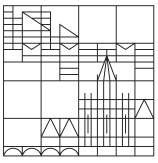
Semi-Automated Detection of Fragmented Rectangular Structures in High Resolution Remote Sensing Images with Application in Archaeology

Dissertation submitted for the degree of

Doctor of Natural Sciences (Dr. rer. nat.)
Doktor der Naturwissenschaften (Dr. rer. nat.)

Presented by
Igor Zingman
at the

Universität Konstanz



Faculty of Sciences
Department of Computer and Information Science

Date of oral examination:	**7.10.2016**
First referee:	**Prof. Dr. Dietmar Saupe**
Second referee:	**PD Dr. Karsten Lambers**
Third referee:	**Prof. Dr. Bastian Goldlücke**

Aknowledgements

I am grateful to my advisers, Dr. Karsten Lambers, whose initiative made this project possible and whose motivation drove this project at all the stages, and Prof. Dr. Dietmar Saupe, for his generous academic support and help in refining various technical solutions throughout my work. I am also grateful to my advisers for their insightful reviews of the published materials.

I would also like to thank Karsten Lambers, Thomas Reitmaier, and Christoph Walser for having an opportunity to explore by foot the regions of the project study in Silvretta Mountains.

I would like to thank Wenke Schimmelpfennig and Brigitte Andres for their interest in the Silvretta Historica project and for providing an image data of Bernese Alps.

I thank Otávio A. B. Penatti for the help in running implementations of extraction of OverFeat and HOG features for the purpose of performance comparison.

I would also like to thank Ms. Anna Dowden-Williams for proofreading some of my publications.

The work was financially supported in part by the European Interreg IV program Alpenrhein-Bodensee-Hochrhein (Silvretta Historica Project), Zukunftskolleg, University of Konstanz, DFG Research Training Group GK-1042 "Explorative Analysis and Visualization of Large Information Spaces", and DFG coloborative research SFB TRR 161 center "Quantitative Methods for Visual Computing".

Abstract

Automated visual analysis has substantially advanced in recent years, allowing a variety of targets to be automatically detected. Remarkably successful algorithms and technologies have been developed, e.g., for face detection and for obstacle detection for autonomous car navigation. In archaeology, however, remote sensing images are still analyzed in the traditional way, by a visual inspection. Such a visual inspection is performed prior to a field survey in order to identify potential sites that may guide later fieldwork. Though this approach saves fieldwork time, visual inspection remains very time consuming and requires the highly concentrated attention of an expert. Due to human fatigue, this approach might be unreliable. Moreover, the visual inspection of image data over vast unexplored areas is not feasible at all. This is especially frustrating, since a large amount of high resolution image data has become available due to recent developments of satellite technology. It is, therefore, very appealing to automate screening of large datasets of remote sensing images.

An automated screening does not aim at replacing visual image interpretation by an archaeologist. Though machine vision algorithms have already become quite powerful at some visual tasks, human vision and its ability to interpret scenes and recognize objects is clearly superior in general. On the other hand, in contrast to a human expert, machine vision algorithms are capable of routinely screening a large amount of imagery and generating plausible candidate locations. These findings can be verified easily and timely by an archaeologist, who can also estimate their potential significance. Such a semi-automated approach can increase the efficiency of an archaeological survey of vast unexplored areas.

In this thesis we develop a semi-automated methodology for detecting unknown rectangular structures, such as archaeological remains of livestock enclosures (LSE), in wide alpine areas covered by high resolution remote sensing images (HRRSI). The LSE structures were of a special interest in several recent archaeological studies, because such architectural remains offer important insights into the origins and historical development of alpine pasture economy. The LSEs have varying sizes and aspect ratios, may be heavily ruined, and may have spectral properties similar to the surrounding terrain and rocks. They appear in HRRSI images as faint fragmented usually approximately rectangular contours on a complex background.

As a part of our methodology we introduce particular image analysis algorithms that are briefly outlined below. We introduce an approach for segmenting out large contextually inappropriate regions of high texture contrast, such as urban areas, forests, or rocky areas. This approach is shown to be superior to other methods in terms of accuracy of segmentation at texture borders and ability to distinguish texture details from individual features, which might be a part of the structures we are searching for. We also introduce a complementary method that extracts individual image features, i.e. linear segments corresponding to walls of ruined LSEs, while suppressing background texture. A quantitative comparison with alternative approaches is also provided.

We propose a method for fast detection of initial candidates. It generates sparse locations that are, at least partially, enclosed by a structures of an arbitrary shape. Image patches at candidate locations are further analyzed based on dedicated rectangularity-size features introduced in the thesis. These features allow capturing rectangular enclosures, even if distorted, incomplete, or fragmented. On the other hand, they are not sensitive to a variety of irrelevant structures, such as isolated corners, line intersections, parallel curves, etc. The LSE structures are detected using a linear classifier fed with the rectangularity-size features. We have designed a dedicated linear classifier that is not prone to overfitting the data even in our case of extremely unbalanced data with only a few positive and a large number of negative examples. We quantitatively compare the effectiveness of the rectangularity-size features for our detection task with other handcrafted features and with state-of-the-art pre-trained deep CNN features.

The flow of the image analysis algorithms, which automatically generates detections and their confidence, is followed by a visual inspection by means of a specially designed graphical user interface (GUI). The GUI allows quick and convenient validation of true detections and rejection of falsely detected sites. We demonstrated the feasibility of our methodology by applying it to two large alpine regions. We were able to detect the LSE structures of interest, some of which were hitherto unknown.

Although, we developed the algorithms for processing HRRSI with a particular archaeological application in mind, they can be used in different domains and for different purposes. We have, for example, briefly discussed the application of some of the algorithms to detecting individual buildings in rural or mountainous areas.

Zusammenfassung

Im Bereich der automatisierten Bildauswertung hat es in den letzten Jahren substanzielle Fortschritte gegeben, so dass heute verschiedenste Zielobjekte automatisch detektiert werden können. Einige bemerkenswert leistungsfähige Algorithmen und Technologien wurden z.B. für die Gesichtserkennung oder die Detektion von Hindernissen durch selbstfahrende Autos entwickelt. Demgegenüber werden in der Archäologie Fernerkundungsbilder weiterhin auf traditionelle Weise ausgewertet, d.h. mittels visueller Überprüfung. Eine solche Auswertung wird im Vorfeld einer Geländeprospektion vorgenommen, um potenzielle Fundstellen zu identifizieren, die als Ausgangspunkte für die Feldarbeiten dienen können. Zwar spart ein solcher Ansatz Zeit während der Feldarbeiten, doch erfordert die visuelle Bildauswertung einen hohen Aufwand an Zeit und Konzentration durch Experten. Aufgrund der Grenzen menschlicher Belastbarkeit kann ein solcher Ansatz unzuverlssig sein. Zudem ist die visuelle Bildauswertung im Falle großer unerforschter Gebiete schlicht nicht leistbar. Dies ist unbefriedigend, da aufgrund neuer Entwicklungen der Satellitentechnologie immer mehr hochaufgelste Bilddaten zur Verfgung stehen. Daher erscheint das automatische Durchsuchen großer Bestände an Fernerkundungsbilddaten als attraktive Option.

Das Ziel einer solchen automatisierten Suche ist es dabei nicht, die visuelle Bildauswertung durch Archäologen zu ersetzen. Algorithmen des maschinellen Sehens sind zwar in einigen Anwendungsbereichen bereits heute sehr leistungsfähig, doch sind ihnen das menschliche Sehen und seine Fähigkeit, Szenen zu interpretieren und Objekte zu erkennen, weiterhin klar überlegen. Im Gegensatz zu menschlichen Experten sind Algorithmen des maschinellen Sehens jedoch in der Lage, routinemäßig große Bilddatenbestände zu durchsuchen und plausible Kandidaten zu lokalisieren. Die Ergebnisse können wiederum einfach und zeitsparend durch Archäologen überprüft werden, die gleichzeitig ihre Bedeutung einschätzen können. Ein solcher halbautomatischer Ansatz kann die Effizienz einer archäologischen Prospektion großer unerforschter Gebiete steigern.

In der vorliegenden Arbeit wird eine halbautomatische Methode zur Detektion bislang unbekannter rechteckiger Strukturen, konkret Ruinen von Viehpferchen, in hochaufgelösten

Fernerkundungsbildern einer weitläufigen alpinen Region entwickelt. Ruinen von Viehpferchen sind für verschiedene archäologische Studien von besonderem Interesse, da solche architektonischen Hinterlassenschaften wichtige Einblicke in die Ursprünge und Entwicklung der Alpwirtschaft erlauben. Die Viehpferche haben verschiedene Größen und Ausrichtungen, sind teilweise stark zerstört, und ihre spektralen Merkmale gleichen oft denen des umgebenden Geländes oder von Felsen. In hochaufgelösten Fernerkundungsbildern erscheinen sie als schwache, unvollständige, annähernd rechteckige Umrisse vor einem komplexen Hintergrund.

Als Teil der hier vorgestellten Methode werden bestimmte Bildanalysealgorithmen neu eingeführt, die im Folgenden kurz erläutert werden. Ein erster Ansatz dient dem Herausfiltern großer für die Aufgabenstellung irrelevanter Bildregionen mit hohem Texturkontrast, wie z.B. moderne Siedlungen, Waldgebiete oder felsige Regionen. Wie gezeigt werden kann, ist dieser Ansatz anderen Methoden im Hinblick auf die Genauigkeit der Segmentierung von Texturgrenzen überlegen, aber auch bei der Unterscheidung zwischen Texturdetails und isolierten Bildmerkmalen, die Teil der gesuchten Strukturen sein können. Eine weitere, komplementäre Methode extrahiert solche isolierten Bildmerkmale, z.B. lineare Segmente, die Mauern von Viehpferchen entsprechen können, und unterdrückt gleichzeitig die Textur des Bildhintergrundes. Beide Methoden werden auch quantitativ mit alternativen Methoden verglichen.

Desweiteren wird eine Methode für eine schnelle Detektion von Kandidaten vorgeschlagen, die als Ausgangspunkte dienen. Diese Methode generiert eine überschaubare Anzahl solcher Punkte, die zumindest teilweise von Strukturen verschiedener Formen umgeben sind. Die Bildbereiche, in denen sich solche Kandidaten befinden, werden sodann weiter auf die Präsenz bestimmter Rechteckigkeitsmerkmale analysiert, die in der vorliegenden Arbeit erstmals vorgestellt werden. Diese Merkmale erlauben die Erfassung von Rechtecken, selbst wenn diese verformt, unvollständig oder unterbrochen sind. Dabei erfassen sie jedoch keine irrelevanten Strukturen wie z.B. einzelne Ecken, Kreuzungen von Linien, parallele Kurven usw. Die Viehpferche werden mit Hilfe eines Linearklassifikators und der Rechteckigkeitsmerkmale detektiert. Der hier vorgestellte Linearklassifikator tendiert auch im hier auszuwertenden stark unausgewogenen Datensatz mit nur wenigen positiven und zahlreichen negativen Beispielen nicht zur Überanpassung. Die Rechteckigkeitsmerkmale werden im Hinblick auf ihre Effizienz für die hier gestellte Detektionsaufgabe quantitativ mit anderen maßgeschneiderten Merkmalen und mit vortrainierten

CNN-Merkmalen, die dem aktuellen Stand der Technik entsprechen, verglichen.

Auf die Abfolge von Bildanalysealgorithmen, die automatisch Detektionen sowie ein Maß für ihre Zuverlässigkeit generieren, folgt die visuelle Überprfung mit Hilfe eine dafür entworfenen graphischen Benutzeroberfläche. Diese erlaubt eine schnelle und bequeme Überprfung korrekter Detektionen und die Erkennung falsch detektierter Fundstellen. Die Anwendbarkeit dieser Methode wird anhand zweier alpiner Regionen demonstriert, in denen Viehpferche von archäologischem Interesse detektiert werden konnten, von denen einige bisher unbekannt waren.

Die hier vorgestellten Algorithmen zur Verarbeitung hochaufgelöster Fernerkundungsbilddaten wurden für eine bestimmte archäologische Anwendung entwickelt. Sie können jedoch auch in anderen Anwendungsbereichen und für andere Zwecke eingesetzt werden, wie eine Beispielanwendung der gleichen Algorithmen zur Detektion einzelner Gebäude in ländlichen und gebirgigen Regionen zeigt.

Parts of this thesis were published in the articles listed below. The articles are followed by chapter numbers in brackets with partially overlapping content.

- Igor Zingman, Dietmar Saupe, and Karsten Lambers. Morphological operators for segmentation of high contrast textured regions in remotely sensed imagery. In *Proc. of the IEEE Int. Geoscience and Remote Sensing Symposium*, pages 3451–3454, Munich, Germany, July 2012 (Ch. 4)

- Karsten Lambers and Igor Zingman. Towards detection of archaeological objects in high-resolution remotely sensed images: the Silvretta case study. In Graeme Earl et al., editors, *Archaeology in the Digital Era, vol. II (e-papers) from the 40th Conf. on Computer Applications and Quantitative Methods in Archaeology, Southampton, March 2012*, pages 781–791. Amsterdam University Press, 2013 (Ch. 1, 3, 4)

- Igor Zingman, Dietmar Saupe, and Karsten Lambers. Detection of texture and isolated features using alternating morphological filters. In *Proc. of the International Symposium on Mathematical Morphology and its applications to image and signal processing (ISMM)*, pages 440–451, Uppsala, Sweden, May 2013 (Ch. 4, 5)

- Igor Zingman, Dietmar Saupe, and Karsten Lambers. Automated search for livestock enclosures of rectangular shape in remotely sensed imagery. In Lorenzo Bruzzone, editor, *Proc. SPIE, Image and Signal Processing for Remote Sensing XIX*, volume 8892, pages 88920F–1 – 88920F–11, Dresden, Germany, 2013 (Ch. 7, 8)

- Igor Zingman, Dietmar Saupe, and Karsten Lambers. A morphological approach for distinguishing texture and individual features in images. *Pattern Recognition Letters*, 47:129 – 138, 2014. Advances in Mathematical Morphology (Ch. 4, 5)

- Igor Zingman, Dietmar Saupe, and Karsten Lambers. Detection of incomplete enclosures of rectangular shape in remotely sensed images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2015 (Ch. 7, 8, 9, 10)

- I. Zingman, D. Saupe, O. A. B. Penatti, and K. Lambers. Detection of fragmented rectangular enclosures in very high resolution remote sensing images. *IEEE Trans. on Geoscience and Remote Sensing*, 54(8):4580–4593, 2016 (Ch. 6, 7, 8, 9, 10, 13)

Contents

Acronyms

AMF	Amplitude Modulation Function
ASF	Alternating Sequential Filters
ASF diff.	ASF difference
AUC	Area Under ROC Curve
CC	Component Count
CNN	Convolutional Neural Networks
GIS	Geographical Information System
GLCM	Gray Level Co-occurrence Matrix
GUI	Graphical User Interface
GVF	Gradient Vector Flow
HRRSI	High Resolution Remote Sensing Images
LBP	Local Binary Patterns
LSE	Livestock Enclosures
MFC	Morphological Feature Contrast
MPP	Marked Point Processes
MTC	Morphological Texture Contrast
non-CRF	non-Classical Receptive Field
ROC	Receiver Operating Characteristic
SE	Structuring Element
SRTM	Space Shuttle Radar Topography Mission
StD	Standard Deviation
SVM	Support Vector Machines
TER	Texture Range

Chapter 1

Introduction

1.1 Background, goals, and challenges

Background and motivation

Automated image analysis has substantially advanced in recent years, allowing a variety of targets to be automatically detected. Remarkably successful algorithms and technologies have been developed for face detection and for obstacle detection for autonomous car navigation, [8, 9]. In archaeology, on the other hand, remote sensing images are still analyzed in the traditional way, i.e. by performing a visual inspection. Such a visual inspection is performed prior to a field survey in order to identify potential sites that may guide later fieldwork. Though this approach saves a considerable amount of fieldwork time, visual inspection remains very time consuming. In addition, it might be not reliable due to human fatigue. Moreover, visual inspection of the image data over vast unexplored areas is not feasible at all. This is especially frustrating, since a large amount of high resolution image data has become available due to recent developments of satellite technology. It is, therefore, very appealing to automate the screening of large datasets of remote sensing images, [10].

It should be noted that automated screening does not aim at replacing visual image interpretation by an archaeologist. Though machine vision algorithms have already become quite powerful at some visual tasks, human vision and its ability to interpret scenes and recognize objects is clearly superior in general. On the other hand, in contrast to

a human expert, machine vision algorithms are capable of routinely screening a large amount of imagery and generating plausible candidate locations. These findings can be verified easily and timely by an archaeologist, who can also estimate their potential significance. Such a semi-automated approach can guide and increase the efficiency of an archaeological survey of vast unexplored areas [2].

This research work accompanies an archaeological project called Silvretta Historica. It was initiated in 2010 (see [11]) by the University of Zurich, the University of Konstanz, and various project partners in the region of the Silvretta Alps located on the border between Austria and Switzerland. The Silvretta Historica project aims at studying human activity in the Silvretta Alps and at promoting cross-border tourism in this region (see [12, 13, 14, 15, 16]). Over several years of archaeological survey, dozens of archaeological sites were discovered. Among them there are sites with the ruins of livestock enclosures (LSE), which offer important insights into the origins and historical development of the alpine pasture economy [13, 16, 17]. Fig. 1.1 shows several examples of these structures.

Goals

The goal of this work is the development of a methodology and corresponding image analysis algorithms for the routine screening of large alpine areas covered by high resolution remote sensing images (either aerial or satellite images) at 0.5 m or higher resolution in order to detect archaeological sites. This work does not target the detection of all possible types of sites of archaeological interest, which is hardly an achievable task. Instead, it specifically focuses on the detection of rectilinear structures, such as the architectural remains of LSEs. These structures are of special interest for the Silvretta Historica project. The developed algorithms should be embedded into a prototypical graphical user interface that allows the efficient inspection of vast and unexplored areas, shows detections and their confidence, and allows quickly rejecting falsely detected sites. The interface should allow the user to adjust various parameters, such as those related to, for example, sensitivity/false detection rates trade-off or the range of the sizes of the structures. The alpine Silvretta region will serve the project as a case study area for the development and exploring the potential of the introduced methodology.

Although, in this thesis, algorithms are developed with an archaeological application in mind, they are not be limited to remote sensing images and can also be used in other geoscience related or completely different applications. As an example, we show in Sec.

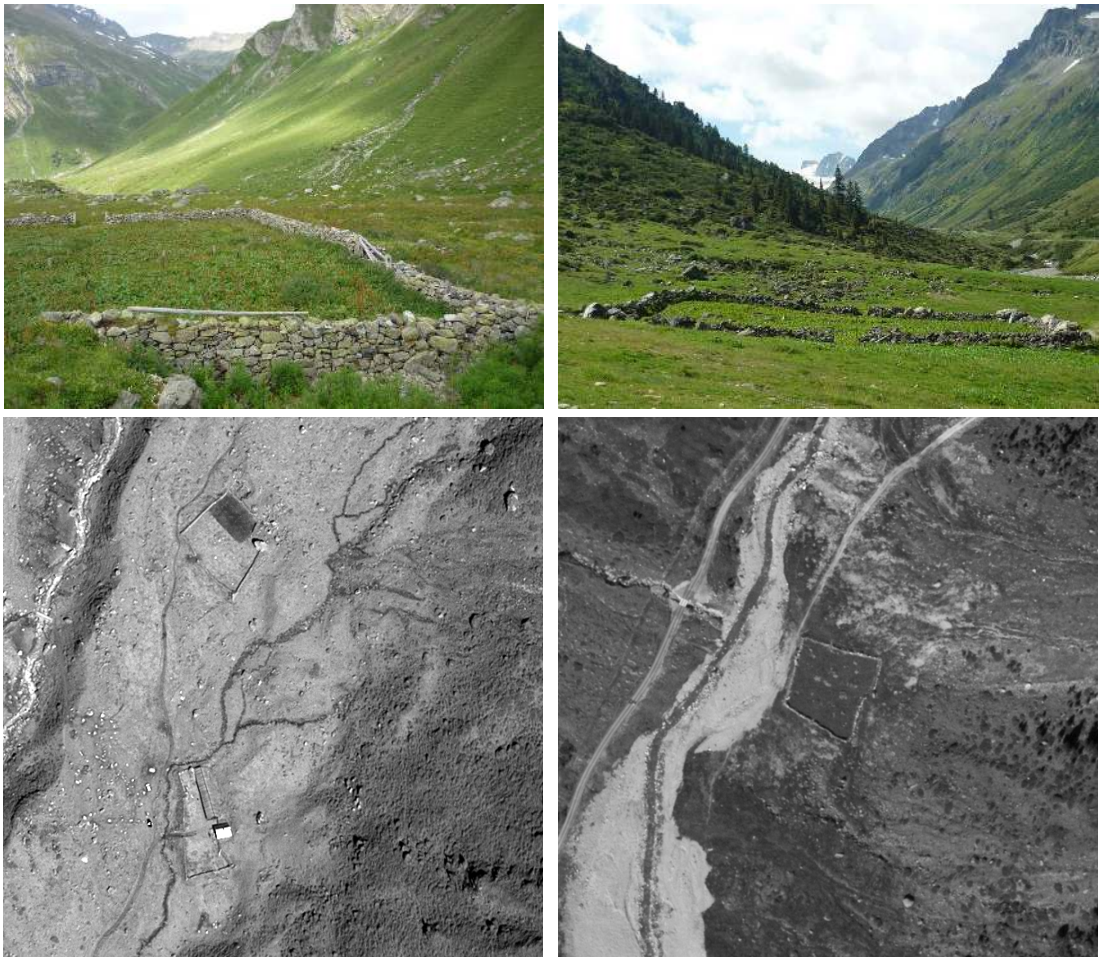


Fig 1.1: Above: Livestock enclosures (LSE) in Silvretta mountains. Below: 600×600 satellite (©GeoEye 2011) and aerial (SWISSTOPO) images of 0.5m resolution with structures corresponding to livestock enclosures in the pictures above. In the left satellite image in the topmost corner of the enclosure structure one can see excavation work. Note that in this image there is a second enclosure structure below.

13 that the introduced rectangularity feature is useful for the detection of buildings in rural or mountainous areas.

Challenges

The LSE structures are usually composed of linear walls that may be heavily ruined. The most common shape of LSEs resembles a rectangular contour with greatly varying size and aspect ratio. The rectangle's angles may deviate from right angles, and the rectangle's

sides may be fragmented. The angle between adjacent fragments of the same side may deviate from 180 degrees. Moreover, the rectangular contours are sometimes incomplete, so that even an entire side may be missing.

The width of a wall in an image at a resolution of 0.5 m does not exceed two pixels. The ruined walls are of low height, which results in low contrast linear features in the images. The spectral properties of LSEs are similar to the spectral properties of the surrounding terrain, rocks, and other irrelevant objects. Though the internal area of enclosures sometimes exhibits distinctive spectral signatures, they are not consistent from site to site and depend on the time an image was captured and on the type of imaging modality. The second row of Fig. 1.1 shows a satellite and an aerial image with structures corresponding to the LSEs shown above in the same figure. Nearby irrelevant structures, such as rivers, trails, or rocks, are often of similar or higher contrast, either due to their larger size (e.g. big rocks) or distinctive spectral properties (e.g. rivers). The detection of such faint enclosures in such a complex terrain is a very challenging task. Even the detection of easily modeled circular soil structures in [18] had very limited success due to their low contrast and the complexity of the terrain. In addition, only a few examples of LSEs (see Ch. 3 for details) are available in our case, which makes the standard approaches that learn from the data inappropriate. Due to the aforementioned difficulties, commonly used methods for the detection of rectilinear structures or rectangles are hardly applicable. In Sec. 1.2.2 we review some common approaches that have been previously used for the detection of different types of rectangular structures.

These problems caused by the complex and variable terrain, and the discrepancies in the appearances of the faint archaeological objects has resulted in the low effectiveness of automated methods [19]. This has also led to a limited number of publications in the literature on the whole topic in general. In Sec. 1.2.1 we elaborate on the most closely related literature on the detection of archeological objects.

1.2 Review of the related literature

Here we briefly review the literature, both that related to the detection of various archaeological sites (Sec. 1.2.1) and that related to the detection of rectangular structures for various purposes (Sec. 1.2.2).

1.2.1 Automated detection of structures in archaeological studies

The image processing techniques developed for archaeological applications have been mostly focused on contrast enhancement and the automated mapping of known sites. For example, in [20], the authors experimented with morphological image processing techniques [21, 22] in order to segment and extract objects of cultural heritage within known archaeological sites. In [23, 24], the extraction and mapping of linear archaeological traces (such as ancient drainage canals, roads, and property divisions visible as crop marks) was addressed. The authors used active contour models [25] initialized by an operator who manually determined the initial clearly visible parts of the relevant archaeological traces (e.g. by giving either seed points, or ellipses circumscribing the traces as in [23] or drawing straight lines in the vicinity of the traces as in [24]). These tasks differ essentially from the task of the detection of new sites in unknown areas, which is the topic of this thesis, with its focus on rectilinear structures such as LSEs.

The widely held belief that the great variation of the archaeological record prevents the use of automated detection methods has led to a limited number of investigations in this field [26]. There have been many case studies that address the identification of probable archeological sites based on their spectral properties. Traditionally, the spectral characteristics of sites of archaeological interest have been used to analyze remote sensing data [27]. Various vegetation indices have been developed in order to identify particular types of sites of archaeological interest (e.g. buried remains) based on crop marks. For example [28] reviews and evaluates a large set of such indices using a case study of the detection of Neolithic tells in the Thessalian plain (Greece). Unfortunately, the spectral properties are neither unique to the sites of our interest nor consistent from site to site. In addition, the small size of the objects (the walls of the structures are usually not more than two pixels wide for images at 0.5 m resolution) does not allow a reliable determination of their spectral properties.

In contrast to spectral properties, the geometrical or shape properties of the LSEs are more useful for their identification. Such properties appear to be more distinctive and do not depend on image modality and conditions under which an image was captured.

To our knowledge, there have been only a few case studies that are to some extent similar in nature to our study (in terms of the data used, the goals to be achieved, etc.)

and that aim at detecting of archaeological sites relying on their geometrical properties as the main features. For example, [29] focused on the detection of high circular tombs in southern Arabia using high resolution Quickbird satellite images. In [18], the detection of circular soil or crop marks in the panchromatic band of high-resolution Quickbird satellite images was addressed. Such circular marks are of archaeological interest because they may be caused by burial mounds. Airborne laser scanning data (ALS) (also called LiDAR) of high resolution, which is an alternative to optical data obtained from passive sensors, has been used in several archaeological studies. For example, in [30], ALS-based digital terrain models (DTMs) were used for the detection of the circular kiln remains that typically result from charcoal production. In addition to the DTMs, the slope and the topographic position index were embedded into the detection flow. Using ALS data, an approach similar to that of [18] was developed in [31] for the detection of circular grave mounds. One of the conclusions of that paper was that the method is useful for the detection of unknown fields of grave structures, but single structures are hard to detect. This would prohibit its use in our case study. To detect candidate objects, the aforementioned methods used template matching or similar techniques that rely on a predefined set of templates or filters. In some of the papers, the detection of candidate objects was followed by feature extraction and classification, which reduced the number of falsely detected objects. In our case, template matching techniques cannot be adopted because of the much higher variability of the appearances of the objects, which cannot be well represented by a reasonably sized set of predefined templates.

In [32], the authors aimed at the detection of spots that might be caused by tell mounds. In contrast to the previously mentioned papers, and to our project, the data used was of an essentially different type, with a coarser resolution that was still sufficient for the detection of the tell mounds. Namely, the authors used the digital elevation model of the Space Shuttle Radar Topography Mission (SRTM). Instead of using predefined templates, the authors proposed to learn such templates from the data, and called them eigenspots or fisherspot images, in analogy to [33]. The very small number of examples of LSEs (see Sec. 3 for details) prevents us from using approaches similar to [32]. Moreover, many other machine learning approaches (e.g. those employed in [18, 31, 29]) are also not appropriate for our task because they require a dataset of examples to learn from that is much larger than that available to us.

It should be noted that most archaeological studies of the kind mentioned are limited to

particular archaeological areas without attempts at wider application. Our project, on the other hand, aims at the development of methods that will be robust enough to be applied to new wide areas that may differ essentially from the primary area of the case study. One exception we are aware of is published in [34], where the authors aimed at the automated mapping of previously unknown settlement mounds (also called tells) over 23000 square km in northeastern Syria using several sources of satellite imagery (CORONA, ASTER) and digital elevation models (SRTM) with spatial resolutions ranging from 15 to 90 meters per pixel.

In this short review of the archaeological literature, we have focused on closely related research in terms of the data used and the goals of the project. For a more general discussion of the use of remotely sensed imagery for archaeological fieldwork, the interested reader is referred to the comprehensive overview in the handbook [35], or to the more recent short overview of the literature [36].

1.2.2 Detection of rectangular structures

There is a large body of work on the detection of rectangular structures in different contexts without relation to archaeology. Examples are the detection of buildings in remote sensing images [37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47], traffic signs [48, 49, 50], and particles of a rectangular shape in cryo-electron microscopy images [51, 52]. The methods used have been based on Markov Random Fields [42, 48], Marked Point Processes [43, 47], search on a graph [39, 4], the Hough Transform and other voting schemes [40, 41, 51, 49, 50], template matching [53], aggregation of local features [43, 46, 44], and heuristic rules [38].

The majority of the publications for the detection of rectangular structures has dealt with buildings in remote sensing images. These methods are also the most relevant to our project. For example, in the graph-based approach in [39], a search for cycles was used to generate building hypotheses. The search was accompanied by an extensive set of rules and thresholds, which limited the robustness of the approach. Markov Random Fields (MRF) were used in [42] to delineate buildings. More recently, a similar approach was used in [48] for the detection of traffic signs in color images. That approach is sensitive to any inaccuracy in the extracted edges and cannot detect incomplete rectangles, as it requires the presence of all four sides of a rectangular structure. Marked point processes (MPP) [54]

have recently become popular for the extraction of various structures in remote sensing images, including buildings (e.g. in [43, 47]). MPPs have proved to be very powerful when applied to real data. However, these stochastic methods are still computationally expensive. As with the MRF, they may not converge to a globally optimal solution, and usually need the careful tuning of a large number of parameters. Attempts have recently been made to address some of these problems, which are crucial for the analysis of large images. In [55], substantial improvements in performance were achieved for the extraction of line networks (roads and rivers). The potential of GPUs was efficiently exploited as well.

An approach for the detection of rectangular contours based on the Hough transform was developed in [41]. The approach relies on certain strict geometrical rules, making it unsuitable for the detection of fragmented or incomplete structures. It may also result in the detection of rectilinear configurations that cannot form a rectangular contour. The detection of such configurations is prevented in our approach by adding a convexity constraint (Sec. 8.1).

In [44], a set of local features that carried local corner information was used to produce a probability map of building rooftops. Unfortunately, in the case of fragmented enclosures, corners are not reliable features. Moreover, local features in general do not suffice in the case of faint contours appearing in a cluttered background. A more global description that takes into account the spatial relations between local features is necessary. For example, in [43, 46], the gradient orientation density function (GODF) was computed from image gradients. A correlation of this function with a mixture of two Gaussians having mean values separated by ninety degrees served as a GODF-based feature indicating the presence of buildings. In Sec. 10.4, we quantitatively compare the rectangularity feature introduced in this thesis with the GODF-based feature applied to the task of detecting LSEs. We have found that this feature is not effective when computed over relatively large windows, where the relative number of points belonging to an enclosure is small.

Although there is a variety of methods developed for building detection, they are not applicable to our task because buildings are much more salient structures. In contrast to building rooftops, the walls of ruined livestock enclosures are narrow and are of low height (low contrast features), may be highly fragmented, or even completely missing. Higher contrast irrelevant structures may appear inside or outside the rectangular structures in the immediate neighborhood. Many of the cues (rooftop color, shadows, 3D cues, etc.)

usually employed in building detection algorithms are not available in our case.

In [56] there was introduced a general method for the detection of man-made structures in satellite images that cover broad areas. The method is applicable to diverse categories of man-made structure. It was not focused on buildings, but was demonstrated to be suitable for the detection of architectural structures. It is based on saliency-gist 238-dimensional feature vectors and support vector machine (SVM) [57, 58] classification. The saliency features [59] are biologically inspired low-level features that locally capture the contrasts in color or intensity that attract human visual attention (and hence are called saliency features). On the other hand, the complementary gist features [60] globally summarize the contextual information of the entire scene. The method based on saliency-gist showed surprisingly good performance in a series of experiments using a relatively small number of examples (50–300 for each of the positive and negative categories) for training.

Our objects of interest commonly appear with lower contrast than the contrast of irrelevant nearby structures, which can make the aforementioned saliency features inefficient. We, therefore, did not investigate the power of saliency-gist features for our task, but instead compared the rectangularity-size features introduced in this thesis with the currently state-of-the-art deep convolutional neural network (CNN) features. It should be noted that training classifiers using feature vectors of a size comparable to or larger than the size of the training set (as was the case in [56]) may overfit the training data (even in the case of well-regularized classifiers, such as SVM). This danger was of special concern in this thesis due to the very small number of available positive examples[1]. To avoid the danger of overfitting, we developed a simple classification technique that can safely use a small number of positives for training, but requires a large number of negatives, which were available in our project.

1.3 Main contributions and thesis organization

Here we point out what are our main contributions and where the corresponding details can be found in this thesis.

In this thesis we introduce a methodology for the semi-automated detection of the

[1]We used nine examples of structures taken from both satellite and areal images. These examples correspond to only five different well preserved enclosures (see Sec. 3 for details).

remains of rectangular structures, which could be of archaeological interest, in wide areas captured by either aerial or satellite imagery. The methodological flow is presented in Ch. 2. We also confirmed the feasibility of this methodology detecting unknown archaeological sites in two mountainous regions, see Ch. 12.1 and Ch. 12.2.

Large areas in remote sensing images that are characterized by high contrast texture (such as urban areas, forests, or rocky mountains) do not present any interest for our project. Therefore, they can be excluded from the regions to be analyzed, which speeds up the whole image processing flow and may significantly reduce the number of false detections. In Ch. 4, we introduce a new approach for the automated detection of such texture regions, which are to be segmented out. We show that in contrast to other approaches, our method does not segment out individual objects and accurately segments the regions in the vicinity of texture borders.

Partially based on a few previous ideas, in Ch. 6 we design a new technique for the generation of candidate locations that are likely to be enclosed, at least partially, by structures of an arbitrary shape.

Novel rectangularity-size features (Ch. 8) that are computed at carefully chosen candidate locations is a key ingredient of our methodology. These features are capable of capturing enclosures of an approximately rectangular shape, even if these are fragmented or incomplete. On the other hand, these features do not respond to various spurious structures such as line intersections, individual corners, etc.

The rectangularity-size features were fed to a classifier that rejects the majority of falsely detected candidate locations. Commonly used classifiers cannot be used here because of the very small number of LSE examples available to us. We, therefore, develop an original technique (Ch. 9) for constructing a simple linear classifier for the detection of rare events (LSEs in our case). The classifier can be trained on a small number of positives, without the danger of overfitting the data. Yet it requires a large number of negative examples, which are available in our case.

The computation of the rectangularity-size features relies on previously extracted linear segments, which could be parts of ruined walls. They usually appear as faint image features of low contrast that are hard to extract. In Ch. 5, we therefore introduce a new method for the extraction of isolated features in images. The distinctive property of this method is that it can distinguish isolated features from elements of texture even of

the same shape with higher contrast. We adapt this method to the extraction of linear features in complex backgrounds and compare it to alternative approaches.

Comparing the discrimination power of the developed rectangularity-size features and the detection performance in our task is not straightforward, because of a very small number of available LSEs. Therefore, in Ch. 10 we develop an evaluation strategy and some particular measures of discrimination ability suitable to such a case, which allow a comparison to other types of features. Among other features we evaluate the deep CNN features that have recently shown remarkable performance in various classification problems. We show that the handcrafted rectangularity-size features outperform the deep CNN features in our task. On the other hand, a few CNN architectures perform surprisingly well. In Sec. 14.1 we briefly discuss promising directions of future research in the further improvement of LSE detection.

Chapter 2

Overview of the proposed methodology

In contrast to spectral properties, the geometrical properties of LSEs appear to be more distinctive and do not depend on the image modality or conditions under which an image was captured. We therefore developed a measure that quantifies the distinctive geometry of approximately rectangular enclosures. Our approach relies on new rectangularity-size features that discriminate rectangular patterns from other structures in a complex cluttered background. The rectangularity feature is based on a prior model of a fragmented rectangle, which is a convex polygon with constrained angles.

Based on the taxonomy of methods for object detection in optical remote sensing images that was recently suggested in [61], our approach partially fits into the two categories, knowledge based and machine learning based methods. We use prior geometrical knowledge to generate candidate proposals and to compute the suitable rectangularity-size features. Based on these features, we apply machine learning technique to reject the majority of false candidates and validate the proper candidates.

In this section we give an overview of the sequential steps of our methodology. We briefly describe each step and its purpose and refer the reader to the chapter where it is fully elaborated.

The methodology developed comprises the following main steps:

A. Filtering out high contrast texture areas, Ch. 4

B. Extraction of short linear features, Ch. 5

C. Detection of candidate locations, Ch. 6

D. Extraction and modeling of linear segments, Ch. 7

E. Computation of rectangularity-size features, Ch. 8

F. Classification and detection, Ch. 9

G. Visual validation of detections using a specially designed GUI, Ch. 11

Given an image, this flow results in a map indicating the most probable locations with their likelihoods of being a structure of interest. An example is given in Fig. 9.2. Such a map is very sparse having zero likelihoods in most regions. The outline of each step from the list above follows next.

A. Filtering out high contrast texture areas. The objects of interest are usually located in smooth grassland areas. Thus, filtering out high contrast textured areas, e.g urban areas, where a large number of false detections may be generated, will significantly reduce the rate of false detections almost without affecting the sensitivity to true examples. Such a preprocessing will also reduce the computational burden since we developed a filtering technique that is much faster than the following algorithms for the detection of geometrical structures.

For this purpose, we developed a morphological texture contrast (MTC) descriptor [5, 3, 1] based on alternating morphological filters [62, 21] that allows filtering out urban areas, forests, rocky mountains, and other high contrast texture regions, while preserving individual or isolated structures. The important property of this descriptor, hardly achieved by other techniques, is that it has low values at isolated features, even if they have high contrast. Another advantage of the MTC is that it allows accurate filtering in the vicinity of texture borders.

B. Extraction of short linear features.

We extract narrow linear bar edge features that are short enough to match highly fragmented walls of the structures of interest. We also call these features ridge (bright) or valley (dark) features. For this purpose, we developed a morphological feature contrast (MFC) operator [5, 3] for the detection of linear structures that also significantly

suppresses the surrounding texture and noise. It is based on morphological filters and has a similar underlying structure as the MTC operator. This operator is capable of extracting isolated structures that stand out from a background texture, which is soil or stone texture in the images we analyze.

C. Detection of candidate locations. We designed a fast method based on the ideas previously appearing in [63, 64, 65, 66] dealing with the detection of shape centered points of interest. Namely, we suggest detecting candidate points by finding the local maxima of the average flux of the gradient field of the Euclidean distance function of the binary map of bar edges [7]. These maximum points usually correspond to junction points of the medial axis of an inverted binary map of the edges. Only local maxima with an average flux greater than a particular value are taken into account.

D. Extraction and modeling of linear segments. A Hough transform [67] computed locally at the candidate points is used to extract linear segments, which are aligned ridge or valley features. Simultaneously, linear segments are modeled with a few parameters.

E. Rectangularity and size features. An undirected graph is constructed, the nodes of which correspond to the linear segments. The graph's edges encode the spatial relations between the linear segments. In particular, we use angle and convexity properties to encode the spatial relations. Due to the construction of the graph, its maximal cliques correspond to geometrically valid configurations of linear segments. The valid configurations are then ranked by a proposed rectangularity measure that encodes the goodness of grouping the segments into a rectangular structure. Configurations that better match a rectangular structure receive a higher rectangularity measure. The proposed rectangularity feature [6, 7] is defined as the maximal rectangularity measure of all valid configurations. The low number of corresponding maximal cliques within the analysis window at a candidate point allows an exact maximization that can be efficiently computed. The resulting rectangularity feature captures the presence of Π-like structures and is robust to their fragmentation and to the deviation of the angles between wall fragments. We also compute an additional feature proportional to the enclosure size, which along with the rectangularity feature, is used as an input to the classifier in the next stage.

F. Classification and detection. We developed a methodology that allows learning

a classier from just a few representatives of enclosure structures and a large number of negative examples [6, 7]. To prevent overfitting we use a linear classifier, which is trained on a (well sampled) distribution of negatives. The positives are treated as deterministic points in the feature space and influence the classifier only via their average. This learning methodology is not limited to low dimensional feature spaces, such as the developed two dimensional rectangularity-size feature space, but directly extends to higher dimensions. We used this methodology in our comparative experiments in order to compare the developed rectangularity-size features with high-dimensional generic features.

The livestock enclosures are detected by thresholding the output of the classifier at the level that ensures that the number false detections per unit area will be below a required limit.

G. Visual inspection of detected structures using a graphical user interface. We built a prototype user interface that allows a user to conveniently explore large images. It shows detections and their confidence (the output likelihood of the classifier for the detected locations), and allows quickly examining and rejecting falsely detected sites. The user interface also allows the adjustment of various parameters, such as the number of false detections to be generated for an analyzed area, the range of structure sizes, and several other parameters of the algorithms. A snapshot of the user interface is given in Fig. 11.1.

Chapter 3

Collected data

Our main source of the data is the high resolution remotely sensed imagery that covers the region of the Silvretta mountains at the border between Austria and Switzerland, Fig. 3.1.

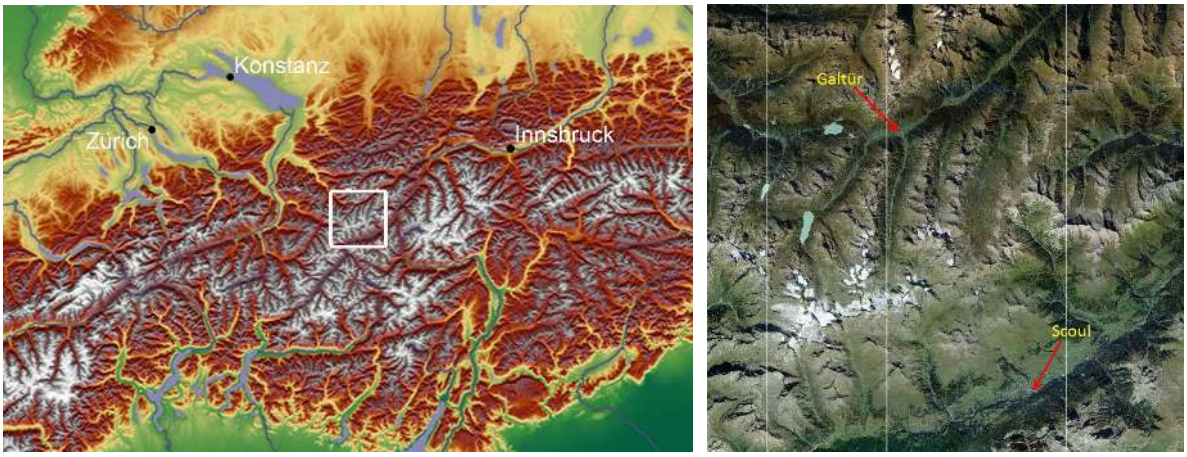


Fig 3.1: Left: The Silvretta Alps (white box) on the Swiss–Austrian border. The parental Silvretta Historica project involved the Universities of Zurich, Konstanz, and Innsbruck, located in the nearby region. Right: Zoomed in region of the Silvretta Alps. The towns of Galtür and Scuol lie on the two different sides of the Swiss–Austrian border.

We used both aerial and satellite images in order to develop a robust approach that is not sensitive to illumination conditions or a particular source of the data. Aerial color ortho-photos were purchased from the Swiss Federal Office of Topography (SWISSTOPO,

Bern). These images were captured in 2005. Satellite imagery is likely to become the preferred data source for regional archaeological research in many areas of the world where alternative data sources, such as aerial images, are not easily available. We, therefore, also purchased satellite imagery through the GAF AG provider. These images were captured in September 2011 by the GeoEye-1 satellite in accordance with our order.

The aerial images have a spatial resolution of 0.5m/pixel for each of the three RGB channels (8 bits per pixel per channel). For satellite images we ordered the bundle product comprising the panchromatic channel and pansharpened color Red, Green, Blue, and Infrared channels. Due to legal regulations, after pansharpening all channels were downsampled to a spatial resolution of 0.5m pixel size (11 bits per pixel). The original resolution for the panchromatic channel was 0.41m/pixel, while for the color channels the resolution was only 1.64m/pixel. At that time of acquisition, 0.5m/pixel resolution was the highest resolution for satellite images available for purchase. Due to narrow structures of our interest (walls of livestock enclosures) coarser resolutions would hardly be sufficient for our purposes.

The satellite and aerial images cover approximately the same area, of about 550 km^2, of the alpine Silvretta mountain range. The satellite imagery was divided into 17 large images, while the aerial imagery comprised 88 smaller images. Dozens of archaeological sites have been discovered in this region over several years of archaeological survey by the Silvretta Historica project (see Sec. 1.1). Among them there are five sites of livestock enclosures (LSE) with well preserved ruins, see Fig. 1.1 and Fig. 3.2. These enclosures were used as representative examples for the class of structures of interest. We localized four of these sites in images taken from both the satellite and the aerial imagery. The fifth site appeared only in the aerial imagery and was outside of the area captured by the satellite imagery. Thus, overall, we had nine image examples of LSEs that came from five different archaeological sites. We did not use other available but heavily ruined LSEs with walls likely to be missed by the segment detection algorithm. Our approach is based on a prior knowledge of the representative canonical shape of structures of our interest. Therefore, including a large proportion of heavily distorted enclosures (relative to all available examples of LSEs) would cause learning inappropriate classifier that would be prone to false detections.

In our research we did not use the color channels because the spectral properties of the sites of interest are not unique and also not consistent from site to site. Moreover, the

Fig 3.2: Aerial images with well preserved LSEs used along with LSEs in Fig. 1.1 for training of the classifier in Ch. 9.

use of color information would reduce the robustness of our algorithms to illumination conditions and increase their sensitivity to a particular choice of the data source. For the case of the satellite imagery, we used the panchromatic channel, which has the highest original resolution, while for the case of the aerial images, we used the red channel.

Though we developed our algorithms to work primarily with remote sensing images of 0.5m/pixel resolution, they can also be applied to images of higher resolution. Some adaptation of the parameters is required in that case. We performed a few additional experiments in Sec. 12.2 with aerial SWISSTOPO images at higher resolution, 0.25 m/pixel. These images were taken above the Bernese Alps in Switzerland and were obtained from our collaborators from the Archaeological Service of Canton Bern. Similarly to the Silvretta Historica project, alpine surveys were carried out in the Bernese Oberland with the goal of increasing the number of sites in the archaeological record [68]. The acquired images cover the region of Interlaken and the region of Oberland East. The 846 images, of size 4000×4000 pixel, were captured in 2008 (Interlaken) and 2007 (Oberland) and together cover about 850 km^2 of the Bernese Alps.

Chapter 4

Filtering texture areas

Objects of our interest usually appear in smooth grassland areas. Therefore, we filter out high contrast texture regions, such as urban areas, forests and rocky areas. Moreover, the majority of false structure detections might be generated in such texture areas, which makes detection and segmentation of these areas a necessary step.

In this chapter we develop a morphological texture contrast (MTC) operator that allows segmentation of texture and non-texture regions in images, irrespectively of illumination and type of texture. We show that in contrast to other approaches, the MTC discriminates between texture details (features that are a part of texture) and individual features in images (e.g. isolated edges of blobs) and provides high accuracy of texture localization. It also does not involve heavy computations. The MTC is developed on the basis of mathematical morphology, which provides a theoretically consistent nonlinear analysis that has proven to be very effective in various applications, including processing of remotely sensed imagery [22]. A comparison with other methods used for texture detection is provided at the end of the chapter.

4.1 Related approaches

In [69] it was proposed to use the difference between maximal and minimal intensities (MaxMin diff.) in a pixel neighborhood for a fast segmentation of an image into textured and non-texture regions. A standard deviation (StD) is also frequently used as

a measure of texture that describes its smoothness [70]. In [71], where the Local Binary Patterns (LBP) approach was developed, the authors also suggested to incorporate a variance based descriptor for texture classification purposes. While the LBP descriptor is related to inherent texture properties, a complementary variance based descriptor measures texture contrast. The amplitude modulation function (AMF), derived from the amplitude-modulation frequency-modulation model [72], can locally capture texture contrast. Although each of the texture contrast descriptors mentioned above can be used to discriminate between texture regions and non-texture areas, later also called smooth areas, they cannot distinguish individual features from texture details.

Several descriptors were suggested to approach this problem. In [73] the difference between closing and opening, called texture range (TER), was suggested to distinguish individual step and ramp edges from texture edges. The TER operator, however, cannot distinguish isolated features, such as ridges and blobs, from texture details of comparable size. Recently, in [74] the PanTex index was developed to detect settlements in panchromatic satellite imagery. The operator is able to distinguish texture areas from individual linear features such as roads or borders between homogenous cultivated fields in satellite images. The PanTex index is defined as a minimal contrast among contrast measures derived from the gray-level co-occurrence matrixes (GLCM) [75], computed for different orientations of displacement vectors. The PanTex method, however, does not distinguish other individual features, such as isolated peaks or small isotropic blobs, from texture. The component count (CC) method [76] is based on the product of two measures computed in small image blocks. The first one is the sum of the number of connected components (component count) in the background and the foreground obtained by simple binarization of image blocks. The second measure is the difference between average intensities in the background and the foreground. This descriptor is supposed to discriminate blocks covering texture and individual step edges at the borders between homogenous regions. A similar idea of counting the number of local extrema (texture primitives) for detection of texture regions was proposed earlier in [77]. Since this method does not take into account contrast of texture primitives, it can be very sensitive to noise. Another disadvantage that all the above texture descriptors, excluding the TER, have in common, is that they extend or blur the borders of texture regions, preventing accurate localization of texture borders. The MTC operator we develop in this chapter does not suffer from the above disadvantages.

4.2 Detection of texture regions: Morphological Texture Contrast (MTC)

We define here the morphological texture contrast (MTC) operator $\psi_{\mathrm{MTC}}(f)$ for distinguishing texture regions in satellite images from smooth areas, which may also contain individual structures that should not to be assigned to texture. The operator has high values in texture regions, but low values at individual features and in smooth regions.

The MTC is based on alternating morphological filters, $\gamma_r\varphi_r$ and $\varphi_r\gamma_r$ [62, 21], which are closing φ followed by opening γ and opening followed by closing, respectively. r denotes the size of the structuring element (SE). Alternating filters are usually employed for noise filtering. We use them to estimate texture envelopes. The difference between upper and lower texture envelopes defines a measure of texture contrast, which can serve as an indicator of the presence of texture

$$\psi_{\mathrm{MTC}}(f) = |\gamma_r\varphi_r(f) - \varphi_r\gamma_r(f)|^+ \ , \tag{4.1}$$

where the argument f denotes a 1D signal or a 2D gray-scale image, and $|\cdot|^+$ is defined as

$$|\nu|^+ \triangleq \begin{cases} \nu, & \nu > 0 \\ 0, & \text{otherwise} \ . \end{cases} \tag{4.2}$$

A remarkable property of these envelopes is that they coincide at individual features, thereby yielding low response at individual features even if they are of high contrast (see an example in Fig. 4.1). Since in the 2D case, $\varphi_r\gamma_r$ and $\gamma_r\varphi_r$ are not ordered [62, 21], a lower envelope $\varphi_r\gamma_r$ might be above an upper envelope $\gamma_r\varphi_r$. However, Proposition 3 below shows that regions where this happens are small in the sense that an erosion with a structuring element of size r completely removes these regions. In the following discussion we will show that r defines the minimal size of texture regions to be detected (see Eq. (5.4)). Therefore, the regions where $\gamma_r\varphi_r - \varphi_r\gamma_r < 0$ are small enough to be considered as non-texture regions. They are correspondingly removed by the $|.|^+$ operator in the definition of ψ_{MTC} above.

Let us denote morphological erosion of a set or a function by ε. Large letters X, Y, A will denote sets. Structuring elements are identical for all morphological operators in the following propositions.

Proposition 1. *The following inequality holds:* $\varepsilon\gamma\varphi \geq \varepsilon\varphi\gamma$.

Proof. We have $\varepsilon\varphi \geq \varepsilon\varphi\gamma$ due to the increasing property of closing and erosion, and antiextensivity of opening. Proposition 1 follows directly from the last inequality and due to $\varepsilon\gamma = \varepsilon$.

Proposition 2. *Given the ordering condition* $g_1(x) < g_2(x), x \in A$ *the following inequality holds:* $[\varepsilon(g_1)](y) < [\varepsilon(g_2)](y), y \in \varepsilon(A)$.

Proof. Let us denote by B_y a structuring element shifted to position y. For $y \in \varepsilon(A)$ we have $[\varepsilon(g_1)](y) = \min_{x \in B_y \subseteq A} g_1(x) < \min_{x \in B_y \subseteq A} g_2(x) = [\varepsilon(g_2)](y)$, where the inequality follows from the given ordering condition.

Proposition 3. *Given the set* $X = \{x : \gamma\varphi < \varphi\gamma\}$, *the set* $Y = \{y : y \in \varepsilon(X)\}$ *is an empty set.*

Proof. From the construction of the sets X,Y and from Proposition 2 with $g_1(x) = [\gamma\varphi(f)](x)$ and $g_2(x) = [\varphi\gamma(f)](x)$, for $x \in X$, it follows that $[\varepsilon\gamma\varphi(f)](y) < [\varepsilon\varphi\gamma(f)](y)$, for $y \in Y$. Since the last inequality contradicts Proposition 1 we conclude that Y is empty.

The results of applying the MTC to an artificial 1D signal and to a remotely sensed image[1] of a forested area are shown in Fig. 4.1(right) and Figs. 4.2(b, c), respectively. Note that individual trees in Fig. 4.2, and individual peaks as well as step edges (front and back of the wide pulse) in Fig. 4.1 were suppressed. Throughout this paper we use square SEs, where the size refers to its side length. The size r of the SE of ψ_{MTC} should be chosen to be larger than the maximal distance between details in textured regions. Features that stand apart from texture details farther than r are treated as individual features and are suppressed correspondingly. In general, we can use different sizes $r_1 \neq r_2$ for SEs of ψ_{MTC},

$$\psi_{\mathrm{MTC}}(f) = |\gamma_{r_2}\varphi_{r_1}(f) - \varphi_{r_2}\gamma_{r_1}(f)|^+. \tag{4.3}$$

The size of the SEs r_1 and r_2 should be chosen such that

$$D_1 < r_1 < D_2,\ r_2 < S_2, \tag{4.4}$$

[1]The satellite image was logarithmically transformed before applying the MTC operator, see Sec. 4.4 for the rationale behind.

where D_1 is the maximal distance between neighboring texture details, D_2 is the minimal distance to isolated features, and S_2 is the minimal size of texture regions. Comparing Figs. 4.2(b) and (c) illustrates how r_2 controls the minimal size of texture regions to be detected. In addition, for the case $r_2 < r_1$, isolated features of size S_1 are not suppressed if $r_2 < S_1 < r_1$. Therefore, choices of r_1 and r_2 with $r_2 < S_1 < r_1$ should be avoided.

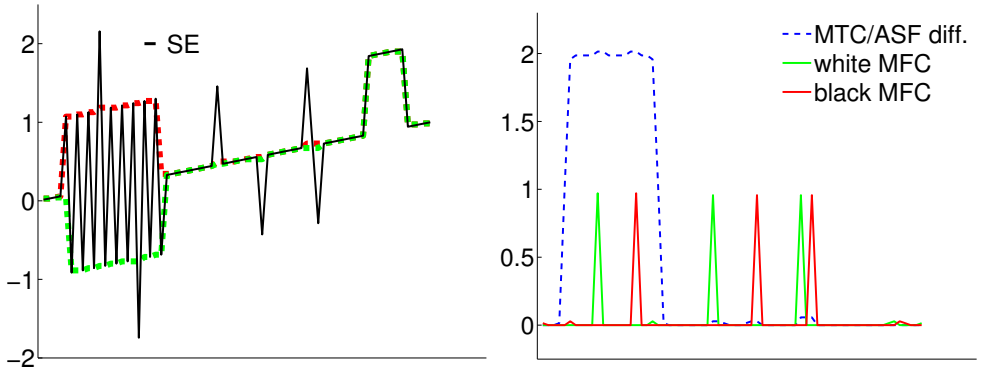
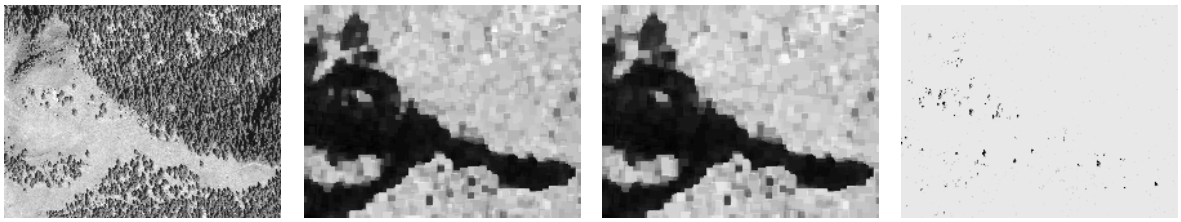


Fig 4.1: Left: An artificial signal composed of a slowly varying component, a texture region, and individual features. Upper and lower envelopes of the texture obtained with alternating morphological filters are shown by red and green dashed lines. Right: Extraction of the texture region and individual features with the MTC and ASF diff. (Sec. 4.2) and the MFC (Sec. 5.1) operators. Note that MTC and ASF diff. yield identical responses for 1D signals (Proposition 4).



(a) (b) (c) (d)

Fig 4.2: (a) Pan-chromatic satellite image of 1500x1150 pixel size (© GeoEye 2011, distributed by e-GEOS). (b) and (c) The MTC descriptor. $r_1 = 30, r_2 = 25$ in (b) and $r_1 = 30$, $r_2 = 35$ in (c). (d) Extraction of individual dark features, i.e. individual trees, using the ψ^-_{MFC} operator with $r_1 = r_2 = 30$ (Sec. 5.1). Note that the trees in forest areas are almost completely suppressed.

Alternatively, we introduce an operator defined as the difference between alternating sequential filters $\varphi\gamma\varphi$ and $\gamma\varphi\gamma$ with identical structuring elements

$$\psi_{\mathrm{ASF}}(f) = \varphi\gamma\varphi(f) - \gamma\varphi\gamma(f) \ . \tag{4.5}$$

Alternating sequential filters $\varphi\gamma\varphi$ and $\gamma\varphi\gamma$ are ordered [62, 21], which ensures that ASF difference (ASF diff.) always yields non negative values, allowing to drop the $|.|^+$ operator used in the MTC. In the next proposition we outline two properties of the ASF diff. that relate it to the MTC.

Proposition 4. *The MTC operator defined in Eq. (4.1)*

1. *has lower response than the ASF diff. with the same SE, i.e.* $\psi_{\text{ASF}}(f) \geq \psi_{\text{MTC}}(f)$.
2. *is identical to the ASF diff. with the same SE in the 1D case.*

The first property holds in general and is directly followed from extensivity of closing and anti-extensivity of opening operators. The second property above follows from the fact that in the 1D case we have $\varphi\gamma\varphi = \gamma\varphi$ and $\gamma\varphi\gamma = \varphi\gamma$ [62]. The ψ_{ASF} operator applied to the artificial 1D signal is shown in the blue dashed line Fig. 4.1. As we expect it has the same response as the ψ_{MFC} operator. In the 2D case the ASF diff. and the MTC operators are not equivalent. However, we will experimentally show in Sec. 4.3 that they perform almost identically in their ability to distinguish texture from individual features. On the other hand, since the MTC operator is much faster than the ASF diff. the MTC operator is preferred in practice.

An important property of the ASF diff. and the MTC operators is that they neither extend nor blur the borders of textured regions, thereby allowing accurate localization of texture borders. This property is illustrated in Fig. 4.3 in the rightmost column. Below we outline three other properties of the the ASF diff. and the MTC that are desirable for texture detection.

Proposition 5. *The MTC operator in Eq. (4.3) and the ASF diff. (denoted in this proposition* ψ *without subscript) are*

1. *bias invariant,* $\psi(f) = \psi(f + a)$,
2. *invariant to signal inversion*[2]*,* $\psi(f) = \psi(a - f)$,
3. *proportional to signal magnitude* $\psi(af) = |a|\psi(f)$,

where $a \in \mathbb{R}$ *is a constant.*

[2]In Mathematical Morphology signal inversion is also referred to as self-complementarity [21, 78].

These properties follow from the general property of morphological operators with flat structuring elements to commute with increasing continuous functions[3] and from the duality of opening and closing. Namely we will need the following properties of morphological operators to proof Proposition 5. Let us denote by ζ morphological opening or closing operators and by ζ^* their dual (opening for closing and closing for opening). Then ζ^* satisfies the following equalities

$$\zeta(f+a) = a + \zeta(f) \tag{4.6}$$

$$\zeta(a-f) = a - \zeta^*(f) \tag{4.7}$$

$$\zeta(af) = a\zeta(f), a \geq 0 \text{ and } \zeta(af) = a\zeta^*(f), a < 0 \tag{4.8}$$

Below we prove Proposition 5 for the MTC operator $\psi_{MTC}(f)$ defined with two size parameters r_1, r_2 in Eq. (4.3). The proofs for the ASF diff. operator are similar in nature.

Proof.

1. Bias invariance follows from Eq. (4.6)

$$\begin{aligned}
\psi_{\text{MTC}}(f+a) &= |\gamma_{r_2}\varphi_{r_1}(f+a) - \varphi_{r_2}\gamma_{r_1}(f+a)|^+ \\
&= |\gamma_{r_2}(a+\varphi_{r_1}(f)) - \varphi_{r_2}(a+\gamma_{r_1}(f))|^+ \\
&= |a+\gamma_{r_2}\varphi_{r_1}(f) - a - \varphi_{r_2}\gamma_{r_1}(f)|^+ \\
&= |\gamma_{r_2}\varphi_{r_1}(f) - \varphi_{r_2}\gamma_{r_1}(f)|^+ = \psi_{\text{MTC}}(f) \ .
\end{aligned}$$

2. Invariance to signal inversion follows from Eq. (4.7)

$$\begin{aligned}
\psi_{\text{MTC}}(a-f) &= |\gamma_{r_2}\varphi_{r_1}(a-f) - \varphi_{r_2}\gamma_{r_1}(a-f)|^+ \\
&= |\gamma_{r_2}(a-\gamma_{r_1}(f)) - \varphi_{r_2}(a-\varphi_{r_1}(f))|^+ \\
&= |a-\varphi_{r_2}\gamma_{r_1}(f) - (a-\gamma_{r_2}\varphi_{r_1}(f))|^+ \\
&= |\gamma_{r_2}\varphi_{r_1}(f) - \varphi_{r_2}\gamma_{r_1}(f)|^+ = \psi_{\text{MTC}}(f) \ .
\end{aligned}$$

[3] In Mathematical Morphology increasing and continuous functions are usually termed anamorphoses [79].

3. Proportionality to signal magnitude follows from Eq. (4.8).
   For $a \geq 0$ we have

$$\begin{aligned}
\psi_{\mathrm{MTC}}(af) &= |\gamma_{r_2}\varphi_{r_1}(af) - \varphi_{r_2}\gamma_{r_1}(af)|^+ \\
&= |a\gamma_{r_2}\varphi_{r_1}(f) - a\varphi_{r_2}\gamma_{r_1}(f)|^+ \\
&= |a(\gamma_{r_2}\varphi_{r_1}(f) - \varphi_{r_2}\gamma_{r_1}(f))|^+ \\
&= a|\gamma_{r_2}\varphi_{r_1}(f) - \varphi_{r_2}\gamma_{r_1}(f)|^+ \\
&= a\psi_{\mathrm{MTC}}(f) = |a|\psi_{\mathrm{MTC}}(f) \ .
\end{aligned}$$

   For $a < 0$ we have

$$\begin{aligned}
\psi_{\mathrm{MTC}}(af) &= |\gamma_{r_2}\varphi_{r_1}(af) - \varphi_{r_2}\gamma_{r_1}(af)|^+ \\
&= |a\varphi_{r_2}\gamma_{r_1}(f) - a\gamma_{r_2}\varphi_{r_1}(f)|^+ \\
&= |a(\varphi_{r_2}\gamma_{r_1}(f) - \gamma_{r_2}\varphi_{r_1}(f))|^+ \\
&= |-a(\gamma_{r_2}\varphi_{r_1}(f) - \varphi_{r_2}\gamma_{r_1}(f))|^+ \\
&= ||a|(\gamma_{r_2}\varphi_{r_1}(f) - \varphi_{r_2}\gamma_{r_1}(f))|^+ \\
&= |a||\gamma_{r_2}\varphi_{r_1}(f) - \varphi_{r_2}\gamma_{r_1}(f)|^+ \\
&= |a|\psi_{\mathrm{MTC}}(f) \ .
\end{aligned}$$

Although, the MTC was developed to discriminate texture and non-texture regions, its multi-scale extension can also be used for classification of different types of texture. The MTC computed for varying sizes of the SE generates a set of features that can be used for this purpose. Since accurate localization is an inherent property of the MTC, we expect that such MTC based classification will be more accurate at texture borders in comparison to other texture classification approaches that involve the computation of summary statistics of dedicated features within a window, such as standard GLCM [75] or a more recent approach based on linear contact distribution [80]. In addition, for the MTC based classification no preprocessing is required to mask out isolated structures (which are not part of any texture class) that may disturb classification results. The extension of the MTC approach to texture classification, as opposed to texture detection considered here, is, however, beyond the scope of this thesis.

4.3 Comparison of texture contrast descriptors

In this section we compare the performance of the MTC operator with the ASF diff., the TER, the CC, the MaxMin difference, the StD, the LBP contrast, the AMF, and the PanTex algorithms. We denote by w the scale parameter required for all algorithms. For the MTC operator it equals the size of the structuring elements $r_1 = r_2$. To allow a consistent comparison, a few algorithms were slightly modified as follows.

In the CC algorithm we avoided several parameters suggested by the authors since they need to be adjusted for each type of image. Specifically, we used the simple product of contrast and number of connected components. Instead of disjoint blocks, a sliding window of the single size w was used to compute the texture measure at each pixel as in the other compared methods. In the PanTex algorithm we used a square root of contrast measure derived from the GLCM matrix. This contrast measure can actually be computed without an explicit calculation of the GLCM matrix. The original PanTex index was designed with a single window of 9 pixels size, which was adjusted to 5m satellite image resolution. Ten displacement vectors with the length varying from $w/9$ to $w\sqrt{5}/9$ were chosen in order to cover the full range of possible orientations. In our comparison we computed the GLCM contrast measure within a window of an arbitrary size w, such that displacement vectors of an approximate length $\frac{w}{7}$ were determined by all pixels on a discrete circle of radius $\frac{w}{7}$. Taking shorter displacement vectors did not significantly change the performance of PanTex, while taking longer vectors reduced the performance estimated in Sec. 4.3.2.

To compute the LBP local variance and the AMF measures, we used a Matlab code available online[4]. In the LBP we used the square root of local variance computed as a variance of $4(w-1)$ equally spaced point samples on a circle of radius Round$(\frac{w}{2})$. This, for example, gives 2, 4, 18 orientations (or the doubled number of directions), and radii 1, 2, 5 for w equal to 2, 3, 10, respectively. The AMF approach does not contain an analysis window, but contains a free parameter, which is the largest period of a sinusoid in the Gabor filters used in the AMF. We set this largest period of a sinusoid to $3w$ pixels. In the quantitative comparison in Sec. 4.3.2 w varies from 10 to 70. The corresponding largest periods of the AMF approach included the recommended value given in the AMF

[4]Matlab sources are available at
http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab
and http://cvsp.cs.ntua.gr/software/texture/

code mentioned above. We also noted that choosing larger values only decreased the performance.

Note that similar to the MTC operator, all the algorithms, after the small modifications described above, except for AMF, fulfill the properties of Proposition 5.

4.3.1 Qualitative comparison

The texture contrast descriptors obtained using the compared transformations are shown in Fig. 4.3. The first two original images are of a size of 512x512 pixels; the third image is a pan-chromatic satellite image of 1300x1100 pixels size[5]; the fourth image is an enlarged part of the third image. $w = 10$ for the first two images and $w = 30$ for the satellite images. For the case of satellite images, we applied all compared operators to logarithmically transformed images. Such preprocessing improves the robustness of the texture contrast descriptors to illumination variations Sec. 4.4, and provides visual results of higher contrast. The performance of the MTC operator and the ASF diff. is undistinguishable. All the descriptors have high values in textured areas and low values in smooth areas. However, contrary to the MTC operator and the ASF diff., the other approaches yield also high responses at isolated features that do not belong to texture. The TER operator is able to suppress step and ramp edges, but yields a high response at isolated ridges and blobs. The PanTex descriptor partially succeeds to suppress isolated curvilinear structures.

To better visualize the accuracy in texture localization, the texture descriptors were superimposed on the enlarged part of the satellite image in the fourth column of Fig. 4.3, where the contrast of red tones is proportional to the values of the descriptors. Since the distribution of descriptor values is strongly bimodal, one can distinguish two major levels of texture descriptors, low and high, that appear as a gray-reddish and saturated red overlaid on the original image. As can be seen from these images, another advantage of the MTC operator and the ASF diff. is that they do not extend the borders of texture regions as other methods do, except the TER. Our implementation of the CC method generates a halo near texture borders and around individual features. This effect does not occur in the original version of the CC method, in which disjoint/overlapping block processing was performed that, however, would not allow accurate texture localization.

[5]© GeoEye 2011, distributed by e-GEOS.

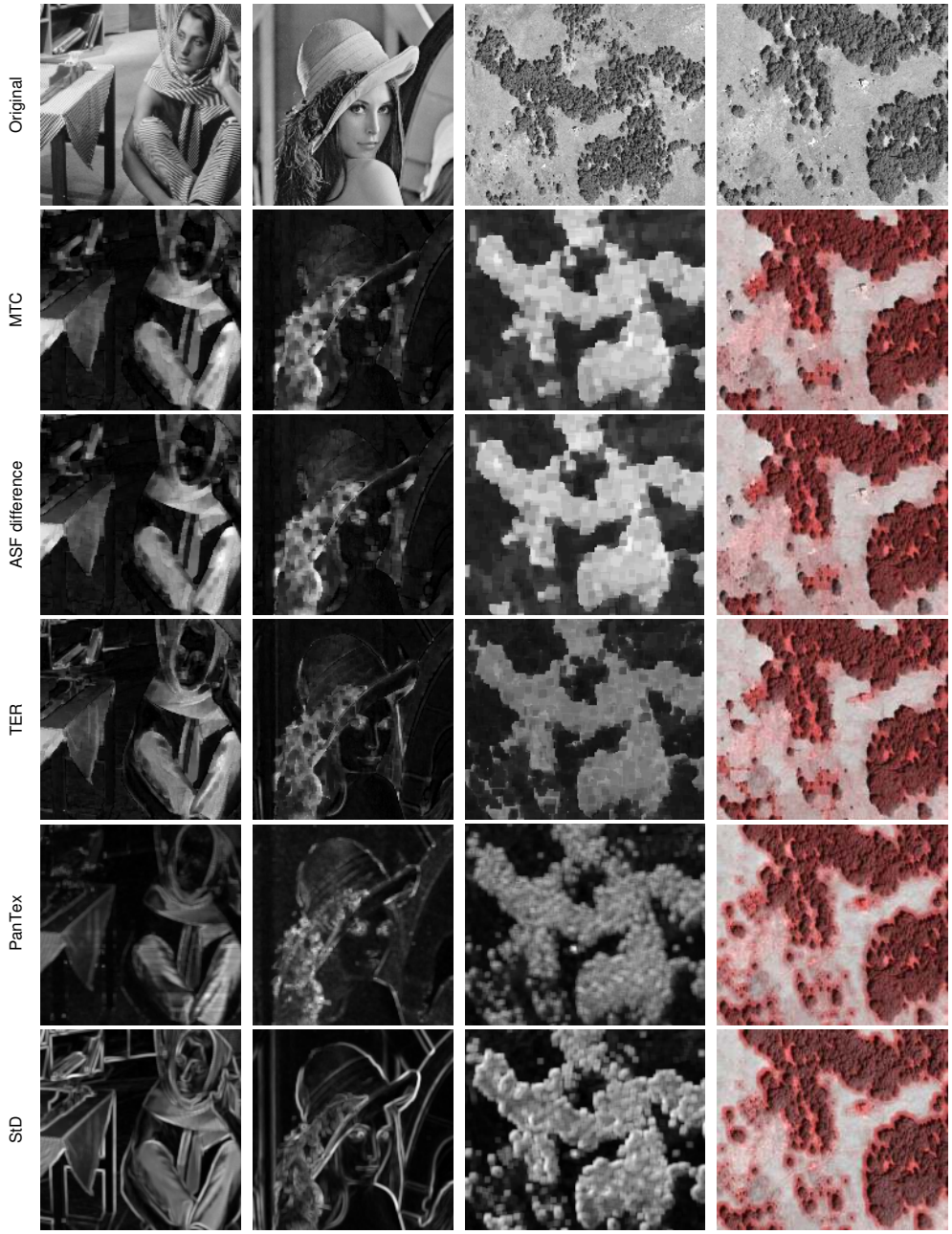


Fig 4.3: Comparison of texture contrast descriptors. The first two images in the first row are the standard test images; the third image is a satellite image with a forested area; the forth image is a zoomed in part of the third image. It was zoomed in for better visualization. All operators generate high values at texture regions and low values at smooth regions. However, the MTC and the ASF diff. yield much lower values at individual features, such as object edges in the standard test figures (first two columns) or blobs corresponding to isolated trees in the satellite image (last two columns). Note also that in contrast to other operators, the MTC and the ASF diff. do not yield high responses in the smooth areas in the vicinity of texture borders, which allows accurate localization of texture areas.

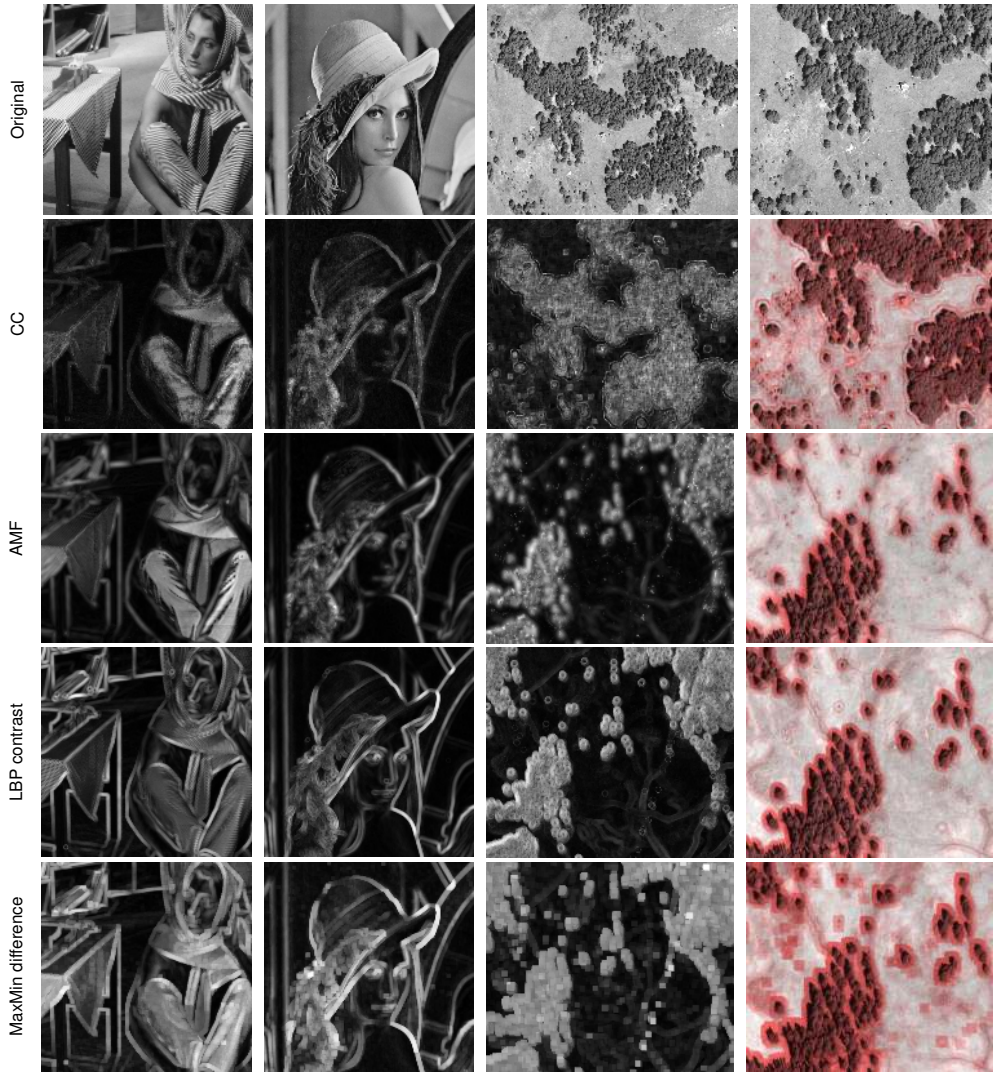


Fig 4.3: Comparison of texture contrast descriptors: figure continuation.

4.3.2 Quantitative comparison

In order to quantitatively compare the ability of the methods to distinguish between texture and non-texture areas we quantify the separability between distributions of texture descriptors in these areas. We used the Fisher criterion [81] that measures the distance between distribution means relative to their compactness. The criterion is given by $\frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}$, where μ denotes the class mean and σ^2 denotes the class variance. Since ground truth data is required to define textured and non-texture regions, we created an artificial data set

of gray-scale images along with corresponding masks of texture regions and non-texture areas, whereby the last also include individual features.

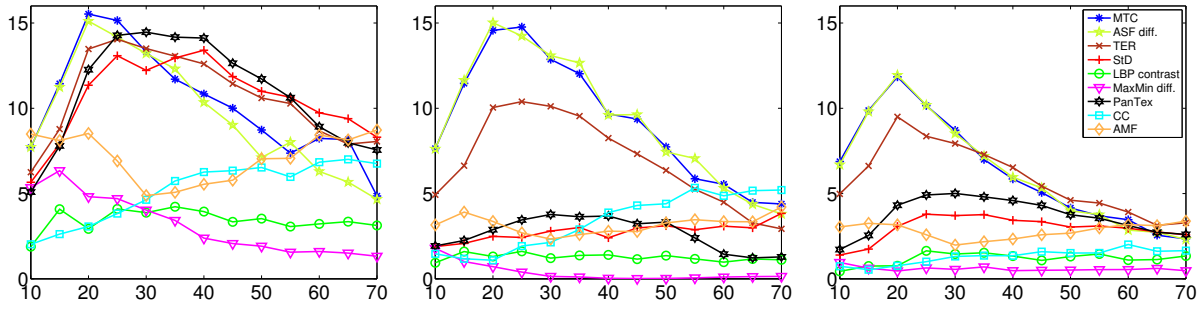
The data set consists of 100 images of 300x300 pixels with circular texture clusters and individual features (clusters may overlap; see the upper-left image in Fig. 5.1). The number of clusters and their diameters were chosen uniformly randomly and varied from 2 to 4 and from 60 to 120 pixels, respectively. The diameter of both individual features and texture details was 5 pixels. Texture details within clusters were placed at positions on a regular grid with random Gaussian offsets. The distance between nodes of the regular grid was set to 9 pixels. The amplitude of texture details and individual features varied randomly with normal distribution. White noise was added followed by smoothing with an averaging filter with a 3x3 kernel. The standard deviation of the noise was equal to one third of the amplitude of the texture details.

In the first two experiments, we set the mean amplitude of individual features equal and triple, respectively, of the amplitude mean of texture details. Figs. 4.4(a, b) show the resulting separability measure for the size parameter w varying from 10 to 70 pixels. A comparison of these figures reveal a high degree of immunity of the MTC and ASF diff. to individual features even of high magnitude. In contrast, the performance of other methods severely decreased in the presence of such individual features.

In the third experiment we restricted the class of non-texture areas to smooth areas adjacent to texture regions and to individual features including their neighborhood. Fig. 4.4(c) shows the separability between such restricted non-texture areas and texture regions when mean amplitude of texture details and individual features is equal. A comparison of Fig. 4.4(c) and Fig. 4.4(a) confirms that the superiority in the performance of the MTC and ASF diff. methods stems from its ability to distinguish texture from isolated features as well as from smooth regions adjacent to texture borders. The disadvantage of the ASF diff. in comparison with the MTC is that the ASF diff. takes almost fifty percent more time to compute. In practice, this makes the MTC operator preferable for texture detection.

4.4 Illumination invariant MTC

In accordance with the multiplicative model of image formation, image gray level values are proportional to the product of illumination and surface reflectance. The former is de-



(a) (b) (c)

Fig 4.4: (a) and (b): The measure of separability between texture and non-texture regions as a function of the scale parameter w. The mean amplitude of individual features is equal to the amplitude of texture details in (a) and tripled in (b). (c) The measure of separability of texture regions from areas around individual features and smooth areas adjacent to texture borders. The mean amplitude of individual features equals the amplitude of texture details.

termined by properties of incident light while the later characterizes the observed scenery. Variation of illumination causes a proportional variation of local contrast measured by a difference of gray levels and correspondingly causes variation in response of ψ_{MTC}. On the other hand, a texture contrast descriptor that describes local contrast by means of gray level ratio, would have the advantage of being insensitive to illumination changes. It is worth mentioning that the brightness perceived by the human visual system is also approximately logarithmically proportional to the light intensity incident on the eye.

To define an illumination invariant morphological texture contrast descriptor we apply ψ_{MTC} to $\log(f)$, rather than f, where f is a non-negative function, e.g. 2D gray scale image. Due to the property of morphological operators with flat structuring elements to commute with an increasing and continuous functions, $\psi_{\mathrm{MTC}}(\log(f))$ becomes logarithmically proportional to the ratio of upper and low texture envelopes:

$$\begin{aligned}\psi_{\mathrm{MTC}}(\log f) &= |\gamma_{r_2}\varphi_{r_1}(\log f) - \varphi_{r_2}\gamma_{r_1}(\log f)|^{+} \\ &= |\log \gamma_{r_2}\varphi_{r_1}(f) - \log \varphi_{r_2}\gamma_{r_1}(f)|^{+} = \left|\log \frac{\gamma_{r_2}\varphi_{r_1}(f)}{\varphi_{r_2}\gamma_{r_1}(f)}\right|^{+} .\end{aligned}$$

Since zero values of f are not allowed in $\psi_{\mathrm{MTC}}(\log(f))$ due to the logarithmic function, in practice they are replaced by small values.

Proposition 6. *The* $\hat{\psi}_{\mathrm{MTC}}(f) = \psi_{\mathrm{MTC}}(\log(f))$ *transformation*

1. *is illumination invariant in the sense that* $\hat{\psi}_{\mathrm{MTC}}(af) = \hat{\psi}_{\mathrm{MTC}}(f)$, *for* $a > 0$,

2. *satisfies* $\hat{\psi}_{\mathrm{MTC}}(m/f) = \hat{\psi}_{\mathrm{MTC}}(f)$, *for* $m > 0$.

Proof.

1. Due to Eq. (4.8) or due to the general property of morphological operators with flat structuring elements to commute with an increasing and continuous functions we have

$$\begin{aligned}\hat{\psi}_{\mathrm{MTC}}(af) &= \left|\log \frac{\gamma_{r_2}\varphi_{r_1}(af)}{\varphi_{r_2}\gamma_{r_1}(af)}\right|^{+} = \left|\log \frac{a\gamma_{r_2}\varphi_{r_1}(f)}{a\varphi_{r_2}\gamma_{r_1}(f)}\right|^{+} \\ &= \left|\log \frac{\gamma_{r_2}\varphi_{r_1}(f)}{\varphi_{r_2}\gamma_{r_1}(f)}\right|^{+} = \hat{\psi}_{\mathrm{MTC}}(f) \ .\end{aligned}$$

2. Due to bias invariance of the MTC (first property in Proposition 5) we have

$$\begin{aligned}\hat{\psi}_{\mathrm{MTC}}(m/f) &= \psi_{\mathrm{MTC}}(\log(m/f)) \\ &= \psi_{\mathrm{MTC}}(\log m - \log f) = \psi_{\mathrm{MTC}}(-\log f)\end{aligned}$$

using the property of the MTC to be proportional to signal magnitude (third property in Proposition 5) we obtain

$$= \psi_{\mathrm{MTC}}(\log f) = \hat{\psi}_{\mathrm{MTC}}(f) \ .$$

In remote sensing images that cover large areas, illumination might be different in different parts of the image. Moreover, images even of neighboring areas might be acquired in deferent times. Therefore, for remote sensing images we used the illumination invariant version of the MTC, $\psi_{\mathrm{MTC}}(\log(f))$, in our experiments. The illumination invariant MTC is also used in our final system for detection of objects of interest in Ch. 12.

Fig. 4.5(b, e) illustrates application of the illumination invariant MTC operator to panchromatic satellite in Fig. 4.5(a) and aerial in Fig. 4.5(d) images. The images contain high contrast textured regions, namely urban and forest areas. One can see that the MTC transformation provides a descriptor that has high values in textured regions and low values within smooth areas. The distribution of the descriptor values is highly bimodal in the examples. Therefore, segmentation of textured regions can be accomplished by simple thresholding of the descriptor values. For example, the segmentation results in Fig. 4.5(c, f) superimposed on the original image were obtained by means of automatic

Otsu thresholding [82]. Note that borders of textured regions are well localized and isolated trees were segmented similarly to smooth areas (green).

In a situation where the area sizes of textured and smooth regions differ considerably an automated thresholding may not work properly. A supervised segmentation scheme can be employed such that training samples of textured and smooth regions are provided manually by user. We designed an interactive interface that conveniently allows a user to draw two categories of regions (that are high contrast texture and smooth areas of a comparable spatial size) to be distinguished from each other. Each category can be defined either by a single polygon or by a composition of several polygons drawn over different texture types (e.g forested and urban texture areas). Using the distribution of the MTC values within these manually defined training regions an appropriate threshold can be reliably found. The whole image can then be segmented into texture and non-texture regions by thresholding values of the MTC descriptor.

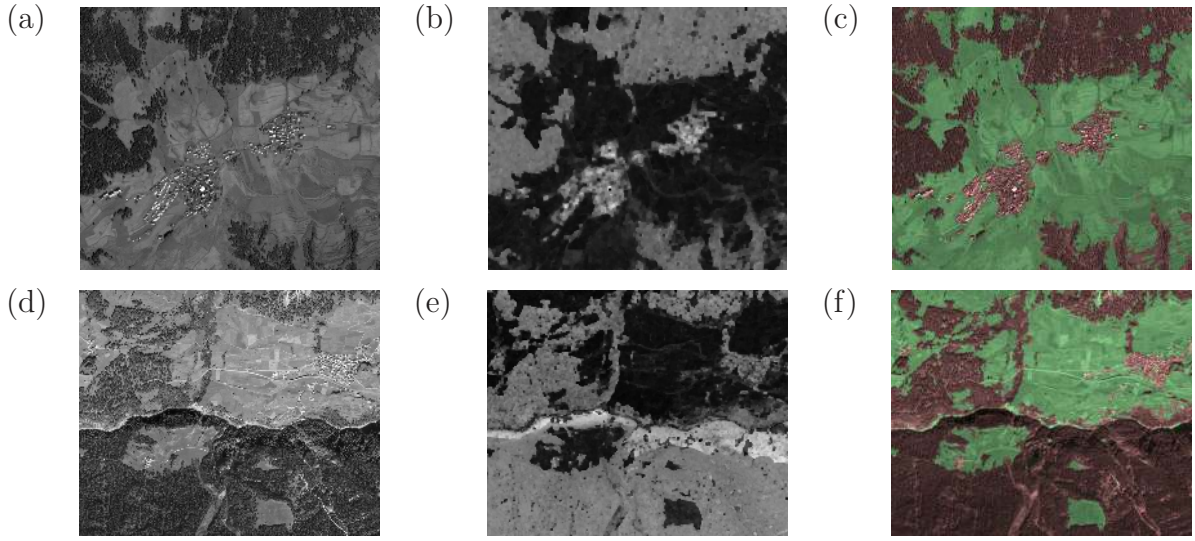


Fig 4.5: a. Pan-chromatic image of 4000x3500 pixel size and 0.5m/pixel resolution captured by the GeoEye-1 satellite (© GeoEye 2011, distributed by e-GEOS). d. Aerial SWISSTOPO image of 6100x5000 pixel size and 0.5m/pixel resolution. Scenery in both images includes high contrast textured regions (urban and forest areas), and comparably smooth field areas. b. and e. The MTC descriptor. c. and f. The segmentation result superimposed on the original image was obtained by automatic thresholding of the MTC descriptor. Brownish areas correspond to high contrast textured regions. Best viewed in digital version.

Chapter 5

Extraction of linear features

In this chapter we focus on detection of isolated features while avoiding detection of parts of neighboring or background texture, i.e. texture details, even if such texture details are similar to features of interest. For example, one may want to detect individual trees distinguishing them from trees of a forest. This problem has mainly been treated in the context of edge detection capable of discarding texture surroundings. For example, recently in [83] a surround inhibition mechanism was introduced to improve edge detection at region boundaries. [84] proposed a normal complexity measure that is able to separate isolated curves and isolated edges from texture in binary images. The paper provides an original theoretical framework, but it is computationally very expensive. In our work, however, we are focused on detection of linear isolated features, which are called here bar edges, or alternatively ridges (bright features) or valleys (dark features).

In Sec. 5.1 we show how the ideas underlying the MTC operator lead to a Morphological Feature Contrast (MFC) operator that aims at the detection of small isolated objects, rather than edges, in textured background. We show how the MFC operator can be incorporated into a scheme for extracting isolated linear features. We show the advantages of this scheme over the approach for the detection of contours with texture background suppression introduced in [83]. Though, we use only gray–scale images for our main task of detection of ruined rectangular structures, in Sec. 5.4 we also show how the MFC operator can be extended to vector-valued images (e.g. multispectral images).

5.1 Extraction of isolated features: Morphological Feature Contrast (MFC)

Using the ideas underlying the MTC operator, below we propose a Morphological Feature Contrast (MFC) operator that extracts isolated features while suppressing texture details of textured background. Using alternating morphological filters, upper and lower texture envelopes were estimated in the MTC approach. To extract bright or dark individual features, we suggest using the difference between the original signal and one of its envelopes, as defined in the following equations

$$\psi^{+}_{\mathrm{MFC}}(f) = |f - \gamma_{r_2}\varphi_{r_1}(f)|^{+} \ , \tag{5.1}$$

$$\psi^{-}_{\mathrm{MFC}}(f) = |\varphi_{r_2}\gamma_{r_1}(f) - f|^{+} \ . \tag{5.2}$$

To extract both types of individual features the sum of two operators should be used

$$\psi_{\mathrm{MFC}} = \psi^{+}_{\mathrm{MFC}} + \psi^{-}_{\mathrm{MFC}} \ . \tag{5.3}$$

We call ψ^{+}_{MFC} and ψ^{-}_{MFC} white and black MFC, respectively. The MFC operators applied to a 1D artificial signal are illustrated in Fig. 4.1. The appropriate sizes r_1, r_2 of the SEs should be chosen such that

$$D_1 < r_1 < D_2, \ S_1 < r_2 < S_2, \tag{5.4}$$

where D_1 is the maximal distance between neighboring texture details, D_2 is the minimal distance to isolated features, S_1 is the maximal size of isolated features, and S_2 is the minimal size of texture regions. For the MFC operators these constraints ensure detection of isolated features and suppression of texture. As stated in the following proposition, the MFC operator ψ_{MFC} shares with the MTC the three properties defined in Sec. 4.2.

Proposition 7. *The ψ_{MFC} operator in Eq. (5.3) fulfills the three properties of Proposition 5, i.e it is bias invariant, invariant to signal inversion, and proportional to signal magnitude.*

These properties make the ψ_{MFC} operator invariant to a constant or slow varying level of signal (DC), signal polarity, and proportional to intensity of an isolated feature. **Proof.**

1. Bias invariance.
   We first prove bias invariance for ψ^+_{MFC} and ψ^-_{MFC} operators

$$\psi^+_{\text{MFC}}(f+a) = \psi^+_{\text{MFC}}(f) \text{ and } \psi^-_{\text{MFC}}(f+a) = \psi^-_{\text{MFC}}(f) \ . \tag{5.5}$$

   The equalities above follow from bias invariance of morphological closings and openings stated in Eq. (4.6), e.g. for ψ^+_{MFC} we have

$$\begin{aligned}\psi^+_{\text{MFC}}(f+a) &= |f+a-\gamma_{r_2}\varphi_{r_1}(f+a)|^+ = |f+a-a-\gamma_{r_2}\varphi_{r_1}(f) \\ &= |f-\gamma_{r_2}\varphi_{r_1}(f)|^+ = \psi^+_{\text{MFC}}(f) \ .\end{aligned}$$

   From Eq. (5.5) it directly follows that $\psi_{\text{MFC}} = \psi^+_{\text{MFC}} + \psi^-_{\text{MFC}}$ is also bias invariant, i.e. $\psi_{\text{MFC}}(f+a) = \psi_{\text{MFC}}(f)$.

2. Invariance to signal inversion.
   We first prove that the following equations are hold

$$\psi^+_{\text{MFC}}(a-f) = \psi^-_{\text{MFC}}(f) \text{ and } \psi^-_{\text{MFC}}(a-f) = \psi^+_{\text{MFC}}(f) \ . \tag{5.6}$$

   The equalities above follow from the property of morphological openings and closings stated in Eq. (4.7), e.g. for $\psi^+_{\text{MFC}}(a-f)$ we have

$$\begin{aligned}\psi^+_{\text{MFC}}(a-f) &= |a-f-\gamma_{r_2}\varphi_{r_1}(a-f)|^+ = |a-f-\gamma_{r_2}(a-\gamma_{r_1}(f))|^+ \\ &= |a-f-a+\varphi_{r_2}\gamma_{r_1}(f)|^+ = |-f+\varphi_{r_2}\gamma_{r_1}(f)|^+ = \psi^-_{\text{MFC}}(f) \ .\end{aligned}$$

   Similarly $\psi^-_{\text{MFC}}(a-f) = \psi^+_{\text{MFC}}(f)$ can be proved. Finally, due to Eq. (5.6) we have

$$\begin{aligned}\psi_{\text{MFC}}(a-f) &= \psi^+_{\text{MFC}}(a-f) + \psi^-_{\text{MFC}}(a-f) \\ &= \psi^-_{\text{MFC}}(f) + \psi^+_{\text{MFC}}(f) = \psi_{\text{MFC}}(f) \ .\end{aligned}$$

3. Proportionality to signal magnitude.
   We first show that for $a \geq 0$ the following equalities are hold

$$\psi^+_{\text{MFC}}(af) = a\psi^+_{\text{MFC}}(f) \quad \text{and} \quad \psi^-_{\text{MFC}}(af) = a\psi^-_{\text{MFC}}(f) \ , \tag{5.7}$$

   while for $a < 0$ the following equalities are true

$$\psi^+_{\text{MFC}}(af) = |a|\psi^-_{\text{MFC}}(f) \quad \text{and} \quad \psi^-_{\text{MFC}}(af) = |a|\psi^+_{\text{MFC}}(f) \ . \tag{5.8}$$

Due to Eq. (4.8) we have for the case of $a \geq 0$

$$\begin{aligned}\psi^{+}_{\mathrm{MFC}}(af) &= |af - \gamma_{r_2}\varphi_{r_1}(af)|^{+} = |af - a\gamma_{r_2}\varphi_{r_1}(f)|^{+} \\ &= a|f - \gamma_{r_2}\varphi_{r_1}(f)|^{+} = a\psi^{+}_{\mathrm{MFC}}(f) \ ,\end{aligned}$$

and for the case of $a < 0$

$$\begin{aligned}\psi^{+}_{\mathrm{MFC}}(af) &= |af - \gamma_{r_2}\varphi_{r_1}(af)|^{+} = |af - a\varphi_{r_2}\gamma_{r_1}(f)|^{+} \\ &= |a(f - \varphi_{r_2}\gamma_{r_1}(f))|^{+} = |-a(\varphi_{r_2}\gamma_{r_1}(f) - f)|^{+} \\ &= |a||(\varphi_{r_2}\gamma_{r_1}(f) - f)|^{+} = |a|\psi^{-}_{\mathrm{MFC}}(f) \ .\end{aligned}$$

Above, we proved only the left parts of Eqs. (5.7, 5.8), i.e. only for the case of the white MFC ψ^{+}_{MFC} operator. The proof for the case of the black MFC ψ^{-}_{MFC} operator is similar.
Thus, for $a \geq 0$ due to Eq. (5.7) we have

$$\begin{aligned}\psi_{\mathrm{MFC}}(af) &= \psi^{+}_{\mathrm{MFC}}(af) + \psi^{-}_{\mathrm{MFC}}(af) = a\psi^{+}_{\mathrm{MFC}}(f) + a\psi^{-}_{\mathrm{MFC}}(f) \\ &= a\psi_{\mathrm{MFC}}(f) \ ,\end{aligned}$$

while for $a < 0$ due to Eq. (5.8) we have

$$\begin{aligned}\psi_{\mathrm{MFC}}(af) &= \psi^{+}_{\mathrm{MFC}}(af) + \psi^{-}_{\mathrm{MFC}}(af) = |a|\psi^{-}_{\mathrm{MFC}}(f) + |a|\psi^{+}_{\mathrm{MFC}}(f) \\ &= |a|\psi_{\mathrm{MFC}}(f) \ .\end{aligned}$$

Therefore, for arbitrary a we can write

$$\psi_{\mathrm{MFC}}(af) = |a|\psi_{\mathrm{MFC}}(f) \ ,$$

which concludes the proof.

In the next Proposition 8 we list additional properties of the MFC that are related to the work in [85]. In that work the authors suggested to use $f - \min(\gamma\varphi(f), f)$ and $\max(\varphi\gamma(f), f) - f$ operators for detection of defects in the noisy background of a metallic surface.

Proposition 8. *The MFC operators obey the following properties*

1. *The white MFC* ψ^{+}_{MFC} *is equivalent to* $f - \min(\gamma_{r_2}\varphi_{r_1}(f), f)$

2. *The black MFC* ψ^{-}_{MFC} *is equivalent to* $\max(\varphi_{r_2}\gamma_{r_1}(f), f) - f$

3. *The MFC* ψ_{MFC} *is equivalent to* $\max(\varphi_{r_2}\gamma_{r_1}(f), f) - \min(\gamma_{r_2}\varphi_{r_1}(f), f)$

Proof.

1. Let us define operator $\hat{\psi} = f - \min(\gamma_{r_2}\varphi_{r_1}(f), f)$.
   For $\gamma_{r_2}\varphi_{r_1}(f) \geq f$ both $\hat{\psi}$ and ψ^{+}_{MFC} operators equal 0 i.e.

   $$\psi^{+}_{\mathrm{MFC}} = |f - \gamma_{r_2}\varphi_{r_1}(f)|^{+} = 0 \text{ and } \hat{\psi} = f - \min(\gamma_{r_2}\varphi_{r_1}(f), f) = 0 \ .$$

   For $\gamma_{r_2}\varphi_{r_1}(f) < f$ we have

   $$\psi^{+}_{\mathrm{MFC}} = f - \gamma_{r_2}\varphi_{r_1}(f) \text{ and } \hat{\psi} = f - \gamma_{r_2}\varphi_{r_1}(f) \ ,$$

   which proofs equivalence of $\hat{\psi}$ and ψ^{+}_{MFC} for arbitrary f.

   The second claim of Proposition 8 can be proved similarly. The third claim directly follows from the first and second claims together.

Fig. 5.1 and Fig. 5.2 show examples of the MFC operator ψ_{MFC} applied to gray-scale images. Additionally, Fig. 4.2(d) shows an example for the MFC operator ψ^{-}_{MFC}. Note that in these and in the following illustrations, dark tones represent high values of transformations extracting isolated features. One can observe that various individual features/objects were highlighted while texture areas were simultaneously suppressed. For example, in the right image in Fig. 5.1 and in the left image in Fig. 5.2 the forest texture area and the texture of the village were suppressed, while isolated buildings (mostly bright roofs) outside the dense village center and isolated trees were preserved in the output image. The MFC operator is capable of suppressing texture areas even if composed of details of higher magnitude and similar shape in relation with the magnitude and shape of individual features. Although, several methods were developed to extract object boundaries (edge features) from textured background, we are not aware of other techniques that perform qualitatively similar to the MFC when extracting blob-like features (as well as features of arbitrary shape).

The ASF diff., the MTC (Ch. 4) and the MFC operators have similar properties (e.g. stated in Proposition 7) and are good in distinguishing texture from isolated features.

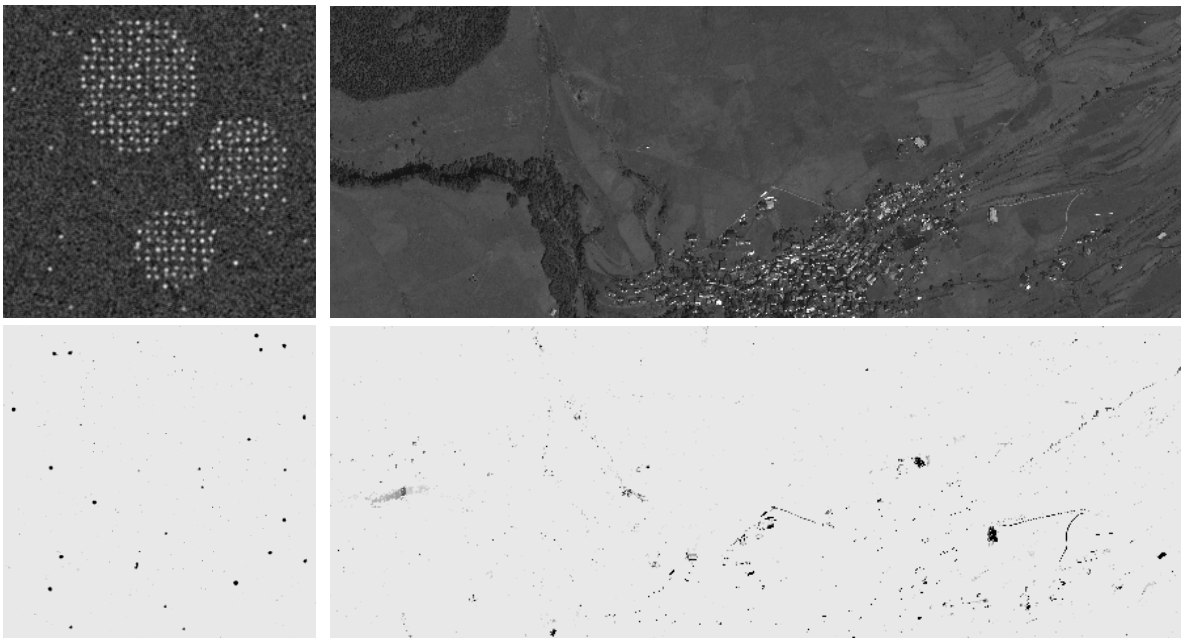


Fig 5.1: First row: 312x312 artificial and 5200x1900 satellite (© GeoEye 2011, distributed by e-GEOS) images. Second row: individual features extracted by means of the MFC operator ψ_{MFC}. $r_1 = 30, r_2 = 10$ for the artificial image and $r_1 = r_2 = 90$ pixels for the satellite image.

The first two are complementary to the MFC operator in the sense that they respond to texture while the MFC responds to individual features. The qualitative behavior of these operators is briefly summarized in Table 5.1, which is conveniently interpreted along with Fig. 4.1. Note also that these operators do not respond to step edges and respond correctly (with 'Low') at smooth regions nearby texture borders and in the vicinity of isolated features.

Table 5.1: Qualitative behavior of the MTC and ASF diff. versus the MFC operators.

	Texture	Isolated features	Isol. features within texture	Smooth regions
MTC & ASF diff.	High	Low	High	Low
MFC	Low	High	High	Low

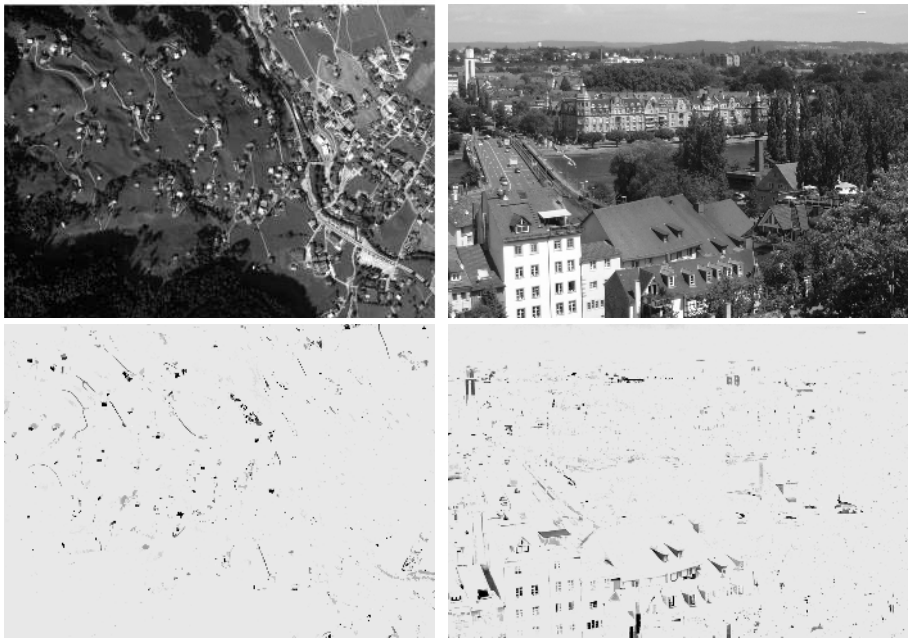


Fig 5.2: First row: 2570x1870 satellite (© GeoEye 2011, distributed by e-GEOS), and 1010x690 digital camera images. Second row: individual features extracted by means of the MFC operator ψ_{MFC}. $r_1 = r_2 = 90$ for the satellite image, and $r_1 = r_2 = 25$ pixels for the digital camera image.

5.2 MFC based extraction of isolated linear features

Above, we have shown that the MFC operators are capable of extracting features of different types with width smaller than r_2. Features of a specific type can be extracted by a sequence of standard morphological transformations, with the structuring element shaped similarly to features. Here we illustrate advantages of the use of the MFC within such a sequence for the case of linear features.

The remote sensing images in Fig. 5.3 (left) contain rectangular structures composed of nearly linear walls that were used as livestock enclosures. The white top-hat transform is commonly used to remove background and emphasize small bright structures in images. Fig. 5.3 (the second column) shows the white top-hat transform followed by a directional filter γ_{lin}. The directional filter is obtained by the point-wise maximum (denoted $\bigvee$) of morphological openings γ_θ with linear SE of a particular length and at different orientations θ

$$\gamma_{\mathrm{lin}} = \bigvee_{\theta} \gamma_\theta \ .$$

This sequence of the top-hat and directional opening[1] operators highlights narrow linear features longer than the length of the linear structuring element. However, texture details are also emphasized. Furthermore, an appropriate threshold setting is required to obtain a binary map of features.

To remove texture while keeping isolated features, the MFC operator ψ^{+}_{MFC} can be applied prior to γ_{lin}.We will call the sequence of the white MFC operator and γ_{lin} the MFC based detector of linear features, see Fig. 5.4. White top-hat followed by the MFC based detector of linear features thresholded at zero level yields a map of linear features with most texture details removed, which is illustrated in an example in Fig. 5.3 in the third column. In our experiments here and in our detector of LSE we used such a sequence of morphological operators (also shown in Fig. 5.4) with a 5x5 SE in the top-hat, $r_1 = 5, r_2 = 10$ in the white MFC, and 15 pixels length for the linear SE at 12 different orientations ranging from 0 to π in the directional opening. Note that threshold tuning was not required to obtain the result. Since we used the white top-hat operator, only bright linear features (ridges) were extracted. Using the black top-hat operator extracts dark linear features (valleys). For the case of linear (elongated) step edges, this approach is adapted by replacing the top-hat transform with the morphological gradient [86].

5.3 Comparison of the MFC and the non-CRF based methods

One of the advantages of the MFC based detector of linear features is that it involves easily tunable geometry related parameters, i.e. r_1, r_2 for the MFC operator, and the length of the structuring element for the directional opening γ_{lin}. These parameters define spatial constraints on the objects to be detected and texture to be suppressed. Moreover, final thresholding can always be performed at zero level, which results in robust detection of linear features in variable scenes and illumination conditions. In contrast, other methods usually involve parameters related to intensities of features or frequency of their appearance. For example, an efficient approach was recently developed for detection of object contours in cluttered scenes by means of biologically motivated non-classical receptive field (non-CRF) inhibition [83]. In this approach an inhibition level needs to be carefully

[1] γ_{lin} fulfills all the properties of algebraic opening [21].

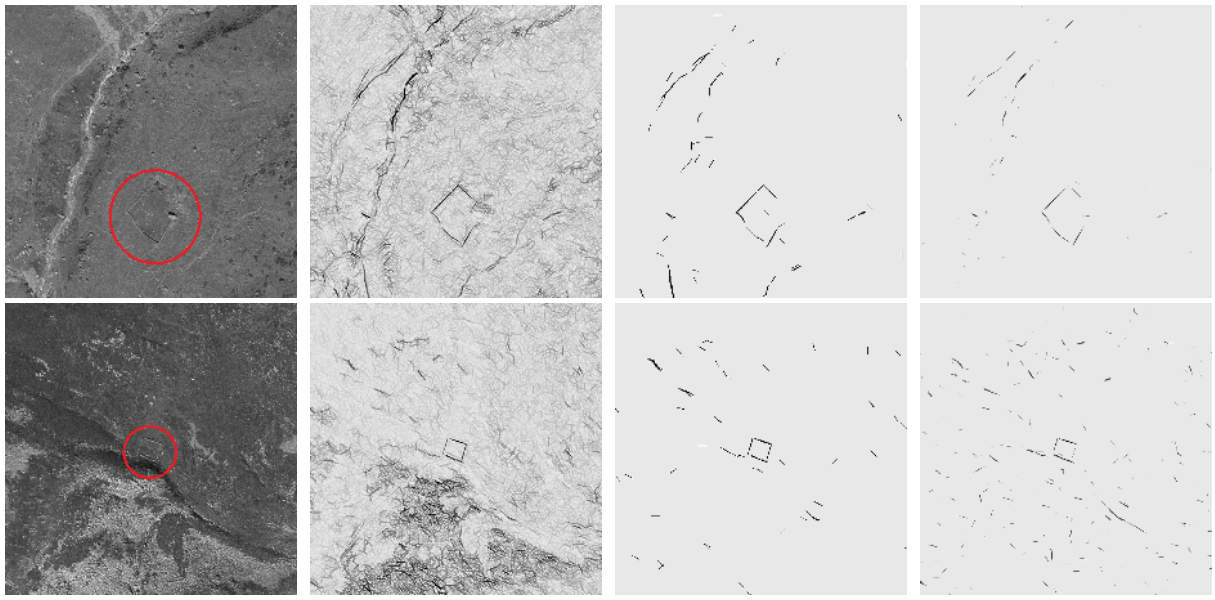


Fig 5.3: First column: Aerial SWISSTOPO images (red channel) of 600x600 pixel size with man-made structures composed of linear walls. Second column: White top-hat transform followed by a directional filter γ_{lin} obtained by the point-wise maximum of morphological openings with linear structuring elements at different orientations. Third column: White top-hat transform followed by the MFC based detector of isolated linear features. Non-zero pixels are shown in black. Fourth column: Non-CRF based detector of isolated linear features.

tuned. The suitable value of this parameter may vary for different images depending on illumination conditions and ratio of features' strength over clutter or texture. A multilevel inhibition technique was suggested in [87] to address this problem. It makes the approach more robust, however, it may reduce the performance when applied to a particular image comparing to the single optimal inhibition level. In addition, a multilevel inhibition involves a fraction constant p with an appropriate value depending on the size and the number of structures of interest in the image.

Below, we compare the MFC based and the non-CRF based detectors of linear features. In this comparison we adapted the isotropic non-CRF inhibition approach [83] for extraction of linear ridge (bar) features by using only an even filter in the Gabor energy. Note, that the Gabor energy is not invariant to constant bias because an even Gabor filter is not a zero-DC filter. The right-most column in Fig. 5.3 shows the result of applying the non-CRF inhibition approach with the parameters chosen for the best visual results, keeping the structures of interest extracted with the lowest level of clutter. In

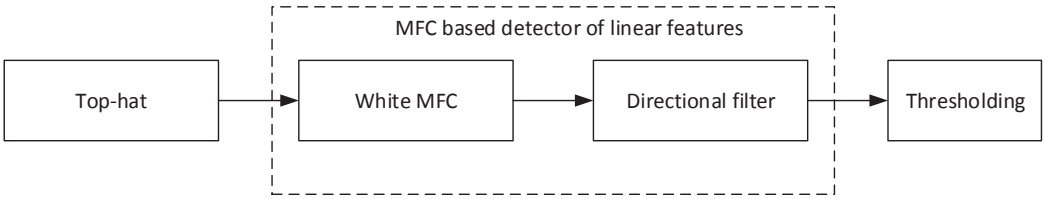


Fig 5.4: The sequence of transformations that extracts linear features while suppressing texture details. The top-hat transform is followed by the MFC based detector of linear features thresholded at 0 level. When the white top-hat is used bright linear features (ridges) are extracted, while in the case of black top-hat dark features (valleys) are extracted.

our example the optimal inhibition level is $\alpha = 1.8$. For the Gabor filter the standard deviation is $\sigma = 2$, the spatial aspect ratio $\gamma = 0.25$, and the wavelength $\lambda = 5.7$. For the inhibition term the standard deviation is $\sigma = 1.5$. For visualization purposes the output of the non-CRF inhibition method was logarithmically transformed.

To quantitatively compare the MFC based and the non-CRF based detectors we generated a set of 26 512x512 artificial images composed of Brodatz textures and patterns of linear features. Ground truth linear features were created as dashed lines with an amplitude equal 0.8, three pixels width and 20 pixels segment length. They were blurred by convolution with a 3x3 averaging kernel and then added to textures (see Fig. 5.5 on the left). The gap between line segments was equal to the length of segments. Line orientation was different for each texture and taken from all possible equally spaced orientations. Prior to the composition, the Brodatz textures were normalized to a unit standard deviation and multiplied by a linear intensity gradient image such that the right texture border was brighter than the left border by four times (see Fig. 5.5 on the left).

Since we have very unbalanced classes (the class of linear features is much smaller in size than the size of texture background), we used receiver operating characteristic (ROC) curves [88] to compare the detection performance of the algorithms. ROC curves were previously used for comparison of edge detectors in [89], which also needed to account for unbalanced classes. A receiver operating characteristic (ROC) curve shows the relationship between true positives detection rate and false positives rate of a particular detector. The true positives rate is the relative number of pixels of linear features that were correctly identified, while the false positive rate is the relative number of pixels of

texture background that were wrongly detected as linear features. An important advantage of ROC curves is that they summarize the performance of detectors for different class priors and detection error costs. Fig. 5.6 shows the resulting ROC curves for low values of false positives, where the MFC based detector is superior. For high false positives the non-CRF approach yields a higher true positives rate. Moreover, the ROC curve of the MFC based detector cannot be generated for true positives rates higher than shown in the figure, because this detector completely removes parts of linear features. This behavior fits visual results in Fig. 5.5 obtained for a couple of particular images (shown in Fig. 5.5 on the left) from the image dataset. For visual purposes, the output of the MFC and non-CRF based detectors was logarithmically transformed. It can be seen that the MFC based approach is more successful in suppressing texture background, but thins linear features.

ROC curves were generated with optimal detector parameters. In our experiments such parameters for the MFC based detector were $r_1 = 10, r_2 = 5$. The length of the linear SE for the directional opening γ_{lin} was equal 5. For the non-CRF based detector optimal parameters were $\sigma = 3.6$, $\gamma = 0.9$, $\lambda = 10.3$ for the Gabor filter and $\sigma = 3$ for the inhibition term. The inhibition level was $\alpha = 1.8$.

5.4 Extension of the MFC operator to vector-valued images

Though, we do not use multispectral images in our system for detection of LSE, in this section we show how the MFC operators can be generalized from gray-scale to vector-valued discrete images, where each pixel is attributed by a vector of values. A multispectral image is an example of such a vector-valued image.

The problem with extending morphological operators to vector-valued images lies in the lack of a natural ordering of vectors. However, some morphological transformations defined in terms of arithmetic differences between morphological operators can naturally be extended to vector-valued images without the need of chosing a vectorial order. Examples of such extended transformations were recently proposed for morphological gradient and for top-hat in [90] and [91]. Using similar ideas, we derive an extended version of the MFC operators below.

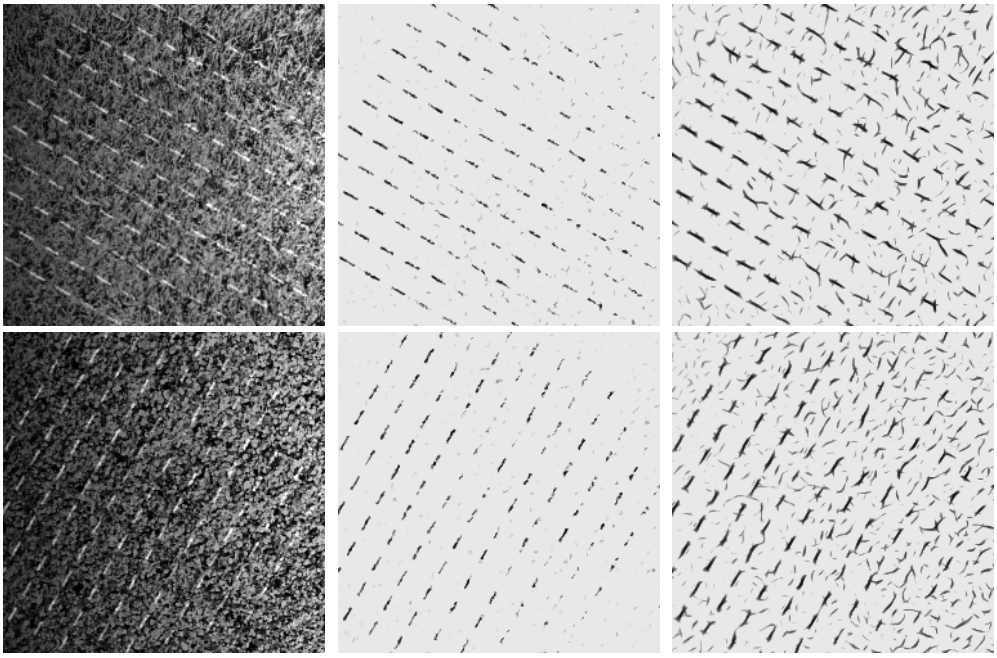


Fig 5.5: Left: Examples of artificial images used for quantitative comparison of the MFC and the non-CRF based detectors of linear features. The images were composed of Brodatz textures and patterns of linear features (see details in Sec. 5.3). Middle: MFC based detector. Right: Non-CRF based detector.

Proposition 9. *The MFC operators $\psi^-_{\mathrm{MFC}}(f), \psi^+_{\mathrm{MFC}}(f)$, defined in Eq. (5.2) and in Eq. (5.1) with structuring elements of sizes r_1 and r_2, can be expressed in the following forms*

$$[\psi^-_{\mathrm{MFC}}(f)](x) = \min_{k\in B^{(2)}_x} \max_{j\in B^{(3)}_k} \min_{i\in B^{(1)}_j} |f(i) - f(x)|^+ \ , \tag{5.9}$$

$$[\psi^+_{\mathrm{MFC}}(f)](x) = \min_{k\in B^{(2)}_x} \max_{j\in B^{(3)}_k} \min_{i\in B^{(1)}_j} |f(x) - f(i)|^+ \ , \tag{5.10}$$

where $B^{(1)}_p$ and $B^{(2)}_p$ are structuring elements of of sizes r_1 and r_2, respectively, shifted to p, and $B^{(3)}$ denotes the structuring element $B^{(1)}$ dilated by $B^{(2)}$.

Proof. Let us denote by δ and ε morphological dilation and erosion, respectively. The MFC operator defined in Eq. (5.2) can be rewritten as follows

$$\begin{aligned}
&[\psi^-_{\mathrm{MFC}}(f)](x) = |[\varepsilon_{r_2}\delta_{r_2}\delta_{r_1}\varepsilon_{r_1}(f)](x) - f(x)|^+ \\
&= |[\min_{k\in B^{(2)}_x} \max_{j\in B^{(3)}_k} \min_{i\in B^{(1)}_j} f(i)] - f(x)|^+ = |\min_{k\in B^{(2)}_x} ([\max_{j\in B^{(3)}_k} \min_{i\in B^{(1)}_j} f(i)] - f(x))|^+
\end{aligned}$$

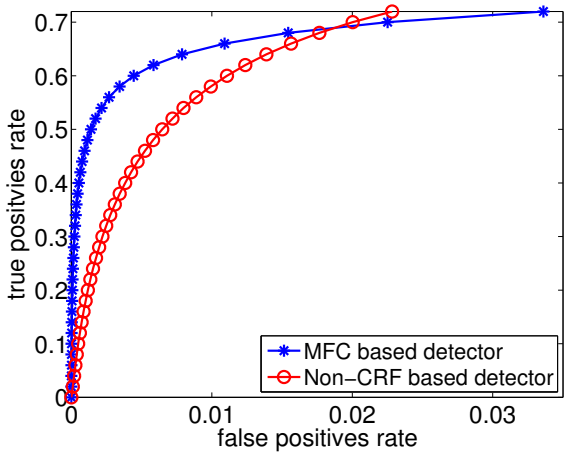


Fig 5.6: ROC curves for the MFC and non-CRF based detectors of linear features.

We note that for an arbitrary function $g(y)$

$$
\begin{aligned}
&|\min_{y\in B_x}(g(y)-f(x))|^+ \\
&= \begin{cases} \min_{y\in B_x}(g(y)-f(x)), & \text{if } \forall y\in B_x : g(y)\geq f(x) \\ 0, & \text{otherwise} \end{cases} \\
&= \min_{y\in B_x}|g(y)-f(x)|^+ \ ,
\end{aligned}
$$

and

$$
\begin{aligned}
&|\max_{y\in B_x}(g(y)-f(x))|^+ \\
&= \begin{cases} \max_{y\in B_x}(g(y)-f(x)), & \text{if } \exists y\in B_x : g(y)\geq f(x) \\ 0, & \text{otherwise} \end{cases} \\
&= \max_{y\in B_x}|g(y)-f(x)|^+ \ .
\end{aligned}
$$

Thus, we can proceed with

$$
\begin{aligned}
&|\min_{k\in B_x^{(2)}}([\max_{j\in B_k^{(3)}}\min_{i\in B_j^{(1)}} f(i)]-f(x))|^+ = \min_{k\in B_x^{(2)}}|[\max_{j\in B_k^{(3)}}\min_{i\in B_j^{(1)}} f(i)]-f(x)|^+ \\
&= \min_{k\in B_x^{(2)}}|\max_{j\in B_k^{(3)}}([\min_{i\in B_j^{(1)}} f(i)]-f(x))|^+ = \min_{k\in B_x^{(2)}}\max_{j\in B_k^{(3)}}|[\min_{i\in B_j^{(1)}} f(i)]-f(x)|^+ \\
&= \min_{k\in B_x^{(2)}}\max_{j\in B_k^{(3)}}|\min_{i\in B_j^{(1)}}(f(i)-f(x))|^+ = \min_{k\in B_x^{(2)}}\max_{j\in B_k^{(3)}}\min_{i\in B_j^{(1)}}|f(i)-f(x)|^+.
\end{aligned}
$$

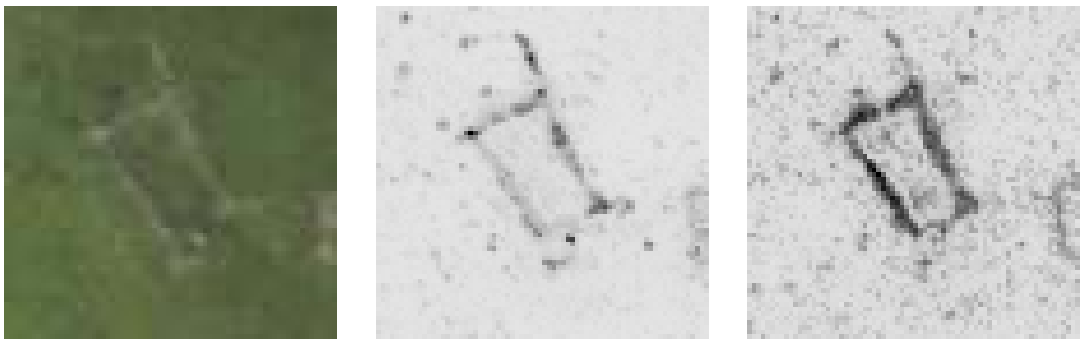
This proves Eq. (5.9). Eq. (5.10) can be proved similarly.

We now define a new vectorial MFC operator $\psi_{\mathrm{MFC}}(\bar{f})$ that applies to vector-valued images $\bar{f}$. We replace the non-negative difference between intensity values in Eq. (5.9) and Eq. (5.10) by a suitable metric distance D between vectors,

$$[\psi_{\mathrm{MFC}}(\bar{f})](x) = \min_{k \in B_x^{(2)}} \max_{j \in B_k^{(3)}} \min_{i \in B_j^{(1)}} D(\bar{f}(x), \bar{f}(i)) \ . \tag{5.11}$$

In contrast to $\psi^+_{\mathrm{MFC}}(f)$ and $\psi^-_{\mathrm{MFC}}(f)$, the $\psi_{\mathrm{MFC}}(\bar{f})$ operator extracts both dark and bright structures when applied to multispectral images. If one is interested in extracting either dark (or bright) structures only, i.e. structures having low (or high) values relative to background in all channels, pseudo-distances may be used. For example, instead of using the D_∞ distance, pseudo distances defined by $D^+_\infty(\bar{f}, \bar{g}) = \max_i |f_i - g_i|^+$ and $D^-_\infty(\bar{f}, \bar{g}) = \max_i |g_i - f_i|^+$ may be employed.

Vectorial operators may or may not be preferable to an independent processing of channels of a vector-valued image followed by integration of the results. Another alternative is to transform a vector valued image to a single channel image before processing. Finding a proper transformation or a way to combine independently processed channels is a task dependent problem, often approached by trial and error. A similar problem appears in the case of vectorial MFC, because a suitable distance must be chosen. Fig. 5.7 shows examples of vectorial MFC with Euclidean $D_2(\bar{f}, \bar{g}) = \|\bar{f} - \bar{g}\|_2$ and angular (spectral) distances $D_\alpha(\bar{f}, \bar{g}) = (\bar{f} \cdot \bar{g})/(\|\bar{f}\|_2 \|\bar{g}\|_2)$. A comparative evaluation of the vectorial MFCs, however, is beyond the scope of this thesis.



(a) (b) (c)

Fig 5.7: An example of application of the vectorial MFC operator to an aerial RGB (SWISSTOPO) image patch with a structure shown in (a). Vectorial MFC with (b) Euclidian distance and (c) angular distance.

Chapter 6

Detection of candidate locations

We detect candidate locations using a map of bar edges (ridges and valleys) that were extracted with the Morphological Feature Contrast (MFC) line detector that extracts linear features, while suppressing texture elements of cluttered background. We also experimented with other approaches [92, 93, 87], but these are either not sensitive enough to extract faint edges of enclosures, or generate lots of clutter edges depending on the parameters used. The parameterless line segment detector of [94], which is known to provide robust results for a large range of images, misses faint edges of ruined enclosures.

Our approach to detection of candidate locations relies on the medial axis of an inverted binary map of edges. The basic idea is to use medial axis junction points as candidate structure locations. The junction points are located at centers of structures that are enclosed from at least three sides, which is the case for LSE structures of interest. On the other hand, spurious structures, such as lines, corners, junctions, and other simple curves, do not generate junction points of the medial axis.

In [65, 95, 96] the medial axis of a shape was extracted by thresholding the average flux of the gradient field of the Euclidean distance function D to the boundary of the shape. The average flux of the gradient field through the boundary $\partial\mathcal{N}$ of a region is defined as the corresponding flux normalized by the length of the boundary

$$F(\nabla D) = \frac{\oint_{\partial\mathcal{N}} \nabla D \cdot \mathbf{n} \, \mathrm{d}L}{\oint_{\partial\mathcal{N}} \mathrm{d}L}, \tag{6.1}$$

where $\mathbf{n}$ denotes the inward normal[1] to the boundary $\partial\mathcal{N}$ and $\mathrm{d}L$ is the boundary element. As the region $\mathcal{N}$ shrinks to a point, the average flux F approaches zero at non-medial points and non-zero values at the medial axis of the shape.

In practice, we detect candidate points by finding local maxima of a discrete approximation of the average flux $F(\nabla D)$ through the boundary of a small disk $\mathcal{N}$, where D is the distance function of the binary edge map. These local maximum points usually correspond to junction points of the medial axis of the inverted binary edge map. Only local maxima with the average flux greater than 0.5 were taken into account. Fig. 6.1 shows examples of detections (in red) overlaid on the average flux, which has positive extrema on the medial axis (white) and negative extrema on the bar edges (black)[2]. Note that there are a few redundant detections due to the discrete nature of computations and a general problem of the sensitivity of medial axes to small details in the boundary.

In a related approach [63], non-maxima suppression was applied to the average flux of the normalized gradient vector flow (GVF) [97, 98] in order to detect medial feature points. Using the GVF instead of the gradient field of the distance transform of the edge map allowed detection of medial feature points directly from the gray-scale image without the need of edge extraction. However, GVF may ignore weak gradients of low contrast structures of interest. In addition, computing GVF might be too slow on large images, depending on the number of predefined iterations.

We combine candidate points separately obtained from the binary maps of ridge and valley edges. The structures that are within a window around the candidate points p are further analyzed. The size of the analysis window should be set proportionally to the size of candidate structures. The adaptive size of the window can be determined proportionally to the value of the distance transform $D(p)$ at a candidate location p, i.e. the smallest distance of p to a candidate structure. Particularly, we use the circular window of radius

$$r = D(p)\sqrt{b^2 + 1}$$

centered at p that covers a fragmented rectangle with small side length $2D(p)$, as shown in Fig. 6.2. The figure shows a fragmented rectangle, the shaded circular window, and the candidate point p. The candidate point is located at the junction point of the rectangle's skeleton, part of which is shown in a dashed line. The value of b in the equation above

[1] In [65, 95, 96] the outward normal was used.

[2] We used valley edges for the left figure and ridge edges for the right figure.

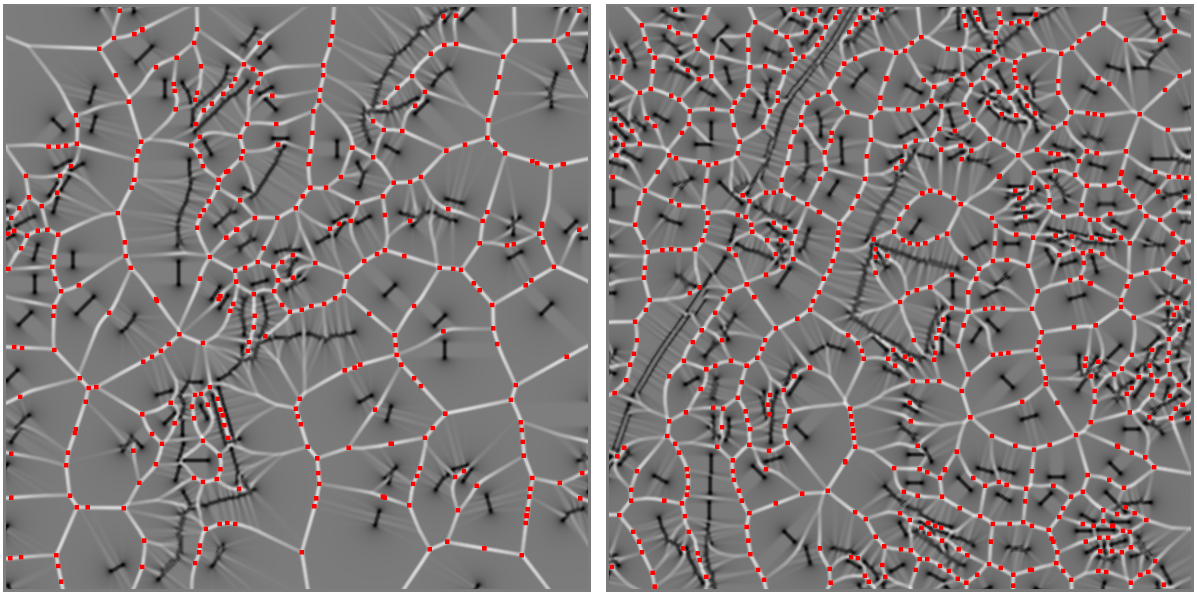


Fig 6.1: Average flux of the gradient field of the distance function computed for edge maps of the images in Fig. 1.1. The medial axis coincides with positive singularities of the flux (white), while edges coincide with negative singularities (black). Local maxima (red points) are used as candidate locations.

can be determined from the largest allowed aspect ratio $a = \frac{D(b+1)}{2D}$ of a rectangle to be covered by the analysis window. In our experiments the aspect ratio a was set to 1.2, which gives $b = 2a - 1 = 1.4$. We discarded all candidate points having a distance D less than 15 or greater than 90 pixels, which limits the distances between opposite walls of the structures to be in between 7.5m and 45m.

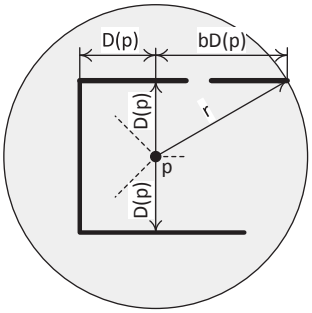


Fig 6.2: Variable radius r of the circular analysis window at a candidate point p is determined from the value of the distance transform $D(p)$ and the predefined maximally allowed aspect ratio $a = (b+1)/2$ of a (fragmented) rectangle (bold line) to be covered by the window. The candidate point p is located at the junction point of the skeleton, part of which is shown in dashed line.

Chapter 7

Extraction and modeling of linear segments

Given a candidate location and edge points accompanied by estimated orientations, we extract and parameterize linear segments, each of which is a group of aligned edge points. Linear segments are represented by a triple of parameters (θ, r, l) found by the use of a local Hough transform centered at the candidate points. We use the Hough transform in the form introduced in [67], where a line is defined by the orientation θ of the normal and a distance r from the origin

$$r = x \cos\theta + y \sin\theta. \tag{7.1}$$

The spatial coordinates of an edge point are x, y. We use the parametrization $\theta \in [0, 360)$ and $r \in (0, \infty)$ of a Hough plane. A peak at (θ, r) in the Hough plane corresponds to a line. The peaks are detected as regional maxima in the Hough plane that was discretized with $\Delta\theta = 3^\circ$ and $\Delta r = 1$ pixel. The parameter l in the triple (θ, r, l) is the number of points that belong to the linear segment. To better relate the parameter l to the length and avoid its dependence on the width of the extracted edges, we perform their thinning [99] prior to clustering in a Hough plane. The thinning reduces the width of edges to one pixel.

Since edges were extracted together with their orientations, r can be directly computed for each edge point (x, y) using Eq. (7.1). Thus, each edge point votes for a single point in the (θ, r) plane instead of voting for a curve as suggested in [67]. This idea, which was used already in [100] for clustering of short ridge features, considerably eases extraction of

meaningful peaks in the Hough plane. This voting technique computed in a local window can actually be considered an extension of edge orientation histograms, which are at the core of most common feature sets used for detection of certain object classes, [101, 102]. In contrast to edge orientation histograms, the (θ, r) plane based technique allows detection of not only dominant orientations of local features, but also their spatial alignment.

Detected in the Hough plane lines correspond to either a single connected linear segment, or to several aligned linear connected components. In the latter case, the connected components with gaps smaller than a predefined threshold are considered a single linear segment (see the segment $\mathcal{S}_j$ in Fig. 8.1). Connected components with larger gaps are considered separate linear segments, which allows for additional valid configurations of linear segments discussed in Sec. 8.1. Though, the idea of setting the minimal gap between linear segments is simple, its implementation on a discrete grid is not straightforward. Therefore, below we give the details on the algorithm we designed in order to generate a set of linear segments from lines detected in the Hough plane.

For each peak in the Hough plane with (θ_i, r_i, l_i)

1. Extract points p_i corresponding to the line i with (θ_i, r_i, l_i)
2. Extract $n_i \geq 1$ connected components $\mathrm{CC}_{i,j=1,\ldots,n_i}$ that belong to the line i
3. Determine end points for each $\mathrm{CC}_{i,j=1,\ldots,n_i}$ by finding maximum and minimum projections of $p \in \mathrm{CC}_{i,j=1,\ldots,n_i}$ on the tangent direction of $\mathrm{CC}_{i,j=1,\ldots,n_i}$ (which differs by 90 degrees from θ_i)
4. Build a graph with nodes corresponding to connected components $\mathrm{CC}_{i,j=1,\ldots,n_i}$. Set an edge between two nodes if the distance between end points of the corresponding connected components is small enough (we used 3 pixels as a predefined threshold)
5. Find $1 \leq k_i \leq n_i$ connected components of the graph (do not confuse with the spatial connected components CC that correspond to nodes of the graph), e.g. using depth-first search. Each connected component of the graph corresponds to a linear segment that we denote $\mathcal{S}$
6. For each of k_i linear segments $\mathcal{S}$ assign θ_i, r_i of the line i, points p that belong to the corresponding subset of the spatial connected components, and recalculate l which is the number of these points.

The set of all $m = \sum_i k_i$ linear segments $\mathcal{S}$ will be further used to infer the presence of a (possibly ruined) rectangular structure at a given candidate location.

Chapter 8

Rectangularity and size features

We introduce the rectangularity f_{R} and size f_{S} features for detection of approximately rectangular enclosure structures. These structures are modeled by convex configurations of linear segments with orientation angles constrained to be close to zero or ninety degrees. The rectangularity feature has highest values for perfect rectangles, but also sensitive to incomplete and fragmented ones. On the other hand, it yields zero value for structures composed of less than three sides, which allows us to avoid a large number of random configurations of linear segments occurring in images with complex or cluttered background. The rectangularity f_{R} and size f_{S} features are computed from a set of linear segments $\mathbf{W} = \{\mathcal{S}_i, i = 1, ..., m\}$ that were already extracted from an image for each of the candidate locations.

8.1 Valid configurations of linear segments

We define a valid configuration of linear segments $\mathbf{C} \subseteq \mathbf{W}$ that can be a part of a rectangular structure. We require angles $\beta_{k,j}$ between linear segments $\mathcal{S}_k, \mathcal{S}_j \in \mathbf{C}$ to be close to either zero, 180°, or right angles. An angle tolerance α will be set to control the strictness of the angle constraint. We define $\beta_{k,j}$ as

$$\beta_{k,j} = \min(|\theta_{\mathcal{S}_k} - \theta_{\mathcal{S}_j}|, 360 - |\theta_{\mathcal{S}_k} - \theta_{\mathcal{S}_j}|). \tag{8.1}$$

Note that $\beta_{j,k} = \beta_{k,j}$ and $\beta \in [0, 180]$, since $\theta \in [0, 360)$.

The angle constraint alone does not suffice to restrict configurations to be perceptually close to rectangles or rectangle parts. We therefore define a second constraint that requires the valid configuration to be nearly convex in the sense that extension of all linear segments of the configuration can form a nearly convex contour. The convexity tolerance t will be defined to control the strictness of the convexity constraint. For a convex configuration of linear segments it is required that a half plane generated by each segment includes all other segments of the configuration. Additionally, we require all these half planes to contain the candidate point around which we search for a rectangular structure. Pair-wise convexity constraints suffice to verify the convexity of a configuration containing the given candidate point. We define the pair-wise convexity measure τ for a pair of linear segments $\mathcal{S}_k, \mathcal{S}_j$, each with corresponding attributes of size $l_{\mathcal{S}}$, orientation $\theta_{\mathcal{S}}$, and distance $r_{\mathcal{S}}$ to the candidate point p_0, as

$$\tau_{k,j} = \max(\tilde{\tau}_{k,j}, \tilde{\tau}_{j,k}), \tag{8.2}$$

$$\tilde{\tau}_{k,j} = \frac{1}{l_j} \sum_{p \in \mathcal{S}_j} H((p - p_0)^T \cdot n_k - r_k), \tag{8.3}$$

where $n_k = (\cos\theta_k, \sin\theta_k)^T$ is the unit normal of $\mathcal{S}_k$ and $H(u)$ is an indicator function

$$H(u) = \begin{cases} 1, & u > 0 \\ 0, & u \leq 0 \end{cases} .$$

$\tilde{\tau}_{k,j}$ measures the relative number of points in the segment $\mathcal{S}_j$ that are behind the segment $\mathcal{S}_k$, relative to the given candidate point p_0 as illustrated in Fig. 8.1. Note that $\tau \in [0, 1]$, and $\tau_{k,j} = \tau_{j,k}$, while $\tilde{\tau}_{k,j} \neq \tilde{\tau}_{j,k}$.

Definition 1. Let $\alpha \in [0, 45], t \in [0, 1]$, a candidate point p_0, and a configuration $\mathbf{C}$ of linear segments be given. If for all pairs $\mathcal{S}_k, \mathcal{S}_j \in \mathbf{C}$, $j \neq k$, one of the inequalities of the angle constraint

$$\beta_{k,j} \leq \alpha \text{ or } |90 - \beta_{k,j}| \leq \alpha \text{ or } 180 - \beta_{k,j} \leq \alpha \tag{8.4}$$

and the convexity constraint

$$\tau_{k,j} \leq t \tag{8.5}$$

both hold, then $\mathbf{C}$ is called a (t,α)-valid configuration located around p_0, and denoted by $\mathbf{C}_{p_0}^{t,\alpha}$.

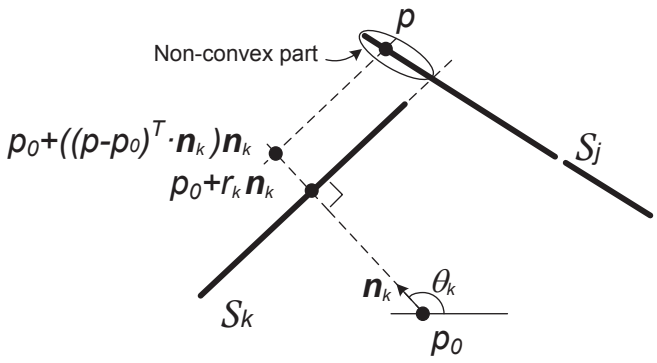


Fig 8.1: The fraction of points p of $\mathcal{S}_j$ that violates the convexity constraint relative to $\mathcal{S}_k$ and p_0 is given by $\tilde{\tau}_{k,j}$ in Eq. (8.3). Note that linear segments can be fragmented having small gaps as in $\mathcal{S}_j$.

For the sake of brevity, we usually omit the indices t, α and the reference point p_0, mentioning that $\mathbf{C}$ is a valid configuration. Valid configurations include not only perfect rectangles or their parts, but also convex polygons or their parts with angles around either 90 or 180 degrees. This is important in practice since approximately rectangular structures are better modeled by such polygons rather than by perfect rectangles.

8.2 Rectangularity measure of a valid configuration

A couple of poorly aligned short segments can be a valid configuration as far as the tolerances t, α allow. There is a need to rank valid configurations according to their similarity to a canonical rectangle. To find and rank valid configurations we construct an undirected graph $\mathbf{G}^w$ from the given set $\mathbf{W}$ of linear segments in a window centered at a candidate point p_0. The graph $\mathbf{G}^w$ has nodes $j = 1, .., m$ corresponding to the segments $\mathcal{S}_1, .., \mathcal{S}_m \in \mathbf{W}$. Each node j is attributed by a triple of parameters (θ_j, r_j, l_j), i.e. orientation, distance to the reference point p_0, and size of the linear segment. An edge $\{k, j\}$ is attributed with the angle $\beta_{k,j}$ and the pair-wise convexity $\tau_{k,j}$ of the corresponding pair of segments $\mathcal{S}_k, \mathcal{S}_j$. An edge $\{k, j\}$ is included in the graph $\mathbf{G}^w$ if $\beta_{k,j}$ and $\tau_{k,j}$ satisfy the constraints in Eqs. (8.4, 8.5). This attributed graph encodes properties of linear segments and their spatial relationships. Due to the graph construction and Definition 1, valid configurations $\mathbf{C}$ correspond to fully connected subgraphs $\mathbf{G}^c$, also called cliques, of the graph $\mathbf{G}^w$.

Below we introduce the new rectangularity measure $\rho(\mathbf{G}^c)$ that ranks a clique $\mathbf{G^c}$ corresponding to a valid configuration $\mathbf{C} \subseteq \mathbf{W}$. We define the measure with the following properties in mind. The rectangularity measure shall yield higher values for configurations with

1. higher degree of convexity given by lower values of the convexity measure τ
2. higher degree of angle alignments given by angles β
3. longer linear segments given by larger l.

In addition, the proposed rectangularity measure shall

4. have the increasing property $\rho(\mathbf{G}_1^c) \leq \rho(\mathbf{G}_2^c)$ for $\mathbf{G}_1^c \subseteq \mathbf{G}_2^c$. Thus, the rectangularity measure of a larger encompassing clique has a higher value
5. yield a zero value for configurations of linear segments with less than three sides of a rectangle. Thus, a non-zero rectangularity indicates existence of at least three-sided structure.

We define the rectangularity measure of a graph clique $\mathbf{G}^c$ in terms of sums over its undirected edges $\{k, j\} \in E^c$

$$\rho(\mathbf{G}^c) = \left(\left(\sum_{\{k,j\} \in E^c} l_k l_j \mathrm{f}_{90}(\beta_{k,j}) \mathrm{f}_{cv}(\tau_{k,j}) \right) \times \left(\sum_{\{k,j\} \in E^c} l_k l_j \mathrm{f}_{180}(\beta_{k,j}) \mathrm{f}_{cv}(\tau_{k,j}) \right) \right)^{\frac{1}{4}}, \quad (8.6)$$

where $\mathrm{f}_{90}, \mathrm{f}_{180}$, and f_{cv} are functions depicted in Fig. 8.2. f_{90} and f_{180} equal zero for angles β that deviate from the mode center larger than the angle tolerance α that is used in constraint Eq. (8.4). f_{cv} equals zero for the convexity measure τ larger than the convexity tolerance t that is used in constraint Eq. (8.5). We used $\alpha = 35^\circ$ and $t = 0.3$. The exact definitions of f_{90}, f_{180}, and f_{cv} are given below with the use of mode-functions $m_{\mu,\delta}$ with peak location μ and spread δ

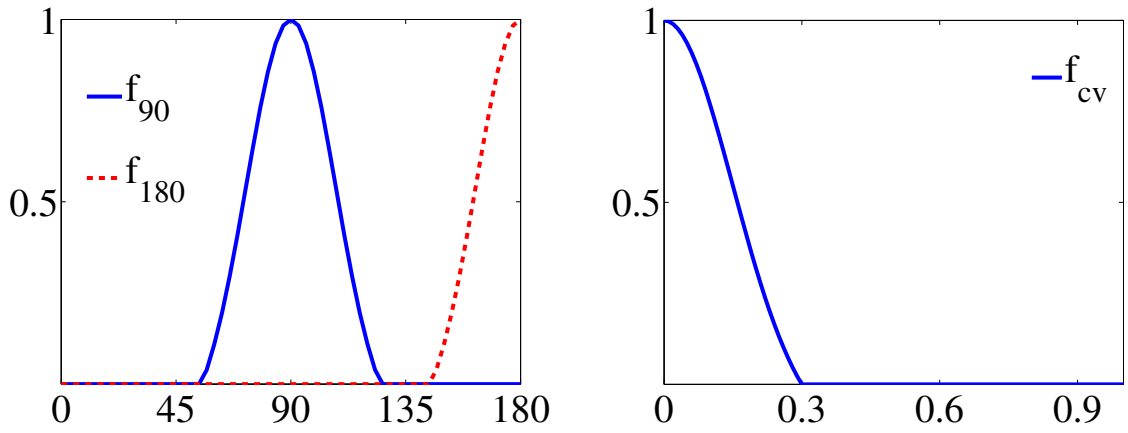


Fig 8.2: Functions f_{90} (left figure, solid blue curve), f_{180} (left figure, dashed red curve), and f_{cv} (right figure) used in the rectangularity measure in Eq. (8.6).

$$\begin{aligned} f_{90}(\beta) &= m_{90,\delta}(\beta), \quad \delta < 45, \\ f_{180}(\beta) &= m_{180,\delta}(\beta), \quad \delta < 45, \\ f_{cv}(\tau) &= m_{0,\delta}(\tau), \end{aligned} \tag{8.7}$$

where

$$m_{\mu,\delta}(u) = \begin{cases} \frac{1}{1-a}(g_{\mu,\sigma}(u) - a), & g_{\mu,\sigma}(u) - a > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{8.8}$$

$$g_{\mu,\sigma}(u) = \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right),$$

and a is determined such that

$$m_{\mu,\delta}(u) = 0, \text{iff } |u - \mu| > \delta.$$

Note that only half of the second mode is present in f_{180}, which is shown in Fig. 8.2. For f_{90} and f_{180} the peak is at $\mu = 90$ and $\mu = 180$ degrees, respectively. The spread $\delta = \alpha$ for both f_{90} and f_{180}, where α is the angle tolerance. For f_{cv} the parameters are $\mu = 0$ and $\delta = t$, where t is the convexity tolerance. The parameter σ controls the peakiness of the function shape and is less influential in practice. In our experiments we set $\sigma = \frac{\alpha}{2}$ for f_{180} and f_{90} functions and $\sigma = \frac{t}{2}$ for f_{cv}.

The first factor of $\rho(\mathbf{G}^c)$ in Eq. (8.6) yields a non-zero value only if the valid configuration $\mathbf{C}$ contains at least one pair of approximately perpendicular linear segments that

fulfill the convexity constraint in Eq. (8.5). The second factor is non-zero only if the valid configuration contains at least one pair of approximately parallel linear segments[1]. The product of these two factors is non-zero only if the valid configuration $\mathbf{C}$ contains at least one pair of parallel and one pair of perpendicular linear segments. The angles between linear segments of these parallel and perpendicular pairs are restricted to be approximately 0, 180, or 90 degrees since $\mathbf{C}$ is a valid configuration with linear segments constrained by Eq. (8.4). Thus, a non-zero rectangularity measure insures a valid configuration $\mathbf{C}$ containing at least one triple of segments (as required in property 5 above) arranged in a Π-like structure. This property allows suppression of a large number of configurations originating from clutter (e.g. lines, corners, junctions etc.).

The introduced rectangularity measure also exhibit the fourth (increasing) property above. This is because the larger encompassing clique can only increase the number of summands in each of the two sums in Eq. (8.6) and because each of the terms (l, f_{90}, f_{cv}) within the summands are all non-negatives. This property essentially speeds up the search for the optimal clique that gives the maximal rectangularity measure, as discussed in Sec. 8.3. It is straightforward to verify that the first three properties above are also satisfied by the introduced rectangularity measure.

Note that the rectangularity measure scales linearly with the spatial size of rectangles. It follows directly from the definition in Eq. (8.6) that scaling a configuration of linear segments of a corresponding graph clique scales its rectangularity measure by the same factor. Therefore, the rectangularity measure scales linearly with the spatial size of rectangles. In fact, for a rectangular structure with perfectly aligned linear segments (the structure can still be fragmented), i.e. for the case of all angles β being equal to either 90 or 180 degrees and all τ being equal to 0, the rectangularity measure reduces to $((L_1 + L_3)(L_2 + L_4)(L_1L_3 + L_2L_4))^{\frac{1}{4}}$, where $L_i, i = 1, .., 4$ are the sums of segment sizes of four sides of a rectangle, such that index pairs $1, 3$ and $2, 4$ correspond to parallel sides. If more sides than one are missing, the expression above equals zero. Note that the rectangularity measure is a function of graph node and edge attributes and does not require explicit partitioning of a valid configuration of linear segments into four subsets corresponding to four sides of a hypothesized rectangle as was required in our preliminary work [4].

[1] f_{cv} in the second term has only a small impact on results. It reduces the rectangularity measure for configurations with badly aligned opposite sides with a non-zero convexity measure.

8.3 Rectangularity feature

Given a set of linear segments $\mathbf{W}$ in an analysis window, we define below the rectangularity feature f_{R} of the corresponding graph $\mathbf{G}^w$. We denote the set of cliques of $\mathbf{G}^w$ as $\mathcal{K}(\mathbf{G}^w)$. The rectangularity feature of $\mathbf{G}^w$ is defined as the maximal rectangularity measure ρ of the cliques in $\mathcal{K}(\mathbf{G}^w)$

$$f_{\mathrm{R}}(\mathbf{G}^w) = \max_{\mathbf{G}^c \in \mathcal{K}(\mathbf{G}^w)} \rho(\mathbf{G}^c). \tag{8.9}$$

The corresponding optimal clique is

$$\mathbf{G}^c_{opt} = \operatorname*{argmax}_{\mathbf{G}^c \in \mathcal{K}(\mathbf{G}^w)} \rho(\mathbf{G}^c). \tag{8.10}$$

Due to the increasing property of ρ (the fourth property of the rectangularity measure stated in Sec. 8.2), the maximum can be searched over the set of maximal cliques[2] only, denoted here by $\mathcal{M}(\mathbf{G}^w)$. That is

$$f_{\mathrm{R}}(\mathbf{G}^w) = \rho(\mathbf{G}^c_{opt}) = \max_{\mathbf{G}^c \in \mathcal{M}(\mathbf{G}^w)} \rho(\mathbf{G}^c). \tag{8.11}$$

Since the set of maximal cliques $\mathcal{M}(\mathbf{G}^w) \subseteq \mathcal{K}(\mathbf{G}^w)$ is much smaller than the set of graph cliques $\mathcal{K}(\mathbf{G}^w)$, the number of times the rectangularity measure ρ needs to be evaluated in Eq. (8.11) is considerably reduced in comparison to Eq. (8.9). Since, in addition, there are efficient algorithms for finding maximal cliques, e.g. [103], we compute the rectangularity feature by an exhaustive search for the maximum in Eq. (8.11).

Fig. 8.3 (left) shows an example of a given set $W = \{\mathcal{S}_1, \mathcal{S}_2, .., \mathcal{S}_6\}$ of linear segments and the optimal configuration $\mathbf{C}_{opt} = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_5\}$ in red, while Fig. 8.3 (right) shows the corresponding graph $\mathbf{G}^w$ and the optimal maximal clique $\mathbf{G}^c_{opt}$ in red. There are two additional maximal cliques $\mathbf{G}^c_1$ and $\mathbf{G}^c_2$ and corresponding valid configurations $\mathcal{C}_1 = \{\mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4, \mathcal{S}_6\}, \mathcal{C}_2 = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4\}$. They, however, have lower rectangularity values $\rho(\mathbf{G}^c_1) < \rho(\mathbf{G}^c_{opt}), \rho(\mathbf{G}^c_2) < \rho(\mathbf{G}^c_{opt})$.

Fig. 8.4 shows a couple of examples of the rectangularity feature computed for the real

[2]Maximal cliques are cliques that are not contained in larger cliques.

satellite and areal images. The first row shows detected bar edges and candidate points[3]. The rectangularity feature f_{R} computed at the candidate points is visualized by colored disks in the second row. Color saturation increases and hue is changing from yellow to red for growing values of the features in accordance with the color bar in the bottom. As expected, high values were obtained at positions of LSE while zero or low values were obtained at most other candidate positions. One can see that the rectangularity feature map is quite sparse. This is partially because the rectangularity feature has zero value for spurious structures with less than three sides.

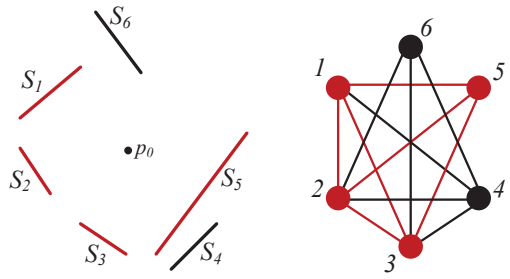


Fig 8.3: Left: A set $W = \{\mathcal{S}_1, \mathcal{S}_2, .., \mathcal{S}_6\}$ of linear segments around a candidate point p_0. Right: A graph $\mathbf{G}^w$ for the set of linear segments. We assume an angle tolerance α such that all angle constraints are satisfied. Several node pairs of the graph are not connected by an edge due to the convexity constraint, which is not satisfied for an assumed convexity tolerance t. The red nodes of the graph are the nodes of the optimal maximal clique $\mathbf{G}^c_{opt}$. The corresponding valid configuration $\mathbf{C}_{opt}$ is marked in red on the left figure.

The rectangularity feature scales with the structure size having lower values for small structures. A detector based on such a feature is prone to dismiss small rectangles. On the other hand, false structures of a small size are more frequent. We, therefore, introduce an additional feature f_{S} proportional to the structure size. In the following chapters we will build and explore a detector in the two-dimensional rectangularity-size feature space. Detection in this two-dimensional feature space results in essentially higher performance in comparison to the threshoding of a single f_{R} feature.

We define the size feature as the size of the structure, which is represented by the

[3]Note that not all of the candidate points are the same as in Fig. 6.1. In contrast to Fig. 6.1, the map of candidate points in Fig. 8.4 resulted from the union of points coming from both valley and ridge edges. On the other hand, candidate points that are to too distant or too close to the edges were removed (see Ch. 6) and do not appear in Fig. 8.4.

optimal clique $\mathbf{G}^c_{opt} \subseteq \mathbf{G}^w$, calculated as follows

$$f_{\mathrm{S}}(\mathbf{G}^w) = \frac{\sum_j l_j r_j}{\sum_j l_j}, \tag{8.12}$$

where the sums are over all nodes of the optimal clique $\mathbf{G}^c_{opt}$. f_{S} is computed as the weighted distance of the linear segments of $\mathbf{C}_{opt}$ from the corresponding candidate point, where the weights are segment sizes.

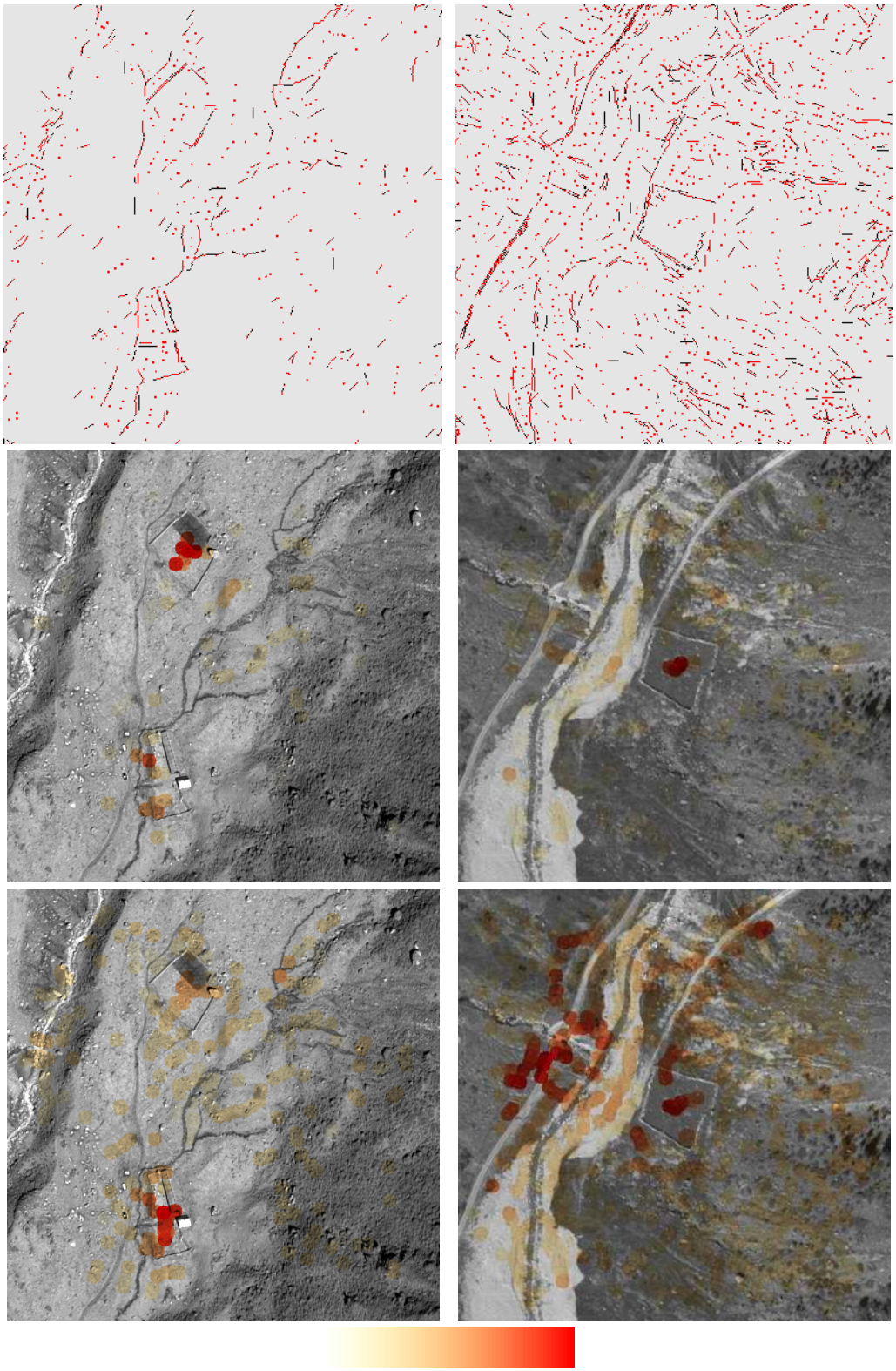


Fig 8.4: First row: Bar edges (black) and candidate points (red) generated from the images in Fig. 1.1. Second row: The rectangularity feature computed at each candidate point and visualized by a colored disk. Third row: The GODF based feature raised to the power of eight (Sec. 10.1). Color saturation increases and hue is changing from yellow to red for growing values of the features in accordance with the color bar in the bottom.

Chapter 9

Classification and detection of LSE

One of the difficulties that arose in our task and may arise in other applications that use methods learning from the data is a very small number of representative examples (see Ch. 3 for details). In this chapter we construct a classifier that can be trained on a small number of representative positives and a large number of readily available negatives. We first explore this classifier in the two-dimensional rectangularity-size feature space. Later, in Ch. 10, we show that this classifier, even though trained on a small number of positives, can perform well in high-dimensional spaces.

Fig. 9.1 on the left shows the frequency of candidate locations as a function of the corresponding rectangularity $f_{\mathrm{R}} \neq 0$ and the size f_S features. The candidate locations were obtained for 17000×11000 pixel satellite image that covers a part of the Silvretta region. The structures are limited in their size by setting two thresholds on the minimal and maximal values of the distance transform as was mentioned in Ch. 6. Real livestock enclosures are very rare in the field and though a few samples of the distribution may correspond to unknown livestock enclosures, the vast majority of the candidates correspond to false structures. We, therefore, will refer to this distribution as distribution of negatives and denote it by X. The scatter plot of the distribution of negatives is shown in cyan on the right of Fig. 9.1. Red samples in this figure are the known livestock enclosures in the Silvretta mountains. Thresholding the feature f_{R} corresponds to a 1D classifier with decision boundary in the feature space $(f_{\mathrm{R}}, f_{\mathrm{S}})$ that is a horizontal line. Learning a general a decision boundary from the available data in the 2D rectangularity-size feature space may improve the trade-off between the sensitivity and the number of false detections in

comparison to one-dimensional case. However, only a few positive examples are available in our case (shown in red in Fig. 9.1 on the right). Therefore, a classification approach should be carefully chosen.

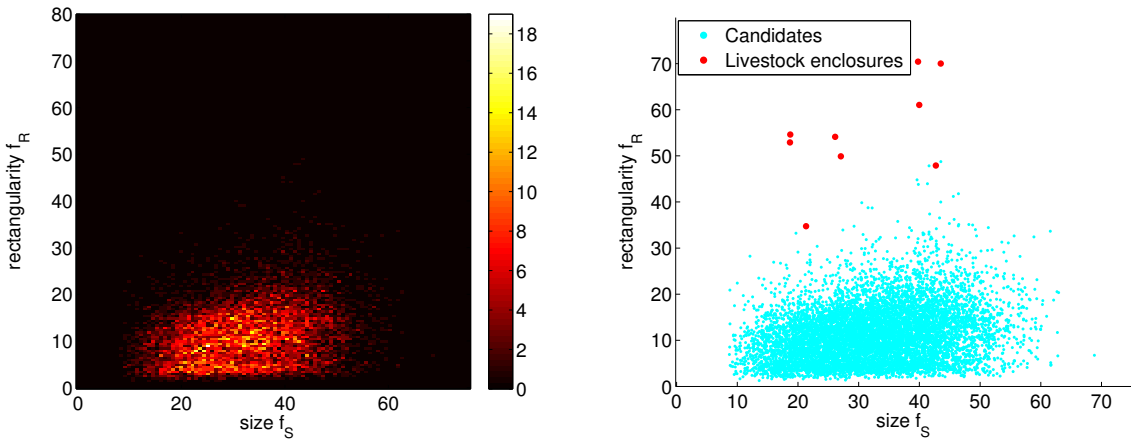


Fig 9.1: Rectangularity-size feature space. Left: Distribution of candidates (with non-zero rectangularity feature) extracted from a satellite image. Right: Scatter plot of the candidates (cyan) corresponding to the distribution on the left and examples of livestock enclosures (red).

The Neyman-Pearson approach, commonly used for detection tasks, is a non-Bayesian decision making that is especially useful when priors are not available or misclassification risks are not comparable [104]. The Neyman-Pearson classification method maximizes the sensitivity of the classifier given an upper bound for the rate of false detections[1]. This strategy is directly applicable to our problem of enclosure detection. It can be interpreted as setting the maximal number of false detections that can be visually verified by an expert, while maximizing the sensitivity of the detection. As for the case of Baysian classifiers, the solution is based on the ratio of class-conditional distributions. Unfortunately, we have a very small number of positive examples, which makes a reliable estimation of the distribution of the target class (positives) impossible.

One-class classifiers are usually employed in situations with available samples from a single class only [105]. Samples from the other class are either not available, difficult to obtain, or very rare. The instances from the second class, which is poorly or not at all represented, are called novelties, outliers, or anomalies. Several approaches were developed in order to approach the one-class classification task. The distribution of the well-described

[1]Alternatively, one may minimize the rate of false detections given a lower bound on the sensitivity.

dominant class can represented by a model of choice. The samples that are very distant from the modeled distribution in accordance with the chosen metric are then assigned to the second class of novelties. Alternatively, the reconstruction error of representing the sample by the chosen model can be used as the measure of novelty [106]. In [107] a support vector data description was developed, where a decision boundary separating the dominant class from novelties is a hypersphere of minimum volume containing samples of the dominant class. An important advantage of this method is its ability to incorporate examples from the class of novelties (if a small number of such examples is available) while learning the decision boundary.

One-class classifiers usually tend to produce a decision boundary that compactly encloses the samples of the well-represented dominant class X. All the other samples are assigned to the second class of novelties. In fact, such classifiers imply a uniform distribution for the class of novelties [108]. In our case and in many other applications, where novelties are positives of the target class that describes a particular category of objects to be detected, the distribution of novelties is far from being uniform. For example, the samples with lower value of f_R and the same value of the f_S have lower probability of being positive examples. Therefore, the one-class classifiers may yield erroneous results, e.g. assigning samples with very low rectangularity values outside of the enclosed distribution dominant distribution to the class of enclosures (positives). A possible solution is to construct a decision boundary that cannot fold up. The simplest choice is a hyperplane, i.e. a linear classification approach.

The linear classifiers are favorable when there is a danger of overfitting the data due to a limited number of available examples. They also are not computationally demanding. Simple linear classifiers may be powerful enough when used together with a few category-specific features as opposed to the use of many generic features [109]. We carefully constructed such rectangularity and size features. The normal w of the separating hyperplane of a linear classifier can be found by means of the Fisher Linear Discriminant analysis (FLD). In this approach, the optimal direction is determined such that the data from two classes projected on w is maximally separated. The separation is measured by the squared distance between class means normalized by the sum of their variances [81, 100]. This approach results in a simple solution represented in terms of class means and covariance matrices. In our case, however, the number of positive examples is very limited and the covariance matrix cannot reliably be estimated.

Below we describe the proposed linear classifier designed such that it does not require estimation of the distribution of positives. We optimize the normal direction w based on the large number of available samples from the dominant class of negatives and just a few examples from the class of positives. Let us define the expected signed distance between a deterministic point y (positive example) and the distribution X of negatives, both projected to the direction w and normalized by the standard deviation of the projected distribution

$$D_w(y, X) \equiv \frac{E_x\left[w^T y - w^T x\right]}{\sqrt{E_x\left[(w^T x - w^T \mu_x)^2\right]}} = \frac{w^T(y - \mu_x)}{\sqrt{w^T C_x w}} \ , \tag{9.1}$$

where μ_x and C_x are the mean and the covariance matrix of the distribution X, respectively. Next, we define the average signed distance between a set of deterministic points $\{y_i, i = 1, ..., n\}$ and the distribution X

$$\bar{D}_w(\{y_i\}, X) \equiv \frac{1}{n}\sum_{i=1}^{n} D_w(y_i, X) \ . \tag{9.2}$$

It follows that due to linearity of $D_w(y, X)$ with respect to y we have

$$\bar{D}_w(\{y_i\}, X) = \frac{w^T(\bar{y} - \mu_x)}{\sqrt{w^T C_x w}} \ , \tag{9.3}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. We now define the optimal direction w as the direction that maximizes the absolute value of the average signed distance between a set of points corresponding to positive examples and the distribution of the dominant class of negatives X, i.e.

$$w_{opt} \equiv \underset{w}{\operatorname{argmax}} \left|\bar{D}_w(\{y_i\}, X)\right| \ . \tag{9.4}$$

From Eqs. (9.3, 9.4) we obtain

$$w_{opt} = \underset{w}{\operatorname{argmax}} \frac{\left|w^T(\bar{y} - \mu_x)\right|}{\sqrt{w^T C_x w}} \ . \tag{9.5}$$

The objective function in Eq. (9.5) is invariant with respect to scale and sign of the vector w. We, therefore, can add constraints $w^T C_x w = 1$ and $w^T(\bar{y} - \mu_x) \geq 0$ that uniquely define the norm and the orientation of w, and solve the constrained optimization problem

$$\tilde{w}_{opt} = \underset{\substack{w^T C_x w = 1 \\ w^T(\bar{y} - \mu_x) \geq 0}}{\operatorname{argmax}} \; w^T(\bar{y} - \mu_x) \ . \tag{9.6}$$

Using Lagrange multipliers it can be shown that maximum is achieved for

$$\tilde{w}_{opt} = \frac{C_x^{-1}(\bar{y} - \mu_x)}{\sqrt{(\bar{y} - \mu_x)^T C_x^{-1}(\bar{y} - \mu_x)}} \ . \tag{9.7}$$

Since, the length of the optimal w is not important we will further use the scaled simpler version of $\tilde{w}_{opt}$

$$w_{opt} = C_x^{-1}(\bar{y} - \mu_x) \tag{9.8}$$

that also solves Eq. (9.5). The obtained direction w_{opt} is similar to the one in the FLD analysis [100]. In contrast to the FLD solution, Eq. (9.8) includes the covariance matrix of the class of negatives only, preferring the solution in the direction of the small variance of negatives. Negatives are well sampled in our problem and their covariance matrix can be robustly estimated. The positives are treated as deterministic points in the feature space and influence the solution only via their average. Literally, the average only weakly guides the solution pointing to the relevant location in the feature space. Note that the signed distance in Eq. (9.3) for w_{opt} given in Eq. (9.8) yields a positive value equal to the Mahalanobis distance $\bar{D}_{w_{opt}}(\{y_i\}, X) = \sqrt{(\bar{y} - \mu_x)^T C_x^{-1}(\bar{y} - \mu_x)}$ with the metric C_x.

The samples of X may include outliers. Therefore, in Eq. (9.8) we use the robust Multivariate Trimming (MVT) estimates of the mean and the covariance matrix [110]. The MVT is an iterative technique with mean and covariance matrices recomputed at each iteration. Given the current estimates of μ_x and C_x the Mahalanobis distance is computed for all the data points. A specified percentage of the observations with the largest Mahalanobis distance is discarded and the remaining data is used to recompute the estimates of μ_x and C_x. The technique is initialized with the sample mean and covariance matrix computed from the whole data. The samples with zero rectangularity f_{R}, which correspond to non-valid configurations, were excluded from such a training procedure. In our experiments we used three iterations and discarded 10% of observations in each iteration.

We will refer to

$$f_{\mathrm{RS}} = \begin{pmatrix} f_{\mathrm{S}} \\ f_{\mathrm{R}} \end{pmatrix}$$

as the rectangularity-size features. Given the optimal direction w_{opt}, the LSE structures are detected by

$$f_{\mathrm{RS}}^T w_{opt} > b, \tag{9.9}$$

where b is a threshold to be set. It determines the tradeoff between the sensitivity and the rate of false detections. The optimal linear feature combination $f_{RS}^T w_{opt}$, which is computed at candidate points, can be seen as a confidence measure of an enclosure structure being present in the area around the candidate point. Fig. 9.2 shows an example of the confidence map of detections obtained for a particular value of b. Detections are marked by a colored square. Color saturation increases and hue is changing from yellow to red for growing values of detection confidence. The 0.5m resolution satellite image of 10220×13350 pixel size covers the area in the Silvretta mountains above Galtür.

Most of the detections are false positives. Note that due to filtering texture areas (Ch. 4) there are no detections in the dense urban area. Among seven detections that are actual enclosures, only a few are of archaeological interest. Two detections in Fig. 9.2 correspond to the livestock enclosures that were shown in Fig. 1.1 on the right and in Fig. 5.3 in the bottom row. In order to quickly inspect and reject numerous false detections we designed an interactive tool described in Ch. 11. The details on the choice of the threshold b, which defines the resulting rate of false positives, and the details on the used training data are given in Sec. 12.1.

Note that learning the optimal feature combination w_{opt} as shown above is not limited to two-dimensional feature spaces, but directly extends to higher dimensions. This will be used in our experiments in Ch. 10 in order to compare the developed features with high-dimensional generic features.

Somewhat similar ideas of using Linear Discriminant Analysis (LDA) adapted to a small number of positives within the context of pedestrian detection appeared in [111]. Relying on the high-dimensional HOG features [101] and LDA, the authors modeled the background class with the mean and covariance matrix learned from unlabeled image patches. Their model trained on a few positives was highly competitive. In contrast to [111], our model was explicitly derived from optimization of Eq. (9.4) that was defined as a way to cope with the settings of the highly unbalanced problem at hand.

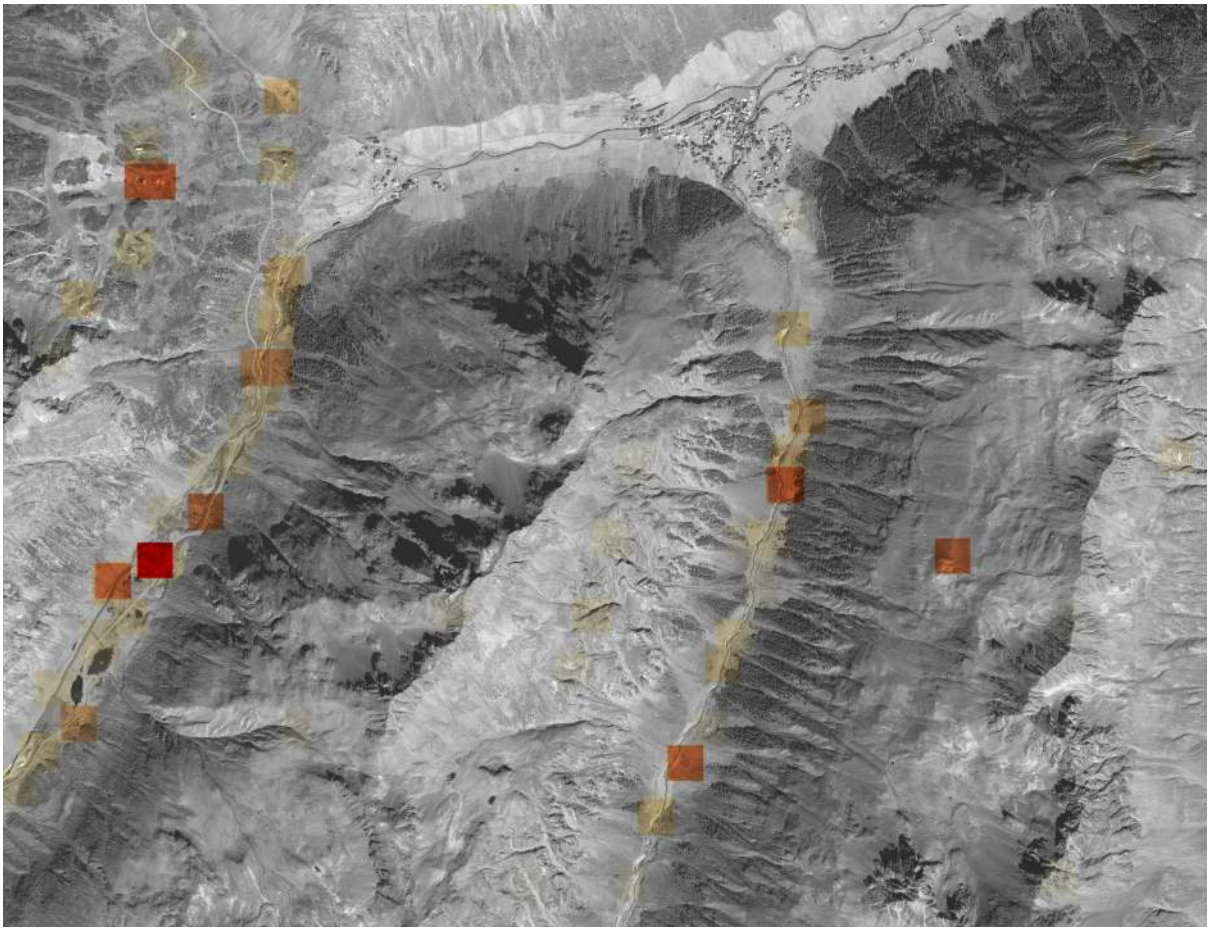


Fig 9.2: Map of detections for the region in the Silvretta mountains above Galtür. Color saturation increases and hue is changing from yellow to red for growing values of detection confidence.

In this section we have developed a detector that can be trained with a small number of positive examples. In Sec. 12.1 we apply this detector to a region in the Silvretta Alp. In section Sec. 12.2 we consider the Bernese Alps case study where no ground truth examples were available. In that case we do not train the detector that generates the rectangularity-size features f_{RS}. Instead, we use the rectangularity feature normalized by the size feature. Note that is possible to train the classier on the available data that came from one region and apply it to another region, provided that the imagery was acquired under similar conditions using the same technology.

Chapter 10

Performance evaluation and feature comparison

In this chapter we evaluate the developed rectangularity f_{R} and rectangularity-size f_{RS} features and provide their comparison with a scalar feature proposed in [43] for building detection and with high-dimensional generic features, namely the histogram of gradients (HOG) features and the features generated by pre-trained deep convolutional neural networks. We will call these features deep CNN features or simply deep features.

10.1 Features for comparison

We first evaluate the discrimination ability of the introduced rectangularity feature f_{R} and provide a comparison with the GODF based feature f_{GODF} recently proposed for building detection in [43]. The GODF, denoted $\lambda(\theta)$, is a weighted gradient orientation histogram with gradient magnitudes as weights and discrete orientation $\theta \in [0, 180)$. The correlation of $\lambda(\theta)$ with a function having two modes separated by 90° served as a GODF based feature f_{GODF} indicating the presence of rectilinear structures. The normalization constant was set such that $\lambda(\theta)$ is a unit vector, which gave us better results than for the normalization constant equal to the sum of the weights used in [43]. More implementation details can be found in [6]. Note that we did not compare the rectangularity feature with the whole approach developed in [43], because it is based on additional features not appropriate

in the case of enclosures. We have also tested other methods for building detection (e.g. [44, 41]) applied to detection of livestock enclosures. Unfortunately, these methods completely failed to detect enclosures. Thus a corresponding quantitative comparison cannot be made.

We use the learning framework developed in Ch. 9 in order to evaluate and compare the introduced rectangularity-size features f_{RS}, high-dimensional HOG feature vectors f_{HOG} [101] and deep convolutional neural networks (CNN) [112, 9] generating so called deep features, denoted f_{CNN} with CNN substituted by the name of a particular CNN architecture. These deep features are neural activations generated by pre-trained CNNs at some intermediate layer of the deep network. Usually these features are extracted either from the last convolutional layer or from one of the following fully connected layers but before the final one. There is mounting evidence that such features generated by CNNs pre-trained on a very large dataset of labeled images have a sufficient representation power to perform recognition tasks on completely different types of target images. Several recent works successfully used deep features in conjunction with either a fully connected neural network [113] or even a simple linear classifier [114, 115, 116] trained on a relatively small target set of images. Moreover, deep features were also shown to be useful for classification of remote sensing images [117]. Such an approach allows us using CNNs even though we have a very limited amount of positive examples to learn from.

We generated deep features using several CNN models pre-trained on the subsets of the ImageNet database [118]. Deep features $f_{\text{Vgg-f}}, f_{\text{Vgg-m}}, f_{\text{Vgg-m-2048}}, f_{\text{Vgg-s}}$[1] were extracted from the CNN architectures described in [119]. $f_{\text{Vgg-deep-16}}$, f_{OverFeat}, and $f_{\text{GoogLeNet}}$ were extracted from the networks described in [120], [121], and [122], respectively. f_{AlexNet} features were extracted from the network described in [123], while f_{CaffeNet} features were generated by an independently trained variation of that network as mentioned in [124], [125]. All deep features, except OverFeat, were computed using the Matlab toolbox MatConvNet [126] using the pre-trained models taken from the webpage [127] accompanying the toolbox. The OverFeat pre-trained model was taken from the webpage [128], which has the source code implementing the deep network presented in [121]. The "fast" network was used. Following recommendations of the authors of the CNN models, except for Over-

[1]We have also experimented with Vgg-m-1024 and Vgg-m-128 CNNs that have a smaller last hidden layer (1024 and 128 versus 4096 neurons in Vgg-m), but they gave worse results compared to Vgg-m. Therefore, we did not consider these features in our comparative experiments.

Feat, we subtracted the mean image of the training dataset from each image presented to the CNNs. Since we detect structures in gray-scale images, while the CNN models require RGB channels as an input, we set each channel equal to the given gray-scale image.

The HOG features were computed for 14×14 pixels cell size (with 7 pixels of overlap from one cell to the next) and 9 orientation bins, which gave us the best results among all configurations with which we experimented. We used the C source code of the HOG implementation available in the VLFeat package [129].

All the features were computed for image regions around the candidate points p. The size of these candidate regions was taken proportionally to the distance transform $D(p)$ as described in Ch. 6. To keep the size of HOG feature vectors constant we resized the candidate regions to 160×160 pixel patches. Deep features were computed for the candidate regions resized to the size required by a particular CNN architecture.

10.2 Measuring discrimination power of the features

Using a particular type of features f, livestock enclosures can be detected with $f > b$ for 1D features ($f_{\mathrm{R}}, f_{\mathrm{GODF}}$) and with $f^T w_{opt} > b$ for multi-dimensional features ($f_{\mathrm{RS}}, f_{\mathrm{HOG}}, f_{\mathrm{CNN}}$), where b is an appropriate threshold to be set. Setting a particular threshold defines the true positive rate (TPR) and the false positive rate (FPR), or correspondingly the number of detected true and false positives (TP and FP). In our case, the effectiveness of the features is their ability to discriminate livestock enclosures from irrelevant structures and clutter. A possible measure of this ability is the minimal number of FP detected with the threshold that insures $TPR \geq \xi$, where ξ is the predefined rate of true positives[2]. We computed FP for $\xi = 1$, denoted in the following by FP_{100}. This was done by setting the detection threshold b to the minimum value of f for 1D features and $f^T w_{opt}$ for multi-dimensional features computed for all positive examples. Obviously, the threshold used to obtain the detection rate $TPR = 1$ on a small number of available examples does not insure a detector with 100% detection rate. However, it allows us to measure and compare the discrimination ability of the features. FP_{100} is related to the spread of the class of positives toward the samples of the class of negatives, similarly to the Fisher criterion of discrimination ability [81]. However, FP_{100} also gives a rough estimate of the

[2]This corresponds to the so-called Neyman-Pearson task [104].

minimal number of false detections per area size that should be allowed in order to have a reasonable detection rate. Unfortunately, the actual detection rate cannot be reliably estimated due to a very small number of positive examples.

We also used an alternative measure of the discrimination ability that is the area under receiver operating characteristic (ROC) curve. It is especially useful in the presence of unbalanced classes [88, 130]. In contrast to FP_{100}, the area under receiver operating characteristic (AUC) does not rely on a particular threshold and a corresponding operating point on the ROC curve, but instead summarizes the detection performance for different values of the threshold. In fact, it is an average of true positive rates estimated for all false positive rates. The AUC has an important statistical property. It equals the probability that a randomly chosen sample y from the population of positives $\mathcal{P}$ has a higher score $f(y)$ (e.g. the rectangularity feature) than the score $f(x)$ for randomly chosen sample x from the population of negatives $\mathcal{N}$, i.e. $AUC(f) = P(f(y \in \mathcal{P}) > f(x \in \mathcal{N}))$. We estimated this probability of a correct ranking by means of the Wilcoxon-Mann-Whitney statistic [131, 130] as

$$\frac{1}{n_P n_N} \sum_{i=1}^{n_P} \sum_{j=1}^{n_N} I(f(y_i \in \mathcal{P}), f(x_j \in \mathcal{N})), \tag{10.1}$$

where n_P and n_N denote the number positive and negative samples, respectively, and $I(u, v)$ is the indicator function defined as

$$I(u,v) = \begin{cases} 1, & u > v \\ 0, & u < v \\ 0.5, & u = v. \end{cases}$$

It should be noted that for the case of normally distributed features $f(x), f(y)$, the AUC has a simple relation to the Fisher criterion [81, 67], which is also frequently used as a separability measure between distributions. Namely, $AUC = \Phi\left(\frac{\mu_p - \mu_n}{\sqrt{\sigma_p^2 + \sigma_n^2}}\right)$, where Φ is the normal cumulative distribution function, evaluated for the Fischer criterion for positive and negative populations with distribution means μ_p, μ_n and standard deviations σ_p, σ_n, respectively.

10.3 Evaluation procedure

In our experiments here we used panchromatic satellite images at 0.5m resolution that cover mountainous regions of the Silvretta Alps. We generated 49584 negative samples for training. The samples were taken around candidate points in a 11000×17000 pixel satellite image. For testing we used 57504 negative samples taken from a different satellite image of 10000×17000 pixel size that covers about 42.5 km^2. Overall only 9 examples of enclosures (positives) taken from aerial and satellite images were available to us. We augmented this data with additional 135 rotated versions of the same enclosures. 16 rotation angles were taken uniformly in the interval $[0, 360)$ degrees. This results in 144 positive examples, which is hardly enough for training and evaluation on separate train and test subsets as we have done with negative samples. In the case of high-dimensional feature vectors, the learned classifier parameters w_{opt} and the estimated performance may largely vary, depending on the selected subset of positives. In order to use most of the positives for training and also make reliable evaluation of the classier performance based on the data not used for training, we perform 9-fold cross validation. Note that we do not have hyper-parameters associated with the classifier that need to be set a priori or optimized on spare data. On each fold we use 16 examples of the same enclosure at different angles for testing and other 128 positives for training.

In the following we report the average value of the performance measures and the standard deviation over the nine folds of cross validation. In addition, we compare the sensitivity of FP_{100} to the reduction in number of positives used for training. To do so we compute the "inverted" 9-folds cross validation where on each fold we use only 16 examples of a single enclosure at different angles for training. For testing we use all 144 positives on each fold. Thus, the results may vary only due to the used training data, since the same data set is used for the performance evaluation. Note that for the case of FP_{100} measure, using all the positives for testing including a single training example can yield only higher (worse) FP_{100}, because the worst positive sample defines FP_{100}.

10.4 Comparative results

The quantitative measures of the discrimination performance of the rectangularity f_{R} and the rectangularity-size features f_{RS} are summarized in Table 10.1. The performance

measures FP_{100} and AUC evaluate the discrimination ability of the features for our task. The FP_{100} (see Sec. 10.2) measures the number of false detections obtained in an area of approximately 42.5km^2, when all available positives are detected. This measure is particulary useful as it helps to decide how many false detections should be allowed in order to have a high detection rate, i.e. the rate that insures detection of at least all available positives examples. The AUC measure yields a performance ranking of different feature types similar to that of FP_{100}. On the other hand, unlike FP_{100}, the absolute values of AUC are quite close to each other giving the wrong impression of similar performance. The high and close values of AUC are due to the ability of the detectors to reject most negatives while detecting a modest number of all available positives. The corresponding ROCs saturate at the maximum detection rate already for small values of false positive rate and differ only for lower false positive rates. Nevertheless, along with the FP_{100}, which is more intuitive and useful for our application, we also provide AUC because it is commonly used for evaluation of detector performance. The last column in Table 10.1 indicates the dimensionality of the features.

Table 10.1 shows that the discrimination ability of the rectangularity-size features f_{RS} is superior to the others. It allowed reduction of false positives by 31% relative to f_{R}. Though effective for building detection, the GODF based feature turned out to be far worse for detecting faint enclosures in cluttered background. This feature is not useful when computed over large windows, where the relative number of points belonging to an enclosure is small. From Fig. 8.4 one can see that the rectangularity feature map (second row) is much sparser than the GODF based feature map (third row). This is partially because the rectangularity feature has zero value for spurious structures with less than three sides, while the GODF based feature may have only small nonzero values for such structures.

To compute the rectangularity-size features f_{RS} we learned w_{opt} from the separate training dataset of 49584 negative examples (see Sec. 10.3). The set of 144 augmented positives used for testing of all the feature types was also used for training the linear classifier. Learning the two-dimensional w_{opt} involves positives only via their average $\bar{y}$ (see Eq. (9.8)) and uses separate large datasets of negatives, therefore it is unlikely that the data is overfitted. Nevertheless, below we carried out another set of experiments, where we avoid the use of the same positives for training and testing by means of cross validation procedure (see Sec. 10.3 for details). In this set of experiments we used cross

Table 10.1: Comparison of discrimination measures for f_{GODF} and the proposed features $f_{\mathrm{R}}, f_{\mathrm{RS}}$.

	FP_{100}	$AUC \times 10^2$	dim.
f_{GODF}	**6862**	99.262	1
f_{R}	**292**	99.967	1
f_{RS}	**201**	99.977	2

Table 10.2: Comparison of discrimination measures for multi-dimensional CNN, HOG, and the rectangularity-size features f_{RS}. 9-fold cross validation was performed on 144 (augmented) positives with either 128 or 16 samples (N_{pos}) used for training. Large separate datasets of negatives were used for training and testing.

feature	$N_{pos} = 128$						$N_{pos} = 16$	dimensionality	output layer
	FP_{100}			$AUC \times 10^2$			FP_{100}		
	mean	std	max	mean	std	min	mean		
f_{RS}	**49.3**	64.3	203	99.976	0.036	99.884	206.3	2	
f_{AlexNet}	**178.6**	373.7	1145	99.943	0.124	99.615	35834.1	4096	fc_7
$f_{\mathrm{Vgg\text{-}f}}$	**195.4**	269.5	847	99.849	0.213	99.358	22699.8	4096	fc_6
$f_{\mathrm{Vgg\text{-}m\text{-}2048}}$	**365.8**	520.6	1688	99.814	0.287	99.091	25895.0	4096	fc_6
$f_{\mathrm{Vgg\text{-}deep\text{-}19}}$	**578.1**	1694.3	5096	99.718	0.826	97.515	32573.9	4096	fc_{17}
f_{CaffeNet}	**609.7**	1153.8	3498	99.771	0.384	99.000	46673.3	4096	fc_6
$f_{\mathrm{Vgg\text{-}m}}$	**631.3**	1039.9	2600	99.752	0.484	98.527	31120.2	4096	fc_6
$f_{\mathrm{GoogLeNet}}$	**1002.4**	2176.5	6549	99.727	0.608	98.161	18603.1	1024	avg pool_{21}
$f_{\mathrm{Vgg\text{-}s}}$	**2911.2**	7097.7	21720	99.353	1.348	95.835	38231.3	4096	fc_6
f_{OverFeat}	**2967.3**	8164.2	24715	99.416	1.554	95.281	37536.7	3072	fc_6
$f_{\mathrm{Vgg\text{-}deep\text{-}16}}$	**3473.0**	10294.5	30925	99.447	1.576	95.244	39626.8	4096	fc_{14}
f_{HOG}	**7472.8**	8352.8	27155	97.926	2.404	92.045	54154.9	4356	
$f_{\mathrm{Rand\text{-}4096}}$	**54039.1**	3124.5	57019.9	49.871	6.694	39.629	57054.1	4096	
$f_{\mathrm{Rand\text{-}2}}$	**54227.0**	2819.5	57087.5	50.736	6.164	41.098	57034.4	2	

validation that allowed us comparison with high-dimensional HOG and deep CNN based features (deep features), which are much harder to keep from overfitting.

Table 10.2 shows the discrimination performance of the rectangularity-size features f_{RS}, high-dimensional histogram of oriented gradients f_{HOG}, and deep features f_{CNN} generated by several pre-trained CNNs. For all CNN architectures, the table gives the layer used to extract features that produced the best result (last column). "fc" adjacent to the

layer number in the table stands for "fully connected". Given a particular set of features, we use the methodology described in Ch. 9 based on training the linear classifier (learning the hyperplane w_{opt}). The table shows mean values, standard deviation, and worse values (max or min) for the discrimination measures FP_{100} and AUC over nine folds of cross validation. On each fold of cross validation 128 positives (augmented from 8 real examples) were used for training and the remaining 16 positives (augmented from the 9th remaining example) for testing (see Sec. 10.3 for details). Mean values of FP_{100} show that the rectangularity-size features f_{RS} outperform all the other features by a large margin. Surprisingly however, two architectures of CNN, AlexNet [123] and Vgg-f [119], provided us with deep features that showed relatively high performance. AlexNet features showed a little lower average number of false positives FP_{100} than Vgg-f features. On the other hand, Vgg-f features showed lower variance of FP_{100} over cross validation folds and therefore might be preferable for our task. The results are remarkable, because the CNNs were trained on a completely different image dataset, while the linear classifier w_{opt} was trained on 128 examples augmented from just 8 real enclosures. Note that no fine-tuning of the pre-trained CNN was performed. These results also indicate that the simple methodology we developed in Ch. 9 for learning from a few positive and a large number of negative examples is useful even for the case of high-dimensional features. Though, as we show below, performance of such features is much more sensitive to the number of positives used for training the linear classifier.

We experimented with CNNs with one or two final fully connected layers or the softmax function of the last layer removed. The table shows the layer that yields the best performing features. For AlexNet the best result was obtained when two final fully connected layers were kept. For all the other CNNs the best results were obtained when features were taken from the first layer that generates data reduced to 1×1 spatial dimension, which is an average pooling for the GoogLeNet and a fully connected layer for all the other CNNs. Our results support the hypothesis that convolutional layers of pre-trained CNNs generate generic features that might be useful for various tasks. In contrast, the final fully connected layers generate task specific features.

Table 10.2 also shows how the performance for all features dropped when only 16 augmented samples were used for training (see Sec. 10.3 for details) within each fold of cross validation. However, in contrast to the other features, the rectangularity-size f_{RS} still yielded relatively high performance, while all the other high-dimensional features

became not useful. This experiment showed high sensitivity of the performance to the number of training examples for the case of high-dimensional features. This also suggests that collecting additional examples might substantially improve their performance.

We also notice that the deeper architectures (Vgg-deep-16, Vgg-deep-19, and GoogLeNet) did not have superior performance for our task. The architecture of CNNs was more important than just their depth, which is in line with a recent observation in [132]. Note that the best performing CNNs AlexNet and Vgg-f have similar architecture [119]. The importance of the particular architecture is also evident from the large variability of the performance of the different CNNs in Table 10.2. Moreover, though CaffeNet and AlexNet are supposed to perform similarly (the first network is a minor variation of the second [125]) they produce substantially different results. The differences in particular training procedures may be responsible for such a discrepancy. The choice of the CNN architecture and training procedure was crucial for our task and seems likely to be critical for other applications. However, it seems that currently there is no established alternative to the trial-and-error based choice of the most suitable architecture for the task at hand.

In Fig. 10.1 we show candidate patches seen by AlexNet and Vgg-f CNNs that generate top responses of the linear classifier $f_{\mathrm{CNN}}^{T} w_{opt}$. The patches were taken out of 57504 negative samples used for testing. These top response patches resemble the structures of interest, indicating that corresponding deep features might be powerful enough to capture the concept of the rectangular enclosures.

For reference purposes, in Table 10.2 we also evaluate performance of random feature vectors f_{Rand} using the same evaluation strategy. The random feature vectors with 2 and 4096 entries of independently identically distributed variables were drawn from the standard normal distribution. As expected, such features give average AUC values close to 0.5. The resulting mean values for false positives FP_{100} are not far from the overall number of samples used for testing (57504).

10.5 Summary

In this chapter, we shown that the rectangularity-size features f_{RS} outperform the rectangularity feature f_{R} in discriminating ruined livestock enclosures from irrelevant structures and clutter. The rectangularity feature f_{R}, in turn, outperforms by a large margin the

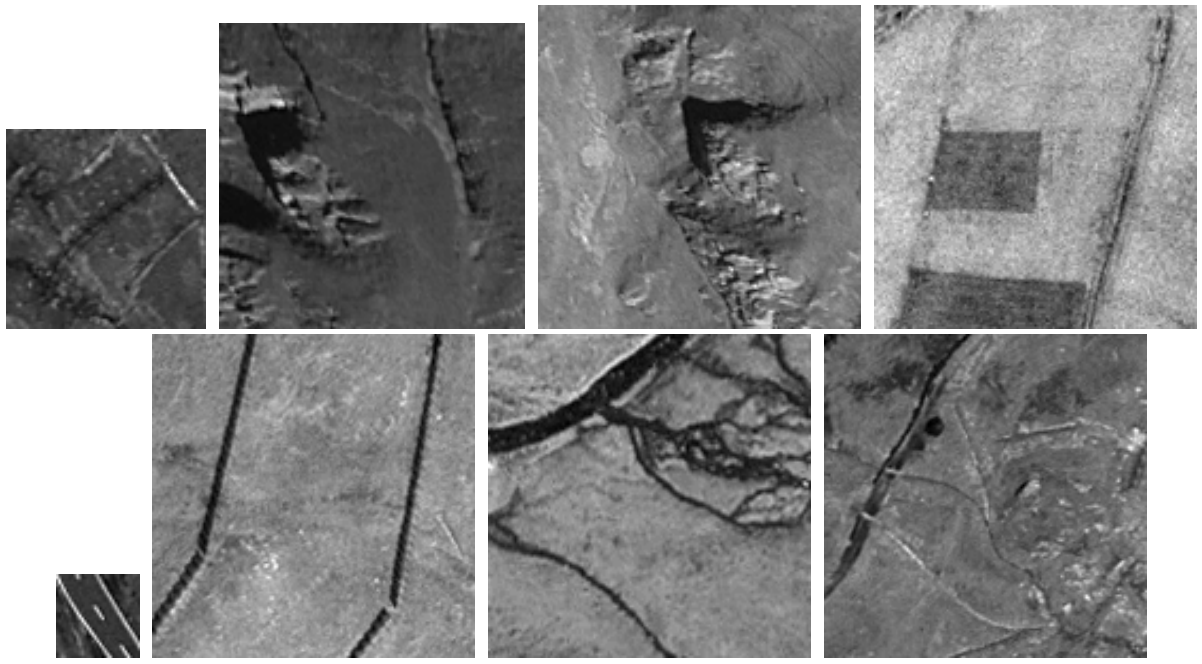


Fig 10.1: Four candidate patches generating highest responses by AlexNet (top row) and Vgg-f (bottom row) architectures of pre-trained CNNs. CNNs were followed by a linear classifier trained on 8 real examples (128 augmented examples) of LSE and large number of negatives. Note that these patches contain structures that are conceptually close to rectangles.

dedicated GODF based feature f_{GODF} that was designed for the building detection. We also shown that the rectangularity-size features are clearly superior to high-dimensional generic features, such as HOG and pre-trained deep CNN features that have recently demonstrated remarkable performance in various tasks. In contrast to the rectangularity-size features, the high dimensional features shown to be very sensitive to the number of available training examples.

Nevertheless, some particular architectures of pre-trained deep CNNs demonstrated promising potential generating features that yield good classification performance already for a small number of positive examples (and a large number of negatives) used for training of our linear classifier. These deep features yielded lower but comparable to the rectangularity-size features performance. We also demonstrated that similarly to the rectangularity-size features the deep features can, to some extent, capture the visual concept of the rectangular enclosure structure.

It should be noted that contrary to the rectangularity-size features, the deep features

do not require a separate stage of extracting bar edges, which may fail in the cases of very low contrasts (e.g. due to low heights of ruined walls). Moreover, given examples of enclosures of non-rectangular shape, we could easily retrain our linear classifier using the same generic deep features. The resulting performance is likely to be improved by learning from more augmented examples using additional transformations, e.g. scalings, flipping, brightness transformations etc. Availability of additional (real) positive examples is certainly critical for improving performance of the deep CNN based detector and may also enable performing fine-tuning of the CNN itself for further gain in performance. The above issues are interesting topics for future research.

Chapter 11

GUI for validation of detections in large areas

In order to be able to quickly reject falsely detected sites we built a prototypical user interface. It allows a user to explore large images, shows detections and their confidence, and allows quickly verifying true detections. Two snapshots of the interface are shown in Fig. 11.1.

Once the detection results have been obtained and the corresponding confidence maps (see an example in Fig. 9.2) generated, one can start examining detections using this GUI. On the left top of the GUI one can see a list of confidence map files in the current directory, which can be chosen in the text window above. The user can click on a file name in order to see the corresponding image in the preview window in the left bottom of the GUI. Clicking on the preview window or double clicking on the name of the file will load an image into the large main window. In this window the user can browse the image with a panning function (left clicking and dragging with the mouse) and zoom in or zoom out with the mouse scroll wheel. Clicking on a particular site of interest moves that site into the center of the main window. The center of the main window is always marked with a small blue marker. The same site is also marked in the preview window, with a blue disk. In addition to zooming in or out, the user can use two buttons in the upper right part of the GUI, which switch between two preset resolutions, namely the full image resolution and the resolution that allows seeing the whole image within the main window. Using the buttons within the "Candidates panel", the user can move through the detections in

order of decreasing confidence. For example, the "Next" button centers the image on the next detected site with lower confidence. The "First" button centers the image around the first detection with the highest confidence. The number above the "First" button shows the sequential number of the detected site and the overall number of detections for the current image. Using the "save/remove" button in the right of the "Candidates panel", the user can save (remove) true detections to (from) the list of findings. When the "Findings" button is pressed on, the user can go through the saved true detections instead of all the original detections. The number above the "save/remove" button indicates the number of saved true detections.

If the user wants to avoid going through the detections in some regions, due to knowing that these regions are of no interest, the "Ignore regions" panel on the left can be used. For example, if there are many detections in the glacier regions, where livestock enclosures are not expected, the user can mark these regions to be ignored. A region of no interest is drawn by first pressing "Draw a polygon" button and then setting the polygon vertices with mouse clicks. The drawn polygon is draggable and resizable. Right clicking on a particular polygon opens a menu that allows the user to delete the polygonal region previously set to be ignored. The "Clear all" button cancels all the defined regions to be ignored. After defining all the regions of no interest, the user should press the "Discard candidates" button. This activates the mode of avoiding the selected regions when the user browses detections with the "Next" and "Previous" buttons. The selected regions of no interest can be saved with the "Save mask" button, to be used later.

To make browsing all the detections smoother, close detections (closer than 30 pixels) are grouped together, so that the user attends to detections that are close to each other only once, using the "Next" and "Previous" buttons. The detections are ordered in decreasing order of their confidence, so that the first shown detection is the detection with the highest confidence. Sometimes this may result in unpleasantly abrupt moves between close detections with different levels of confidence and other distant locations. To prevent such behavior, close but not clustered detections (detections within a distance between 30 and 120 pixels) are shown after each other irrespectively of their confidence.

The GUI tool allows fast browsing of original detections at an average speed of about 1.5 seconds per detection in our own experience. This depends on various factors, such as the image content, the number of images that cover the analysed area (loading the next image takes some time), etc.

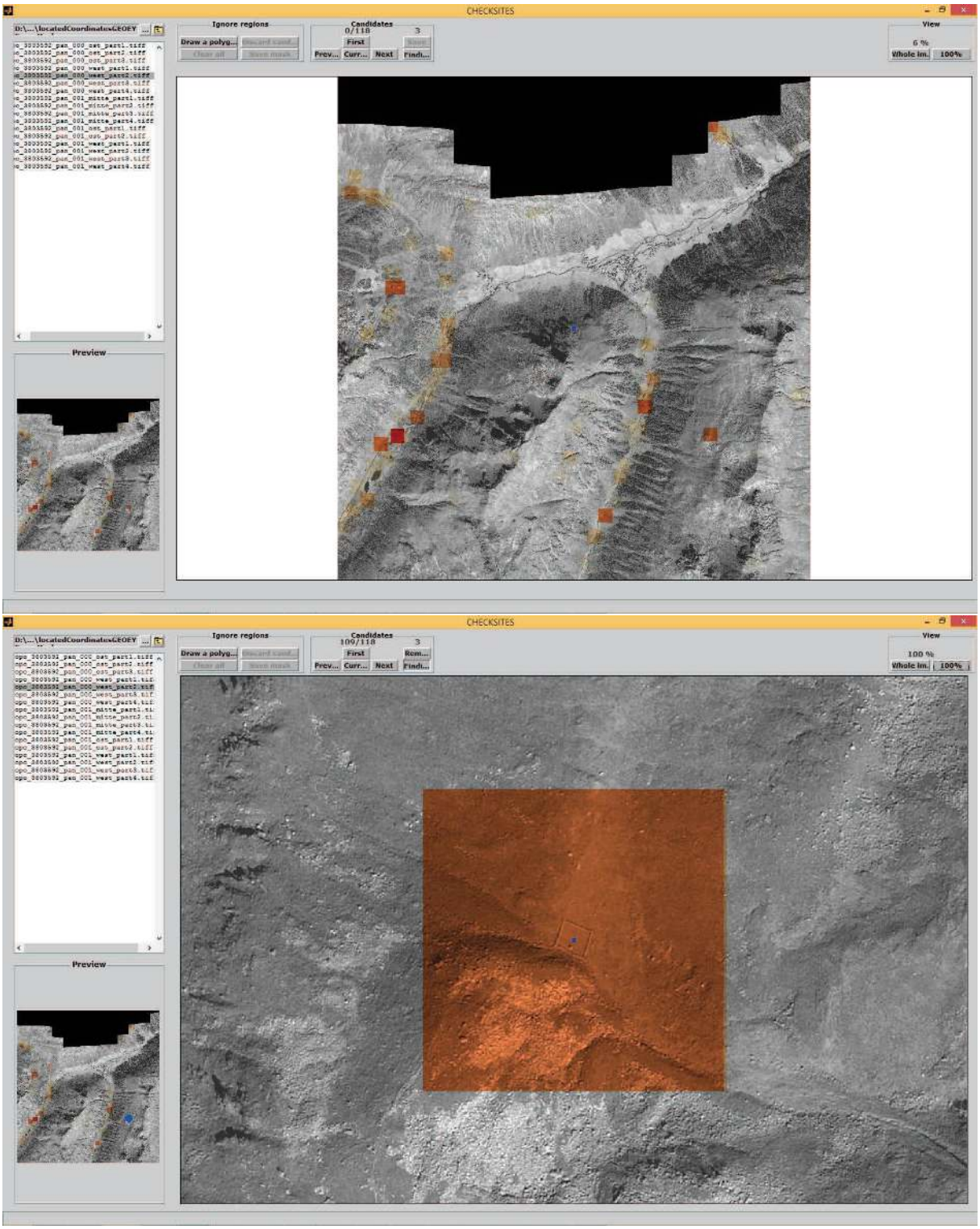


Fig 11.1: GUI that allows fast visual validation of structures of interest and rejection of false detections. Above: The GUI after loading a confidence map of detections. Among 118 original detections there are three LSE structures of archaeological interest and seven other enclosures. Below: The GUI after centering the image at one of the detected structures of interest. The image is zoomed in to the full resolution. The large blue circle in the bottom-left preview image shows the position of the detection.

Chapter 12

Application to detection of LSE: Results

12.1 Case study: The Silvretta Alps

In Ch. 10 we estimated the performance of our LSE detector in terms of FP_{100} and AUC using the satellite image that covers 42.5 km^2 and a small number of positives available to us. In this chapter we illustrate the use of the LSE detector in practice by reporting the results of applying the detector to a region in the Silvretta Alps [11] with a size of about 550 km^2. First, we used panchromatic images at 0.5 m resolution captured by the GeoEye1 satellite. We then repeated our experiments using the red channel of SWISSTOPO aerial images also of 0.5m resolution that covered a slightly larger area of the same region. The size parameters were set so that automatic detection would target structures whose distance between opposite walls varies from 7.5 m to 45 m (see Ch. 6 for details).

As mentioned in Ch. 4, a large number of false candidates can be generated in textured regions (e.g. urban areas or forests). Since we are only interested in livestock enclosures, which sparsely appear in grassland areas, high contrast texture regions need to be segmented out. For this purpose, we used the illumination invariant Morphological Texture Contrast (MTC) descriptor (Ch. 4), which allows the preservation of individual structures. The size parameters r_1 and r_2 in the MTC were set to 30 and 60 pixels,

respectively. The required texture masks were generated using the MTC operator with the Otsu thresholding method [82] for each of the 17 satellite images and for each of the 88 aerial images that cover the analyzed area of the Silvretta Alps. Using the generated texture masks, urban areas, forests, rocky mountains, and other high contrast texture regions were filtered out.

To learn the optimal w_{opt} of the linear classifier, we used nine available examples of livestock enclosures. In the experiment with the satellite imagery, we used 49584 negative examples[1] of structures around candidate points extracted from an 11000×17000 pixel satellite training image. In the experiment with the aerial imagery, we used 72245 negatives examples extracted from an 8750×6000 pixel aerial training image.

The threshold parameter b for the LSE detector in Eq. (9.9) can be set based on the number of allowed false detections. For example, based on Tables (10.1, 10.2), for the satellite image covering an area of 42.5 km^2 that was used in our comparative experiments in Sec. 10.4, the number of allowed false detections should be at least ~ 200. Table 10.1 shows that the detector generates 201 false detections when the detector's threshold b was set to the level that ensures the detection of all the available positives including augmented ones. In the corresponding experiment, all the available positives were used for testing and for training. In the experiment summarized in Table 10.2, where the data used for testing was excluded from training, the maximal number of 203 false detections was obtained in one of the folds of the cross validation. 200 false detections in 42.5 km^2 corresponds to $550/42.5 \times 200 = 2588$ false detections in an area of 550 km^2, which is covered by the images of the Silvretta mountains. In our experiments, we here set the detection threshold b at a lower level, in such a way that the number of generated detections was 5000, which made the detector almost two times as sensitive.

After visual inspection of the detected sites with the user interface, described in Ch. 11, we found 31 structures resembling livestock enclosures. Locations of these findings are shown as red points in Fig. 12.1. Corresponding geographical coordinates and small images of adjacent areas are available in digital form as supplementary material. Several of these detections were found to be livestock enclosures that were hitherto unknown. The corresponding examples are shown in Fig. 12.2. Several detected structures were recognized by our colleagues, archaeologists, as known structures. Some findings are not clear and need further verification in the field. An example of such an enclosure is shown

[1]The same training image was used in Ch. 10 to generate negative examples.

in Fig. 12.2 in the rightmost bottom picture.

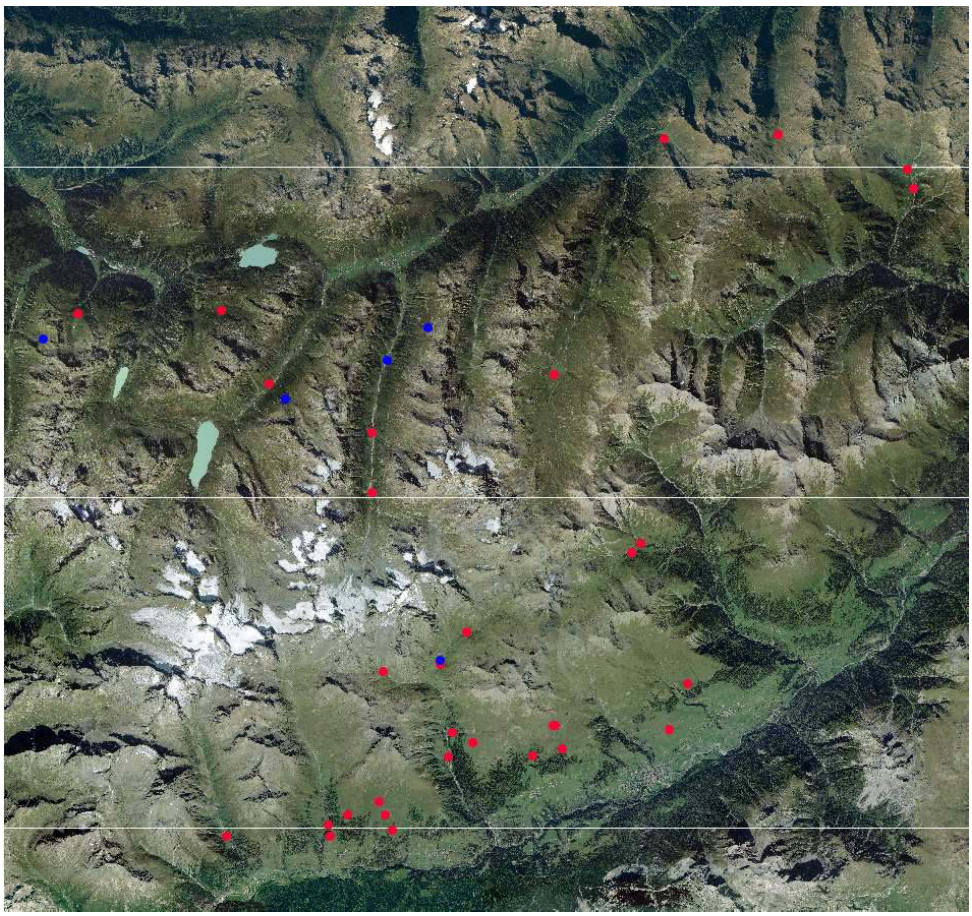


Fig 12.1: Detections in the Silvretta mountains. Red points correspond to the detected structures. Blue points correspond to the LSEs used for training.

The areas covered by the satellite images do not exactly coincide with those covered by the aerial images. Partly due to this fact, some LSEs were detected only in the aerial images, and some only in the satellite images. In general, the aerial images that we used were of higher contrast than the satellite images, resulting in a larger number of detected candidates.

Fig. 12.3 shows some typical false detections. False detections are usually caused by streams, trails, roads and modern man-made structures. The use of 3D data (e.g. LiDAR or stereo image pairs) would allow us to avoid some of the false detections. Spectral features might also be useful for discarding false detections caused by streams. The use

of geo-referenced digital elevation models may help discard false detections located on steep slopes, where LSE usually are not built. One could also use elevation models to define the relevant heights of the terrain. Improvements such as these, based on the use of additional sources of data, are outside the scope of this thesis. However, they may comprise the future continuation of this research.

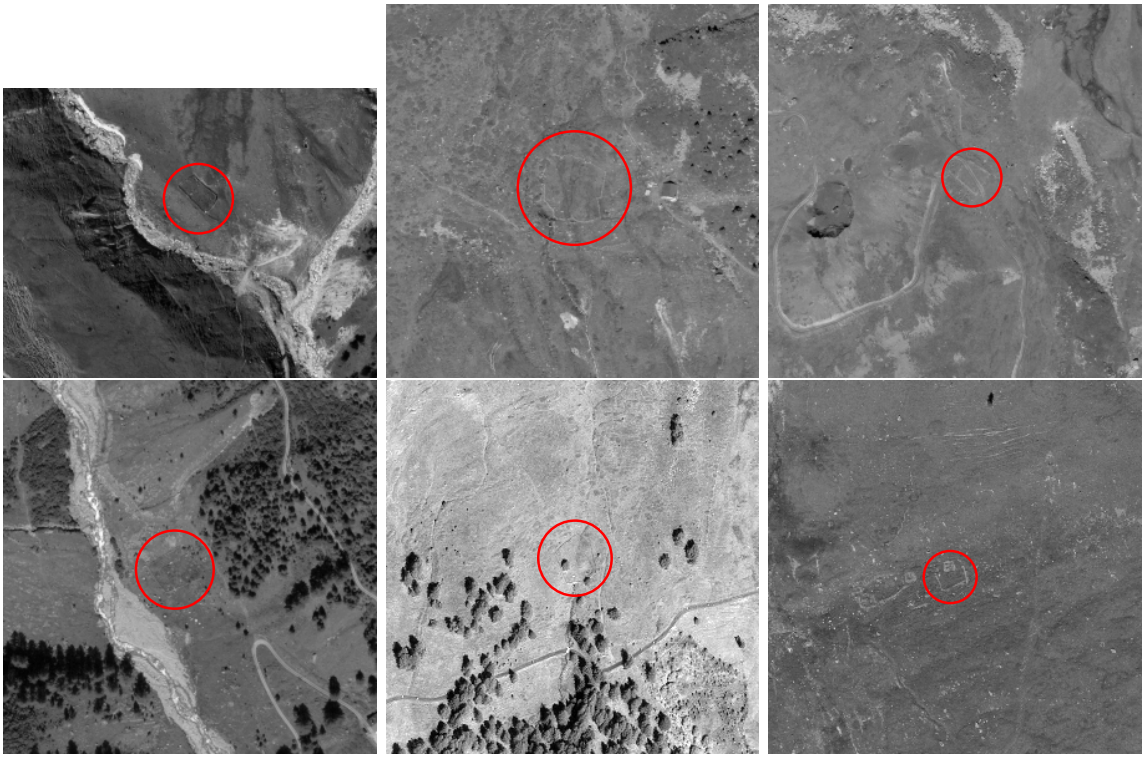


Fig 12.2: Previously unknown enclosures that were detected using the methodology developed in this thesis. Best viewed in digital version.

In our experiments, we used Matlab software and a PC equipped with an Intel Core i5 3.3 GHz Quad-Core processor and 32GB RAM. Processing 550km^2 area of the Silvretta Alps (including all the stages), which is equivalent to processing a single 53000×53000 pixel image, took three hours and forty minutes for the case of the satellite imagery and about 13 hours for the case of the aerial imagery. Such a difference in running times is due to the higher contrast in the aerial imagery, which resulted in a larger number of initial candidates to be processed.

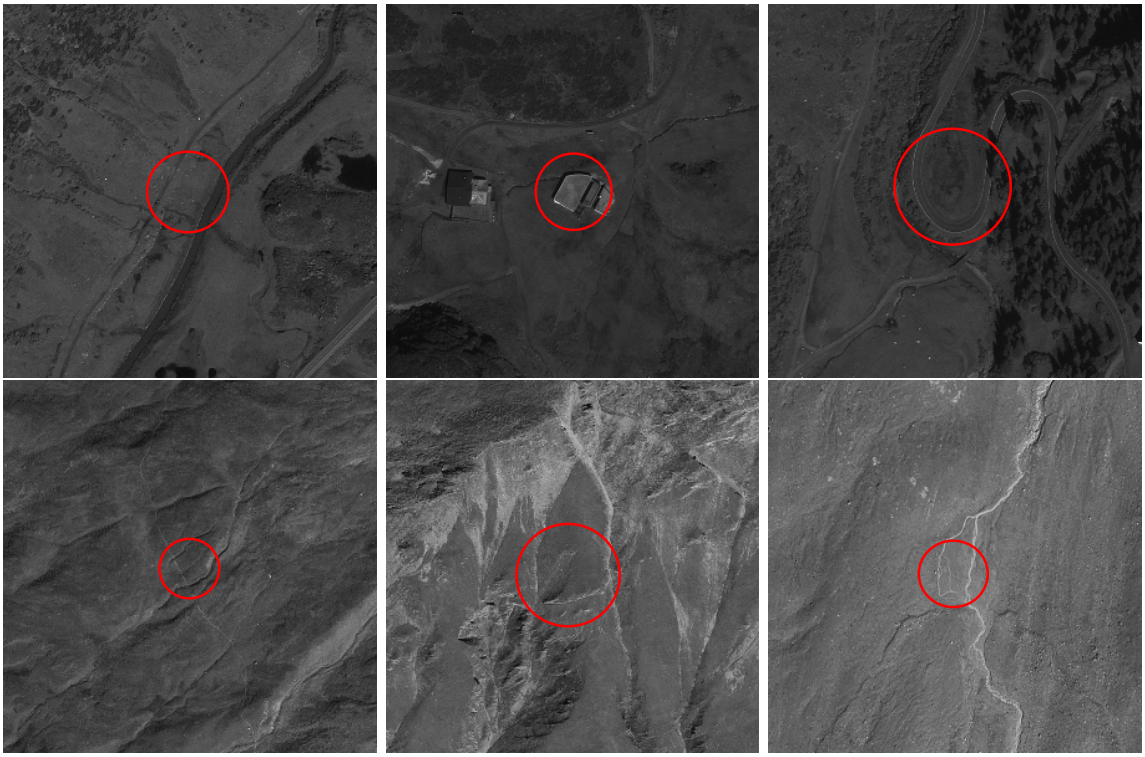


Fig 12.3: Typical false detections. Best viewed in digital version.

12.2 Case study: Bernese Alps

Although we developed and tested (Ch. 10, Sec. 12.1) our approach using the 0.5m/pixel resolution imagery, here we report supplementary experiments using imagery of a higher resolution. We applied our algorithms to 0.25m/pixel aerial imagery that covers about 850 km^2 of the Bernese Alps (see more details on the source of the data in Ch. 3). We use exactly the same set of parameters as for the case of the Silvretta Alps (Sec. 12.1). Due to the difference in resolution, the use of the same parameters implies that the structures being detected are half the size of those in the experiments with the images of the Silvretta Alps. Thus, we target structures with an approximate distance between the walls varying from 3.75m to 22.5m. In contrast to the Silvretta region, for the Bernese Alps we do not have examples of rectangular enclosures of interest. We, therefore, cannot train the rectangularity-size features f_{RS}. Instead, we use the rectangularity feature normalized by the size feature, i.e. we use an $f_{\mathrm{R}}/f_{\mathrm{S}}$ descriptor. Since the rectangularity feature scales with the size of the structure, such normalization reduces the dependence of the

rectangularity feature on the size of the candidate structure. Similarly to the experiments with images from the Silvretta mountains, we allowed the detection of 5000 structures. Using our GUI for fast validation of the detections we found 19 interesting structures resembling livestock or garden enclosures and hut remains. Locations of these findings are shown as red points in Fig. 12.4. Corresponding geographical coordinates and small images of adjacent areas are available in the digital form as a supplementary material. Some of the structures were already known to our archaeological partners from Canton Bern in Switzerland. A few examples of these findings are shown in Fig. 12.5.

Fig 12.4: Detections in the Bernese mountains.

In this section we applied our approach without any kind of adaptation to the imagery with a different resolution obtained from a different source and acquired for a different region. However, we were still able to detect a plurality of the structures of interest. These results suggest that our approach based on the rectangularity feature is robust to the discrepancy between different data sources and correspondingly to the differences in illumination, geometric distortions, resolutions, and other factors. Such a robustness is crucial in practice.

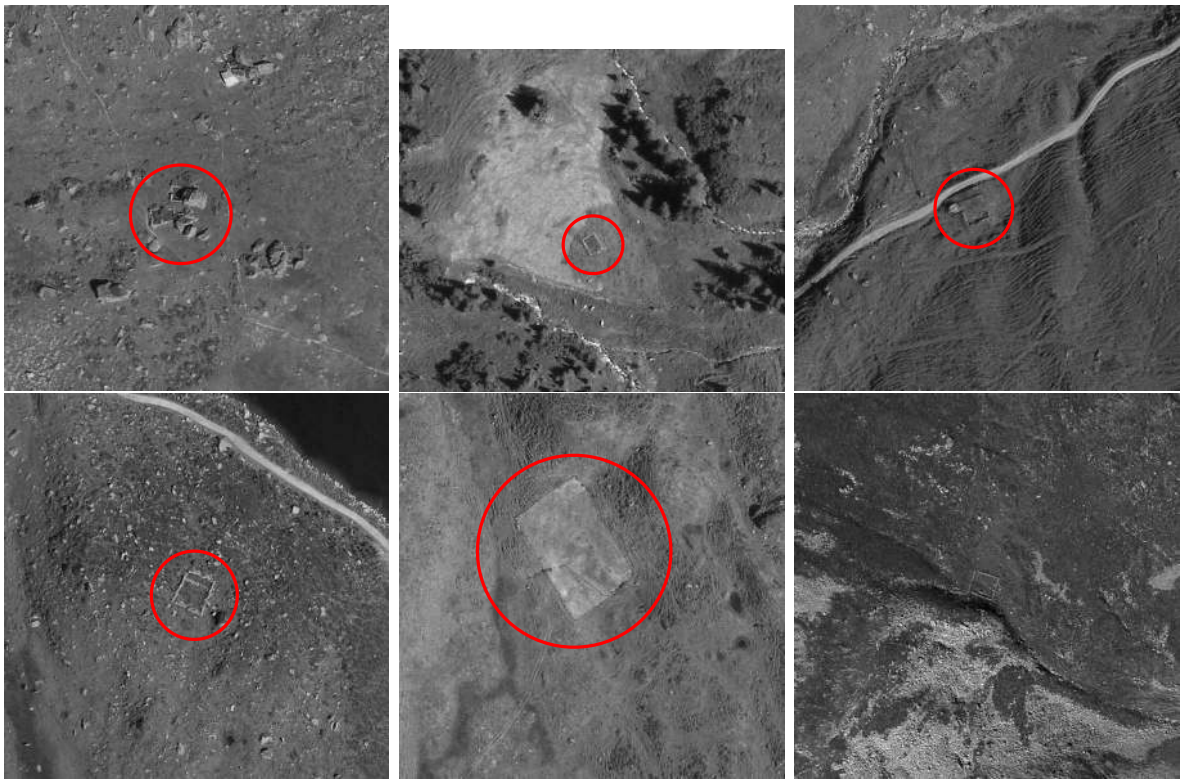


Fig 12.5: Examples of enclosures and hut remains detected in the region of the Bernese Alps using images of 0.25m/pixel resolution and the same methodology as for the case of the 0.5m/pixel images of the Silvretta Alps. No representative examples of rectangular structures taken from the images of the Bernese Alps were available to us. Therefore, the detection was based on the $f_{\mathrm{R}}/f_{\mathrm{S}}$ descriptor instead of the rectangularity-size features f_{RS} (which require at least a few positive examples for training). Best viewed in digital version.

Chapter 13

Application to detection of buildings

Though the rectangularity feature was developed for the particular task of detecting ruined livestock enclosures, it can also be used for other tasks. Here we briefly illustrate its application to building detection. Since the MFC based line detector extracts bar edges only, we replaced it with the line segment detector of [94], which also extracts step edges that are more appropriate for the detection of buildings. In the case of strong object contrasts (unlike the case of livestock enclosures) it reliably detects object borders and yields a relatively small number of edges caused by clutter.

Once the edges were extracted, we used the same workflow as for the LSEs in order to generate the rectangularity feature f_{R}. Since we do not have a labeled training dataset for buildings, we did not experiment with the rectangularity-size features f_{RS}. We also do not expect that it can be essentially better than f_{R}, because the variability of buildings sizes is much smaller than for the case of the LSEs. Nevertheless, we still took into account the dependence of the f_{R} on the size of the structure by normalizing the rectangularity feature as $f_{\mathrm{R}}/f_{\mathrm{S}}$. Fig. 13.1 illustrates $\ln(f_{\mathrm{R}}/f_{\mathrm{S}})$ computed for a SWISSTOPO 4000×4000 aerial image of 0.25m resolution taken over the Bernese Alps. The logarithm was taken in order to make weak detections more visible. We used the same parameters as before for the 0.5m resolution images, except for the maximal size of a building structure. The maximal size was reduced to 65 pixels, while the minimal size was kept at 15 pixels. One can see that most of the buildings were detected. On the other hand, there are false detections mostly caused by occasional configurations of forest and field edges or roads. In urban areas our detector can produce many false detections in between adjacent buildings or other man-

made structures. Therefore, in its original form, the detector might be more appropriate for rural or mountainous areas when high sensitivity is needed for the detection of possibly occluded individual rare structures. For better performance it can be adapted to detect buildings (instead of LSEs) by, for example, incorporating region and/or corner cues. Such adaptations, however, are outside the scope of this thesis, as is the quantitative evaluation of the performance for building detection tasks.

Note that while the building detection problem was treated using the enclosure detector, the detection of enclosures cannot be treated using building detectors. In general, methods for building detection are not suitable for our case because of the considerably lower heights (resulting in low feature contrasts) and feature sizes (ruined walls versus building rooftops) and due to the absence of various cues (roof colors, roof homogeneity, shadows, 3-D cues, etc.). Some walls or parts of walls may be missing or may also be missed in the edge extraction (the width of the linear features does not exceed two pixels in images of 0.5 m resolution). Various irrelevant structures (trails, streams, rocks, etc.) with sizes or/and reflectance properties similar to those of enclosure walls may occasionally form rectilinear configurations. In contrast to enclosures, building rooftops are much more distinctive structures. As an example, we have shown in Sec. 10.4 that the GODF-based feature used for the detection of buildings reveals a poor discrimination ability for the LSE detection task.

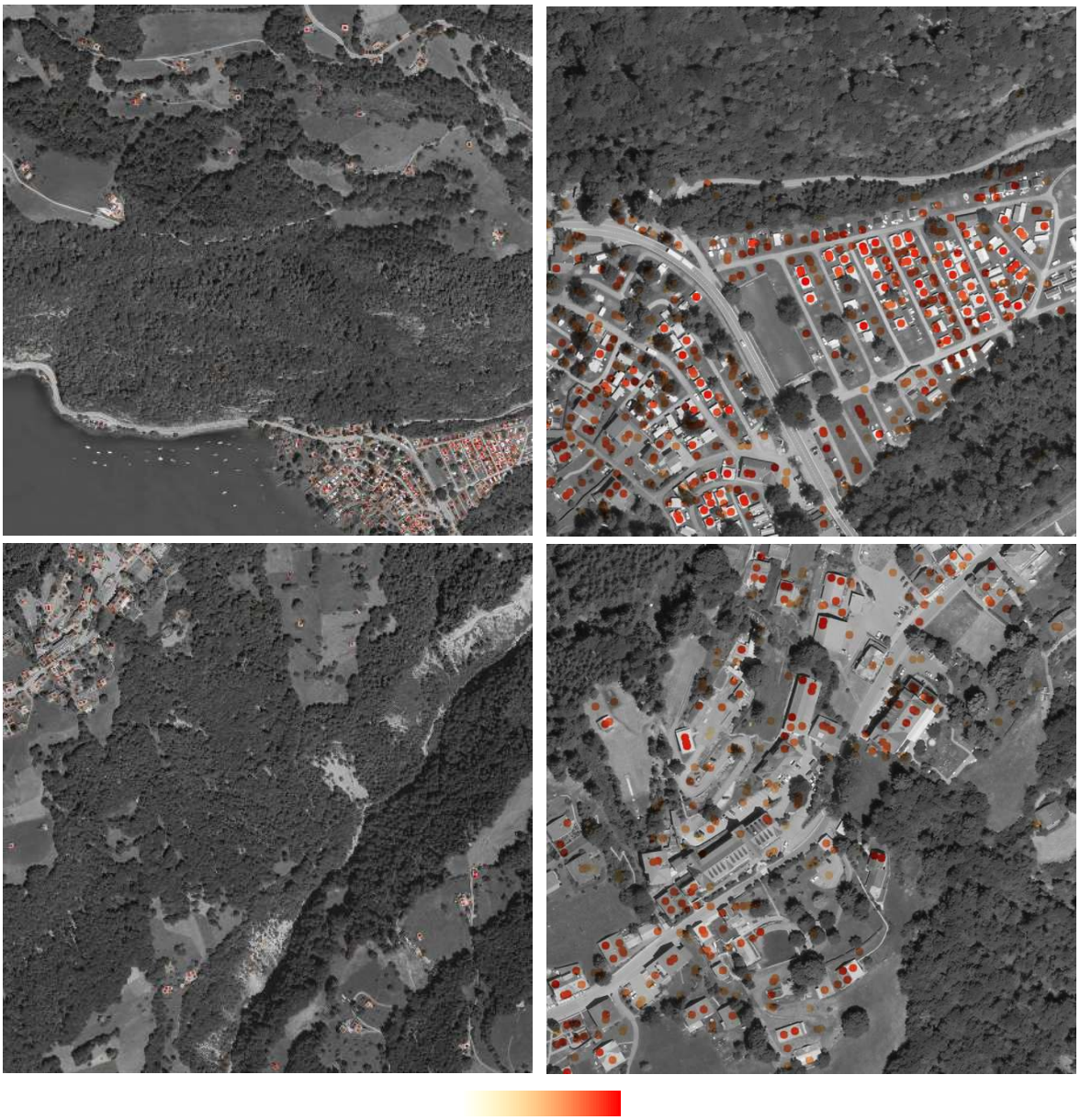


Fig 13.1: Left: Building detections in 4000 × 4000 aerial (SWISSTOPO) images of 0.25m resolution visualized by colored disks. Right: Enlarged parts of the dense urban areas of the images on the left. Color saturation increases and hue changes from yellow to red for increasing values of $\ln(f_R/f_S)$ in accordance with the color bar at the bottom.

Chapter 14

Conclusions

We developed a complete methodology for the semi-automated detection of ruins of livestock enclosures (LSE) in mountainous regions. We have verified the feasibility of this methodology by detecting previously unknown enclosures in two large study areas. This semi-automated approach consists of a sequence of dedicated algorithms for the automated scanning of large areas and the detection of plausible candidate locations. These candidates are further visually inspected by an expert using a specially developed graphical user interface (GUI). This GUI allows a convenient and quick validation of the automatically detected structures. Our methodology can effectively guide archaeological prospection that aims at detecting new sites with remains of LSEs. Such a guidance is especially valuable in large unexplored areas with little or no archaeological record.

In this thesis we contributed new machine vision algorithms that are especially suitable for the detection of LSEs in alpine areas. We have also shown that some of the developed algorithms can be used for the conceptually similar task of detecting buildings in rural or mountainous areas. On the other hand, we explained in Ch. 13 that the methods used for the detection of buildings are not suitable for our task. The algorithms introduced in this thesis may also be used in various computer vision applications that are very different from our task. Below we very briefly outline the nature of these novel algorithms. A more comprehensive overview can found in Sec. 1.3 and Ch. 2.

We proposed an effective method for the generation of proposals for object detection tasks and a simple classifier that is not prone to overfitting, even when trained with only a few positive examples and a sufficient number of negatives. Our novel algorithms for

texture detection, the extraction of ridge and valley features in a complex background, and the generation of handcrafted rectangularity-size features were all separately evaluated and compared with alternative techniques. In this comparison we pointed out properties important for our task that are not achieved by the other techniques. For the evaluation of the different features that feed a detector, we proposed a technique suitable for the case with a very small number of positive examples, when the sensitivity of the detector cannot be reliably estimated. The evaluation has shown that our detector, which is based on the rectangularity-size features, clearly outperforms alternative features for our task. The closest performance was obtained with the state-of-the-art deep features generated by a particular architecture of pre-trained convolutional neural networks. This might be a interesting direction for future research aiming at further improvement of the detection performance.

14.1 Future work

We have found that the linear classifier that is fed with generic features generated by several pre-trained deep CNN architectures results in a well-performing detector. Although these generic deep CNN-based features did not perform as well as the introduced rectangularity-size features, they may be particularly useful for detection of LSEs of very low contrast. Unlike the rectangularity-size features, they do not require a separate stage of extracting bar edges, which may fail in cases of very low contrast (e.g., due to the low heights of ruined walls). Additionally, deep features based detector is not limited to the detection of enclosures of an approximately rectangular shape. A linear classifier can be easily retrained using examples of enclosures of non-rectangular shape, if they are available.

We also expect that the reported performance of deep CNN-based features is likely to be improved by learning from more augmented examples using various transformations in addition to the rotations. These can be e.g. flipping, brightness transformations, and scaling. Using patches at several coarser scales would result in an analysis of broader areas around candidate structures, which would add contextual information to the treatment.

The availability of additional (real) positive examples is certainly critical for improving the performance of a deep-CNN-based detector and may also enable fine-tuning the CNN

itself for further improvements in performance. In fact, the newly detected LSEs in Sec. 12.1 together with their augmented instances could already be used to improve the training of the detector.

There are several promising directions that could further improve the performance of the overall system using additional complementary sources of the geo-referenced data. For example, using airborne laser scanning data (ALS) or stereo pairs would allow distinguishing the features in images caused by spectral contrasts (streams) and by land relief (stones, architectural remains made of stones). We used only one global filter, which filtered out large areas of high texture contrast, such as urban or forest areas. Other global filters can be designed. For example, large areas covered by glaciers can be segmented out by using spectral information. This would be a useful preprocessing since occasional groups of numerous cracks in glaciers may result in falsely detected structures. Provided with digital elevation models, one can also filter out areas with slopes that are too steep, where LSEs are not likely to exist.

It should be noted that the detection of other architectural remains of an approximately rectangular shape, such as the remains of huts, can also be addressed using a suitable adaptation of the methods developed in this thesis. In fact, in Sec. 12.2, we have shown that applying our methodology to images of higher resolution results in the detection of the remains of some types of huts or buildings. Lastly, the developed algorithms can be embedded into the Geospatial Information System (GIS). This would provide geo-referencing tools that associate image coordinates with geographical location, allow a combining of different sources of the data, and offer an interface archaeologists are familiar with.

Bibliography

[1] Igor Zingman, Dietmar Saupe, and Karsten Lambers. Morphological operators for segmentation of high contrast textured regions in remotely sensed imagery. In *Proc. of the IEEE Int. Geoscience and Remote Sensing Symposium*, pages 3451–3454, Munich, Germany, July 2012.

[2] Karsten Lambers and Igor Zingman. Towards detection of archaeological objects in high-resolution remotely sensed images: the Silvretta case study. In Graeme Earl et al., editors, *Archaeology in the Digital Era, vol. II (e-papers) from the 40th Conf. on Computer Applications and Quantitative Methods in Archaeology, Southampton, March 2012*, pages 781–791. Amsterdam University Press, 2013.

[3] Igor Zingman, Dietmar Saupe, and Karsten Lambers. Detection of texture and isolated features using alternating morphological filters. In *Proc. of the International Symposium on Mathematical Morphology and its applications to image and signal processing (ISMM)*, pages 440–451, Uppsala, Sweden, May 2013.

[4] Igor Zingman, Dietmar Saupe, and Karsten Lambers. Automated search for livestock enclosures of rectangular shape in remotely sensed imagery. In Lorenzo Bruzzone, editor, *Proc. SPIE, Image and Signal Processing for Remote Sensing XIX*, volume 8892, pages 88920F–1 – 88920F–11, Dresden, Germany, 2013.

[5] Igor Zingman, Dietmar Saupe, and Karsten Lambers. A morphological approach for distinguishing texture and individual features in images. *Pattern Recognition Letters*, 47:129 – 138, 2014. Advances in Mathematical Morphology.

[6] Igor Zingman, Dietmar Saupe, and Karsten Lambers. Detection of incomplete enclosures of rectangular shape in remotely sensed images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2015.

[7] I. Zingman, D. Saupe, O. A. B. Penatti, and K. Lambers. Detection of fragmented rectangular enclosures in very high resolution remote sensing images. *IEEE Trans. on Geoscience and Remote Sensing*, 54(8):4580–4593, 2016.

[8] Richard Szeliski. *Computer vision: algorithms and applications*. Springer, 2010.

[9] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[10] David C. Cowley. In with the new, out with the old? Auto-extraction for remote sensing archaeology. In *Proceedings of SPIE 8532. Remote sensing of the ocean, sea ice, coastal waters, and large water regions*, 2012.

[11] Karsten Lambers and Thomas Reitmaier. Silvretta Historica: satellite-assisted archaeological survey in an Alpine environment. In F. Contreras, M. Farjas, and F. J. Melero, editors, *CAA 2010 Fusion of cultures: Proceedings of the 38th annual conference on Computer Applications and Quantitative Methods in Archaeology, Granada, Spain, April 2010*. Oxford: Archaeopress, 2013.

[12] Christoph Walser and Karsten Lambers. Human activity in the silvretta massif and climatic developments throughout the holocene. *eTopoi. Journal for Ancient Studies*, pages 55–62, 2012.

[13] Karsten Lambers and Igor Zingman. Am Boden, aus der Luft, aus dem All: Prospektion archäologischer Fundstellen in der Silvretta. In Thomas Reitmaier, editor, *Letzte Jäger, erste Hirten: Hochalpine Archäologie in der* S*ilvretta*, pages 71–85. Chur: Archäologischer Dienst Graubünden, 2012.

[14] Thomas Reitmaier. *Letzte Jäger, erste Hirten: Hochalpine Archäologie in der* S*ilvretta*. Chur: Archäologischer Dienst Graubünden, 2012.

[15] Michael Kasper, editor. *Silvretta historica : Zeitreise durch die Silvretta*. Heimatschutzverein Montafon, Schruns, Austria, 2013.

[16] Katja Kothieringer, Christoph Walser, Benjamin Dietre, Thomas Reitmaier, Jean Nicolas Haas, and Karsten Lambers. High impact: early pastoralism and environmental change during the Neolithic and Bronze Age in the Silvretta Alps

(switzerland/austria) as evidenced by archaeological, palaeoecological and pedological proxies. *Zeitschrift für Geomorphologie, Supplementary Issues*, 59(2):177–198, 2015.

[17] Francesco Carrer, André Carlo Colonese, Alexandre Lucquin, Eduardo Petersen Guedes, Anu Thompson, Kevin Walsh, Thomas Reitmaier, and Oliver E Craig. Chemical analysis of pottery demonstrates prehistoric origin for high-altitude alpine dairying. *Plos one*, 11(4), 2016.

[18] Øivind Due Trier, Siri Øyen Larsen, and Rune Solberg. Automatic detection of circular structures in high-resolution satellite images of agricultural land. *Archaeological Prospection*, 16:1–15, 2009.

[19] Jesse Casana. Regional-scale archaeological remote sensing in the age of big data. *Advances in Archaeological Practice*, 2(3):222–233, 2014.

[20] M. Jahjah and C. Ulivieri. Automatic archaeological feature extraction from satellite VHR images. *Acta Astronautica*, 66:1302–1310, 2010.

[21] Pierre Soille. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag Berlin, 2nd edition, 2003.

[22] Pierre Soille and Martino Pesaresi. Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 40:2042–2055, September 2002.

[23] Tiziana D'Orazio, Filippo Palumbo, and Cataldo Guaragnella. Archaeological trace extraction by a local directional active contour approach. *Pattern Recognition*, 45(9):3427–3438, 2012.

[24] Benedetto Figorito and Eufemia Tarantino. Semi-automatic detection of linear archaeological traces from orthorectified aerial images. *International Journal of Applied Earth Observation and Geoinformation*, 26:458–463, 2014.

[25] Tony F Chan and Luminita A Vese. Active contours without edges. *IEEE transactions on Image processing*, 10(2):266–277, 2001.

[26] Rosa Lasaponara and Nicola Masini. Remote sensing in archaeology: From visual data interpretation to digital data manipulation. In Rosa Lasaponara and Nicola Masini, editors, *Satellite remote sensing: a new tool for archaeology*, pages 3–16. Springer Netherlands, 2012.

[27] John Richards. *Remote sensing digital image analysis: An introduction*. Springer, 2013.

[28] Athos Agapiou, Dimitrios D. Alexakis, Maria Stavrou, Apostolos Sarris, Kyriakos Themistocleous, and Diofantos G. Hadjimitsis. Prospects and limitations of vegetation indices in archeological research: the neolithic thessaly case study. *Proc. SPIE*, 8893:88930D–88930D–10, 2013.

[29] Jared Schuetter, Prem Goel, Joy McCorriston, Jihye Park, Matthew Senn, and Michael Harrower. Autodetection of ancient arabian tombs in high-resolution satellite imagery. *International journal of remote sensing*, 34(19):6611–6635, 2013.

[30] Anna Schneider, Melanie Takla, Alexander Nicolay, Alexandra Raab, and Thomas Raab. A template-matching approach combining morphometric variables for automated mapping of charcoal kiln sites. *Archaeological Prospection*, 22(1):45–62, 2015.

[31] Øivind Due Trier, Maciel Zortea, and Christer Tonning. Automatic detection of mound structures in airborne laser scanning data. *Journal of Archaeological Science: Reports*, 2:69 – 79, 2015.

[32] Bjoern H. Menze, B. Michael Kelm, and Fred A. Hamprecht. From Eigenspots to Fisherspots: Latent spaces in the nonlinear detection of spot patterns in a highly varying background. *Advances in Data Analysis*, pages 255–262, 2007.

[33] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[34] Bjoern H. Menze and Jason A. Ur. Mapping patterns of long-term settlement in Northern Mesopotamia at a large scale. In *Proceedings of the National Academy of Science of the United States*, pages E 778–787, 2012.

[35] Sarah H Parcak. *Satellite remote sensing for archaeology*. Routledge, 2009.

[36] Athos Agapiou and Vasiliki Lysandrou. Remote sensing archaeology: Tracking and mapping evolution in european scientific literature from 1999 to 2015. *Journal of Archaeological Science: Reports*, 4:192 – 200, 2015.

[37] Helmut Mayer. Automatic object extraction from aerial imagery - a survey focusing on buildings. *Computer Vision and Image Understanding*, 74(2):138–149, 1999.

[38] Chungan Lin and Ramakant Nevatia. Building detection and description from a single intensity image. *Computer Vision and Image Understanding*, 72(2):101–121, 1998.

[39] Taejung Kim and Jan-Peter Muller. Development of a graph-based approach for building detection. *Image Vision Comput.*, 17(1):3–14, 1999.

[40] A. Croitoru and Y. Doytsher. Right-angle rooftop polygon extraction in regularised urban areas: Cutting the corners. *The Photogrammetric Record*, 19(108):311–341, 2004.

[41] Claudio Rosito Jung and Rodrigo Schramm. Rectangle detection based on a windowed Hough transform. In *Proceedings of the Computer Graphics and Image Processing (SIBGRAPI), XVII Brazilian Symposium*, pages 113–120, 2004.

[42] Santhana Krishnamachari and Rama Chellappa. Delineating buildings by grouping lines with MRFs. *IEEE Transactions on Image Processing*, 5(1):164–168, 1996.

[43] Csaba Benedek, Xavier Descombes, and Josiane Zerubia. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(1):33–50, 2012.

[44] Beril Sirmacek and Cem Unsalan. A probabilistic framework to detect buildings in aerial and satellite images. *IEEE Tran. Geoscience and Remote Sensing*, 49(1-1):211–221, 2011.

[45] Beril Sirmacek and Cem Unsalan. Urban-area and building detection using SIFT keypoints and graph theory. *IEEE T. Geoscience and Remote Sensing*, 47(4):1156–1167, 2009.

[46] A. Manno-Kovacs and T. Sziranyi. Multidirectional building detection in aerial images without shape templates. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, (1):010000–232, May 2013.

[47] Mathias Ortner, Xavier Descombes, and Josiane Zerubia. A marked point process of rectangles and segments for automatic analysis of digital elevation models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(1):105–119, 2008.

[48] Yangxing Liu, Takeshi Ikenaga, and Satoshi Goto. An MRF model-based approach to the detection of rectangular shape objects in color images. *Signal Processing*, 87(11):2649–2658, 2007.

[49] Christoph Gustav Keller, Christoph Sprunk, Claus Bahlmann, Jan Giebel, and Gregory Baratoff. Real-time recognition of US speed signs. In *Intelligent Vehicles Symposium*, pages 518–523. IEEE, 2008.

[50] Gareth Blake Loy and Nick Mark Barnes. Fast shape-based road sign detection for a driver assistance system. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan*, pages 70–75, 2004.

[51] Yuanxin Zhu, Bridget Carragher, Fabrice Mouche, and Clinton S. Potter. Automatic particle detection through efficient hough transforms. *IEEE Trans. Med. Imaging*, 22(9):1053–1062, 2003.

[52] Zeyun Yu and Chandrajit Bajaj. Detecting circular and rectangular particles based on geometric feature detection in electron micrographs. *Journal of Structural Biology*, 145(12):168 – 180, 2004.

[53] Hankyu Moon, Rama Chellappa, and Azriel Rosenfeld. Optimal edge-based shape detection. *IEEE Transactions on Image Processing*, 11(11):1209–1227, 2002.

[54] Xavier Descombes and Josiane Zerubia. Marked point process in image analysis. *Signal Processing Magazine, IEEE*, 19(5):77–84, 2002.

[55] Yannick Verdie and Florent Lafarge. Detecting parametric objects in large scenes by Monte Carlo sampling. *International Journal of Computer Vision*, 106(1):57–75, 2014.

[56] Zhicheng Li and Laurent Itti. Saliency and gist features for target detection in satellite images. *IEEE Transactions on Image Processing*, 20(7):2017–2029, 2011.

[57] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, Inc., New York, USA, 1995.

[58] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001.

[59] Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, 2001.

[60] Christian Siagian and Laurent Itti. Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):300–312, 2007.

[61] Gong Cheng and Junwei Han. A survey on object detection in optical remote sensing images. {*ISPRS*} *Journal of Photogrammetry and Remote Sensing*, 117:11 – 28, 2016.

[62] Jean Serra and Luc Vincent. An overview of morphological filtering. *Circuits, Systems, and Signal Processing*, 11:47–108, 1992.

[63] David Engel and Cristbal Curio. Scale-invariant medial features based on gradient vector flow fields. In *International Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008.

[64] David. Engel and Cristobal Curio. Shape centered interest points for feature grouping. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 9–16, 2010.

[65] Kaleem Siddiqi, Sylvain Bouix, Allen Tannenbaum, and Steven W Zucker. Hamilton-jacobi skeletons. *International Journal of Computer Vision*, 48(3):215–231, 2002.

[66] Eli Horn and Igor Zingman. System and method for detecting anomalies in a tissue imaged in-vivo, December 16 2014. US Patent 8,913,807.

[67] Richard O. Duda and Peter E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, January 1972.

[68] Brigitte Andres. *Alpine Summer Farms Upland Animal Husbandry and Land Use Strategies in the Bernese Alps (Switzerland)*. 2012.

[69] Itshak Dinstein, AC Fong, LM Ni, and KY Wong. Fast discrimination between homogeneous and textured regions. In *Proc. of Int. Conf. on Pattern Recognition*, pages 361–363, Montreal, Canada, 1984.

[70] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. 2002.

[71] Timo Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:971–987, July 2002.

[72] Iasonas Kokkinos, Georgios Evangelopoulos, and Petros Maragos. Texture analysis and segmentation using modulation features, generative models and weighted curve evolution. *IEEE Trans. Pattern Analyis and Machine Intelligence*, 31:142–157, Jan. 2009.

[73] P. W. Verbeek, H. A. Vrooman, and L. J. Van Vliet. Low-level image processing by max-min filters. *Signal Processing*, 15(3):249–258, 1988.

[74] Martino Pesaresi, Andrea Gerhardinger, and Francois Kayitakire. A robust built-up area presence index by anisitropic rotation-invariant textural measure. *IEEE Journal of selected topics in applied earth observation an remote sensing*, 1:180–192, Sept. 2008.

[75] Robert M. Haralick, K. Shanmugam, and Itshak Dinstein. Textural features for image classification. *IEEE Trans. Systems Man and Cybernetics*, 6:610–621, 1973.

[76] Ruth Bergman, Hila Nachlieli, and Gilit Ruckenstein. Detection of textured areas in natural images using an indicator based on component counts. *J. Electronic Imaging*, 17(4):043003-1–043003-13, 2008.

[77] Kalle Karu, Anil K. Jain, and Ruud M. Bolle. Is there any texture in the image? *Pattern Recognition*, 29:1437–1446, 1996.

[78] Pierre Soille. Beyond self-duality in morphological image analysis. *Image and Vision Computing*, 23:249–257, Feb. 2005.

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

[80] I. Epifanio and P. Soille. Morphological texture features for unsupervised and supervised segmentations of natural landscapes. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4):1074–1083, April 2007.

[81] Keinosuke Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press., 1990.

[82] Nobuyuki Otsu. A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, January 1979.

[83] Cosmin Grigorescu, Nicolai Petkov, and Michel Westenberg. Contour detection based on non-classical receptive field inhibition. *IEEE Trans. Image Processing.*, 12:729–739, 2003.

[84] Benoit Dubuc and Steven Zucker. Complexity, confusion, and perceptual grouping. Part II: Mapping complexity. *J. Mathematical Imaging and Vision*, 15:83–116, 2001.

[85] P. Salembier. Comparison of some morphological segmentation algorithms based on contrast enhancement - application to automatic defect detection. In *European Signal Processing Conference*, pages 833–836, Barcelona, Spain, Sept. 1990.

[86] Jean-Franois Rivest, Pierre Soille, and Serge Beucher. Morphological gradients. *J. Electronic Imaging*, 2(4):326–336, 1993.

[87] Giuseppe Papari and Nicolai Petkov. An improved model for surround suppression by steerable filters and multilevel inhibition with application to contour detection. *Pattern Recognition*, 44:1999 – 2007, 2011.

[88] Tom Fawcett. An introduction to ROC analysis. *Pattern Recogn. Lett.*, 27(8):861–874, June 2006.

[89] Kevin W. Bowyer, Christine Kranenburg, and Sean Dougherty. Edge detector evaluation using empirical ROC curves. *Computer Vision and Image Understanding*, 84(1):77–103, 2001.

[90] Andrian N. Evans and Xin U. Liu. A morphological gradient approach to color edge detection. *IEEE Trans. Image Processing*, 15:1454–1463, 2006.

[91] Allan Hanbury. The morphological top-hat operator generalised to multi-channel images. In *Proc. of the Int. Conf. on Pattern Recognition*, pages 672–675, August 2004.

[92] Tony Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–156, 1998.

[93] Cosmin Grigorescu, Nicolai Petkov, and Michel Westenberg. Contour and boundary detection improved by surround suppression of texture edges. *Image and Vision Computing*, 22:609 – 622, 2004.

[94] Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: A fast line segment detector with a false detection control. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(4):722–732, April 2010.

[95] Stephen M Pizer, Kaleem Siddiqi, Gabor Székely, James N Damon, and Steven W Zucker. Multiscale medial loci and their properties. *International Journal of Computer Vision*, 55(2-3):155–179, 2003.

[96] Pavel Dimitrov, James N Damon, and Kaleem Siddiqi. Flux invariants for shape. In *Proceedings of the the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–835. IEEE, 2003.

[97] Chenyang Xu and Jerry L Prince. Snakes, shapes, and gradient vector flow. *Image Processing, IEEE Transactions on*, 7(3):359–369, 1998.

[98] Chenyang Xu and Jerry L Prince. Generalized gradient vector flow external forces for active contours. *Signal processing*, 71(2):131–139, 1998.

[99] Louisa Lam, Seong-Whan Lee, and Ching Y. Suen. Thinning methodologies-a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):869 – 885, 1992. p.879.

[100] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. John Willey & Sons, 1973.

[101] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893. IEEE, 2005.

[102] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. of Computer Vision*, 60(2):91–110, 2004.

[103] Coen Bron and Joep Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Commun. ACM*, 16(9):575–577, September 1973.

[104] Michail I Schlesinger and Vaclav Hlavac. *Ten lectures on statistical and structural pattern recognition*. Springer, 2002.

[105] David M.J. Tax. One-class classification. PhD Thesis, Delft University of Technology, Delft, 2001.

[106] Heiko Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition*, 40(3):863–874, March 2007.

[107] David M. J. Tax and Robert P. W. Duin. Support vector data description. *Mach. Learn.*, 54(1):45–66, January 2004.

[108] Gilles Blanchard, Gyemin Lee, and Clayton Scott. Semi-supervised novelty detection. *J. Maching Learning Research*, 11:2973–3009, December 2010.

[109] M. Vidal-Naquet and S. Ullman. Object recognition with informative features and linear classification. In *Proc. of the International Conference on Computer Vision*, pages 281–288 vol.1, Oct 2003.

[110] S. J. Devlin, R. Gnanadesikan, and J. R. Kettenring. Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association*, 76(374):354–362, 1981.

[111] Bharath Hariharan, Jitendra Malik, and Deva Ramanan. Discriminative decorrelation for clustering and classification. In *Proc. of the European Conference on Computer Vision*, pages 459–472. Springer, 2012.

[112] Simon Haykin. *Neural networks and learning machines*, chapter 4.17. 3 edition, 2009.

[113] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[114] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. DeCAF: A deep convolutional activation feature for generic visual recognition. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 647–655, June 2014.

[115] Ali S Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 512–519. IEEE, 2014.

[116] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Region-based convolutional networks for accurate object detection and segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2015.

[117] Otávio A. B. Penatti, Keiller Nogueira, and Jefersson A dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In *Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2015.

[118] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, pages 1–42, April 2015.

[119] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *British Machine Vision Conference*, 2014.

[120] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*, May 2015.

[121] Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, and Yann LeCun. OverFeat: Integrated recognition, localization and detection using convolutional networks. In *International Conference on Learning Representations (ICLR)*, April 2014.

[122] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[123] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[124] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross B. Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *CoRR*, abs/1408.5093, 2014.

[125] Caffe Model Zoo. `http://caffe.berkeleyvision.org/model_zoo`.

[126] A. Vedaldi and K. Lenc. Matconvnet – convolutional neural networks for matlab. In *Proceeding of the ACM Int. Conf. on Multimedia*, 2015.

[127] Pre-trained CNN models. `http://www.vlfeat.org/matconvnet/pretrained` [Accessed: 18.11.2015].

[128] Pre-trained OverFeat. `https://github.com/sermanet/OverFeat` [Version: v04-2].

[129] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/` [Version: 0.9.16].

[130] Wojtek J. Krzanowski and David J. Hand. *ROC Curves for Continuous Data*. Chapman & Hall/CRC, 2009.

[131] James A. Hanley and Barbara J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, 1982.

[132] Bojan Pepik, Rodrigo Benenson, Tobias Ritschel, and Bernt Schiele. What is holding back convnets for detection? In Juergen Gall, Peter Gehler, and Bastian Leibe, editors, *37th German Conference on Pattern Recognition (GCPR)*, pages 517–528. 2015.