, 2008; Das et al., 2011] or only use methods that rely on similar strategies [Hinz & Baumgartner, 2003; Mena & Malpica, 2005; Poulain et al., 2010; Porway et al., 2010]. Thus, these approaches do not achieve the same generality as approaches that apply a *sequential combination concept*, e.g. [Zhang, 2004; Grote et al., 2012]. However, the independence of the parameters of the methods makes frequent adaptations to new datasets much easier.

**Feature-level fusion:** The last combination concept discussed here is the combination of different road models by machine learning methods on feature level. Machine learning methods have already been described in the context of high level fusion, i.e. for combining the outputs of independently operating detection methods that itself might rely on heuristics. Furthermore, machine learning has



been mentioned in connection with colour and texture-based models which basically means that the underlying feature spaces just cover colour or texture properties. Colour and texture features are combined frequently by stacking the respective feature vectors, e.g. [Zhang & Couloigner, 2006]. Mnih & Hinton [2010] also follow this idea, but define the feature space in a much more general way. In particular, they consider the grey values of all  $64 \times 64$  pixels in a local neighbourhood. As they use RGB images with 1.2 m GSD, the resulting feature vector has 12,288 dimensions and represents an area of about  $5,000 \text{ m}^2$  in object space. For the experiments a neural network classifier with millions of parameters is learned on the basis of approximately 1.2 million training samples, extracted automatically on the basis of a  $500 \text{ km}^2$  subset of a road database. Due to the huge amount of parameters mainly computational aspects are discussed in [Mnih & Hinton, 2010]. The authors report that training was resolved within three days. The advantage of such a strategy is the automated parametrisation. A potential disadvantage is the high amount of required training data that must represent the incredible large number of possibilities of the  $64 \times 64$  pixel set.

### 1.3.3. Working with road databases

Das et al. [2011] reviewed experiments from various state-of-the-art road detection approaches and noticed that the test datasets are usually very small and comprise less than 200 road objects spread over 2–8 small scenes.<sup>5</sup> Commonly achieved detection ratios range from 75% to 85% and the topology of road junctions is often unconsidered. Furthermore, automatically extracted geometries often do not match the aesthetic requirements of road maps as they look blurred. In conclusion, state-of-the-art solutions are basically not good enough for automatic road database acquisition. However, a few approaches have been published that contribute some degree of automation to tasks related to existing road databases. Three such tasks will be discussed in the following: *geometric enhancement*, *semi-automatic verification* and *semi-automatic detection of missing roads*.

**Geometric enhancement:** A number of approaches geometrically correct existing road databases, e.g. initiated by the national mapping agencies of France [Bordes et al., 1997], Switzerland [Zhang, 2004] and Japan [Koutaki et al., 2006; Fujimura et al., 2008]. All of these approaches limit the search space in accordance with the given road geometries and the expected maximum spatial discrepancies. Beyond that, some authors use further knowledge from the inaccurate database to tune their algorithms. For instance, Bordes et al. [1997] adjust the parameters of their method by additional object-specific knowledge such as *road type* and *road width*, stored in the geometrically inaccurate database. Koutaki et al. [2006] use the inaccurate database to initialize an active contour model. Fujimura et al. [2008] generate correction hypotheses on the basis of the geometrically inaccurate database by assuming only parallel displacements which preserves the shape and hence the aesthetic level of the man-made road geometries contained in the database.

The approaches to geometric database enhancement demonstrated high success rates, i.e.  $\approx 85\%$  of the database roads in rural areas could be shifted to an acceptable position. A practical problem might be that these approaches do not indicate the database roads that are still inaccurate ( $\approx 15\%$ )

<sup>5</sup>The exceptions from that observation are the experiments published in [Gerke & Heipke, 2008] with over 530 road objects, the experiments published in [Mnih & Hinton, 2010] with two scenes of 50 and 28  $\text{km}^2$  size, respectively, and the experiments published in [Das et al., 2011] with 200 small scenes.

which means that a human operator has to check 100% of the database to correct the remaining 15% of the roads.

**Semi-automatic verification:** Such approaches automatically check whether a road database object (a road hypothesis) can be found in the image and whether the required positional accuracy is achieved. Potential errors are forwarded to a human operator who finally edits the database.

The automatic component suggested by Wiedemann & Mayer [1996] evaluates the grey value profiles perpendicular to the road axes stored in the database with respect to their correspondence with the parallel edge model described in Section 1.3.1. Klang [1998] proposed an approach that was applied to the Swedish national road database. First, road junctions are detected in the vicinity of the positions indicated by the database. Based on this result the junctions are used to initialize an active contour model. Finally, a comparison of the extraction result and the corresponding database provides the human operator with a number of potential errors. The results indicate that both methods are able to detect database errors. Unfortunately, Wiedemann & Mayer [1996]; Klang [1998] do not provide a quantitative evaluation.

Gerke et al. [2004] describe a verification approach that was applied to the German national database. First, lines are detected in the vicinity of the road hypotheses stored in the database. The *width* attribute is used to adjust the underlying line-detector [Steger, 1998] to each road object individually. Finally, a road object is accepted to be *correct* if a line is detected over the whole length of a road hypothesis. All other road hypotheses are considered as potential errors and have to be checked by a human operator. This also means that correct road hypotheses that cannot be explained by the line-based model fall into the second category. Gerke & Heipke [2008] enhanced the evaluation component in order to also detect inconsistencies with respect to road shape and width. This component provides a classification for each road hypothesis with respect to the classes *correct* and *incorrect*. Additionally, road hypotheses for which the line detection algorithm does not provide a solution over the full length, are classified as *potentially incorrect*. Beyond that, Gerke & Heipke [2008] do not discriminate between *incorrect* and *potentially incorrect objects* and consider both categories as *rejected* objects. Extensive experiments involving 530 road objects demonstrated the applicability of the approach in rural areas. In particular, 65% of the roads were correctly accepted while a sensitivity test with artificially generated errors resulted in a 100% rejection rate.

Poulain et al. [2010] propose a verification approach particularly for urban areas. Classifiers, relying on different input data such as SAR *corner lines*, *parallel edges*, *shadow*, *buildings*, *vegetation* and *sealed surface* are applied. Then, each database road object is projected to the classification results considering its *centreline* and *width*. Finally, each road hypothesis is classified on the basis of the detection results as being either *correct* or *incorrect*. Results based on the French national database show that their approach classifies 95% of the correct hypotheses as *correct*. A sensitivity test with artificially generated errors resulted in a 92% rejection rate, i.e. 92% of the non-roads are classified as *incorrect*. The lower rejection rate compared to [Gerke & Heipke, 2008] can be explained by the fact that Poulain et al. [2010] do not question the ability of the underlying road model to explain a situation. Instead, insufficiently modelled situations will also be classified, either as *correct* or as *incorrect*. This point must be seen as critical against the background of potential cartographic applications as it means that 8% of the artificially generated database errors remained undetected.

**Semi-automatic detection of missing roads:** Such approaches explicitly search for roads that are omitted in an existing database and forward their output as so-called *update hints* to a human operator who finally edits the database. Reasons for missing roads might be the construction of new roads or roads having been overlooked in prior updates. From the application viewpoint, *verification* and *the detection of missing roads* have to be performed together, e.g. for updating a database. In this regard, Auclair-Fortier et al. [2001] extended the approach of Klang [1998] by an additional method that detects lines in the entire scene. The resulting line segments are filtered mainly according to two criteria, a lower bound for the length and the requirement of a connection to the verified road network. The authors tested their method using the national Canadian road database. Unfortunately, they do not provide a quantitative evaluation of their experiments, but it becomes clear that the search for new roads is much more challenging than verifying existing database entries. Poulain et al. [2010] extend their verification method by a line-based extraction algorithm. The resulting line segments are classified by the same approach as in their verification component, just extended by an additional feature, the distance to the verified road network. The line segments classified to be *correct* finally represent the update hints. The experiments show that their approach detects 60% of the missing road objects, but they also have a false alarm rate of 50%. Beyen et al. [2012] describe a method for detecting new roads in rural areas. For that purpose, they apply landcover classification with respect to the two classes *settlement* and *vegetation*. As the authors assume new roads and buildings to be closely related, the classified settlement regions that neither correspond to roads nor to buildings in the old database are interpreted as update hints. The experiments show a detection rate of about 60% and a false alarm rate of 40%. The quantitative analyses of all authors show that *detecting missing roads* is more challenging than *road verification* and *geometric enhancement*. One reason is the lack of available prior knowledge for missing roads but also the fact that the cited approaches just consider a few road properties, i.e. a line based model [Auclair-Fortier et al., 2001; Poulain et al., 2010] or a colour-texture-based model [Beyen et al., 2012].

#### 1.3.4. Discussion

This thesis concentrates on *road verification* rather than on *geometric enhancement* or *the detection of missing roads*. The review of the related work has shown that this focus is justified because the three tasks favour different strategies. Also authors dealing with two of those tasks at once solve the problems based on different strategies that introduce different kinds of prior knowledge, e.g. [Poulain et al., 2010; Auclair-Fortier et al., 1999].

Gerke & Heipke [2008] demonstrate that a semi-automatic strategy can be implemented to solve the verification problem, even if the underlying road detection methods have a limited performance. The approaches described in [Gerke & Heipke, 2008] and [Poulain et al., 2010] show that the verification problem can effectively be transferred into a classification problem with respect to the two states *correct* and *incorrect*. This strategy allows integrating prior knowledge from the existing, possibly partially incorrect database, e.g. to specify road models for the detection by introducing *centreline geometry* and *road width*.

A deficit of both approaches, [Gerke & Heipke, 2008] and [Poulain et al., 2010], is their restriction

to specific context regions such as rural and urban areas. The reasons for the restriction can be found in the underlying road models that concentrate on just a few road properties. Zhang [2004] and Grote et al. [2012] present approaches to *road detection* (not *road verification*) that integrate many more different road properties. This is realized by combining methods based on different road models that were discussed in Section 1.3.1. These approaches can potentially deal with different context areas. However, the sequential combination concept applied by Zhang [2004]; Grote et al. [2012] is not useful to solve large scale verification problems as it requires complex parametrisation, and thus leads to unacceptable manual efforts for the adaptation of the methods to new scenes. As discussed in Section 1.3.2 concepts based on a combination of independent methods are better suited with respect to parametrisation efforts but have not been realized for many different road properties. One reason for that is the difficulty of defining consistent quality measures that control the impact of the single outputs on the combined solution.

In the related literature quality measures usually refer to the question whether an observation in the imagery corresponds to an underlying road model. How the correspondence can be determined, mainly depends on the applied methods. For instance, [Hinz & Baumgartner, 2003; Poulain et al., 2010] use by-products of the *detection* processes while other authors such as Gerke & Heipke [2008] also consider additional knowledge. As discussed in Section 1.3.2 the consistency of the quality measures is the most critical part for this kind of combination strategy. The more different the detection strategies of the methods are and the larger the number of the methods is, the more challenging is that problem. Therefore, most of the state-of-the-art approaches either restrict the number of methods, e.g. [Gerke & Heipke, 2008; Das et al., 2011], or only use methods that rely on similar detection strategies, e.g. [Hinz & Baumgartner, 2003; Poulain et al., 2010; Porway et al., 2010].

In conclusion, mainly two reasons can be identified why the state-of-the-art approaches do not meet the objectives formulated for this thesis:

- Road verification approaches, such as [Gerke & Heipke, 2008; Poulain et al., 2010] focus on just a few road properties that can only be observed under specific circumstances, e.g. within specific context regions.
- Road detection approaches, such as [Zhang, 2004; Grote et al., 2012] consider more different road properties but their concept to combine multiple road models leads to inappropriate parametrisation efforts.

Furthermore, it has to be noticed that none of the approaches in the related work explicitly considers *ignorance* as a possible solution even though some of those approaches explicitly consider existence of *ignorance* with respect to interim results [Mena & Malpica, 2005; Gerke, 2005; Gerke & Heipke, 2008; Poulain et al., 2010]. However, due to the objective of general applicability, *knowledge* about the presence of *ignorance* is of vital importance.

## 2. Basics

The basic strategy followed in this thesis is to solve a classification problem based on multiple independent methods that all provide *uncertain information* about the same classification problem. In the literature different meanings can be found for *uncertain* and for *information*. Therefore, the meaning of this phrase needs to be defined for this thesis:

**Definition:** In accordance with the Probability Theory, *uncertainty* means that the state of a random variable is expressed by a probability distribution. The term *information*, in this thesis, stands for any probability distribution that refers to the main classification problem. Hence, the phrase *uncertain information* means that the state of a primal probability distribution (the *information*) is expressed by a secondary probability distribution. A primal probability distribution has two possible states: Either it is *interpretable*<sup>1</sup> or it is *not interpretable* for the main classification problem.

Two major scientific problems arise from that. One problem concerns the combination of independent and *uncertain information* in order to find an optimal solution for the classification problem. In this thesis, this problem is tackled by a probabilistic approach based on the *Dempster-Shafer Theory*, which will be introduced in the first section of this chapter. The second problem concerns the determination of the required uncertainties. In this respect, supervised learning methods define their model on the basis of training data. If the training data does not properly represent a test sample, the solution provided by such a method is *not interpretable* in the context of the main classification problem. In Chapter 3 the probability of an SVM classifier to provide *interpretable* and *not interpretable* outputs will be defined on the basis of the feature space distance of a test sample to the training data. In the second section of this chapter a metric for the feature space distance on the basis of a *Support Vector Data Description* will be presented.

### 2.1. Combination of uncertain information

#### 2.1.1. The reasoning problem

Considering a classification problem with  $n$  mutually exclusive classes  $C_i$  with  $1 \leq i \leq n$ , the so-called *frame of discernment*  $\Theta$  is defined as:

$$\Theta = \{C_1, C_2, \dots, C_n\} \quad (2.1)$$

---

<sup>1</sup>Alternatively, one might think of the primal probability distribution having two possible states *correct* and *erroneous*, but note that if a functional relation of a given (*erroneous*) primal probability distribution and their *correct* version were known, the primal probability distribution would be *interpretable* by this functional relation. Hence, *not interpretable* also means that this functional relation is *unknown*, which finally means that it is impossible to infer anything useful from the primal probability distribution.

Further let consider  $\Theta$  to be a non-empty and exhaustive set where probabilities  $P$  are defined for each element  $C_i$ . The probabilities  $P(C_i) \in \mathbb{R}$  have to satisfy the *three Kolmogorov axioms*<sup>2</sup>: Any set  $F \subseteq \Theta$  must satisfy:

1. *non-negativity*:

$$P(A) \geq 0 \quad \forall A \in F \quad (2.2)$$

2. *normalization*:

$$P(\Theta) = 1 \quad (2.3)$$

3. *finite additivity*:

$$P(A \cup B) = P(A) + P(B) \quad \forall A, B \in F \wedge A \cap B = \emptyset \quad (2.4)$$

where  $\emptyset$  denotes the empty set.

Let there be  $S$  information sources that provide independent information to the introduced problem in the form of probabilities  $P$ . In the literature such a problem is often named a *statistical reasoning problem*, and the information is frequently denoted as *evidence* to the reasoning problem. Some authors such as Shafer [1976] also use the phrase *degrees of believe* instead of *evidences*. Both notations shall emphasize the fact that the *probabilities* are interpreted in *Bayesian* sense and not in *Frequentist* sense.

The most prominent reasoning approach is *Bayesian inference* which is completely founded in *Probability Theory*. In accordance with [Klein, 2004], each source  $s$  is interpreted to provide information in form of *likelihoods*  $P(x_s|C_i)$  for all  $i$ . Bayesian reasoning searches for  $P(C_i|x_1 \cap x_2 \cap \dots \cap x_S)$ , i.e. the probability for  $C_i$  given all available evidence. Based on the assumption of independence the *Combination Rule of Bayesian inference* is defined as:

$$P(C_i|x_1 \cap x_2 \cap \dots \cap x_S) = \frac{\left[ \prod_{1 \leq s \leq S} P(x_s|C_i) \right] \cdot P(C_i)}{\sum_{1 \leq j \leq n} \left( \left[ \prod_{1 \leq s \leq S} P(x_s|C_j) \right] \cdot P(C_j) \right)} \quad \forall C_i \in \Theta \quad (2.5)$$

where  $P(C_i)$  denotes the prior probability of a class  $C_i$ . The denominator normalizes the term so that the resulting posteriors satisfy the second Kolmogorov axiom (cf. Equation 2.3). The denominator becomes small if the information coming from the different sources is largely contradicting<sup>3</sup>. Note that total contradiction is impossible by design because in this case, the denominator would become zero. In Equation 2.5 the ability of an information source to solve the classification problem is expressed by the likelihoods  $P(x_s|C_i)$ . For instance, a *weak classifier*, i.e. one that is only slightly better than a random decision will result in likelihoods close to a uniform distribution. As a consequence, the information provided by such a classifier nearly corresponds to the prior probabilities. In order to prevent confusion, the term *weak* must consciously be discriminated from the

<sup>2</sup>The three axioms were defined in 1933 by Andrey N. Kolmogorov. The notation used in this thesis is in accordance with [Shafer, 1976].

<sup>3</sup>In this context, contradictions are a consequence of the independence and not of erroneous likelihoods.

term *uncertain* here.

Many applications require a hard decision for one of the classes  $C_i \in \Theta$  as their final result. In case of Bayesian inference the *maximum a posteriori probability* (MAP) criterion maps the resulting distribution to a distinct optimal state. In this respect, the difference of the resulting posteriors describes a property of the mapping which is sometimes denoted as an *uncertainty measure*. However, this phrase is also avoided here to prevent confusions with the preceding definition of *uncertain information*.

Next, it will be assumed that an information source  $s$  provides *interpretable* likelihoods  $P(x_s|C_i)$  for all  $i$ , only with the (known) probability  $P_s$ . This means that the actual class label is either *unknown* (with a probability  $1 - P_s$ ) or just *uncertain* (with a probability  $P_s$ ). However, a joint probability of  $P(x_s|C_i)$  and  $P_s$  does not exist in the classical Probability Theory which further means that Equation 2.5 is not applicable for the given problem. Fortunately, the problem can be solved with the Dempster-Shafer Theory (DST) that provides an expression for the joint probability and an adequate extension to Equation 2.5. In the three subsequent sections the characteristics of a DST-based reasoning in contrast to Bayesian reasoning will be introduced. A practical comparison of Bayesian reasoning and the DST is provided in the Appendix A.1.

### 2.1.2. The Dempster-Shafer Theory

The foundations of the DST were laid by Dempster [1967] and [1968]. Shafer [1976] outreached the theory to a broader community but also delimited their field of application. Since then, numerous elaborations and applications of the DST have been developed. Today, the term *Dempster-Shafer Theory* commonly addresses these works in their entirety, and thus includes many variations, e.g. [Gordon & Shortliffe, 1984; Smets, 1990; Dezert, 2002].

**Representation:** Referring to the initially defined reasoning problem with the frame of discernment  $\Theta$  (cf. Equation 2.1), the state space defined in the DST is defined as the power set of  $\Theta$ . This set, commonly denoted as  $2^\Theta$  includes all possible subsets of  $\Theta$  and the empty set  $\emptyset$ . For the case of  $n = 2$  classes  $\{C_1\}$  and  $\{C_2\}$ , the state space is defined as follows:

$$2^\Theta = \{C_1, C_2, C_1 \cup C_2, \emptyset\} \quad (2.6)$$

In analogy to probabilities, the amount of evidence that an information source assigns to an element  $A \in 2^\Theta$  is expressed by a *probability mass*<sup>4</sup>  $m_s(A)$ . The probability mass function  $m$  for a set  $A \in 2^\Theta$  has the following properties:

$$m(A) \in \mathbb{R} \wedge m(A) \geq 0 \quad \forall A \in 2^\Theta \quad (2.7)$$

$$\sum_{A \in 2^\Theta} m(A) = 1 \quad (2.8)$$

$$m(\emptyset) = 0 \quad (2.9)$$

<sup>4</sup>In the original publications of Dempster only the term *lower probability* can be found. Later, authors such as Shafer [1976] used the expression *mass* in analogy to the physical mass of an object to consider the evidence as a quantity that can move around, be split up, and combined.

As Shafer [1976] points out, the constraints 2.7–2.9 correspond to the Kolmogorov axioms (2.2–2.4), except the considered state space is  $2^\Theta$  for probability masses and  $\Theta$  for probabilities. In this regard, probability masses are also additive (cf. axiom 2.8), but not necessarily for states  $A \in \Theta$ , i.e. the states representing the classes  $C_i$ . Based on  $\Theta \subseteq 2^\Theta$  Haenni [2005] points out that the DST includes the Bayesian reasoning as a special case, where non-zero probability masses are only assigned to the states  $A \in \Theta$ . This is why Dempster [1968] introduced his theory as a generalization of Bayesian inference. Despite those similarities, some notations have very different meanings, e.g. in the Probability Theory the notation  $P(C_1 \cup C_2)$  means the evidence assigned to a set of two states  $\{C_1\}$  and  $\{C_2\}$ , whereas in the DST the notation  $m(C_1 \cup C_2)$  means the evidence assigned to a single state  $\{C_1 \cup C_2\}$  which excludes the states  $\{C_1\}$ ,  $\{C_2\}$ . In the DST evidences assigned to sets of states are expressed by help of additional functions, e.g. *support* (*sp*) and *plausibility* (*pl*) that will be described below.

**Dempster's rule of combination:** If  $S$  independent information sources provide probability masses  $m_s(B_s)$  with  $B_s \in 2^\Theta$  they are combined by applying *Dempster's rule* [Guan & Bell, 1991] which is the counter part to the Bayesian combination rule discussed above:

$$m(A) = \frac{\sum_{\cap_{s=1}^S B_s = A} \left[ \prod_{1 \leq s \leq S} m_s(B_s) \right]}{\sum_{\cap_{s=1}^S B_s \neq \emptyset} \left[ \prod_{1 \leq s \leq S} m_s(B_s) \right]} \quad \forall A \in \{2^\Theta \setminus \emptyset\} \quad (2.10)$$

In Equation 2.10 the sums are taken over the valid permutations described by the terms under the symbol for sum, i.e. the products are only defined for these permutations. The permutations defined in the enumerator are required to have the intersection  $A$ , while the permutations defined in the denominator are only required to have an intersection different from the empty set. Hence, the denominator normalizes the probability masses such that the resulting probability masses satisfy the constraint 2.8. According to Equation 2.5, the denominator becomes small if the information largely contradicts but is undefined in case of total contradiction. Dempster's combination rule corresponds to the Bayesian combination rule if non-zero probability masses are only assigned to the states  $A \in \Theta$ .

**Conflict mass:** The degree to which the different sources contradict can be expressed by the conflict mass [Shafer, 1976]:

$$K = \sum_{\cap_{s=1}^S B_s = \emptyset} \left[ \prod_{1 \leq s \leq S} m_s(B_s) \right] \quad (2.11)$$

The conflict mass  $K \in \mathbb{R} \wedge 0 \leq K < 1$  represents a meta information of the fusion process, and thus might be considered for internal evaluations. The conflict mass has the following relation to



the denominator in Equation 2.10:

$$K = 1 - \sum_{\cap_{s=1}^S B_s \neq \emptyset} \left[ \prod_{1 \leq s \leq S} m_s(B_s) \right] \quad (2.12)$$

**Support and Plausibility:** Given the probability masses  $m(B)$  for all  $B \in 2^\Theta$  the *support*  $sp(A)$  and the *plausibility*  $pl(A)$  can be defined for all  $A \in 2^\Theta$ :

$$sp(A) = \sum_{B \subseteq A} m(B) \quad (2.13)$$

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad (2.14)$$

The *support* to a set  $A$  aggregates all probability masses that directly provide evidence to  $A$ , whereas *plausibility* aggregates all probability masses that do not directly provide evidence against  $A$ . Based on Equations 2.7–2.9 and 2.13–2.14  $sp(A)$  and  $pl(A)$  have the following properties:

$$sp(A), pl(A) \in \mathbb{R} \wedge 0 \leq sp(A), pl(A) \leq 1.0 \quad \forall A \in 2^\Theta \quad (2.15)$$

$$\sum_{A \in 2^\Theta} sp(A) \geq 1.0, \sum_{A \in 2^\Theta} pl(A) \geq 1.0 \quad (2.16)$$

$$sp(\emptyset), pl(\emptyset) = 0 \quad (2.17)$$

In general *support* and *plausibility* are not additive (cf. Equations 2.16) which stands in contrast to *probability masses* (cf. Equation 2.8). Based on Equations 2.13 and 2.14, the following relations of  $sp(A)$  and  $pl(A)$  can be found:

$$sp(A) \leq pl(A) \quad (2.18)$$

$$sp(A) = 1 - pl(\bar{A}) \quad (2.19)$$

In Equation 2.19  $\bar{A}$  denotes the complementary set of  $A$  with respect to  $2^\Theta$ . *Support* and *plausibility* are the basis for the so-called *uncertainty interval*<sup>5</sup> shown in Figure 2.1. This *uncertainty interval* leads to a simplified representation of the probability mass distribution which might be helpful for analyses where the state space  $2^\Theta$  has large cardinality.



Figure 2.1.: The uncertainty interval.

<sup>5</sup>The notation *interval* is misleading because the *uncertainty interval* does not show the lower and upper bounds of the probability masses (or probabilities), instead it shows the probability mass distribution aggregated to only three sets. For any state  $A \in 2^\Theta$  the corresponding *uncertainty interval* has the same meaning: The left part shows the probability mass for  $A$  being *true*, the right part shows the probability mass for  $A$  being *false* and the centre part shows the probability mass for *not knowing* the state of  $A$ .

### 2.1.3. Decision making

Many applications of statistical reasoning and almost every scientifically oriented benchmark test finally require a hard decision for only one class  $C_i$  which means a hard decision for a state  $A \in \Theta$ . This implies a mapping from the intentionally defined state space  $2^\Theta$  back to  $\Theta$ , for which a variety of different mapping functions have been defined in the literature.<sup>6</sup> The three most prominent mapping functions, the *maximum support rule* [Shafer, 1976], the *maximum plausibility rule* [Shafer, 1976] and the *maximum mean rule* [Guan & Bell, 1991] will subsequently be discussed. In this regard, it is mainly interesting how these mapping functions deal with probability masses assigned to states  $A \in \Theta$  in contrast to the states  $A \in 2^\Theta \wedge A \notin \Theta$ .

**Maximum support rule:** The maximum support rule defines the decision state  $D$  as follows:

$$D = \arg \max_{A \in \Theta} sp(A) \quad (2.20)$$

This rule represents the predominantly applied mapping function in the related literature, e.g. [Lee et al., 1987; Gerke, 2005; Rottensteiner et al., 2007]. It simply provides the class  $C_i$  (a state  $A \in \Theta$ ) that has the largest probability mass.<sup>7</sup> It basically means that probability masses assigned to states that represent *ignorance* are ignored. In this respect, it was experimentally analysed by different authors, e.g. Lee et al. [1987]; Gerke [2005] that the *maximum support rule* leads to similar results as Bayesian reasoning, if the existence of erroneous probability distributions is basically ignored.

le Hégarat-Masclé et al. [1997] extend the *maximum support rule* (cf. Equation 2.20) by adding a *rejection-state* to the decision space:

$$D = \begin{cases} \arg \max_{A \in \Theta} sp(A) & | sp(A) > sp(\bar{A}) \\ rejection & | \text{otherwise} \end{cases} \quad (2.21)$$

Equation 2.21 implies that if the probability mass assigned to the complement (the set  $\bar{A}$ ) is larger than the probability mass assigned to the class with maximum probability mass (a state  $A \in \Theta$ ), the decision is *rejected*. le Hégarat-Masclé et al. [1997] demonstrated that their decision rule is applicable for a state space with a cardinality of  $\|2^\Theta\| = 8$ . However, it is an open question if this strategy is transferable to other problems.

**Maximum plausibility rule:** The maximum plausibility rule defines the decision state  $D$  as follows:

$$D = \arg \max_{A \in \Theta} pl(A) \quad (2.22)$$

Barnett [1991] favours this rule over the *maximum support rule* (cf. Equation 2.20) by arguing that the probability masses assigned to the states  $A \in 2^\Theta \wedge A \notin \Theta$  represent a potential voting against the correctness of the states  $A \in \Theta$ , and thus they need to be taken into account. The *maximum*

<sup>6</sup>This is in contrast to the Probability Theory, where the mapping is mostly realized by the maximum a posteriori probability (MAP) decision rule. This rule assumes the state space and the decision space are equivalent and its result is commonly denoted as Bayesian-optimal.

<sup>7</sup>As only the states  $A \in \Theta$  are considered the use of *probability mass* or *support* (in Equation 2.13) would lead to same result.

*plausibility rule* would lead to similar results as Bayesian reasoning, if the potentially erroneous likelihoods were replaced by a uniform distribution.

**Maximum mean rule:** The maximum mean rule defines the decision state  $D$  as follows:

$$D = \arg \max_{A \in \Theta} (sp(A) + pl(A)) \quad (2.23)$$

Poulain et al. [2011] suggest this mapping function by arguing that it represents a trade-off between the *maximum support rule* (cf. Equation 2.20) and the *maximum plausibility rule* (cf. Equation 2.22). Beyond that, it is hard to grasp its underlying assumption; it can merely be stated that the probability masses assigned to the states  $A \in \Theta$  are weighted somewhat higher than those that are assigned to the states  $A \in 2^\Theta \wedge A \notin \Theta$ .

**An alternative decision rule:** The mapping functions discussed above, do not necessarily provide DST-*optimal* decisions because there might always be a state  $A \in 2^\Theta$  with larger probability mass than the decision state  $D$ . Instead, the resulting decisions can be interpreted as *Bayesian-optimal* under specific assumptions, e.g. by assuming a uniform distribution. However, neither of the cited authors seems to be aware of the assumptions implicitly made by the decision rules or at least had interest to discuss this problem. Another critical question is: Why using the DST in favour of Bayesian reasoning if an answer representing *ignorance* is not of interest?

For the approach proposed in this thesis a new decision rule is defined that does not restrict the decision space to classes  $C_i$  (states  $A \in \Theta$ ). Instead the decision space includes all states  $A \in 2^\Theta$ . The new decision rule, the so-called *maximum probability mass rule*, is defined as follows:

$$D = \arg \max_{A \in 2^\Theta} m(A) \quad (2.24)$$

Equation 2.24 provides a DST-*optimal* result and does not imply additional assumptions, nor does it correspond to some Bayesian solution. However, the uncommon decision space needs to be discussed.

- Decisions  $D = A$  with  $A \in \Theta$  still have the common meaning: The state  $D$  represents the most likely class.
- Decisions  $D = A$  with  $A = \Theta$  mean that most likely *nothing is known*. This aspect will play a key role for the human-machine interface presented in Chapter 3.

A second specific property of the decision space defined in this thesis is the state *invalid* that is considered as additional state in the decision space. If information from different sources is available, the conflict of the information can be considered to indicate problems of the approach. In the DST the *conflict mass*  $K$  (cf. Equation 2.11) is a measure of conflicting information, whereas for Bayesian reasoning the denominator in Equation 2.5 can be interpreted in a similar way [Haenni, 2005]. In theory, conflicts are justified by the independence of the information sources. However, Haenni [2005] points out that in practical applications model errors may also lead to large conflict masses. In this regard, some authors, e.g. Zadeh [1984] and Ruspini et al. [1992], presented examples giving counter-intuitive results that emphasize typical problems of DST-based applications. In these examples, highly conflicting probability masses are combined which leads to a small value in the

denominator of Equation 2.10. Consequently, some authors suggested improvements to the DST and the given examples. A rather important work on this topic was provided by Smets [1990] who argued that the presence of a high conflict mass indicates that a possible answer has been overlooked. Thus, the frame of discernment can no longer be assumed to be exhaustive. As a consequence, Smets [1990] suggested an alternative approach that relaxes the strict closed-world assumption of the classical DST. The work presented by Dezert [2002] shows that a high conflict mass can also be caused by the presence of non-exclusive states in the frame of discernment. Therefore, Dezert [2002] proposed enhancements to the classical DST that basically allow to deal with non-exhaustive and non-exclusive sets. Haenni [2005] argues that the counter-intuitive results are directly related to the counter-intuitive assumption that highly conflicting information might be completely reliable, and thus suggests adjustments to the examples. In summary, nearly all attempted corrections of the counter-intuitive examples lead to the conclusion that the problem is caused by a misapplication of the DST; either the frame of discernment is not exhaustive and exclusive or the probability masses are not defined in an adequate way. This conclusion is in accordance with Haenni [2005] who remarks that a high *conflict mass* is a possible consequence of a violation of the pre-conditions for the DST. In this thesis a high *conflict mass* (cf. Equation 2.11) is interpreted as an indicator for an erroneous DST framework, and thus is considered as an exclusive state of the decision space called *invalid*. Using Equations 2.11 and 2.24 the new decision rule is defined as:

$$D = \begin{cases} \arg \max_{A \in 2^\Theta} m(A) & | K < g \\ \text{invalid} & | \text{otherwise} \end{cases} \quad (2.25)$$

where  $g \in \mathbb{R} \wedge 0 \leq g \leq 1$  is a parameter that needs to be defined in accordance with the cardinality of  $2^\Theta$  and the probability for an erroneous DST framework.

#### 2.1.4. Derivation of probability masses

An essential question for any application based on the DST is how to define the probability masses. Several models and strategies to define these probability masses have been developed.

le Hégarat-Masclé et al. [1997] consider a reasoning problem for different classes  $C_i$ , where a set of probabilistic classifiers depending on different data  $x_s$  provide probabilistic measures  $P(C_i|x_s)$ . Basically they assign the available probabilities directly to probability masses, i.e.  $m_s(C_i) = P(C_i|x_s)$ . However, le Hégarat-Masclé et al. [1997] define exceptions from this kind of assignment if two classes  $\{C_i\}, \{C_j\}$  cannot clearly be distinguished by a classifier. For these pairs of classes they assign the full evidence to the state  $\{C_i \cup C_j\}$  which leads to  $m_s(C_i) = m_s(C_j) = 0$  and  $m_s(C_i \cup C_j) = P(C_i|x_s) + P(C_j|x_s)$ . The authors also present a second alternative, splitting up the evidence to three states  $\{C_i\}, \{C_j\}, \{C_i \cup C_j\}$ , which leads to  $m_s(C_i) = m_s(C_j) = m_s(C_i \cup C_j) = 3^{-1} (P(C_i|x_s) + P(C_j|x_s))$ .

In contrast to le Hégarat-Masclé et al. [1997], Lee et al. [1987] define a value  $\alpha_s \in \mathbb{R} \wedge 0 \leq \alpha_s \leq 1$  for each classifier, interpreting  $\alpha_s$  to be the expected correctness of the classification outputs. The values for  $\alpha_s$  are experimentally defined on the basis of a training dataset using cross validation,

and thus automatically adjust the approach to different datasets based on the following definition:

$$m_s(A) = \begin{cases} \alpha_s \cdot P(C_i|x_s) & | A = C_i \\ 1 - \alpha_s & | A = \Theta \\ 0 & | \text{otherwise} \end{cases} \quad \forall A \in 2^\Theta \quad (2.26)$$

In Equation 2.26 the first case defines the probability masses for all  $A \in \Theta$ , i.e. all classes  $C_i$  contained in  $\Theta$  (cf. Equation 2.1). The second case defines the probability mass assigned to the state of *ignorance*  $\Theta$ , i.e. the state corresponding to  $\{C_1 \cup C_2 \cup \dots \cup C_n\}$ . The third case defines the probability masses for all remaining states, i.e. the states  $A$  with  $A \subset \Theta \setminus A \in \Theta$  that represent *partial ignorance*.

Gerke [2005] mainly follows the concept introduced by Lee et al. [1987] but uses a different strategy to define the  $\alpha_s$ . In [Gerke, 2005] the  $\alpha_s$  describe the ability of the respective classifier to contribute to the main classification problem. The model for the  $\alpha_s$  is based on a heuristic.

All three approaches discussed so far interpret probability masses on the basis of probabilistic measures related to  $\Theta$ . As soon as the  $\alpha_s$  in Equation 2.26 are interpreted as probabilities<sup>8</sup> that describe the states of the  $P(C_i|x_s)$ , those approaches are in accordance with the theory of *multi-valued mapping* described in [Dempster, 1967].

Apart from Dempster's concept of *multi-valued mapping*, several strategies for defining probability masses have been developed. For instance, Rottensteiner et al. [2007] define the probability masses without probabilities as the basis. They present five independent two-class classifiers that directly provide probability masses to different pairs  $A, \bar{A} \in 2^\Theta$  with respect to  $\Theta = \{C_1, C_2, C_3, C_4\}$ . One important aspect is that neither  $A$  nor  $\bar{A}$  must represent class  $C_i \in \Theta$ ; instead they are just assumed to represent complementary sets of classes. Each classifier relies on specific features  $x_s$  and provides  $m_s(A)$ ,  $m_s(\bar{A})$  and  $m_s(\Theta)$  in accordance with the function depicted in the left part of Figure 2.2. The set of variables  $\{a, b, c, d, d'\}$  denotes the parameters of the functions that rely on a heuristic model.

Another approach that defines probability masses without a probabilistic background is described in [Tupin et al., 1999; Poulain et al., 2011]. Similarly to [Rottensteiner et al., 2007] several two-class classifiers are defined. Each of them provides probability masses  $m_s(A)$ ,  $m_s(\bar{A})$  and  $m_s(\Theta)$  using the model depicted in the right part of Figure 2.2. The structure of the model is based on the assumption that the classification output is more useful far away from the decision threshold than near the decision threshold. Tupin et al. [1999] determine the parameters  $\{a, b, c, d, d'\}$  by fitting the three functions to a histogram drawn from a training data set. Poulain et al. [2011] apply an optimization to determine the parameters, where the objective function aggregates two terms. The first term minimizes the squared empirical error and the second term maximizes the square of the correct class assignments. Combining both objectives is important because minimizing the empirical error alone has a trivial solution:  $m(\Theta) = 1.0$ , whereas only maximizing correct class assignments does not induce assignments to (partial) ignorance.

<sup>8</sup>Note that the  $\alpha_s$  fulfil the three Kolmogorov axioms (cf. Equations 2.2–2.4).

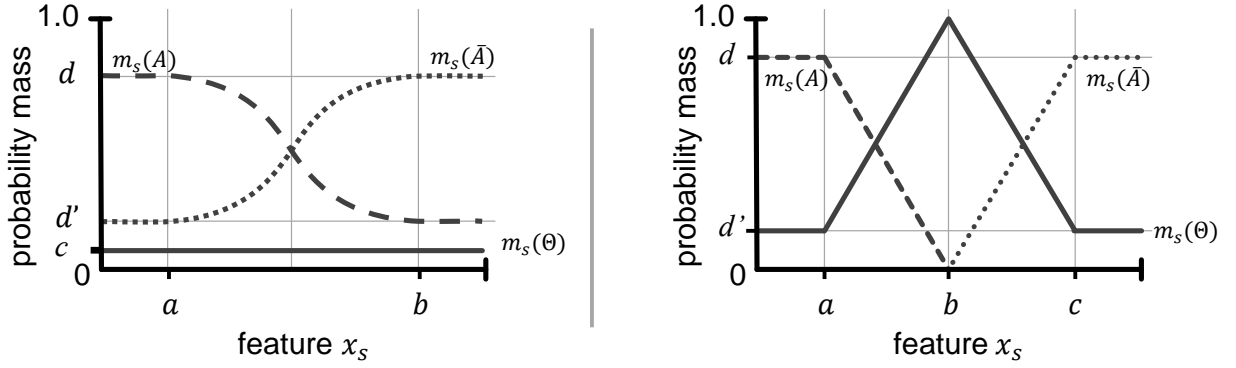


Figure 2.2.: Models for probability masses. Left: [Rottensteiner et al., 2007]. Right: Poulain et al. [2011].

## 2.2. An uncertainty metric for Support Vector Machines

In the field of remote sensing, support vector machine classifiers (SVM; Vapnik, 1995) have empirically been proven to be effective (e.g. in Huang et al., 2002; Foody & Mathur, 2004; Oommen et al., 2008; Bruzzone & Marconcini, 2010). Bruzzone & Marconcini [2010] attribute the success of SVMs to the following properties:

- the convexity of the objective function which allows to find an optimal solution
- the sparse representation of the model that is restricted to a subset of the training data
- the capability of addressing classification problems without defining class-related distributions
- the capability to solve not linearly separable classification problems.

Noticing other state-of-the-art machine learning approaches such as AdaBoost [Freund & Schapire, 1995] and random forests [Breiman, 2001] to be comparably effective for classification problems in the field of remote sensing [Schindler, 2012], two variants of SVMs will be applied in this thesis. The first variant is a conventional two-class SVM which is applied in the same way as in the related work, e.g. [Fujimura et al., 2008]. The second variant is a specific type of a one-class SVM which is the basis for the uncertainty metric discussed in this section. The following introductions refer to this second variant even if some of the conceptional aspects hold out for SVMs in general.

### 2.2.1. The problem of partly non-representative training data

Machine learning methods define a model on the basis of training data. Having representative training data is often considered to be a pre-condition, which is only realistic if the *closed world assumption* is correct. For many remote sensing problems this is not necessarily the case, and thus the problem is worth to be considered.

Figure 2.3 shows a situation in which a conventional two-class SVM is applied. The (symbolic) two-dimensional feature space is separated by a hypersurface (bold black line) which is learned from

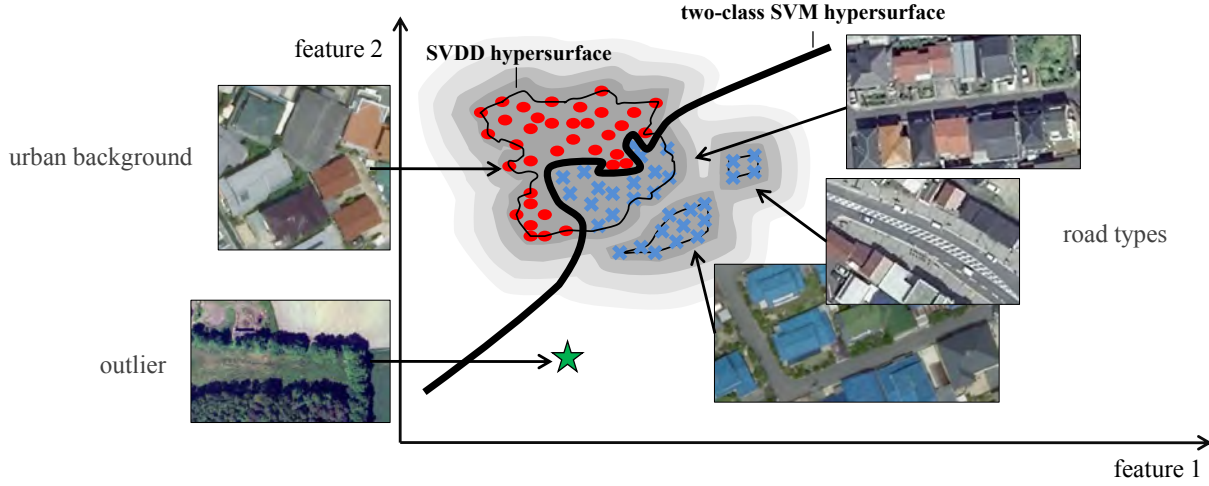


Figure 2.3.: Motivation for SVDD.

training samples corresponding to the two classes: *road* (blue) and *non-road* (red). A test object to be classified, such as the one indicated by the green star, may be far away from the training clusters, and as a consequence, the classification result may be erroneous without any indication by the classification algorithm. In particular, the distance to the two-class SVM hypersurface separating the two classes (bold black line) does not provide an appropriate measure to answer the question whether the training data are representative for a test sample or not.

This problem can be interpreted as a one-class problem with respect to the training data that include both classes *roads* and *non-road*. A test sample that resides far away from the training data as the green star depicted in Figure 2.3 represents an *outlier* while a test sample located near the training data would represent an *inlier*. The Support Vector Data Description (SVDD, Tax & Duin, 2004)<sup>9</sup> is a one class classifier that allows to separate *inliers* from *outliers* by only providing training data for the *inliers*. In Figure 2.3 the SVDD hypersurface is drawn by thin black lines around the training data. The grey regions inside the SVDD hypersurface indicate positions of *inliers* while the rest of the feature space indicates possible positions of *outliers*. During the testing phase, points outside the indicated regions, such as the green star corresponding to a row of trees, are considered to be *outliers*. Note that the conventional two-class SVM classifies it as a correct road. However, the SVDD hypersurface is not necessarily a good decision surface to separate *inliers* from *outliers* because test samples very near the SVDD hypersurface, i.e. those that reside in shaded area depicted in Figure 2.3, might still be represented by the training data. In this thesis, the SVDD hypersurface will deliver the basis for a metric, but will not be the metric itself. This requires some additional investigations of kernel space metrics that will be presented in Section 2.2.4. Before that, the theoretical and practical basics of the SVDD will be presented briefly.

<sup>9</sup>Tax & Duin [2004] show that an alternative one-class SVM introduced by Schölkopf et al. [2001] leads to similar classification results if comparable parameters are selected. However, the metrics discussed in this thesis are only valid for the method described in [Tax & Duin, 2004].

### 2.2.2. The Support Vector Data Description

In accordance with Tax & Duin [2004], the SVDD approach constructs a hypersphere (a sphere in a multi-dimensional space) of minimum volume enclosing most of the training data (cf. Figure 2.4). The problem of finding the hypersphere of minimum volume represented by its centre  $a$  and radius  $R$ , is formulated as:

$$\begin{aligned} \min_{R,a,\xi_i} \quad & R^2 + C \cdot \sum_{1 \leq i \leq n} \xi_i \\ \text{s.t.} \quad & (||x_i - a||^2 \leq R^2 + \xi_i) \wedge (\xi_i \geq 0) \quad \forall i = 1, \dots, n \end{aligned} \quad (2.27)$$

In Equation 2.27 the training data are denoted as feature vectors  $x_i \in X^{\text{train}}$  where the training dataset  $X^{\text{train}}$  has the cardinality  $n$ . The slack variables  $\xi_i$  allow a number of training samples to lie outside the hypersphere, and thus make the approach robust against isolated samples. The parameter  $C$  is a trade-off constant controlling the relative importance of the two terms.

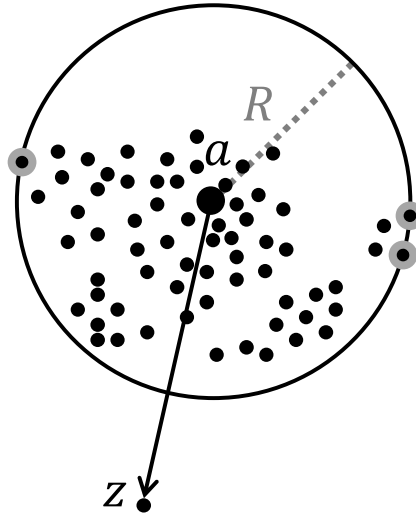


Figure 2.4.: A 2D-Hypersphere (centre  $a$ , radius  $R$ ) with enclosed training data and test object  $z$  outside the sphere.

**Training of the SVDD:** The constrained minimization problem in Equation 2.27 is solved by introducing the Lagrangian multipliers  $\alpha_i$  and  $\gamma_i$ . This leads to the following formulation:

$$\min_{R,a,\alpha,\gamma,\xi} L(R, a, \alpha, \gamma, \xi) \quad (2.28)$$

with

$$\begin{aligned} L(R, a, \alpha, \gamma, \xi) = & R^2 + C \sum_{1 \leq i \leq n} \xi_i - \sum_{1 \leq i \leq n} [\alpha_i (R^2 + \xi_i - (\langle x_i, x_i \rangle - 2\langle x_i, a \rangle + \langle a, a \rangle))] - \sum_{1 \leq i \leq n} [\gamma_i \xi_i] \\ \text{s.t.} \quad & (\alpha_i \geq 0) \wedge (\gamma_i \geq 0) \quad \forall i \end{aligned} \quad (2.29)$$



where  $\langle \cdot, \cdot \rangle$  denotes the inner product of two vectors. Setting the partial derivatives in Equation 2.29 to zero gives the following constraints:

$$\frac{\partial L(R, a, \alpha, \gamma, \xi)}{\partial R} = 0 : \quad \sum_{1 \leq i \leq n} \alpha_i = 1 \quad (2.30)$$

$$\frac{\partial L(R, a, \alpha, \gamma, \xi)}{\partial a} = 0 : \quad a = \sum_{1 \leq i \leq n} \alpha_i x_i \quad (2.31)$$

$$\frac{\partial L(R, a, \alpha, \gamma, \xi)}{\partial \xi_i} = 0 : \quad C - \alpha_i - \gamma_i = 0 \quad (2.32)$$

Using Equation 2.32 the Lagrange multipliers  $\gamma_i$  can be substituted by  $\gamma_i = C - \alpha_i$  which leads to the new constraint  $0 \leq \alpha_i \leq C$ . Re-substituting Equations 2.30–2.32 into 2.29 leads to the *Lagrangian dual form* of the optimization problem:

$$\max_{\alpha} L(\alpha) \quad (2.33)$$

with

$$L(\alpha) = \left[ \sum_{1 \leq i \leq n} \alpha_i \langle x_i, x_i \rangle - \sum_{1 \leq i, j \leq n} \alpha_i \alpha_j \langle x_i, x_j \rangle \right] \quad (2.34)$$

s.t.  $(0 \leq \alpha_i \leq C) \quad \forall i \wedge \sum_{1 \leq i \leq n} \alpha_i = 1$

Equation 2.33 describes a convex quadratic programming problem with the Lagrangian multipliers  $\alpha_i$  as the unknowns. For such a problem efficient algorithms have been proposed, for instance the one described in [Burges, 1998], which delivers solutions that are globally optimal. In the literature this solution is frequently denoted as a sparse solution, because the  $\alpha_i$  that correspond to training samples inside the optimal hypersphere are equal to zero, and thus play no role for the definition of the hypersphere centre  $a$  (cf. Equation 2.31). The training samples  $x_i \in X^{\text{train}}$  corresponding to non-zero Lagrangian multipliers  $\alpha_i$  are called *support vectors*  $X^{\text{SV}}$  of the hypersphere. Depending on the parameter  $C$ , a fraction of the support vectors still reside outside the hypersphere. These support vectors  $x_i$  can easily be identified by their coefficients  $\alpha_i = C$  [Tax & Duin, 2004]. Only the so-called *bounded support vectors*  $X^{\text{bndSV}} = \{x_i \in X^{\text{train}} | 0 < \alpha_i < C\}$  reside on the hypersphere surface, and can thus be used to determine the radius:

$$R^2 = \|x_l - a\|^2 \quad \forall x_l \in X^{\text{bndSV}} \quad (2.35)$$

Plugging Equation 2.31 into 2.35 gives:

$$R^2 = \langle x_l, x_l \rangle - 2 \sum_{1 \leq i \leq n} [\alpha_i \langle x_l, x_i \rangle] + \sum_{1 \leq i, j \leq n} [\alpha_i \alpha_j \langle x_i, x_j \rangle] \quad \forall x_l \in X^{\text{bndSV}} \quad (2.36)$$

In theory the distances of all bounded support vectors  $X^{\text{bndSV}}$  to the centre  $a$  are identical but for practical reasons there might be small divergences. Thus, commonly the mean distance to all bounded support vectors  $X^{\text{bndSV}}$  is calculated to define the radius  $R$ . The computation of  $R$  concludes the training step, and hence provides the model for the test case.

**The test case:** A test sample  $z$  lies inside or on the bound of the previously defined hypersphere if it satisfies the following inequality:

$$R^2 - \|z - a\|^2 \geq 0 \quad (2.37)$$

Plugging Equation 2.31 into 2.37 defines the squared distance to the centre:

$$\|z - a\|^2 = \langle z, z \rangle - 2 \sum_{1 \leq i \leq n} [\alpha_i \langle z, x_i \rangle] + \sum_{1 \leq i, j \leq n} [\alpha_i \alpha_j \langle x_i, x_j \rangle] \quad (2.38)$$

For a small number of support vectors the computational complexity of the test is very low which is an advantage of the SVDD over comparable one-class methods such as kernel density estimation [Tax & Duin, 2004].

**The implicit feature space mapping:** As a simply shaped form like a hypersphere is not necessarily a good description, the data are mapped to a new representation with a function  $\Phi : \mathbb{R}^d \rightarrow \mathcal{H}$ , where  $\mathbb{R}^d$  is the  $d$ -dimensional input feature space and  $\mathcal{H}$  is some (possibly infinite-dimensional) Hilbert space<sup>10</sup> [Burges, 1998]. The hypersphere model is assumed to fit much better to the mapped feature vectors  $\Phi(x_i)$  in  $\mathcal{H}$  than to the original feature vectors  $x_i$  in  $\mathbb{R}^d$ . However, an explicit computation in  $\mathcal{H}$  could be problematic as the dimensionality of an appropriate space  $\mathcal{H}$  might be very high. Analogously to the conventional two-class SVM, the *Kernel Trick* is applied to overcome this problem. In Equations 2.34, 2.36 and 2.38, the inner products  $\langle \cdot, \cdot \rangle$  are replaced by a kernel function  $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ . Now, the formulation as the Lagrangian dual form (Equation 2.34) and the consideration of squared distances in Equations 2.36 and 2.38 becomes clear, as all the feature vectors only occur in the form of inner products. Therefore, the Kernel Trick allows to carry out the whole SVDD approach without explicitly defining  $\mathcal{H}$  and  $\Phi$ , and thus without much additional computational complexity.

In accordance with Mercer's condition [Vapnik, 1995], any symmetric and positive semi-definite function represents an inner product in a Hilbert space, and thus is a valid kernel function. Hence, various kernel functions have been developed but not all meet the initial assumption, according to which the training data should (implicitly) be mapped into a spherically shaped area in  $\mathcal{H}$ . Tax & Duin [2004] showed that the radial basis function (RBF) satisfies the assumption quite well. The RBF is defined as follows:

$$K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \quad \text{s.t. } \sigma \neq 0 \quad (2.39)$$

Note that in general the inner product is a similarity measure in  $\mathcal{H}$ , and hence the kernel function is also a similarity measure based on  $\mathbb{R}^d$ . In case of the RBF kernel (Equation 2.39) the similarity of two feature vectors depends on their Euclidean distance  $\|x_i - x_j\|$  in  $\mathbb{R}^d$ . Using the RBF kernel, also denoted as Gaussian kernel, simply corresponds to the assumption that Euclidean distances in input feature space  $\mathbb{R}^d$  are locally meaningful. The term *local* is related to the parameter  $\sigma$  that will be discussed later. From a theoretical viewpoint, such an assumption seems to be quite meaningful for many applications, and therefore it is not a big surprise that also experimental analyses mostly

<sup>10</sup>A Hilbert space is any linear space with an inner product that is complete with respect to the corresponding norm. A Euclidean space is a Hilbert space with finite dimension.

demonstrate superiority of the RBF kernel compared to alternative kernel functions, e.g. [Huang et al., 2002]. Those alternative kernel functions often imply a more specific structure of the input feature space. The most prominent alternative, the polynomial kernel  $K(x, y) = (\langle x, y \rangle + 1)^p$  relies on the inner product in  $\mathbb{R}^d$ , which itself depends on the distances  $\|x\|$  and  $\|y\|$  from the origin in  $\mathbb{R}^d$ , and thus differences between samples that are far away from the origin are weighted higher than those near the origin. The impact of this weighting effect depends on the value of  $p$  as Tax [2001] has theoretically proven. In this thesis, only the RBF kernel will be applied because its theoretical background let expect transferability between different datasets.

Introducing the RBF property that  $K(x, x) = 1$ , the optimization problem in Equation 2.33 can be rewritten as:

$$\begin{aligned} \max_{\alpha} \quad & - \sum_{1 \leq i, j \leq n} \left[ \alpha_i \alpha_j \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \right] \\ \text{s.t.} \quad & (0 \leq \alpha_i \leq C) \forall i \wedge \sum_{1 \leq i \leq n} \alpha_i = 1 \end{aligned} \quad (2.40)$$

In accordance with Equations 2.36 and 2.38 the hypersphere radius  $R_{\mathcal{H}}$  and the Euclidean distance of a test sample  $\Phi(z)$  to the hypersphere centre  $a_{\mathcal{H}}$  are defined as follows:

$$R_{\mathcal{H}}^2 = 1 - 2 \sum_{1 \leq i \leq n} \left[ \alpha_i \exp \left( -\frac{\|x_l - x_i\|^2}{2\sigma^2} \right) \right] + \sum_{1 \leq i, j \leq n} \left[ \alpha_i \alpha_j \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \right] \quad (2.41)$$

$$\|\Phi(z) - a_{\mathcal{H}}\|_{\mathcal{H}}^2 = 1 - 2 \sum_{1 \leq i \leq n} \left[ \alpha_i \exp \left( -\frac{\|z - x_i\|^2}{2\sigma^2} \right) \right] + \sum_{1 \leq i, j \leq n} \left[ \alpha_i \alpha_j \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \right] \quad (2.42)$$

The index  $\mathcal{H}$  in Equations 2.41 and 2.42 shows that these variables are explicitly defined in  $\mathcal{H}$  and not in  $\mathbb{R}^d$ . However, the entities in  $\mathcal{H}$  are all defined by feature vectors given in  $\mathbb{R}^d$ . Finally, a test sample is classified as an *inlier* if it satisfies the following inequality:

$$R_{\mathcal{H}}^2 - \|\Phi(z) - a_{\mathcal{H}}\|_{\mathcal{H}}^2 \geq 0 \quad (2.43)$$

The implicit feature space mapping leads to a tighter description of the training data in  $\mathbb{R}^d$  than the original form. How tight the description is and how many support vectors are required to define it mainly depends on the parameters  $\sigma$  and  $C$  that will be discussed below.

### 2.2.3. Hyperparameter adjustment

If the RBF kernel is applied, the two hyperparameters  $\sigma$  (Equation 2.39) and  $C$  (Equation 2.27) have to be adjusted to find an appropriate solution for a specified problem. For conventional two-class SVMs, several algorithms for an automatic adjustment exist, where the possible parameter combinations are systematically evaluated on the basis of cross validation, e.g. [Hsu et al., 2003]. As cross validation cannot be applied to the one-class problem, theoretical models have to be defined. Therefore, a deeper understanding of the hyperparameters is required.

**Kernel width  $\sigma$ :** The character of the solution of the optimization problem in Equation 2.40 depends on the term:

$$\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2.44)$$

For all  $i = j$  the Term 2.44 equals to one, irrespective of what value is chosen for  $\sigma$ . For all  $i \neq j$  the Term 2.44 depends on the kernel width  $\sigma$ . For very small values of  $\sigma$  in the order of  $\min_{i,j} \|x_i - x_j\| \forall i \neq j$ , the Term 2.44 converges to zero for all  $i \neq j$ . As a consequence, the elements with  $i = j$  dominate the optimization, a case for which Tax & Duin [2004] have shown that Equation 2.40 has a stable maximum at  $\alpha_i = n^{-1}$  for all  $i$ , i.e. each training sample is a support vector. Hence, selecting such a small value for  $\sigma$  will lead to over-fitting. For the other extreme case, where  $\sigma$  is selected very large, i.e. in the order of  $\max_{i,j} \|x_i - x_j\|$ , the Term 2.44 converges to one for all  $i \neq j$  or is equal to one for all  $i = j$ . Tax & Duin [2004] have shown that Equation 2.40 then has a stable maximum that always corresponds to a spherical description in  $\mathbb{R}^d$ . Hence, selecting such a large value for  $\sigma$  will lead to under-fitting. Tax & Duin [2004] argue that the best choices for  $\sigma$  are moderate values, e.g.  $\text{mean}_{i,j} \|x_i - x_j\| \forall i \neq j$ . As  $\sigma$  marks the inflection point of a Gaussian function that is centred at zero (cf. Equation 2.44), the term is most sensitive to distances  $\|x_i - x_j\| \approx \sigma$ , a case for which Tax & Duin [2004] have shown that only the few most dissimilar training samples, i.e. those with largest  $\|x_i - x_j\|$  become support vectors. As this agrees rather well with the objectives of most applications, many authors such as Park et al. [2007] use the following definition for  $\sigma$  that will also be applied in this thesis:

$$\sigma = \sqrt{\frac{1}{n(n-1)} \sum_{1 \leq i,j \leq n} \|x_i - x_j\|^2} \quad (2.45)$$

Figure 2.5 displays the relationship between  $\sigma$ , the amount of support vectors and the tightness of the SVDD boundary, where the training samples  $X^{\text{train}}$  are drawn as black dots and the support vectors  $X^{\text{SV}}$  are highlighted by grey circles, while the decision surfaces are symbolized by black lines.

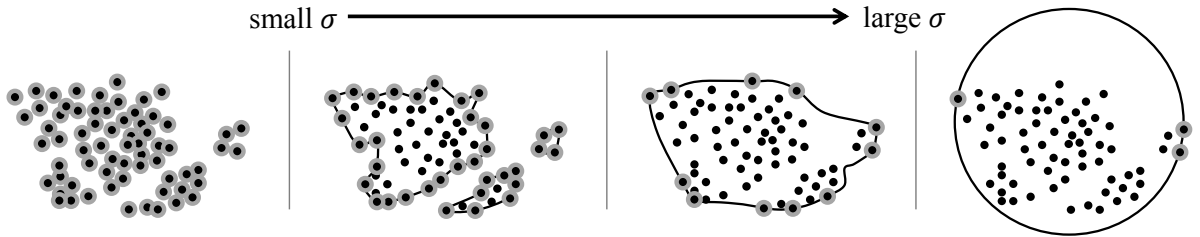


Figure 2.5.: Influence of kernel width  $\sigma$  on the SVDD.

**Regularization parameter  $C$ :** The second parameter to be adjusted is  $C$  which balances the length of the radius with the tolerated empirical error introduced as  $\xi = \{\xi_1 + \dots + \xi_n\}$  in Equation 2.27. All training samples outside the SVDD hypersphere are characterized by Lagrangian multipliers  $\alpha_i = C$ . Additionally, the Lagrangian multipliers are limited by the constraint  $\sum_{1 \leq i \leq n} \alpha_i = 1$  (Equation 2.40) which allows to define a relation between the number of training samples outside the SVDD hypersphere  $n^{\text{out}}$  and  $C$ , namely  $C = (n^{\text{out}})^{-1}$  [Tax & Duin, 2004]. Hence, an empirical error  $\xi = 0$  can be enforced by setting  $C > 1.0$ , whereas a value  $0 < C \leq 1.0$  leads to an empirical error  $\xi > 0$ ,

which allows the most isolated training samples to reside outside the SVDD hypersphere. Tax & Duin [2004] use the relation between  $n^{\text{out}}$  and  $C$  to define:

$$C = \frac{1}{\nu \cdot n} \quad (2.46)$$

where  $\nu$  is the fraction of training samples outside the SVDD hypersphere. They argue that a proper selection of  $\nu$  has to go in accordance with the overlap of the *inlier* and the *outlier* distributions. Hence, as soon as the relative amount of this overlap  $\nu$  is known the optimal value for  $C$  follows from Equation 2.46.

#### 2.2.4. Kernel space metric

Due to the kernel mapping, two different representations have to be considered: Firstly, the original input space, here denoted as  $\mathbb{R}^d$ , where all the feature vectors are defined. Secondly, the possibly infinite dimensional Hilbert space, here denoted as  $\mathcal{H}$ , where the Euclidean distances of the feature vectors to the hypersphere centre are defined (cf. Equations 2.41 and 2.42). The input feature space is usually defined in accordance with a specific model, and thus allows a natural interpretation of those Euclidean distances or at least of their relative amounts. In contrast, distances in the possibly infinite dimensional Hilbert space do not have that clear geometrical interpretation. Thus, the goal is to derive the distance of a test sample  $z$  to the SVDD hypersurface in  $\mathbb{R}^d$ . However, the SVDD hypersurface potentially has a complex shape in  $\mathbb{R}^d$  which makes its computation in  $\mathbb{R}^d$  difficult. Hence, it is advantageous to determine the closest distance in  $\mathcal{H}$ , where the SVDD hypersurface has a simple geometrical shape (a hypersphere) and then transfer the distance back to  $\mathbb{R}^d$  where it can be interpreted more easily. For that purpose, some general properties of the mapping and the relations of the representations in  $\mathbb{R}^d$  and in  $\mathcal{H}$  are required.

Even if  $\Phi$  is not explicitly known, some useful properties of the feature vector representation in  $\mathcal{H}$  can be defined, because the inner product in  $\mathcal{H}$  is known by its functional relation to the Euclidean distance in  $\mathbb{R}^d$ , see [Burges, 1999]. In case of the RBF, this means that all input feature vectors  $x_i \in \mathbb{R}^d$  are mapped onto a  $d$ -dimensional manifold in  $\mathcal{H}$ . It is further known that this  $d$ -dimensional manifold has a constant distance to the origin that is equal to one, which is easy to show as for all  $x_i \in \mathbb{R}^d$  the norm<sup>11</sup> becomes one; Figure 2.6 shows the situation in 2D. The SVDD hypersphere covers the part of the manifold that represents the training data in  $\mathcal{H}$ . Any vectors that may not reside on the manifold, such as the hypersphere centre  $a_{\mathcal{H}}$  in Figure 2.6, have no representation in the input feature space.<sup>12</sup> Hence, distances to  $a_{\mathcal{H}}$ , i.e. the radius  $R_{\mathcal{H}}$  of the hypersphere (Equation 2.41) and the distance  $\|\Phi(z) - a_{\mathcal{H}}\|_{\mathcal{H}}$  (Equation 2.42) do not have a straight-forward interpretation in  $\mathbb{R}^d$  either.

With respect to the representation in  $\mathcal{H}$ , the situation is rather simple. A point  $s_{\mathcal{H}}$  on the decision surface that has the closest distance to a test sample  $\Phi(z)$  is situated at a straight line through the hypersphere centre  $a_{\mathcal{H}}$  and the test sample  $\Phi(z)$ . The distance of  $s_{\mathcal{H}}$  to the hypersphere centre  $a_{\mathcal{H}}$  is given by the radius  $R_{\mathcal{H}}$ . But caution, the point  $s_{\mathcal{H}}$  does not necessarily reside on the manifold, and

<sup>11</sup>The norm  $\|x_i\|$  of a real vector  $x_i$  equals the squared root of the inner product of the vector with itself.

<sup>12</sup>In the literature feature vectors  $x_i \in \mathbb{R}^d$  are denoted as pre-images of vectors  $\Phi(x_i) \in \mathcal{H}$ , e.g. [Williams, 2002].

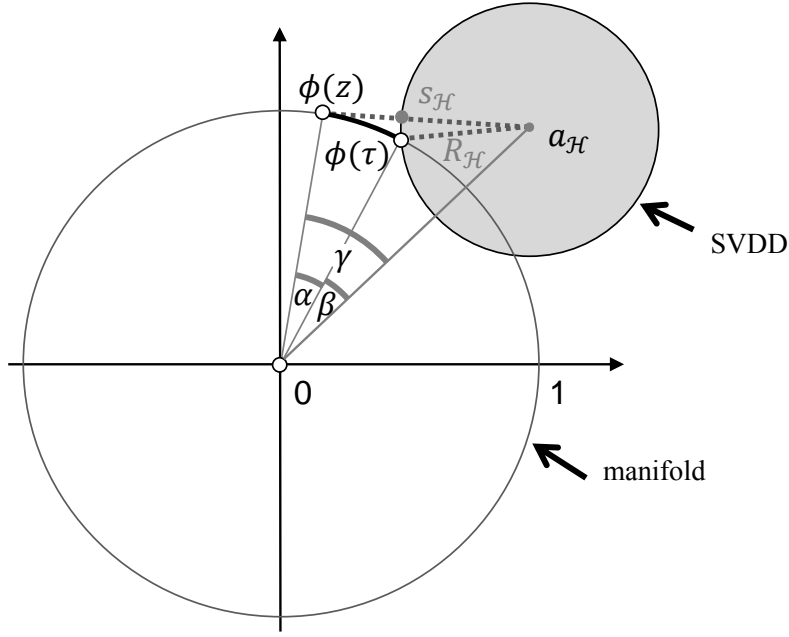


Figure 2.6.: 2D projection of the SVDD.

thus might have no representation in  $\mathbb{R}^d$ . Therefore, an additional constraint has to be defined: The point has to reside on the manifold. The point subsequently denoted by  $\Phi(\tau)$  is the nearest point to  $\Phi(z)$  that resides on the SVDD hypersurface and on the manifold (cf. Figure 2.6). Its representation in  $\mathbb{R}^d$  (pre-image) is denoted by  $\tau$ .

The following relationship is required to show that searching for the shortest distance of  $z$  and  $\tau$  in  $\mathcal{H}$  is equivalent to searching for their shortest distance in  $\mathbb{R}^d$ :

$$\|\Phi(z) - \Phi(\tau)\|_{\mathcal{H}}^2 = 2 \left( 1 - \exp \left( -\frac{\|z - \tau\|_{\mathbb{R}^d}^2}{2\sigma^2} \right) \right) \quad (2.47)$$

Equation 2.47 is derived by expanding the squared  $\mathcal{H}$ -distance  $\|\Phi(z) - \Phi(\tau)\|_{\mathcal{H}}^2$  to  $\langle \Phi(z), \Phi(z) \rangle + \langle \Phi(\tau), \Phi(\tau) \rangle - 2 \langle \Phi(z), \Phi(\tau) \rangle$  and replacing inner products by the RBF kernel [Burges, 1999].

Guo et al. [2009] have proven that  $\Phi(z)$ ,  $\Phi(\tau)$  and  $a_{\mathcal{H}}$  are coplanar, and hence the three angles  $\{0, \Phi(z), a_{\mathcal{H}}\}$ ,  $\{0, \Phi(\tau), a_{\mathcal{H}}\}$  and  $\{0, \Phi(z), \Phi(\tau)\}$  corresponding to the three triangles are related by:

$$\alpha = \gamma - \beta \quad (2.48)$$

The triangle sides  $\|\Phi(z) - a_{\mathcal{H}}\|_{\mathcal{H}}$  and  $R_{\mathcal{H}}$  are defined by Equations 2.41 and 2.42, and the sides  $\|\Phi(z)\|$ ,  $\|\Phi(\tau)\|$  are equal to one. Hence, only  $\|a_{\mathcal{H}}\|$  is missing to determine  $\beta$  and  $\gamma$ , and thus  $\alpha$  (cf. Figure 2.6). The sphere centre  $a_{\mathcal{H}}$  is defined as follows (cf. Equation 2.31):

$$a_{\mathcal{H}} = \sum_{1 \leq i \leq n} \alpha_i \Phi(x_i) \quad (2.49)$$

Hence, the squared distance of  $a_{\mathcal{H}}$  to the origin in  $\mathcal{H}$  is defined as:

$$\|a_{\mathcal{H}}\|^2 = \sum_{1 \leq i, j \leq n} [\alpha_i \alpha_j \langle \Phi(x_i), \Phi(x_j) \rangle] \quad (2.50)$$

The inner products in Equation 2.49 are replaced by the RBF kernel, which leads to:

$$\|a_{\mathcal{H}}\|^2 = \sum_{1 \leq i, j \leq n} \left[ \alpha_i \alpha_j \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \right] \quad (2.51)$$

Hereafter, all required triangle sides are defined, and the law of cosines allows to determine  $\beta$  and  $\gamma$  by:

$$\cos \beta = \frac{1 + \|a_{\mathcal{H}}\|^2 - R_{\mathcal{H}}^2}{2\|a_{\mathcal{H}}\|} \quad (2.52)$$

$$\cos \gamma = \frac{1 + \|a_{\mathcal{H}}\|^2 - \|\Phi(z) - a_{\mathcal{H}}\|_{\mathcal{H}}^2}{2\|a_{\mathcal{H}}\|} \quad (2.53)$$

Plugging Equations 2.52–2.53 into Equation 2.48 leads to the angle  $\alpha$ , which can be used to define the required distance in  $\mathcal{H}$ :

$$\|\Phi(z) - \Phi(\tau)\|_{\mathcal{H}}^2 = 2 - 2 \cos \alpha \quad (2.54)$$

Plugging Equation 2.54 into Equation 2.47 with some rearrangement gives the input space distance to the decision boundary:

$$\|z - \tau\|_{\mathbb{R}^d} = \sigma \sqrt{-2 \cdot \ln(\cos(\gamma - \beta))} \quad (2.55)$$

Equation 2.55 yields the Euclidean input space distance of any test sample  $z$  to the decision surface. Figure 2.6 refers to a test sample  $z$  outside the SVDD hypersphere. It is easy to see in Equations 2.48 and 2.54 that the preceding definitions for the distance also hold true for test samples inside the SVDD hypersphere. In order to prevent ambiguities, a signed distance  $f_{\text{SVDD}}(z)$  is defined based on the original decision rule introduced in [Tax & Duin, 2004] (cf. Equation 2.43):

$$f_{\text{SVDD}}(z) = \begin{cases} -\|z - \tau\|_{\mathbb{R}^d} & | R_{\mathcal{H}}^2 - \|\Phi(z) - a_{\mathcal{H}}\|_{\mathcal{H}}^2 \geq 0 \\ +\|z - \tau\|_{\mathbb{R}^d} & | \text{otherwise} \end{cases} \quad (2.56)$$

In the next chapter the probability for a particular test sample  $z$  being classified on the basis of a valid model, defined through a two-class SVM, is determined on the basis of distance  $f_{\text{SVDD}}(z)$ .





## 3. New methodology

The new methodology that will be presented in this chapter is mainly inspired by the work described in [Gerke, 2006; Gerke & Heipke, 2008]. Preliminary results of the new method have been published in [Ziems et al., 2010, 2011a, 2011b and 2012].

This chapter is structured as follows: In the first section the basic strategy will be introduced, followed by a detailed description of the approach in Sections 3.2 and 3.3. The chapter concludes with discussions of some practical (Section 3.4) and theoretical (Section 3.5) aspects of the new approach.

### 3.1. Basic strategy

The goal of this thesis is the automated verification of road databases on the basis of up-to-date imagery. In this context, *verification* refers to the question of whether a road object can be found in the image or not and whether its required positional accuracy is achieved or not. In accordance with Gerke & Heipke [2008], a human operator is integrated into the workflow who investigates potentially wrong road objects and corrects them if necessary. In accordance with [Gerke & Heipke, 2008; Poulain et al., 2010], the verification problem is interpreted as a classification problem by considering the two possible states, *correct* and *incorrect*, for every road object stored in the database. The approach to solve the classification problem is new. It is based on the following core ideas:

1. Combine several classifiers, all relying on different road models described in the literature.
2. Identify assumptions made by the underlying model of each classifier and check the correspondence of the actual situation in the image with theoretical assumptions made by the model. Use the correspondence to determine the so-called *model-uncertainty* which describes the state of the underlying model being either *applicable* or *not applicable*.
3. Map the two state spaces  $\{correct, incorrect\}$  and  $\{applicable, not\ applicable\}$  to a new state space that considers the three alternative states for every road object stored in the database  $\{correct, incorrect, unknown\}$ . Combine the mapped outputs from all classifiers by applying Dempster's rule and consider the *conflict mass* to identify problems of the reasoning approach, while flagging the output as *invalid*.
4. Classify each road hypothesis into *correct*, *incorrect*, *unknown* and *invalid* and forward this result to a human operator who has the freedom to deal with the states *unknown* and *invalid*.

Compared to other state-of-the-art methods the main novelty of this approach lies in its strictly modular concept. Each classifier works as a completely independent module, and hence represents

a verification approach on its own. The so-called *decision level fusion* on the basis of the DST is also novel even though other state-of-the-art road verification approaches used the DST to combine lower level information, e.g. [Gerke, 2005; Poulain et al., 2010]. Another unique characteristic of this approach is that the decision space explicitly includes the state *unknown* which exploits the specific ability of the DST to express *ignorance* in a straight-forward manner. The road models and detection strategies of the verification modules correspond to established state-of-the-art road detection approaches. In this respect, the contribution lies in transferring those detection strategies into strategies that are useful for the verification problem. Furthermore, some of the state-of-the-art strategies are expanded to make them more effective for the overall approach, e.g. by introducing prior knowledge of the road database or by focusing on weak points of other modules. Finally, the analysis and the mathematical definition of the *model-uncertainties* for several state-of-the-art road detection approaches are a particular contribution of this thesis.

Figure 3.1 gives an overview of the semi-automatic workflow starting with the *input road database* and the *imagery*, and ending with the *corrected road database*.

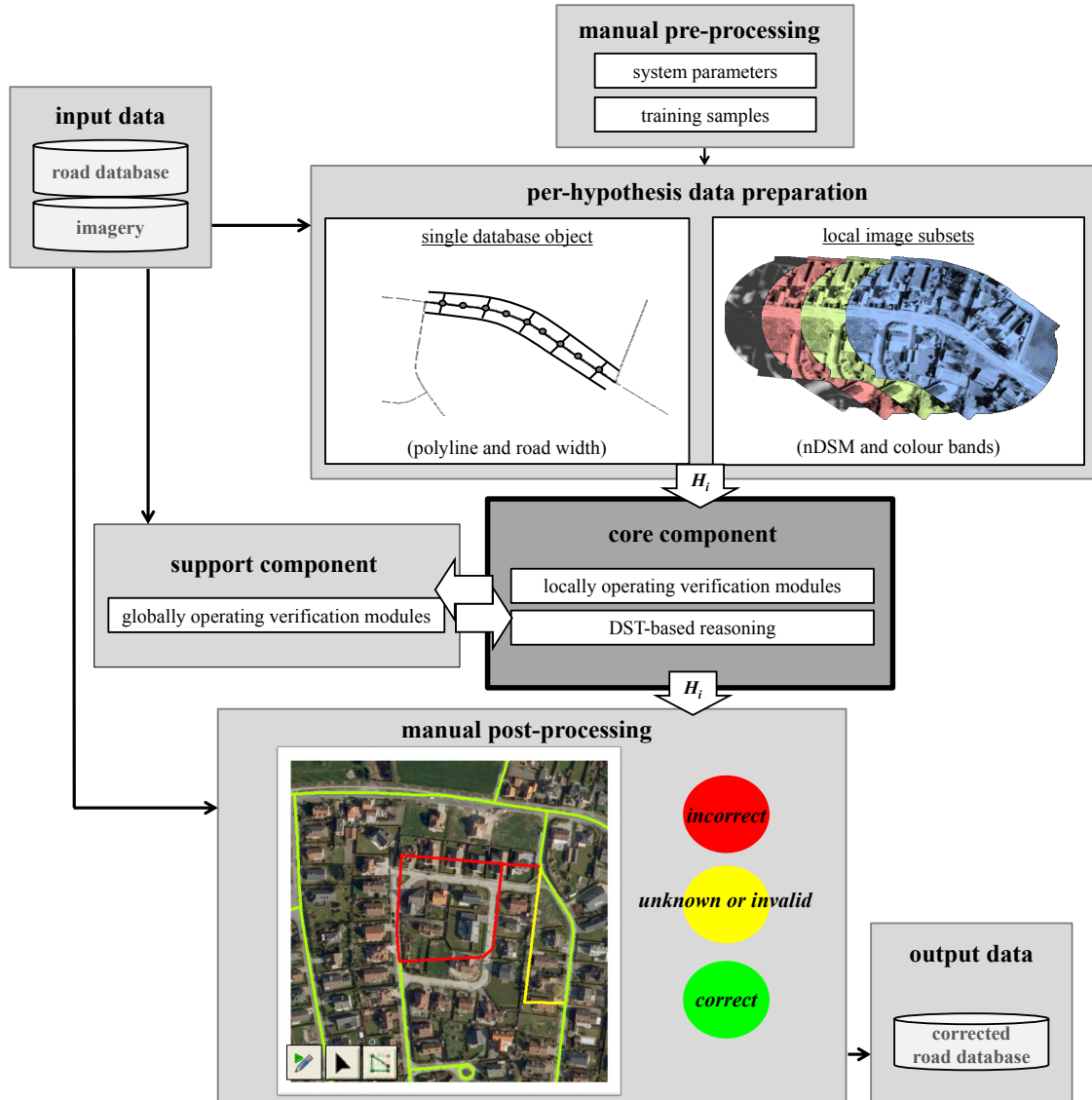


Figure 3.1.: Overview of the proposed approach.

Two of the boxes depicted in Figure 3.1 correspond to interactive tasks that require a human operator. Firstly, the *manual pre-processing* where the approach has to be prepared to the actual problem. Secondly, the *manual post-processing*, where the indicated database errors have to be corrected. The other boxes represent fully automatic processing steps. The *per-hypothesis data preparation* automatically defines image subsets showing the local surroundings of each single road object stored in the database. Then the data are forwarded to the *core component* that includes all verification modules operating locally and the reasoning approach. The *core component* is applied to each road object separately, and thus finally provides a decision for each road object. The *support component* contains the modules operating globally, i.e. those that are not restricted to a per-hypothesis analysis. Therefore, the *support component* has access to the whole database and is allowed to exchange intermediate results with the *core component*.

In the following, all components depicted in Figure 3.1 will be introduced briefly.

**Input road database:** For the thesis it is assumed that the input road database has an object-based structure in which a road object is defined by its centreline and a unique value for the road width. Further it is required that a single road object does not include junctions. Road objects that are connected must have the same coordinates at their end points. The input road databases may have different requirements for spatial accuracy which is considered by a parameter  $D^{xy}$  denoting the required accuracy for the road centreline and  $D^{width}$  denoting the required accuracy for the road width attribute. As the actual spatial accuracies of road databases are often different from what is required, two additional parameters are defined: The root mean square (RMS) error of the road centreline  $D^{rms(xy)}$  and the RMS error of the road width attribute  $D^{rms(width)}$ . These four so-called *system parameters* are listed in Table 3.1 and will subsequently be used in different equations. Below more system parameters will be introduced and reviewed all together in Section 3.4.

$D^{xy}$	[m]	required accuracy of the road centreline
$D^{width}$	[m]	required accuracy of the width attribute
$D^{rms(xy)}$	[m]	RMS error of the road centreline
$D^{rms(width)}$	[m]	RMS error of the road width

Table 3.1.: System parameters: Spatial accuracy of the database.

**Input imagery:** For this thesis it is assumed that the geo-reference of the input imagery corresponds to the *road database* and that the image is orthorectified. Beyond that, different kinds of imagery with a GSD of about 0.2–2.0 m can be used. In particular, panchromatic, colour<sup>1</sup>, NDVI or nDSM (aerial or satellite) images are considered as possible inputs. As will be shown later, the characteristics of the input imagery have consequences to the applicability of the verification modules. The actual GSD of the imagery is introduced as a system parameter (cf. Table 3.2).

$GSD$	[m]	ground sampling distance of the imagery
-------	-----	---

Table 3.2.: System parameter: Image resolution.

**Manual pre-processing:** This step is important to adapt the overall approach to different input datasets. Here, a human operator has to define the *system parameters* and the *training dataset*.

<sup>1</sup>For simplification, the subsequent discussions refer to images having three bands such as RGB or IRRG.

According to Tables 3.1 and 3.2, the system parameters describe general properties of the input data, while the training dataset is required to adapt the approach of the radiometric properties of a scene. The related training algorithms will be discussed together with the respective verification modules in Section 3.3. In summary, training samples for three different object classes are required: *roads*, *buildings* and *grassland objects* (cf. Table 3.3).

$X^{\text{roads}}$	training data set consisting of image regions that correspond to roads
$X^{\text{buildings}}$	training data set consisting of image regions that correspond to buildings
$X^{\text{grassland}}$	training data set consisting of image regions that correspond to grassland

Table 3.3.: Training datasets.

**Per-hypothesis data preparation:** Usually, the number of road hypotheses to be verified will be large ( $> 1,000$ ), but the core component considers each database hypothesis individually. The automatic pre-processing step generates a hypothesis  $H_i$  for each road object stored in the database. For practical reasons road objects that are longer than 100 m are subdivided into different hypotheses of similar lengths. The index  $i$  denotes the particular road hypothesis. A set  $H_i$  includes nine elements:

$$H_i = \{h_i^{\text{xy}}, h_i^{\text{width}}, h_i^{\text{length}}, R_i^{\text{H}}, R_i^{\text{H}\bullet}, R_i^{\bullet}, R_i^{\text{H}||}, R_i^{[\text{H}\cdot]}, R_i^{\text{H}[\cdot]}\} \quad (3.1)$$

Table 3.4 describes the meanings of all elements while Figure 3.2 provides a symbolical<sup>2</sup> interpretation.

$h_i^{\text{xy}}$	2D coordinate list of the centreline
$h_i^{\text{width}}$	width hypothesis
$h_i^{\text{length}}$	length hypothesis
$R_i^{\text{H}}$	set of image pixels that are representing the road according to the database
$R_i^{\text{H}\bullet}$	set of image pixels that contains all possible positions of the road hypothesis
$R_i^{\bullet}$	set of pixels that only show the road hypothesis
$R_i^{\text{H}  }$	set of pixels that show all possible positions of the road borders
$R_i^{[\text{H}\cdot]}$	set of pixels that show the road hypothesis and its context area
$R_i^{\text{H}[\cdot]}$	set of pixels that show the context area, but not the road hypothesis
$\Rightarrow H_i = \{h_i^{\text{xy}}, h_i^{\text{width}}, h_i^{\text{length}}, R_i^{\text{H}}, R_i^{\text{H}\bullet}, R_i^{\bullet}, R_i^{\text{H}  }, R_i^{[\text{H}\cdot]}, R_i^{\text{H}[\cdot]}\}$	

Table 3.4.: Input for verification modules.

The element  $h_i^{\text{xy}}$  denotes the *road centreline*, i.e. a 2D coordinate list indicating the points of a polyline. The elements  $h_i^{\text{width}}$  and  $h_i^{\text{length}}$  denote the *road width* and the *road length*.  $R_i^{\text{H}}$  denotes the set of pixels corresponding to the database information (*road centreline*, *road width*), i.e. a direct projection of the database entry into the image space.  $R_i^{\text{H}\bullet}$  denotes the set of pixels corresponding to the database entry considering the tolerated spatial inaccuracies. To obtain  $R_i^{\text{H}\bullet}$ , morphological

<sup>2</sup>Figure 3.2 shows a simplified road hypothesis that only relies on a straight polyline and whose shape is simplified at both ends. In general, polylines have more complex shapes that lead to image subsets of more complex shapes.

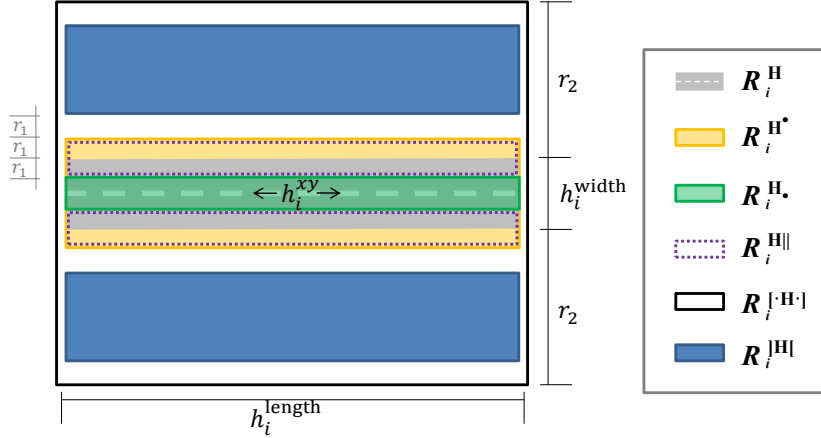


Figure 3.2.: Definition of image subsets, related to a single road hypothesis.

dilation is applied to  $R_i^H$  using a circular structuring element with radius  $r_1$ , which is defined as:

$$r_1 = 3 \left( D^{\text{rms}(\text{xy})} + \frac{1}{2} D^{\text{rms}(\text{width})} \right) + \text{GSD} \quad (3.2)$$

In Equation 3.2, the RMS errors of *road centreline* and *road width* are considered together with a one-pixel distance (GSD) that compensates the effects of mixed pixels. The decimal factor 3 in Equation 3.2 is motivated by the three-sigma rule that states that nearly all values lie within three standard deviations of the mean in a normal distribution.  $R_i^{H\bullet}$  is determined by morphological erosion of  $R_i^H$  using the same structuring element as for  $R_i^{H\bullet}$ , and thus includes only pixels that must correspond to the road if the database entry was correct. Consequently, the set  $R_i^{H||} = R_i^{H\bullet} \setminus R_i^{H\bullet}$  defines the tolerated spatial positions of the road borders. The set  $R_i^{[H\cdot]}$  is defined by morphological dilation of the set  $R_i^H$  using a circular structuring element with radius  $r_2$  and removing pixels corresponding to  $R_j^{H\bullet} \in H_j \forall j \neq i$ . Here,  $r_2$  corresponds to the width of the local context region that has to be considered. As the optimal value for  $r_2$  depends on the particular scene, it is introduced as another system parameter that has to be defined manually (cf. Table 3.5).

$r_2$	[m]	width of considered local context region, perpendicular to the road centreline hypothesis
-------	-----	---

Table 3.5.: System parameter: Local context definition.

Removing pixels corresponding to  $R_j^{H\bullet} \in H_j$  excludes pixels corresponding to other road hypotheses that have to be ignored for the per-hypothesis analysis.  $R_i^{[H]}$  is defined by morphological erosion of the set  $R_i^{[H\cdot]} \setminus R_i^{H\bullet}$  with a circular structure element of radius  $r_1$  (cf. Equation 3.2) and thus explicitly represents the local context region.

**Core and support components:** In accordance with Figure 3.3 the core component contains nine independently and locally operating verification modules that rely on different road models. Each verification module  $vm$  receives the pre-processed dataset  $H_i$  and provides two different outputs: Firstly, the *verification output* given by a probability function  $P_i^{vm} : \Theta_v \rightarrow [0, 1]$ . Secondly, the *model-uncertainty output* given by another probability function  $P_i^{vm} : \Theta_u \rightarrow [0, 1]$ . Both functions refer to different state spaces denoted by indices  $v$  and  $u$ . They are mapped to a new state space  $2^\Theta$  for which probability mass functions  $m_i^{vm} : 2^\Theta \rightarrow [0, 1]$  are defined for all verification modules  $vm$ . These outputs are combined using Dempster's rule and then mapped to a decision space that

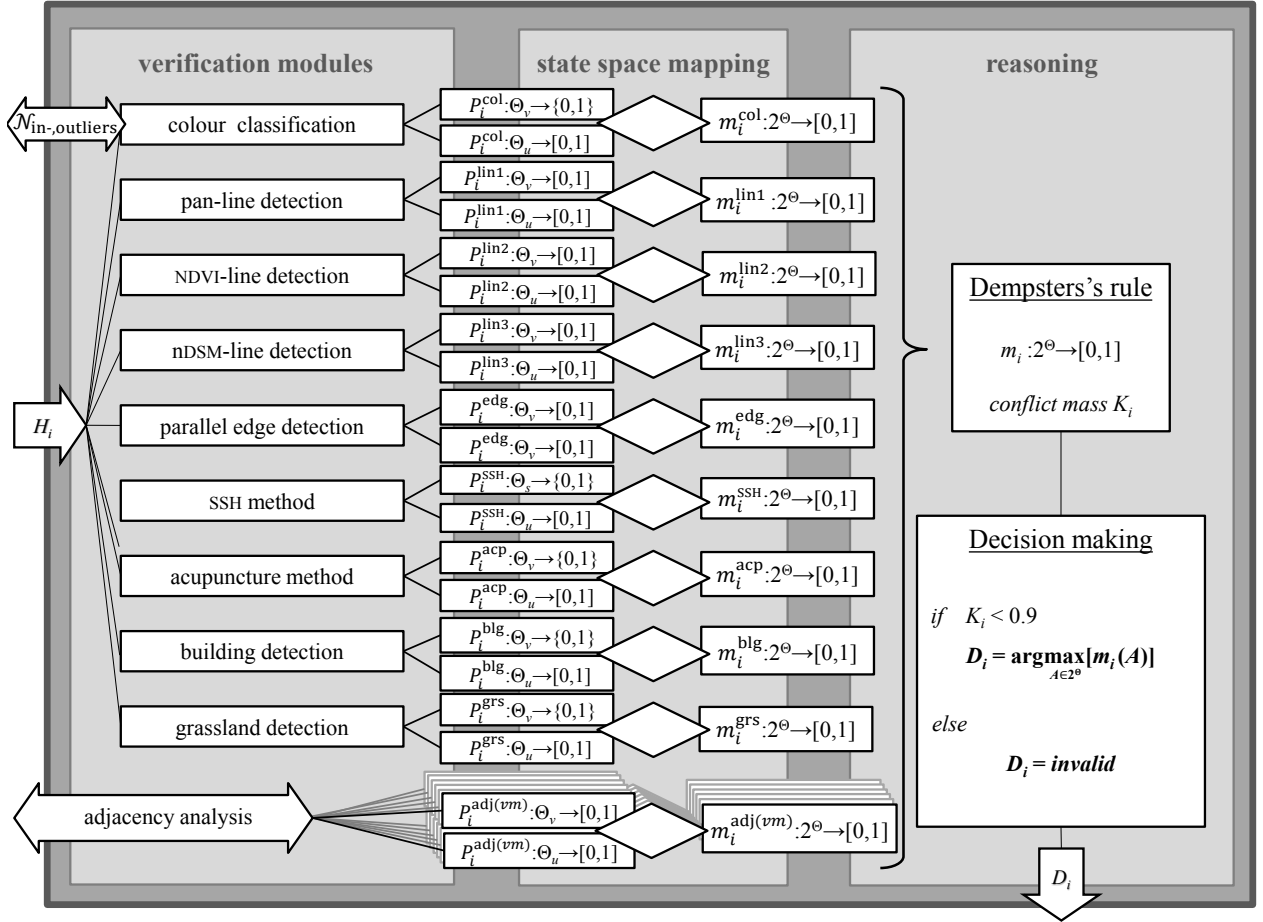


Figure 3.3.: Core component of the proposed approach.

includes four states  $\{correct, incorrect, unknown, invalid\}$ . The decision  $D_i$  corresponding to a hypothesis  $H_i$  is then forwarded to the manual post processing. Additionally, the core component exchanges intermediate results with the support component which is not restricted to local image pixel sets. From the support component the colour classification module receives the parameters for the distributions  $\mathcal{N}_{inliers}$ ,  $\mathcal{N}_{outliers}$  and the adjacency analysis provides additional outputs for *verification* and *model-uncertainty*. A detailed description of the core and the support components will be given in Sections 3.2 and 3.3, respectively.

**Manual post processing:** The input imagery and the database to be verified are visualized in a standard GIS environment together with the colour-coded decision outputs. The basic idea is that the human operator can concentrate on a few objects only, instead of having to check the whole scene. How the search over automatically verified road objects can be effectively implemented, is discussed in [Beyen et al., 2008], and will not be dealt with in this thesis. Whether or not a road object has to be investigated by the human operator depends on its decision state  $\{correct, incorrect, unknown, invalid\}$  and on the cartographic task to be solved. Two scenarios will be particularly considered in the experiments of this thesis. The first scenario focuses on a high database quality where the human operator has to check every non-verified object, i.e. road hypotheses assigned to the states *incorrect*, *unknown* and *invalid*. This scenario largely corresponds to the scenario defined in [Gerke & Heipke, 2008]. The second scenario focuses on the reduction of

manual labour. Accordingly, a human operator only checks the detected errors, i.e. only road hypotheses assigned to the state *incorrect*. Both scenarios do not distinguish the states *unknown* and *invalid*, and thus a traffic light scheme with red (*incorrect*), yellow (*unknown* or *invalid*) and green (*correct*) is sufficient to visualize the automatically achieved outputs (cf. Figure 3.1). However, the states *unknown* and *invalid* determine different problems that will be stressed in Chapter 4.

## 3.2. The fusion framework

In this section, first the basic idea of the reasoning approach will be presented (Section 3.2.1) which is followed by an introduction to the principle of determining model-uncertainties (Section 3.2.2).

### 3.2.1. Reasoning strategy

The verification modules provide two kinds of probabilistic outputs:

1. The *verification output*, a probability function  $P : \Theta_v \rightarrow [0, 1]$  with:

$$\Theta_v = \{correct, incorrect\} \quad (3.3)$$

2. The *model-uncertainty*, a probability function  $P : \Theta_u \rightarrow [0, 1]$  with:

$$\Theta_u = \{applicable, not applicable\} \quad (3.4)$$

Subsequently, the states *applicable* and *not applicable* will be abbreviated by  $a$  and  $n/a$ .

**The verification output:** This output in form of a probability distribution provides a solution to the main classification problem, according to which a road hypothesis might be either *correct* or *incorrect*. The output is specific for each road hypothesis, and thus depends on some data term  $x_v^{vm}(H_i)$  derived from the local dataset included in  $H_i$ . The function  $x_v^{vm}$  represents the subjective detection strategy related to a verification module  $vm$ . As the road model and the detection strategy of each verification module are rather specific, its individual output is not useful in a general sense. Hence, the verification output is conditioned to the *applicability* of the underlying road model which leads to the following interpretation of what each module effectively provides:

$$P(road = correct | x_v^{vm}(H_i), model = a.) = 1 - P(road = incorrect | x_v^{vm}(H_i), model = a.) \quad (3.5)$$

**The model-uncertainty:** This probability distribution, refers to the two possible states of the underlying road model that might be either *applicable* or *not applicable*. As the model-uncertainty is specific for each road hypothesis, it is also referred to a specific data term  $x_u^{vm}(H_i)$  derived from the local dataset included in  $H_i$ . The function  $x_u^{vm}$  represents the analysis strategy of a particular verification module  $vm$  for the data term and thus the model-uncertainty is expressed as follows:

$$P(model = a. | x_u^{vm}(H_i)) = 1 - P(model = n/a | x_u^{vm}(H_i)) \quad (3.6)$$

**The unknown distribution:** In order to combine the outputs of the different verification modules, the conditioning of the *model* to the state *applicable* needs to be removed:

$$P(\text{road} = \text{correct} | x_v^{vm}(H_i)) = P(\text{road} = \text{correct} | x_v^{vm}(H_i), \text{model} = a.)P(\text{model} = a. | x_u^{vm}(H_i)) \\ + \underbrace{P(\text{road} = \text{correct} | x_v^{vm}(H_i), \text{model} = n/a.)P(\text{model} = n/a | x_u^{vm}(H_i))}_? \quad (3.7)$$

In Equation 3.7 the term labelled with the question mark is *unknown*. In particular, it is expected that there is neither an experimental nor a heuristical basis to define that term. If there was such a basis, the verification outputs would have a general character and the conditioning on the state of the model would be unnecessary because it could be replaced by another more comprehensive model, which is *applicable*. A possible solution to Equation 3.7 is replacing the *unknown* distribution by a uniform distribution such that:

$$P(\text{road} = \text{correct} | x_v^{vm}(H_i), \text{model} = n/a) = P(\text{road} = \text{incorrect} | x_v^{vm}(H_i), \text{model} = n/a) = 0.5 \quad (3.8)$$

However, this definition is inapplicable for the given problem as will be demonstrated experimentally in Chapter 4. To overcome the problem, the two probabilistic representations are transferred into a representation based on the DST, where  $P(\text{road} = \text{correct} | x_v^{vm}(H_i), \text{model} = n/a)$  does not need to be specified.

**Definition of probability masses:** In the DST the state space is defined by the power set of  $\Theta_v$ :

$$2^\Theta = \{\text{correct}, \text{incorrect}, \text{correct} \cup \text{incorrect}, \emptyset\} \quad (3.9)$$

The state  $\{\text{correct} \cup \text{incorrect}\}$  denotes the *ignorance* in discriminating the two states  $\{\text{correct}\}$  and  $\{\text{incorrect}\}$ . The probability mass functions  $m_i : 2^\Theta \rightarrow [0, 1]$  are defined by mapping both kinds of outputs, the *verification output* and the *model-uncertainty output*, to probability mass distributions in  $2^\Theta$ . In accordance with the discussions in Section 2.1.4 the mapping is defined as follows:

$$m_i^{vm}(\text{road} = \text{correct}) = P(\text{road} = \text{correct} | x_v^{vm}(H_i), \text{model} = a.)P(\text{model} = a. | x_u^{vm}(H_i)) \quad (3.10)$$

$$m_i^{vm}(\text{road} = \text{incorrect}) = P(\text{road} = \text{incorrect} | x_v^{vm}(H_i), \text{model} = a.)P(\text{model} = a. | x_u^{vm}(H_i)) \quad (3.11)$$

$$m_i^{vm}(\text{road} = \text{correct} \cup \text{incorrect}) = P(\text{model} = n/a | x_u^{vm}(H_i)) \quad (3.12)$$

$$m_i^{vm}(\text{road} = \emptyset) = 0 \quad (3.13)$$

**Fusion and decision making:** Dempster's rule (cf. Equation 2.10) is applied to combine the probability mass functions  $m_i^{vm} : 2^\Theta \rightarrow [0, 1]$  of all verification modules *vm* to a single probability mass function  $m_i : 2^\Theta \rightarrow [0, 1]$  for each road hypothesis  $H_i$ . Additionally, the conflict mass  $K_i$  is determined in accordance with Equation 2.11. Decision making is implemented analogous to Algorithm 1, where the three states  $\{\text{correct}, \text{incorrect}, \text{correct} \cup \text{incorrect}\}$  are mapped onto the decision space by applying Equation 2.25. In Section 2.1.3 the explicit consideration of large conflict masses for the mapping was motivated by its potential to uncover internal errors of the approach. A value of 0.9 for the related threshold  $K_i$  turned out to be appropriate for the experiments applied for this thesis.



**Algorithm 1** Decision rule.

---

```

if  $K_i < 0.9$  then
  if  $\operatorname{argmax}_{2^\Theta}(m_i) = \{correct\}$  then
     $D_i = correct$ 
  else if  $\operatorname{argmax}_{2^\Theta}(m_i) = \{incorrect\}$  then
     $D_i = incorrect$ 
  else if  $\operatorname{argmax}_{2^\Theta}(m_i) = \{correct \cup incorrect\}$  then
     $D_i = unknown$ 
else
   $D_i = invalid$ 

```

---

In this thesis the term *unknown* is used for the state  $\{correct \cup incorrect\}$  because it highlights the presence of different levels of knowledge. In this context the first level represents *uncertain knowledge*, given in form of a probability function that describes the state of the road object (*correct*, *incorrect*). The second level represents *ignorance* which means that the probability function that describes the state of the road object (*correct*, *incorrect*) is *unknown*; such a decision reads: *Given all evidences, the state of the road object is most likely unknown.*<sup>3</sup>

**3.2.2. Model-uncertainties**

As discussed in Section 1.3 state-of-the-art methods, and thus also the applied verification modules, rely on rigid assumptions, e.g. they induce specific context areas or appearances of roads. The basic idea is that only if these assumptions are satisfied the model is *applicable* and provides an interpretable verification output. If at least one of the assumptions is *violated* the model is *not applicable* and the verification output is not interpretable for the given task. Accordingly, the state of a model can be defined by analysing the states of the critical assumptions  $a_k$  that might be either *satisfied* or *violated*:

$$P(a_k = satisfied | x_{u,k}^{vm}(H_i), \theta) = 1 - P(a_k = violated | x_{u,k}^{vm}(H_i), \theta) \quad (3.14)$$

The state of a critical assumption  $a_k$  depends on a feature vector  $x_{u,k}^{vm}(H_i)$  and on a parameter set  $\theta$ . Considering  $K$  critical assumptions  $a_k$  with  $1 \leq k \leq K$ , the corresponding distribution is defined:

$$P(model = a. | x_u^{vm}(H_i)) = P(a_1 = satisfied \cap a_2 = satisfied \cap \dots \cap a_K = satisfied) \quad (3.15)$$

$$\begin{aligned} & \Downarrow \\ P(model = a. | x_u^{vm}(H_i)) &= \prod_{1 \leq k \leq K} P(a_k = satisfied | x_{u,k}^{vm}(H_i), \theta) \end{aligned} \quad (3.16)$$

Equation 3.16 requires independence of the different  $a_k$  which is assumed because the feature vectors  $x_{u,k}^{vm}(H_i)$  vary with  $k$ . The probabilities  $P(a_k = satisfied | x_u(H_i), \theta)$  are always modelled by sigmoid functions parametrised with  $\theta$ . As shown in Figure 3.4 a sigmoid function allows a smooth transition from the state *violated* (left) to the state *satisfied* (right).

The *critical assumptions* are identified with the descriptions in the original literature or other

---

<sup>3</sup>Bammer & Smithson [2008] provide a comprehensive discussion about levels knowledge and their probabilistic interpretations.

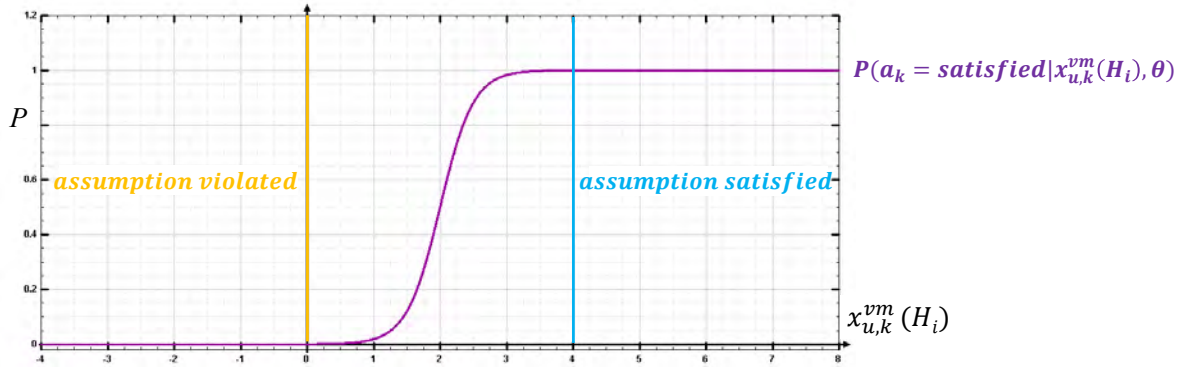


Figure 3.4.: Sigmoid function.

empirical analyses such as those carried out in benchmark tests. In accordance with those knowledge sources the parameters of the sigmoid functions  $\theta$  are defined on a heuristical or a statistical basis. In Section 3.3, four different approaches to define the sigmoid parameters will be used, the *generative parametrisation approach*, the *discriminative parametrisation approach*, the *smooth heuristic parametrisation approach* and the *rigid heuristic parametrisation approach*. Each approach is particularly useful for different kinds of assumptions that will be dealt with below.

**The generative parametrisation approach:** If probability density functions for feature values  $x_{u,k}^{vm}(H_i)$  can be defined by Gaussians such that:

$$p(x_{u,k}^{vm}(H_i) | a_k = \text{satisfied}) = \mathcal{N}(x_{u,k}^{vm}(H_i) | \mu_{sat}, \sigma_{sat}) \quad (3.17)$$

$$p(x_{u,k}^{vm}(H_i) | a_k = \text{violated}) = \mathcal{N}(x_{u,k}^{vm}(H_i) | \mu_{vio}, \sigma_{vio}) \quad (3.18)$$

are known a sigmoid function is defined on the basis of the Bayes' theorem which leads to the following form (cf. Bishop, 2006):

$$P(a_k = \text{satisfied} | x_{u,k}^{vm}(H_i), \theta_{\text{gen}}) = \frac{\mathcal{N}(x_{u,k}^{vm}(H_i) | \mu_{sat}, \sigma_{sat}) P(\text{sat})}{\mathcal{N}(x_{u,k}^{vm}(H_i) | \mu_{sat}, \sigma_{sat}) P(\text{sat}) + \mathcal{N}(x_{u,k}^{vm}(H_i) | \mu_{vio}, \sigma_{vio}) P(\text{vio})} \quad (3.19)$$

In Equation 3.19  $P(\text{sat})$  and  $P(\text{vio})$  denote the prior's. The *generative parametrisation approach* requires to define the parameter set  $\theta_{\text{gen}} = \{\mu_{sat}, \sigma_{sat}, P(\text{sat}), \mu_{vio}, \sigma_{vio}, P(\text{vio})\}$ . Subsequently, this parametrisation approach will always be used if there is a heuristical or a statistical basis to define the parameters of the Gaussian likelihoods.

**The discriminative parametrisation approach:** In this case, the sigmoid function is defined using the following form (cf. Bishop, 2006):

$$P(a_k = \text{satisfied} | x_{u,k}^{vm}(H_i), \theta_{\text{disc}}) = \frac{1}{1 + \exp(-t)} \quad (3.20)$$

where  $t$  represents a quadratic function depending on  $x_{u,k}^{vm}(H_i)$ :

$$t = t_1 [x_{u,k}^{vm}(H_i)]^2 + t_2 [x_{u,k}^{vm}(H_i)] + t_3 \quad (3.21)$$

Equations 3.19 and 3.20 are actually equivalent. According to Bishop [2006] their parameters are related by:

$$t_1 = \frac{1}{2\sigma_{sat}^2} - \frac{1}{2\sigma_{vio}^2} \quad (3.22)$$

$$t_2 = \frac{\mu_{vio}}{2\sigma_{vio}^2} - \frac{\mu_{sat}}{2\sigma_{sat}^2} \quad (3.23)$$

$$t_3 = \frac{\mu_{sat}^2}{2\sigma_{sat}^2} - \frac{\mu_{vio}^2}{2\sigma_{vio}^2} - \ln P(sat) + \ln P(vio) - \ln \frac{1}{\sigma_{sat}\sqrt{2\pi}} + \ln \frac{1}{\sigma_{vio}\sqrt{2\pi}} \quad (3.24)$$

One advantage of the *discriminative approach* over the *generative approach* is that the number of parameters is reduced from six to three:  $\theta_{disc} = \{t_1, t_2, t_3\}$ . This allows to define simplified regression models for the parameter training, and thus the *discriminative parametrisation approach* is subsequently used in cases where all sigmoid parameters must be learned from a single training dataset.

**The smooth heuristic parametrisation approach:** A disadvantage of the *discriminative parametrisation approach* is that the parameters  $\{t_1, t_2, t_3\}$  do not have an intuitive meaning which makes implementing heuristics difficult. In contrast, the *generative approach* may allow a heuristic definition, but one has to be aware of the assumption that the likelihoods are Gaussians which will sometimes be problematic. As discussed in Section 1.3 some promising road detection approaches rely on heuristics, often realized through Fuzzy-membership functions, e.g. [Wiedemann & Mayer, 1996; Bacher & Mayer, 2005; Grote et al., 2012]. In this respect, the *smooth heuristic parametrisation approach* is defined to implement simple heuristics comparable to fuzzy membership functions, such that for  $x_{u,k}^{vm}(H_i) = d_1$  the critical assumption is *satisfied* and for  $x_{u,k}^{vm}(H_i) = d_2$  the critical assumption is *violated*. This heuristic is used to define the sigmoid parameters  $\{t_1, t_2, t_3\}$  in Equation 3.20:

$$t_1 = 0 \quad (3.25)$$

$$t_2 = \frac{16}{d_1 - d_2} \quad (3.26)$$

$$t_3 = \frac{8(d_2^2 - d_1^2)}{(d_1 - d_2)^2} \quad (3.27)$$

Considering the relation of the sigmoid parameters  $\{t_1, t_2, t_3\}$  to the *generative parametrisation approach* (cf. Equations 3.22–3.24), the parameters  $\{d_1, d_2\}$  have the following relations to the Gaussian parameters:

$$\mu_{sat} = d_1 \quad (3.28)$$

$$\mu_{vio} = d_2 \quad (3.29)$$

$$\sigma_{sat} = \sigma_{vio} = \frac{1}{4} \cdot |\mu_{sat} - \mu_{vio}| \quad (3.30)$$

$$P(sat) = P(vio) \quad (3.31)$$

This *smooth heuristic approach*, associated with the parameter set  $\theta_{smooth} = \{d_1, d_2\}$ , at least yields a smooth transition of the sigmoid between  $d_1$  and  $d_2$ . Subsequently, this parametrisation approach will be used if an appropriate heuristic for  $d_1$  and  $d_2$  is available.

**The rigid heuristic parametrisation approach:** The *rigid heuristic approach* is a further simplification of the *smooth heuristic approach*. It is expected that a single value  $T$  can be defined that clearly discriminates the two states of an assumption. Hence,  $P(a_k = \text{satisfied} | x_{u,k}^{vm}(H_i), \theta_{\text{rigid}})$  might be either zero or one. This corresponds to a sigmoid function based on Gaussians with infinitely small standard deviations, i.e.  $\lim_{\sigma_{sat}, \sigma_{vio} \rightarrow 0} P(a_k = \text{satisfied} | x_{u,k}^{vm}(H_i), \theta_{\text{gen}})$  (cf. Equation 3.19). In this case the sigmoid function can be approximated by a delta function  $\delta(x_{u,k}^{vm}(H_i) \geq T)$ , with only one parameter  $T$ . The setting for the *rigid heuristic approach* will further be denoted as  $\theta_{\text{rigid}} = \{T\}$ . This model will be used if such a simple heuristic seems to be sufficient.

In the subsequent sections the indices {gen, disc, smooth, rigid} will be used to emphasize the different parametrisation approaches. Figure 3.5 exemplary shows realizations of the different parametrisation approaches for basically the same sigmoid function. For the *generative approach* it is:

$$P(a_1 = \text{satisfied} | x_{u,1}^{vm}(H_i), \theta_{\text{gen}} = \{4.0, 1.0, 0.8, 4.0, 1.0, 0.8\}) \quad (3.32)$$

where the Gaussian likelihoods have the following definition:

$$\mathcal{N}(x_{u,1}^{vm}(H_i) | \mu_{sat} = 4.0, \sigma_{sat} = 1.0) \quad (3.33)$$

$$\mathcal{N}(x_{u,1}^{vm}(H_i) | \mu_{vio} = 0.0, \sigma_{vio} = 1.0) \quad (3.34)$$

For the *discriminative approach* and the *smooth heuristic approach* it is:

$$P(a_1 = \text{satisfied} | x_{u,1}^{vm}(H_i), \theta_{\text{disc}} = \{0.0, 4.0, -8.0\}) \quad (3.35)$$

$$P(a_1 = \text{satisfied} | x_{u,1}^{vm}(H_i), \theta_{\text{smooth}} = \{4.0, 0.0\}) \quad (3.36)$$

and for the *rigid heuristic approach* merely an approximation can be defined by:

$$P(a_1 = \text{satisfied} | x_{u,1}^{vm}(H_i), \theta_{\text{rigid}} = \{2\}) \quad (3.37)$$

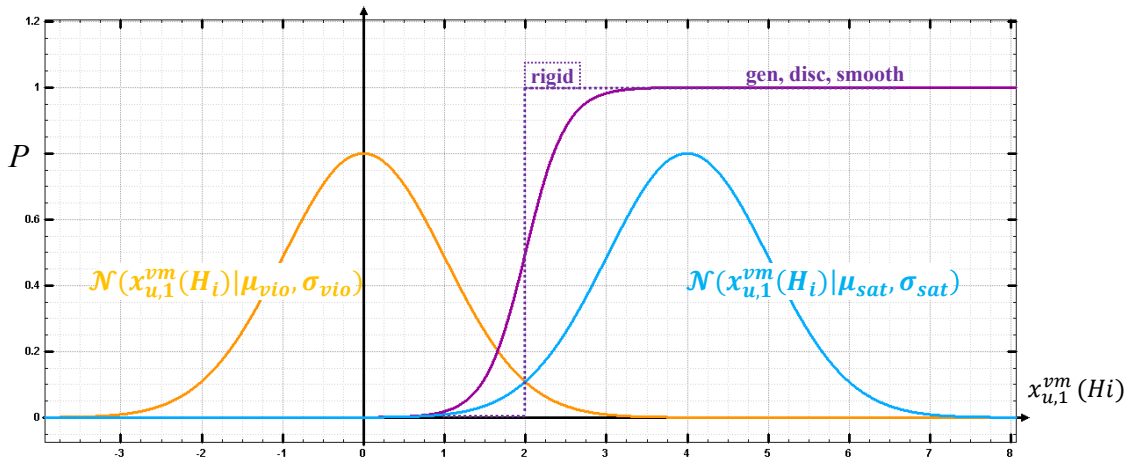


Figure 3.5.: parametrisation approaches for the sigmoid function.

### 3.3. Verification modules

In general, all verification modules rely on state-of-the-art road detection approaches. However, new elements have been added to make these approaches more useful for the verification task. Subsequently, all modules will be introduced by their underlying *models*, their *verification strategies* and their *model-uncertainties*.

#### 3.3.1. Colour classification

The colour classification module is mainly based on the work presented in Fujimura et al. [2008] who developed an approach to correct geometric errors of roads caused by parallel displacements. While the authors restrict their approach to rural areas, the extension presented below operates in urban areas as well.

#### Model

The model basically assumes that image regions belonging to roads can be identified on the basis of their radiometric properties that can be specified in advance, e.g. on the basis of training samples. It is further assumed that such image regions have homogeneous radiometric properties and therefore can be represented by a single grey value for each colour band. While this holds true for the original approach described in [Fujimura et al., 2008], the proposed expansion relaxes the homogeneity assumption slightly so that image regions are represented by two colour vectors. This allows to cover situations where roads are partly affected by shadows which most frequently occurs in urban areas. Here, the partly bright and shadowed road surface is assumed to be characterized by two specific colours.

The model is defined for colour images with a GSD  $\leq 0.5$  m.

#### Verification strategy

In accordance with Fujimura et al. [2008], a two-class SVM is applied to distinguish *road* and *background* image regions. The following paragraphs describe the definition of the *feature vectors*, the *training* and the definition of the *verification output*.

**Definition of the feature vectors:** Initially, a grey value histogram is computed for each colour band inside the image region corresponding to the set  $R_i^{H\bullet} \in H_i$  showing only the road surface if the database hypothesis is correct (cf. Table 3.4). Fujimura et al. [2008] use the mean grey value in each colour band to define a colour feature vector for the respective image region. In this thesis, a Gaussian Mixture Model (GMM) is used to describe the distribution of the grey values in each colour band. For each colour band a GMM is defined on basis of the grey values  $g \in R_i^{H\bullet}$ :

$$f_{\text{GMM}}(g) = \sum_{1 \leq q \leq Q} w_q \mathcal{N}(g | \mu_q, \sigma_q) \quad (3.38)$$

where  $Q$  denotes the number of Gaussian components  $\mathcal{N}(g | \mu_q, \sigma_q)$  and  $w_q$  denotes the mixing coefficients. The GMM is derived by the so-called *expectation maximization method* introduced in

[Dempster et al., 1977] which is explained in the Appendix A.2. Only two of the Gaussian components are assumed to represent the homogeneous road surface that is partly bright and partly shadowed. It is further assumed that these Gaussian components are characterized by rather large mixing coefficients  $w_q$  and low standard deviations  $\sigma_q$ . However, the GMM might also represent other objects such as road markings or vehicles by separate Gaussian components. In order to extract the components that represent the road surface only, Equation 3.38 is altered to:

$$f_{\text{GMM}^*}(g) = \sum_{1 \leq q \leq Q} w_q^* \mathcal{N}(g | \mu_q, \sigma_q) \quad (3.39)$$

where the parameters  $\{\mu_q, \sigma_q\}$  correspond to the initial GMM in Equation 3.38 and the new mixing coefficients  $w_q^*$  are defined for all  $q$  as:

$$w_q^* = \begin{cases} w_q & |w_q \geq 0.1 \wedge \sigma_q < 5 \text{ bit} \wedge q = q_{\max} \\ w_q & |w_q \geq 0.1 \wedge \sigma_q < 5 \text{ bit} \wedge q = q_{2\text{nd}} \\ 0 & |\text{otherwise} \end{cases} \quad (3.40)$$

with

$$q_{\max} = \arg \max_q w_q \quad (3.41)$$

$$q_{2\text{nd}} = \arg \max_{q \setminus q_{\max}} w_q \quad (3.42)$$

This definition leads to non-zero mixing coefficients  $w_q^*$  if the respective Gaussians satisfy three constraints. The first two constraints  $w_q \geq 0.1$  and  $\sigma_q < 5 \text{ bit}$  are defined to suppress Gaussian components that represent smaller objects such as cars or road markings. The threshold of 5 bit for the Gaussian kernel width refers to 8 bit images with 256 possible grey values. The third constraint is only satisfied for the two Gaussian components having the largest mixing coefficient and the second largest mixing coefficients, respectively. Definition 3.40 leads to either *two*, *one* or *zero* mixing coefficients  $w_q^*$  unequal to zero.

In case of *two* non-zero mixing coefficients  $w_q^*$ , the two corresponding  $\mu_q$  values are used as part of the feature vector in an ascending order. This is important because it means that the first vector element always represents a darker region and the second vector element always represents a brighter region. As the procedure is repeated for each colour band, the six-dimensional feature vector has the structure: *dark* (band 1), *bright* (band 1), *dark* (band 2), *bright* (band 2), *dark* (band 3), *bright* (band 3). The idea behind this order is that the image regions are usually not very saturated and the main differences are due to the brightness. In other words, most pixels in an image can be associated with vectors in RGB colour space that are predominantly located near the diagonal in the RGB colour space. While this assumption seems to be quite realistic for most situations in remote sensing images where the colour bands are strongly correlated, its violation would lead to non-unique<sup>4</sup> feature vectors, which is demonstrated in Figure 3.6, a problem that will be discussed

<sup>4</sup>In theory one could circumvent this problem by defining a GMM in 3D. However, the available amount of pixels of approximately 5,000–10,000 per road hypothesis is much too small to learn the nine parameters per Gaussian component simultaneously in contrast to only three parameters per Gaussian component.

in the context of the model-uncertainty.

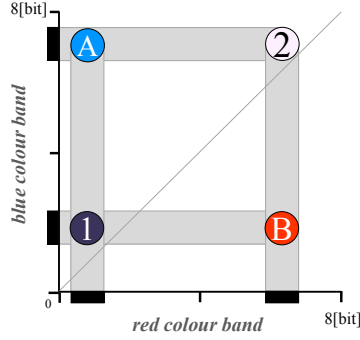


Figure 3.6.: The feature vector representing the clusters 1 and 2 has the same marginal distributions (black bars) as the feature vector representing the clusters A and B. The circles show the colour according to the position in the feature space.

In case of *one* non-zero mixing coefficient  $w_q^*$  the corresponding  $\mu_q$  is used twice as a part of the feature vector. This strategy avoids feature vectors of different dimensions which is important for applying the SVM. In accordance with the initially discussed model, such situations are interpreted as road surfaces that are either completely bright or completely in the shadow.

The third scenario is that no Gaussian component satisfies the constraints in Equation 3.40. In this case the feature vector is not defined, which is considered by the model-uncertainty that will be introduced below.

Figure 3.7 shows the feature vectors definition for several road hypotheses. The left column shows the road hypotheses, the centre column the respective RGB histograms and the right column the derived feature vectors. The  $\mu_q$  for which  $w_q \geq 0.1 \wedge \sigma_q < 5 \text{ bit}$  holds true even if they are unconsidered are set in brackets, while the additionally considered  $\mu_q$  are marked by a plus symbol. The first question that may arise is why only two values  $\mu_q$  are considered for each colour band. The main reason is that each additional feature increases the required amount of training samples and the benefit of an additional mean value is expected to be rather small because many image regions can be represented sufficiently by only one or two mean values, e.g. examples a,b,c in Figure 3.7. In this context it is worth noting that objects that are small compared to the GSD, e.g. cars or road markings, result in Gaussians with large standard deviations because they are mainly represented by mixed pixels (cf. Figure 3.7c). Hence, the mean values do not represent the colour properties of small objects well. This leads to another question: Why are only the mean values considered and not the other GMM parameters  $w_q, \sigma_q$ ? One reason can be seen from the examples depicted in the Figures 3.7a and b. The appearance of the two correct road regions in the image is rather different because the relative amounts of shadowed and bright road parts are different. However, this difference only affects the parameters  $w_q$  and  $\sigma_q$  and not the  $\mu_q$ . Hence, the proposed strategy requires fewer training samples because the relative amount of shadow is nearly irrelevant.

Finally, the restriction to (only) six features preserves the basic idea of the original method proposed by Fujimura et al. [2008] who defined a relatively simple model that is still able to provide good results without a high amount of training data. A possible disadvantage of the introduced two-colour model can be seen in Figure 3.7d, where an incorrect road hypothesis is insufficiently described by only two Gaussians. The ideal case would be an incorrect road region that is completely

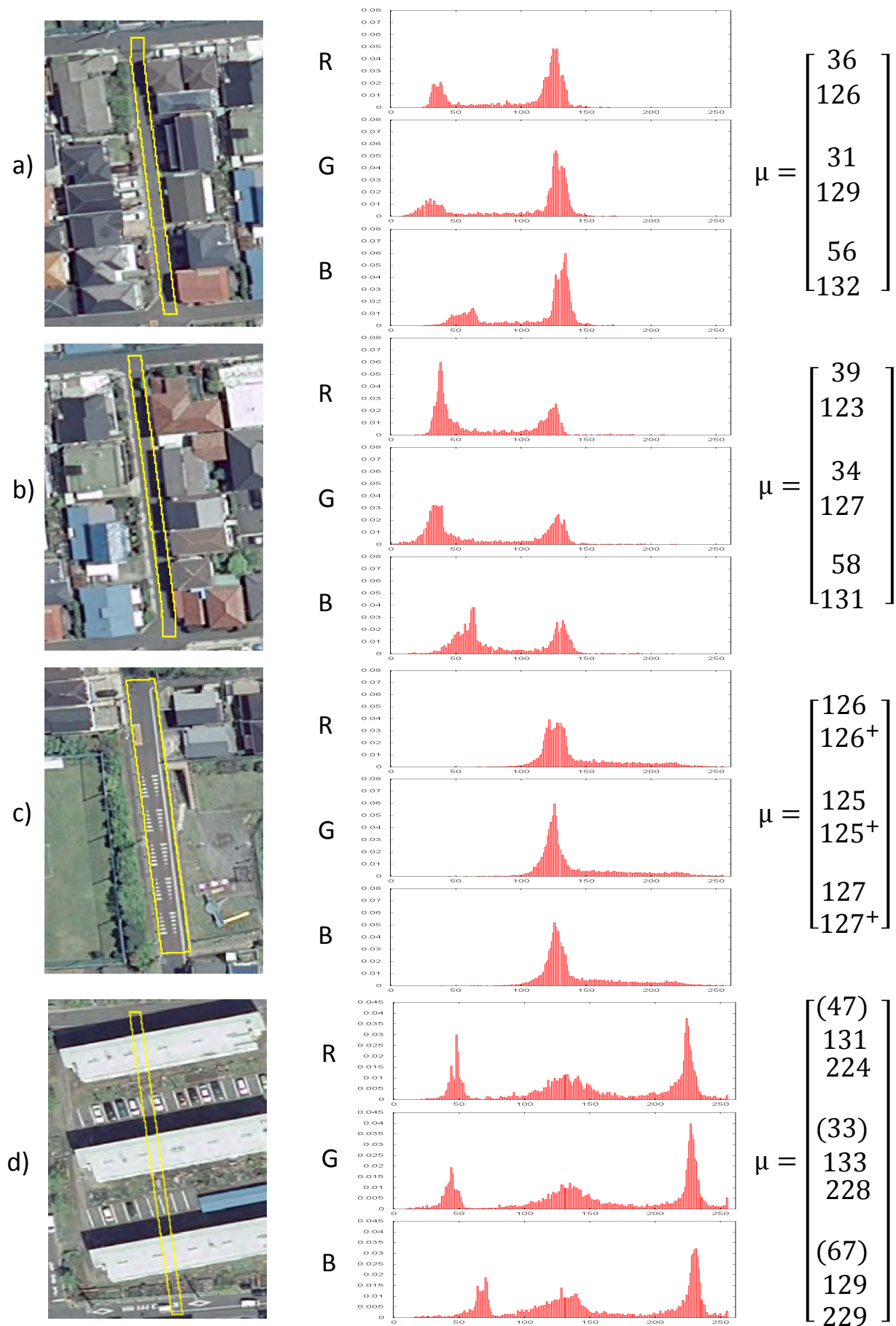


Figure 3.7.: Colour module feature extraction for four different road hypotheses.



covered by a maximum of two background objects. Despite that problem the proposed extension prepares the approach for dealing with shadows which is a weak point of most state-of-the-art road detection approaches. As the other verification modules do not consider shadow in an explicit way, this expansion is of central importance for the overall verification approach as will be shown by the experiments in Chapter 4.

**SVM training:** For the classification a  $\nu$ -SVC [Schölkopf et al., 2000]<sup>5</sup> in the implementation of the open-source library LIBSVM [Chang & Lin, 2001] is trained on the basis of a set of feature vectors  $X^{\text{svmtrain}}$ . The training dataset contains correct road regions but also background regions such as buildings and grassland. The basis for the training dataset are the image regions corresponding to sets  $X^{\text{roads}}$ ,  $X^{\text{buildings}}$ ,  $X^{\text{grassland}}$  (cf. Table 3.3). However, only feature vectors whose computation satisfies a pre-defined set of critical assumptions are considered in  $X^{\text{svmtrain}}$  which will be dealt with below in the context of model-uncertainty. Based on  $X^{\text{svmtrain}}$  an RBF kernel is used for the  $\nu$ -SVC (cf. Equation 2.39) whose hyperparameter  $\sigma$  is determined by Equation 2.45. The second hyperparameter  $\nu$  is determined by a step-wise search for  $\nu = \{0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30\}$  based on three-fold cross-validation [Chang & Lin, 2001].

**Classification and verification output:** A feature vector  $z_i$  corresponding to  $x_v^{\text{col}}(H_i)$  is classified by the previously trained model. The binary classification output of the  $\nu$ -SVC is interpreted directly to define the probabilities required for the given verification problem:

$$P(\text{road} = \text{correct} | x_v^{\text{col}}(H_i), \text{model} = a.) = \begin{cases} 1.0 & |\nu - \text{SVC classified road} \\ 0.0 & |\nu - \text{SVC classified background} \end{cases} \quad (3.43)$$

The output of the  $\nu$ -SVC is a binary class assignment, and consequently also the verification output is binary. Some approaches, e.g. the one in [Platt, 2000] have demonstrated that for two-class SVMs probabilistic distributions can be defined on the basis of the distance to the hyperplane. However, in this thesis this possibility is not investigated further.

## Model-uncertainty

Four *critical assumptions* have been identified for the colour classification module. The fourth assumption is specific for each colour band and thus considered three times which leads to, in total, seven critical assumptions.

**Critical assumption 1: The training samples represent the test data.** As already discussed in Section 2.2, this is a critical assumption frequently made by applications of machine learning. In order to detect potential violations, an SVDD is trained with the same training data as the  $\nu$ -SVC. In contrast to the  $\nu$ -SVC both classes *road* and *background* are aggregated to one *inlier* class. The respective kernel function is selected in full accordance with the  $\nu$ -SVC. The regularization parameter is set to:

$$C = (0.01 \cdot n)^{-1} \quad (3.44)$$

<sup>5</sup>Schölkopf et al. [2000] introduced the  $\nu$ -SVC as a variant of the conventional two-class SVM described by Vapnik [1995]. The  $\nu$ -SVC replaces the regularization parameter  $C$  by a parameter  $\nu$ . Using  $\nu$  has the advantage that it can be interpreted as the probability that a training sample is located on the wrong side of the hyperplane.

As  $n$  denotes the number of training samples, this will usually result in a low degree of generalization, i.e. the SVDD hypersphere will be determined so that there will only be a few training samples outside the SVDD hypersphere (cf. Section 2.2.3). In accordance with Equation 2.56 the signed feature space distance  $f_{\text{SVDD}}(z_i)$  of each test sample  $z_i$  is obtained. As described in Section 2.2.3  $f_{\text{SVDD}}(z_i)$  denotes the distance of a test sample  $z_i$  to the training data represented by the SVDD surface. Basically, a distance  $f_{\text{SVDD}}(z_i) \leq 0$  indicates that  $z_i$  is represented by training data while a distance  $f_{\text{SVDD}}(z_i) > 0$  indicates that  $z_i$  is not represented by the training data. Consequently,  $f_{\text{SVDD}}(z_i)$  allows to evaluate the fulfilment of the first critical assumption, and hence is directly interpreted as the indicator feature to this critical assumption:

$$x_{u,1}^{\text{col}}(H_i) = f_{\text{SVDD}}(z_i) \quad (3.45)$$

Next, the distribution  $P(a_1 = \textit{satisfied} | x_{u,1}^{\text{col}}(H_i), \theta)$  has to be defined. This is difficult because the typical distance from the SVDD surface to an *outlier* is basically unknown. Note that the *outliers* were not part of the training. As a consequence, knowledge about the distribution of the *outliers* is required. In the test phase (the per-hypothesis analysis) *outliers* are very likely to appear. In order to make that knowledge available, the SVDD *test data analysis* is introduced as an extra module outside the *core component* (cf. Figure 3.3) which shall be explained briefly.

As a part of the *support component* this module has access to the distances  $f_{\text{SVDD}}(z_j)$  of all test samples  $z_j$ . Hence, a set of all distances is defined:  $f_{\text{SVDD}}(z_j) \in X^{\text{SVDDtest}}$ , for which the two subsets for *inliers*  $X^{\text{in}}$  and *outliers*  $X^{\text{out}}$  with  $X^{\text{in}}, X^{\text{out}} \subseteq X^{\text{SVDDtest}}$  are defined by applying the natural decision threshold introduced by Tax & Duin [2004]:

$$\begin{aligned} X^{\text{in}} &= \{f_{\text{SVDD}}(z_j) \in X^{\text{SVDDtest}} | f_{\text{SVDD}}(z_j) \leq 0\} \\ X^{\text{out}} &= \{f_{\text{SVDD}}(z_j) \in X^{\text{SVDDtest}} | f_{\text{SVDD}}(z_j) > 0\} \end{aligned} \quad (3.46)$$

Due to the hard decision threshold, the distributions of the  $f_{\text{SVDD}}(z_j)$  for  $X^{\text{in}}$  and  $X^{\text{out}}$  do not overlap. Now, it is supposed that these two distributions do not represent the true distributions for the *inliers* and the *outliers*. Instead, the true distributions are assumed to correspond to Gaussians. Therefore,  $X^{\text{in}}$  and  $X^{\text{out}}$  are used to estimate the parameters for the respective Gaussians  $\mathcal{N}(f_{\text{SVDD}}(z_j) | \mu_{\text{in}}, \sigma_{\text{in}}, P_{\text{in}})$  and  $\mathcal{N}(f_{\text{SVDD}}(z_j) | \mu_{\text{out}}, \sigma_{\text{out}}, P_{\text{out}})$  that are exemplary shown in Figure 3.8. The determined Gaussian parameters are plugged into Equation 3.19 which corresponds to the generative parametrisation approach introduced in Section 3.2.2:

$$P(a_1 = \textit{satisfied} | x_{u,1}^{\text{col}}(H_i), \theta_{\text{gen}}\{\mu_{\text{in}}, \sigma_{\text{in}}, P_{\text{in}}, \mu_{\text{out}}, \sigma_{\text{out}}, P_{\text{out}}\}) \quad (3.47)$$

One consequence of the definition in Equation 3.47 is that the decision threshold  $f_{\text{SVDD}}(z) = 0$  suggested by Tax & Duin [2004] is adapted according to the test data. Empirical analyses with different datasets have shown that this adapted decision threshold represented by  $f_{\text{SVDD}}(z) | P(a_1 = \textit{satisfied} | x_{u,1}^{\text{col}}(H_i), \theta_{\text{gen}}) = 0.5$  is mostly larger than zero, the decision threshold suggested by Tax & Duin [2004]. This finding corresponds to the intuition that test samples in the vicinity of the training data are still covered by the trained model. The example depicted in Figure 3.8 has the new decision threshold approximately at 1.2. The unit of distance drawn on the x-axis in Figure 3.8

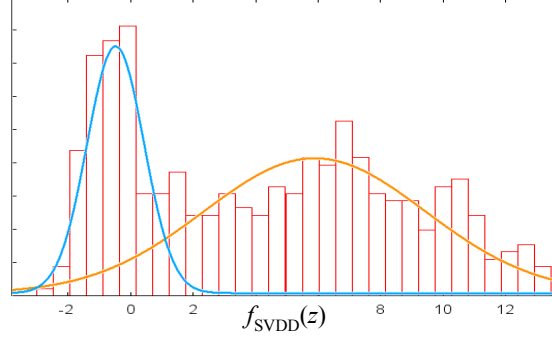


Figure 3.8.: Histogram of feature space distances to the SVDD hypersurface, computed from 1,154 test samples and overlaid by the estimated distributions  $\mathcal{N}_{\text{in}}$  in blue and  $\mathcal{N}_{\text{out}}$  in orange.

corresponds to the average distance of training data from the SVDD hypersurface.

**Critical assumption 2: The feature vector definition is unique.** As mentioned above, the chosen order of the vector elements requires the colour bands to be positively correlated. This problem can be geometrically described by considering plausible colour feature vectors only to reside in a small margin around the diagonal in RGB space. Outside that region the feature vector definition is not unique. Subsequently, *saturation* and *intensity* are used to indicate the positions of the vectors in the RGB space. For each pixel  $p_b \in R_i^{\text{H}\bullet}$  saturation  $S(p_b)$  and intensity  $I(p_b)$  are determined in accordance with Burger & Burge [2005]<sup>6</sup>:

$$\begin{bmatrix} M_1 \\ M_2 \\ I_1 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3.48)$$

$$S(p_b) = \sqrt{M_1^2 + M_2^2} \quad (3.49)$$

$$I(p_b) = \frac{R + G + B}{3} \quad (3.50)$$

The presence of off-diagonal colour vectors is indicated by a high variance of saturation, considering zero to be the expectation value for a feature value near the diagonal in RGB space:

$$\text{Var}_s(R_i^{\text{H}\bullet}) = \frac{1}{|R_i^{\text{H}\bullet}|} \sum_{p_b \in R_i^{\text{H}\bullet}} [S(p_b)]^2 \quad (3.51)$$

This value needs to be set in relation to the differences in direction of the RGB-diagonal. Hence, the variance of intensity  $\text{Var}_I(R_i^{\text{H}\bullet})$  is defined by considering the mean intensity  $\bar{I}(R_i^{\text{H}\bullet})$  to be the expectation value:

$$\text{Var}_I(R_i^{\text{H}\bullet}) = \frac{1}{|R_i^{\text{H}\bullet}|} \sum_{p_b \in R_i^{\text{H}\bullet}} \left[ I(p_b) - \bar{I}(R_i^{\text{H}\bullet}) \right]^2 \quad (3.52)$$

<sup>6</sup>The definition of Burger & Burge [2005] is different from the standard definition for saturation  $S(p_j)$  as there is no normalization over the intensity  $R + G + B$ .

The fulfilment of the critical assumption is evaluated by the relative amount of both variances which only becomes relevant if two dominant Gaussians have been observed:

$$x_{u,2}^{\text{col}}(H_i) = \begin{cases} 0.0 & | \text{one dominant Gaussian} \\ \frac{\text{Var}_s(R_i^{\text{H}\bullet})}{\text{Var}_i(R_i^{\text{H}\bullet})} & | \text{two dominant Gaussians} \end{cases} \quad (3.53)$$

The critical assumption is satisfied if  $x_{u,2}^{\text{col}}(H_i) = 0.0$  because this means that only one combination of the vector components is possible. In contrast, the critical assumption is questionable if  $x_{u,2}^{\text{col}}(H_i) = 1.0$  because this means that the clusters 1, 2 and A, B in Figure 3.6 are equally plausible. Furthermore, it can be expected that the probability of satisfying the critical assumption increases smoothly with increasing values  $x_{u,2}^{\text{col}}(H_i)$ . This heuristic is used to define the parameters of the sigmoid function such that:

$$P(a_2 = \text{satisfied} | x_{u,2}^{\text{col}}(H_i), \theta_{\text{smooth}} \{0.0, 1.0\}) \quad (3.54)$$

**Critical assumption 3: There is sufficient colour contrast.** This critical assumption concerns the colour contrast between *road* and *background* image regions. In areas with large parking lots or building roofs that have similar colour properties as roads, the  $\nu$ -SVC tends to classify background regions as roads. In order to evaluate the contrast locally, the adjoint regions on both sides of the road hypothesis are classified in the same way as the original region ( $R_i^{\text{H}\bullet}$ ). For that purpose, two additional image regions are defined by displacing the original region in a direction perpendicular to the road centreline by twice of the road width. As background regions corresponding to grassland or buildings are expected to be next to a road, sufficient contrast is only supposed to occur if both displaced regions are classified as *background*. Accordingly, if the road hypothesis region is classified as *road* and at least one of the displaced regions is also classified as *road*, the indicator feature  $x_{u,3}^{\text{col}}(H_i)$  is set to zero and set to one otherwise. Only in the latter case the critical assumption is supposed to be satisfied. This heuristic is rigid, and thus the rigid parametrisation approach is chosen to define the sigmoid function, respectively:

$$P(a_3 = \text{satisfied} | x_{u,3}^{\text{col}}(H_i), \theta_{\text{rigid}} = \{1.0\}) \quad (3.55)$$

**Critical assumptions 4–6: The GMM\* represents the actual distribution.** As can be seen in Figure 3.7d, image regions may be poorly represented by two Gaussians, while other image regions may not fit to a GMM at all. Another problem is that the expectation maximization method does not necessarily provide an optimal GMM for a set of grey values. Consequently, the  $f_{\text{GMM}^*}(g)$  (cf. Equation 3.39) might not be a good representation for the stated problem. In order to describe this case, the indicators to the critical assumptions 4–6 are defined based on *Pearson's correlation coefficient* [Press et al., 2007] between the distribution introduced by  $f_{\text{GMM}^*}(g)$  and the actual distribution

$f_{AD}(g)$  introduced by the histogram of the respective colour band:

$$x_{u,4-6}^{\text{col}}(H_i) = \begin{cases} \frac{\sum_{1 \leq g \leq 256} f_{AD}(g) f_{GMM^*}(g)}{\sqrt{\sum_{1 \leq g \leq 256} (f_{AD}(g))^2} \cdot \sqrt{\sum_{1 \leq g \leq 256} (f_{GMM^*}(g))^2}} & | \sum_{1 \leq g \leq 256} f_{GMM^*}(g) > 0 \\ 0.0 & | \text{otherwise} \end{cases} \quad (3.56)$$

The definition in Equation 3.56 is similar for all three colour bands, and thus holds true for the three indicator features  $x_{u,4-6}^{\text{col}}$ . It has to be noted that  $f_{GMM^*}(g)$  does not necessarily describe a complete density distribution because generally  $\sum_{1 \leq g \leq 256} f_{GMM^*}(g) \leq 1.0$  has to be taken into account. One consequence is that the case where  $\sum_{1 \leq g \leq 256} f_{GMM^*}(g) = 0$  needs to be considered (cf. second case in Equation 3.56) because it corresponds to the third scenario where no Gaussian component satisfies the constraints (cf. Equation 3.40). Another consequence is that those *incomplete distributions* also let  $x_{u,4-6}^{\text{col}}(H_i)$  decrease with  $\sum_{1 \leq g \leq 256} f_{GMM^*}(g)$ . In particular  $\lim_{\sum f_{GMM^*}(g) \rightarrow 0} [x_{u,4-6}^{\text{col}}(H_i)] = 0$ . Both circumstances are used to define a heuristic, according to which the critical assumptions 4–6 are fulfilled if  $x_{u,4-6}^{\text{col}}(H_i) = 1.0$  but violated if  $x_{u,4-6}^{\text{col}}(H_i) = 0.0$ . Furthermore, it can be expected that the probability of satisfying the critical assumptions smoothly increases with increasing values  $x_{u,4-6}^{\text{col}}(H_i)$ . This heuristic is implemented to define the sigmoid function as follows:

$$P(a_{4-6} = \text{satisfied} | x_{u,4-6}^{\text{col}}(H_i), \theta_{\text{smooth}} = \{1.0, 0.0\}) \quad (3.57)$$

It is worth noting that these definitions probably violate the assumption of independence in Equation 3.16, because the features  $x_{u,4-6}^{\text{col}}(H_i)$  are defined by the same model (cf. Equation 3.56) based on (probably) positive correlated data. Consequently,  $P(\text{model} = n/a | x_u^{\text{col}}(H_i))$  may be defined too large, which would reduce the effect of the colour classification module on the final verification result.

**Criterion for training dataset definition:** The discussed critical assumptions with  $k = \{2, 4, 5, 6\}$  explicitly denote problems of the feature vector definition. In the training dataset  $X^{\text{SVMtrain}}$  only those feature vectors are included that satisfy the following inequality:

$$\prod_{k=\{2,4,5,6\}} [P(a_k = \text{satisfied} | x_{u,k}^{\text{col}}(R_t), \theta)] \geq 0.8 \quad (3.58)$$

## Discussion

The most important aspect of the colour verification module is its ability to verify roads that are partly in the shadow. Unlike other state-of-art approaches, shadow effects are not only modelled as a possible disturbance for the appearance of the road surface. Instead, the shadowed surface is considered in a similar way as bright areas, contributing features to the road region. Apart from this, the model is rather simple and does not cover all imaginable situations. In particular, it does not cover road hypotheses that are characterized by more than two different colours, which however, leads to a violation of the critical assumptions 4–6. Such situations may occur more frequently for

incorrect road hypotheses, e.g. those that correspond to more than one topographic object, such as in the example shown in Figure 3.7d. However, road databases can be expected to be widely correct, and thus this problem is expected to be less relevant. An even more problematic aspect is the colour contrast (cf. critical assumption 3), because sometimes whole city-districts are characterized by building roofs that have similar radiometric properties as roads. Note that the model-uncertainty considers also this scenario.

Another important aspect of the colour verification module is that the model is instantiated on the basis of training data, which is advantageous because radiometric properties of roads and even more those of background objects can hardly be defined manually. Finally, the introduced concept for the model-uncertainty allows a human operator to select partly non-representative training data without direct consequences for the correctness of the final verification result.

### 3.3.2. Line detection

The three modules: *pan-line detection*, *NDVI-line detection* and *nDSM-line detection* share the same strategy, and hence are described together in this section. Each module relies on different input data, namely panchromatic, NDVI and nDSM images. If a panchromatic image is not available, the colour band with the shortest wavelength is selected instead, which turned out to be the best choice based on empirical analysis. NDVI denotes the normalized digital vegetation index and nDSM the normalized digital surface model; both are encoded by 8 bit images.

The underlying models correspond to the works described in [Gerke & Heipke, 2008] for pan-line detection, in [Gerke & Busch, 2005] for NDVI-line detection and in [Hinz & Baumgartner, 2003] for nDSM-line detection. The strategy mainly corresponds to [Gerke & Heipke, 2008] who developed a verification approach that focuses on rural areas. Extensions compared to this approach are the use of different input data (NDVI, nDSM) and to a large extent the approach to determine the model-uncertainty.

## Models

The basic model which holds true for all three modules assumes that a road appears as a line of homogeneous grey values that has a specific width in the imagery of about 1–2 m GSD. Depending on the input data the models have different interpretations:

**Pan-line detection:** In homogeneous context areas the roads appear as bright lines in the panchromatic image. The brightness of these lines can be specified in form of a scene-dependent grey value interval which can be defined on the basis of a few training samples selected from the same scene.

**NDVI-line detection:** In vegetated areas the roads appear as dark lines in the NDVI image. The maximum brightness of the lines can be specified in form of a scene-dependent value which can be defined on the basis of a few training samples selected from the same scene.

**nDSM-line detection:** In dense urban areas roads appear as dark lines in the nDSM image. The maximum brightness of the lines can be generally specified in form of a value that represents

the threshold between above-ground objects such as buildings and trees and the ground surface.

### Verification strategy

In accordance with [Gerke & Heipke, 2008], lines are extracted from the image and then verified by comparing the extraction results with the database hypothesis. It is important to note that the three line-based modules operate independently. This is different to [Gerke & Heipke, 2008] who introduced independent detection methods but only one verification method that considers all extracted geometries simultaneously. This means that in this work the pan-based module only relies on lines detected in the panchromatic image, and so on. This difference is a consequence of the modular concept which is expected to lead to simpler parametrisation as discussed in Section 3.1. This aspect becomes relevant here, because the overall approach proposed here combines ten different verification methods and not only two as in [Gerke & Heipke, 2008].

**Line extraction:** The line detector introduced by Steger [1998] is applied to image subsets of the panchromatic, NDVI or nDSM input image, where the extent of those image subsets is defined by the set  $R_i^{[H]} \in H_i$  (cf. Table 3.4). In the right column of Figure 3.9 three examples for such image subsets are shown. Steger [1998] defined a set of parameters that allow adjustments of his approach to different problems. These parameters are used to specify the line models with respect to the input imagery (pan, NDVI and nDSM), but also to the radiometric scene properties and the road width stored in the database. Some of the parameters are learned on the basis of the training datasets (cf. Table 3.3); others are fixed in accordance with the introduced road models. The fixed parameters are interpreted as parts of the underlying road models. They are not interpreted as parameters of the proposed approach, and thus will not be discussed beyond this section. The parameters for the Steger-algorithm are given in Table 3.6.

	pan	NDVI	nDSM
line width	$h_i^{\text{width}}$	$1.5 \cdot h_i^{\text{width}}$	$1.5 \cdot h_i^{\text{width}}$
target	bright line	dark line	dark line
$T_{\text{low}}$	<i>trained</i>	$\hat{=}$ -1.0	$\hat{=}$ 0.0 [m]
$T_{\text{high}}$	<i>trained</i>	<i>trained</i>	$\hat{=}$ 4.0 [m]
contrast high	<i>trained</i>	<i>trained</i>	$\hat{=}$ 4.0 [m]
contrast low	<i>trained</i>	<i>trained</i>	$\hat{=}$ 3.0 [m]

Table 3.6.: Parameter settings for the Steger line detection algorithm with respect to the three possible inputs: panchromatic, NDVI and nDSM.

The first parameter in Table 3.6, the *line width*, is set in accordance with the database hypothesis  $h_i^{\text{width}} \in H_i$  (cf. Table 3.4). The parameter *line width* is among others responsible for adapting the Steger-algorithm to a GSD < 1.0 m. As paved roadsides and side walks do not show a contrast to the roads in NDVI and nDSM images, the *line width* is enlarged compared to the database information. The second parameter *target* = {*bright line*, *dark line*} is a trivial consequence of the model, and thus a constant for each type of input data. The parameter  $T_{\text{low}}$  denotes the minimum grey value allowed for pixels that correspond to the line. In case of NDVI and nDSM the thresholds are fixed to the theoretical minima. For a better understanding, in Table 3.6 the values for NDVI and nDSM

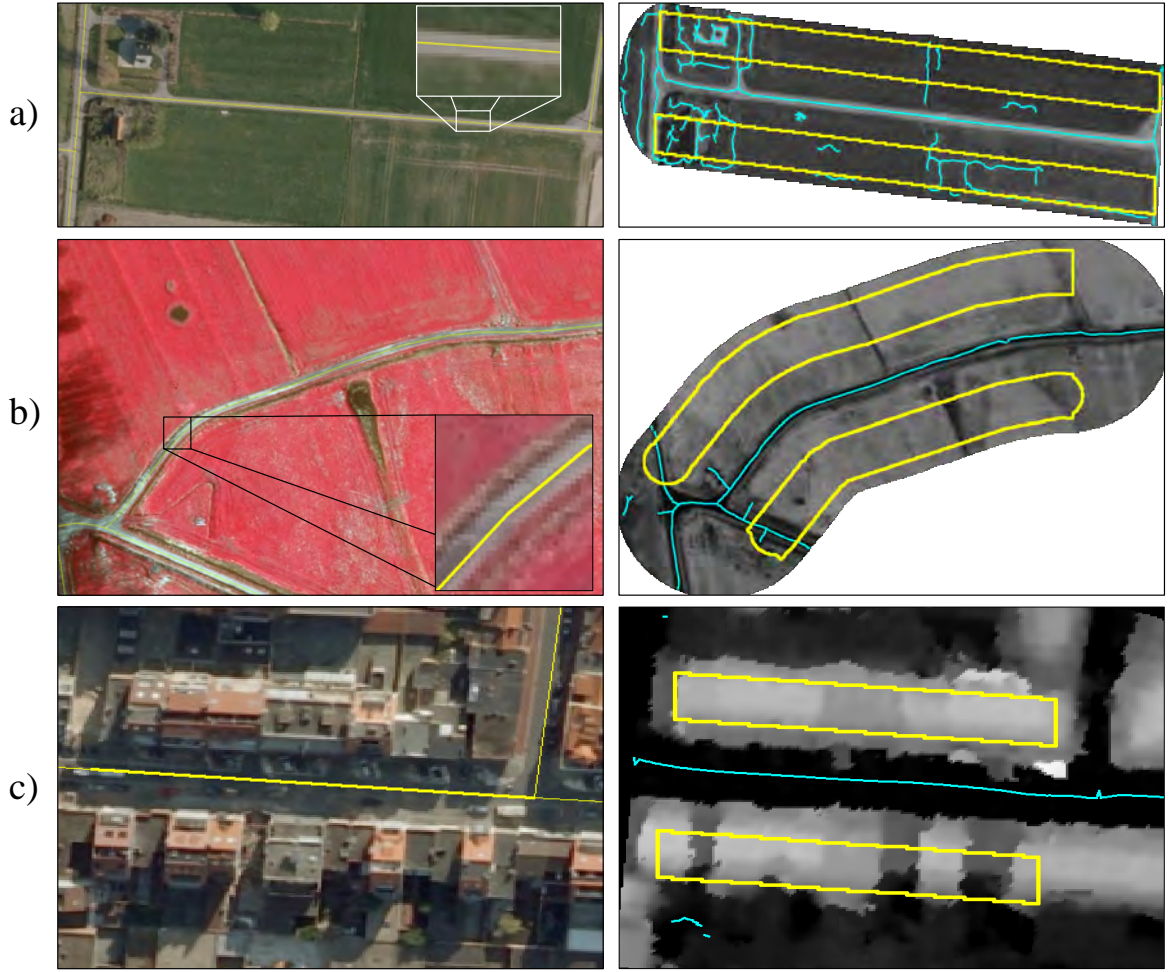


Figure 3.9.: Pan-line detection (a), NDVI-line detection (b) and nDSM-line detection (c). Left: Road hypotheses given in the database (yellow). Right: Outputs of the Steger-algorithm (cyan) and the context buffer (yellow) defined by the set  $R_i^H$ .

are given with their original units, emphasized with the  $[\hat{=}]$  symbol. The parameter  $T_{\text{high}}$  denotes the maximum grey value for pixels that correspond to the line. In case of nDSM the value is fixed to 4 m, which is large enough to include trucks on the road surface and small enough to exclude small buildings and trees. The parameter *contrast high* denotes the minimum contrast between grey values that correspond to the line and grey values next to the line. The parameter *contrast low* denotes the maximum difference of grey values corresponding to the same line primitive. In case of nDSM the values are fixed to 4 m and 3 m, respectively, which is in accordance with the former discussion about trucks and buildings. All parameters denoted by *trained* in Table 3.6 are automatically determined on the basis of the training dataset for roads ( $X^{\text{roads}}$ , Table 3.3). The training algorithm analyses the grey value histograms from image regions that correspond to the training data and their local background. Further details of the training algorithm are given in [Ziems et al., 2007].

The right column in Figure 3.9 shows the line primitives (cyan) extracted by the different line detection modules. All extracted line primitives that lie near the road hypothesis are subsequently considered for the verification, while all others are ignored. A line primitive  $y_p \in Y^{\text{steger}}$  is considered here if it satisfies the constraint  $\text{reg}(y_p) \cap R_i^{H\bullet} > 0$ , where the function *reg* provides the set of pixels corresponding to the argument, i.e. all pixels of the input image that are intersected by a line



primitive  $y_p$ ; see Figure 3.10 for an example.

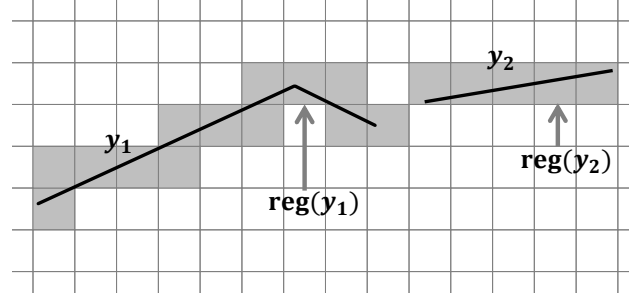


Figure 3.10.: Principle for the definition of pixel sets corresponding to line primitives;  $\text{reg}$  provides sets of pixels (shaded grey squares) under the line primitives  $y_1$  and  $y_2$ .

**Comparison of extracted lines with the database hypothesis:** This step is identical to the way described in [Gerke, 2006; Gerke & Heipke, 2008]. In the following, a brief summary of their work will be presented in order to motivate the subsequent discussions.

Initially, each extracted line primitive  $y_p \in Y^{\text{steger}}$  is projected to the centreline hypothesis  $h_i^{xy} \in H_i$ , resulting in the projected primitive  $y'_p$  for each  $y_p \in Y^{\text{steger}}$ . This projection is an orthogonal projection with respect to the centreline  $h_i^{xy} \in H_i$ ; see Figure 3.11 for an example.

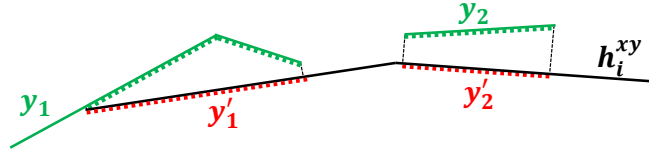


Figure 3.11.: Principle of the orthogonal projection of extracted primitives on the centreline hypothesis [Gerke, 2006]. The line primitives  $y_1$  and  $y_2$  are projected onto the centreline hypothesis  $h_i^{xy}$  which results in the projected primitives  $y'_1$  and  $y'_2$ .

In order to check whether  $y_p$  corresponds to the road hypothesis, and thus whether  $y_p$  and  $y'_p$  can be assumed to be identical, two statistical tests are applied. The first test checks the identity of the *shape*, providing a probability  $P(\text{shape} = \text{identical} | y_p, y'_p)$ . For that purpose, the differences of the first eight translation and rotation invariant line moments are determined. The second test checks the identity of the *orientation* providing a probability  $P(\text{orientation} = \text{identical} | y_p, y'_p)$ . The *orientation* is defined as the angle from the start and the end points of  $y'_p$  and  $y_p$ , respectively. Both tests are considered to be independent, and thus the identity of  $y'_p$  and  $y_p$  is defined as the joint probability (see Figure 3.12 for the possible constellations):

$$P(\text{primitives} = \text{identical} | y_p, y'_p) = P(\text{shapes} = \text{identical} | y_p, y'_p) \cdot P(\text{orientations} = \text{identical} | y_p, y'_p) \quad (3.59)$$

The outputs related to all line primitives  $y_p \in Y^{\text{steger}}$  are combined to one solution concerning the whole road hypothesis represented by  $P(\text{road} = \text{correct} | Y^{\text{steger}})$ . The length  $l'_p$  of a projected line primitive  $y'_p$  controls its impact on the combined output such that:

$$P(\text{road} = \text{correct} | Y^{\text{steger}}) = \frac{1}{N} \cdot \sum_p [l'_p P(\text{primitives} = \text{identical} | y_p, y'_p)] \quad (3.60)$$

where  $N$  denotes a normalizing constant to make  $P(\text{road} = \text{correct} | Y^{\text{steger}})$  a probability.

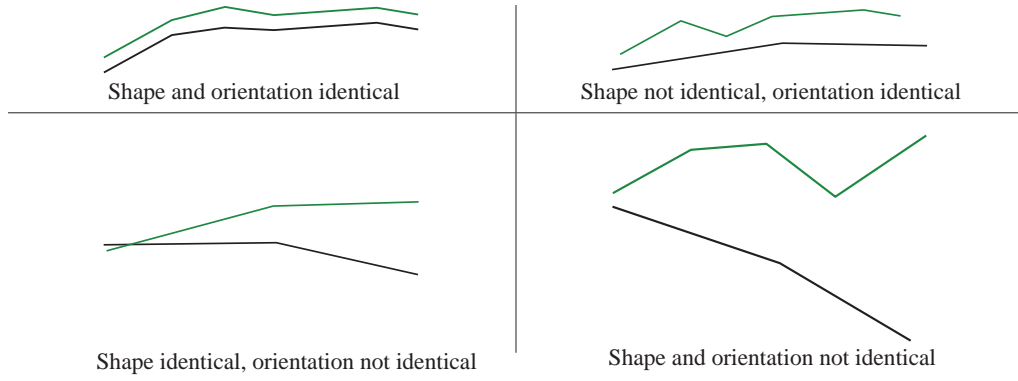


Figure 3.12.: Identity of shape and orientation of two line primitives [Gerke, 2006].

**Verification output:** While Gerke & Heipke [2008] interpret  $P(\text{road} = \text{correct} | Y^{\text{steger}})$  already as a posterior probability, in this thesis it is interpreted as the module-specific verification output under the assumption that the respective line model is *applicable* for the dataset  $H_i$ . Hence, the outputs for the three line detection modules are defined as follows:

$$P(\text{road} = \text{correct} | x_v^{\text{lin1}}(H_i), \text{model} = a.) \Leftrightarrow P(\text{road} = \text{correct} | Y^{\text{steger}}(\text{pan})) \quad (3.61)$$

$$P(\text{road} = \text{correct} | x_v^{\text{lin2}}(H_i), \text{model} = a.) \Leftrightarrow P(\text{road} = \text{correct} | Y^{\text{steger}}(\text{NDVI})) \quad (3.62)$$

$$P(\text{road} = \text{correct} | x_v^{\text{lin3}}(H_i), \text{model} = a.) \Leftrightarrow P(\text{road} = \text{correct} | Y^{\text{steger}}(\text{nDSM})) \quad (3.63)$$

### Model-uncertainty

Many authors discuss the strengths and limitations of line-based models with respect to global context areas. Gerke & Heipke [2008] restrict their pan-based line model to rural areas, where the surroundings of the roads can be assumed to be homogeneous. With respect to the NDVI-based line model this means that homogeneous vegetation is a precondition, and thus this model is basically restricted to areas of rich vegetation. The model assumption of nDSM-based line detection approaches is that there are buildings on both sides of the road which implies dense urban areas. Another assumption of all three models is that a road surface appears more or less homogeneous (cf. Steger-parameter *contrast low* in Table 3.6). For instance, tyre marks on the asphalt, road markings or cars contradict that assumption in case of the pan-line detection. An advantage of the NDVI-line model is that such situations do not violate the model assumptions, but vegetation areas on the road surface or overhanging trees still do. Considering that the line detection modules will be applied to different context areas and geographical regions without restrictions, not all road hypotheses will necessarily correspond to the line models. Consequently, the assumption that roads appear as such lines is basically questionable, and thus considered as the first critical assumption for the model-uncertainty.

Another frequently discussed problem in the literature is that line detection approaches produce a comparably large amount of false alarms (cf. discussion in Section 1.3.1). A reason for that is the compliance of other landscape objects with the line model. Hence, the assumption implicitly made by the line detection modules that their underlying model is discriminative, is also questionable.

**Critical assumption 1: Roads correspond to the Steger-line-model.** If the assumption of homogeneous road surfaces is violated, the line detector simply does not have any response, even if a road exists at the considered position. This will also happen if the local surroundings appear heterogeneous so that their contrast with the road is too low to be detected (cf. Steger parameter *contrast high* in Table 3.6). Thus, a road or parts thereof might be overlooked by the verification step, and as a consequence the verification output relies on incomplete data. This problem is also considered by Gerke & Heipke [2008] in form of the coverage  $c_i$  with:

$$c_i = \frac{h_i^{\text{length}'}}{h_i^{\text{length}}} \quad (3.64)$$

where  $h_i^{\text{length}} \in H_i$  denotes the total length of the road centreline hypothesis and  $h_i^{\text{length}'}$  the length of the centreline covered by projected line primitives  $y'_p$ . This definition of the coverage is directly interpreted as the indicator feature to the critical assumptions:

$$x_{u,1}^{\text{lin1-3}}(H_i) = c_i \quad (3.65)$$

[Gerke & Heipke, 2008] further define a rigid lower bound  $c_i = 0.8$ . Accordingly in this thesis, the probability of the state satisfying the critical assumption is defined on the basis of the rigid heuristic model:

$$P(a_1 = \text{satisfied} | x_{u,1}^{\text{lin1-3}}, \theta_{\text{rigid}} = \{0.8\}) \quad (3.66)$$

**Critical assumption 2: Only roads correspond to the Steger line model.** This is a critical assumption as non-road objects such as rivers, rows of trees, elongated buildings or gateways to buildings may also appear as lines, very similar to roads. In the literature, this problem is often referred to as the false alarm problem of the line-based road detection approaches (cf. discussion in Section 1.3.1). Against the background of verification, this is usually less critical, because only image regions  $R_i^{\text{H}^\bullet} \in H_i$  are considered. Only if many of these objects appear within this relatively small image subset they gain influence to the verification result. Hence, the probability that other linear objects occur in these small regions, is interpreted to correspond to the probability that the critical assumption is *violated*. In accordance with Lu et al. [2008], the probability of the occurrence of linear structures in remote sensing images can be related to the image entropy determined for the local surroundings of a road hypothesis. Hence, the indicator feature for the critical assumption is defined by the image entropy:

$$x_{u,2}^{\text{lin1-3}}(H_i) = - \sum_{T_{\text{low}} \leq g \leq T_{\text{high}}} f_{\text{H}}(g) \cdot \log f_{\text{H}}(g) \quad (3.67)$$

where  $f_{\text{H}}(g)$  denotes the frequency of the grey values  $g$  defined in form of an 8 bit histogram for the image region  $R_i^{\text{H}^\bullet} \in H_i$ , corresponding to the yellow boxes in Figure 3.9. The domain in Equation 3.67 is restricted to the parameters  $T_{\text{low}}, T_{\text{high}}$  because they define the lower and upper bounds of pixels for the line primitives similar to the parameter set for the Steger-algorithm (cf. Table 3.6).

Now, a functional model for the sigmoid (cf. Equation 3.20) is required that explains the dependency between the probability of linear structures and the image entropy in Equation 3.67. However,

it turns out to be difficult to find a heuristic for that. It can merely be supposed that a larger image entropy corresponds to a larger probability of linear structures because the Steger-algorithm is sensitive to large image gradients that may also lead to a larger image entropy. Due to the imprecise heuristics, the functional model for the sigmoid is defined on an statistical basis which leads to:

$$P(a_2 = \text{satisfied} | x_{u,2}^{\text{lin1-3}}(H_i), \theta_{\text{disc}} = \{t_1, t_2, t_3\}) \quad (3.68)$$

The sigmoid parameters  $\{t_1, t_2, t_3\}$  are determined in accordance with the approach described by Platt [2000] who assumes  $t_1 = 0$ . This basically corresponds to the assumption that the likelihoods  $\mathcal{N}(x_{u,2}^{\text{lin1-3}}(H_i) | \mu_{\text{sat}}, \sigma_{\text{sat}})$  and  $\mathcal{N}(x_{u,2}^{\text{lin1-3}}(H_i) | \mu_{\text{vio}}, \sigma_{\text{vio}})$  in Equation 3.19 have equal standard deviations (cf. Equation 3.22). Here, the same simplification of the sigmoid model is assumed. According to [Platt, 2000] the optimal parameters  $t_2$  and  $t_3$  are determined by optimizing the objective:

$$\min_{t_2, t_3} - \sum_{1 \leq q \leq n} \left( \kappa_q \log \left[ \frac{1}{1 + \exp(t_2 \cdot E(y_q^{\text{train}}) + t_3)} \right] + (1 - \kappa_q) \log \left[ 1 - \frac{1}{1 + \exp(t_2 \cdot E(y_q^{\text{train}}) + t_3)} \right] \right) \quad (3.69)$$

with

$$n = |Y_{\text{sat}}^{\text{train}}| + |Y_{\text{vio}}^{\text{train}}| \quad (3.70)$$

$$\kappa_q = \begin{cases} 1 & |y_q^{\text{train}} \in Y_{\text{sat}}^{\text{train}} \\ 0 & |y_q^{\text{train}} \in Y_{\text{vio}}^{\text{train}} \end{cases} \quad (3.71)$$

The required training datasets  $Y_{\text{sat}}^{\text{train}}, Y_{\text{vio}}^{\text{train}} \subseteq Y^{\text{train}}$  are defined without further human interactions. They are based on the training dataset  $X^{\text{roads}}$  used to train the Steger line parameters (cf. Table 3.6), but their definition requires some additional steps. First, a *training image* consisting of rectangular image patches that enclose the training dataset  $X^{\text{roads}}$  is defined. Then, the Steger-algorithm is applied to the whole training image and not only to the surroundings of the road hypotheses as before. The resulting line primitives are used to define the training dataset with  $y_p^{\text{train}} \in Y^{\text{train}}$  for all  $1 \leq p \leq |Y^{\text{train}}|$ . Figure 3.13 shows a training image and the detected line primitives by different colours. The shown training image corresponds to a 0.5 m GSD pan image that consist of  $7,000 \times 7,000$  pixels.

Given the detected line primitives  $Y^{\text{train}}$  and the training dataset  $X^{\text{roads}}$  denoting a set of pixels that corresponds to true roads, the two sets  $Y_{\text{sat}}^{\text{train}}$  and  $Y_{\text{vio}}^{\text{train}}$  are defined as follows:

$$Y_{\text{sat}}^{\text{train}} = \{y_p^{\text{train}} \in Y^{\text{train}} | (\text{reg}(y_p^{\text{train}}) \cap X^{\text{roads}}) = |\text{reg}(y_p^{\text{train}})| \wedge (\text{reg}(Y^{\text{train}} \setminus y_p^{\text{train}}) \cap R_p^{[Y]}) = 0\} \quad (3.72)$$

$$Y_{\text{vio}}^{\text{train}} = \{y_p^{\text{train}} \in Y^{\text{train}} | (\text{reg}(y_p^{\text{train}}) \cap X^{\text{roads}}) = 0 \wedge (\text{reg}(Y^{\text{train}} \setminus y_p^{\text{train}}) \cap R_p^{[Y]}) > 0\} \quad (3.73)$$

According to Equation 3.72, the set  $Y_{\text{sat}}^{\text{train}}$  includes all line primitives for which the corresponding image regions lie completely within the training dataset and where the local surroundings are free of other line primitives. According to Equation 3.73, the set  $Y_{\text{vio}}^{\text{train}}$  includes all line primitives for which the corresponding image regions lie completely outside the training dataset and where the local surroundings contain other line primitives. The expression  $R_p^{[Y]}$  denotes a set of pixels corresponding to the surrounding of a line primitive  $y_p$ , which is defined in the same way as the set  $R_i^{[H]} \in H_i$

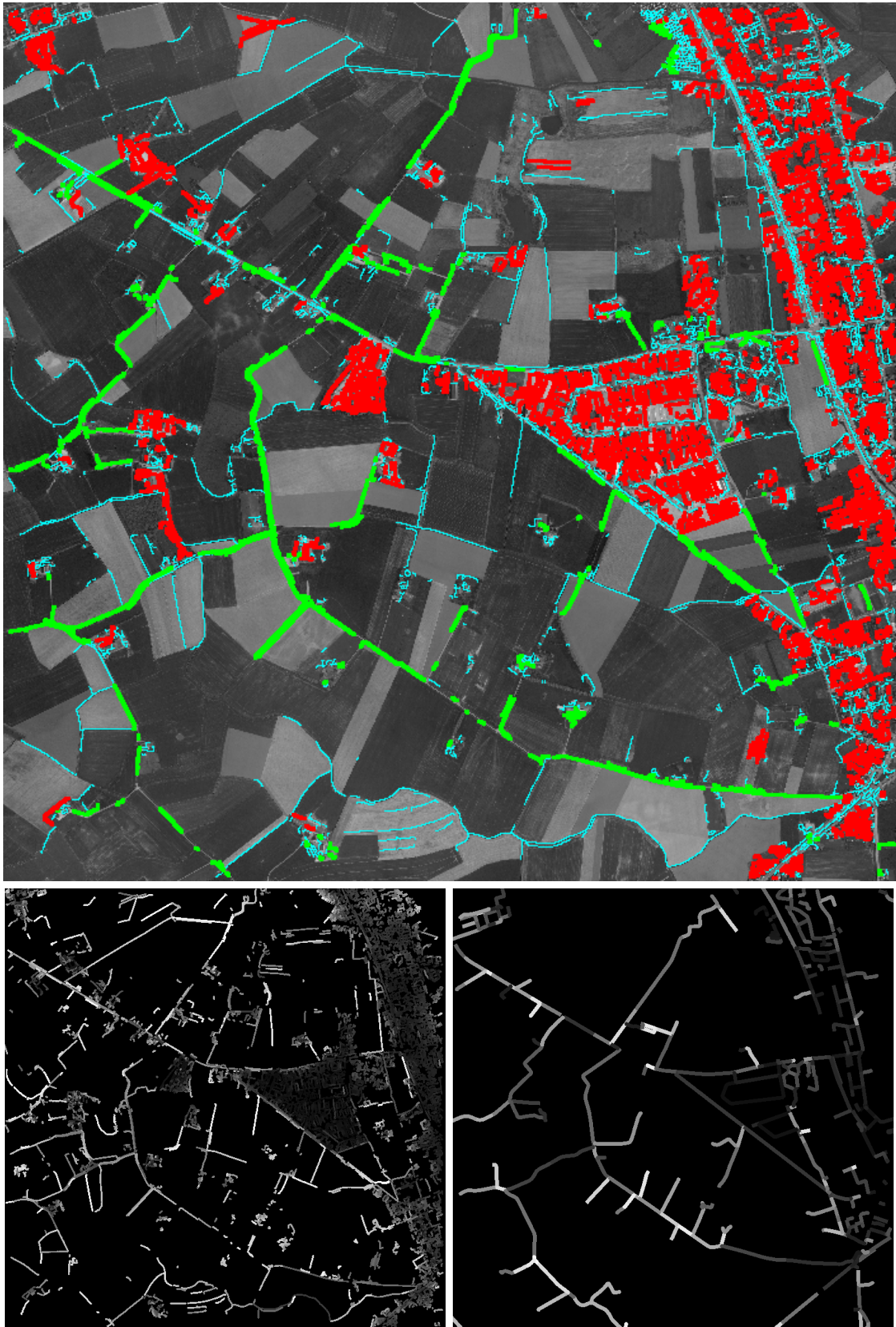


Figure 3.13.: Example for training data generation and application of a logistic regression model with  $Y_{sat}^{train}$  (green),  $Y_{vio}^{train}$  (red),  $Y^{train} \setminus (Y_{sat}^{train} \cup Y_{vio}^{train})$  (cyan) and the derived probabilities in the images displayed at the bottom (large probabilities signify bright grey values).

(cf. Table 3.4) for the centreline hypothesis  $h_i^{xy} \in H_i$ . In Figure 3.9 the sets  $Y_{sat}^{train}$  (green) and sets  $Y_{vio}^{train}$  (red) are shown together with the line primitives that are assigned to neither of two sets (cyan). The latter cannot be assigned automatically, and thus are ignored. A manual assignment is too time consuming, because the set  $Y^{train}$  usually has a large cardinality. For the training image depicted in Figure 3.13 it is  $|Y^{train}| \approx 12,000$ . Figure 3.13 further confirms the expectation that the pan-based line extraction results in many false alarms in the settlements, indicated by the high frequency of red line primitives in such areas. In Equation 3.69,  $E(y_q^{train})$  denotes the image entropy of the local surroundings of a line primitive  $y_q \in \{Y_{sat}^{train} \cup Y_{vio}^{train}\}$ , which is defined as follows:

$$E(y_q^{train}) = - \sum_{T_{low} \leq g \leq T_{high}} f_{Y^I}(g) \cdot \log f_{Y^I}(g) \quad (3.74)$$

In Equation 3.74  $f_{Y^I}(g)$  denotes the frequency of the grey values  $g$  based on an 8 bit histogram defined by the set  $R_q^{Y^I}$ . Figure 3.14 shows the frequency of extracted line primitives  $y_p$  having a certain entropy computed from their surroundings  $R_p^{Y^I}$ . The plot is superimposed with the determined sigmoid function. In the lower part of Figure 3.13, the resulting probabilities are shown, on the left side for the line primitives  $y_p^{train} \in Y^{train}$ , and on the right side for the road hypotheses  $h_j^{xy} \in H_j$ . In both images brighter grey levels correspond to higher probabilities of satisfying the second critical assumption.

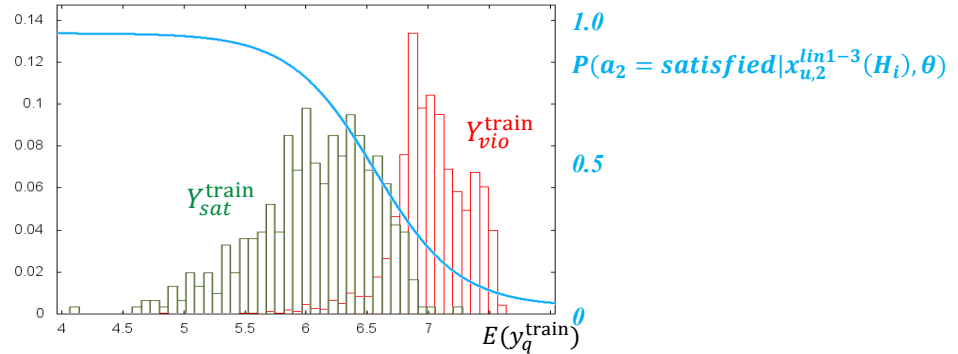


Figure 3.14.: Histograms over the image entropy (green, red) and the resulting sigmoid function.

## Discussion

The area of application for the three line detection modules is expected to be different. For instance, the pan-line detection is mainly useful for rural areas, while the nDSM-line detection is exclusively useful for dense urban areas.

The proposed strategy includes many aspects of the approach introduced by Gerke & Heipke [2008]. One difference is that not only the detection components, but the entire strategy is encapsulated, starting with the detection, via the evaluation of the hypotheses up to the definition of the model-uncertainty. While this encapsulation is important to restrict the complexity of the parametrisation, it also has disadvantages. A single road hypothesis can only be verified if it is fully in accordance with at least one of the models. As a consequence, road hypotheses that partly appear in dense urban area and partly in a park with grassland are not covered by the line-based models.

Another difference to [Gerke & Heipke, 2008] is the handling of the second critical assumption

discussed above. Gerke & Heipke [2008] overcome the problem by restricting the approach to rural context areas based on prior knowledge stored in the database. The proposed strategy evaluates the situation for each road hypothesis individually. This is advantageous if, for instance, homogeneous image regions appear within urban areas or heterogeneous image regions appear in rural areas.

Finally, it is important to note that the line detection modules only have a model for database errors that are related to relatively small divergences of *shape* or *orientation*. Other database errors related to larger divergences or the non-existence of roads are not considered, and thus leave the automatic solution in the state of *ignorance*.

### 3.3.3. Parallel edge detection

The parallel edge detection module is based on [Heipke et al., 1995], whose main objective was road detection in panchromatic images in rural and sub-urban context. Here, the approach is extended to the verification problem.

#### Model

The model basically assumes that the lateral road borders appear as parallel edge pairs in the image and that the distance of those edges from each other is approximately known. This model is designed to deal with panchromatic images with a GSD of about 0.2–0.5 m.

#### Verification strategy

The verification strategy consists of three main stages: The extraction of the parallel edges, the comparison of the extracted edges with the database and the derivation of the verification output.

**Parallel edge extraction:** The Canny edge detector [Canny, 1986] is applied to the images showing the local surroundings of the road hypothesis defined by the set  $R_i^{[H]} \in H_i$  (cf. Table 3.4). The resulting edge primitives  $y_p \in Y^{\text{canny}}$  are further analysed to determine the road border candidates  $Y^{\text{borders}} \subseteq Y^{\text{canny}}$ . An edge primitive  $y_p \in Y^{\text{canny}}$  is assigned to  $Y^{\text{borders}}$  if a pair of edge primitives  $y_p, y_q \in Y^{\text{canny}}$  exists that satisfies the following five constraints:

$$\text{reg}(y_p) \cap R_i^{H||} > 0 \quad (3.75)$$

$$\text{reg}(y_q) \cap R_i^{H||} > 0 \quad (3.76)$$

$$y_p \text{ is parallel to } y_q \quad (3.77)$$

$$\text{reg}(y'_{p,q}) \cap \text{reg}(y_q) > 0 \quad (3.78)$$

$$\left(h_i^{\text{width}} - D^{\text{width}}\right) \leq \text{dist}(y_p, y_q) \leq \left(h_i^{\text{width}} + D^{\text{width}}\right) \quad (3.79)$$

where the function  $\text{reg}$  again denotes the set of pixels intersected by its argument (cf. Figure 3.10),  $y'_{p,q}$  denotes the orthogonal projection from  $y_p$  on  $y_q$ , and the function  $\text{dist}$  provides the minimal Euclidean distance for two arguments. The interval in the constraint 3.79 considers the road width hypothesis  $h_i^{\text{width}} \in H_i$  but also the tolerated discrepancy of this attribute defined for a particular database (cf. Table 3.1). Figure 3.15 shows a possible solution for  $Y^{\text{borders}}$  and a road centreline



hypothesis  $h_i^{xy} \in H_i$ , respectively. The line primitives  $y_1, y_2, y_3, y_4$  satisfy constraints 3.75–3.76 because they lie at least partly within the shaded image region representing the set  $R_i^{H||} \in H_i$ , but only the line primitives  $y_1, y_2$  and  $y_3$  also fulfil the constraints 3.77–3.79. The red line segment  $y'_{1,2}$  in Figure 3.15 shows the orthogonal projection of  $y_1$  on  $y_2$ .

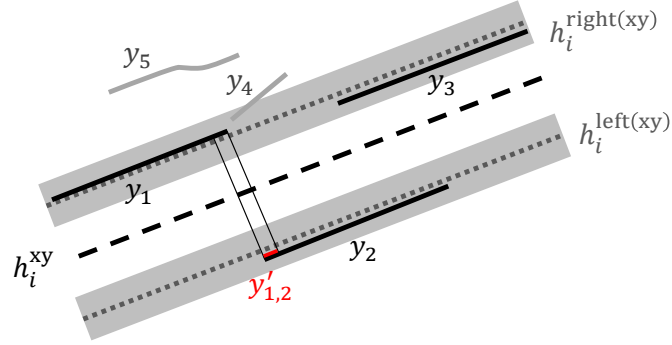


Figure 3.15.: Determination of the road border candidates.

**Comparison of the road border candidates with the database hypothesis:** In order to evaluate the road hypothesis on the basis of the border candidates  $Y^{\text{borders}}$ , hypotheses for the road borders  $h_i^{\text{left}(xy)}$  and  $h_i^{\text{right}(xy)}$  are defined by shifting  $h_i^{xy} \in H_i$  perpendicular to its main direction by half the road width ( $\pm 0.5 \cdot h_i^{\text{width}}$ ). Furthermore, the sets  $Y^{\text{left}(borders)}, Y^{\text{right}(borders)} \in Y^{\text{borders}}$  are defined for each road border hypothesis:

$$Y^{\text{left}(borders)} = \{y_p \in Y^{\text{borders}} \mid \text{dist}(y_p, h_i^{\text{left}(xy)}) \leq \text{dist}(y_p, h_i^{\text{right}(xy)})\} \quad (3.80)$$

$$Y^{\text{right}(borders)} = \{y_p \in Y^{\text{borders}} \mid \text{dist}(y_p, h_i^{\text{left}(xy)}) > \text{dist}(y_p, h_i^{\text{right}(xy)})\} \quad (3.81)$$

where  $\text{dist}$  still provides the minimum Euclidean distance between two primitives. Similarly to the verification strategy for the line detection modules, the road border hypotheses are compared with the detected edge primitives using the approach described in Section 3.3.2. Here, the comparison is applied for each road border hypothesis separately, so that two probability distributions  $P(\text{left border} = \text{correct} \mid Y^{\text{left}(borders)})$  and  $P(\text{right border} = \text{correct} \mid Y^{\text{right}(borders)})$  are defined.

**Verification output:** Finally, the two probability distributions are combined to a joint distribution representing the verification output. This is realized as the mean of both distributions, which results in:

$$\begin{aligned} P(\text{road} = \text{correct} \mid x_v^{\text{edg}}(H_i), \text{model} = a.) &= 0.5 \cdot P(\text{left border} = \text{correct} \mid Y^{\text{left}(borders)}) \\ &\quad + 0.5 \cdot P(\text{right border} = \text{correct} \mid Y^{\text{right}(borders)}) \end{aligned} \quad (3.82)$$

## Model-uncertainty

As discussed in the review of the related work in Section 1.3.1, parallel edge detection methods often provide results of relatively low completeness. It is a well known problem of the underlying model that quite often the road sides are not observable because of occlusions by trees, bushes, parked vehicles or due to shadow effects. The advantage of parallel edge detection methods over other approaches is that their results usually show low false alarm rates. Even though the model is discriminative to most background objects that appear in remote sensing images, under specific



circumstances, e.g. in the presence of elongated buildings, false alarms possibly occur. Consequently, two critical assumptions are identified.

**Critical assumption 1: The road borders can always be observed in the image.** If the road borders cannot be detected completely, the verification output relies on incomplete data. Therefore, the coverage of the road centreline hypothesis  $h_i^{\text{xy}} \in H_i$  is considered in a very similar way as with the line detection modules (cf. Equation 3.64). For the parallel edge detection module, this means that the road border candidates  $y_p \in Y^{\text{borders}}$  are projected on the centreline hypothesis  $h_i^{\text{xy}}$ . The projections on  $h_i^{\text{xy}}$  define the projected length  $h_i^{\text{length}'}$ . This also means that both sets  $Y^{\text{right(borders)}}$  and  $Y^{\text{left(borders)}}$  are considered simultaneously. Hence, there is no difference whether two road sides or only one road side have been observed. With respect to Figure 3.15 this means that the edge primitives  $y_1, y_2, y_3$  cover nearly the whole road centreline. The value for  $h_i^{\text{length}'}$  and the length of the hypothesis  $h_i^{\text{length}} \in H_i$  define the indicator feature for the critical assumption:

$$x_{u,1}^{\text{edg}}(H_i) = \frac{h_i^{\text{length}'}}{h_i^{\text{length}}} \quad (3.83)$$

The current problem is rather similar to the problem discussed concerning the first critical assumption made for the line detection modules in Section 3.3.2, so that the same rigid parametrisation approach for the functional relation is chosen here, which leads to:

$$P(a_1 = \textit{satisfied} | x_{u,1}^{\text{edg}}(H_i), \theta_{\text{rigid}} = \{0.8\}) \quad (3.84)$$

**Critical assumption 2: Only roads correspond to the parallel edge model.** As discussed above, the parallel edge model is discriminative to more background objects than the line models. Hence, in general this assumption is less critical compared to the second critical assumption identified for the line detection modules. The assumption is questionable only with respect to a few specific objects such as very long buildings. Therefore, the presence of these specific objects is explicitly analysed and not considered in form of the image entropy (cf. Section 3.3.2). If problematic objects appear in a scene frequently, it can be expected that some of the extracted edges contained in the set  $Y^{\text{canny}}$  correspond to such objects. Considering these edges as elements of the set  $Y^{\text{xborders}}$ , they are defined as any edge primitive  $y_p \in Y^{\text{canny}}$  for which a pair of edge primitives  $y_p, y_q \in Y^{\text{canny}}$  exists that satisfies the following five constraints:

$$\text{reg}(y_p) \cap (R_i^{[\text{H}\cdot]} \setminus R_i^{\text{H}||}) > 0 \quad (3.85)$$

$$\text{reg}(y_q) \cap (R_i^{[\text{H}\cdot]} \setminus R_i^{\text{H}||}) > 0 \quad (3.86)$$

$$y_p \text{ is parallel to } y_q \quad (3.87)$$

$$\text{reg}(y'_{p,q}) \cap \text{reg}(y_q) > 0 \quad (3.88)$$

$$(h_i^{\text{width}} - D^{\text{width}}) \leq \text{dist}(y_p, y_q) \leq (h_i^{\text{width}} + D^{\text{width}}) \quad (3.89)$$

The only difference to the constraints 3.75–3.79 is the mirrored search space. In the situation depicted in Figure 3.15, only the one edge primitive  $y_5$  fulfils the new constraints 3.85–3.89. The

presence of edge primitives  $y_p \in Y^{\text{xborders}}$  is quantified by length  $h_i^{\text{xlength}'}$  that is defined as the length of all projections  $y_p \in Y^{\text{xborders}}$  to the centreline hypothesis  $h_i^{\text{xy}} \in H_i$ . The indicator feature for the second critical assumption is defined as follows:

$$x_{u,2}^{\text{edg}}(H_i) = \frac{x_i^{\text{xlength}'}}{x_i^{\text{length}'}} \quad (3.90)$$

If  $h^{\text{xlength}'}$  is very small compared to  $h_i^{\text{length}'}$ , it is expected that this assumption is *satisfied*. In this case,  $x_{u,2}^{\text{edg}}(H_i)$  is near zero. In contrast, if  $h^{\text{xlength}'}$  is very large, i.e. if  $h^{\text{xlength}'} \approx h_i^{\text{length}'}$ , the critical assumption is *violated*. Furthermore, it is expected that the probability of violating the assumption increases smoothly with increasing values  $x_{u,2}^{\text{edg}}(H_i)$ , and thus the sigmoid function is defined as follows:

$$P(a_2 = \text{satisfied} | x_{u,2}^{\text{edg}}(H_i), \theta_{\text{smooth}} = \{0.0, 1.0\}) \quad (3.91)$$

## Discussion

The parallel edge detection module connects a part of the road detection strategy described in [Heipke et al., 1995], with the verification component introduced by Gerke & Heipke [2008]. The detected edge primitives are compared to road border hypotheses the same way as the detected line primitives are compared to the centreline hypotheses. In contrast to the line detection modules, the parallel edge detection module is expected to deal much better with heterogeneous context areas. A weak point of the model is the road borders being partially occluded often, which usually leads to incomplete results. Other authors such as Baumgartner et al. [1999] try to compensate this problem by integrating other model properties for certain road parts, e.g. line primitives or texture. However, the basic strategy followed in this thesis prevents such a low level combination.

### 3.3.4. The SSH method

SSH stands for Sum of Similarities of Histograms. The SSH module is mainly based on the work presented by Fujimura et al. [2008] who developed an approach to geometrically correct road databases. Fujimura et al. [2008] assumed the errors to be caused by parallel displacements with a maximum discrepancy of about 30 m.

## Model

The SSH model assumes that image regions that correspond to roads have different texture properties compared to their local surroundings. The texture properties are object-based, i.e. they refer to image regions defined on the basis of *widths* and *lengths* of those road objects. The model is designed to deal with panchromatic imagery with a GSD of about 0.2–2.5 m.

## Verification strategy

The SSH computation is similar to [Fujimura et al., 2008], while the object definition and the output are adapted to the verification problem.

**Parallel region definition:** Starting from the set  $R_i^{H^\bullet} \in H_i$  that represents the possible positions of the road hypothesis in the image (cf. Table 3.4), a number of additional image regions are defined by shifting the set  $R_i^{H^\bullet}$  in both directions perpendicular to the centreline  $h_i^{xy} \in H_i$ . The additional image regions are assumed to represent the local context, if the road hypothesis is correct. The range of shifts is defined by  $h_i^{\text{width}} + 2r$ , which equals the width of the set  $R^{H^\bullet}$  (cf. Equation 3.2). The number of all considered image regions  $S$  depends on the width  $2r'$  of the image subset  $R_i^{[H]} \in H_i$  (cf. Table 3.2):

$$S = \text{int} \left( \frac{2r'}{h_i^{\text{width}} + 2r} \right) + 1 \quad (3.92)$$

In Equation 3.92, the function  $\text{int}$  defines the integer part of a real argument. The  $S$  non-overlapping image regions all have similar sizes and shapes; see yellow boxes in Figure 3.16 for an example. Each image region is denoted by an index  $s$  with  $1 \leq s \leq S$ . The index follows an ascending order from left<sup>7</sup> to right so that  $s = 1$  denotes the leftmost region,  $s = 0.5(S + 1)$  the road hypothesis and  $s = S$  the rightmost region.

**The SSH:** Based on histograms  $f_s(g)$  derived for each region  $s$ , the sum of similarities of histograms  $SSH(s)$  is defined:

$$SSH(s) = \sum_{1 \leq t \leq S} (1 - \delta_{s,t}) BC(s, t) \quad \forall 1 \leq s \leq S \quad (3.93)$$

$$\text{with} \quad BC(s, t) = \sum_{1 \leq g \leq 256} \sqrt{f_s(g) f_t(g)} \quad (3.94)$$

where  $BC(s, t)$  denotes the Bhattacharyya distance [Bhattacharyya, 1943] of the two histograms  $s$  and  $t$ , and  $\delta$  is the Dirac function. The lower part of Figure 3.16 shows the  $SSH(s)$  for the examples in the upper row as a function of the index  $s$ . The model says that the minimum  $SSH(s)$  indicates an image region that contains a road (cf. green arrows in Figure 3.16).

**The verification output:** The expectation is that the histogram corresponding to a roads is mostly dissimilar from the rest. Hence, the centre region with index  $s = 0.5(S + 1)$  must be represented by the minimum  $SSH(s)$  if the road hypothesis is correct. This circumstance is used to define the verification output:

$$P(\text{road} = \text{correct} | x_v^{\text{SSH}}(H_i), \text{model} = a.) = \begin{cases} 1.0 & | 0.5(S + 1) = s_{\min} \\ 0.0 & | \text{otherwise} \end{cases} \quad (3.95)$$

with

$$s_{\min} = \arg \min_s SSH(s) \quad (3.96)$$

The definition in Equation 3.95 allows the verification output to be either zero or one.

<sup>7</sup>Here, the definitions of *left* and *right* depend on the orientation of the centreline  $h_i^{xy} \in H_i$ .

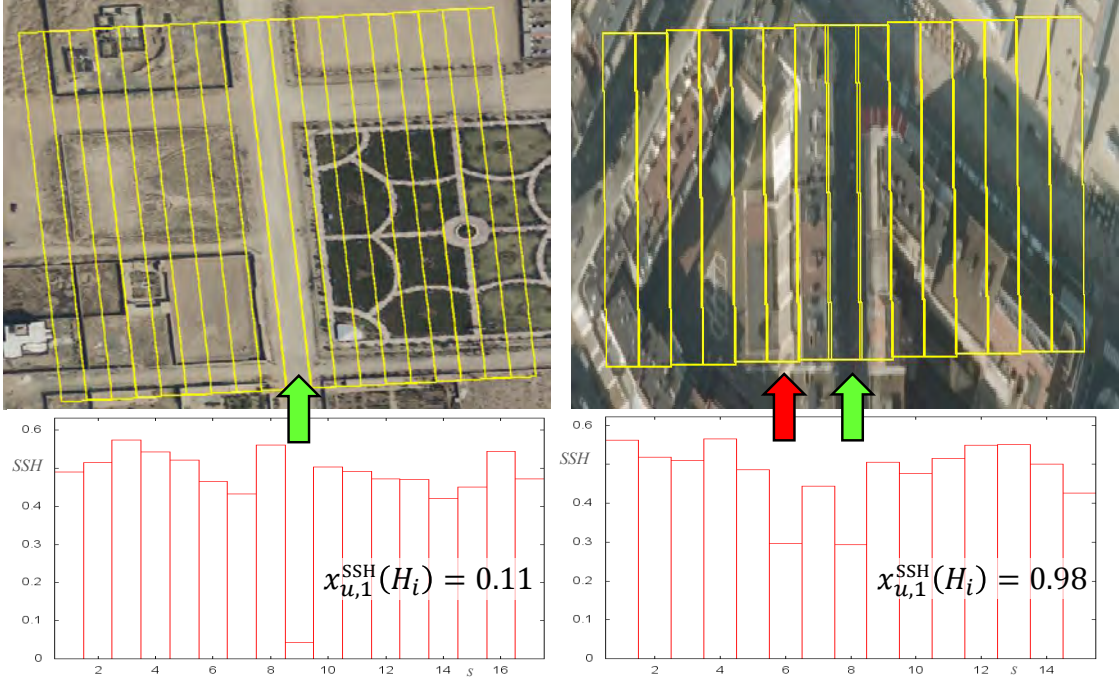


Figure 3.16.: Demonstration of the SSH method. Left: A situation that fits well to the underlying model. Right: A situation that is not covered by the SSH-model. The green arrows indicate the position of the road. The red arrow indicates a second minimum related to a bright façade.

### Model-uncertainty

The underlying model implicitly assumes that all context regions (all regions except the road region) are at least similar to some other context regions. If the surrounding area is heterogeneous in orthogonal direction to the road hypothesis, this assumption is *violated*. In such cases, the minimum SSH might denote some context region. The assumption is also *violated* if other elongated objects, e.g. a row of trees or water canals, appear in a direction parallel to the road hypothesis. A further critical assumption of the underlying model is that a database hypothesis can either be *correct* or just affected by a parallel displacement. This assumption is realistic for the scenario described in [Fujimura et al., 2008], but not necessarily for the verification problem discussed in this thesis. For instance, if an error is caused by a wrong shape or orientation or if there is no road at all, the minimum SSH corresponds to some other object(s) than a road.

In all discussed scenarios, violating a critical assumption is expected to lead to a non-significant minimum, either because all regions are nearly equally dissimilar or because a second region is also dissimilar to the rest. Therefore, the discussed critical assumptions are substituted to one.

**Critical assumption 1: The determined minimum is significant.** The indicator feature for the significance of the determined minimum is defined as follows:

$$x_{u,1}^{SSH}(H_i) = \begin{cases} \frac{|\overline{SSH} - \min_s SSH(s)|}{\overline{SSH} - \min_{s \setminus s_{\min}} SSH(s)} & |\overline{SSH}| < \min_{s \setminus s_{\min}} SSH(s) \\ 0.0 & \text{otherwise} \end{cases} \quad (3.97)$$

In Equation 3.97  $\overline{SSH}$  defines the mean SSH of all regions. The term in the denominator denotes the difference of the mean and the second smallest SSH of all regions. When the denominator equals zero, the indicator is zero. The critical assumption is expected to be fulfilled if the following relations hold true:

$$\min_s SSH(s) \ll \min_{s \setminus s_{\min}} SSH(s) \quad (3.98)$$

$$\min_s SSH(s) \ll \overline{SSH} \quad (3.99)$$

$$\min_{s \setminus s_{\min}} SSH(s) \approx \overline{SSH} \quad (3.100)$$

If the relations 3.98–3.100 hold true,  $x_{u,1}^{SSH}(H_i)$  is close to zero, otherwise  $x_{u,1}^{SSH}(H_i)$  becomes larger, but not larger than 1.0. Furthermore, it is expected that the probability of violating the critical assumption increases smoothly with increasing values for  $x_{u,1}^{SSH}(H_i)$ . This heuristic is plugged into the functional model for the sigmoid in Equation 3.20, which leads to:

$$P(a_1 = \textit{satisfied} | x_{u,1}^{SSH}(H_i), \theta_{\text{smooth}} = \{0.0, 1.0\}) \quad (3.101)$$

Figure 3.16 shows two examples providing verification outputs with different model-uncertainties. The situation on the left fulfils the critical assumption with probability  $P(a_1 = \textit{satisfied} | x_{u,1}^{SSH}(H_i) = 0.11, \theta) = 0.99$ . The situation on the right violates the critical assumption because another image region showing a bright façade also results in a small SSH. The situation leads to a low probability, with  $P(a_1 = \textit{satisfied} | x_{u,1}^{SSH}(H_i) = 0.98, \theta) = 0.0005$ .

## Discussion

The strength of the SSH module compared to the other modules is its ability to deal with low image resolutions of about 2.5 m GSD. This specific property is related to the object-based analysis strategy that integrates the texture properties over the whole length of a road hypothesis that is usually a large area with a sufficient number of pixels. Another strength of the SSH module is that it can deal with low contrast conditions in the imagery because the differences between the histograms integrate over the number of considered regions. Thus, tiny differences between the road-related histogram and context-related histograms become relevant even if the differences between context regions are larger. In the left part of Figure 3.16, it can be seen that the model is able to deal well with heterogeneous context areas if the surroundings do not change completely for each considered region. Fujimura et al. [2008] have also shown that the surroundings can be completely different on both sides of the road.

One difference of the proposed strategy to the original approach is related to the image region definition. Fujimura et al. [2008] defined the width of the image regions to be one pixel, which corresponds to regions of 2.5 m width in their application. The proposed definition with the set  $R_i^{H^\bullet} \in H_i$  usually results in much larger regions. This difference is related to the difference between the tasks. Fujimura et al. [2008] searches for the actual displacement. The advantage of using  $R_i^{H^\bullet}$  for the verification problem is that the central image region is covering the road region in any case, whereas all shifted regions represent only the surroundings, if the database hypothesis is *correct*.

### 3.3.5. Acupuncture method

The acupuncture module is mainly based on the work of Youn et al. [2008] who developed a road detection method that focuses on urban and sub-urban areas.

#### Model

The underlying model assumes that an image region which corresponds to a road shows a lower amount of edge structures orthogonal to the road centreline than the local surroundings, where such structures appear more frequently, mainly due to buildings. The model relies on panchromatic imagery with a GSD of 0.2–1.0 m.

#### Verification strategy

**Edge image:** The Canny edge-extractor [Canny, 1986] is applied to the image subset defined by  $R_i^{[H]} \in H_i$  (cf. Table 3.4).

**Acupuncture nails:** What Youn et al. [2008] called *acupuncture nails* are basically groups of lines with one pixel spacing across the whole image. In [Youn et al., 2008], these lines are defined in accordance with the two main road directions that are determined by another method. In this thesis, the *acupuncture nails* are defined on the basis of the centreline hypothesis  $h_i^{xy} \in H_i$ , where  $h_i^{xy}$  directly defines the first *acupuncture nail*. All additional *nails* are defined by shifting  $h_i^{xy}$  orthogonal to the main direction of  $h_i^{xy}$ . The distance of the shifts is always one pixel, independent of the image resolution. Depending on the width of the considered context region  $R_i^{[H]}$  (parameter  $r_2$ , Table 3.5) and the given image resolution (parameter GSD, Table 3.2), the number of *acupuncture nails* is defined as:

$$U = \text{int} \left( \frac{2r'}{\text{GSD}} \right) + 1 \quad (3.102)$$

where  $\text{int}$  again defines the integer part of its argument. In the subsequent discussions, each *acupuncture nail* will be denoted by an index  $u$  with  $1 \leq u \leq U$ . The index has ascending order from left to right so that  $u = 1$  corresponds to the leftmost *nail*,  $u = 0.5(U + 1)$  to the centreline hypothesis  $h_i^{xy}$  and  $u = U$  to the rightmost *nail*. This notation is rather similar to the one used for the SSH module. The difference is that many indices  $u$  potentially represent the road hypothesis because the distance between the *acupuncture nails* is only one pixel. Considering the required accuracy  $D^{xy}$  of the database (cf. Table 3.1), a tolerance interval for the road hypothesis can be defined as follows:

$$0.5(U + 1) - \text{int} \left( \frac{D^{xy}}{\text{GSD}} \right) \leq u \leq 0.5(U + 1) + \text{int} \left( \frac{D^{xy}}{\text{GSD}} \right) \quad (3.103)$$

Figure 3.17a displays an example for an edge image, whereas Figure 3.17b shows the superimposition of the edge image with the *acupuncture nails* defined on the basis of a correct road hypothesis.

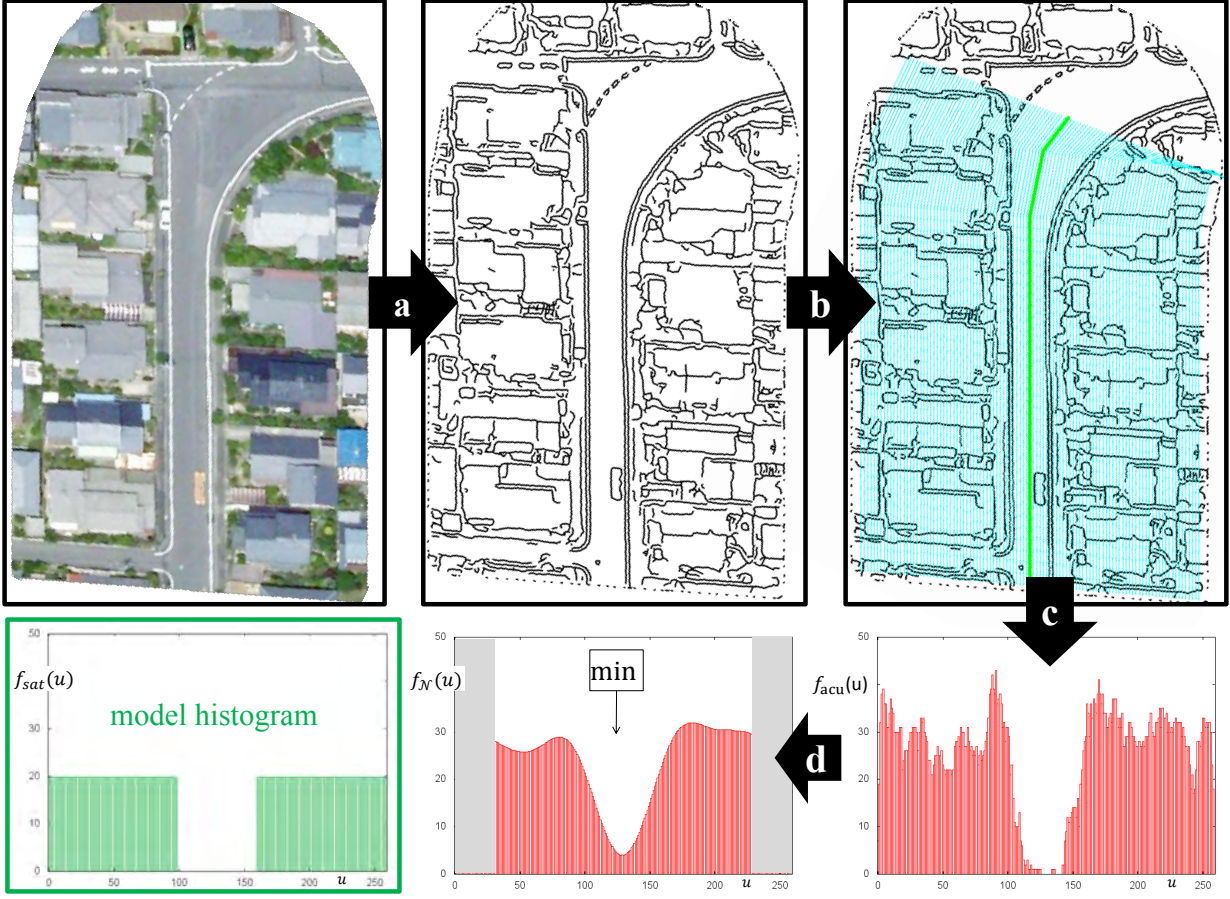


Figure 3.17.: Demonstration of the acupuncture method in a situation that fits well to its underlying model because the road appears homogeneous and local context is characterized by several detached houses. The extracted edges are drawn in black; the *acupuncture nail* with the minimum number of intersections is drawn in green while the other *nails* are drawn in cyan. The meaning of the plots in the second row is described in the text.

**Histogram of intersections and Gaussian filtering:** A histogram is defined that counts the intersections  $f_{acu}(u)$  of each *acupuncture nail*  $1 \leq u \leq U$  with the extracted edges. An example for this is shown in Figure 3.17c. In order to enhance spatial accuracy of the original method, the resulting histogram is filtered by a Gaussian kernel, whose kernel width is defined in accordance with the road width hypothesis  $h_i^{width} \in H_i$  (cf. Table 3.4):

$$f_N(u) = f_{acu}(u) * \mathcal{N}(u | \sigma = h_i^{width}) \quad (3.104)$$

An example for the smoothed histogram  $f_N(u)$  is shown in Figure 3.17d. Due to the Gaussian filtering, the centreline of a road appears as a clear minimum even if more than one *acupuncture nail* representing the road surface has zero intersections with the detected edges. In Figure 3.17 the *acupuncture nail* with a minimum number of intersections is highlighted by a bold green line.

**Verification output:** In accordance with Equation 3.103, a road hypothesis is *correct* if the centreline hypothesis  $h_i^{xy}$  corresponds to the acupuncture nail with minimum  $f_N(u)$  by taking into account the accuracy  $D^{xy}$  required for the database. This binary decision is used to define the verification

output, which is also binary:

$$P(\text{road} = \text{correct} | x_v^{\text{acu}}(H_i), \text{model} = a.) = \begin{cases} 1.0 & |u_{\text{hyp}} - \text{int}(\frac{D^{\text{xy}}}{GSD}) \leq u_{\text{min}} \leq u_{\text{hyp}} + \text{int}(\frac{D^{\text{xy}}}{GSD}) \\ 0.0 & \text{otherwise} \end{cases}$$

(3.105)

with

$$u_{\text{hyp}} = 0.5(U + 1) \quad (3.106)$$

$$u_{\text{min}} = \arg \min_u \{f_{\mathcal{N}}(u)\} \quad (3.107)$$

## Model-uncertainty

The model assumes recurring regular structures in the vicinity of roads, which is mainly true for urban areas with a lot of small houses that are in the focus of Youn et al. [2008]. With respect to the verification problem discussed in this thesis, this is a critical assumption because homogeneous context areas such as grassland, crop fields or huge industrial buildings have to be taken into account. For such situations, the module is not unlikely to provide erroneous outputs. A second critical assumption is that fewer structures appear in road regions, which might be *violated* in the presence of too many cars, road markings, trees or shadow effects. Furthermore, similar to the SSH module, database errors are assumed to correspond only to parallel shifts. However, database errors related to wrong orientations or shapes probably occur, and thus this assumption might be *violated*.

If at least one of the three identified critical assumptions is *violated*, the histogram  $f_{\text{acu}}(u)$  differs from a histogram that would be expected by the underlying model. Consequently, those critical assumptions can be substituted by only one.

**Critical assumption 1: The actual histogram corresponds to the expectations of the model.** Consider a situation that perfectly *satisfies* the assumptions of the underlying model, represented by histogram  $f_{\text{sat}}(u)$ , where the frequency of intersections is zero if an *acupuncture nail* corresponds to a road region and is large for the context area. In order to determine how many intersections should be expected in the context area, a synthetic situation is defined, which assumes the presence of detached houses to occur every ten meters on both sides of the road. Each house is further expected to have the same orientation as the centreline hypothesis  $h_i^{\text{xy}}$  and to be represented by only two edges that intersect the *acupuncture nails* (cf. Figure 3.18). Hence, the expected number of intersections  $o$  corresponding to the synthetic situation is:

$$o = \frac{h_i^{\text{length}}}{5.0 \text{ m}} \quad (3.108)$$

An example for such a model histogram is shown in the lower left corner in Figure 3.17. The frequency  $f_{\text{sat}}(u)$  of the model histogram depends on the road hypothesis because the underlying model depends on the road width hypothesis. Furthermore,  $f_{\text{sat}}(u)$  depends on the position of the minimum. Based on these thoughts, firstly the domain of the histogram representing the expected road has to be defined. Let the index  $u_{\text{left}}$  represent the leftmost and  $u_{\text{right}}$  the rightmost *acupuncture nails* that are expected to intersect only the road surface by considering  $h_i^{\text{width}} \in H_i$  (cf. Table 3.4)



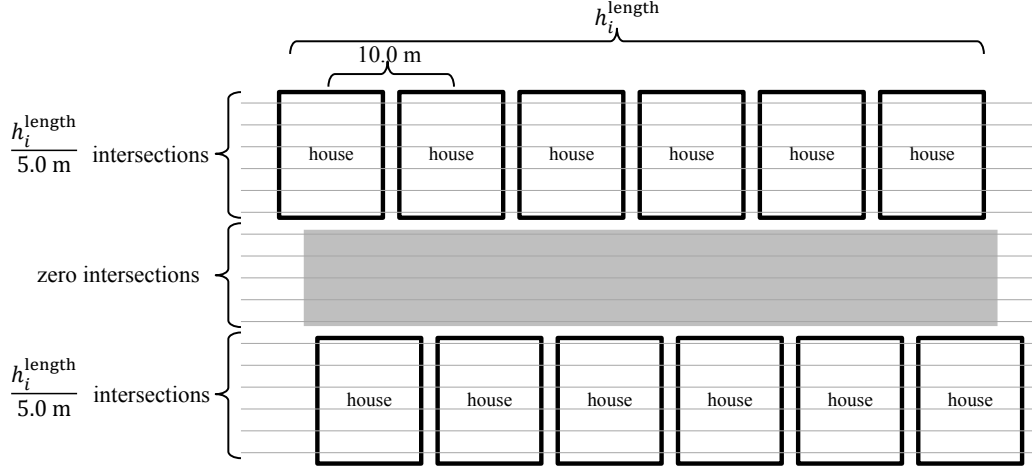


Figure 3.18.: Synthetic model for the number of intersections.

and  $u_{\min}$  (cf. Equation 3.107):

$$u_{\text{left}} = u_{\min} - \text{int} \left( \frac{0.5 \cdot h^{\text{width}}}{GSD} \right) \quad (3.109)$$

$$u_{\text{right}} = u_{\min} + \text{int} \left( \frac{0.5 \cdot h^{\text{width}}}{GSD} \right) \quad (3.110)$$

Then, a threshold operation is applied on  $f_{\text{acu}}(u)$  to avoid penalizing larger numbers of intersections:

$$f_{\text{T}}(u) = \begin{cases} f_{\text{acu}}(u) & |f_{\text{acu}}(u) \leq o \\ o & |\text{otherwise} \end{cases} \quad \forall 1 \leq u \leq U \quad (3.111)$$

Finally,  $f_{\text{T}}(u)$  and  $f_{\text{sat}}(u)$  are set in relation by considering the ratio of the areas under the densities approximated by the two histograms. This leads to the indicator feature for the critical assumption, which is defined as follows:

$$x_{u,1}^{\text{acu}}(H_i) = \left[ \frac{\sum_{1 \leq u \leq u_{\text{left}}} f_{\text{T}}(u)}{a_{\text{left}}} \right]_{\text{left}} \cdot \left[ 1 - \frac{\sum_{u_{\text{left}} < u < u_{\text{right}}} f_{\text{T}}(u)}{a_{\text{road}}} \right]_{\text{road}} \cdot \left[ \frac{\sum_{u_{\text{right}} \leq u \leq U} f_{\text{T}}(u)}{a_{\text{right}}} \right]_{\text{right}} \quad (3.112)$$

with

$$a_{\text{left}} = u_{\text{left}} \cdot o \quad (3.113)$$

$$a_{\text{road}} = (u_{\text{right}} - u_{\text{left}} - 1) \cdot o \quad (3.114)$$

$$a_{\text{right}} = (U - u_{\text{right}} + 1) \cdot o \quad (3.115)$$

In Equation 3.112, each factor considers an interval of the histograms. The first interval represents the left context area, the second interval represents the road region and the third interval represents the right context area. This is important because the context areas might have different characteristics on both road sides. As  $f_{\text{sat}}(u)$  is a constant within each interval, the areas  $a_{\text{left}}$ ,  $a_{\text{road}}$  and  $a_{\text{right}}$  can be defined as areas of rectangles, whereas the areas corresponding to  $f_{\text{T}}(u)$  require the

computation of the sums over the histogram domains.

If the road region is characterized by only a few intersections, e.g.  $f_{acu}(u) \approx 0$  and if the context areas on both sides are characterized by many intersections, e.g.  $f_T(u) \approx o$  the initially defined critical assumptions are *satisfied*. In this case  $x_{u,1}^{acu}(H_i)$  is nearly one (cf. Equation 3.112). In contrast, if the context region of at least one road side is homogeneous, e.g. because of the presence of a crop field, grassland or a large industrial building,  $x_{u,1}^{acu}(H_i)$  is near zero, which denotes a violation of the critical assumptions. The same result is reached if the road region is represented by many intersections, e.g.  $f_{acu}(u) \approx o$ . Between these two extremes the probability of satisfying the critical assumptions is expected to change smoothly, and thus the functional model for the sigmoid is:

$$P(a_1 = \textit{satisfied} | x_{u,1}^{acu}(H_i), \theta_{\text{smooth}} = \{1.0, 0.0\}) \quad (3.116)$$

According to Equation 3.116, the situation in Figure 3.17 *satisfies* the critical assumptions with high probability  $P(a_1 = \textit{satisfied} | x_{u,1}^{acu}(H_i) = 0.95, \theta) = 0.99$  because the road region is rather homogeneous and the surroundings show many detached houses. In contrast, Figures 3.19 and 3.20 show two examples where the underlying model is most likely not applicable. The situation in Figure 3.19 *satisfies* the critical assumption only with low probability  $P(a_1 = \textit{satisfied} | x_{u,1}^{acu}(H_i) = 0.09, \theta) = 0.001$  because of the comparably low response for edge detection in the local context. The situation in Figure 3.20 also *satisfies* the critical assumption with low probability  $P(a_1 = \textit{satisfied} | x_{u,1}^{acu}(H_i) = 0.28, \theta) = 0.03$  because of the comparably high response for edge detection within the road region. As can be deduced from the largely displaced bold red lines, the verification outputs for the examples depicted in Figures 3.19 and 3.20 indeed assign the wrong state.

## Discussion

The acupuncture model is especially designed to deal with sub-urban context areas that are characterized by many small detached houses. As this is a weak point of most of the other modules, it can be expected that the acupuncture module provides complementary information, and therefore is important for the overall approach.

Despite that being mentioned, the model itself is rather specific. For instance, in (partly) rural areas with homogeneous surroundings the model is *not applicable*. Failure must also be expected for dense urban areas with high buildings that cast shadows on the road surface. In this case, the shadows may lead to similar numbers of intersections with the *acupuncture nails* as the buildings themselves.

The proposed verification strategy extends the original approach described in [Youn et al., 2008] by a novel definition of the *acupuncture nails*. Youn et al. [2008] use an automatic method that extracts the two main directions in the image and hence requires a grid-like structure of the road network. As the proposed strategy uses the road hypothesis to define the orientation and the shapes of the *acupuncture nails*, the approach can also deal with arbitrary road network structures. Another extension of the original method is introduced by the Gaussian filtering (cf. Equation 3.104), which makes the method more effective for smaller GSD because it explicitly determines the centreline and not road region.

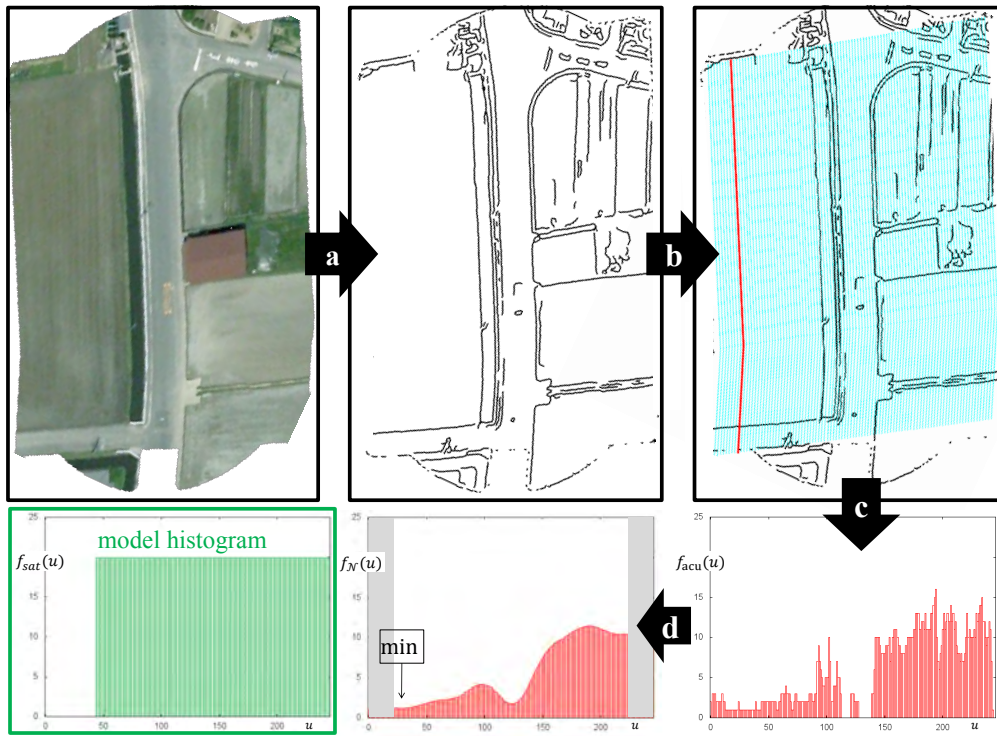


Figure 3.19.: Demonstration of the acupuncture method in a situation that does not fit its underlying model because the context region appears to be homogeneous. The meaning of the sub-figures corresponds to Figure 3.17.

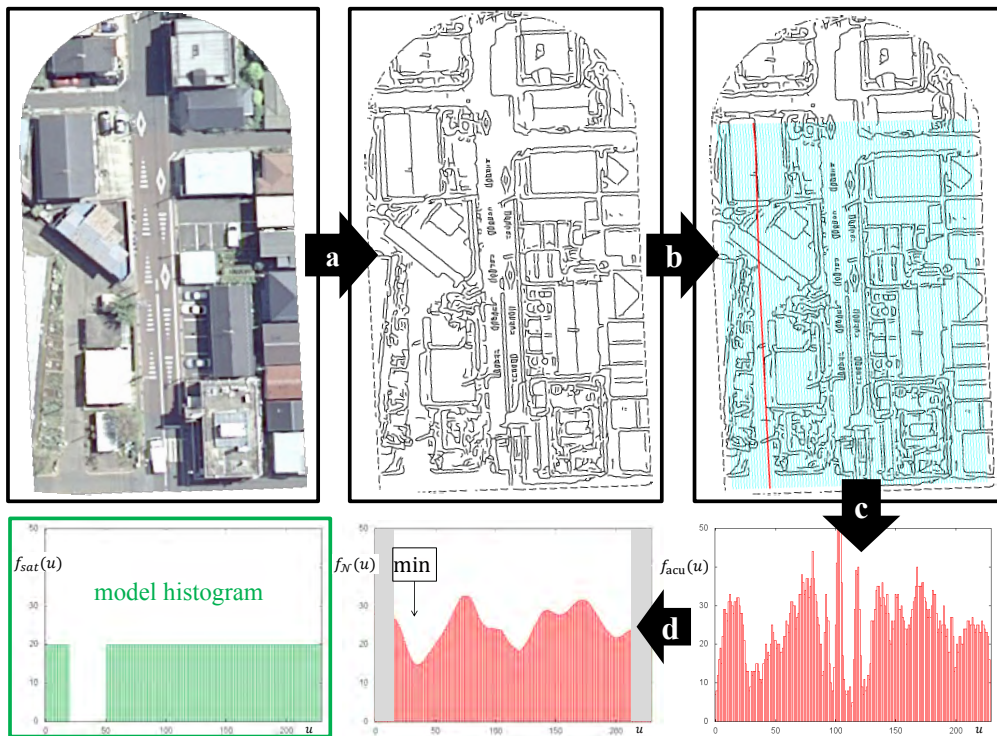


Figure 3.20.: Demonstration of the acupuncture method in a situation that does not fit its underlying model because the road region is characterized by many road markings and vehicles. The meaning of the sub-figures corresponds to Figure 3.17.

### 3.3.6. Building and grassland detection

The modules *building detection* and *grassland detection* will be discussed together as they follow very similar strategies. The difference between them mainly concerns the definition of the verification outputs, which will be described separately.

The strategy is mainly based on the work described in [Rottensteiner et al., 2007] who developed an approach to detect buildings on the basis of high resolution images and laserscanner data.

#### Models

The models basically assume that the objects *roads*, *buildings* and *grassland* are disjunct, i.e. they cannot exist in the same place in the image. For this purpose, different authors, e.g. Youn et al. [2008] and Poulain et al. [2010] explicitly considered such objects in order to define image regions where roads probably do not exist.

With respect to the verification of outdated road databases, it is further assumed that temporal changes of the road network are closely related to changes of buildings and grassland areas. Hence, database errors related to temporal changes can be explicitly detected if a road hypothesis overlaps with a new building or a new grassland object.

As the road model relies on buildings and grassland these objects have to be detected in the imagery, which requires models for buildings and grassland, respectively. For the two modules the models are defined in accordance with [Rottensteiner et al., 2007]:

**Buildings:** These topographic objects have a minimum height above the terrain that can be represented by a specific grey value in the nDSM image. Furthermore, buildings are non-vegetation areas and therefore have a relatively small response for the NDVI.

**Grassland:** The height of grassland approximately corresponds to the terrain height that can be represented by a specific grey value in the nDSM image. Furthermore, grassland objects are areas of vital vegetation and therefore have a relatively large response for the NDVI.

The modules are designed to deal with nDSM and NDVI images that correspond to a GSD  $\leq 1$  m.

#### Verification strategy

First, the imagery is classified in order to determine sets of pixels for building and grassland areas. This step is based on the work described in Rottensteiner et al. [2007]. Then, these sets are compared with the database information, and consequently the verification outputs are defined.

**Image classification:** Initially four sets of pixels are defined. The set  $R^{\text{BUT}}$  denotes elevated areas such as *buildings* or *trees*. The set  $R^{\text{GUR}}$  denotes areas close to the ground such as *grassland* or *roads*. The set  $R^{\text{GUT}}$  denotes vegetation such as *grassland* or *trees*. The set  $R^{\text{BUR}}$  denotes non-vegetation such as *buildings* or *roads*. Unlike in [Rottensteiner et al., 2007], a simple threshold is applied to define the four sets by only considering the local surroundings of the road hypothesis represented

by pixels  $p_b \in R_i^{[H]}$  (cf. Table 3.4):

$$R^{\text{BUT}} = \{p_b \in R_i^{[H]} | \text{nDSM}(p_b) \geq 4.0 \text{ m}\} \quad (3.117)$$

$$R^{\text{GUR}} = \{p_b \in R_i^{[H]} | \text{nDSM}(p_b) < 4.0 \text{ m}\} \quad (3.118)$$

$$R^{\text{GUT}} = \{p_b \in R_i^{[H]} | \text{NDVI}(p_b) \geq T_{\text{high}}(\text{NDVI})\} \quad (3.119)$$

$$R^{\text{BUR}} = \{p_b \in R_i^{[H]} | \text{NDVI}(p_b) < T_{\text{high}}(\text{NDVI})\} \quad (3.120)$$

The threshold of 4.0 m for the nDSM heights (cf. Equations 3.117 and 3.118) is based on a simple heuristic, which seeks to separate terrain with vehicles and small bushes from buildings and trees. The definition of a comparable heuristic for the NDVI is difficult because the exact calibration of the underlying colour bands is often unknown. Therefore, the threshold  $T_{\text{high}}(\text{NDVI})$  is defined based on training data. For that purpose, the approach for determining the Steger-line parameters described in [Ziems et al., 2007] is applied in a similar fashion to the NDVI-line detection module (cf. Section 3.3.2). Hence, the required value  $T_{\text{high}}(\text{NDVI})$  is equal to the value of  $T_{\text{high}}(\text{NDVI})$  in Table 3.3.

The set  $R^{\text{B}}$  representing buildings and the set  $R^{\text{G}}$  representing grassland are defined as intersections of the sets defined in Equations 3.117–3.120:

$$R^{\text{B}} = R^{\text{BUT}} \cap R^{\text{BUR}} \quad (3.121)$$

$$R^{\text{G}} = R^{\text{GUR}} \cap R^{\text{GUT}} \quad (3.122)$$

**Verification output for building detection:** In accordance with the initially defined model, a road hypothesis is *incorrect* if a building (partly) occurs in the image region  $R_i^{\text{H}\bullet} \in H_i$  (cf. Table 3.4). Consequently, the set  $\{R^{\text{B}} \cap R_i^{\text{H}\bullet}\}$  represents the contradiction set between the database hypothesis and the image classification output. The cardinality of this contradiction set is used to define a binary verification output:

$$P(\text{road} = \text{correct} | x_v^{\text{blg}}(H_i), \text{model} = a.) = \begin{cases} 0.0 & |R^{\text{B}} \cap R_i^{\text{H}\bullet}| > 0 \\ 1.0 & \text{otherwise} \end{cases} \quad (3.123)$$

Figure 3.21 shows a situation where a small built-up area has been constructed. The outdated road database information is shown in the left part of Figure 3.21. The right part shows the nDSM image superimposed by the detected building segments  $R^{\text{B}}$  (cyan outlines), the determined contradiction set  $\{R^{\text{B}} \cap R_i^{\text{H}\bullet}\}$  (red) and the non-contradiction set  $\{R_i^{\text{H}\bullet} \setminus R^{\text{B}}\}$  in green.

**Verification output for grassland detection:** While buildings and roads are absolutely disjunct, grassland objects are allowed to appear on road surfaces if their area is below the minimum mapping unit defined for the respective database. This is related to cartographic generalization which may allow for small traffic islands to be interpreted as parts of the road. In this thesis, the minimum mapping unit is not directly considered. Instead, the squared value for the required spatial accuracy  $D^{\text{xy}}$  (cf. Table 3.1) is introduced as an approximation for the minimum mapping unit. Hence, the definition of the verification output for the grassland detection module differs from the one for building detection (cf. Equation 3.123) in two points: Firstly, in the definition of the contradiction



Figure 3.21.: Demonstration of building detection module for an outdated road database. Contradictions (red) with the database information (yellow) and the detected building segments (bright regions with cyan outlines). Left: RGB image. Right: nDSM image.

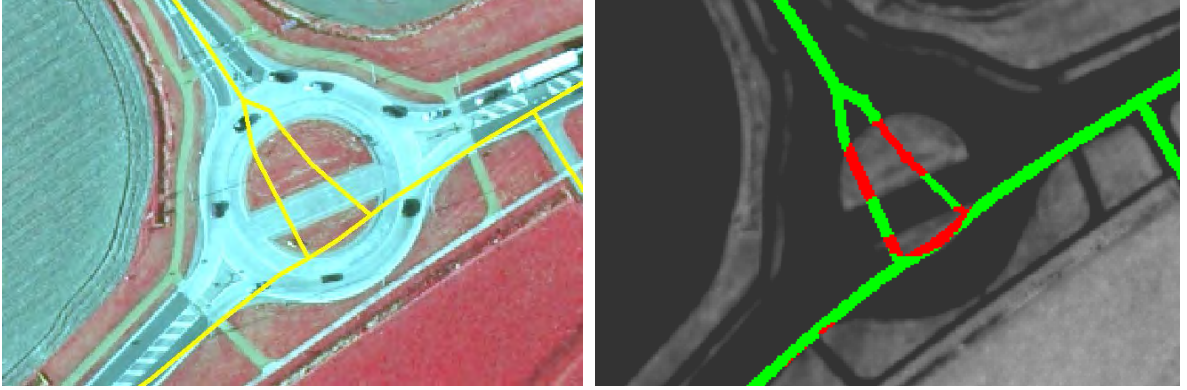


Figure 3.22.: Demonstration of the grassland detection module. Contradictions (red) with the database information (yellow). Left: IRRG image. Right: NDVI image.

set  $\{R^G \cap R_i^{H\bullet}\}$ , and secondly in its tolerated cardinality:

$$P(\text{road} = \text{correct} | x_v^{\text{grs}}(H_i), \text{model} = a.) = \begin{cases} 0.0 & |R^G \cap R_i^{H\bullet}| > (D^{\text{xy}})^2 \\ 1.0 & \text{otherwise} \end{cases} \quad (3.124)$$

Figure 3.22 shows a situation where a roundabout has recently been constructed. The outdated road database is depicted in the left part of Figure 3.22 for an IRRG image. The right part shows the NDVI image superimposed by the contradiction set  $\{R^G \cap R_i^{H\bullet}\}$  (red) and the non-contradiction set  $\{R_i^{H\bullet} \setminus R^G\}$  in green.

### Model-uncertainty

The definitions of the verification outputs in Equations 3.123 and 3.124 have two considerable consequences. Firstly, a road hypothesis is assigned to the state *incorrect* with a probability of 1.0 if at least one single pixel  $p_b \in \{R^B \cap R_i^{H\bullet}\}$  (building detection) or a few pixels  $p_b \in \{R^G \cap R_i^{H\bullet}\}$  (grassland detection) exist. This implies the assumption that the image classification is correct with a probability of 1.0. This is critical because the NDVI values might be inaccurate, or the nDSM might be geometrically inaccurate. Another consequence of the definitions in Equations 3.123 and 3.124

is that any road hypothesis for which the contradiction set is empty (building detection) or has small cardinality (grassland detection) is assigned to the state *correct* with a probability of 1.0. In this case, not only the correctness of the image classification is a critical assumption, but also the expectation that any database error can be explained by a redevelopment of buildings or grassland areas.

**Critical assumption 1: The vegetation threshold is clearly discriminative.** Rottensteiner et al. [2007] pointed out that due to a disadvantageous error propagation, the NDVI is affected by high standard deviations in weakly illuminated areas. If the standard deviation of the NDVI is high, the assumption that the decision threshold (cf. Equations 3.119 and 3.120) allows a clear separation is questionable. Therefore, the standard deviation of the NDVI defined on the basis of the Gaussian law of error propagation is considered as the indicator feature:

$$x_{u,1}^{\text{blg,grs}}(H_i) = \frac{2 \cdot \sqrt{g_{\text{red}}^2 (\sigma_{\text{ir}}^{\text{grey}})^2 + g_{\text{ir}}^2 (\sigma_{\text{red}}^{\text{grey}})^2}}{(g_{\text{ir}} + g_{\text{red}})^2} \quad (3.125)$$

with

$$g_{\text{ir}} = \begin{cases} \frac{1}{|R^{\text{B}} \cap R_i^{\text{H}\bullet}|} \cdot \sum_{p_b \in \{R^{\text{B}} \cap R_i^{\text{H}\bullet}\}} \text{ir}(p_b) & | \text{ building detection} \\ \frac{1}{|R^{\text{G}} \cap R_i^{\text{H}\bullet}|} \cdot \sum_{p_b \in \{R^{\text{G}} \cap R_i^{\text{H}\bullet}\}} \text{ir}(p_b) & | \text{ grassland detection} \end{cases} \quad (3.126)$$

$$g_{\text{red}} = \begin{cases} \frac{1}{|R^{\text{B}} \cap R_i^{\text{H}\bullet}|} \cdot \sum_{p_b \in \{R^{\text{B}} \cap R_i^{\text{H}\bullet}\}} \text{red}(p_b) & | \text{ building detection} \\ \frac{1}{|R^{\text{G}} \cap R_i^{\text{H}\bullet}|} \cdot \sum_{p_b \in \{R^{\text{G}} \cap R_i^{\text{H}\bullet}\}} \text{red}(p_b) & | \text{ grassland detection} \end{cases} \quad (3.127)$$

where  $g_{\text{ir}}$  and  $g_{\text{red}}$  denote the means of the grey values of the red and infrared bands in the contradiction sets. The standard deviations  $\sigma_{\text{ir}}^{\text{grey}}$  and  $\sigma_{\text{red}}^{\text{grey}}$  have to be defined by a human operator, and thus are *system parameters* of the proposed approach (cf. Table 3.7).

$\sigma_{\text{ir}}^{\text{grey}}$	[bit]	standard deviation of grey values of the infrared band
$\sigma_{\text{red}}^{\text{grey}}$	[bit]	standard deviation of grey values of the red band

Table 3.7.: System parameter: Standard deviations of grey-values.

Rottensteiner et al. [2007] found a heuristic, whereupon a standard deviation of 25% of the NDVI is critical, while a standard deviation near zero is not critical at all. Based on these two values, they describe the model-uncertainty by a smooth transition between these two values, which approximately corresponds to a sigmoid function. In close accordance to their definition, the probability of satisfying the critical assumption is defined as follows:

$$P(a_1 = \text{satisfied} | x_{u,1}^{\text{blg,grs}}(H_i), \theta_{\text{smooth}} = \{0.0, 0.25\}) \quad (3.128)$$

**Critical assumption 2: The surface model is spatially accurate.** This assumption is critical because dense matching methods used for the DSM determination tend to provide inaccurate outputs near



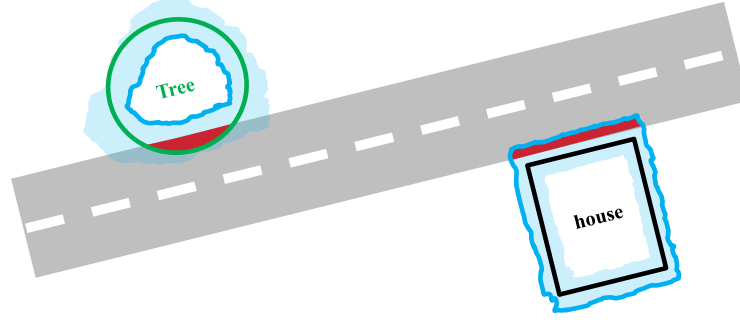


Figure 3.23.: Possible problems of an inaccurate nDSM. Blue lines: The 4 m level curve. Red: Contradictory sets. Shaded blue: Considered  $3\sigma$  tolerance of the 4 m level curve.

large height gradients such as at the 4 m level curve.<sup>8</sup> The problem is described by an additional system parameter, the standard deviation of the building borders  $\sigma_{\text{nDSM}}^{\text{xy}}$  (cf. Table 3.8).

$\sigma_{\text{nDSM}}^{\text{xy}}$	[m]	standard deviation of the spatial position of building borders in the nDSM
------------------------------------	-----	--

Table 3.8.: System parameter: nDSM accuracy.

Figure 3.23 shows possible problems of a spatially inaccurate nDSM. If a house, whose true position (black outline) next to the road is detected on the basis of a spatially inaccurate nDSM, then the set  $R^B$  might be displaced in direction towards the road (cf. Figure 3.23). As a consequence, the road hypothesis is assigned to the state *incorrect* because the contradiction set is not empty although the road hypothesis is *correct*. In contrast, a displacement in the opposite direction (direction away from the road) does not have that effect.

If a tree (green outline in Figure 3.23) is located near the road, this usually does not have any effect on the verification outputs because the presence of trees is not considered. However, if pixels corresponding to a tree are misclassified as *grassland*, then the state *incorrect* may be assigned to a road hypothesis although it is *correct* because the cardinality of the contradiction set may exceed the threshold  $(D^{\text{xy}})^2$ . These types of misclassifications are expected if the 4 m level curve is displaced in direction away from the road. Both situations are indicated in Figure 3.23, where the set  $R^G$  is defined on the basis of a spatially inaccurate nDSM. In contrast, a displacement in the opposite direction (direction towards the road) does not have that effect.

Consequently, the critical directions of the spatial inaccuracies of the 4 m level curve are the opposite for both modules. If the spatial uncertainty of the 4 m level curve corresponds to the blue shaded area depicted in Figure 3.23, the presence of the contradiction sets in those areas would be also considered as to be uncertain, and thus represents a valuable uncertainty metric for the given problem.

According to the previous discussion the indicator feature is defined on the basis of the spatial distance of contradictory sets to the 4 m level curve:

$$x_{u,2}^{\text{blg}}(H_i) = \max_{p_b \in \{R^B \cap R_i^{\text{H}\bullet}\}} d^{\text{BUT}}(p_b) \quad (3.129)$$

$$x_{u,2}^{\text{grs}}(H_i) = \max_{p_b \in \{R^G \cap R_i^{\text{H}\bullet}\}} d^{\text{GUR}}(p_b) \quad (3.130)$$

<sup>8</sup>Matching and subsequent filter operations also lead to other kinds of errors that are not considered here.



In Equations 3.129 and 3.130,  $d^{\text{BUT}}(p_b)$  and  $d^{\text{GUR}}(p_b)$  denote the distances of a pixel  $p_b$ , being part of the contradictory sets, to the 4 m level curve. These distances are defined by applying the positive Euclidean distance transform described in [Soille, 2003] to the two sets  $R^{\text{BUT}}$  and  $R^{\text{GUR}}$ , respectively. Both sets are inverse to each other (cf. Equations 3.117 and 3.118), and thus the direction with respect to the 4 m level curve is considered. It is important to note that only the maximum distance of all pixels  $p_b$  contained in the contradiction set is considered because the related pixels are expected to contradict the road hypothesis with the maximum probability. Thus, if the indicator features are larger than  $3\sigma_{\text{nDSM}}^{\text{xy}}$ , the xy-positions of the contradictory sets are expected to be accurate enough to satisfy the critical assumption, while for values near zero the assumption is questionable. Furthermore, it is expected that the probability of satisfying the critical assumption increases smoothly with increasing values  $x_{u,2}^{\text{blg,grs}}(H_i)$ . This heuristic is plugged into Equation 3.20 which leads to the functional model:

$$P(a_2 = \textit{satisfied} | x_{u,2}^{\text{blg,grs}}(H_i), \theta_{\text{smooth}} = \{3\sigma_{\text{nDSM}}^{\text{xy}}, 0.0\}) \quad (3.131)$$

**Critical assumption 3: Database errors can always be explained by a redevelopment of buildings or grassland.** The assumption can only be interpreted to be *satisfied* if a building or grassland object has really been observed. Therefore, the (negative) verification outputs are used to define the indicator feature for this critical assumption:

$$x_{u,3}^{\text{blg,grs}}(H_i) = -P(\textit{road} = \textit{correct} | x_v^{\text{blg,grs}}(H_i), \textit{model} = a.) \quad (3.132)$$

Due to the binary character of the verification outputs, the functional relation approximated by the sigmoid function is:

$$P(a_3 = \textit{satisfied} | x_{u,3}^{\text{blg,grs}}(H_i), \theta_{\text{rigid}} = \{0.0\}) \quad (3.133)$$

Consequently, the building and grassland detection modules have only an effect on the final verification result, if their verification output assigns the state *incorrect* to the road hypothesis.

## Discussion

Both the building and the grassland detection modules explicitly search for evidence against the correctness of road hypotheses, which is an essential difference to all other modules discussed so far. Both modules are relevant for the overall approach as they focus on specific types of errors that are not necessarily covered by the other modules.

Building detection and grassland detection are implemented as two independent modules, even though they mainly share the same strategy. Hence, image classification is applied twice for each road hypothesis. Despite that computational overhead, the independence of the modules is meaningful because grassland and building objects have a different relation to road objects: Roads and buildings are disjunct, but roads and grassland are not entirely because generalisation effects have to be taken into account.

The identified problem of spatially incorrect nDSM is closely related to the production process of those data. Therefore, using interim results of the production process such as the probability for a

successful matching of pairs of pixels would be interesting. However, in this thesis this possibility is not further investigated.

### 3.3.7. Adjacency analysis

In contrast to the other verification modules introduced so far, the adjacency analysis module does not consider local image information. Instead, the verification outputs of the other verification modules for neighbouring road hypotheses are taken into account. The special status of the adjacency analysis among the verification modules can be seen through the workflow shown in Figures 3.1 and 3.3, where the adjacency analysis module is located outside the core component. Despite that difference, the adjacency module also provides the typical per-hypothesis outputs that are propagated back to the core component.

The verification strategy presented in this section does not have an explicit equivalent in the literature but is related to works that consider road network characteristics in post-processing, e.g. [Zhang & Couloigner, 2006].

## Model

The model assumes the structure of the road network to be defined by its function, according to which roads provide effective links between places. Thus, the position of a single road object is related to the whole road network and in particular to its direct neighbours. It is further assumed that redevelopment projects leading to outdated road databases have an extent larger than a single road object, and thus do affect multiple road objects.

## Verification strategy

The verification strategy of the module consists of two steps. First, the network topology is analysed in order to identify adjacent road hypotheses for a given road hypothesis. Secondly, the verification outputs of all other modules for the adjacent road hypotheses are transferred to the currently considered road hypothesis.

**Topology of the road network:** All road hypotheses stored in the database form a graph. The road centreline hypotheses  $h_i^{xy} \in H_i$  are interpreted as the edges of the graph. Their end points are interpreted as the nodes of the graph. Road centreline hypotheses that share a common end point are connected to the same graph node. Subsequently, any road centreline hypothesis  $h_j^{xy}$  with  $j \in X_i^{\text{adj}}$  is considered to be adjacent to a road centreline hypothesis  $h_i^{xy}$ , if it is connected to any of the end points of  $h_i^{xy}$ . Here, the set  $X_i^{\text{adj}}$  contains the indices of the adjacent road hypotheses. The number of road centreline hypotheses that are connected to the same node defines the *degree* of that node.

**Definition of the verification output:** First, it is assumed that for each adjacent road hypothesis verification outputs of the form  $P(\text{road} = \text{correct} | x_v^{vm}(H_j), \text{model} = a.)$  are available. Furthermore, a set  $X^{\text{locmod}}$  is defined that includes the names of the other (locally operating) verification modules. In fact, the adjacency analysis module just collects these outputs and assigns them to the road

hypothesis  $H_i$  as its own verification output. Accordingly, the verification output of the adjacency module is a set of probability functions  $P_j^{vm} : \Theta_v \rightarrow [0, 1]$ :

$$\mathbf{P}_v^{\text{adj}} = \{P(\text{road} = \text{correct} | x_v^{vm}(H_j), \text{model} = a.)\} \quad (3.134)$$

$$\forall vm \in X^{\text{locmod}}, j \in X_i^{\text{adj}}$$

The set  $\mathbf{P}_v^{\text{adj}}$  has the cardinality  $|\mathbf{P}_v^{\text{adj}}| = |X^{\text{locmod}}| \cdot |X_i^{\text{adj}}|$ . The whole set takes part in the reasoning, which means that the number of verification outputs assigned to a road hypothesis  $H_i$  grows by a factor  $(1 + |X_i^{\text{adj}}|)$ . Figure 3.24 demonstrates the introduced strategy for a hypothetical road hypothesis  $H_3$  that has four neighbours.

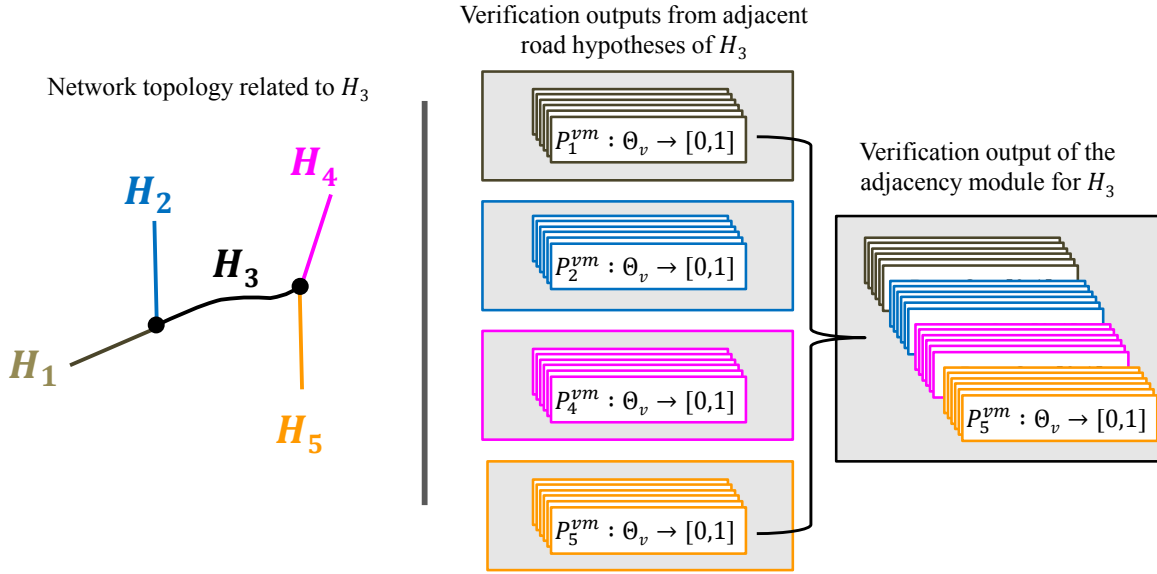


Figure 3.24.: Vector stacking principle of the adjacency method. The colours indicate the information stemming from the different adjacent road hypotheses.

### Model-uncertainty

The underlying model of the adjacency module implicitly assumes that changes, such as those related to redevelopment projects, affect more than a single road hypothesis, which is critical if the road hypotheses are longer than the extent of the area of change is. Furthermore, it is implicitly assumed that a road hypothesis is completely embedded into road network, which is critical for hypotheses representing dead end roads. Thirdly, according to the previous discussions the transferred verification outputs are still uncertain.

**Critical assumption 1: Changes of the road network do not only affect single road hypotheses.** This assumption is critical for road hypotheses that are larger than the expected spatial extent of those changes. The length of a road hypothesis  $h_i^{\text{length}} \in H_i$  (cf. Table 3.4) is introduced as indicator feature such that:

$$x_{u,1}^{\text{adj}}(H_i) = h_i^{\text{length}} \quad (3.135)$$

The spatial extent of the expected changes has to be specified by the human operator as an additional system parameter  $S^{\text{change}}$  (Table 3.9).

$S^{\text{change}}$	[m]	expected spatial extent of changes in the road network
---------------------	-----	--

Table 3.9.: System parameter for the adjacency module.

For  $h_i^{\text{length}}$  near zero the critical assumption is expected to be satisfied. In contrast, for  $h_i^{\text{length}} > S^{\text{change}}$  the critical assumption is questionable. Furthermore, it is expected that the probability of satisfying the critical assumption increases smoothly for decreasing values of  $h_i^{\text{length}}$ . This heuristic is introduced in Equation 3.20, which leads to a functional model for sigmoid of the form:

$$P(a_1 = \text{satisfied} | x_{u,1}^{\text{adj}}(H_i), \theta_{\text{smooth}} = \{0.0, S^{\text{change}}\}) \quad (3.136)$$

**Critical assumption 2: The roads are always completely embedded in the road network.** This is critical if a road hypothesis is not connected to two junctions because even big redevelopment projects may only affect the outer limit of a dead-end road. Thus, the indicator feature  $x_{u,2}^{\text{adj}}(H_i)$  is defined as the number of nodes that are connected with the road hypothesis  $H_i$  and that have a degree larger than one. Only if a road is connected with other roads at both ends, it is in accordance with the model assumption. This heuristic allows a clear separation of the two possible states of the critical assumption, which leads to a rigid parametrisation given by the delta function:

$$P(a_2 = \text{satisfied} | x_{u,2}^{\text{adj}}(H_i), \theta_{\text{rigid}} = \{2.0\}) \quad (3.137)$$

**Transfer of model-uncertainty:** The model-uncertainty outputs  $P(\text{model} = a | x_u^{\text{vm}}(H_j))$  given from the verification modules  $\text{vm} \in X^{\text{locmod}}$  for all  $j \in X_i^{\text{adj}}$  are considered by a set  $\mathbf{P}_u^{\text{adj}}$  including  $|\mathbf{P}_u^{\text{adj}}| = |X^{\text{locmod}}| \cdot |X_i^{\text{adj}}|$  probability functions  $P_j^{\text{vm}} : \Theta_u \rightarrow [0, 1]$ :

$$\mathbf{P}_u^{\text{adj}} = \{P(a_1 = s | x_{u,1}^{\text{adj}}(H_i), \theta) \cdot P_{\text{adj}}(a_2 = s | x_{u,2}^{\text{adj}}(H_i), \theta) \cdot P(\text{model} = a | x_u^{\text{vm}}(H_j)) | \text{vm} \in X^{\text{locmod}}, j \in X_i^{\text{adj}}\} \quad (3.138)$$

Consequently, the set  $\mathbf{P}_u^{\text{adj}}$  (Equation 3.138) defines the model-uncertainties corresponding to the verification outputs  $\mathbf{P}_v^{\text{adj}}$  (Equation 3.134). Both  $\mathbf{P}_u^{\text{adj}}$  and  $\mathbf{P}_v^{\text{adj}}$  define a set of probability mass functions  $m_i^{\text{adj}(\text{vm})} : 2^\Theta \rightarrow [0, 1]$  (cf. Figure 3.3).

## Discussion

The other verification modules only consider the imagery in the vicinity of a particular road hypothesis, but changes to the road network are usually not restricted to single roads. For instance, changes caused by redevelopment projects affect larger areas with potentially many road hypotheses. An example for this can be seen in the right part of Figure 3.25. On the other hand, the probability is rather small that a short road hypothesis such as the one in the left part of Figure 3.25 is *incorrect* if all surrounding road hypotheses are *correct*. The proposed strategy is simple as it merely transfers the verification outputs from one hypothesis to the adjacent ones. Accordingly, information coming from the adjacent road hypotheses is considered by the overall reasoning approach. The characteristics of the transferred information are indicated by the colours of the circles in Figure 3.25. For the situation shown in the left part of Figure 3.25, the proposed strategy works well because all information consistently indicates that the short road hypothesis is *correct*, which it is indeed.

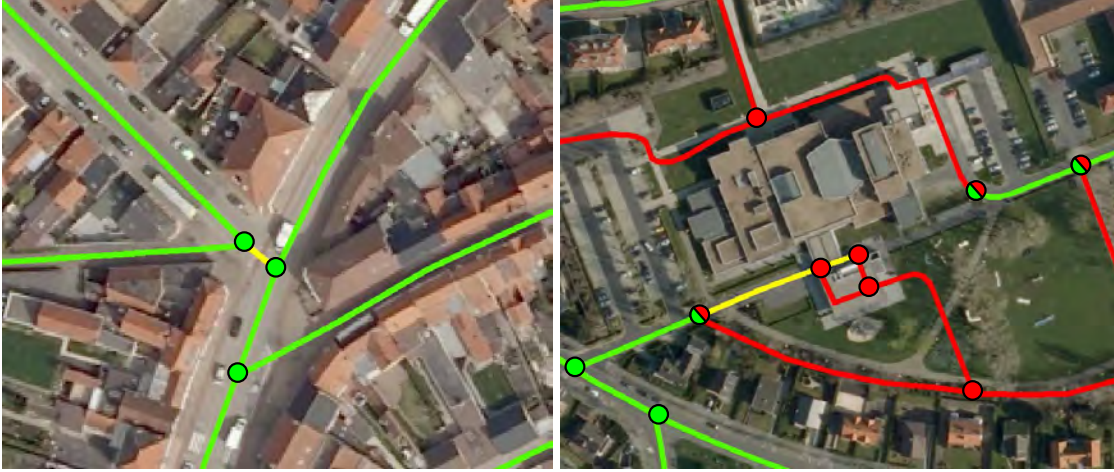


Figure 3.25.: A typical problem of the proposed per-hypothesis strategy without the adjacency module is that the states of short road hypotheses often remain *unknown* (yellow). Left: A situation with only *correct* (green) road hypotheses in the neighbourhood. Right: A situation with mostly *incorrect* (red) road hypotheses in the neighbourhood.

For the situation shown in the right part of Figure 3.25, the circles with two colours indicate conflicting information at the margin of the redeveloped area. This is problematic if the conflict mass becomes large, i.e.  $K_i \geq 0.9$  (cf. Algorithm 1) because then the respective hypothesis is assigned to state *invalid*, which means that the information provided by the other modules is ignored. Hence, the adjacency analysis might affect the performance of the overall approach for road hypotheses at the border of redevelopment projects. From the viewpoint of a practical application this might be less problematic as the human operator will probably have a look at those objects when correcting the adjacent one. On the other hand, this problem is relaxed by the introduced concept for the model-uncertainty that lowers the effect of the adjacency analysis for longer road hypotheses. The definition of the parameter  $S^{\text{change}}$  (cf. Table 3.9) directs the main task of the adjacency analysis to find solutions for short road hypotheses, where it is expected to be most beneficial because this is a weak point of the other modules. A reason for the weakness is that short road hypotheses correspond to small amounts of pixels in the imagery, which makes a verification more difficult.

Further, it has to be noted that the proposed strategy of the adjacency module is still not global as only the direct neighbours of a road hypothesis are involved. In the related work, several approaches can be found that apply a global network analysis, e.g. [Fischler et al., 1981; Wiedemann & Ebner, 2000; Poullis & You, 2010]. The main reason why such global concepts are not realized in this thesis is the fact that the current approach only concentrates on verification, and thus does not consider new roads for which a global network analysis is most important. Furthermore, a global approach will suffer from ignoring the new roads that potentially represent alternative connections.

## 3.4. Practical aspects of the approach

### 3.4.1. System parameters

Throughout the preceding sections, several so-called *system parameters* have been introduced. The adjustment of these system parameters allows adapting the verification approach to different types

of road databases, imagery, geographic regions and scene properties. Table 3.10 summarizes all the introduced system parameters. The table is subdivided into three thematic groups describing *database*, *imagery* and *scene* properties. The first group describing the database properties contains the parameters for the required database accuracy  $D^{xy}$  and  $D^{width}$  that are responsible for the test criteria of the verification modules. The parameters  $D^{rms(xy)}$  and  $D^{rms(width)}$  denote the actual accuracy of the database geometries. In cases where  $D^{xy} \gg D^{rms(xy)}$  and  $D^{width} \gg D^{rms(width)}$  holds true, using  $D^{rms(xy)}$  and  $D^{rms(width)}$  instead of  $D^{xy}$  and  $D^{width}$  enhances the performance of the overall approach because the image regions can be defined more precisely. The parameter  $S^{change}$  denotes the expected spatial extent of changes in the road network. Errors related to smaller changes might be overlooked because the verification results from the adjacent road hypotheses gain increasing influence on shorter road hypotheses.

The second group of parameters has to be set in accordance with the available input imagery. The parameter for the GSD is mainly used for conversions between parameters defined in image space and parameters defined in object space, e.g. for the definitions of the image sets contained in  $H_i$  (cf. Table 3.4) and the definition of the number of the *acupuncture nails* in Equation 3.102. The parameters  $\sigma_{ir}^{grey}$  and  $\sigma_{red}^{grey}$  denote the standard deviations of the grey values of the red and infrared colour bands. In this thesis, the values are interpreted as meta data available for the imagery. The parameter  $\sigma_{ndsm}^{xy}$  denotes the standard deviation of the x and y coordinates from the ground truth in the ndsm image.

The third group of parameters refers to the particular scene properties. The parameter  $r_2$  defines the extent of the local context area considered for the verification of a road hypothesis. It is a basic requirement that the context properties mostly do not change within the distance  $r_2$  in a direction perpendicular to the road hypothesis. Thus, it is important that  $r_2$  is not too large. In other words, buildings next to a road should be included but a crop field behind the buildings should not. However, a small value for  $r_2$  is also problematic because some modules such as the SSH and the acupuncture method integrate the image properties over the context area, and thus are less effective if  $r_2$  is defined too small. The correct values for the parameter  $r_2$  depends on the scene characteristics, i.e. it describes how frequently the context area in a scene changes.

In addition to the parameter  $r_2$ , there are a plenty of other parameters responsible for scene adaptation that are not listed in Table 3.10, e.g. the Lagrange multipliers for the SVM-based classification approaches, the GMM parameters and the Steger-line parameters. All these parameters have mainly the task to adjust the approach to radiometric scene properties and all of them are instantiated on the basis of the training datasets for *roads*, *buildings* and *grassland objects* (cf. Table 3.3).

Another parametrisation issue to be discussed here is related to the applicability of the verification modules depending on the given input data. As can be seen in the first three columns in Table 3.11, the individual requirements of the verification modules are rather different. In order to make the approach flexible for different input datasets, the verification modules can be disabled if the respective requirements are not fulfilled. Because every module represents an encapsulated verification approach on its own the overall approach provides results even if only one verification module (other than the adjacency analysis) is enabled. Given the fact that the imagery can always be down-sampled, Table 3.11 only includes the upper limits for the GSD.

<i>database</i>	$D^{xy}$	[m]	required accuracy of the road centreline
	$D^{width}$	[m]	required accuracy of the width attribute
	$D^{rms(xy)}$	[m]	RMS error of the road centreline
	$D^{rms(width)}$	[m]	RMS error of the road width
	$S^{change}$	[m]	expected spatial extent of changes in the road network
<i>imagery</i>	$GSD$	[m]	ground sampling distance of the imagery
	$\sigma_{ir}^{grey}$	[bit]	standard deviation of grey values of the infrared colour band
	$\sigma_{red}^{grey}$	[bit]	standard deviation of grey values of the red colour band
	$\sigma_{nDSM}^{xy}$	[m]	standard deviation of the spatial position of building borders in the nDSM
<i>scene</i>	$r_2$	[m]	width of considered context region, perpendicular to the road centreline hypothesis

Table 3.10.: Parameters of the proposed approach.

verification modules	image type	GSD [m]	primary field of application	v. output	critical assumptions	
colour classification	colour	$\leq 0.5$	urban areas, can deal with shadow	binary	generative smooth heur. rigid heur.	1 4 1
pan-line detection	pan	$\leq 2$	open landscape, without any settlements	distrib.	discriminative rigid heur.	1 1
NDVI-line detection	IR, red	$\leq 2$	dense vegetation areas	distrib.	discriminative rigid heur.	1 1
nDSM-line detection	nDSM	$\leq 2$	dense urban area with attached houses	distrib.	discriminative rigid heur.	1 1
parallel edge detection	pan	$\leq 0.5$	rural area including small settlements	distrib.	smooth heur. rigid heur.	1 1
SSH method	pan	$\leq 2.5$	rural area including small settlements, deals with weak contrast	binary	smooth heur.	1
acupuncture method	pan	$\leq 1.0$	urban area with detached houses	binary	smooth heur.	1
building detection	nDSM, IR, red	$\leq 0.5$	development areas for buildings and roads	binary	smooth heur. rigid heur.	2 1
grassland detection	nDSM, IR, red	$\leq 0.5$	development areas for roads and grassland	binary	smooth heur. rigid heur.	2 1
adjacency analysis	-	-	short road objects	-	smooth heur. rigid heur.	1 1
combination of all	pan, colour, IR, nDSM	$\leq 0.5$	different context areas, different road types	5 binary 4 distrib.	generative discriminative smooth heur. rigid heur.	1 3 12 8
						24

Table 3.11.: Summary of the module specific properties.

### 3.4.2. Implementation issues

From the human operator’s viewpoint, the proposed verification approach consists of three parts. The first part is interactive, i.e. the human operator defines the input data (road database and imagery), adjusts the system parameters and provides the training samples. The second part is the automatic image analysis, which concludes with a distinct decision for each road hypothesis. The third part is again interactive as it requires the human operator to investigate at least a subset of the road hypotheses.

When processing a large amount of data, i.e. thousands of road objects covering hundreds of square kilometres in object space, the automatic part might be required to run over night, which allows a human operator to proceed with the third part at the following working day. Hence, there is no demand for a real-time application. However, the implemented per-hypothesis verification strategy is ideal to be scaled up as it breaks down the verification problem into small sub-tasks. Each fully encapsulated sub-task only needs to consider a small image subset  $R_i^{[H]} \in H_i$  that will usually be much smaller than one square kilometre in object space, and thus does not require more than a few seconds of processing time for a single CPU. This opens up the possibility of parallel processing to reduce the computation time further, if required.

## 3.5. Discussion

To conclude this chapter, the advantages and the limitations of the proposed verification approach will be discussed.

A core characteristic of the proposed approach is its modular structure: Ten independent road verification approaches are combined to get an overall solution. The road model for each module focuses on a single road property, and is therefore comparably simple. The restriction to simple models has the consequence that one module alone is not able to deal with every imaginable situation, i.e. all possible appearances of *correct* and *incorrect* road hypotheses. However, different road properties are considered by the individual modules. For instance, the *pan-* and *NDVI-line detection modules* are designed for open landscape and rural areas; the *acupuncture module* for urban areas with detached houses and the *nDSM-line detection module* for dense urban areas. Furthermore, some of the modules are designed to be useful in situations where the other modules will fail, e.g. the *SSH module* solves situations with weak contrast and the *colour classification module* is able to deal with shadows. The idea behind the modular concept is that nearly any situation in the imagery can be explained by simple models, but not necessarily by the same simple model. The advantage of this strategy is that complex interactions between the model properties do not need to be considered, which leads to a relatively simple parametrisation as was shown with the small set of system parameters in Table 3.10. There is also a disadvantage in that: Not every situation might be that simple. For instance, a road that is only partly observed by its parallel borders and partly by its colour properties is not covered by the combined model, because it is not covered by at least one single model.

The models and strategies for the modules are mostly defined in close accordance to successful state-of-the-art road detection approaches. The most notable extension is the development of the



*colour classification module* using a GMM to deal with partly shadowed roads. Another important contribution of this work is the development of a verification strategy for the approaches that were originally developed for road detection and not for road verification. Prior knowledge from the database is introduced to define exact road hypotheses (cf. Table 3.4) and to specify the models of the detection algorithms. The latter point is interesting, as it makes the respective models more effective compared to their original versions, e.g. by a precise definition of the line width (cf. Section 3.3.2, *line detection modules*) or the precise definition of orientation and shape (cf. Section 3.3.5, *acupuncture module*).

Another important aspect of the new approach concerns the proposed model-uncertainties that are always based on an analysis of the fulfilment of the critical assumptions made for a particular road model. In summary, 24 critical assumptions are considered. All proposed analyses are new, except the line coverage assumptions (Equations 3.66 and 3.84) which were introduced by Gerke et al. [2004].

The most notable analysis of a critical assumption concerns the assumption of representative training data, which was introduced for the *colour classification module* (Section 3.3.1). This concept can be transferred to many classification problems in remote sensing applications apart from the context of road verification because assuming training data to be representative is quite often too optimistic.

The proposed combination concept allows a single module (a state-of-the-art approach) to violate its subjective *closed world assumption*. However, also the new (combined) approach is based on a *closed world assumption* because the considered critical assumptions have been identified based on the model descriptions, empirical results given in original publications and the experience gained in benchmark tests, which is nothing else then another closed world. In other words: In the context of world-wide applicability some critical assumptions may have been overlooked.

It is further worth noting that both the *verification outputs* and the *model-uncertainties* are defined in terms of probabilities. One advantage of this strategy is that the Probability Theory, which provides well-known concepts for classification and information fusion can be applied. The probabilistic concept is realized for the whole framework: Starting with the definition of the module-specific outputs up to the DST-based decision logic.

The DST framework explicitly expresses *uncertainty* and *ignorance* by only one distribution. This ability is used to define a decision space that also considers *unknown* as a possible state. The mapping to this decision space is realized by the *maximum probability mass rule* (Equation 2.24), which is a specific aspect of the proposed approach compared to the related works where the states expressing *ignorance* are mostly ignored by applying the *maximum support rule* (Equation 2.20). In contrast, the idea followed in this thesis is that *ignorance* is resolvable by querying additional information sources such as a human mind.



## 4. Experiments

In this chapter, initially six datasets that serve as the basis for the experiments will be described and discussed in the context of system parametrisation. As a sideline the dataset descriptions illustrate potential applications of the proposed approach. Afterwards, three different sets of experiments will be presented, each focussing on a specific aspect of the new methodology.

### 4.1. Datasets

The six input datasets used as the basis for the experiments show different properties concerning imagery, vector data, context areas, road types and geographic regions. Subsequently, all datasets will be briefly described.

**EuroSDR-Ikonos:** This dataset was provided by European Spatial Data Research (EuroSDR) as the basis for a benchmark test for road detection approaches [Mayer et al., 2006]. The imagery consists of two four band (IR,RGB) pan-sharpened Ikonos images with 1.0 m GSD. The reference datasets also provided by EuroSDR represent the centrelines of the correct road objects. The images are subdivided into three scenes of  $1,600 \times 1,600$  pixels, each showing an area of the Kosovo (cf. Figures A.2–A.4). The three scenes are described by Mayer et al. [2006] as:

- Ikonos1-sub1: sub-urban hilly scene of medium complexity (total length of the road network: 18 km)
- Ikonos3-sub1: flat rural scene of low complexity (total length of the road network: 8 km)
- Ikonos3-sub2: rural hilly scene of low complexity (total length of the road network: 6 km).

**EuroSDR-Aerial:** This dataset was also provided by EuroSDR and belongs to the same benchmark test as the EuroSDR-Ikonos dataset [Mayer et al., 2006]. The imagery consists of three scanned aerial images (RGB) with 0.5 m GSD, showing three different areas in Switzerland. Each aerial image consist of  $4,000 \times 4,000$  pixels (cf. Figures A.5–A.7). Reference data for the three scenes were made available by EuroSDR. The three scenes are described by Mayer et al. [2006] as:

- Aerial1: sub-urban scene of high complexity (total length of the road network: 43 km)
- Aerial2: hilly rural scene of medium complexity (total length of the road network: 22 km)
- Aerial3: hilly rural scene of low complexity (total length of the road network: 23 km).

**ISPRS-Vaihingen:** This dataset is part of a benchmark for urban object detection approaches, conducted by the International Society for Photogrammetry and Remote Sensing (ISPRS) described in [Rottensteiner et al., 2012]. In this thesis, a four band image (IR,RGB) with 0.2 m GSD and an nDSM of 0.2 m point spacing was derived from the original imagery provided by the German Society for Photogrammetry, Remote Sensing, and Geoinformation [Cramer, 2010]. The reference dataset, provided by ISPRS, comprises centrelines of 99 road objects with an overall length of 9 km. The image, having  $4,423 \times 5,890$  pixels, shows an urban area of medium to high density of development (cf. Figure A.8).

**MGCP-Algiers:** This dataset is a part of the Multinational Geospatial Co-production Program (MGCP). In accordance with [Farkas, 2009], the goal of MGCP is to generate a world-wide topographic map of a scale of 1:50,000 on the basis of remote sensing data. The data are acquired by different organizations all over the world. The road data represent the most important part of the maps<sup>1</sup> and their quality has to be checked by an independent authority after acquisition. This task is carried out on the basis of the same remote sensing data that were used for data acquisition. So far, the comparison of the road database with the remote sensing data is still carried out manually. As this is a very time-consuming task, the MGCP-organizers are looking for methods that allow automation of this work. For that purpose, they provided a test dataset to evaluate the approach presented in this thesis. The test dataset consists of an Ikonos image with four colour bands (IR,RGB) that are pan-sharpened with 1.0 m GSD. The image has  $12,000 \times 24,000$  pixels. Furthermore, the corresponding subset of the MGCP road database that contains 1,378 paved roads with solid surface and 909 dirt roads was provided. The overall length of the road network is 750 km, which is six times larger than the total road length of the EuroSDR and ISPRS datasets together. The scene corresponds to an area of  $200 \text{ km}^2$  and shows the coast of a region near Algiers in Algeria (cf. Figure A.9). A few settlements are situated within an agricultural plain, while the hilly terrain in the interior is covered by natural vegetation. The manual quality check of MGCP identified 33 incorrect road objects with an overall length of 13 km. The errors identified manually will be used as the reference for the proposed approach to evaluate its ability to detect errors under real conditions.

**GSI-Uruga:** This dataset is a part of the Japanese authoritative topographic cartographic database (Fundamental Geospatial Data, which also is introduced into the Digital Japan Basic Map (Map Information)), covering the whole urban area of Japan and corresponding to a mapping scale of 1:2,500. The data structure is standardized, but the data are produced by different local government agencies and thus show heterogeneous quality. Therefore, the Geo-spatial Information Authority of Japan (GSI) wants to establish quality control measures that ensure a consistent quality standard. For that purpose, GSI uses up-to-date aerial images that are available for the entire area of interest. However, due to the high efforts required for a complete manual comparison of the database with the images for the whole country, a manual procedure does not seem realistic. In order to reduce the huge manual efforts for this quality check, this task has to be automated without compromising the data quality. Thus, GSI provided a test dataset to evaluate the approach proposed in this thesis with respect to their task. The test dataset consists of an aerial image (RGB) with 0.2 m GSD that has

<sup>1</sup>It is worth noting that the MGCP database does not include all roads and buildings by design, i.e. less important connections of the road network and buildings are allowed to be missing.

$3,800 \times 4,600$  pixels. The respective subset of the Digital Japan Basic Map contains 396 road objects with an overall length of 15 km. The scene shows a sub-urban area, where significant differences between the database and the image appear due to a large redevelopment zone that stretches from the north-west to the south-east of the scene (cf. Figure A.10). Based on an interactive quality check of the road database, 330 road objects were declared to be correct, whereas 66 were found to be incorrect. This assessment serves as a reference for the automatic verification of an outdated road database.

**NGI-Zeebrugge:** This dataset is part of the Belgian road reference database, corresponding to a mapping scale of 1:25,000. The road database is maintained by the Belgian National Geographical Institute (NGI). For the future it is planned to update the road database every three years based on up-to-date imagery. Therefore, NGI investigates strategies to automate the updating process, and thus provided a test dataset to evaluate the approach proposed in this thesis with respect to their task. The test dataset consists of a colour image mosaic (IRRG) with  $32,000 \times 24,000$  pixels and 0.5 m GSD. Furthermore, the dataset consists of an nDSM that was computed on the basis of dense matching from stereo images (the same images that are the basis for the image mosaic). The corresponding subset of the road database contains 5,836 road objects with 620 km length. The test data correspond to an area of  $134 \text{ km}^2$  and show Zeebrugge and its surrounding area (cf. Figure A.11). The region is nearly planar and shows compact settlements of different density, large industrial facilities near the harbour and rural regions. With respect to the number of road objects, the Zeebrugge dataset is the largest of the test, whereas the overall length of the road network is slightly smaller than the MGCP-Algiers dataset (620 vs. 750 km). For the evaluation a road database extract from 1997 is used, while the imagery was acquired in 2011. Hence, many discrepancies related to recent road network improvements can be found, e.g. extensions to multi-lane roads and new road junctions that connect by-pass roads with new built-up areas. Based on an interactive quality check of a road database, whose latest update was in 2008, 5,259 road objects were declared to be correct, whereas 577 were found to be incorrect. It is worth noting that, in 2008, a professional cartographer required 20 full working days for the manual update of that database subset.

## Discussion of the datasets

The six datasets have rather different properties. The GSD of the considered imagery ranges from 1.0 m for the Ikonos images to 0.2 m for the high resolution aerial images. Only for some datasets, an infrared band or an nDSM are available. Furthermore, the datasets represent different geographic regions with different characteristics.

The vector data differ with respect to the expected database errors. The three benchmark datasets are error-free, the MGCP-Algiers dataset contains just a few database errors stemming from data acquisition, the GSI-Uruga dataset contains many database errors related to one single redevelopment project that is not contained in the database, and the NGI-Zeebrugge dataset contains predominantly database errors related to lots of smaller redevelopment projects of the road network that were carried out over a period of 14 years.

Another difference between the datasets originates from their purpose. The benchmark datasets were designed to evaluate different road detection algorithms under controlled conditions. They

consist of seven relatively small image subsets where each shows roads of different appearances. The datasets MGCP-Algiers and GSI-Uraga represent a typical cartographic task, i.e. the *quality control* of a database acquired by a third party. The MGCP-Algiers dataset is specific, as it is acquired using the same imagery that is available for quality control. Accordingly, errors related to temporal differences are widely ruled out, just as errors related to incorrect geo-referencing. In contrast, the GSI-Uraga dataset shows a development area under construction. This is interesting with respect to the introduced concept of model-uncertainty because such situations are not represented by any of the introduced models. The NGI-Zeebrugge dataset represents another typical cartographic task, i.e. verification as part of a *database update*. The temporal difference between database and imagery is 14 years, and thereby unusually large. This has the advantage that the basis for analysing the ability of the method to detect typical database errors is larger.

A further important aspect is the size of the datasets. In particular the datasets MGCP-Algiers and NGI-Zeebrugge are large compared to the benchmark datasets as well as those datasets that are commonly applied in the related work [Das et al., 2011]. Therefore, the numbers computed for these datasets can be expected to be more stable.

## Parameter settings

Table 4.1 shows the parametrisation of the proposed approach for the six datasets in accordance with Tables 3.10 and 3.11 that were discussed in Chapter 3. The differences between the required database accuracies  $D^{xy}$ ,  $D^{width}$  for the three cartographic datasets are related to the different mapping scales of those databases. For the benchmark datasets, such specifications are not available, and thus the required values are set to 3 m, i.e. the rounded average value of the other cartographic datasets. For the datasets NGI-Zeebrugge and GSI-Uraga the parameter  $S^{change}$  is set in accordance with the diameter of typical redevelopment projects that occur in those scenes. As for the three benchmark datasets and the MGCP-Algiers dataset, database errors related to redevelopment projects can be ruled out,  $S^{change}$  is set to a relatively small value of 10 m. The parameters  $\sigma_{ir}^{grey}$ ,  $\sigma_{red}^{grey}$  and  $\sigma_{nDSM}^{xy}$  are set in accordance with the properties of the input data. They are required for the modules building detection and grassland detection which are only applied to the datasets ISPRS-Vaihingen and NGI-Zeebrugge. The parameter  $r_2$  is set to a small value (15 m) for all those datasets where dispersed settlements occur frequently. For the datasets, NGI-Zeebrugge and GSI-Uraga  $r_2$  is set relatively large because here the settlements form large compact regions. The training datasets consisting of building and grassland areas are only required if the colour classification module is applied, which is only the case if the image GSD is at least 0.5 m.

Due to the different input data properties, some verification modules are disabled for some datasets. Only for the datasets ISPRS-Vaihingen and NGI-Zeebrugge, all modules are enabled because only for these two datasets are all required input data available.

		EuroSDR-Ikonos, Kosovo	EuroSDR-Aerial, Switzerland	ISPRS-Vaihingen, Germany	MGCP-Algiers Algeria	GSI-Uruga, Japan	NGI-Zeebrugge, Belgium
<i>data properties</i>	input imagery	RGB,IR	RGB	IRRG,nDSM	RGB,IR	RGB	IRRG,nDSM
	platform	Satellite	Aerial	Aerial	Satellite	Aerial	Aerial
	GSD	1.0 m	0.5 m	0.2 m	1.0 m	0.2 m	0.5 m
	database mapping scale	-	-	-	1:50,000	1:2,500	1:25,000
	covered area	7.7 km <sup>2</sup>	12 km <sup>2</sup>	1.1 km <sup>2</sup>	200 km <sup>2</sup>	0.7 km <sup>2</sup>	134 km <sup>2</sup>
	number of road objects	308	696	99	2,287	396	5,836
	overall road length	32 km	88 km	9 km	750 km	15 km	620 km
<i>database</i>	$D^{xy}$	3 m	3 m	3 m	5 m	2 m	3 m
	$D^{width}$	3 m	3 m	3 m	5 m	2 m	3 m
	$D^{rms(xy)}$	0.33 m	0.33 m	0.33 m	0.33 m	0.2 m	0.66 m
	$D^{rms(width)}$	0.66 m	0.66 m	0.66 m	0.66 m	0.33 m	0.66 m
	$S^{change}$	10 m	10 m	10 m	10 m	100 m	30 m
<i>imagery</i>	$GSD$	1 m	0.5 m	0.2 m	1.0 m	0.2 m	0.5 m
	$\sigma_{ir}^{grey}$	-	-	2.1	-	-	1.8
	$\sigma_{red}^{grey}$	-	-	1.9	-	-	1.7
	$\sigma_{nDSM}^{xy}$	-	-	0.5 m	-	-	1.33 m
<i>scene</i>	$r_2$	15 m	15 m	15 m	15 m	30 m	30 m
	$X^{roads}$	x	x	x	x	x	x
	$X^{buildings}$	-	x	x	-	x	x
	$X^{grassland}$	-	x	x	-	x	x
<i>verification modules</i>	colour classification	-	x	x	-	x	x
	pan-line detection	x	x	x	x	x	x
	NDVI-line detection	x	-	x	x	x	x
	nDSM-line detection	-	-	x	-	-	x
	parallel edge detection	x	x	x	x	x	x
	SSH method	x	x	x	x	x	x
	acupuncture method	x	x	x	x	x	x
	building detection	-	-	x	-	-	x
	grassland detection	-	-	x	-	-	x
	adjacency analysis	x	x	x	x	x	x

Table 4.1.: Data properties and parameter settings ('x' denotes enabled).

## 4.2. Evaluation of the proposed methodology

In this Section three sets of experiments will be presented. The first set (Section 4.2.1) mainly focuses on the SVDD-based uncertainty metric as one of the key contributions of this thesis. The second set (Section 4.2.2) focuses on the proposed fusion framework and the performance of the verification modules in comparison with state-of-the-art road detection approaches. The third set (Section 4.2.3) focuses on real cartographic datasets and thus analyses the applicability of the proposed approach with respect to the main objectives of the thesis.

### 4.2.1. The SVDD-based uncertainty metric

An important contribution of this thesis is the proposed SVDD-based uncertainty metric (cf. Sections 2.2 and 3.3.1). It was argued that, with an increasing feature space distance  $f_{\text{SVDD}}(z_i)$  of a test sample  $z_i$  to the training samples  $X^{\text{svmtrain}}$  (cf. Equation 2.56), the probability for *ignorance* increases. According to this argumentation, the proposition of misclassifications is also expected to increase with the feature space distance. In order to check this hypothesis, the classification outputs of the  $\nu$ -SVC are evaluated with respect to different feature space distances to the SVDD hypersurface. Figure 4.1 displays the idea of the evaluation strategy subsequently described. In accordance with Figure 2.3, the feature vectors related to the training data are depicted in red (*non-roads*) and blue (*roads*), whereas the feature vectors related to potential *outliers* ( $f_{\text{SVDD}}(z_i) > 0$ ) are depicted in green. The thin grey lines represent different decision thresholds between *inliers* (represented by the training data) and *outliers* (not represented by the training data). By increasing a rigid distance threshold  $\beta \bar{d}_{\text{SVDD}}$ , more and more test samples symbolized by the green stars will be classified as *inliers*. Henceforth, it has to be checked if test samples displayed in a brighter colour will be misclassified with larger probability than the darker ones that reside nearer to the SVDD hypersurface.

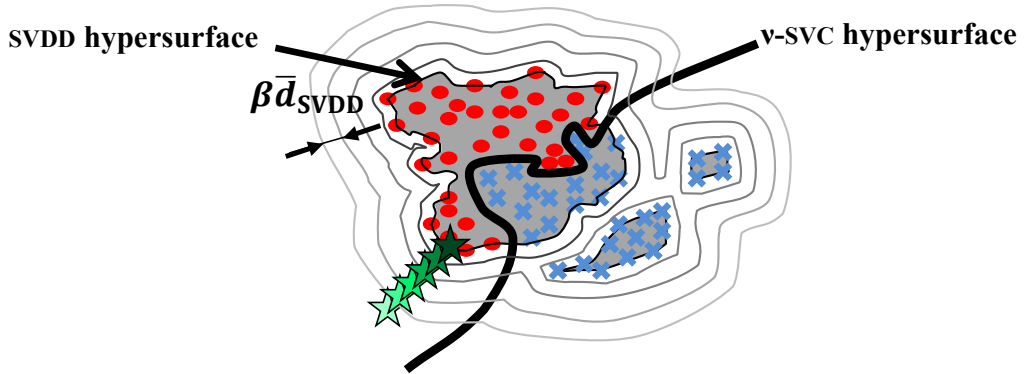


Figure 4.1.: SVDD-based outlier detection applying different thresholds for the distance to the SVDD hypersurface.

### Evaluation strategy

Only some parts of the colour classification module (cf. Section 3.3.1) are required for this set of experiments: The colour feature extraction based on the GMM<sup>\*</sup>, the  $\nu$ -SVC for classifying *roads* and *non-roads* and the SVDD-based kernel metric to determine the input feature space distance. In Section 3.3.1, six critical assumptions were considered for the model-uncertainty. However, only



the first critical assumption, i.e. *The training samples represent the test data*, is in the focus of the current set of experiments. In order to exclude other effects, the state of the other critical assumptions needs to be fixed. For that reason, only road hypotheses that satisfy the following inequality are considered further on:

$$\prod_{k=\{2,4,5,6\}} \left[ P(a_k = \textit{satisfied} | x_{u,k}^{\text{col}}(H_i), \theta) \right] \geq 0.8 \quad (4.1)$$

The constraint 4.1 refers to the critical assumptions  $k = \{2, 4, 5, 6\}$  that deal with the uncertainty of the feature vector definition (cf. Section 3.3.1).

The GSI-Uraga dataset, introduced in Section 4.1 serves as the basis for this set of experiments. This dataset is particularly useful to demonstrate the performance of the SVDD-based uncertainty metric as it is characterised by a large redevelopment area under construction, which has obviously different colour properties compared to the rest of the scene.

The required training dataset is deliberately defined to be partly non-representative for the scene. It consists of 36 road objects that are all placed in the centre of the scene, indicated by the blue lines in Figure 4.2. The 72 non-road objects used for training are generated from the road objects by a parallel shift in both directions perpendicular to the main axis of the road objects. Hence, the training data have a fairly homogeneous appearance while the characteristics of the redevelopment area and the forest areas in the periphery of the scene are not represented.

In the test case, all road database objects in the GSI-Uraga dataset that satisfy the Constraint 4.1 and that do not correspond with the training data in the centre are classified. For the remaining 276 (of 396), road hypotheses feature vectors  $z_i$  and feature space distances  $f_{\text{SVDD}}(z_i)$  (cf. Equation 2.56) are computed. According to the feature space distances  $f_{\text{SVDD}}(z_i)$ , the test samples  $z_i$  are classified as *inliers* or *outliers*. Unlike in [Tax & Duin, 2004], the decision threshold to discriminate *inliers* and *outliers* is not set to  $f_{\text{SVDD}}(z_i) = 0$ . Instead, the threshold is defined as a variable that will be adapted in the experiments. A test sample  $z_i$  is classified as *outlier* if:

$$f_{\text{SVDD}}(z_i) > \beta \bar{d}_{\text{SVDD}} \quad (4.2)$$

where  $\bar{d}_{\text{SVDD}}$  denotes the average distance of all feature vectors used for training that reside inside the SVDD hypersphere:

$$\bar{d}_{\text{SVDD}} = \frac{1}{|X^{\text{trainSVM}} \setminus X^{\text{SV(SVDD)}}|} \sum_{x_i \in X^{\text{trainSVM}} \setminus X^{\text{SV(SVDD)}}} f_{\text{SVDD}}(x_i) \quad (4.3)$$

Equation 4.3 explicitly excludes the support vectors ( $X^{\text{SV(SVDD)}}$ ) that are characterized by feature space distances  $\geq 0$  to the SVDD hypersurface. The variable  $\beta = [0, \infty]$  controls the distance threshold, whereas  $\bar{d}_{\text{SVDD}}$  provides an intuitive unit for its value.

The classification output of the  $\nu$ -SVC (*road*, *non-road*) is only considered if the respective test sample  $z_i$  is classified as an *inlier*. Given the classification output and the reference dataset described in Section 4.1, a confusion matrix can be defined:

		combination of SVDD and $\nu$ -SVC		
		road	non-road	outlier
reference	road	<i>RR</i>	<i>RN</i>	<i>RO</i>
	non-road	<i>NR</i>	<i>NN</i>	<i>NO</i>

Table 4.2.: Confusion matrix for the combined colour classification approach.

According to Table 4.2 two evaluation parameters are defined:

$$\textit{classification correctness} = \frac{RR + NN}{RR + NN + RN + NR} \quad (4.4)$$

$$\textit{classification completeness} = \frac{RR + NN + RN + NR}{RR + NN + RN + NR + NO + RO} \quad (4.5)$$

The *classification correctness* (cf. Equation 4.4) indicates the relative amount of correct class assignments with respect to all class assignments. Here, *outliers* are not considered. The optimal value for the *classification correctness* is 100%, which means that all class assignments are correct. The *classification completeness* (cf. Equation 4.4) indicates the relative amount of *inliers*.

## Results

The reference dataset contains 330 correct road hypotheses and 66 incorrect ones. Applying the constraint 4.2 and removing the hypotheses that were part of the training data leads to 227 test samples that correspond to *roads* (correct road hypotheses) and 49 test samples correspond to *non-roads* (incorrect road hypotheses).

Table 4.3 shows the confusion matrices of the results for the  $227 + 49 = 276$  test samples determined by using four different values  $\beta = \{0, 1, 2, \infty\}$ . Figure 4.2 shows a graphical representation of the results.

In the first scenario shown in Table 4.3, the  $\nu$ -SVC is combined with the SVDD using  $\beta = 0$ . Hence, only the  $\nu$ -SVC outputs of feature vectors from inside the SVDD hypersphere are considered. The second scenario considers also samples a bit outside of the SVDD hypersphere, i.e.  $\beta = 1$ . As expected, the *classification correctness* decreases moderately compared to the first scenario ( $-1.1\%$ ) while the *classification completeness* increases more clearly ( $+7.7\%$ ). This means that the majority of the classification outputs considered in addition to those of the first test are indeed correct. The third scenario with  $\beta = 2$  shows the same trend with respect to the *classification correctness* ( $-1.1\%$  compared to the second scenario), but the *classification completeness* increases less clearly than before ( $+4.4\%$  compared to the second scenario). The last scenario considers the  $\nu$ -SVC result alone, which leads to a *classification correctness* of 69.5% and an obvious *classification completeness* of 100%.

Figure 4.3 shows the *classification completeness* and the inverse<sup>2</sup> of the *classification correctness* for different values of  $\beta$ . The *classification completeness* increases most significantly for values  $\beta \leq 2$ , whereas the *classification correctness* is nearly constant. In contrast, for values  $\beta \geq 4$

<sup>2</sup>The inverse is chosen here as it better visualises the relation of both curves.



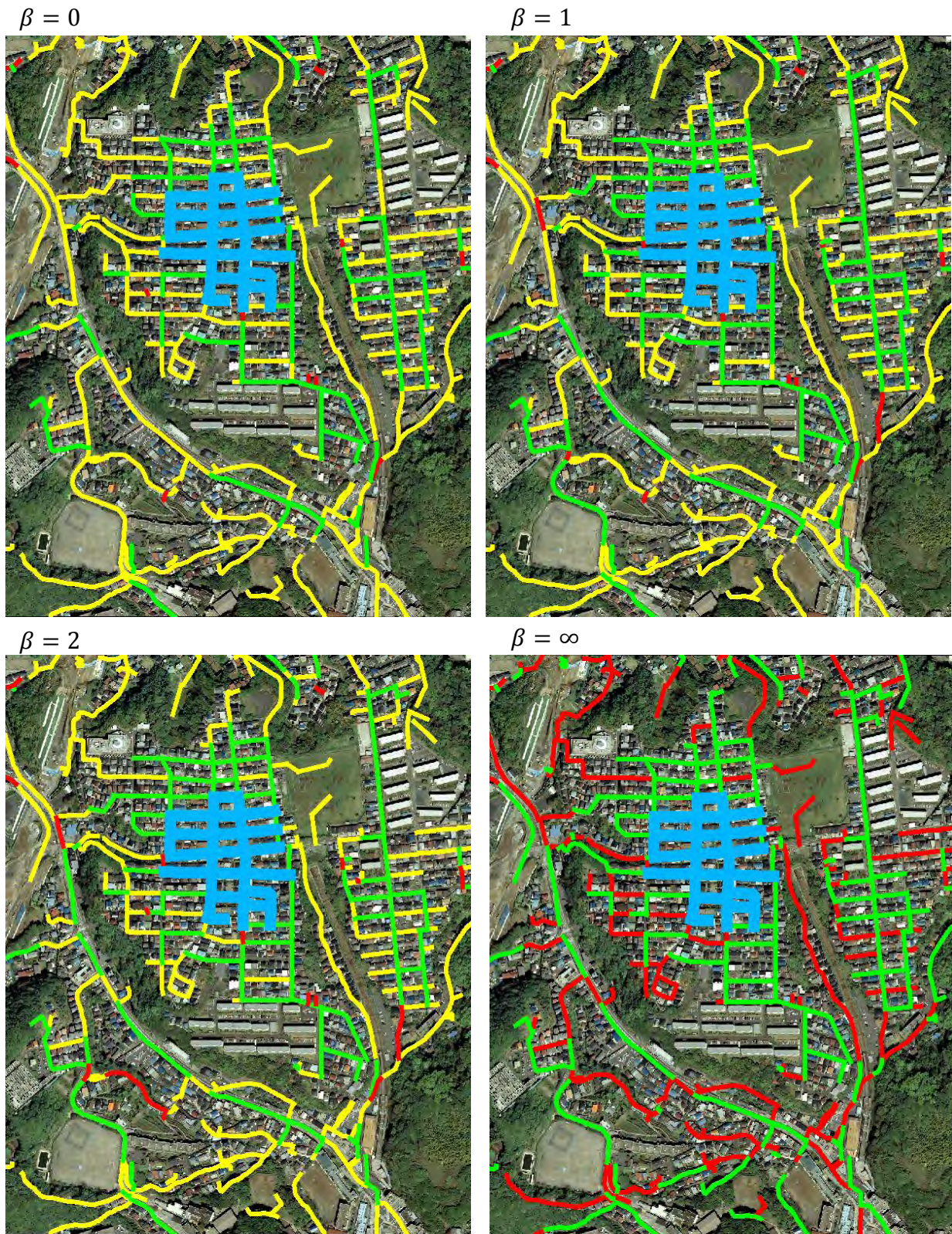
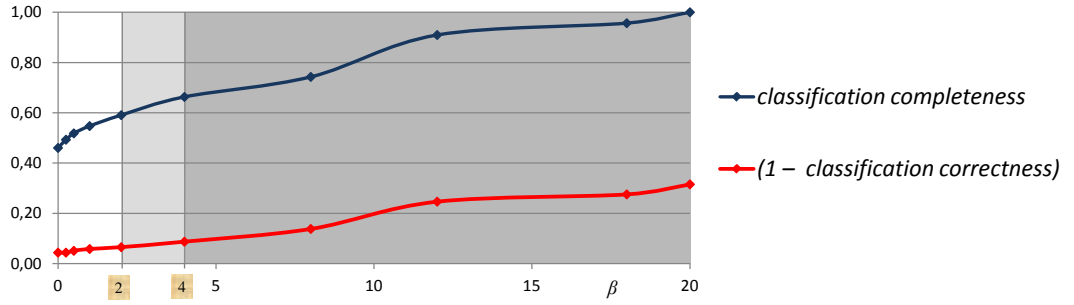


Figure 4.2.: Combination of two  $\nu$ -SVC and SVDD with different values of  $\beta$  (training data in blue, correct class assignments in green, incorrect assignments in red and *outliers* in yellow).



reference	proposed approach				
	<i>road</i>	<i>non-road</i>	<i>outlier</i>		
	combination of $\nu$ -SVC and SVDD with $\beta = 0$				
	<i>road</i>	113 (41%)	12 (4%)	102 (37%)	<i>classification correctness</i> = 95.7%
	<i>non-road</i>	0	2 (1%)	47 (17%)	<i>classification completeness</i> = 46.0%
	combination of $\nu$ -SVC and SVDD with $\beta = 1$				
	<i>road</i>	130 (47%)	16 (6%)	81 (29%)	<i>classification correctness</i> = 94.6%
	<i>non-road</i>	0	5 (2%)	44 (16%)	<i>classification completeness</i> = 54.7%
	combination of $\nu$ -SVC and SVDD with $\beta = 2$				
	<i>road</i>	138 (50%)	17 (6%)	72 (26%)	<i>classification correctness</i> = 93.5%
<i>non-road</i>	1 (1%)	7 (3%)	41 (15%)	<i>classification completeness</i> = 59.1%	
only $\nu$ -SVC without SVDD ( $\beta = \infty$ )					
<i>road</i>	154 (56%)	73 (27%)	-	<i>classification correctness</i> = 69.5%	
<i>non-road</i>	14 (5%)	35 (13%)	-	<i>classification completeness</i> = 100%	

Table 4.3.: Confusion matrices and evaluation parameters for different scenarios defined in the text.

Figure 4.3.: Influence of the distance threshold  $\beta \bar{d}_{\text{SVDD}}$  on the *inverse classification correctness* and the *classification completeness*.

the *classification completeness* merely increases slightly faster than the *classification correctness* decreases.

The results for the four scenarios introduced in Table 4.3 are shown in Figure 4.2. What Figure 4.2 does not show are the states of the road hypotheses. Looking at the four depicted scenarios, it becomes clear that the *roads* that have a similar appearance to the training objects in the centre of the scene are classified as *inliers*, and that they largely receive the correct class assignments. The *outliers* are largely found in areas where, due to different surroundings, the roads have a different appearance than the training data, e.g. in the forest where correct road hypotheses have a different appearance or in the redevelopment zone where incorrect road hypotheses have a different appearance.

## Discussion

First, the experiment confirms the expectation that a combination of a  $\nu$ -SVC with the SVDD achieves a significant reduction of misclassifications. If the SVDD decision threshold is set to zero as suggested

by Tax & Duin [2004], the *classification correctness* is increased from 69.5%, achieved without the SVDD, to 95.7% with the SVDD.

Secondly, the expected dependency of the proportion of misclassifications on the distance  $f_{\text{SVDD}}(z_i)$  was confirmed for the applied test dataset. For feature space distances  $\beta \bar{d}_{\text{SVDD}}$  with  $0 \leq \beta \leq 2$  it turned out that the *classification completeness* increases with a decreasing  $f_{\text{SVDD}}(z_i)$ , which basically means that the closer a test sample resides to the SVDD hypersurface, the larger the probability becomes for a correct class assignment.

In contrast, for feature space distances  $\beta \bar{d}_{\text{SVDD}}$  with  $\beta \geq 4$  the *classification completeness* increases constantly stronger than the *classification correctness* decreases, which basically means that the probability for a correct class assignment does not change with the feature space distance. In particular, approximately 25% of the samples far away from the SVDD (for  $\beta \geq 4$ ) are correctly classified independently of the feature space distance to the SVDD hypersurface. The reason for this observation can be seen in Table 4.3, where for  $\beta = \infty$  the  $\nu$ -SVC classifies most of the samples as *non-roads*, whereas the reference denotes the majority of these samples as *roads*. The predominant assignment to the class *non-roads* can be explained by the fact that regions corresponding to *non-roads* have more heterogeneous colour properties than those that correspond to *roads*, and thus the feature vectors corresponding to *non-road* regions span a larger sub-space in feature space. Consequently, the half-spaces separated by the  $\nu$ -SVC hypersurface do not have the same size, and thus arbitrarily located test samples appear with higher probability in the larger half-space, i.e. as *non-roads*.

For the experiment a value  $\beta = 2$  is a good choice to separate *inliers* from *outliers*, and thus it is a good choice to reduce the probability of misclassifications. However, this value cannot be transferred to another scene because the distribution of the outliers might be different. If the distribution of the *outliers* (and *inliers*) were similar enough, the conditional probabilities  $P(\text{road} = \text{correct} | x_v^{\text{col}}(H_i), \text{model} = n/a) = 0.75$  and  $P(\text{road} = \text{incorrect} | x_v^{\text{col}}(H_i), \text{model} = n/a) = 0.25$  could be plugged into Equation 3.7 and classical Bayesian inference could be applied.

In order to conclude the discussion, it is neither a good choice to transfer the determined optimal  $\beta$  from one scene to another nor it is easy to find a general heuristic for the value of  $\beta$ . In this thesis, the problem is overcome with the analysis of the test data distributions introduced in Section 3.3.1 (cf. Equations 3.45–3.47). A disadvantage of this strategy is that it disturbs the per-hypothesis strategy because it requires  $f_{\text{SVDD}}(z_i)$  for all database hypotheses. However, it also provides a scene-specific and optimal distance threshold, which is a considerable advantage.

#### 4.2.2. The data fusion strategy

The proposed fusion concept of combining independent verification approaches at decision level is one of the important contributions of this thesis. The experiments presented in this section were conducted to answer the following three questions:

- Which verification module is most successful in which kind of scene?
- By which margin is the combined solution better than the individual solutions?
- How good is the combined solution compared to solutions of the related works?

## Configurations

The experiments are conducted using the benchmark datasets introduced in Section 4.1, i.e. EuroSDR-Ikonos, EuroSDR-Aerial and ISPRS-Vaihingen. These datasets are particularly useful to evaluate the fusion concept as they are designed to evaluate different road detection approaches. In each case, the relatively small test scenes restrict the road detection problem to specific road types and context areas, which makes it easier to analyse the particular strengths of the verification modules under different circumstances.

The reference datasets from the benchmark tests alone are not useful as an input for a verification approach. Therefore, database errors are artificially generated by rotating the reference datasets by 180 degrees and adding these roads to the database. Consequently, the input data contain the same number of *correct* and *incorrect* road hypotheses, and thus allow observing the capability of detecting database errors in accordance with the experiments described in [Gerke & Heipke, 2008; Poulain et al., 2010].

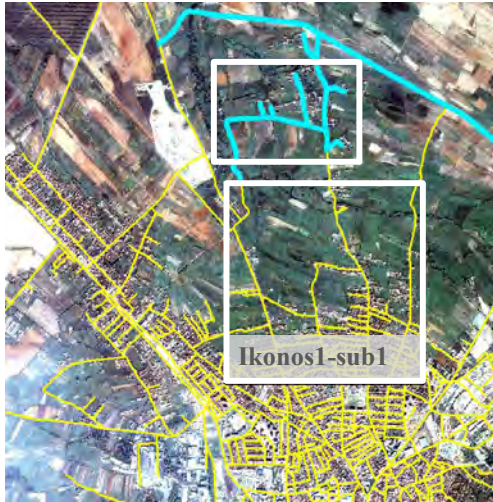
The system configurations for the three different datasets have already been given in Table 4.1. The training data for the three datasets are depicted in Figure 4.4. The upper picture in Figure 4.4 shows the selected training samples for the EuroSDR-Ikonos dataset, which only requires road regions ( $X^{\text{roads}}$ ) because the verification modules that require other datasets are disabled (cf. Table 4.1). The training data are selected from an area next to EuroSDR-Ikonos1-sub1. This is possible because the input images of the EuroSDR-Ikonos dataset are much larger than the three sub-images used in the benchmark test. For the EuroSDR-Aerial dataset, the training samples have to be defined inside one of the three scenes because no additional images of this set are available, which is shown in the centre of Figure 4.4. The training data required for the pan-line detection module and the colour classification module consist of image regions for each class ( $X^{\text{roads}}$ ,  $X^{\text{buildings}}$ ,  $X^{\text{grassland}}$ ). As for the EuroSDR-Aerial dataset and for the ISPRS-Vaihingen dataset independent training data are not available either, and thus they are defined inside the test scene, which is shown in the lower part of Figure 4.4.

All training datasets are chosen from geographically restricted areas in order to simulate a realistic situation where a human operator is not able to invest much time to always select representative training data. For the scenes EuroSDR-Aerial1 and ISPRS-Vaihingen, parts of the benchmark data overlap with training data, which might corrupt the evaluation. In order to keep the results from the proposed approach comparable to those analysed in the benchmark test, this problem is simply ignored here. The effect of ignoring the overlap in the evaluation results is small as will be shown at the end of this section.

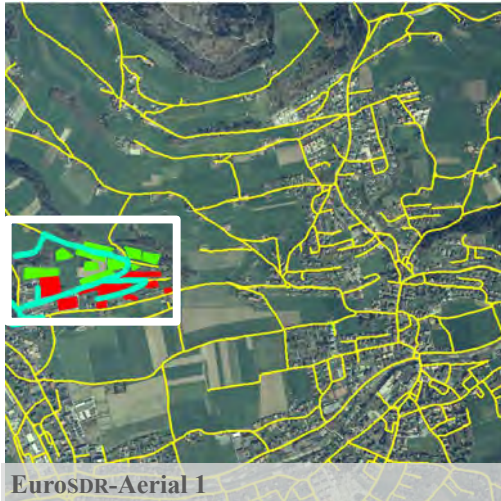
## Evaluation strategy

The main idea is to apply each verification module as an individual verification algorithm. Consequently, the results achieved by the verification modules can be compared to each other, which provides hints about the impact of the single road models on the combined solutions. Furthermore, these results can be compared with the results achieved in the benchmark test and the results achieved with the overall approach including all enabled verification modules. The latter allow

EuroSDR-Ikonos



EuroSDR-Aerial



ISPRS-Vaihingen

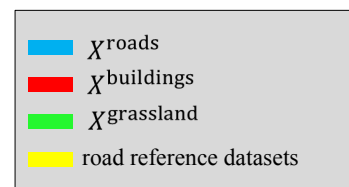


Figure 4.4.: Training datasets for the benchmark datasets.

conclusions about the performance of the fusion framework developed in this thesis.

In accordance with [Mayer et al., 2006], the verification results, provided by the individual modules and the combined solution are analysed on the basis of a confusion matrix:

		automatic approach	
		<i>correct</i>	<i>incorrect, unknown, invalid</i>
reference	<i>correct</i>	true positives ( <i>TP</i> )	false negatives ( <i>FN</i> )
	<i>incorrect</i>	false positives ( <i>FP</i> )	true negatives ( <i>TN</i> )

Table 4.4.: Confusion matrix for the benchmark test.

Based on the confusion matrix (cf. Table 4.4) the evaluation parameters *completeness* and *correctness* are defined as follows:

$$completeness = \frac{TP}{TP + FN} \quad (4.6)$$

$$correctness = \frac{TP}{TP + FP} \quad (4.7)$$

While Table 4.5 provides a general definition for *TP*, *FN*, *FP* and *TN*, differences between *road detection* and *road verification* approaches have to be taken into account because the evaluation of the *road detection* approaches is required to define a positional tolerance for the centre axis. This leads to slightly different definitions that are detailed in Table 4.5: The left column shows the original definitions used in [Mayer et al., 2006], while the right column shows definitions used for the road verification problem.

Mayer et al. [2006] consider both datasets, the reference and the detected roads, by their centrelines, and thus refer their definitions to the road length (cf. Table 4.5). Usually, the proposed approach would indicate an evaluation based on object counts similar to the evaluation presented in Section 4.2.1 because a human operator usually works at object level, e.g. motivated in [Beyen et al., 2008]. In order to make the results in [Mayer et al., 2006] comparable with the verification results, the length of the road hypotheses is considered for the evaluation in this section.

For both problems (*detection* and *verification*), the resulting *completeness* (cf. Equation 4.6) can be interpreted as a measure for how good a road model is capable of explaining the roads of the reference dataset. Therefore, the *completeness* will be used as the basis for comparing the verification modules with each other but also for a comparison with the results achieved by road detection approaches that participated the benchmark test [Mayer et al., 2006].

In contrast, the *correctness* (cf. Equation 4.7) does not allow a comparison with the road detection approaches because the possibilities to produce *false positives* (*FP*) are different for both problems. A verification module might only produce *false positives* for road hypotheses that are part of the rotated reference dataset, whereas a road detection approach might produce *false positives* (*FP*) in the entire image except for the buffered reference dataset. As the rotated reference dataset is usually far smaller than the image, the *correctness* is expected to be much larger for the verification modules than for the road detection approaches. Thus, the *correctness* will only be used as a basis for the comparison of the different verification modules and the combined solution, but not for a comparison with the road detection approaches that participated at the benchmark test [Mayer et al., 2006].



	road detection [Mayer et al., 2006]	road verification
true positives ( <i>TP</i> )	length of the reference dataset that lies within a buffer around the detected roads (the buffer width corresponds to the road width plus the spatial tolerance)	accumulated length of all road hypotheses corresponding to the reference dataset that are assigned to the state <i>correct</i>
false negatives ( <i>FN</i> )	length of the reference dataset that does not lie within a buffer around the detected roads	accumulated length of all road hypotheses corresponding to the reference dataset that are assigned to the states <i>incorrect</i> , <i>unknown</i> or <i>invalid</i>
false positives ( <i>FP</i> )	length of the detected roads that does not lie within a buffer around the reference dataset	accumulated length of all non-road hypotheses (corresponding to the rotated reference dataset) that are assigned to the state <i>correct</i>
true negatives ( <i>TN</i> )	-	accumulated length of all non-road hypotheses that are assigned to the states <i>incorrect</i> , <i>unknown</i> or <i>invalid</i>

Table 4.5.: Definitions of the evaluation criteria for road detection and road verification approaches.

In a semi-automatic scenario, road verification approaches naturally prioritise a large *correctness* over a large *completeness* because any *false positives* corresponds to database errors that remain undetected, whereas *false negatives* just represent correct hypotheses that have to be manually investigated without necessity. In this regard, Gerke & Heipke [2008] and Poulain et al. [2010] presented experiments with rotated reference datasets which they called *sensitivity analysis*, for which Gerke & Heipke [2008] achieved a *correctness* of 1.0 and Poulain et al. [2010] a *correctness* of about 0.91. In contrast, road detection approaches without a concrete practical application as a background, e.g. those described in [Mayer et al., 2006], consider *completeness* and *correctness* with the same priority and hence, try to keep them in balance. In this thesis, the main objective is to achieve results with a *correctness* larger than 0.91, whereas the second objective is to achieve results with a large *completeness*, e.g. in the same order of magnitude as the best results of the benchmark test [Mayer et al., 2006].

## Results

Tables 4.6–4.8 show the results of the verification modules in decreasing order with respect to their *completeness*, followed by the combined solution and the best result of the EuroSDR benchmark test.<sup>3</sup> The column *rk* denotes the rank with respect to the *completeness*. The signed values in brackets denote the exclusive contributions of the modules for the combined solution, i.e. the difference for the *completeness* of combined solutions achieved with and without the respective module.

Figures 4.5–4.30 show the results achieved by the most successful approach of the benchmark test, the results achieved by the combined solution, and the results achieved by the 2–4 most useful verification modules. In all Figures 4.5–4.30, the *true positives* (*TP*) are marked in green, which means that the green colour indicates situations where the underlying road model is capable of explaining a road in the image. In contrast, the *false negatives* (*FN*) marked in red emphasize situations where the respective road model is not capable of explaining an existing road. The *false*

<sup>3</sup>At the time of writing (November 2013), road detection results for the ISPRS benchmark test are not yet available.

rk	module	<i>completeness</i>	<i>correctness</i>
EuroSDR-Ikonos1-sub1			
1	acupuncture method	0.57 [+0.29]	0.998
2	pan-line detection	0.38 [+0.10]	0.995
3	SSH method	0.14 [+0.01]	1.000
4	NDVI-line detection	0.13 [+0.01]	0.995
5	parallel edge detection	0.07 [+0.02]	1.000
6	adjacency analysis	0.02 [+0.01]	0.841
combined solution		0.79	0.995
benchmark [Beumier & Lacroix, 2006]		0.48	0.69
EuroSDR-Ikonos3-sub1			
1	SSH method	0.70 [+0.13]	1.000
2	pan-line detection	0.65 [+0.06]	1.000
3	acupuncture method	0.25 [+0.14]	1.000
4	NDVI-line detection	0.09 [+0.00]	1.000
5	parallel edge detection	0.00 [+0.00]	n/d
5	adjacency analysis	0.00 [+0.00]	n/d
combined solution		1.00	1.000
benchmark [Bacher & Mayer, 2005]		0.81	0.87
EuroSDR-Ikonos3-sub2			
1	SSH method	0.70 [+0.25]	1.000
2	pan-line detection	0.65 [+0.13]	1.000
3	NDVI-line detection	0.45 [+0.00]	1.000
4	acupuncture method	0.09 [+0.03]	1.000
5	parallel edge detection	0.00 [+0.00]	n/d
5	adjacency analysis	0.00 [+0.00]	n/d
combined solution		0.95	1.000
benchmark, Hedmann and Hinz		0.86	0.89

Table 4.6.: Results, EuroSDR-Ikonos dataset.

*positives* (*FP*), denoting the critical errors for a verification approach, are marked in blue, whereas the rotated reference is depicted by shaded thin grey lines.

The style of the tables and figures presented in this section is in close accordance with [Mayer et al., 2006]. In the following the results for the individual scenes will be discussed separately.

**EuroSDR-Ikonos1-sub1:** This is the most challenging scene of the benchmark test. The acupuncture method, which was originally developed for an urban context and imagery with 1 m GSD, performs best. In particular, it performs better by  $\Delta = 0.57 - 0.48 = 0.09$  than the benchmark approach [Beumier & Lacroix, 2006]. The benchmark approach basically relies on a line detector that is applied to the green colour band; the detected line primitives are filtered with respect to criteria such as minimum length, maximum curvature and maximum NDVI. Concerning the *completeness*, the benchmark approach slightly outperforms the second-best module, the pan-line detection module ( $\Delta = 0.48 - 0.38 = 0.10$ ). A reason for this result can be seen by comparing both results (cf. Figures 4.5 and 4.8); the advantage of the benchmark approach is mainly related to several short roads in the settlement area. Indeed, some of these roads have also been detected by the pan-line detection module but for these line primitives the second critical assumption of the line models: *Only roads correspond to the Steger line model* (cf. Section 3.3.2) is violated with large probability. For that reason, the respective road hypotheses were assigned to the state *unknown* and do not appear as *true positives* (*TP*). All other modules achieve comparably low values for the *completeness*, which corresponds to the results for the most approaches computed in the benchmark test [Mayer et al., 2006].

The combined solution mainly benefits from the fusion of the results from the acupuncture module and the pan-line detection module, and consequently outperforms the benchmark approach by  $\Delta = 0.79 - 0.48 = 0.31$  in the *completeness*. In Figures 4.7–4.9, it can be seen that the acupuncture method is most successful in the urban part of the scene, whereas the pan-line detection module is most successful in the rural part. Four other modules achieve completeness values  $< 0.15$ , however, they still provide considerable contributions to the combined solution.

The advantage of the proposed approach over the benchmark approach is mainly related to the acupuncture method, which does not have a correspondence in the benchmark test (it was published by Youn et al. [2008] two years after the test results had been published). It can be further stated that even those modules that produce results with low *completeness* still provide results with high *correctness*, which is interpreted as a success of the proposed concept of model-uncertainty.

In conclusion, the proposed approach provides very good results when considering the problems of other approaches with that scene. In fact, 79% of the roads in the scene can be explained by the combined model. However, this result also shows the limitations: The combined model is still far away from explaining the entire scene. The missing 21% of the roads are mostly narrow roads ( $< 4$  m road width), either characterized by extremely low contrast (left picture in Figure 4.10) or by changing context conditions such that none of the enabled modules alone succeed (right picture in Figure 4.10).

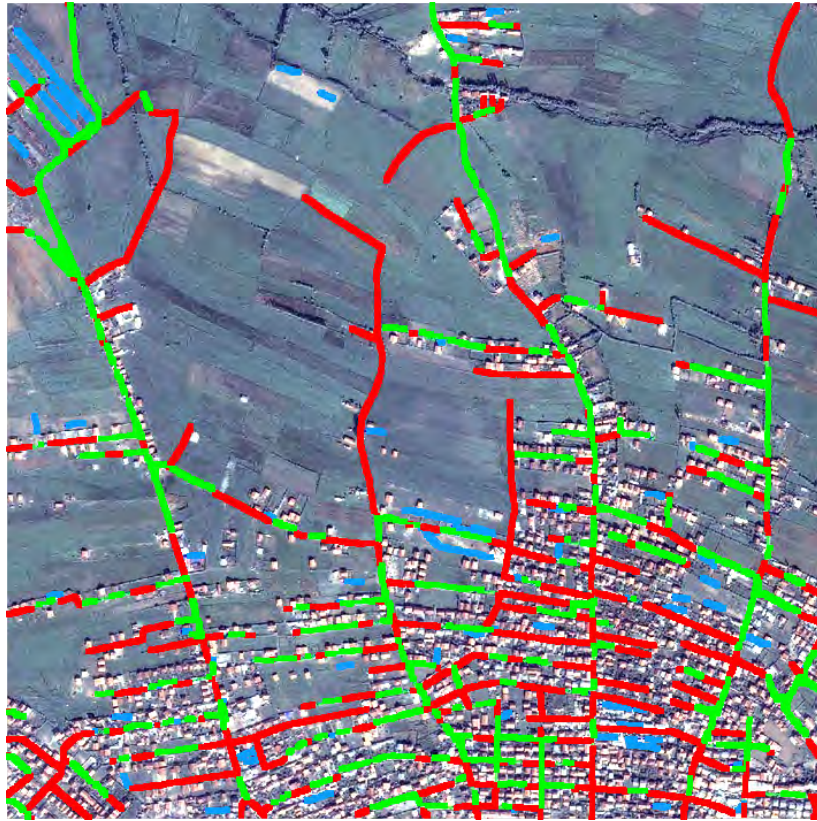


Figure 4.5.: EuroSDR-Ikonos1-sub1, [Beumier & Lacroix, 2006] (*TP* green, *FN* red, *FP* blue). The picture, originating from the benchmark test was adapted for a better visualization.

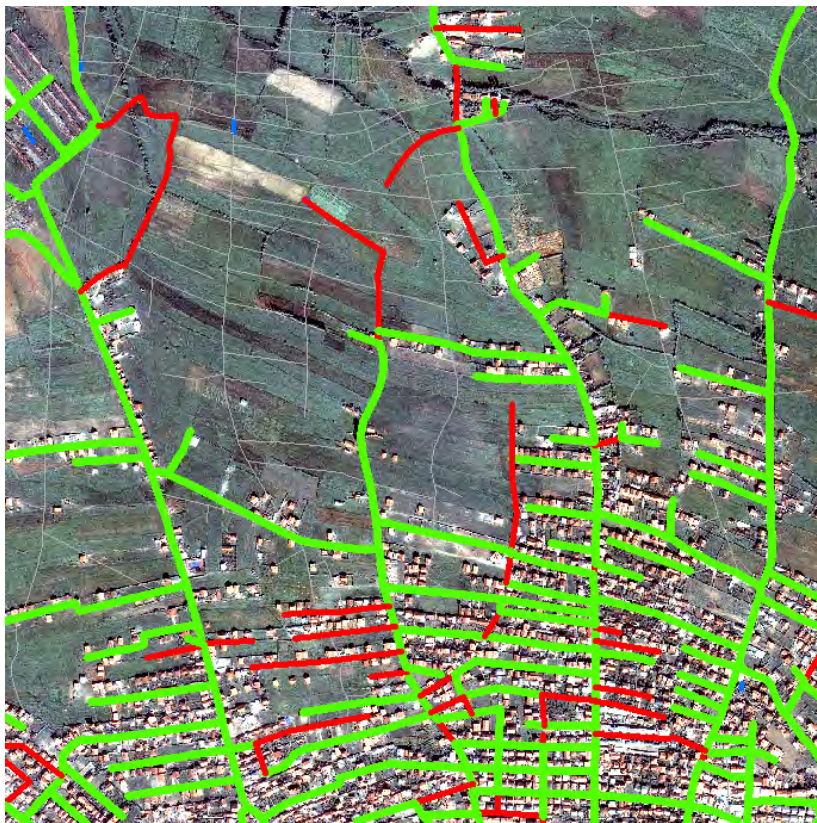


Figure 4.6.: EuroSDR-Ikonos1-sub1, combined solution (*TP* green, *FN* red, *FP* blue).



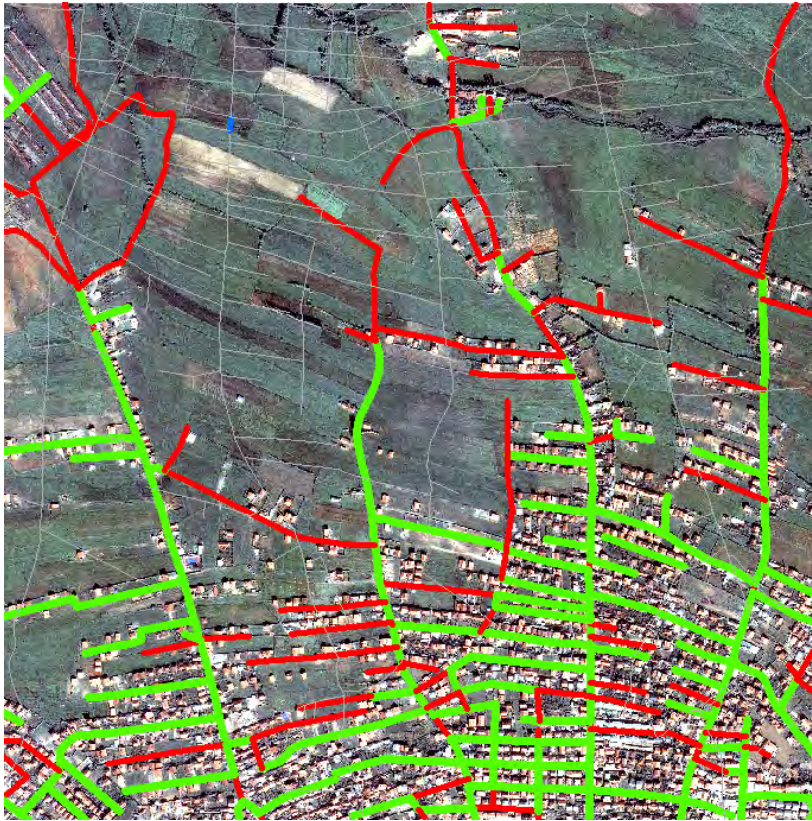


Figure 4.7.: EuroSDR-Ikonos1-sub1, acupuncture method ( $TP$  green,  $FN$  red,  $FP$  blue).

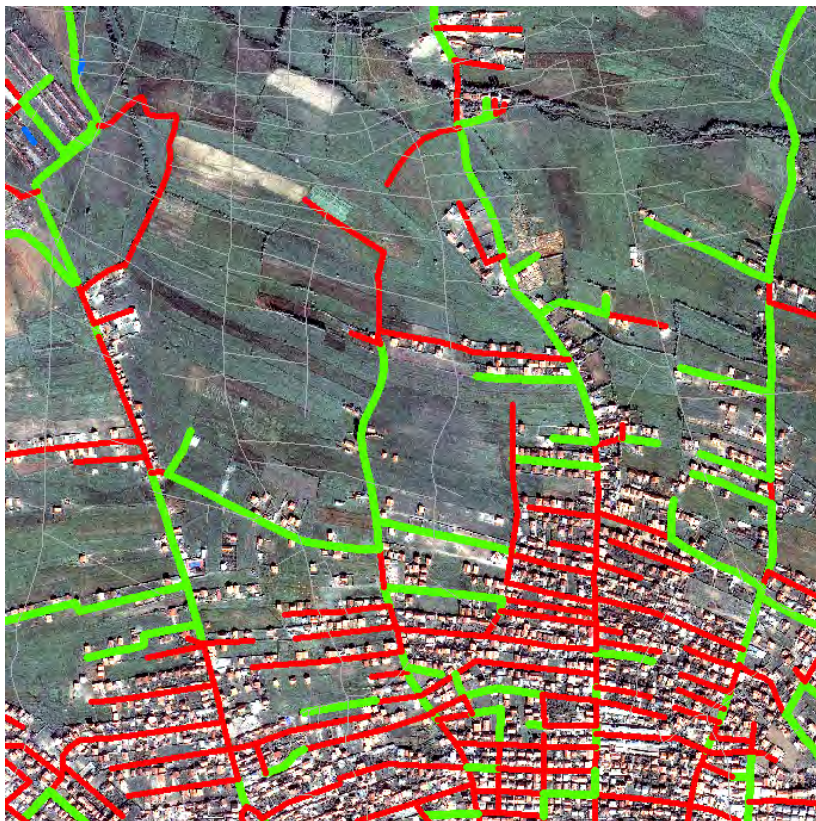


Figure 4.8.: EuroSDR-Ikonos1-sub1, pan-line detection ( $TP$  green,  $FN$  red,  $FP$  blue).



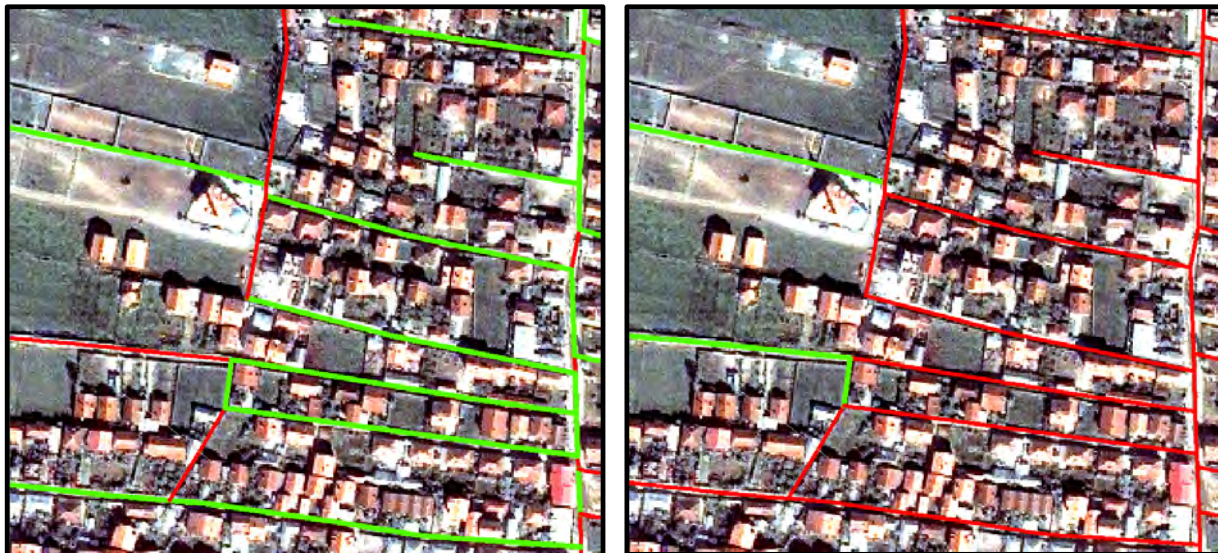


Figure 4.9.: Subsets of EuroSDR-Ikonos1-sub1 showing the transition area between settlement and open landscape ( $TP$  green,  $FN$  red) Left: Acupuncture method. Right: Pan-line detection.



Figure 4.10.: Subsets of EuroSDR-Ikonos1-sub1 showing situations where all enabled modules fail ( $TP$  green,  $FN$  dotted red). Left: Low contrast. Right: Changing context conditions.

**EuroSDR-Ikonos3-sub1&2:** The results for both image subsets turn out to be rather similar (cf. Figures 4.11–4.14). The SSH method achieves the best result with a *completeness* of 0.70, followed by the pan-line detection module with a *completeness* of 0.65 and the NDVI-line detection module with *completeness* of 0.09 and 0.45 for the two scenes. The acupuncture method achieves a *completeness* of 0.09 and of 0.25, respectively, and thus provides only solutions for the small settlements. The best benchmark results achieve a *completeness* values of 0.81 and of 0.86, respectively, which is better than any of the verification modules alone.

Both benchmark approaches, i.e. [Bacher & Mayer, 2005] for EuroSDR-Ikonos3-sub1 and *Hedmann and Hinz*<sup>4</sup> for EuroSDR-Ikonos3-sub2, basically apply the line-based approach described in [Wiedemann & Ebner, 2000]. The considerable differences are due to the definition of the input data. Whereas Wiedemann & Ebner [2000] restrict their approach to panchromatic images, both extensions use the available colour bands (RGB,IR) to extract line primitives. Bacher & Mayer [2005] artificially generate a fifth input image on the basis of fuzzy colour classification applied to the original colour bands and *Hedmann and Hinz* define the NDVI image as an additional input. The authors of the benchmark conclude that line-based models are most capable to deal with these kinds of scenes, which basically corresponds to the observations made for the verification modules. Mayer et al. [2006] further conclude that the different colour representations, i.e. colour bands, NDVI and colour classification image provide complementary information, and hence make those approaches more useful than others. This observation is not confirmed by the evaluation of the verification modules. For instance, the NDVI-line detection module did not explain any additional road hypotheses compared to the pan-line detection module. Instead, it turned out that the SSH module, which does not use colour information, can deal better with heterogeneous context areas than the line-based verification modules (cf. right part of EuroSDR-Ikonos3-sub2).

While the benchmark approaches achieved better results than the individual verification modules, the combined solution outperforms them by achieving larger *completeness*, for EuroSDR-Ikonos3-sub1 by  $\Delta = 1.0 - 0.81 = 0.19$  and for EuroSDR-Ikonos3-sub2 by  $\Delta = 0.95 - 0.86 = 0.09$ . The improvement is mainly related to the acupuncture method, which provides results for the small settlements that were most challenging for the other approaches. Finally, using the combined solution, the best results of the whole test were achieved for these two scenes (EuroSDR-Ikonos3-sub1&2), which also corresponds to the observation of Mayer et al. [2006], who say that six of the tested road detection approaches achieved a *completeness* values of approximately 0.8. Finally, the achieved results indicate that the proposed models sufficiently explain the road network in these scenes.

---

<sup>4</sup>The benchmark approach for EuroSDR-Ikonos3-sub2 was developed by *Karin Hedmann* and *Stefan Hinz* from the Institute for Photogrammetry and Cartography, Technische Universität München, Germany. The approach is detailed in [Mayer et al., 2006] on the basis of a questionnaire. Later publications by *Karin Hedmann* concentrate on road detection in SAR images, e.g. [Hedman, 2010].





Figure 4.11.: EuroSDR-Ikonos3-sub1 ( $TP$  green,  $FN$  red,  $FP$  blue). Left: The benchmark [Bacher & Mayer, 2005]. Right: The combined solution.



Figure 4.12.: EuroSDR-Ikonos3-sub1 ( $TP$  green,  $FN$  red,  $FP$  blue). Left: SSH method. Right: Acupuncture method.



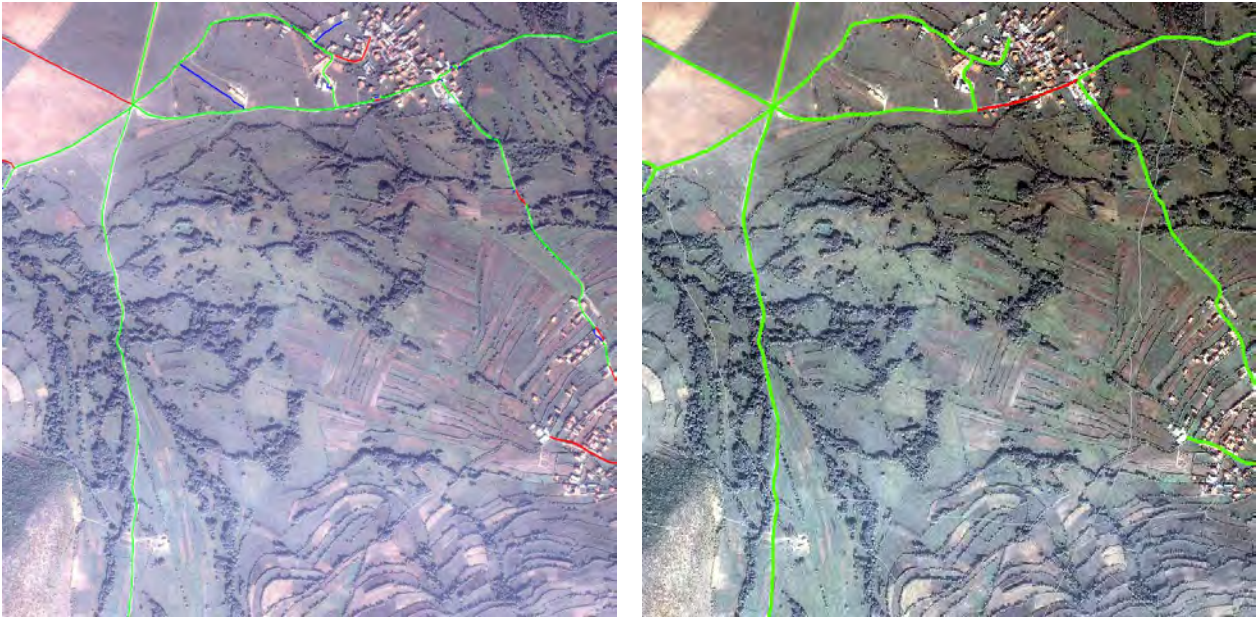


Figure 4.13.: EuroSDR-Ikonos3-sub2 (*TP* green, *FN* red, *FP* blue). Left: The benchmark Hedmann and Hinz. Right: The combined solution.

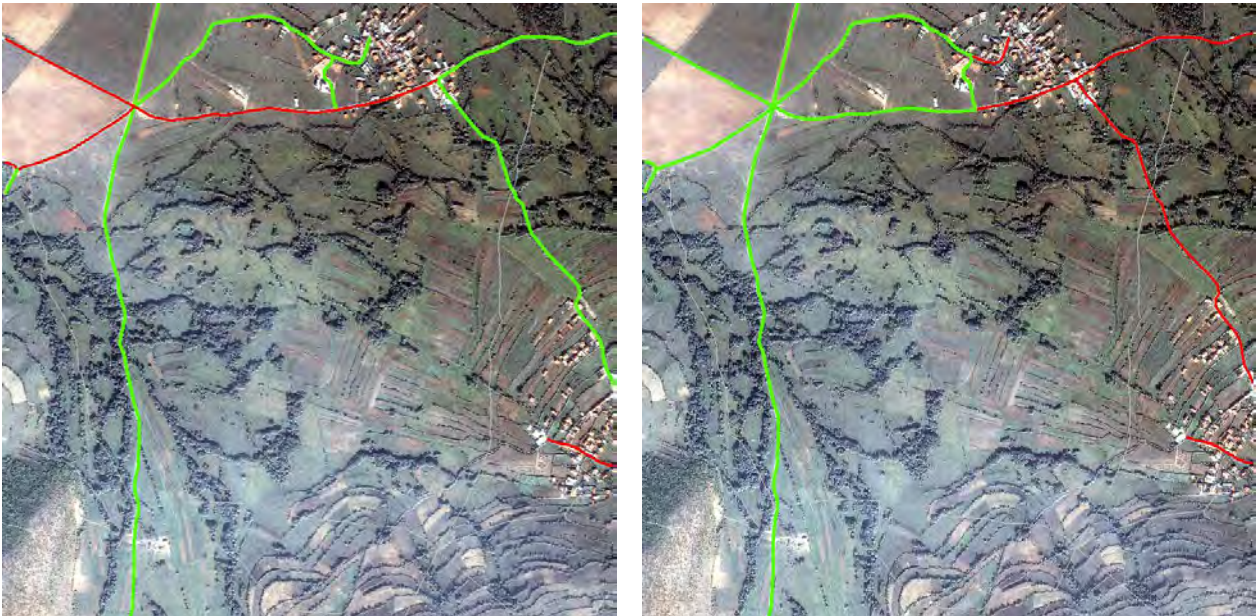


Figure 4.14.: EuroSDR-Ikonos3-sub2 (*TP* green, *FN* red, *FP* blue). Left: SSH method. Right: Pan-line detection.

rk	module	<i>completeness</i>	<i>correctness</i>
EuroSDR-Aerial1			
1	pan-line detection	0.39 [0.14]	0.995
2	colour classification	0.29 [0.08]	1.000 [+0.018]
3	SSH method	0.25 [0.05]	0.983
4	acupuncture method	0.22 [0.06]	0.970
5	parallel edge detection	0.21 [0.03]	1.000
6	adjacency analysis	0.03 [0.02]	1.000
combined solution		0.81	0.987
benchmark [Zhang & Couloigner, 2006]		0.51	0.49
EuroSDR-Aerial2			
1	pan-line detection	0.77 [0.37]	1.000
2	parallel edge detection	0.29 [0.01]	1.000
3	colour classification	0.20 [0.02]	1.000 [+0.002]
4	SSH method	0.14 [0.00]	1.000
5	acupuncture method	0.03 [0.01]	0.982
6	adjacency analysis	0.00 [0.00]	1.000
combined solution		0.85	0.999
benchmark, Gerke and Heipke		0.65	0.82
EuroSDR-Aerial3			
1	pan-line detection	0.79 [0.30]	1.000
2	parallel edge detection	0.41 [0.04]	0.998
3	SSH method	0.23 [0.01]	1.000
4	colour classification	0.04 [0.00]	1.000 [+0.002]
5	acupuncture method	0.03 [0.01]	0.969
6	adjacency analysis	0.01 [0.01]	1.000
combined solution		0.96	0.999
benchmark, Gerke and Heipke		0.72	0.77

Table 4.7.: Results, EuroSDR-Aerial dataset.

**EuroSDR-Aerial1:** This scene shows different context regions such as *sub-urban*, *rural* and *forest* areas and different types of roads such as major roads, minor roads and paths. Due to the varying appearance of roads, each module is able to solve parts of the scene (cf. Figures 4.16–4.20). Consequently, no module achieves a good result on its own (cf. Figure 4.16). This conclusion is in accordance with the EuroSDR test, where the benchmark approach published in [Zhang & Couloigner, 2006] also considers different road properties (cf. Figure 4.15). In particular *colour*, *texture* and *shape* properties are considered by different classification steps that are arranged in a sequential fusion concept (cf. Section 1.3.2). For the proposed approach, these properties are represented by the colour classification module and the SSH method. The larger *completeness* value of the combined solution compared to the benchmark ( $\Delta = 0.81 - 0.51 = 0.30$ ) can be interpreted as a success of integrating more road properties.

A difference to the results for Ikonos scenes is that the acupuncture method is not able to explain the settlements entirely because shadow effects affect its performance in many situations (cf. Figure 4.21).

Another considerable effect is the increase in *correctness* by 0.018 due to the colour classification module (cf. Table 4.7).<sup>5</sup> As the colour classification module frequently classified road hypotheses related to the rotated reference as background, those hypotheses are identified as *incorrect*.

Despite the fact that the combined solution outperforms the benchmark approach, the scene EuroSDR-Aerial1 also reveals the limitations of the proposed approach that is only able to explain 81% of the road network. It turns out that mainly one road model is missing, namely a road model that explains roads in the shadow that additionally show no colour-contrast to their surroundings (cf. Figure 4.21).

---

<sup>5</sup>In Table 4.7 the number in brackets behind *correctness* values shows the difference for *correctness* of the combined solution with and without the respective verification module.



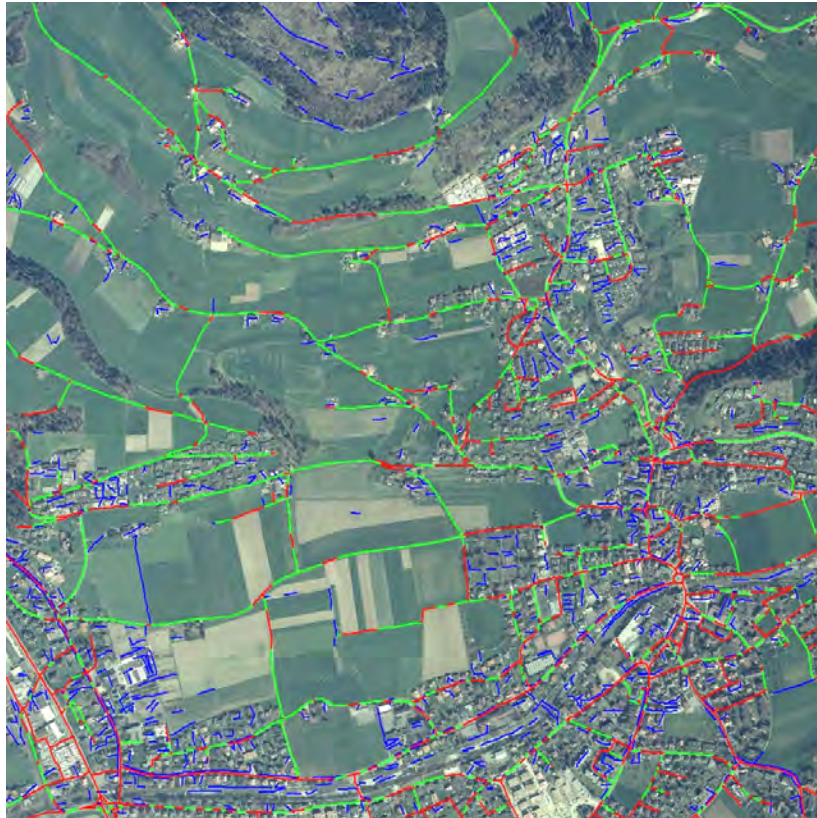


Figure 4.15.: EuroSDR-Aerial1, [Zhang & Couloigner, 2006] (*TP* green, *FN* red, *FP* blue).

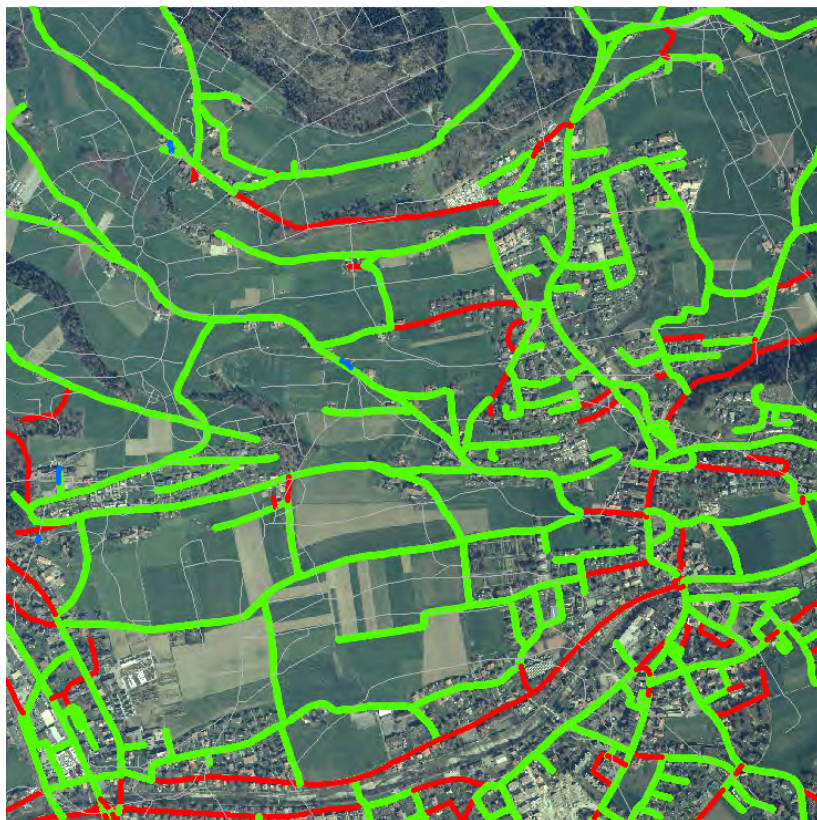


Figure 4.16.: EuroSDR-Aerial1, combined solution (*TP* in green, *FN* in red, *FP* in blue).



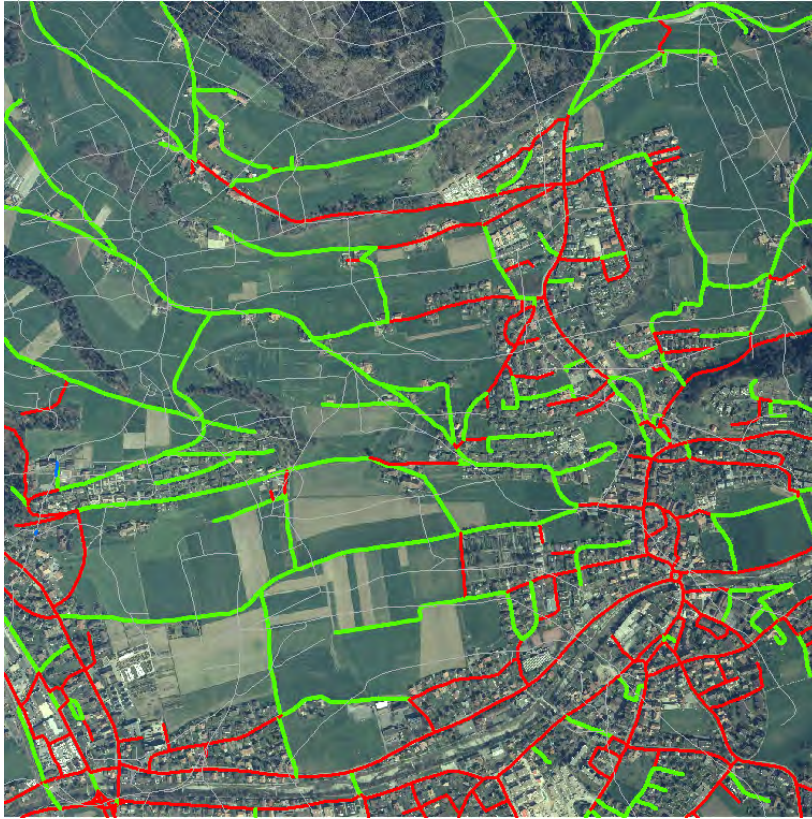


Figure 4.17.: EuroSDR-Aerial1, pan-line detection (*TP* green, *FN* red, *FP* blue).



Figure 4.18.: EuroSDR-Aerial1, colour classification (*TP* green, *FN* red, *FP* blue).





Figure 4.19.: EuroSDR-Aerial1, SSH method (*TP* green, *FN* red, *FP* blue).



Figure 4.20.: EuroSDR-Aerial1, acupuncture method (*TP* green, *FN* red, *FP* blue).



Figure 4.21.: Subsets of EuroSDR-Aerial1 (*TP* green, *FN* red, *FP* blue). Left: Acupuncture module. Right: Colour classification.

**EuroSDR-Aerial2&3:** Both scenes, depicted in Figures 4.22–4.29, mainly show rural areas, where the roads have good contrast to the background. Due to these conditions, the modules pan-line detection and parallel edge detection perform best, which corresponds to the conclusions given by Mayer et al. [2006] for the benchmark test, where the approach of *Gerke and Heipke*<sup>6</sup> performed best. Their approach combines exactly these road models, i.e. the line-based approach from [Wiedemann & Ebner, 2000] with the approach from [Baumgartner et al., 1999]. In particular, the line-based approaches are most effective for thin roads in grassland or crop land areas, whereas the parallel-edge-based approaches are most effective for the wider roads.

Except for the areas, where small forests occlude parts of the road network, the *combined solution* turns out to be nearly perfect. This also shows that the proposed approach deals well with simple situations, which is worth noting because in the EuroSDR benchmark test, the more complex approaches, i.e. those that consider many different road properties, did not work that well in the simpler scenes.

In conclusion, if one blanks out the forests, the introduced road models are mainly capable of explaining the scenes EuroSDR-Aerial 2&3. Only a few roads in the settlements, where shadows appear on the road surfaces and the contrast to the surroundings is low, could not be explained with the introduced models.

<sup>6</sup>The benchmark approach for EuroSDR-Aerial2&3 was developed by *Markus Gerke* and *Christian Heipke*, Institute for Photogrammetry and GeoInformation, Leibniz University, Hannover, Germany. The approach is detailed in [Mayer et al., 2006] on the basis of a questionnaire table. The approach is closely related but not identical to the work presented in [Gerke, 2006; Gerke & Heipke, 2008].



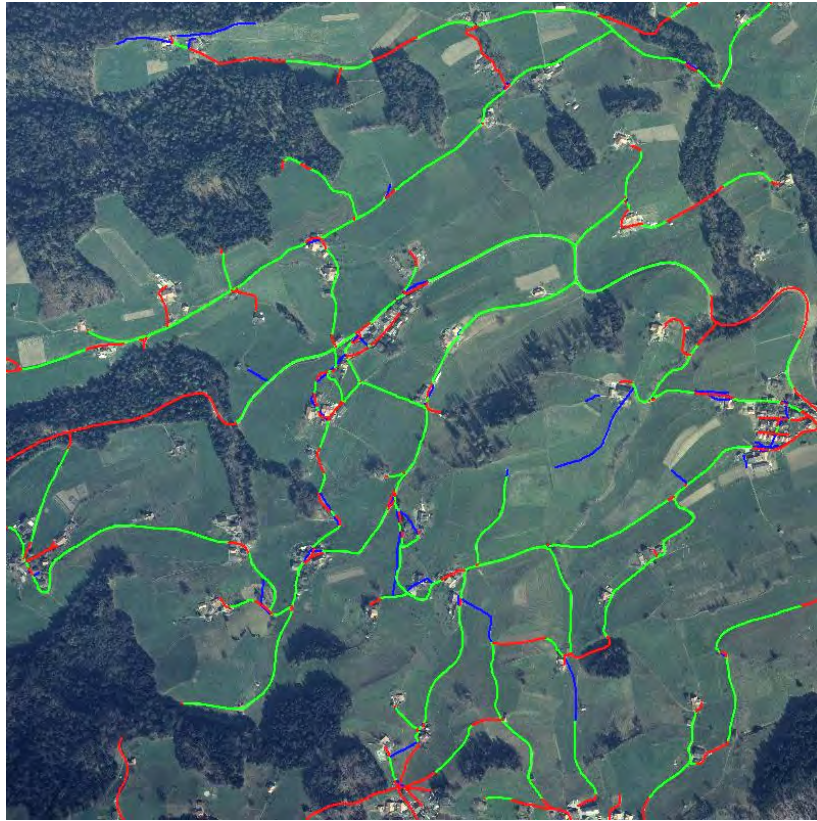


Figure 4.22.: EuroSDR-Aerial2, Gerke and Heipke (*TP* green, *FN* red, *FP* blue).

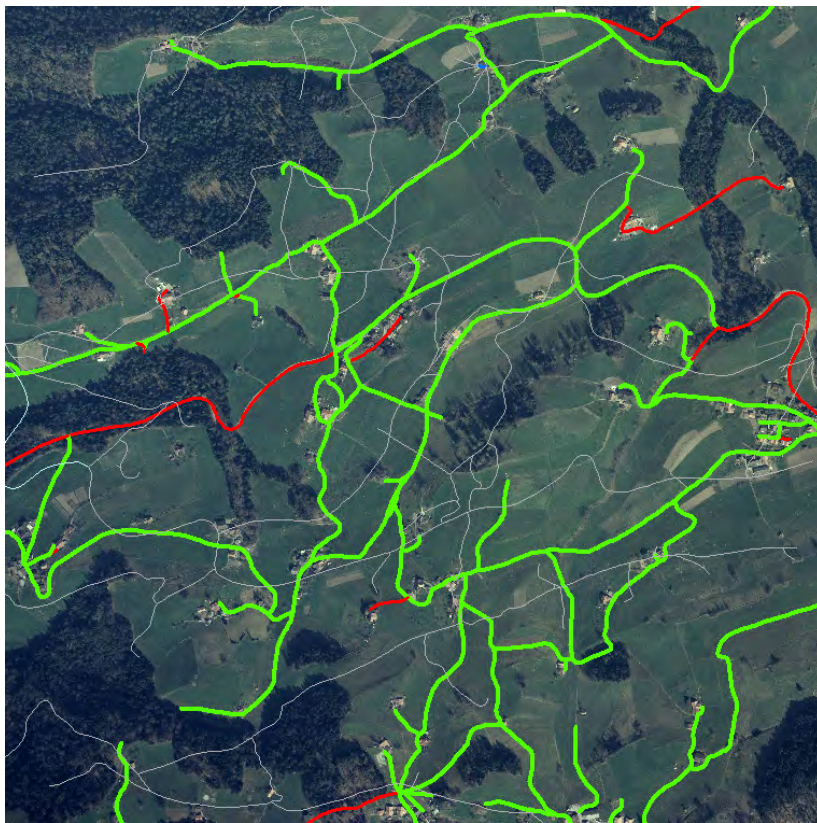


Figure 4.23.: EuroSDR-Aerial2, combined solution (*TP* green, *FN* red, *FP* blue).



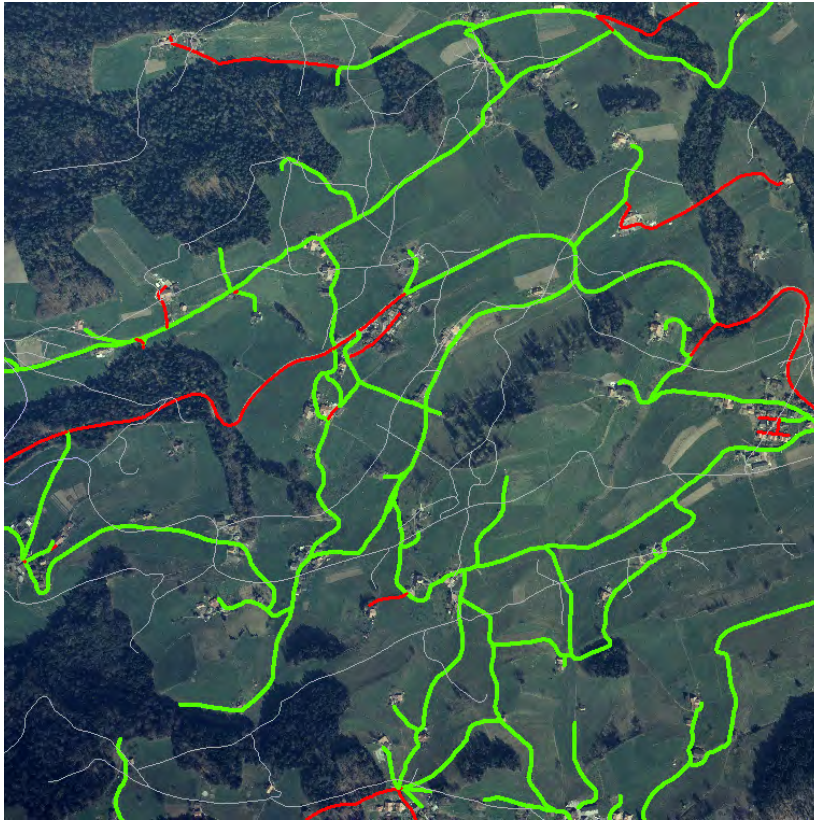


Figure 4.24.: EuroSDR-Aerial2, pan-line detection ( $TP$  green,  $FN$  red,  $FP$  blue).

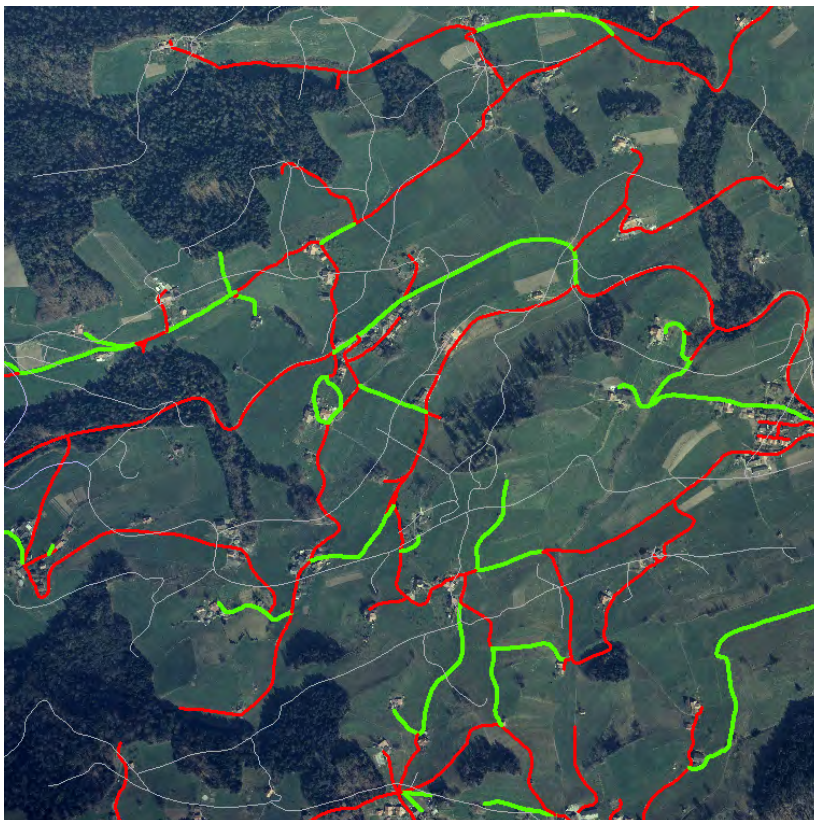


Figure 4.25.: EuroSDR-Aerial2, parallel edge detection ( $TP$  green,  $FN$  red,  $FP$  blue).





Figure 4.26.: EuroSDR-Aerial3, Gerke and Heipke (*TP* green, *FN* red, *FP* blue).



Figure 4.27.: EuroSDR-Aerial3, combined solution (*TP* green, *FN* red, *FP* blue).





Figure 4.28.: EuroSDR-Aerial3, pan-line detection ( $TP$  green,  $FN$  red,  $FP$  blue).

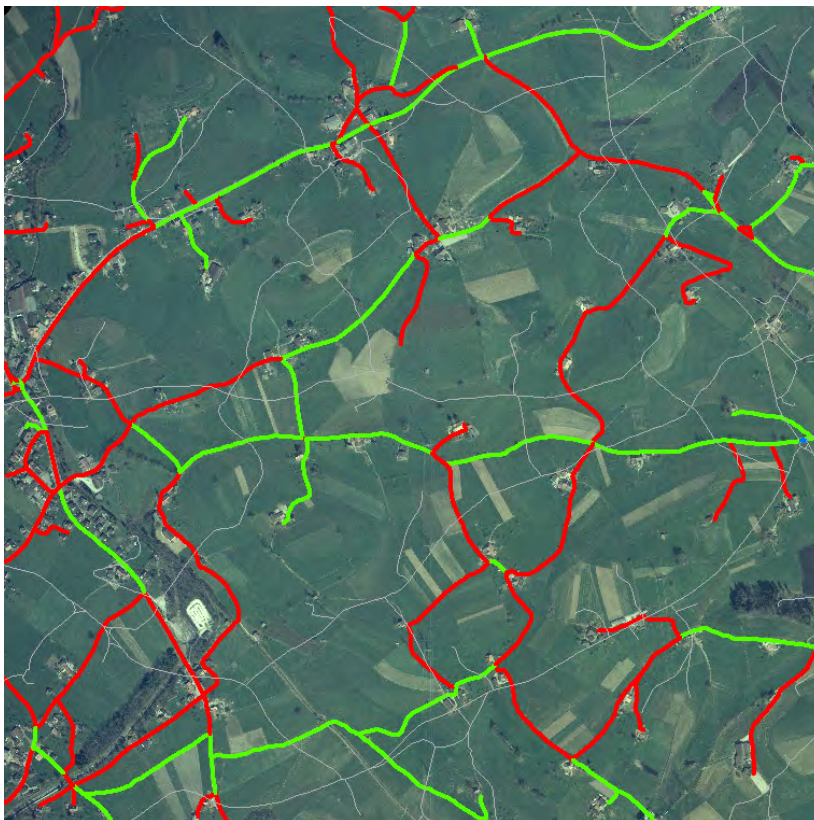


Figure 4.29.: EuroSDR-Aerial3, parallel edge detection ( $TP$  green,  $FN$  red,  $FP$  blue).

rk	module	<i>completeness</i>	<i>correctness</i>
1	acupuncture method	0.76 [+0.14]	0.995
2	nDSM-line detection	0.48 [+0.06]	0.977
3	SSH method	0.36 [+0.02]	0.982
4	colour classification	0.35 [+0.02]	0.961
5	parallel edge detection	0.30 [+0.01]	1.000
6	adjacency analysis	0.01 [+0.01]	1.000
7	NDVI-line detection	0.00 [+0.00]	n/d
7	pan-line detection	0.00 [+0.00]	n/d
7	grassland detection	0.00 [+0.00]	n/d [+0.002]
7	building detection	0.00 [+0.00]	n/d [+0.024]
combined solution		0.98	0.998

Table 4.8.: Results, ISPRS-Vaihingen.

**ISPRS-Vaihingen:** The ISPRS-Vaihingen dataset is the only benchmark dataset that allows all modules to be enabled. The best modules in this scene are the acupuncture method with a *completeness* of 0.76 and the nDSM-line detection module with a *completeness* of 0.48. Both modules were particularly developed to deal with urban areas. The acupuncture method performs slightly better for wider roads, whereas the nDSM-line detection method performs better for narrow roads (cf. Figure 4.30). The other modules still provide some valuable information so that the combined solution turns out to be nearly perfect with a *completeness* of 0.98. The building detection module and the grassland detection module identify most road hypotheses from the rotated reference as *incorrect* because nearly all of those hypotheses intersect buildings or grassland areas. Hence, by integrating these two modules the *correctness* of the combined solution increases by 0.026 from 0.972 to 0.998 (cf. Table 4.8), whereas their contribution with respect to the *completeness* is zero by design (cf. Section 3.3.6). The evaluation of the ISPRS-Vaihingen dataset shows very promising results, especially as this scene is quite challenging due to shadow effects and changing colour properties of the background.

## Discussion

The experiments presented in this section have shown that the most verification modules are capable of explaining at least some parts of the road network. It turned out that the observations concerning these dependencies correspond mainly to those made by Mayer et al. [2006]. Exceptions to this rule are the acupuncture method and the SSH method, which did not exist when the EuroSDR test was carried out. In particular, exactly these two modules considerably outperform the best methods of the benchmark in some scenes. Further significant improvements were achieved by the fusion, i.e. by the *combined solution*, which always outperformed the best methods of the benchmark test. These improvements were larger for scenes with heterogeneous properties, e.g. those that show more than one type of context areas. For four of the seven scenes a *completeness* of at least 0.95 could be achieved. For the other three scenes a *completeness* of at least 0.79 were achieved, which is still





Figure 4.30.: ISPRS-Vaihingen ( $TP$  green,  $FN$  red,  $FP$  blue). Left: Combined solution. Right: Acupuncture method.



Figure 4.31.: ISPRS-Vaihingen ( $TP$  green,  $FN$  red,  $FP$  blue). Left: nDSM-line detection. Right: SSH method.



relatively good when considering the results described in [Mayer et al., 2006], but which also leaves room for further improvement. The second objective, a constantly high *correctness* of at least 0.987 was also achieved for the combined solutions. It turned out that the *correctness* of the combined solution was mostly larger than the lowest *correctness* for the single solutions. This means that *false positives* originating from single modules were frequently compensated by other modules.

combined solution	<i>completeness</i>	<i>correctness</i>
EuroSDR-Aerial1		
complete (cf. Table 4.7)	0.8069	0.9869
without training data	0.8068	0.9883
ISPRS-Vaihingen		
complete (cf. Table 4.8)	0.9821	0.9976
without training data	0.9804	0.9972

Table 4.9.: Results of the combined solution with all test data (complete) and without the test data that overlap with the training data.

So far, not all verification modules were proven to be important for the combined solutions. In particular, it turned out that the NDVI-line detection and the parallel edge detection did not provide much complementary information compared to the pan-line detection. In contrast, the acupuncture method and the pan-line detection are most important, which can be seen by the relevantly large values in the brackets in Tables 4.7–4.8.

In two scenes (EuroSDR-Aerial1 and ISPRS-Vaihingen), the training data partially overlap the test data. Table 4.9 shows the results of the combined solution with and without the overlapping test data. As can be seen, their influence on the results is rather small (below 0,003 in all cases). Due to the design of the benchmark datasets, they may not be optimal for evaluating the ability of the proposed approach to detect realistic database errors. Furthermore, the benefit of considering the states *unknown* and *invalid* have not been demonstrated so far. Both aspects will be particularly tackled in the next section.

#### 4.2.3. The overall approach in a cartographic context

The final set of experiments investigates the practical impact of the proposed approach, i.e. the human efforts potentially saved during road database verification when using the automatic component, as well as the road database quality before and after the automated verification approach has been applied.

#### Configurations and evaluation strategy

In order to check the practical impact of the proposed approach, experiments are carried out with the three cartographic datasets MGCP-Algiers, GSI-Uruga, NGI-Zeebrugge described in Section 4.1. Table 4.1 shows the parametrisation with respect to the three datasets. The required training data will be shown below together with the results.

The reference road databases that were discussed in Section 4.1 can be directly compared to the

verification outputs of the proposed approach, which leads to the confusion matrix shown in Table 4.10. The rows denote the reference, according to which a road hypothesis can be either *correct* or *incorrect*. The columns denote the automatically assigned state of the road hypothesis *correct*, *incorrect*, *unknown* and *invalid*. Here, the analysis refers to the number of road hypotheses and not to the accumulated length of the roads as in the previous section.

		proposed approach			
		<i>correct</i>	<i>unknown</i>	<i>invalid</i>	<i>incorrect</i>
reference	<i>correct</i>	<i>CC</i>	<i>CU</i>	<i>CV</i>	<i>CI</i>
	<i>incorrect</i>	<i>IC</i>	<i>IU</i>	<i>IV</i>	<i>II</i>

Table 4.10.: Confusion matrix for the cartographic datasets.

The evaluation parameters *classification correctness* and *classification completeness* are defined on the basis of the confusion matrix (cf. Table 4.10):

$$\text{classification completeness} = \frac{CC + II + CI + IC}{CC + CU + CV + CI + IC + IU + IV + II} \quad (4.8)$$

$$\text{classification correctness} = \frac{CC + II}{CC + II + IC + CI} \quad (4.9)$$

The *classification completeness* (cf. Equation 4.8) indicates the proportion of road hypotheses that are assigned to the states *correct* or *incorrect*. The *classification correctness* (cf. Equation 4.9) indicates the proportion of correct assignments. Both parameters have an optimum at 100% but only the *classification correctness* is required to come close to this optimum to make the semi-automatic approach practically relevant.

Furthermore, according to the reference dataset the *classification correctness* is considered under two conditions: First, only *correct* road hypotheses are considered and secondly, only *incorrect* road hypotheses are considered:

$$\text{classification completeness}(\text{correct roads}) = \frac{CC}{CC + CU + CV + CI} \quad (4.10)$$

$$\text{classification completeness}(\text{incorrect roads}) = \frac{II}{IC + IU + IV + II} \quad (4.11)$$

Subsequently, the two parameters (cf. Equations 4.10 and 4.11)<sup>7</sup> will be used to show the relative importance of the verification modules for the combined solution in a similar way as in Section 4.2.2.

Another question to be answered is how good the results of the semi-automatic approach are, i.e. how good they are after a human operator has corrected the database according to the output of the automatic component. Two interactive settings are subsequently analysed and compared. The first setting (cf. Table 4.11) represents a *conservative setting* where a human operator checks every unconfirmed road hypothesis, i.e. each road hypothesis that is not explicitly assigned to the state *correct*. The second setting (cf. Table 4.12) represents a *low-effort setting* where a human operator only checks the indicated errors, i.e. only road hypotheses that are explicitly assigned to the state *incorrect*.

<sup>7</sup>Note that Equations 4.10 and 4.11 can also be interpreted as conditional probabilities  $P(\text{approach says correct} | \text{reference says correct})$  and  $P(\text{approach says incorrect} | \text{reference says incorrect})$ .

		proposed approach	
		<i>correct</i>	<i>unknown, invalid, incorrect</i>
reference	<i>correct</i>	true positives ( $TP^{\text{obj}}$ )	false negatives ( $FN^{\text{obj}}$ )
	<i>incorrect</i>	false positives ( $FP^{\text{obj}}$ )	true negatives ( $TN^{\text{obj}}$ )

Table 4.11.: Conservative setting.

		proposed approach	
		<i>correct, unknown, invalid</i>	<i>incorrect</i>
reference	<i>correct</i>	true positives ( $TP^{\text{obj}}$ )	false negatives ( $FN^{\text{obj}}$ )
	<i>incorrect</i>	false positives ( $FP^{\text{obj}}$ )	true negatives ( $TN^{\text{obj}}$ )

Table 4.12.: Low-effort setting.

In the following, it is assumed that a human operator does not commit any errors, meaning: Any erroneous road hypothesis that is forwarded to the human operator is counted as a corrected database error. It is further assumed that the human operator only concentrates on the indicated road hypotheses and not on adjacent road hypotheses or something else. The four matrix elements  $TP^{\text{obj}}, FN^{\text{obj}}, FP^{\text{obj}}, TN^{\text{obj}}$  introduced in Tables 4.11 and 4.12 have a similar meaning as the evaluation parameters  $TP, FN, FP, TN$  introduced in Table 4.5, except for the fact that they refer to numbers of road hypotheses and not to the lengths of the roads. In order to discuss the practical relevance of the proposed approach, the following evaluation parameters are defined on the basis of the four matrix elements:

$$\text{prior db quality} = \frac{TP^{\text{obj}} + FN^{\text{obj}}}{TP^{\text{obj}} + FN^{\text{obj}} + FP^{\text{obj}} + TN^{\text{obj}}} \quad (4.12)$$

$$\text{posterior db quality} = \frac{TP^{\text{obj}} + FN^{\text{obj}} + TN^{\text{obj}}}{TP^{\text{obj}} + FN^{\text{obj}} + FP^{\text{obj}} + TN^{\text{obj}}} \quad (4.13)$$

$$\text{corrected db errors} = \frac{TN^{\text{obj}}}{FN^{\text{obj}} + TN^{\text{obj}}} \quad (4.14)$$

$$\text{automation} = \frac{TP^{\text{obj}} + FP^{\text{obj}}}{TP^{\text{obj}} + FN^{\text{obj}} + FP^{\text{obj}} + TN^{\text{obj}}} \quad (4.15)$$

The *prior db quality* (cf. Equation 4.12) indicates the relative amount of correct road hypotheses before the semi-automatic verification takes place, whereas the *posterior db quality* (cf. Equation 4.13) indicates the relative amount of correct hypotheses after semi-automatic verification has been carried out. *Corrected db errors* (cf. Equation 4.14) indicates the relative amount of incorrect road hypotheses seen by the human operator, and *automation* (cf. Equation 4.15) indicates the relative amount of road hypotheses that can be ignored by a human operator during the verification.

An optimal result is characterized by a *posterior db quality* of 100%, by an *automation* of 100% minus the *prior db quality* and by 100% *corrected db errors* (if the *prior db quality* is smaller than 100%).



## Results

**MGCP-Algiers:** The left side in Figure 4.32 shows the RGB input image, which is superimposed by the database errors (red) from the manual reference dataset. The training data (blue lines) consist of road segments selected at the centre of the scene that overlap with 35 road hypotheses of the MGCP-Algiers dataset. Training samples for the *buildings* and *grassland* are not required as the respective modules are disabled (cf. Table 4.1). At the right hand side of Figure 4.32, the output of the proposed approach is visualised in a traffic light system, which means that road hypotheses assigned to the state *correct* are marked in green, road hypotheses assigned to the states *unknown* or *invalid* are marked in yellow and road hypotheses assigned to the state *incorrect* are shown in red. It turned out that throughout the scene most of the hypotheses were assigned to the state *correct*, while a far smaller part is assigned to the state *unknown* or *invalid* and only one single road hypothesis is assigned to the state *incorrect*. This single road hypothesis was misclassified by the NDVI-line detection module. The characteristics of such *misclassifications* will be analysed in detail after the results for the three cartographic datasets have been presented.

Figure 4.33 shows a subset of Figure 4.32 where most of the correct roads appearing in a challenging urban area are assigned to the state *correct*. Figure 4.34 shows another subset indicating typical database errors that appear in this scene. It becomes clear why the database errors could not be detected by the proposed approach; the database errors do not correspond to errors concerning *orientation*, *shape* or *parallel shifts* that potentially can be explained by the enabled modules (pan-line detection, NDVI-line detection, acupuncture method, SSH method). A detection of these errors would require the other (disabled) modules, e.g. grassland detection and colour classification module for which the input data of the MGCP-Algiers dataset were insufficient.

Table 4.13 shows the quality of the results for the 2,254 road hypotheses of the MGCP-Algiers dataset, excluding the 35 road hypotheses that overlap with the training data. The numbers in brackets show the proportions of the road hypotheses and the accumulated road lengths corresponding to the road hypotheses, respectively. In Table 4.13, the upper confusion matrix shows the evaluation for the automatic component. It can be seen that except for one, all class assignments are actually correct (*classification correctness*  $\approx 100\%$ ), which supports the observations for the benchmark datasets where the *correctness* was generally large ( $> 98\%$ ). However, with 69% the *classification completeness* is relatively small compared to the results for the benchmark datasets. This can be explained by the specific properties of the MGCP-Algiers dataset: The dirt roads (cf. Section 4.1) mostly have weak contrast to the surrounding crop fields and bush land. Hence, most of the verification modules do not perform well with these types of roads. Table 4.13 also shows that the 33 incorrect road hypotheses are all assigned to the state *unknown*, which means that none of the database errors is explicitly detected by the automatic component alone.

The two subsequent confusion matrices in Table 4.13 show the evaluation of the results achieved by the semi-automatic approach. Most notably, in the *conservative setting* 100% of the database errors are corrected because the human operator investigates all road hypotheses automatically assigned to the state *unknown*. This also means that 31% of the road hypotheses have to be manually investigated, which leads to the automation ratio of only 69%. The *low-effort setting* achieves an automation ratio of nearly 100% but does not lead to any database enhancement. This fact seems

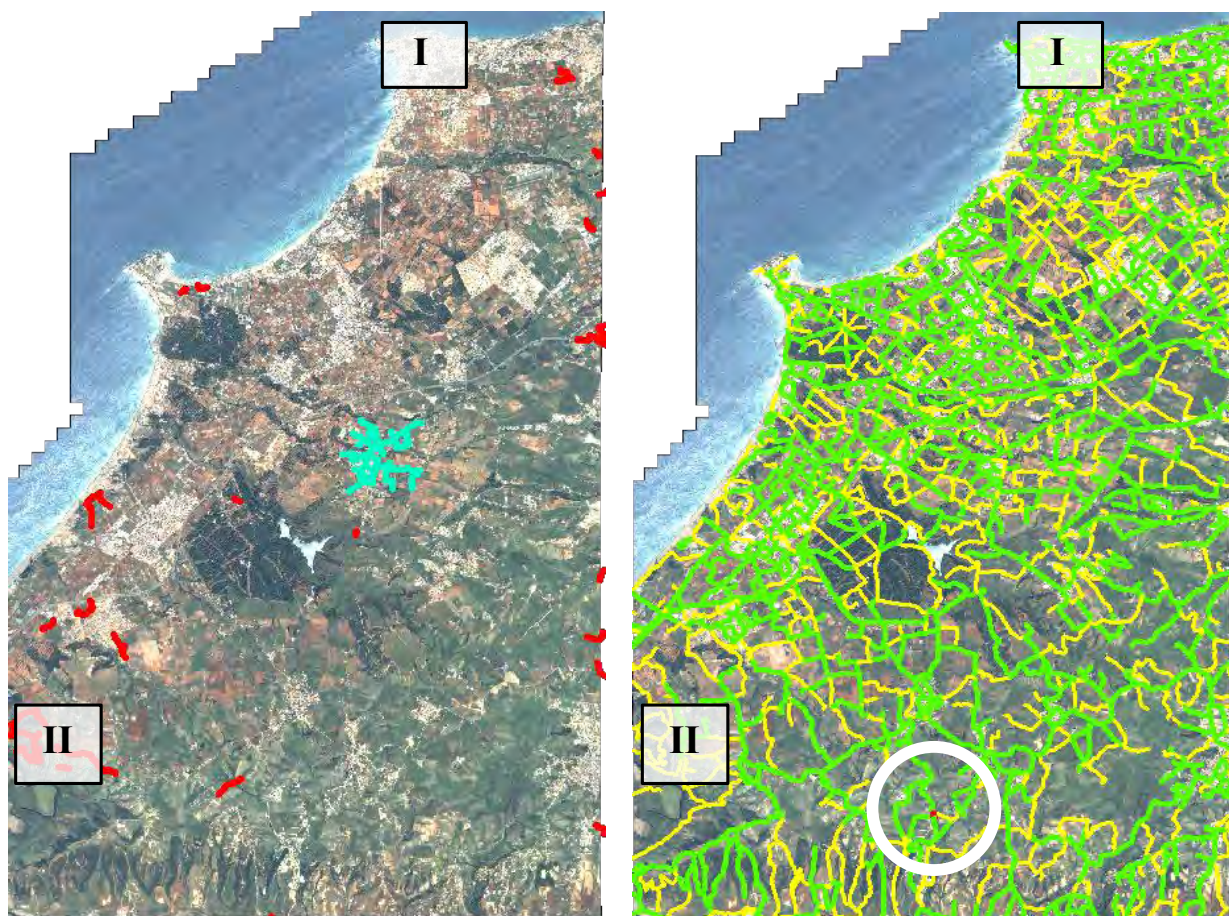


Figure 4.32.: MGCP-Algiers dataset. Left: Database errors indicated by the reference (red) and training data (cyan). Right: The verification results (*correct* in green, *unknown* and *invalid* in yellow, *incorrect* in red with white circle). Box I is detailed in Figure 4.33. Box II is detailed in Figure 4.34.

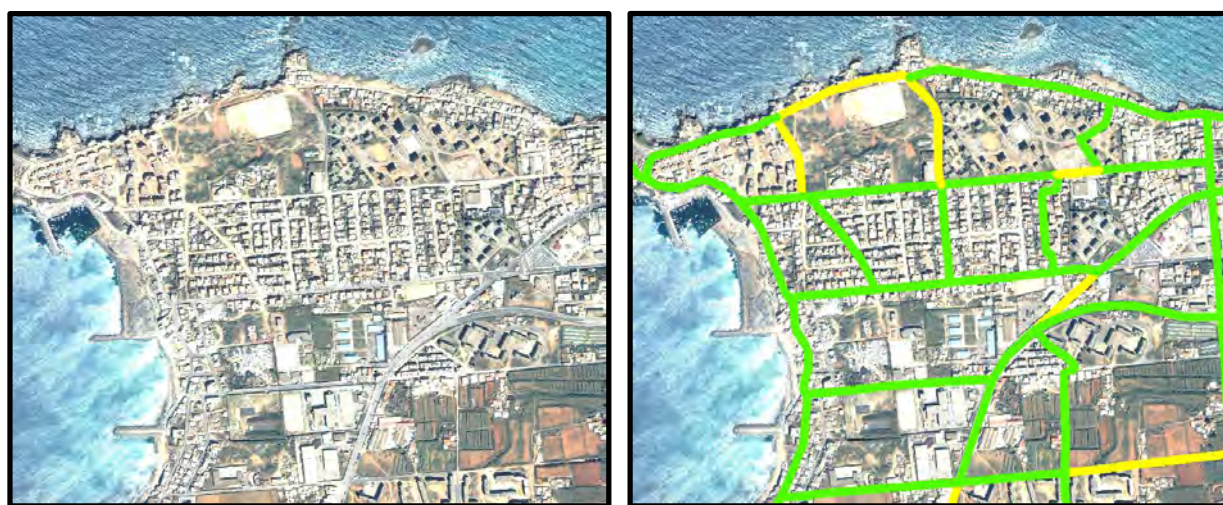


Figure 4.33.: Subsets of MGCP-Algiers showing a typical verification result in an urban area. Left: Input image. Right: Result of the proposed approach.



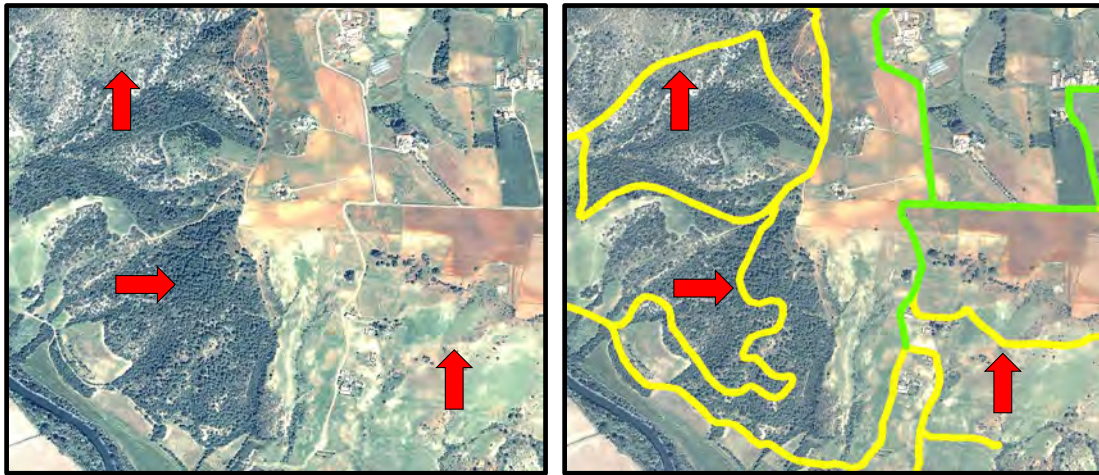


Figure 4.34.: Subsets of MGCP-Algiers showing some database entries (red arrows) that do not have a correspondence in the image. Left: Input image. Right: Verification result.

		proposed approach			
reference	automatic component:	<i>correct</i>	<i>unknown</i>	<i>invalid</i>	<i>incorrect</i>
	<i>correct</i>	1,546 (69%) [449 km]	674 (30%) [275 km]	0 [0 km]	1 (0.04%) [0.068 km]
	<i>incorrect</i>	0 [0 km]	33 (1.5%) [13 km]	0 [0 km]	0 [0 km]
	conservative setting:	<i>correct</i>	<i>unknown,</i>	<i>invalid,</i>	<i>incorrect</i>
	<i>correct</i>	1,546 (69%) [449 km]	675 (30%) [275 km]		
	<i>incorrect</i>	0 [0 km]	33 (1.4%) [13 km]		
	low-effort setting:	<i>correct,</i>	<i>unknown,</i>	<i>invalid</i>	<i>incorrect</i>
	<i>correct</i>	2,220 (98.5%) [734 km]			1 (0.04%) [0.068 km]
	<i>incorrect</i>	33 (1.4%) [13 km]			0 [0 km]
		<i>classification completeness</i> $\approx 69\%$ <i>classification correctness</i> $\approx 100\%$  <i>prior db quality</i> $\approx 98.5\%$ <i>posterior db quality</i> = 100% <i>automation</i> $\approx 69\%$ <i>corrected db errors</i> = 100%			
		<i>prior db quality</i> $\approx 98.5\%$ <i>posterior db quality</i> $\approx 98.5\%$ <i>automation</i> $\approx 100\%$ <i>corrected db errors</i> = 0%			

Table 4.13.: MGCP-Algiers dataset, confusion matrices and evaluation parameters.

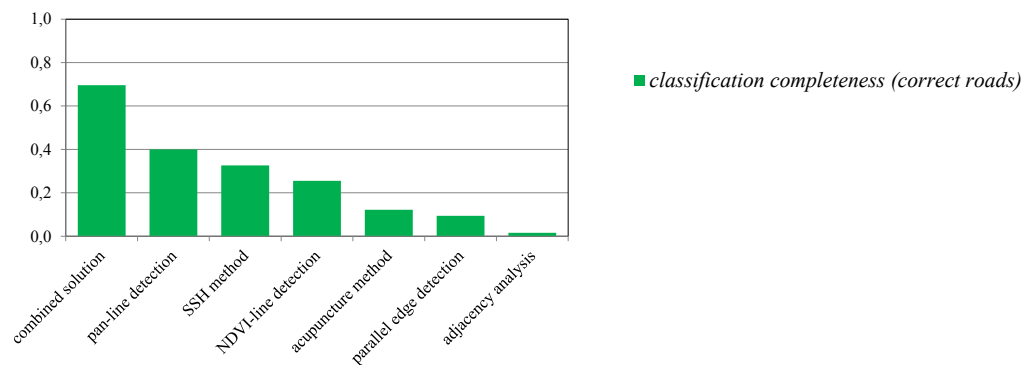


Figure 4.35.: MGCP-Algiers, success of the individual verification modules.

to be odd, but with 98.5% the *posterior db quality* is relatively large, and therefore might already be sufficient for many cartographic tasks. A general discussion of the two settings will be given at the end of this section.

The diagram in Figure 4.35 shows the conditional classification correctness (cf. Equations 4.10 and 4.11) for each verification module applied to the MGCP-Algiers dataset in comparison with the combined solution. It turned out that the pan-line detection, the SSH method and the NDVI-line detection cover the largest proportions of the correct road hypotheses. In particular, the SSH method turned out to be quite successful with the dirt roads that have weak contrast to their surroundings. In correspondence with the observations made in Section 4.2.2, the acupuncture method is successful in urban areas that are however, of minor relevance in the MGCP-Algiers dataset. The parallel edge detection module is able to explain many of the larger multi lane roads that appear in this dataset. This road type is hardly explained by any other road model, and thus the parallel edge detection turned out to be rather relevant for the combined solution.

**GSI-Uraga:** Figure 4.36 shows the results for the GSI-Uraga dataset. At the left side of Figure 4.36, the RGB input image is superimposed by the reference database errors in red. It can be seen that the majority of the database errors is located in the vicinity of the redevelopment zone that stretches from the north-west to the south-east of the scene. The training data, mainly important for the *colour classification module*, are basically the same as those for the experiments described in Section 4.2.1 (cf. Figure 4.2). Only four image regions selected in the north-west of the scene were added to represent the redevelopment zone as a part of the background class. The right part of Figure 4.36 shows the output of the proposed approach, i.e. hypotheses assigned to the states *correct* in green, *unknown* or *invalid* in yellow and *incorrect* in red.

Figure 4.37 shows a subset of Figure 4.36, including a part of the redevelopment zone with the four additional training regions. Figure 4.38 shows a subset of the centre of the scene, where despite the considerable shadow effects, most of the correct road hypotheses are successfully assigned to the state *correct*.

The confusion matrix in the upper row of Table 4.14 shows the evaluation results for the automatic component with the GSI-Uraga dataset, excluding the 36 road hypotheses that overlap with the training data. With 93%, the *classification correctness* is smaller compared to the results of the MGCP-Algiers dataset. This is mainly related to a specific problem with the colour classification module, which misclassified narrow road regions as *background* due to occlusions by buildings. The reason for those occlusions is mostly due to an oblique viewing angle from the sensor to these roads, which will be discussed later. The *classification completeness* amounts to 75%, and thus is larger than the results of the MGCP-Algiers dataset but still smaller than those for the benchmark datasets (cf. Section 4.2.2).

The second matrix in Table 4.14 shows the evaluation for the *conservative setting* that leads to a *posterior db quality* of 97.5% with an *automation* of 66%. This also means a significant enhancement of the database quality from 81.7% to 97.5% by correcting 86% of the database errors. The *low-effort setting* achieves an automation of 91%, while *posterior db quality* of 88.1% is much smaller than the case of the *conservative setting*. Here, only 35% of the database errors are corrected, which means an enhancement of the database quality from 81.7% to 88.1%.



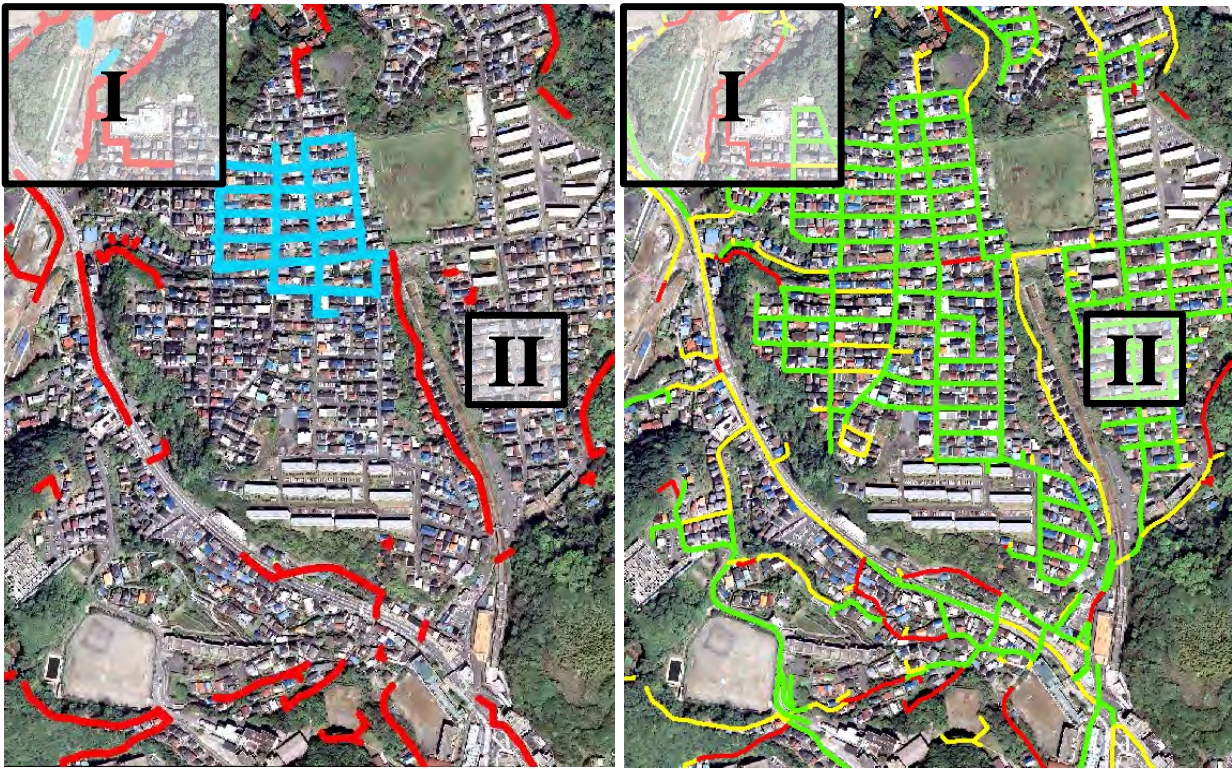


Figure 4.36.: GSI-Uraga. Left: Database errors indicated by the reference (red) and training data (cyan). Right: The verification results (*correct* in green, *unknown*, *invalid* in yellow, *incorrect* in red). Box I is detailed in Figure 4.37. Box II is detailed in Figure 4.38.



Figure 4.37.: Subset of GSI-Uraga showing a redevelopment area under construction. Left: Image with referenced errors (red) and training samples (cyan). Right: Result of the proposed approach.





Figure 4.38.: Subset of GSI-Uraga showing most of the road hypotheses are verified despite significant shadow effects. Left: Input image. Right: Verification result.

		proposed approach				
reference	automatic component:	<i>correct</i>	<i>unknown</i>	<i>invalid</i>	<i>incorrect</i>	
	<i>correct</i>	229 (64%) [9 km]	54 (15%) [2 km]	1 (0.3%) [0.02 km]	10 (2.8%) [0.3 km]	<i>classification completeness</i> $\approx$ 75%
	<i>incorrect</i>	9 (2.5%) [0.2 km]	34 (9%) [0.9 km]	0 [0 km]	23 (6.4%) [0.9 km]	<i>classification correctness</i> $\approx$ 93%
	conservative setting:	<i>correct</i>	<i>unknown,</i>	<i>invalid,</i>	<i>incorrect</i>	
	<i>correct</i>	229 (64%) [9 km]		65 (18%) [2.3 km]		<i>prior db quality</i> $\approx$ 81.7%
	<i>incorrect</i>	9 (2.5%) [0.2 km]		57 (16%) [1.9 km]		<i>posterior db quality</i> $\approx$ 97.5%
						<i>automation</i> $\approx$ 66%
						<i>corrected db errors</i> $\approx$ 86%
	low-effort setting:	<i>correct,</i>	<i>unknown,</i>	<i>invalid</i>	<i>incorrect</i>	
	<i>correct</i>		284 (79%) [11.0 km]		10 (2.8%) [0.3 km]	<i>prior db quality</i> $\approx$ 81.7%
	<i>incorrect</i>		43 (11.9%) [1.2 km]		23 (5.8%) [0.9 km]	<i>posterior db quality</i> $\approx$ 88.1%
						<i>automation</i> $\approx$ 91%
						<i>corrected db errors</i> $\approx$ 35%

Table 4.14.: GSI-Uraga, confusion matrices and evaluation parameters.

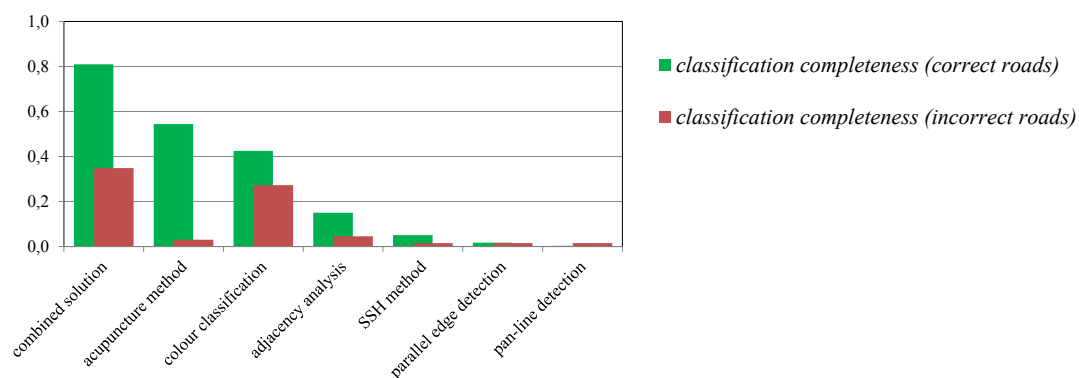


Figure 4.39.: GSI-Uraga, success of the individual verification modules.



The diagram in Figure 4.39 shows the success of the verification modules in the GSI-Uraga scene. The results are consistent with the evaluation for the benchmark datasets in urban context areas (cf. Section 4.2.2). The acupuncture method is able to explain approximately 58% of the roads, which is more than any other verification module. The colour classification module is the only (enabled) module that is designed to deal with shadow effects on the road surface. Accordingly, the colour classification module is also quite successful in this scene as it is able to explain 42% of the roads. The comparably good result of the adjacency analysis is related to the relatively large parameter  $S^{\text{change}}$ . Further, it can be seen that the colour classification method successfully detects approximately 22% of the database errors in that scene. The acupuncture method and the adjacency analysis also detect some of the database errors, so that the combined solution achieves a detection ratio of 35%.

**NGI-Zeebrugge:** The image at the top of Figure 4.40 shows an RGB<sup>8</sup> image superimposed by the database errors indicated in the reference. Due to the large temporal discrepancy of the road database and the imagery of 14 years, many database errors related to changes in the road network can be found. The majority of the database errors appears inside or near the larger settlements. The training data are derived from the existing reference database in the centre of the scene, which is detailed in Figure 4.41. The image at the bottom of Figure 4.40 shows the verification results of the automatic component, which is predominately green (road hypotheses assigned to the state *correct*) throughout the scene.

The subsets depicted in Figures 4.42 and 4.43 show typical database errors that occur in the scene. Thanks to the available infrared band and the nDSM, such database errors are explicitly detected by the building detection and the grassland detection modules that could not be used for the datasets MGCP-Algiers and GSI-Uraga.

The confusion matrix at the top of Table 4.15 shows the evaluation results for the automatic component with the NGI-Zeebrugge dataset, accounting, for the exclusion of the 39 road hypotheses that overlap with the training data. The *classification completeness* amounts to 82% and the *classification correctness* to 96%. One reason for that good result is the high quality of the input imagery that has relatively small GSD (0.5 m) and contains an infrared band and an nDSM, which means that all ten verification modules could be enabled (cf. Table 4.1). Only about 3% of the road hypotheses are misclassified (cf.  $CI, IC$  in Table 4.10), which will be discussed later on.

The second matrix in Table 4.15 shows the evaluation for the *conservative setting* that leads to a *posterior db quality* of 99.3% with an *automation* of 74%. The database quality could be enhanced from 90.2% to 99.3%, which means that 93% of the database errors could be corrected. The third matrix in Table 4.15 shows the evaluation for the *low-effort setting*, which still allows correcting 55% of the database errors with a degree of *automation* of 92%. This results in an improvement of the database quality from 90.2% to 95.5%, which might be adequate for many cartographic tasks.

The diagram in Figure 4.44 shows the success rates of the individual verification modules are more similar than for the datasets MGCP-Algiers and GSI-Uraga. The main reason for this might be the heterogeneity of the scene that contains by far the most road hypotheses. The NGI-Zeebrugge dataset

<sup>8</sup>For a better visualisation, an RGB image mosaic from 2008 was selected instead of the IRRG image from 2011 that was used for the computation (cf. Figure A.11).

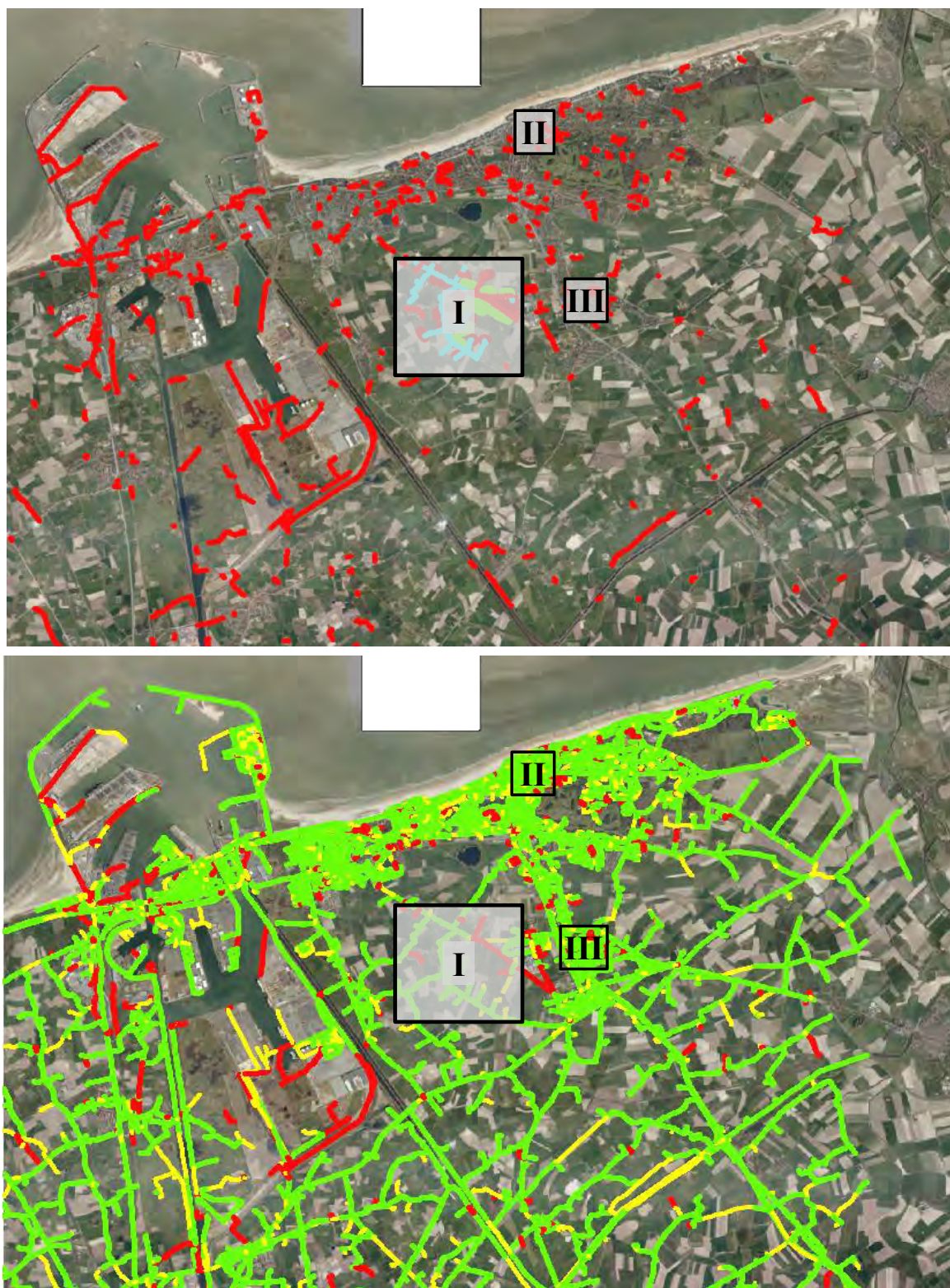


Figure 4.40.: NGL-Zeebrugge. Top: Database errors indicated by the reference in red. Bottom: The verification results (*correct* in green, *unknown*, *invalid* in yellow, *incorrect* in red). The training data (Box I) are detailed in Figure 4.41, while Figure 4.42 (Box II) and Figure 4.43 (Box III) show further interesting subsets in more detail.





Figure 4.41.: Subsets of NGI-Zeebrugge showing the training dataset (road segments in blue, buildings in red, grassland in green).

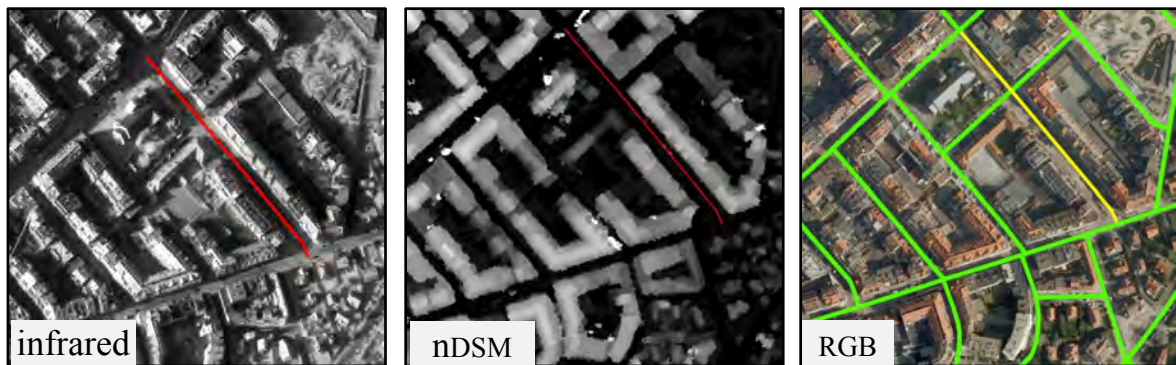


Figure 4.42.: A subset of NGI-Zeebrugge dataset showing a road hypothesis (red) that is *incorrect* because the former road has been relocated by over 8 m from its original position. At its former location, a centre strip (grassland) has been created. The automatic component assigns the road hypothesis to the state *invalid* (yellow), because grassland module and the nDSM-line detection module provide contradicting information. All other road hypotheses in the subset are assigned to the state *correct* (green).

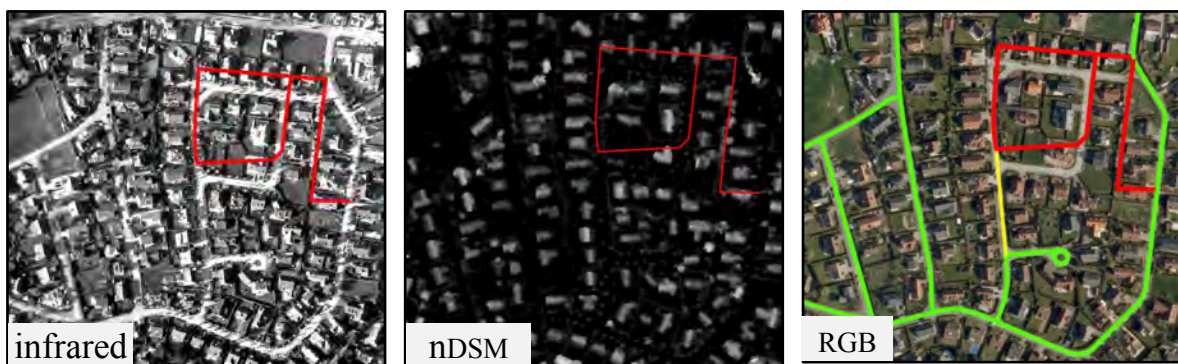


Figure 4.43.: Subsets of NGI-Zeebrugge showing several *incorrect* road hypotheses because of a recently developed building area that led to a new positioning of the roads.

reference	proposed approach					
	automatic component:	<i>correct</i>	<i>unknown</i>	<i>invalid</i>	<i>incorrect</i>	<i>classification completeness</i> $\approx$ 82% <i>classification correctness</i> $\approx$ 96%
	<i>correct</i>	4,263 (74%) [454 km]	782 (13%) [75 km]	53 (0.9%) [7.5 km]	129 (2.2%) [8.8 km]	
	<i>incorrect</i>	39 (0.7%) [2.3 km]	168 (3%) [17 km]	51 (0.9%) [8 km]	312 (5.4%) [38 km]	
	conservative setting:	<i>correct</i>	<i>unknown,</i>	<i>invalid,</i>	<i>incorrect</i>	
	<i>correct</i>	4,263 (74%) [454 km]	964 (16.6%) [9 km]			<i>prior db quality</i> $\approx$ 90.2% <i>posterior db quality</i> $\approx$ 99.3%
	<i>incorrect</i>	39 (0.7%) [2.3 km]	531 (9.2%) [63 km]			<i>automation</i> $\approx$ 74% <i>corrected db errors</i> $\approx$ 93%
	low-effort setting:	<i>correct,</i>	<i>unknown,</i>	<i>invalid</i>	<i>incorrect</i>	<i>prior db quality</i> $\approx$ 90.2% <i>posterior db quality</i> $\approx$ 95.5% <i>automation</i> $\approx$ 92% <i>corrected db errors</i> $\approx$ 55%
	<i>correct</i>	5,098 (88%) [536 km]			129 (2.2%) [8.8 km]	
	<i>incorrect</i>	258 (4.5%) [28 km]			312 (5.4%) [38 km]	

Table 4.15.: NGI-Zeebrugge, confusion matrices and evaluation parameters.

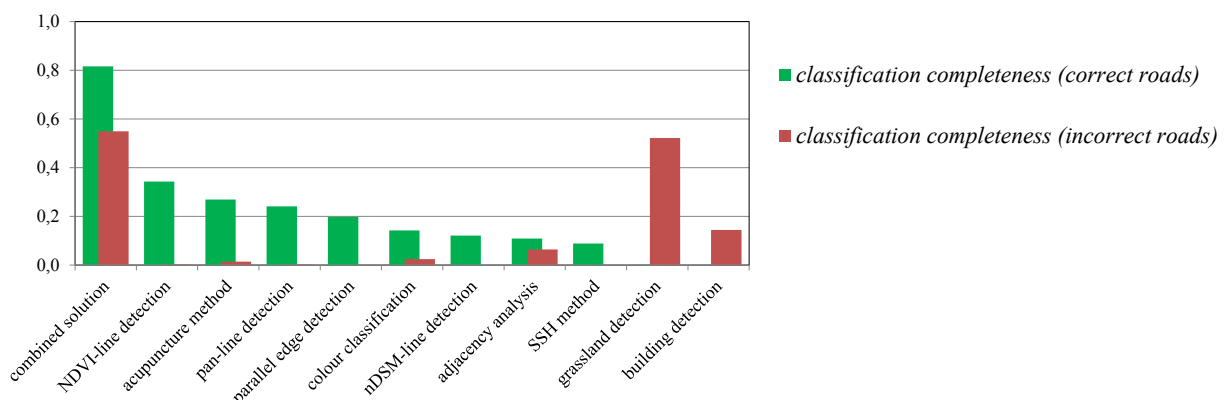


Figure 4.44.: NGI-Zeebrugge, success of the individual verification modules.

shows different context areas, e.g. the inner city of Zeebrugge with numerous commercial and sub-urban areas in the surroundings, but also rural areas with smaller and larger villages. Furthermore, many different road types can be found in the scene, e.g. narrow side roads in sub-urban areas, main roads with a lot of road markings but also regional and inter-regional roads in the rural areas. Three interesting aspects can be seen from the diagram in Figure 4.44: First, the importance of the NDVI-line detection module, which is in contrast to the observations from the benchmark datasets in Section 4.2.2, and secondly, the relative success of the adjacency analysis module even though the parameter  $S^{\text{change}}$  was set to a relatively small value ( $S^{\text{change}} = 30$  m; cf. Table 4.1), which can be explained by the large number of short road objects that occur as parts of complex junctions. Thirdly, the grassland detection module alone is able to explain more than 50% of the database errors. This is possible because new junctions and extensions of single lane roads to multi-lane roads are frequently related to the planting of grass in the vicinity of the new roads.

### Performance of the automatic component

In this section the results achieved by automatic components will be assessed before the performance of semi-automatic approaches (*conservative* and *low-effort settings*) will be discussed.

**Matched pairs:** This category encompasses all successful class assignments, i.e. *correct* road hypotheses that are automatically assigned to the state *correct* and *incorrect* road hypotheses that are automatically assigned to the state *incorrect*. Considering the three test datasets, the proportion of *matched pairs* increases with the number of the applied modules. It is 69% for MGCP-Algiers (six modules), 70% for GSI-Uraga (seven modules) and 79% for NGI-Zeebrugge (ten modules).<sup>9</sup> Some typical examples were already shown in Figures 4.33, 4.37, 4.38, 4.42 and 4.43. All these examples represent situations that comply with at least one of the introduced models.

**Misclassifications:** This category encompasses all wrong class assignments. Basically, two types of *misclassifications* can be distinguished: First, *type-I errors* where *correct* road hypotheses are assigned to the state *incorrect*, and secondly, *type-II errors* where *incorrect* road hypotheses are assigned to the state *correct*. The effect of both types of errors is similar to the *conservative* and the *low-effort settings*, i.e. in both settings, type-I errors decrease the *automation* (cf. Equation 4.15) and type-II errors decrease the *posterior db quality* (cf. Equation 4.13). For this reason Gerke & Heipke [2008], rated type-II errors to be more critical than type-I errors. However, neither Gerke & Heipke [2008] nor the approach proposed in this thesis include a specific strategy to minimize type-II errors in favour of type-I errors. Nevertheless, the errors are not balanced, i.e. it turned out that more than 75% of the misclassifications correspond to type-I errors and less than 25% to type-II errors. Both types of errors are related to particular properties of the verification modules that will be discussed in the following.

**Misclassifications – type-I errors:** Figure 4.45 shows three examples for the most typical type-I errors that were observed during the experiments. Figure 4.45a shows a situation where the acupuncture method detects a parallel shift, because the main road, affected by a traffic jam, matches perfectly

<sup>9</sup>The proportion of *matched pairs* corresponds to the product of *classification completeness* and *classification correctness*.

with its background model, whereas its side lane better matches the road model. Consequently, the acupuncture method detects a parallel shift and assigns the *correct* road hypothesis to the state *incorrect*. Figure 4.45b shows a situation in which the colour classification module misclassified a narrow road region as *background* due to occlusions by buildings.<sup>10</sup> This problem occurred rather frequently in the GSI-Uraga dataset and is mainly responsible for the type-I errors shown in Table 4.14. Figure 4.45c shows a typical problem found in the NGI-Zeebrugge dataset where the building detection module misclassified vehicles on the road surface as *buildings*. The problem is related to inaccurate nDSM heights and mainly responsible for a large proportion of the type-I errors in the NGI-Zeebrugge dataset (cf. Table 4.14).<sup>11</sup>

**Misclassifications – type-II errors:** Such errors can hardly be categorized as they appear for very different reasons. Three examples of type-II errors are depicted in Figure 4.46. Figure 4.46a shows a situation in which the adjacency analysis module incorrectly assigns a short road object to the state *correct* because all adjacent road hypotheses are *correct*. In Figure 4.46b the acupuncture method does not indicate the changed road direction and in Figure 4.46c an omitted road object randomly matches with a linear structure in a carport at Zeebrugge harbour. This carport matches with the road model of the parallel edge detection module, and hence the *incorrect* road hypothesis is assigned to the state *correct*.

**Assignment to the state *unknown*:** This category describes situations that do not fit well to any of the introduced road models. With respect to the *conservative setting*, the proportion of hypotheses assigned to the state *unknown* affects the *automation* (cf. Equation 4.15), whereas in the *low-effort setting* the proportion of hypotheses assigned to the state *unknown* affects the *posterior db quality* (cf. Equation 4.13). Analysing the results from all three datasets, it turned out that the percentage of hypotheses assigned to the state *unknown* decreases with the number of applied modules. It is 31% for MGCP-Algiers dataset (six modules), 24% for GSI-Uraga dataset (seven modules) and only 16% for NGI-Zeebrugge dataset (ten modules). Furthermore, it turned out that the percentage of hypotheses assigned to the state *unknown* is relatively large for *incorrect* road hypotheses. One reason for this might be that the introduced road models represent roads better than the occurring database errors. Figure 4.47 shows some typical examples for road hypotheses assigned to the state *unknown*. Figure 4.47a shows a forested area where the roads are mostly not visible. Figure 4.47b shows a beach promenade where the abrupt transition from high buildings to an open area is not covered by any of the introduced models. Figure 4.47c shows roads at a gas terminal with an unusual appearances in terms of road width, curvature and contrast.

**Assignments to the state *invalid*:** A combination of largely contradicting probability mass distributions results in high conflict masses (cf. Equation 2.11). As discussed in Section 2.1.3, high conflict masses are considered as an indicator for internal errors of the reasoning. Therefore, any verification output achieving a conflict mass  $\geq 0.9$  is mapped to the decision state *invalid*, independent of the probability masses (cf. Algorithm 1). This mapping was motivated by a reduction of misclassifications. For the datasets MGCP-Algiers and GSI-Uraga this strategy obviously did not work, as

<sup>10</sup>This problem could be solved by using true orthophotos or by considering the viewing angle within the model or the model-uncertainty definition.

<sup>11</sup>This problem could be solved by using ALS data.



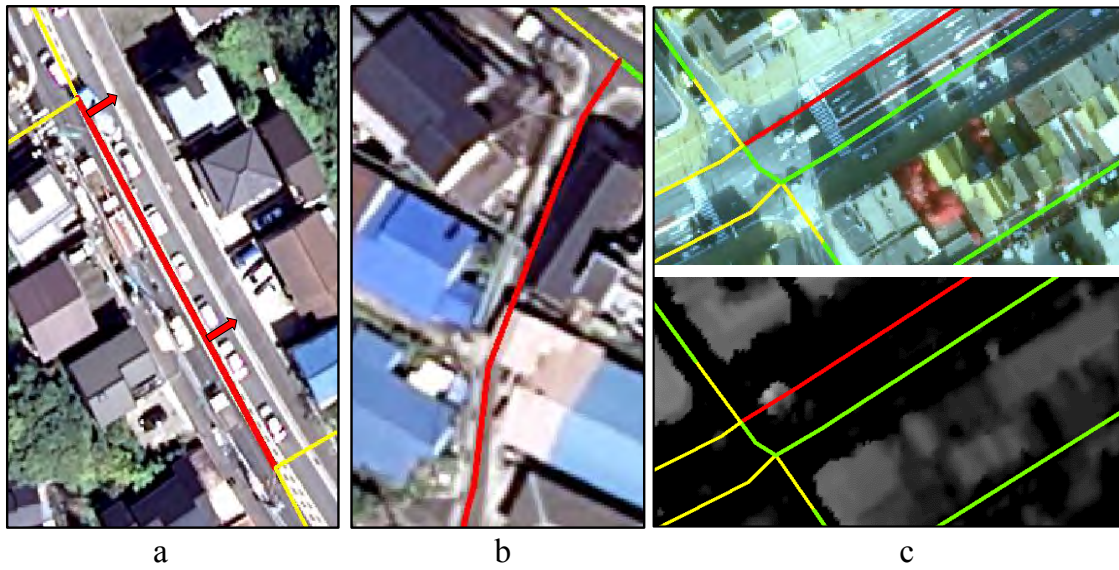
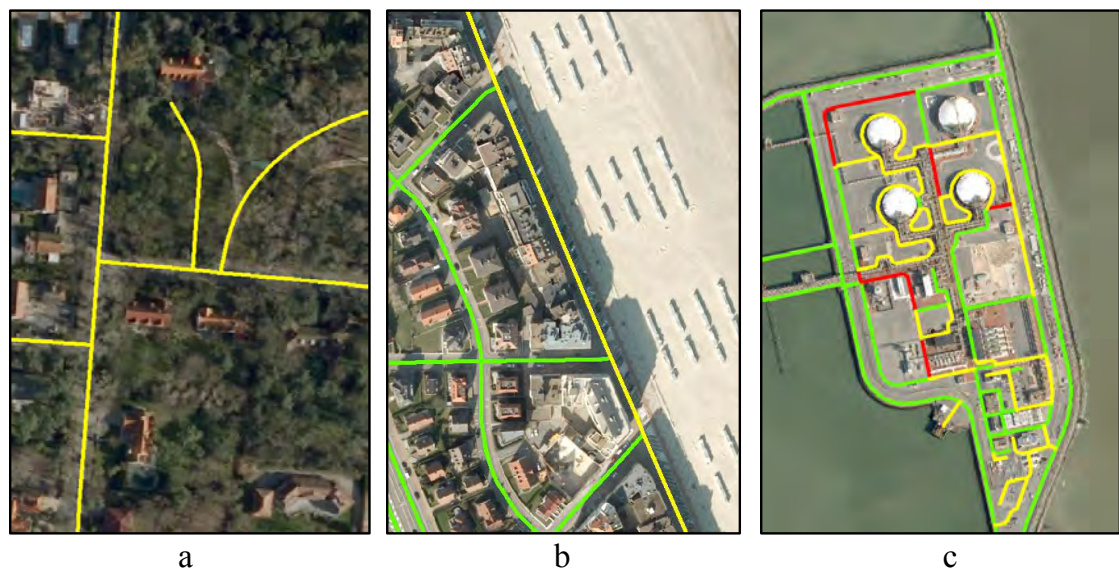


Figure 4.45.: Typical type-I errors (red).



Figure 4.46.: Examples for type-II errors (purple).

Figure 4.47.: Examples to road hypotheses assigned to the state *unknown* (yellow).

only one road hypothesis was assigned to the state *invalid* while misclassifications did occur indeed (cf. Tables 4.13 and 4.14). For the NGI-Zeebrugge dataset, 1.8% (104) of the road hypotheses are assigned to the state *invalid*. For these 104 road hypotheses, Table 4.16 shows an evaluation of the results that would be achieved if the proposed strategy concerning the conflict mass was not applied. It turns out that 43 of the 104 road hypotheses would be assigned to the wrong state. This also means that 61 road hypotheses would be assigned to their correct states. This result principally corresponds to the expectations formulated in Section 2.1.3, according to which high conflict masses do not directly infer misclassifications. Table 4.16 also shows that far more type-I errors are prevented by the additional state *invalid* than type-II errors.

		proposed approach	
		<i>correct</i>	<i>incorrect</i>
reference	<i>correct</i>	11	42 <i>type-I errors</i>
	<i>incorrect</i>	1 <i>type-II errors</i>	50

Table 4.16.: Confusion matrix of the results achieved without considering the state *invalid*. The evaluation considers only the 104 road hypotheses of the NGI-Zeebrugge dataset that were originally assigned to the state *invalid* (cf. Table 4.15).

Figure 4.48 shows some typical examples for road hypotheses assigned to the state *invalid*. Figure 4.48a refers to the same problem of the building detection module that was discussed above, i.e. vehicles on the road surface are classified as buildings. If another module assigned the same road hypothesis to the state *correct*, as does the parallel edge detection module in Figure 4.48a, the conflict mass exceeds the threshold. Figure 4.48b shows a situation where a short part of a road hypothesis is *incorrect*, which is detected by the grassland module. However, the pan-line detection module and the NDVI-line detection module assign the same road hypothesis to the state *correct* as it is correct for more than 80% of its length. The 80% threshold defined for the respective critical assumption (cf. Equation 3.66) is obviously too small to correctly describe the given situation. Furthermore, the independence of the information provided by the two line detection modules that rely on the same problematic model-uncertainty definition (the 80% threshold) is questionable. Hence, the reasoning is affected by at least two considerable problems, which lead to a high *conflict mass*. Figure 4.48c shows a bridge that is detected as a *building* by the building detection module so that, consequently, the road hypothesis is assigned to the state *incorrect*. The same road hypothesis is assigned to the state *correct* by the colour classification module and by the parallel edge detection module. In contrast to the other examples this situation does not represent a problem with the model-uncertainty concept or a violation of the independence assumption, but it represents a problem of the state space definition.<sup>12</sup>

<sup>12</sup>This problem could be solved by considering *bridge* as an exclusive state of a road hypotheses. In this hypothetical scenario, the building detection module would just provide evidence for the set of states  $bridge \cup incorrect$ , which is in accordance with the basic idea of a bridge detector described in [Clode et al., 2005].

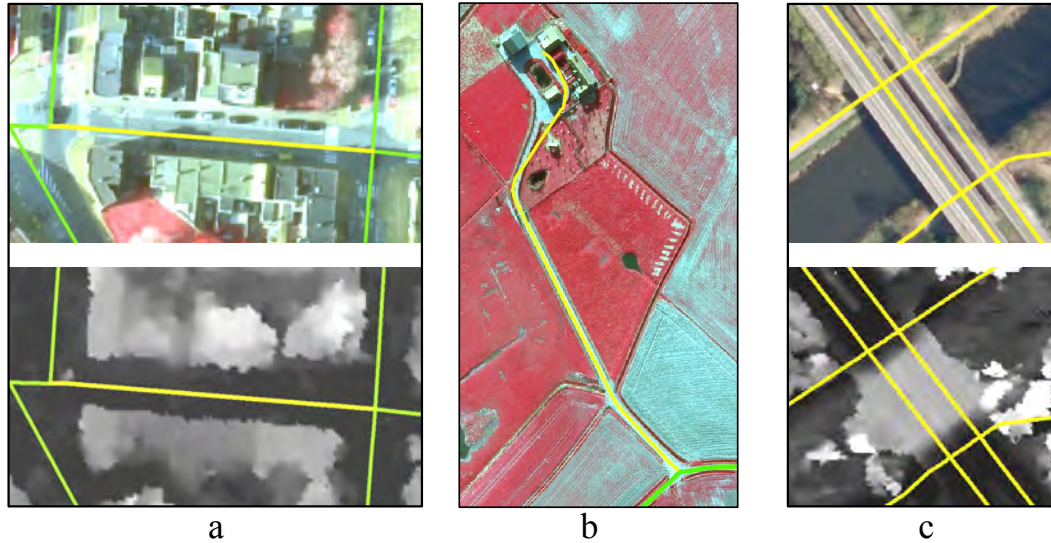


Figure 4.48.: Examples for road hypotheses assigned to the state *invalid* (yellow).

### Results of the semi-automatic approach

**Conservative setting versus low-effort setting:** The settings defined in Tables 4.11 and 4.12 represent two extreme strategies for integrating a human operator into the process. The *conservative setting* focuses on the *posterior db quality* (cf. Equation 4.13), whereas the *low-effort setting* focuses on the degree of *automation* (cf. Equation 4.15). For the cartographic task *update*, one would mainly be interested in achieving a large *posterior db quality*, and thus would tend to define a setting that corresponds to the proposed *conservative setting* that achieved a *posterior db quality* of 97.5–100.0% for the three cartographic datasets and a degree of *automation* of 66–74%.<sup>13</sup> In contrast, for the cartographic task *quality control* one would mainly be interested in achieving a large *automation* because the quality control unit typically does not have the same human resources as the update unit. The *low-effort setting* leading to a degree of *automation* of 91–100% for the experiments with the three cartographic datasets seems to be reasonable. However, the *low-effort setting* also showed a number of problems if the input imagery has low resolution and does not contain an nDSM or an infrared band as is the case for the MGCP-Algiers and GSI-Uruga datasets. In this thesis, the two interactive settings were mainly defined to show the flexibility of the proposed approach concerning different requirements. Basically, one might think of an *adaptive setting*, where a human operator starts with investigating road hypotheses that are assigned to the state *incorrect* and then proceeds with investigating road hypotheses assigned to the states *unknown* or *invalid* until a maximum working time is reached (in an *update* scenario) or until a critical number of database errors has been identified (in a *quality assessment* scenario).

**Conservative setting versus [Gerke & Heipke, 2008]:** The conservative setting corresponds to the strategy proposed by Gerke & Heipke [2008], according to which every road hypothesis that is not classified as *correct* has to be inspected by a human operator. As Gerke & Heipke [2008] presented the results of their experiments by the same type of confusion matrix as defined in this thesis (cf. Table 4.11), their results are in principle comparable to those of the conservative setting, even if

<sup>13</sup>For the task of database *update* also the missing roads (newly built roads) have to be identified in addition to the verification of existing database entries; a problem that is not considered in this thesis.



the datasets are not the same. The experiments presented in [Gerke & Heipke, 2008] deal with the verification of a real subset of the national German road database Authoritative Topographic Cartographic Information System (ATKIS) that corresponds to a mapping scale of 1:25,000. The imagery consist of aerial RGB images with a GSD of about 0.3 m. The test scene shows a challenging rural area in Germany, where approximately 25% of the road objects are occluded by trees. Table 4.17 shows the results for three applied datasets MGCP-Algiers, GSI-Uraga and NGI-Zeebrugge in contrast to results computed from the confusion matrix presented in [Gerke & Heipke, 2008]. The results have approximately the same order of magnitude. In the category *corrected db errors* the proposed approach outperforms the approach from Gerke & Heipke [2008]. However, this statement is not significant as in case of [Gerke & Heipke, 2008] the value only relies on a small number of database errors ( $\approx 10$ ), which can be seen from the comparably large *prior db quality* (98%) and the total number of road objects (530). Unlike the approach of Gerke & Heipke [2008], the main advantage of the new approach is its general applicability as it is not restricted to rural areas.

	[Gerke & Heipke, 2008]	conservative setting		
datasets	ATKIS, RGB, 0.3 m GSD	MGCP-Algiers	GSI-Uraga	NGI-Zeebrugge
number of evaluated road objects	530	2,254	360	5,797
<i>prior db quality</i>	98%	98.5%	81.6%	90.2%
<i>posterior db quality</i>	99%	100%	97.5%	99.3%
<i>automation</i>	66%	69%	66%	74%
<i>corrected db errors</i>	35%	100%	86%	93%

Table 4.17.: Comparison to the results of Gerke & Heipke [2008].

## Final comments

In contrast to the results of the analysis in Section 4.2.2, it turned out that the verification modules NDVI-line detection and parallel edge detection also provide valuable information (cf. Figures 4.35 and 4.44). Furthermore, the relevance of the verification modules building detection and grassland detection could be shown as they are mainly responsible for detecting realistic database errors (cf. Figure 4.44).

The experiments with cartographic datasets show that the new approach has the potential to be used for a practical application. The key for that is the relatively large *classification correctness* of 93–100% for the three datasets that allows a user to trust the results of the automatic component, even if one applies it to challenging scenes and provides partly non-representative training data. Hence, a human operator can concentrate on investigating difficult spots and editing database errors. The achieved degree of *automation* of at least 69% means that a human operator is able to screen a database three times faster, which further means that a human operator is able to screen three times more in a given time period or the same area three times more often.

The preceding analysis also showed some problems of the proposed approach, e.g. that the introduced model-uncertainty analysis is still incomplete in some cases. Finally, it turned out that input imagery with smaller GSD and the availability of infrared band and nDSM significantly increased the degree of automation, which may justify the higher expenses for acquiring such data.

## 5. Conclusions and outlook

In this thesis a new approach for automatic road database verification based on remote sensing data has been presented. In contrast to existing approaches, the applicability of the proposed approach is not restricted to specific road types, context areas or geographic regions. The general applicability is based on a combination of several state-of-the-art road detection and road verification approaches that are particularly useful for different situations. The performance of the approach was evaluated on the basis of larger datasets than other state-of-the-art road verification approaches. In the following, the main aspects of the proposed approach will be discussed.

**Verification modules:** Ten independently operating methods have been implemented as so-called verification modules that all represent specific road models and processing strategies. Models and strategies are in accordance with different state-of-the-art approaches that have experimentally proven to be useful in different situations. Some of the modules are based on heuristics and some are implicitly defined on the basis of training data. The models based on heuristics mostly rely on general properties of roads, e.g. the parallelism of the road borders or the fact that the road surface is situated on the terrain. The statistical models mostly rely on scene-dependent properties of roads, e.g. colour and radiometric contrast. Furthermore, each module requires specific input data such as satellite images, high resolution aerial images, infrared bands or nDSM data.

Experiments carried out in ten scenes representing different context regions, road types and geographic regions show that the implemented road models cover 69–100% of the roads, depending on the scene. For six of the ten scenes results from other state-of-the-art road detection approaches are available from a benchmark test. It turned out that in all six scenes, the proposed approach outperforms the best approaches of the benchmark test.

However, the evaluation also shows potential for further improvement, because even if all ten verification modules are activated, up to 18% of the roads are not covered by the combination of all road models.

**Reasoning approach:** Each verification module represents a probabilistic classifier that assigns each road stored in the database to the states *correct* and *incorrect*. This is mostly realized by extracting features from a local image region indicated by the database and then checking the correspondence of these features with the underlying road model. However, the modules are not expected to provide valuable classification outputs in all situations because their underlying models are based on specific assumptions about the appearance of roads in the imagery that may be violated. Consequently, the *model-uncertainty* is determined for each module, indicating the probability whether the underlying model is *applicable* or not. In accordance with the Dempster-Shafer Theory, the *verification result* and the *model-uncertainty* are mapped to a new state space that considers *correct*, *incorrect* and *unknown* as the possible states of a road hypothesis. Dempster’s rule is applied to combine the

distributions provided by the different verification modules.

The performance of the verification modules has been evaluated and compared to the performance of the combined approach. It turned out that the combined solutions are always better than the module-specific solutions. A considerable strength of the introduced fusion concept is its relatively simple parametrisation because the parameters of the modules are independent. Changing a parameter of one module does not affect the performance of another module. Experiments have shown that the proposed approach is transferable to different datasets with only moderate efforts.

**Model-uncertainty:** For each verification solution the uncertainty of the state of its model (*applicable*, *not applicable*) is expressed by a probability distribution. A model is only considered being *applicable* if all assumptions made for the model definition do hold for the given situation in the data. The state of an assumption (*satisfied*, *violated*) is also expressed by a probability distribution, which is determined by a sigmoid function. The parameters of the sigmoid function are defined on the basis of theoretical and experimental concepts given in the original literature or in benchmark analyses. Due to the literature, their definition includes concepts from the Bayesian and Fuzzy-Set Theory as well.

The *model-uncertainty* mainly has the task to prevent the overall approach from relying on verification modules that potentially provide incorrect classification outputs. The experiments have shown that for the ten test scenes the *classification correctness* is consistently large (93–100%), and hence the new concept turned out to work rather well even for challenging scenes. However, the evaluation have also shown that considering additional assumptions would further enhance the achieved results.

**SVDD-based uncertainty metric:** Another important contribution of this thesis is the model-uncertainty defined for an SVM classifier being at the core of one of the verification modules. An SVM classifier defines the model implicitly on the basis of training samples. The question whether such a learned model is *applicable* refers to the question whether the training data are representative for a test sample. Using an SVDD classifier to describe the training data, the relation of a test sample to the set of training samples is described by a sigmoid function of its feature space distance to the SVDD hypersurface in the input feature space. The sigmoid parameters are automatically adjusted in accordance with the input dataset.

A set of experiments validated the relation of the determined feature space distances and the *classification correctness*. It could be shown that with an increasing feature space distance to the training data, the *correctness* of the SVM outputs decrease significantly, which supports the basic idea that the probability for *ignorance* increases with the feature space distance to the training samples.

**User interaction:** The statistical reasoning provides a solution in form of a probability mass distribution for the states *correct*, *incorrect* and *unknown*. In order to make this solution interpretable for a human operator, this distribution is mapped to a decision space consisting of four exclusive states *correct*, *incorrect*, *unknown* and *invalid*. The state *invalid* signalizes problems of the reasoning, and thus represents situations where the automatic component might be not suitable to solve the verification problem. Two interactive settings have been defined. In the *conservative setting* a human



operator checks roads assigned to the states *incorrect*, *unknown* and *invalid*. In the *low-effort setting* a human operator only checks roads assigned to the state *incorrect*. In any case the human operator has the task to edit the road database if an error has been spotted.

The experiments have shown that the conservative setting leads to good database qualities: 97.5–100% of the roads are correct after the approach has been applied. In order to achieve this result, 25–31% of the roads had to be investigated by a human operator. For the low-effort setting, the resulting database quality is lower; only 88.1–98.5% of the objects are correct after the approach has been applied, but only up to 9% of the roads need to be investigated by a human operator. The differences in the two settings show the flexibility of the proposed approach concerning different cartographic problems that are characterized by different database qualities and degrees of automation.

Finally, it can be concluded that the main objective of this thesis has been met successfully. The experiments indicate a general applicability of the new approach by showing good results for different datasets that represent different context areas, road types, geographic regions and road databases. In this regard, it is worth noting that the proposed approach has been realized as a prototype software implemented in C++ that has been installed at the German Federal Agency for Geodesy and Cartography (BKG) as a part of the WiPKA project [Helmholz et al., 2012].

## Outlook

The incorporation of *additional verification modules* representing further road models seems to be interesting as still not all potential appearances of roads (and appearances of database errors) are covered by the implemented verification modules. In particular, additional models are required that are suitable to explain roads that lie at the border line of different context areas. Such models could be designed by applying existing models for each road side separately and then fuse their outputs. Further methods that could be implemented as additional verification modules can be found in the related literature, for instance the approach presented by Gerke & Heipke [2008] who detect rows of trees to infer the position of occluded roads or the approach presented by Porway et al. [2010] who detect cars to infer the position of roads. Both methods are suitable for specific situations that may further improve the *classification completeness* of the combined solution.

Another promising possibility to enhance the proposed approach is to add verification modules that use completely different information sources such as up-to-date GPS-tracks recorded by standard cars, independent vector databases or planning data for redevelopment projects. The use of such information sources has the potential advantage that roads that are usually not visible in the images could be classified, e.g. those that lie in forested areas.

Furthermore, it would be interesting to add modules that specifically classify road junctions. Promising approaches to be integrated are [Auclair-Fortier et al., 1999; Ravanbakhsh et al., 2008; Grote et al., 2012; Lin et al., 2012].

The experiments have mainly shown two aspects missing in the model-uncertainties: First, the used nDSM turned out to have partially lower accuracy than expected. This problem can be overcome by analysing the nDSM generation process that is based on dense matching. Secondly, it was shown in the experiments that buildings occluding narrow roads can pose an additional problem which

could be compensated by using the viewing angle in the respective model or in the analysis for the uncertainty-model.

One open point with respect to a practical use of the method is the development of a *flexible interface* between the automatic component and the human operator. In this regard, Becker et al. [2012] proposed a strategy for the verification of landcover databases, where the landcover objects are forwarded to the human operator in ascending order of their similarity to the correct landcover objects. This idea could be transferred to the approach described in this thesis using the derived probability masses instead of a similarity measure. However, the definition of the order is not trivial because the three states *correct*, *incorrect* and *unknown* need to be taken into account.

Training samples are required to make the proposed method applicable for different scenes. Their definition is a task of a human operator, and thus affects the automation potential of the approach. Therefore, it is interesting to automate the training data generation. This could be realised by using the existing database where the majority of the database entries are assumed to be *correct*. In this regard, Büschenfeld & Ostermann [2012] proposed a training strategy for an SVM classifier that automatically removes support vectors that probably correspond to the wrong class.

In this thesis *missing database entries* are not considered. For the cartographic problem of *updating road databases*, the missing roads also need to be identified. However, the review of the related work has shown that approaches which try to detect the missing roads directly, achieve results that are currently not good enough to be used in practice, even if the verified road network is used as the basis for the process, e.g. [Auclair-Fortier et al., 2001; Poulain et al., 2010]. Whereas a direct detection of the missing roads is too challenging, indirect detection methods are more promising. For instance, Beyen et al. [2012] demonstrate that more than 50% of the missing roads can be effectively detected by detecting new built-up areas based on a landcover classification. The underlying model of this strategy assumes that new roads are located near or inside larger redevelopment projects involving recently constructed buildings, parking lots and other man-made structures that are much easier to detect than roads alone. Another possibility is to use additional information such as GPS-tracks acquired by cars or planning data of redevelopment projects that allow the definition of update-hints to be investigated by a human operator.

## A. Appendix

### A.1. The cat in the box – Bayesian inference versus Dempster-Shafer Theory

The reasoning problem described below is intended to show the difference between Bayesian inference and the Dempster-Shafer Theory.<sup>1</sup>

#### The physical experiment

Imagine that a physicist has placed a living cat and a physical experiment in a box. The experiment has two possible outcomes, where one has catastrophic consequences for the cat. The chances for both outcomes of the experiment are equal, and hence the cat in the box may be *dead* or *alive* after the experiment is over. There is no possibility to look inside the box, but the physicist has equipped the box with two sensors  $S_1$  and  $S_2$ , each being connected to a control lamp outside the box. After the critical part of the experiment is over, each lamp has two possible states, either it is *green* or it is *red*. As the physicist is very busy, he asks a statistician to observe the experiment and report the outcome with respect to the state of the cat. The physicist gives the statistician just three facts about the sensors:

- If a lamp is *green*, the cat is *alive* with a probability of 0.99.
- If a lamp is *red*, the cat is *dead* with a probability of 0.99.
- The sensors may be *defect* though. In case of a defect sensor, the connected lamp still has the two possible states *green* and *red* but their meaning is *unknown*. The probability that sensor  $S_1$  is defect is 0.2. The probability that sensor  $S_2$  is defect is 0.4.

A while after the physicist has left the room, the experiment has reached its critical moment and the control lamps light up. The lamp  $L_1$  connected to sensor  $S_1$  was turned to *green* and the lamp  $L_2$  connected to sensor  $S_2$  was turned to *red*. Now the statistician is asked to describe the state of the cat.

#### The problem of probably defect sensors

Due to the fact that the lamps connected to defect sensors do not remain dark, the state of the cat (*dead*, *alive*) depends not only on the colours of the lamps (*green*, *red*) but also on the states of the two sensors (*working*, *defect*). Furthermore, the physicist did not explicitly say that the state of the

<sup>1</sup>The given example is inspired by [Smets, 1994].

cat is independent from the states of the sensors. For instance, imagine a scenario in which the cat could somehow affect the functionality of the sensors.

### Preliminary considerations

The frame of discernment describing the state of the *cat* is defined as:

$$\Theta = \{alive, dead\} \quad (A.1)$$

In accordance with the design of the physical experiment, the prior probability for the state of the cat is:

$$P(cat = alive) = P(cat = dead) = 0.5 \quad (A.2)$$

The probability distributions describing the states of the sensors  $S_1$  and  $S_2$  are:

$$P(S_1 = working) = 1 - P(S_1 = defect) = 0.8 \quad (A.3)$$

$$P(S_2 = working) = 1 - P(S_2 = defect) = 0.6 \quad (A.4)$$

According to the observed states of the lamps  $L_1 = green$  and  $L_2 = red$ , the statistician knows:

$$P(cat = alive|L_1 = green, S_1 = working) = 1 - P(cat = dead|L_1 = green, S_1 = working) = 0.99 \quad (A.5)$$

$$P(cat = alive|L_2 = red, S_2 = working) = 1 - P(cat = dead|L_2 = red, S_2 = working) = 0.01 \quad (A.6)$$

However, the statistician does not know the corresponding distributions for defect sensors:

$$P(cat = alive|L_1 = green, S_1 = defect) = 1 - P(cat = dead|L_1 = green, S_1 = defect) = ? \quad (A.7)$$

$$P(cat = alive|L_2 = red, S_2 = defect) = 1 - P(cat = dead|L_2 = red, S_2 = defect) = ? \quad (A.8)$$

### The Bayesian reasoning approach

After the preconditions have been defined, a Bayesian statistician is usually looking for the probability that the cat is alive given all available evidence, which is defined by the *Bayesian rule of combination*<sup>2</sup>:

$$\begin{aligned} &P(cat = alive|L_1 = green, L_2 = red) = \\ &= \frac{P(cat = alive|L_1 = green)P(cat = alive|L_2 = red)}{P(cat = alive|L_1 = green)P(cat = alive|L_2 = red) + P(cat = dead|L_1 = green)P(cat = dead|L_2 = red)} \end{aligned} \quad (A.9)$$

<sup>2</sup>Due to the uniform prior (cf. Equation A.2) the likelihoods could be replaced by the corresponding posteriors without changing the meaning but it simplifies some notations below.

The probabilities that show up in Equation A.9 are defined through Equations A.5–A.8, which can be expressed as follows:

$$P(cat = a.|L_1 = g.) = \underbrace{P(cat = a.|L_1 = g., S_1 = w.)}_{0.99} \underbrace{P(S_1 = w.)}_{0.8} + \underbrace{P(cat = a.|L_1 = g., S_1 = df.)}_{?} \underbrace{P(S_1 = df.)}_{0.2} \quad (\text{A.10})$$

$$P(cat = d.|L_1 = g.) = \underbrace{P(cat = d.|L_1 = g., S_1 = w.)}_{0.01} \underbrace{P(S_1 = w.)}_{0.8} + \underbrace{P(cat = d.|L_1 = g., S_1 = df.)}_{?} \underbrace{P(S_1 = df.)}_{0.2} \quad (\text{A.11})$$

$$P(cat = a.|L_2 = r.) = \underbrace{P(cat = a.|L_2 = r., S_2 = w.)}_{0.01} \underbrace{P(S_2 = w.)}_{0.6} + \underbrace{P(cat = a.|L_2 = r., S_2 = df.)}_{?} \underbrace{P(S_2 = df.)}_{0.4} \quad (\text{A.12})$$

$$P(cat = d.|L_2 = r.) = \underbrace{P(cat = d.|L_2 = r., S_2 = w.)}_{0.99} \underbrace{P(S_2 = w.)}_{0.6} + \underbrace{P(cat = d.|L_2 = r., S_2 = df.)}_{?} \underbrace{P(S_2 = df.)}_{0.4} \quad (\text{A.13})$$

It turns out that the required probabilities remain partly *unknown* because the probabilities, emphasized by the question marks are *unknown*. Substituting the *unknowns* by intervals between zero and one leads to the following solutions:

$$0.792 \leq P(cat = alive|L_1 = green) \leq 0.992 \quad (\text{A.14})$$

$$0.008 \leq P(cat = dead|L_1 = green) \leq 0.208 \quad (\text{A.15})$$

$$0.006 \leq P(cat = alive|L_2 = red) \leq 0.406 \quad (\text{A.16})$$

$$0.594 \leq P(cat = dead|L_2 = red) \leq 0.994 \quad (\text{A.17})$$

Plugging the intervals A.14–A.17 into Equation A.9 leads to:

$$0.022 \leq P(cat = alive|L_1 = green, L_2 = red) \leq 0.988 \quad (\text{A.18})$$

The interval A.18 represents the final solution of the Bayesian reasoning approach. This solution covers all the possibilities the physicist intentionally left open, and hence is an appropriate probabilistic description for the state of the cat. It is worth noting that, regardless of how the lamps had lit up, the result would always have been similar, an interval between almost zero and almost one.

One possibility to overcome the problem of such broad intervals is to define four alternative solutions, each under a specific condition:

$$P(cat = alive|L_1 = green, L_2 = red) = \begin{cases} 0.60 & |S_1 = working, S_2 = working \\ 0.99 & |S_1 = working, S_2 = defect \\ 0.01 & |S_1 = defect, S_2 = working \\ 0.50 & |S_1 = defect, S_2 = defect \end{cases} \quad (\text{A.19})$$

where the probabilities for the conditions are known by:

$$P(S_1 = working, S_2 = working) = 0.48 \quad (\text{A.20})$$

$$P(S_1 = working, S_2 = defect) = 0.32 \quad (\text{A.21})$$

$$P(S_1 = defect, S_2 = working) = 0.12 \quad (\text{A.22})$$

$$P(S_1 = defect, S_2 = defect) = 0.08 \quad (\text{A.23})$$

Another possibility to overcome the problem of broad intervals is to replace the *unknowns* (cf. Equations A.7 and A.8) by a uniform distribution:

$$P(cat = alive|L_1 = green, S_1 = defect) = P(cat = dead|L_1 = green, S_1 = defect) = 0.5 \quad (A.24)$$

$$P(cat = alive|L_2 = red, S_2 = defect) = P(cat = dead|L_2 = red, S_2 = defect) = 0.5 \quad (A.25)$$

Plugging Equations A.24 and A.25 into Equations A.10–A.13 leads to the following distributions:

$$P(cat = alive|L_1 = green) = 1 - P(cat = dead|L_1 = green) = 0.892 \quad (A.26)$$

$$P(cat = alive|L_2 = red) = 1 - P(cat = dead|L_2 = red) = 0.206 \quad (A.27)$$

Now the Bayesian combination rule (cf. Equation A.9) leads to:

$$P(cat = alive|L_1 = green, L_2 = red) = 0.682 \quad (A.28)$$

Note that the strategy of introducing uniform distributions is in accordance with the *principle of maximum entropy* [Giuasu & Shenitzer, 1985], which says that an *unknown* probability distribution most probably corresponds to a uniform distribution. This principle refers to the *second law of thermodynamics* according to which entropy almost always increases. However, even if this physical law represents an appropriate concept here, it also says that there is a second-most probable distribution, a third-most probable distribution and so on. Hence, introducing a uniform distribution restricts the generality of the solution, which further means that the achieved solution is probably wrong.

## The Dempster-Shafer Theory

Dempster [1968] introduced a statistically founded reasoning approach, which he called: *A generalization of Bayesian inference*. With respect to the given task, the preliminary considerations, including the frame of discernment (cf. Equation A.1), the *knowns* (cf. Equations A.2–A.6) and the *unknowns* (cf. Equations A.7 and A.8) correspond to the standard Bayesian approach. Beyond that, Dempster [1967] defined a calculation scheme based on *lower bounds of probabilities* that many other authors denoted as *probability masses*  $m$ . For the given problem these *probability masses*  $m$  are defined as follows:

$$m_{L_1}(cat = alive) = P(cat = alive|L_1 = green, S_1 = working)P(S_1 = working) \quad (A.29)$$

$$m_{L_1}(cat = dead) = P(cat = dead|L_1 = green, S_1 = working)P(S_1 = working) \quad (A.30)$$

$$m_{L_2}(cat = alive) = P(cat = alive|L_2 = red, S_2 = working)P(S_2 = working) \quad (A.31)$$

$$m_{L_2}(cat = dead) = P(cat = dead|L_2 = red, S_2 = working)P(S_2 = working) \quad (A.32)$$

The right terms in Equations A.29–A.32 equal the left terms in Equations A.14–A.17. This means



that the *probability masses*  $m$  of Dempster's approach correspond to the lower bounds of the intervals of the Bayesian approach (cf. intervals A.14–A.17).<sup>3</sup> <sup>4</sup> Calculation leads to:

$$m_{L_1}(cat = alive) = 0.792 \quad (\text{A.33})$$

$$m_{L_1}(cat = dead) = 0.008 \quad (\text{A.34})$$

$$m_{L_2}(cat = alive) = 0.006 \quad (\text{A.35})$$

$$m_{L_2}(cat = dead) = 0.594 \quad (\text{A.36})$$

Dempster [1968] found that *probability masses* must be referred to a different state space than standard probabilities, i.e. a state space that contains all subsets of  $\Theta$  (cf. Equation A.1). Accordingly, the statistician defines the new state space based on  $\Theta$  (cf. Equation A.1):

$$2^\Theta = \{alive, dead, dead \cup alive, \emptyset\} \quad (\text{A.37})$$

The explicit consideration of the state  $\{dead \cup alive\}$  remarks the prominent difference of Dempster's approach compared to Bayesian inference.<sup>5</sup>

Because of  $P(cat = dead \cup alive) = 1.0$  also the lower bound of  $P(cat = dead \cup alive)$  must be equal to one, in general. However, according to the new state space definition  $m(cat = dead \cup alive)$  refers to the state  $\{dead \cup alive\}$  but not to the states  $\{dead\}$  and  $\{alive\}$ . In accordance with Equations A.33–A.36, this leads to the following definitions:

$$m_{L_1}(cat = d. \cup a.) = 1.0 - \underbrace{P(cat = d. | L_1 = g., S_1 = w.)P(S_1 = w.)}_{0.008} - \underbrace{P(cat = a. | L_1 = g., S_1 = w.)P(S_1 = w.)}_{0.792} \quad (\text{A.38})$$

$$m_{L_2}(cat = d. \cup a.) = 1.0 - \underbrace{P(cat = d. | L_2 = r., S_2 = w.)P(S_2 = w.)}_{0.594} - \underbrace{P(cat = a. | L_2 = ., S_1 = w.)P(S_2 = w.)}_{0.006} \quad (\text{A.39})$$

and further calculation leads to:

$$m_{L_1}(cat = dead \cup alive) = 0.2 \quad (\text{A.40})$$

$$m_{L_2}(cat = dead \cup alive) = 0.4 \quad (\text{A.41})$$

<sup>3</sup>Note that if the *lower* and the *upper bound* of a *probability* would be equal, Dempster's approach would provide the same solution as the corresponding Bayesian approach. With respect to the given example, it means that if the interval widths (cf. intervals A.14–A.17) would be zero, i.e. both sensors are *working* with 1.0 probability, Dempster's approach would have no benefit compared to the standard Bayesian approach.

<sup>4</sup>It is further worth noting that Dempster [1967] explicitly used the notation *lower probabilities*  $P_*$  which nicely shows the relation to the lower bounds of the probability intervals. In the 1970s the notation *probability masses*  $m$  has been established. The main intention of changing Dempster's original notation was related to criticisms from the field of *Frequentist statistics*, whereas from a *Bayesian* viewpoint Dempster's notation is OK.

<sup>5</sup>The additional state  $\{dead \cup alive\}$  means that the state of the cat is *unknown*; it does not mean that the state of the cat is *uncertain*, which it is anyway because Equations A.5 and A.6 are considered.

Finally, applying Dempster's combination rule leads to:

$$\begin{aligned}
& m_{L_1, L_2}(cat = alive) = \\
&= \frac{m_{L_1}(cat = a.)m_{L_2}(cat = a.) + m_{L_1}(cat = a.)m_{L_2}(cat = d. \cup a.) + m_{L_1}(cat = d. \cup a.)m_{L_2}(cat = a.)}{1 - [m_{L_1}(cat = a.)m_{L_2}(cat = d.) + m_{L_1}(cat = d.)m_{L_2}(cat = a.)]} \\
&= 0.610
\end{aligned} \tag{A.42}$$

$$\begin{aligned}
& m_{L_1, L_2}(cat = dead) = \\
&= \frac{m_{L_1}(cat = d.)m_{L_2}(cat = d.) + m_{L_1}(cat = d.)m_{L_2}(cat = d. \cup a.) + m_{L_1}(cat = d. \cup a.)m_{L_2}(cat = d.)}{1 - [m_{L_1}(cat = a.)m_{L_2}(cat = d.) + m_{L_1}(cat = d.)m_{L_2}(cat = a.)]} \\
&= 0.239
\end{aligned} \tag{A.43}$$

$$\begin{aligned}
& m_{L_1, L_2}(cat = dead \cup alive) = \\
&= \frac{m_{L_1}(cat = d. \cup a.)m_{L_2}(cat = d. \cup a.)}{1 - [m_{L_1}(cat = a.)m_{L_2}(cat = d.) + m_{L_1}(cat = d.)m_{L_2}(cat = a.)]} \\
&= 0.151
\end{aligned} \tag{A.44}$$

The concept of Dempster's combination rule widely corresponds to the concept of the Bayesian combination rule, which becomes more obvious in the matrix interpretation. In Table A.1 the entries of the matrix correspond to the products of the row and column headings, and thus represent the intersections of the corresponding sets. The main difference of both combination rules lies in the normalization: The Bayesian combination rule normalizes over the bold entries (cf. denominator in Equation A.9), whereas Dempster's combination rule also normalizes over the bottom-right entry (cf. denominator in Equations A.42–A.44).<sup>6</sup>

## Discussion

Table A.2 summarizes the alternative representations discussed above. Each Bayesian approach has particular disadvantages: The first solution, i.e. *the broad interval* is nearly meaningless, the solution based on *4-fold conditioning* is confusing<sup>7</sup>, while the solution which is based on the *uniform distribution* is probably wrong. The DST-solution reads: The *cat* is *alive* with a probability mass of 0.61, *dead* with a probability mass of 0.24 and its state is *unknown* with a probability mass of 0.15. Hence, the DST approach overcomes the disadvantages of the three Bayesian solutions by expressing *uncertainty* and *ignorance* in a single distribution.

<sup>6</sup>Another difference is that for the Bayesian combination rule always some of row and column headings have to be equal to zero.

<sup>7</sup>This statement would be more convincing for a larger number of sensors to be combined.

	$m_{L_2}(cat = alive) = 0.006$	$m_{L_2}(cat = dead) = 0.594$	$m_{L_2}(cat = dead \cup alive) = 0.40$
$m_{L_1}(cat = alive) = 0.792$	<b>0.0048</b>	$\emptyset$	<b>0.3168</b>
$m_{L_1}(cat = dead) = 0.008$	$\emptyset$	<b>0.0047</b>	<b>0.0032</b>
$m_{L_1}(cat = dead \cup alive) = 0.2$	<b>0.0012</b>	<b>0,1188</b>	0.0800

Table A.1.: The orthogonal sum.

The Bayesian reasoning approach:

*the broad interval:*  $0.022 \leq P(cat = alive|L_1 = green, L_2 = red) \leq 0.988$

*4-fold conditioning:*  $P(cat = alive|L_1 = green, L_2 = red) = \begin{cases} 0.60 & |S_1 = working, S_2 = working \\ 0.99 & |S_1 = working, S_2 = defect \\ 0.01 & |S_1 = defect, S_2 = working \\ 0.50 & |S_1 = defect, S_2 = defect \end{cases}$

$$P(S_1 = working, S_2 = working) = 0.48$$

$$P(S_1 = working, S_2 = defect) = 0.32$$

$$P(S_1 = defect, S_2 = working) = 0.12$$

$$P(S_1 = defect, S_2 = defect) = 0.08$$

*uniform distribution:*  $P(cat = alive|L_1 = green, L_2 = red) = 0.682$

The Dempster-Shafer Theory:

$$m_{L_1=green, L_2=red}(cat = alive) = 0.610$$

$$m_{L_1=green, L_2=red}(cat = dead) = 0.239$$

$$m_{L_1=green, L_2=red}(cat = dead \cup alive) = 0.151$$

Table A.2.: Comparison of the alternative reasoning approaches described in the text.

## A.2. Gaussian Mixture Models for colour-based road verification

The goal of the method described below is to find one or two Gaussian components that represent the road surface in the marginal distributions of an RGB colour space. Therefore, a Gaussian Mixture Model (GMM) is defined for each colour band using the Expectation-Maximization algorithm (EM) described in [Dempster et al., 1977]. In the following, GMM and EM will be described according to [Bishop, 2006], and their parametrisation will be motivated with respect to the colour classification problem dealt with in Section 3.3.1.

### Gaussian Mixture Model

A GMM is defined as a linear combination of Gaussians given a grey value  $g$  as an argument:

$$f_{\text{GMM}}(g) = \sum_{1 \leq q \leq Q} w_q \mathcal{N}(g | \mu_q, \sigma_q) \quad \text{s.t.} \quad \sum_{1 \leq q \leq Q} w_q = 1 \quad (\text{A.45})$$

where  $Q$  denotes the number of Gaussian components and  $\mathcal{N}(g | \mu_q, \sigma_q)$  denotes the  $q$ -th Gaussian component with the two parameters  $\mu_q$  and  $\sigma_q$ . The parameters  $w_q \in \mathbb{R}$  with  $w_q \geq 0$  are the mixing coefficients of GMM, modelling the contributions of the individual components to the joint distribution. As the marginal distributions of the colour space are one-dimensional, the considered Gaussian components are also one-dimensional, and thus have the form:

$$\mathcal{N}(g | \mu_q, \sigma_q) = \frac{1}{\sigma_q \sqrt{2\pi}} \cdot \exp \left( -\frac{(g - \mu_q)^2}{2\sigma_q^2} \right) \quad (\text{A.46})$$

**The optimization problem:** The goal is to find a function  $f_{\text{GMM}}$  (cf. Equation A.45) that optimally describes the grey values  $g_j$  of the pixels  $p_j \in R^{\text{H}\bullet}$ , where  $R^{\text{H}\bullet}$  denotes the set of pixels corresponding to the road hypothesis. As Dempster et al. [1977] point out, the function  $f_{\text{GMM}}(g)$  in Equation A.45 can be interpreted as a probability function similar to a single Gaussian. Hence, Dempster et al. [1977] apply Maximum Likelihood Estimation (MLE) to find the GMM parameters that maximize the probability of the data. Accordingly, the *log likelihood* of the GMM is defined for the set  $g_j$  with  $1 \leq j \leq n$  and  $n = |R^{\text{H}\bullet}|$ :

$$\max_{\mu, \sigma, w} L(\mu, \sigma, w) \quad (\text{A.47})$$

with

$$\begin{aligned} L(\mu, \sigma, w) &= \sum_{1 \leq j \leq n} \ln \left\{ \sum_{1 \leq q \leq Q} [w_q \mathcal{N}(g_j | \mu_q, \sigma_q)] \right\} \\ \text{s.t.} \quad &\sum_{1 \leq q \leq Q} w_q = 1 \end{aligned} \quad (\text{A.48})$$

In order to solve the maximum likelihood problem the constraint in Equation A.48 needs to be included using Lagrange multipliers  $\lambda$ :

$$\max_{\lambda, \mu, \sigma, w} L(\lambda, \mu, \sigma, w) \quad (\text{A.49})$$

with

$$L(\lambda, \mu, \sigma, w) = \sum_{1 \leq j \leq n} \ln \left\{ \sum_{1 \leq q \leq Q} [w_q \mathcal{N}(g_j | \mu_q, \sigma_q)] \right\} + \lambda \left( \sum_{1 \leq q \leq Q} w_q - 1 \right) \quad (\text{A.50})$$

The partial derivatives of the log likelihood  $L(\lambda, \mu, \sigma, w)$  with respect to the unknown GMM parameters  $\mu_q, \sigma_q$  and  $w_q$  read:

$$\frac{\partial L(\lambda, \mu, \sigma, w)}{\partial \mu_q} = \sum_{1 \leq j \leq n} \left\{ \frac{w_q \mathcal{N}(g_j | \mu_q, \sigma_q)}{\sum_{1 \leq k \leq Q} w_k \mathcal{N}(g_j | \mu_k, \sigma_k)} \cdot \frac{\mu_q - g_j}{\sigma_q^2} \right\} \quad (\text{A.51})$$

$$\frac{\partial L(\lambda, \mu, \sigma, w)}{\partial \sigma_q} = \sum_{1 \leq j \leq n} \left\{ \frac{w_q \mathcal{N}(g_j | \mu_q, \sigma_q)}{\sum_{1 \leq k \leq Q} w_k \mathcal{N}(g_j | \mu_k, \sigma_k)} \cdot \frac{(g_j - \mu_q)^2 - \sigma_q^2}{\sigma_q^3} \right\} \quad (\text{A.52})$$

$$\frac{\partial L(\lambda, \mu, \sigma, w)}{\partial w_q} = \sum_{1 \leq j \leq n} \left\{ \frac{\mathcal{N}(g_j | \mu_q, \sigma_q)}{\sum_{1 \leq k \leq Q} w_k \mathcal{N}(g_j | \mu_k, \sigma_k)} \right\} + \lambda \quad (\text{A.53})$$

Usually, setting the partial derivatives of the log likelihood with respect to the unknowns to zero provides a solution for unknowns. Unfortunately, neither of the partial derivatives in Equations A.51–A.53 have a closed form as is the case for MLE with a single Gaussian. The reason is that the logarithm in Equation A.50 appears in front of the sum and not in front of the product. Hence, the logarithm does not lead to an elimination of the remaining variables. To overcome the problem, Dempster et al. [1977] introduced so-called *latent variables*  $\gamma(z_{jq})$  that assign a grey value with index  $1 \leq j \leq n$  to a Gaussian component with index  $1 \leq q \leq Q$ :

$$\gamma(z_{jq}) = \frac{w_q \mathcal{N}(g_j | \mu_q, \sigma_q)}{\sum_{1 \leq k \leq Q} [w_k \mathcal{N}(g_j | \mu_k, \sigma_k)]} \quad \forall j, q \quad (\text{A.54})$$

Based on the definition in Equation A.54 the effective number of grey values assigned to a Gaussian component  $q$  is defined by:

$$N_q = \sum_{1 \leq j \leq n} \gamma(z_{jq}) \quad \forall q \quad (\text{A.55})$$

Plugging Equations A.54 and A.55 into Equations A.51 and A.52 and rearrangement after setting the partial derivatives to zero gives:

$$\mu_q = \frac{1}{N_q} \cdot \sum_{1 \leq j \leq n} [\gamma(z_{jq})g_j] \quad (\text{A.56})$$

$$\sigma_q^2 = \frac{1}{N_q} \cdot \sum_{1 \leq j \leq n} [\gamma(z_{jq})(g_j - \mu_q)^2] \quad (\text{A.57})$$

Setting the left term in Equation A.53 to zero gives:

$$0 = \sum_{1 \leq j \leq n} \left\{ \frac{\mathcal{N}(g_j | \mu_q, \sigma_q)}{\sum_{1 \leq k \leq Q} [w_k \mathcal{N}(g_j | \mu_k, \sigma_k)]} \right\} + \lambda \quad (\text{A.58})$$

If both sides in Equation A.58 are multiplied by  $w_q$  and summed over  $q$ , the constraint  $\sum w_q = 1$  provides the relation  $\lambda = -n$ . Using this relation and plugging Equation A.55 into Equation A.58 gives after some rearrangement:

$$w_q = \frac{N_q}{n} \quad (\text{A.59})$$

Imagine that if the latent variables  $\gamma(z_{jq})$  were known, then the optimal GMM parameters could be determined, because then Equations A.56, A.61 and A.59 would have a closed form. However, for the definition of the latent variables  $\gamma(z_{jq})$ , the GMM parameters need to be known (cf. Equation A.54).

## Expectation and maximization

To overcome the described problem Dempster et al. [1977] introduced the EM algorithm that follows an iterative strategy, i.e. it alternatively adjusts the GMM parameters or the latent variables. The algorithm is subdivided into the three steps: *Initialization*, *expectation* and *maximization*.

**Initialization:** The first step requires an approximate solution for the GMM parameters. Usually, this is realized as the output of another method such as K-means [Bishop, 2006]. However, in the context of the colour module the task is rather simple, because the data  $g_j$  are one-dimensional, and hence a simple heuristic is sufficient. This is realized by four equally spaced Gaussian components that have equal mixing coefficients and a low overlap (cf. top row in Figure A.1). In particular, the initial parameters are defined with respect to a marginal distribution of an 8 bit band of a colour image ( $g_j = \{1, \dots, 256\} \forall j$ ) such that for all  $q = \{1, 2, 3, 4\}$ :

$$\mu_q = 32 + 32q \quad (\text{A.60})$$

$$\sigma_q = 16 \quad (\text{A.61})$$

$$w_q = \frac{1}{4} \quad (\text{A.62})$$

The definition  $Q = 4$  is in contradiction with the initially formulated goal of determining just one or two Gaussian components. However, more components have to be taken into account at the start



because vehicles, trees or road markings might also appear as rather small Gaussian components. These components will finally be ignored but their consideration in the EM algorithm is important as they bias the components to be determined if they are not considered. Empirical tests with different image resolutions and road types indicated an initialization with  $Q = 4$  components to be appropriate for standard road segments.

**Expectation:** According to Equation A.54 the latent variables  $\gamma(z_{jq})$  are computed from the previously defined GMM parameters.

**Maximization:** According to Equations A.56, A.61 and A.59 the optimal GMM parameters are computed from the previously defined latent variables.

**Iteration:** Beginning with the *initialization*, the algorithm alternates between *expectation* and *maximization* until a maximum number of iterations  $T_{it}$  has been reached or the changes to the parameters are smaller than  $T_\delta$ . These thresholds, representing a trade off between performance and computation time are set to:

$$T_{it} = 50 \quad (\text{A.63})$$

$$T_\delta = 0.1 \quad (\text{A.64})$$

It is worth noting that the maximization step alone provides an optimal solution given the data but the EM algorithm does not, because it might get stuck at a local minimum. Figure A.1 illustrates iterative parameter adaptation with an example derived from the red band of the image depicted in Figure 3.7d. Superfluous components, such as the component  $q = 3$  in Figure A.1, are often characterized by large standard deviations  $\sigma_q$  or small mixing coefficients  $w_q$ .

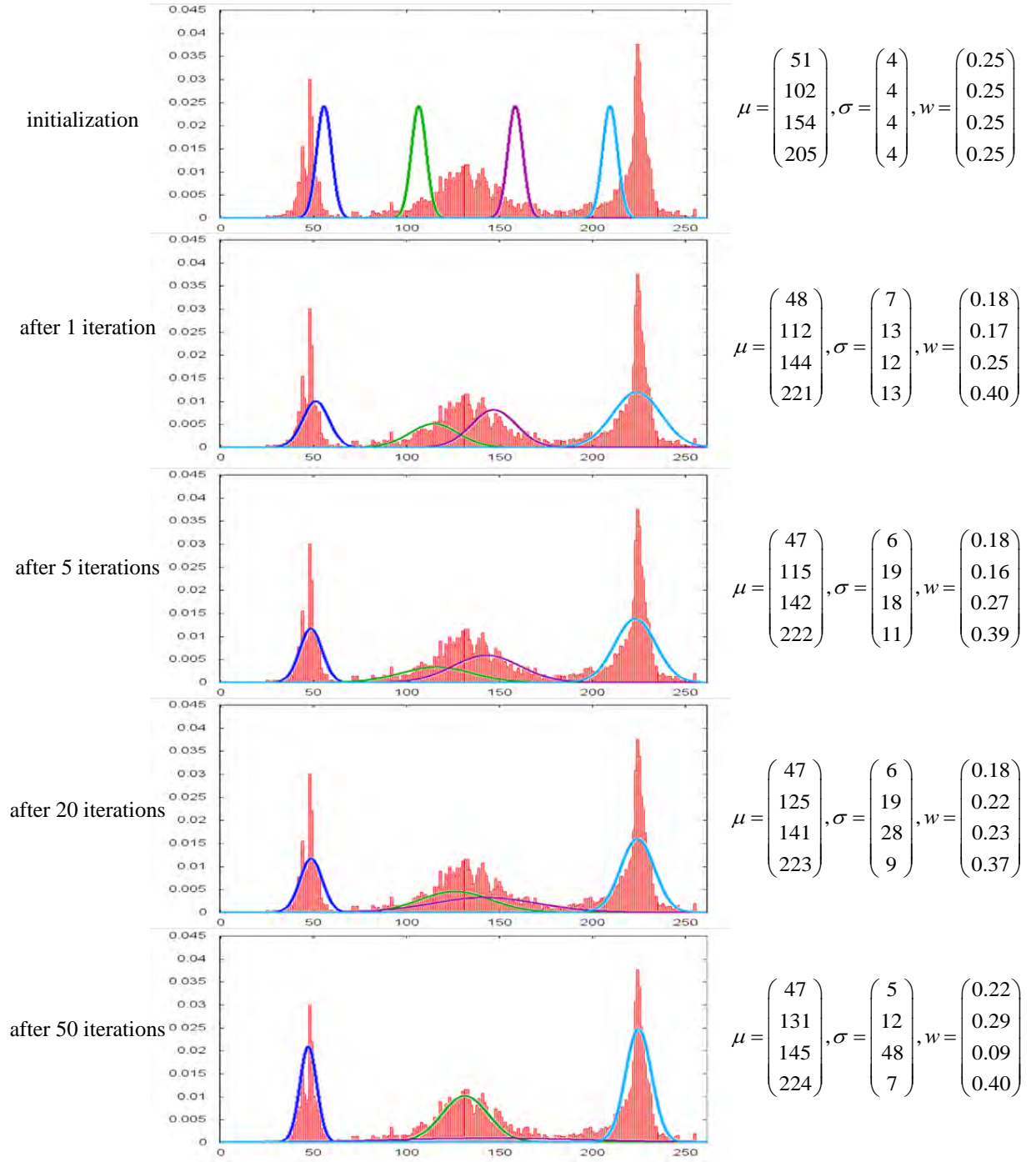


Figure A.1.: Iterative approximation of a Gaussian Mixture Model using the Expectation Maximization algorithm with the parameters described in the text.

### A.3. Images used for the experiments described in the thesis



Figure A.2.: EuroSDR-Ikonos1sub1:  $1.6 \times 1.6$  km, 1 m GSD, Kosovo.





Figure A.3.: EuroSDR-Ikonos3sub1:  $1.6 \times 1.6$  km, 1 m GSD, Kosovo.



Figure A.4.: EuroSDR-Ikonos3sub2:  $1.6 \times 1.6$  km, 1 m GSD, Kosovo.





Figure A.5.: EuroSDR-Aerial1:  $2.0 \times 2.0$  km, 0.5 m GSD, Switzerland.



Figure A.6.: EuroSDR-Aerial2:  $2.0 \times 2.0$  km, 0.5 m GSD, Switzerland.



Figure A.7.: EuroSDR-Aerial3:  $2.0 \times 2.0$  km, 0.5 m GSD, Switzerland.





Figure A.8.: ISPRS-Vaihingen:  $0.88 \times 1.18$  km, 0.2 m GSD, Germany.





Figure A.9.: MGCP-Algiers:  $12.0 \times 24.0$  km, 1.0 m GSD, Algeria.





Figure A.10.: GSI-Uraga:  $0.76 \times 0.92$  km, 0.2 m GSD, Japan.





Figure A.11.: NGI-Zeebrugge:  $12.0 \times 16.0$  km, 0.5 m GSD, Belgium.

# Bibliography

- Auclair-Fortier MF, Ziou D, Armenakis C, Wang S (1999) *Survey of Work on Road Extraction in Aerial and Satellite Images*. Sherbrooke, Qc, Canada: Universite de Sherbrooke, Technical Report 247.
- Auclair-Fortier MF, Ziou D, Armenakis C, Wang S (2001) Automated Correction and Updating of Road Databases from High-Resolution Imagery. *Canadian Journal of Remote Sensing*, 27 (1).
- Bacher U, Mayer H (2005) Automatic road extraction from multispectral high resolution satellite images. In: *Stilla U, Rottensteiner F, Hinz S (Eds) CMRT*. ISPRS, 36 (3/W24).
- Bammer G, Smithson M (2008) *Uncertainty and Risk: Multidisciplinary Perspectives*. Earthscan risk and society series. Earthscan.
- Barnett JA (1991) Calculating Dempster-Shafer Plausibility. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13: 599–602.
- Baumgartner A, Steger C, Mayer H, Eckstein W, Ebner H (1999) Automatic Road Extraction Based on Multi-Scale, Grouping, and Context. *Photogrammetric Engineering & Remote Sensing*, 65 (7): 777–785.
- Becker C, Pahl M, Ostermann J (2012) Mono-Temporal GIS Update Assistance System based on unsupervised Coherence Analysis and Evolutionary Optimisation. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Copernicus GmbH, 1 (7): 233–238.
- Beumier C, Lacroix V (2006) Road extraction for EuroSDR contest. In: *Proceedings of the SPIE edited by Lorenzo Bruzzone*. Image and Signal Processing for Remote Sensing, 6365: 335–338.
- Beyen J, Henrion J, Velde SVD (2008) Selection and customization of an integrated Digital Photogrammetric Workstation + GIS configuration and optimization of the interoperability within the workflow for updating the Belgian topographical reference database. In: *IntArchPhRS. Band XXXVII, Part B*: 335–338.
- Beyen J, Ziems M, Müller S, Roovers S, Heipke C (2012) Semi-automatic update and quality control of road databases. In: *Proceedings of the 4th GEO-BIA*: 443–448.
- Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35: 99–109.
- Bishop CM (2006) *Pattern Recognition and Machine Learning*. New York, USA: Springer.
- Bordes G, Giraudon G, Jamet O (1997) Automatic road extraction from grey-level images based on object database. *Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, 3072: 110–118.
- Box GEP, Tiao GC (1992) *Bayesian Inference in Statistical Analysis (Wiley Classics Library)*. Wiley-Interscience, reprint edition.
- Breiman L (2001) Random Forests. In: *Machine Learning*. Kluwer Academic Publishers, 45: 5–32.
- Bruzzone L, Marconcini M (2010) Domain adaptation problems: a DASVM classification technique and a circular validation strategy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32 (5): 770–787.
- Burger W, Burge MJ (2005) *Digitale Bildverarbeitung: Eine Einführung mit Java und ImageJ (eXamen.press)*. Springer, 1 edition.
- Burges CJC (1998) A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge. Discovery*, 2 (2): 121–167.
- Burges CJC (1999) Geometry and invariance in kernel based methods. In: Smola A, Bartlett PL, Schölkopf B, Schuurmons D (eds) *Advances in kernel methods: support vector learning* (pp. 89–116). Cambridge, MA, USA: MIT Press.
- Büschendorf T, Ostermann J (2012) Automatic Refinement of Training Data for Classification of Satellite Imagery. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Science*. Copernicus GmbH, 1 (7): 117–122.
- Butenuth M, Frey D, Nielsen AA, Skriver H (2011) Infrastructure assessment for disaster management using multi-sensor and multi-temporal remote sensing imagery. *International Journal of Remote Sensing*, 32 (23): 8575–8594.
- Butenuth M, Heipke C (2012) Network snakes: graph-based object delineation with active contour models. *Machine Vision and Applications*, 23: 91–109.
- Canny J (1986) A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8 (6): 679–698.

- Chang CC, Lin CJ (2001) LIBSVM: A library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> (accessed 5/5/4010).
- Clode S, Rottensteiner F, Kootsookos P, Zelniker E (2007) Detection and Vectorization of Roads from Lidar Data. *Photogrammetric Engineering & Remote Sensing*, 73(5) (5): 517–536.
- Clode S, Rottensteiner F, Kootsookos PJ (2005) Improving City Model Determination By Using Road Detection From Lidar Data. In: *Joint Workshop of ISPRS and the German Association for Pattern Recognition (DAGM)*: 159–164.
- Cramer M (2010) The DGPF test on digital aerial camera evaluation – overview and testdesign. *Photogrammetrie Fernerkundung Geoinformation*, 2: 73–82.
- Das S, Mirnalinee TT, Varghese K (2011) Use of Salient Features for the Design of a Multi-stage Framework to Extract Roads From High-Resolution Multispectral Satellite Images. *IEEE Transactions on Geoscience and Remote Sensing*, 49 (10): 3906–3931.
- Dempster AP (1967) Upper and Lower Probabilities Induced by a Multi-valued Mapping. *Annals of Mathematical Statistics*, 38: 325–339.
- Dempster AP (1968) A Generalization of Bayesian Inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30 (2): 205–247.
- Dempster AP, Laird NM, Rubin DB (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39 (1): 1–38.
- Dezert J (2002) Foundations for a new theory of plausible and paradoxical reasoning. *Information and Security Journal*, 9: 2002.
- Doucette P, Agouris P, Stefanidis A, Musavi M (2001) Self-Organised Clustering for Road Extraction in Classified Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 55 (5–6): 347–358.
- Farkas I (2009) Key points and the most significant documents in the production of the High Resolution Vector Data (HRVD) within the Multinational Geospatial Coproduction Program (MGCP). *Academic and Applied Research in Military Science (AARMS)*, 8 (1): 141–149.
- Fischler MA, Tenenbaum JM, Wolf HC (1981) Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. *Computer Graphics and Image Processing*, 15 (3): 201–223.
- Foody GM, Mathur A (2004) A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42 (6): 1335–1343.
- Freund Y, Schapire RE (1995) A decision-theoretic generalization of on-line learning and an application to boosting. In: *Proceedings of the Second European Conference on Computational Learning Theory*: 23–37.
- Fua P (1996) Model-Based Optimization: an Approach to Fast Accurate and Consistent Site Modeling from Imagery. In: *RADIUS: Image Understanding for Imagery Intelligence, O. Firschein and T. Strat*: 12915–2.
- Fujimura H, Ziems M, Heipke C (2008) Degeneralization of Japanese road data using satellite imagery. *Photogrammetrie Fernerkundung Geoinformation*, 5: 363–373.
- Gamba P, Dell’Acqua F, Lisini G (2006) Improving urban road extraction in high-resolution images exploiting directional filtering, perceptual grouping, and simple topological concepts. *Geoscience and Remote Sensing Letters, IEEE*, 3 (3): 387–391.
- Gerke M (2005) Automatic Quality Assessment of GIS Road Data using Aerial Imagery - A Comparison between Bayesian and Evidential Reasoning. In: *Stilla U, Rottensteiner F, Hinz S (Eds) CMRT. ISPRS, XXXVI*: 171–177.
- Gerke M (2006) *Automatic Quality Assessment of Road Databases Using Remotely Sensed Imagery*. PhD thesis, Series C, No. 599. Deutsche Geodätische Kommission.
- Gerke M, Busch A (2005) Verification of a Digital Road Database using Ikonos Imagery. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. ISPRS, XXXVI-1/W3*: 11–17.
- Gerke M, Butenuth M, Heipke C, Willrich F (2004) Graph-supported verification of road databases. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58 (3-4): 152–165.
- Gerke M, Heipke C (2008) Image-based quality assessment of road databases. *International Journal of Geographical Information Science*, 22 (8): 871–894.
- Göpfert J, Rottensteiner F, Heipke C (2011) Using snakes for the registration of topographic road database objects to ALS features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66 (6): 858–871.
- Gordon J, Shortliffe EH (1984) The Dempster-Shafer theory of evidence. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, 3: 832–838.



- Grote A, Heipke C, Rottensteiner F (2012) Road Network Extraction in Suburban Areas. *The Photogrammetric Record*, 27 (137): 8–28.
- Gruen A, Li H (1995) Road extraction from aerial and satellite images by dynamic programming. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50 (4): 11–20.
- Guan JW, Bell DA (1991) *Evidence Theory and Its Applications*. New York, USA: Elsevier Science Inc.
- Guiasu S, Shenitzer A (1985) The principle of maximum entropy. *The Mathematical Intelligencer*, 7 (1): 42–48.
- Guo SM, Chen LC, Tsai SH (2009) A boundary method for outlier detection based on support vector domain description. *Pattern Recognition*, 42 (1): 77–83.
- Haenni R (2005) Shedding new light on Zadeh's criticism of Dempster's rule of combination. In: *Information Fusion, 2005 8th International Conference*. IEEE, 2: 6 pp.
- Haverkamp D (2002) Extracting straight road structure in urban environments using IKONOS satellite imagery. *Optical Engineering*, 41 (9): 2107–2110.
- Hedman K (2010) *Statistical fusion of multi-aspect Synthetic Aperture Radar data for automatic road extraction*. Number 654 in PhD thesis, Series C. Deutsche Geodätische Kommission.
- Heipke C, Steger C, Multhammer R (1995) A Hierarchical Approach To Automatic Road Extraction From Aerial Imagery. In: *Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II*: 222–231.
- Helmholz P, Becker C, Breitzkopf U, Büschenfeld T, Busch A, Braun C, Grünreich D, Müller S, Ostermann J, Pahl M, Rottensteiner F, Vogt K, Ziemis M, Heipke C (2012) Semi-automatic Quality Control of Topographic Data Sets. *Photogrammetric Engineering & Remote Sensing*, 78 (9): 959–972.
- Hinz S, Baumgartner A (2003) Automatic extraction of urban road networks from multi-view aerial imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58 (1-2): 183–98.
- Hsu CW, Chang CC, Lin CJ (2003) *A Practical Guide to Support Vector Classification*. Department of Computer Science, National Taiwan University, Technical Report 2.
- Hu X, Tao CV, Hu Y (2004) Automatic Road Extraction From Dense Urban Area by Integrated Processing of High Resolution Imagery and LIDAR Data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35 (Part B3): 288–292.
- Huang C, Davis LS, Townshend JRG (2002) An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23 (4): 725–749.
- Jin X, Davis CH (2005) An integrated system for automatic road mapping from high-resolution multispectral satellite imagery by information fusion. *Information Fusion*, 6 (4): 257–273.
- Kass M, Witkin A, Terzopoulos D (1988) Snakes: Active contour models. *International Journal of Computer Vision*, 1 (4): 321–331.
- Klang D (1998) Automatic detection of changes in road databases using satellite imagery. In: *International Archives of Photogrammetry and Remote Sensing, Vol. 32*: 293–298.
- Klein LA (2004) *Sensor and Data Fusion: A Tool for Information Assessment and Decision Making (SPIE Press Monograph Vol. PM138)*. SPIE- International Society for Optical Engineering.
- Koutaki G, Uchimura K, Hu Z (2006) Refining road map using active shape model from aerial images. In: *Vision Geometry*. SPIE, 6066 (XIV): 3079–3089.
- Kuo-Tu K (1995) *A Hybrid Road Identification System Using Image Processing Techniques and Back Propagation Neural Networks*. Mississippi State, MS, USA: Mississippi State University, Technical report.
- le Hégarat-Masclé S, Bloch I, Vidal-Madjar D (1997) Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 35 (4): 1018–1031.
- Lee T, Richards J, Swain P (1987) Probabilistic and Evidential Approaches for Multisource Data Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 25 (3): 283–293.
- Lin X, Shen J, Liang Y (2012) Semi-Automatic Road Tracking using Parallel Angular Texture Signature. *Intelligent Automation & Soft Computing*, 18 (8): 1009–1021.
- Lu D, Batistella M, Moran E, de Miranda EE (2008) A Comparative Study of Landsat TM and SPOT HRG Images for Vegetation Classification in the Brazilian Amazon. *Photogrammetric Engineering & Remote Sensing*, 74 (3): 311–321.
- Ma L, Li J, Chen J (2008) Upgrading fundamental GIS databases for navigation from high resolution satellite imagery. In: *IntArchPhRS. Band XXXVII, Part B*: 335–338.
- Mayer H (1998) *Automatische Objektextraktion aus digitalen Luftbildern*. Number 494 in PhD thesis, Series C. Deutsche Geodätische Kommission.

- Mayer H, Baltsavias E, Bacher U (2006) Automated extraction, refinement, and update of road databases from imagery and other data. In: *Report Commission 2 on Image Analysis and Information Extraction; EuroSDR Official Publication 50*: 217–280.
- McKeown DMJ, Denlinger JL (1988) Cooperative methods for road tracking in aerial imagery. In: *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR '88., Computer Society Conference on*: 662–672.
- Mena J (2003) State of the art on automatic road extraction for GIS update: a novel classification. *Pattern Recogn. Lett.*, 24 (16): 3037–3058.
- Mena J, Malpica J (2005) An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. *Pattern Recognition Letters*, 26 (9): 1201–1220.
- Mnih V, Hinton GE (2010) Learning to detect roads in high-resolution aerial images. In: *Proceedings of the 11th European Conference on Computer Vision: Part VI*: 210–223.
- Nevatia R, Babu KR (1980) Linear feature extraction and description. *Computer Graphics and Image Processing*, 13 (3): 257–269.
- Oommen T, Misra D, Twarakavi N, Prakash A, Sahoo B, Bandopadhyay S (2008) An Objective Analysis of Support Vector Machine Based Classification for Remote Sensing. *Mathematical Geosciences*, 40: 409–424.
- Park J, Kang D, Kwok J, Lee SW, Hwang BW, Lee SW (2007) SVDD -based pattern denoising. *Neural Computation*, 19 (7): 1919–1938.
- Peng T, Jermyn I, Prinnet V, Zerubia J (2010) Extended Phase Field Higher-Order Active Contour Models for Networks: its Application to Road Network Extraction from VHR Satellite Images. *International Journal of Computer Vision*, 88 (1): 111–128.
- Platt J (2000) Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In: Smola A, Bartlett PL, Schölkopf B, Schuurmons D (eds) *Advances in Large Margin Classifiers* (pp. 61–74). Cambridge, MA, USA: MIT Press.
- Porway J, Wang Q, Zhu SC (2010) A Hierarchical and Contextual Model for Aerial Image Parsing. *International Journal of Computer Vision*, 88 (2): 254–283.
- Poulain V, Inglada J, Spigai M, Tournet JY, Marthon P (2010) High resolution optical and SAR image fusion for road database updating. In: *IEEE International Geoscience and Remote Sensing Symposium, Hawaii, USA*: 2747–2750.
- Poulain V, Inglada J, Spigai M, Tournet JY, Marthon P (2011) High-Resolution Optical and SAR Image Fusion for Building Database Updating. *IEEE Transactions on Geoscience and Remote Sensing*, 49 (8): 2900–2910.
- Poullis C, You S (2010) Delineation and geometric modeling of road networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65 (2): 165–181.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Numerical recipes : the art of scientific computing*. Cambridge University Press, 3 edition.
- Quam LH (1978) *Road Tracking and Anomaly Detection In Aerial Imagery*. 333 Ravenswood Ave., Menlo Park, CA 94025: AI Center, SRI International, Technical Report 158.
- Ravanbakhsh M, Heipke C, Pakzad K (2008) Road junction extraction from high resolution aerial imagery. *The Photogrammetric Record*, 23 (124): 405–423.
- Rottensteiner F, Sohn G, Jung J, Gerke M, Bailard C, Benitez S, Breitkopf U (2012) The ISPRS Benchmark on Urban Object Classification and 3D Building Reconstruction. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-3: 293–298.
- Rottensteiner F, Trinder J, Clode S, Kubik K (2007) Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62 (2): 135–149.
- Ruskoné R, Airault S (1997) Toward an automatic extraction of the road network by local interpretation of the scene. In: *Photogrammetric Week 97. Eds: D. Fritsch and D. Hobbie. Wichmann Verlag, Heidelberg*: 147–157.
- Ruspini EH, Lowrance JD, Strat TM (1992) Understanding evidential reasoning. *International Journal of Approximate Reasoning*, 6 (3): 401–424.
- Schindler K (2012) An Overview and Comparison of Smooth Labeling Methods for Land-Cover Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50 (11): 4534–4545.
- Schölkopf B, Platt JC, Shawe-Taylor JC, Smola AJ, Williamson RC (2001) Estimating the Support of a High-Dimensional Distribution. *Neural Computing*, 13 (7): 1443–1471.
- Schölkopf B, Smola AJ, Williamson RC, Bartlett PL (2000) New Support Vector Algorithms. *Neural Computing*, 12 (5): 1207–1245.
- Shafer G (1976) *A Mathematical Theory of Evidence*. Princeton University Press.

- Smets P (1990) The Combination of Evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (5): 447–458.
- Smets P (1994) What is Dempster-Shafer's model? In: Yager RR, Kacprzyk J, Fedrizzi M (eds) *Advances in the Dempster-Shafer theory of evidence* (pp. 5–34). New York, NY, USA: John Wiley & Sons, Inc.
- Soille P (2003) *Morphological Image Analysis: Principles and Applications*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2 edition.
- Song M, Civco D (2004) Road extraction using SVM and image segmentation. *Photogrammetric Engineering & Remote Sensing*, 70 (12): 1365–1371.
- Steger C (1998) An Unbiased Detector of Curvilinear Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (2): 113–125.
- Tax DMJ (2001) *One-class classification*. PhD thesis. Delft University of Technology, Netherlands.
- Tax DMJ, Duin RPW (2004) Support Vector Data Description. *Machine Learning*, 54 (1): 45–66.
- Tupin F, I.Bloch, Maitre H (1999) A First Step Toward Automatic Interpretation of SAR Images Using Evidential Fusion of Several Structure Detectors. *IEEE Transactions on Geoscience and Remote Sensing*, 37 (3): 1327.
- Tupin F, Maitre H, Mangin JF, marie Nicolas J, Pechersky E (1998) Detection of Linear Features in SAR Images: Application to Road Network Extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 36 (2): 434–453.
- Vapnik VN (1995) *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc.
- Wiedemann C, Ebner H (2000) Automatic Completion And Evaluation Of Road Networks. In: *International Archives of Photogrammetry and Remote Sensing*, XXXIII Part B3/2: 979–986.
- Wiedemann C, Mayer H (1996) Automatic Verification of Roads in Digital Images Using Profiles. In: *18th Annual Symposium of the German Association for Pattern Recognition DAGM 1996*: 609–618.
- Williams CKI (2002) On a Connection between Kernel PCA and Metric Multidimensional Scaling. *Machine Learning*, 46: 11–19.
- Youn J, Bethel JS, Mikhail EM, Lee C (2008) Extracting urban road networks from high-resolution true orthoimage and Lidar. *Photogrammetric Engineering & Remote Sensing*, 74 (2): 227–238.
- Zadeh L (1965) Fuzzy sets. *Information and Control*, 8 (3): 338–353.
- Zadeh L (1984) Book Review: A Mathematical Theory of Evidence. *AI Magazine*, 5 (3): 81–83.
- Zhang C (2004) Towards an operational system for automated updating of road databases by integration of imagery and geodata. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58 (3-4): 166–186.
- Zhang Q, Couloigner I (2006) Automated road network extraction from high resolution multi-spectral imagery. In: *Proceedings of the ASPRS Annual Conference, Reno, Nevada*
- Ziems M, Beyen J, Müller S, Roovers S, Heipke C (2011a) Multiple-Model based update of Belgian reference road database. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. ISPRS, XXXVIII-part4: 73–81.
- Ziems M, Bretkopf U, Heipke C, Rottensteiner F (2012) Multiple Model Based Verification of Road Data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1 (3): 329–334.
- Ziems M, Fujimura H, Heipke C, Rottensteiner F (2010) Multiple-model based verification of Japanese road data. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. ISPRS, XXXVII-4-8-2/W9: 13–19.
- Ziems M, Gerke M, Heipke C (2007) Automatic road extraction from remote sensing imagery incorporating prior information and colour segmentation. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. TUM Munich, XXXVI-3 W49A: 141–147.
- Ziems M, Heipke C, Rottensteiner F (2011b) SVM-based road verification with partly non-representative training data. In: *6th Urban Remote Sensing Event (JURSE)*: 37–40.



# Curriculum vitae – Marcel Ziems

## *work experience and education*

since May 2013	collaborator, Geoinformation Agency of Lower Saxony
August 2006 – May 2013	PhD student and scientific collaborator, Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover
April 2012 – May 2012	scientific collaborator, Pontifical Catholic University, Rio de Janeiro, Brasilien
June 2009 – Octobre 2009	scientific collaborator, Geographical Survey Institute (GSI), Tsukuba, Japan
May 2004 – June 2006	internship, Geoinformation Agency of Lower Saxony
August 2002 – March 2004	student assistant, Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover
Octobre 1997 – Septembere 2003	course, geodesy and geoinformatics, Leibniz Universität Hannover
Septembre 1996 – June 1997	military service, Federal Armed Forces, Germany
Septembre 1983 – June 1996	high school, Goetheschule Kyritz, Gymnasium Friedrich Ludwig Jahn, Kyritz

## *private*

August 2010	birth daughter, Rika Ziems
July 2007	birth son, Lasse Ziems
June 2007	marriage, Franziska Ziems née Schnabl
January 1977	born in Pritzwalk parents: Hans-Wolf Ziems, Christiane Ziems née Klünder





# Acknowledgements

First and foremost I thank my advisor Christian Heipke for providing careful guidance throughout my PhD. It is due to his expertise, patience and faith that I could evolve my own research ideas.

Additionally, I largely benefited from inspiring discussions with Franz Rottensteiner who contributed many structural and mathematical aspects to this work. Furthermore, I would also like to thank Monika Sester and Jörn Ostermann for acting as referees of this thesis.

I thank all my colleagues and friends at the Institute of Photogrammetry and GeoInformation (IPI) at the Leibniz Universität Hannover for the creative and kind working atmosphere. I am particularly thankful to Sönke Müller I shared the office and lots of challenging problems with. I thank Markus Gerke for providing the basis of my research and for patiently introducing me into the thesis-related project work. I want to thank Uwe Breitkopf for scripting the execution and analysis of all experimental settings which brought me into the comfortable situation to found my thesis on one of the most comprehensive data bases in this research field without much workload.

Moreover, I thank my colleagues and friends from the project at the Institut für Informationsverarbeitung (TNT) at the Leibniz Universität Hannover Martin Pahl, Christian Becker, Thorsten Büschenfeld and Karsten Vogt for the professional and kind collaboration. Many insights into programming and computer vision have been initiated by this great team.

I also thank my friends Ann-Kristin and Martin as well as my friend friend Nestor for proofreading the English text.

I gratefully acknowledge the financial support of the Federal Agency for Cartography and Geodesy of Germany (BKG) and in particular the personal commitment of Dietmar Grünreich who initiated and organized the thesis-related project. Furthermore, I want to thank Carola Braun and her team for their financial support and beta-testing the developed software. I would also like to express my sincere gratitude to the Geo-spatial Information Authority of Japan (GSI), and in particular Hidenori Fujimura for the fruitful scientific exchange and a most memorable time in Japan. Furthermore, I benefited a lot from the collaboration with the Belgian National Geographical Institute (NGI) as they provided lots of reference data. In particular it was Jan Beyen who showed me many practical aspects of manual road mapping, which significantly influenced my work. Moreover, I thank the organizers of the benchmark test Helmut Mayer and Uwe Bacher for providing test and reference data.

Finally, I bow to my family. I thank my parents Christiane and Hans-Wolf as well as my brother Christoph for their holistic support. I thank my dear children Lasse and Rika for their permission to write the thesis and most of all, I thank my dear wife Franziska for keeping everything running whenever it was necessary.

Although our intellect always longs for clarity and certainty,  
our nature often finds uncertainty fascinating.

– Carl von Clausewitz

# Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover

*(Eine vollständige Liste der Wiss. Arb. ist beim Geodätischen Institut, Nienburger Str. 1, 30167 Hannover erhältlich.)*

Nr. 286	JARECKI, Focke:	Spurkreuzungspunktdifferenzen zur Validierung satellitengradiometrischer Messungen (Diss. 2010)
Nr. 287	SCHRAMM, Matthias:	Unüberwachte lineare spektrale Entmischung mit Anwendungen für Baumdichteabschätzungen in semiariden Regionen Afrikas (Diss. 2010)
Nr. 288	ALSAYEL, Mohammed Y.:	Towards Standardizing International Boundary Determination and Quality Consequences on surveying and Mapping (Diss. 2010)
Nr. 289	TAVAKKOLI S., Seyed Mohammad:	Multi-Temporal Classification of Crops Using ENVISAT ASAR Data (Diss. 2011)
Nr. 290	VENNEGEERTS, Harald:	Objektraumgestützte kinematische Georeferenzierung für Mobile-Mapping-Systeme (Diss. 2011)
Nr. 291	BULATOV, Dimitri:	Textured 3D reconstruction of urban terrain from UAV-borne video sequences (Diss. 2011)
Nr. 292	GROTE, Anne:	Automatic road network extraction in suburban areas from aerial images (Diss. 2011)
Nr. 293	ALOEID, Abdalla:	Assessment of Matching Algorithms for Urban DSM Generation from Very High Resolution Satellite Stereo Images (Diss. 2011)
Nr. 294	FANG, Xing:	Weighted Total Least Squares Solutions for Applications in Geodesy (Diss. 2011)
Nr. 295	WEGNER, Jan Dirk:	Detection and height estimation of buildings from SAR and optical images using conditional random fields (Diss. 2011)
Nr. 296	GÖPFERT, Jens :	Snakes for Adapting GIS Road and River Objects to Airborne Laser Scanning Data (Diss. 2011)
Nr. 297	MUHLE, Daniel:	Gegenseitige Orientierung von Mehrkamerasystemen mit nicht überlappendem Sichtfeld (Diss. 2011)
Nr. 298	SIRIBA, David Nyangau:	Congflation of Provisional Cadastral and Topographical Datasets (Diss. 2012)
Nr. 299	HELMHOLZ, Petra:	Verifikation von Ackerland- und Grünlandobjekten eines topographischen Datensatzes mit monotemporalen Bildern (Diss. 2012)
Nr. 300	DENG, Zhiguo:	GPS Meteorology with Single Frequency Receivers (Diss. 2012)
Nr. 301	ZEIBLER, Maik:	Zur Ermittlung von Bodenrichtwerten bei fehlenden Kaufpreisen unbebauter Grundstücke (Diss. 2012)
Nr. 302	PAFFENHOLZ, Jens-André:	Direct geo-referencing of 3D point clouds with 3D positioning sensors (Diss. 2012)
Nr. 303	WEINBACH, Ulrich:	Feasibility and impact of receiver clock modeling in precise GPS data analysis (Diss. 2012)
Nr. 304	HEIKER, Andrea:	Mutual validation of Earth orientation parameters, geophysical excitation functions and second degree gravity field coefficients (Diss. 2012)
Nr. 305	VOIGT, Christian:	Astrogeodätische Lotabweichungen zur Validierung von Schwerefeldmodellen (Diss. 2013)
Nr. 306	THIELE, Antje:	3D Building Reconstruction From High Resolution Multi-Aspect Interferometric Synthetic Aperture RADAR Data (Diss. 2013)
Nr. 307	ALRAJHI, Muhamad:	Semi-Automated Procedure for Orthophoto Generation from High Resolution Satellite Imagery (Diss. 2013)
Nr. 308	WANG, Jin:	Towards deformation monitoring with terrestrial laser scanning based on external calibration and feature matching methods (Diss. 2013)
Nr. 309	NAEIMI, Majid:	Inversion of satellite gravity data using spherical radial base functions (Diss. 2013)
Nr. 310	SEATOVIC, Dejan:	Methods for Real-Time Plant Detection in 3-D Point Clouds (Diss. 2013)
Nr. 311	SCHUNERT, Alexander:	Assignment of Persistent Scatterers to Buildings (Diss. 2014)
Nr. 312	GUERCKE, Richard:	Optimization Aspects in the Generalization of 3D Building Models (Diss. 2014)
Nr. 313	ZIEMS, Marcel:	Automatic Verification of Road Databases using multiple Road Models (Diss. 2014)

*Die Arbeiten werden im Rahmen des wissenschaftlichen Schriftenaustausches verteilt und sind nicht im Buchhandel erhältlich. Der Erwerb ist zu einem Stückpreis von € 25,00 bei den herausgebenden Instituten möglich.*