

Leibniz Universität Hannover
Studiengang Geodäsie und Geoinformatik
Institut für Photogrammetrie und GeoInformation

Bachelorarbeit

Gebäudedetektion in hochaufgelösten Farbinfrarot-Luftbildern in urbanen Gebieten

Alina Maas

Erstprüfer: PD. Dr. techn. Franz Rottensteiner
Zweitprüfer: Dipl.-Ing. Marcel Ziem

Hannover, 2012

Inhaltsverzeichnis

1	Einleitung	2
2	Überblick über die Multispektral-Klassifizierung	3
3	Methodik	5
3.1	Vorverarbeitung	6
3.1.1	Digitales Oberflächenmodell	6
3.1.2	True Orthophoto	10
3.1.3	Normalisiertes digitales Oberflächenmodell	11
3.2	Maximum Likelihood Verfahren	12
3.3	Support Vector Machines	13
3.4	Gebäudeerkennung mit der Dempster-Shafer Theorie	17
4	Experimentelle Untersuchungen	21
4.1	Testdaten und Evaluierungsmethoden	21
4.2	Vorgangsweise	24
4.3	Ergebnisse und Vergleich	27
4.3.1	Bearbeitungszeit der Klassifikationen	28
4.3.2	Visueller Vergleich der Evaluierungsergebnisse	29
4.3.3	Completeness, Correctness, Quality	36
4.3.4	Geometrische Gebauigkeiten	43
4.3.5	Potential der Methoden und Daten	44
5	Zusammenfassung und Ausblick	45
6	Literaturverzeichnis	47
7	Anhang	49

1. Einleitung

Eine der wesentlichen Aufgaben der Photogrammetrie ist die automatische Klassifikation der Landbedeckung. Aus den klassifizierten Bildern können später z.B. Karten erstellt werden, oder sie dienen als Grundlage für die Erkennung von Veränderungen. Besonders wichtig ist die Automatisierung der Klassifikation aufgrund des steigenden Datenvolumens. Eine visuelle Interpretation ist sehr zeitaufwendig und somit ab einer gewissen Datenmenge nicht mehr wirtschaftlich. Aus diesem Grund wird versucht, die Bildinterpretation durch einen menschlichen Beobachter weitgehend zu vermeiden. Besonders in der Objekterkennung wird daher an effizienten Methoden geforscht, um so mindestens teilweise die Verarbeitung der Bilder zu automatisieren. Auch die Benutzer verlangen automatisierbare und an die Bedürfnisse angepasste Verfahren zur Objekterkennung (Graf, 2009).

Diese Arbeit beschäftigt sich mit der Gebäudeerkennung in hochaufgelösten Luftbildern. Als Grundlage der Untersuchung stehen Luftbilder in zwei Auflösungen von drei unterschiedlich charakterisierten Interessensgebieten innerhalb einer Stadt zur Verfügung. Die Bilder wurden mit einer modernen digitalen großformatiger Luftbildkamera produziert und weisen einen grünen, roten und infraroten Kanal auf. Durch moderne Matchingverfahren werden aus den Bildern für jedes Interessengebiet ein normalisiertes digitales Oberflächenmodell sowie ein true Orthophoto erstellt. Diese werden für einen späteren Vergleich sowohl mit einer sehr hohen Auflösung, als auch mit einer etwas geringeren Auflösung erstellt.

Die Gebäudeerkennung erfolgt dann auf Basis von drei verschiedenen Klassifikationsansätzen, die zur Unterscheidung der Objekte die Höheninformation des Oberflächenmodells sowie die radiometrische Information aus dem true Orthophoto nutzen. Die drei Verfahren unterscheiden sich stark in ihrer Methodik und Anwendung, sodass auch die Ergebnisse unterschiedlich ausfallen. Das Maximum-Likelihood-Verfahren (Allbertz, 1991) ist ein älteres, relativ einfaches Standardverfahren, während die Support-Vektor-Machines Klassifikation (Heinert, 2010) aktueller und rechenintensiverer ist und derzeit ein etabliertes Verfahren für diverse Klassifikationsaufgaben darstellt. Die Gebäudeerkennung nach Dempster-Shafer (Rottensteiner et al., 2005) ist ebenfalls eine aktuelle Methode.

Das Ziel dieser Arbeit ist es, das Potential der Daten und der Klassifikationsmethoden zu untersuchen. Es wird einerseits darauf eingegangen, wie gut die Ergebnisse der einzelnen Verfahren bei der Gebäudeerkennung mit den gegebenen Daten sind und wo Probleme auftreten können. Die

Evaluierung der Ergebnisse erfolgt mithilfe von manuell erstellten Referenzdaten. Die klassifizierten Gebäude werden mit der Referenz verglichen, sodass verschiedene Qualitätsmaße bestimmt werden können. Die Qualitätsmaße und die Klassifikationsergebnisse der drei Methoden werden nun miteinander verglichen. Dadurch können einzelne Vor- und Nachteile in Erfahrung gebracht und die Unterschiede diskutiert werden. Es wird auch auf Fehlklassifikationen eingegangen, und deren mögliche Gründe werden erörtert. Schließlich erfolgt durch einen Vergleich der Klassifikationsergebnisse basierend auf Daten mit unterschiedlicher Auflösung.

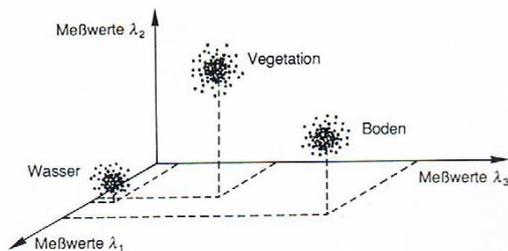
In der Arbeit wird anfangs ein kurzer Überblick der Multispektral-Klassifizierung gegeben, gefolgt von einer Beschreibung der Methodik. In dieser werden zunächst die nötigen Vorarbeiten, wie z.B. die Erstellung des Oberflächenmodells, erläutert, und anschließend wird auf die verschiedenen Klassifikationsmethoden eingegangen. In Abschnitt 4 wird die experimentelle Untersuchung, die im Rahmen dieser Arbeit durchgeführt wurde, dokumentiert. Dabei werden die genutzten Testdaten, die Evaluierungsmethode, sowie die Durchführung der Experimente beschrieben. Ebenfalls zu den experimentellen Untersuchungen gehört die Präsentation und Analyse der Ergebnisse. Schließlich folgen eine kurze Zusammenfassung und ein Ausblick auf ein mögliches weiteres Vorgehen.

2. Überblick über die Multispektral-Klassifizierung

Bei der Klassifikation werden einzelne Bildprimitive (d.h. Pixel oder Segmente) aufgrund ihrer charakteristischen Eigenschaften, auch Merkmale genannt, bestimmten Klassen durch einen Klassifikator zugeordnet. Ist nur ein Merkmal, z.B. ein Grauwert, vorhanden kann lediglich eine 1D-Analyse vorgenommen werden, z.B. die Trennung der Klassen durch einen Schwellenwert. Dies kann jedoch sehr fehlerhaft sein (Kraus, 1990), weswegen üblicherweise mehrere Merkmale erfasst werden. In der Bildanalyse sind diese Merkmale meist verschiedene spektrale Kanäle sowie Funktionen dieser spektralen Kanäle. Bei der Analyse multispektraler Daten zur Unterscheidung der Objektklassen wird das Verfahren Multispektral-Klassifizierung genannt (Albertz, 1991). Im Folgenden wird die Multispektral-Klassifizierung, wie in Albertz (1991) beschrieben, genauer erläutert.

Die beobachteten Bildmerkmale spannen einen *Merkmalsraum* auf, wobei die Dimension dieses Raumes der Anzahl der Merkmale entspricht. Bei einem Multispektralbild mit k Kanälen entsteht also ein k -dimensionaler Merkmalsraum, bei dem jeder Kanal einer Koordinatenachse entspricht. Da die Merkmale desselben Objekts ähnliche Eigenschaften aufweisen, bilden sich im Merkmalsraum

verschiedene Punkthaufen, von denen im Idealfall jeder einer Objektklasse entspricht. Lassen die Punktwolken sich eindeutig trennen, erlaubt das die Zuordnung der Bildprimitive zu den entsprechenden Klassen. In Abbildung 2.1 ist dies für ein 3-Kanalbild gezeigt. In der Praxis ist dies nicht so einfach, da sich Punktwolken, die mehrerer Klassen entsprechen, oft überlappen.



*Abbildung 2.1
3D-Merkmalraum mit den Punktwolken die
den Klassen Wasser, Vegetation und Boden
zugeordnet werden
[Albertz, 1991]*

Für die Umsetzung der Multispektral-Klassifizierung können verschiedene Klassifikationsverfahren verwendet werden. Es wird dabei zwischen den überwachten und unüberwachten Klassifikationsverfahren unterschieden, sowie zwischen pixel- und segmentbasierter Klassifikation.

Die überwachten Klassifikationsverfahren benötigen Vorinformationen in Form von Trainingsgebieten.

Trainingsgebiete sind Referenzflächen, die bereits den einzelnen Klassen zugeordnet wurden. Dies erfolgt z.B. manuell oder durch andere Quellen. Aus diesen Trainingsgebieten lernt der Klassifikator die Unterscheidung der einzelnen Klassen. Aus diesem Grund ist es wichtig, dass die Trainingsgebiete möglichst fehlerfrei sind und dass für jede Klasse mindestens ein Trainingsgebiet vorhanden ist. Laut de Lange (2006) ist es außerdem wichtig, die Trainingsgebiete mit Bedacht, auch im Bezug auf das genutzte Klassifikationsverfahren, zu wählen. Die Gebiete sollten nur Merkmale einer Klasse aufweisen und die ganze Vielfalt der einzelnen Merkmale widerspiegeln.

Die Merkmale der zu klassifizierenden Bildprimitive werden mit den Informationen aus den Trainingsgebieten verglichen und dann der entsprechenden Klasse zugeordnet. Einige Standardalgorithmen sind z.B. das Maximum-Likelihood-Verfahren (s. Abschnitt 3.2.), das Minimum-Distance-Verfahren und das Quaderverfahren, welche in (Albertz, 1991) genauer erläutert werden. Ein weiteres überwachtes Verfahren ist die Methode der Support Vector Machines (s. Abschnitt 2.4.).

Die unüberwachte Klassifikation unterteilt ein Bild in eine vordefinierte Anzahl von Klassen mit ähnlichen Merkmalen. Dabei wird kein Vorwissen über die Daten, wie z.B. Trainingsgebiete, vorausgesetzt. Der Benutzer gibt lediglich die Anzahl der Klassen und, bei iterativen Verfahren, die Anzahl der Iterationen sowie eine Konvergenzschwelle vor. Mit diesen Angaben werden dann die Daten in Ballungen gruppiert, die keine bekannte Bedeutung besitzen müssen. Dennoch können sie

Zusammenhänge aufzeigen, die einem Menschen ggf. nicht aufgefallen wären, besonders bei Bildern mit mehr als drei Kanälen. Die unüberwachte Klassifikation dient allerdings meist nur als Vorstufe zur überwachten Klassifikation. Mit den entstandenen Ballungen kann z.B. festgelegt werden wo und wie viele Trainingsgebiete für die jeweilige Klasse benötigt werden oder wie viele Klassen mit den gegebenen Merkmalen sinnvoll unterschieden werden können. Zwei bekannte Beispiele der unüberwachten Klassifikation sind das k-means- und das ISODATA-Verfahren (Sörgel, 2012).

Ein Verfahren zur Klassifikation kann entweder segment- oder pixelbasiert durchgeführt werden (z.B. Helmholz, 2012).

Bei pixelbasierten Methoden wird die Klassifikation auf Grund von Merkmalen eines Pixels und ggf. dessen lokaler Umgebung durchgeführt. Bei den zu klassifizierenden Bildprimitiven handelt es sich also um Pixel. Die Ergebnisbilder dieser Verfahren können jedoch unter Umständen einen starken Salz-und-Pfeffer-Effekt¹ aufweisen.

Bei segmentbasierten Methoden erfolgt zunächst eine Segmentierung, also eine Aufteilung des Bildes in zusammenhängende, homogene Regionen. Die Merkmale dieser Segmente werden dann zur Klassifikation genutzt. Da hier bereits größere Regionen klassifiziert werden, kann der Salz-und-Pfeffer-Effekt nicht auftreten. Ein weiterer Vorteil dieser Methode ist, dass mehrere Pixel zur Bestimmung der Merkmale eines Segmentes beitragen und so die Bestimmung der Merkmale robuster wird. Durch die Segmentierung können aber auch Nachteile entstehen. Besonders bei Bildern, die einen schlechten Kontrast aufweisen, kann es vorkommen, dass die Klassengrenze nicht mit einer Segmentgrenze übereinstimmt und so z.B. ein Segment zwei oder mehrere Klassen beinhaltet.

3. Methodik

Im Rahmen dieser Arbeit werden drei verschiedene Klassifikationsmethoden zur Gebäudeerkennung an unterschiedlichen Gebieten getestet. Die Ergebnisse der Klassifikation sollen in einem übergeordneten Koordinatensystem vorliegen. Dafür muss die Klassifikation auf der Basis geokodierter Bilder durchgeführt werden. Bei der Geokodierung wird das digitale Bildmaterial in einen geographischen Raumbezug gestellt, welches auch die Eliminierung von system- oder geländeabhängigen Verzerrungen beinhaltet. Bei den geokodierten Bildern ist jedem Pixel eine Objektkoordinate (x, y, ggf.z) zugeordnet (Prinz, 2007).

¹ Hier: Auftreten einzelner Pixel, deren Klassenzugehörigkeit sich von den Pixeln der näheren Umgebung unterscheidet.

Als Ausgangsdaten stehen jedoch keine geokodierten Bilder der zu klassifizierenden Gebiete zur Verfügung, sondern mehrere multispektrale Luftbilder, deren absolute Orientierung bekannt ist. Am Anfang werden daher für jedes Interessengebiet aus den Einzelbildern ein digitales Oberflächenmodell und ein true Orthophoto erstellt, die als Grundlage der Klassifikationen dienen. Die genauen Verfahrensschritte, sowie die Methodik der einzelnen Klassifikationsmethoden der Vorbearbeitung werden im Folgenden genauer beleuchtet. Für die Gebäudeerkennung wurden das Maximum-Likelihood-Verfahren, das SVM-Verfahren, sowie eine Methode nach Dempster-Shafer verwendet.

3.1. Vorverarbeitung

Zunächst wird aus den Bildern durch digitale Bildzuordnung (digital image matching) ein digitales Oberflächenmodell (DOM) erzeugt. Hier wird das DOM als Raster modelliert, bei dem jedes Pixel einen bestimmten Punkt mit den Raumkoordinaten $(x, y, z(x,y))$ auf der Erdoberfläche oder auf einem sich auf der Erdoberfläche befindenden Objekt repräsentiert. Mit diesem DOM kann ein normalisiertes digitales Oberflächenmodell (nDOM) approximiert werden, welches lediglich die Höhen von Objekten über dem Gelände zeigt. Aus dem DOM und den gegebenen Luftbildern kann zusätzlich ein Farbinfrarot-True Orthophoto (TOP) erstellt werden. Das TOP ist ein maßstabsgetreues und verzerrungsfreies digitales Bild, bei dem Objekte, wie z.B. Gebäude, lagerichtig dargestellt werden.

Da auch das nDOM und TOP geokodiert sind, können aus ihnen für jeden Punkt die Höhe über der Geländeoberfläche und die spektrale Reflexionseigenschaften im grünen, roten und infraroten Bereich ermittelt werden. Die Höhe, die Reflexionseigenschaft und von ihnen abgeleitete Größen dienen als Merkmale zur Klassifikation jedes Punktes der Oberfläche.

3.1.1. Digitales Oberflächenmodell

Ein DOM beschreibt die Geländeoberfläche einschließlich aller sich darauf befindenden Objekte, wie Bebauung und Vegetation. Dargestellt wird das DOM z.B. durch ein digitales Bild, welches ein gleichmäßiges Höhenraster mit den positionsabhängigen Höhen $z(x,y)$ darstellt (GI-Lexikon, 2001).

Mit der digitale Bildzuordnung (digital image matching) (Heipke, 2010), ist es möglich, 3D Koordinaten aus mindestens zwei Bildern, die das gleiche Gebiet abdecken, zu erzeugen. Dabei werden homologe Punkte in zwei oder mehreren Bildern bestimmt, aus denen bei bekannter

Orientierung, wie in Abbildung 3.1 gezeigt, durch einen räumlichen Vorwärtsschnitt die 3D Koordinaten eines Punktes p ermittelt werden können. Wichtig ist, dass die Bilder von unterschiedlichen Standpunkten aufgenommen werden, damit die Abbildungsstrahler beider Bilder nicht identisch sind.

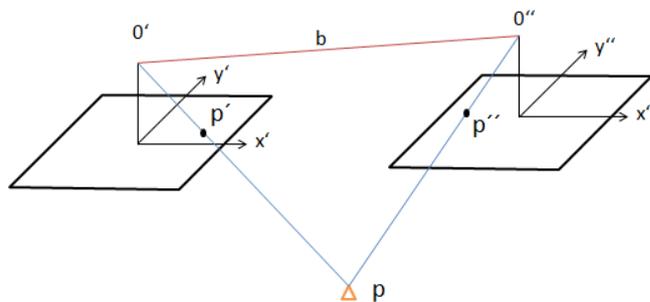


Abbildung 3.1

Vorwärtsschnitt zur Erzeugung eines 3D-Bildpunktes
[Heipke,2011]

Beim *dichten Matching* („dense matching“) wird für jedes Pixel in dem ersten Bild ein 3D-Punkt bestimmt. Dafür wird für jedes Pixel p' im ersten Bild der entsprechende Bildpunkt p'' im zweiten Bild gesucht. Die Bestimmung des homologen Bildpunktes im zweiten Bild kann bei einer Suche, die das komplette Bild abdeckt, unter Umständen nicht eindeutig sein und eine hohe Laufzeit aufweisen. Dies kann durch die Verwendung der Epipolargeometrie, wie in (Heipke, 2010) erläutert, verbessert werden. Ist der Punkt p' in einem Bild bekannt, muss der Punkt im zweiten Bild auf der Epipolarlinie liegen. Dadurch wird der Suchraum von dem kompletten Bild (2D) auf die Epipolarlinie (1D) reduziert. Durch eine Transformation zweier Bilder in allgemeiner Lage in zwei Koordinatensysteme, deren Achsen parallel sind und deren x-Achsen parallel zu Basis b liegen, entstehen zwei Epipolarbilder. Die Epipolarlinien entsprechen dann den Bildzeilen. Epipolarbilder haben den Vorteil, dass mit ihnen eine schnellere Berechnung möglich ist als mit den Originalbildern, da der Suchbereich auf eine Zeile begrenzt wird.

Semiglobal Matching (Hirschmüller, 2008) ist ein pixelbasiertes Verfahren zum Matching zweier Bilder, welches mit den oben erläuterten Epipolarbildern I' und I'' arbeitet. Da die Bildpunkte auf einer Zeile liegen und somit $y' = y''$ ist, kann ein Bildpunkt im zweiten Bild $p'' = [x'', y'']$ durch die Koordinaten des Punktes im erstem Bild $p' = [x', y']$ und einer Disparität d parametrisiert werden: $p'' = [x' - d, y']$. Diese Disparität muss beim dichten Matching nun für jedes Pixel im ersten Bild bestimmt werden. Durch die Angabe der minimalen und maximalen Disparität entsteht mit der Größe des ersten Bildes ein Disparitätsraum in Form eines Quaders, bei dem jedes Voxel (x', y', d) ein Punktpaar darstellt. Durch eine Kostenfunktion werden für jedes Pixel (x', y') und alle möglichen homologen Pixel im zweitem Bild, d.h. für alle möglichen Werte $d(x', y')$, Kosten C zwischen den beiden Bildern bestimmt. Die Kosten spiegeln die Ähnlichkeit der Grauwerte der beiden Pixel wieder.

Niedrigen Kosten weisen darauf hin, dass es sich um homologe Bildpunkte handelt, während hohe Kosten bei vermutlich nicht korrespondierenden Punkten auftreten. Die Kostenfunktion kann z.B. durch Mutual Information (Hirschmüller, 2008) oder durch absolute Differenzen der Grauwerte für p' und p'' bestimmt werden. Ein weiteres Beispiel ist die Kostenfunktion von Birchfield und Tomasi (1998).

Ziel ist es nun, in dem Disparitätsraum eine ggf. leicht gekrümmte Fläche mit minimalen Kosten zu finden. Dadurch entsteht ein Disparitätsbild $D = D(x', y')$, welches jedem Pixel (x', y') eine Disparität $d(x', y')$ zuordnet.

Aufgrund des möglicherweise mehrfachen Auftretens eines Grauwertes bzw. aufgrund von zufälligen Fehlern in den Grauwerten innerhalb einer Zeile kann es vorkommen, dass ein falsch zugeordnetes Bildpunktpaar die gleichen oder sogar geringere Kosten aufweist als ein richtig zugeordnetes Paar. Aus diesem Grund werden die benachbarten Pixel durch eine Glattheitsbedingung mit einbezogen.

Eine Energiefunktion $E(D)$, abhängig von dem Disparitätsbild D , vereint die pixelweisen Kosten mit einer zusätzlichen Kostenfunktion E_{glatt} , welche Disparitätsschwankungen penalisiert und somit zu einer Glättung des Disparitätsbildes führt. Das Disparitätsbild D wird so bestimmt, dass die Energiefunktion $E(D)$ minimal ist.

$$E(D) = \sum_{p'} (C(p', D_{p'}) + E_{glatt})$$

$$E_{glatt} = \sum_{p'' \in N_{p'}} P_1 T[|D_{p'} - D_{p''}| = 1] + \sum_{p'' \in N_{p'}} P_2 T[|D_{p'} - D_{p''}| > 1]$$

Der erste Term beschreibt dabei die Kosten C als Funktion des Disparitätsbildes D . Der zweite Term besteht aus der Glattheitsbedingung, die bei identischen Disparitäten $D_{p'} = D_{p''}$ den Wert 0 liefert und ansonsten einen Strafterm P_1 oder P_2 . Die Wahl des Strafterms wird über die Funktion $T[\cdot]$, welche bei wahren Argument 1 und andernfalls 0 zurückgibt, von dem Disparitätsunterschied $|D_{p'} - D_{p''}|$ an benachbarter Pixeln p' und p'' entschieden. Der Strafterm P_1 wird bei Disparitätsunterschieden von 1 bezüglich Pixel in der Nachbarschaft $N_{p'}$ von p' verwendet, während der Strafterm P_2 bei größeren Sprüngen genutzt wird, wobei $P_1 \leq P_2$ gilt. Die Tatsache, dass für kleine Disparitätsunterschiede ein kleiner Strafterm verwendet wird, erlaubt eine Anpassung an geneigte oder gekrümmte Flächen. Der Strafterm P_2 wird zusätzlich noch durch $P_2 := \frac{P_2^0}{1 + \|\nabla I'\|}$ an den Grauwertgradienten $\nabla I'$ des ersten Bildes angepasst, um größere Sprünge bei starken Grauwertunterschieden zuzulassen. Ist P_2 nun kleiner als P_1 , wird P_2 durch P_1 ersetzt, sodass immer $P_1 \leq P_2$ gilt.

Die Bestimmung von D aus der Minimierung von $E(D)$ muss approximativ erfolgen. Beim Semiglobal Matching (SGM) (Hirschmüller, 2008) werden analog zur dynamischen Programmierung (DP) Kosten entlang von 1D Pfaden akkumuliert, wobei die Glattheitsterme entlang der Pfade bei der

Akkumulation berücksichtigt werden. Um dem „streaking“-Effekt des ursprünglichen DP-Verfahrens zu entkommen, werden beim SGM die Kosten in verschiedenen Richtungen akkumuliert. Das Disparitätsbild D ergibt sich aus den akkumulierten Kosten, indem an jeder Position (x', y') die Disparität mit minimalen Kosten gewählt wird.

Es kann vorkommen, dass sich die Werte von mehreren lokalen Minima der akkumulierten Kosten kaum unterscheiden. Daher werden nur Minima akzeptiert, die sich um mindestens $Q_{max}\%$ von dem zweitkleinsten Minima unterscheiden. Andernfalls wird das Pixel als „leer“ markiert.

Zur Kontrolle des ermittelten Disparitätsbildes D werden die Epipolarbilder I' und I'' getauscht, sodass I'' das Referenzbild darstellt und in I' die homologen Punkte gesucht werden. In dieser Form wird erneut durch SGM ein Disparitätsbild D_2 erstellt. Für jede Position wird nun die Disparität $d_1 = D(x', y')$ aus dem ersten Disparitätsbild und die entsprechende Disparität $d_2 = D_2(x' + d_1, y')$ aus dem 2. Disparitätsbild bestimmt. Bei einer korrekten Zuordnung ist $|d_1 - d_2| = 0$. Aus diesem Grund wird d_1 nur anerkannt wenn $|d_1 - d_2| \leq 1$ und andernfalls verworfen. Eine starke Abweichung der Disparitäten entsteht z.B. bei homogenen Flächen oder durch Verdeckungen.

Schließlich folgt eine Segmentierung des Disparitätsbildes durch Region Growing². Als Homogenitätskriterium dient der Disparitätsunterschied zwischen dem aktuellen Pixel und dessen unmittelbaren Nachbarn, der einen gegebenen Schwellenwert S_{diff} nicht überschreiten darf. Segmente, die kleiner als ein vordefinierter Schwellenwert S_{seg} sind, werden nicht genutzt.

Das Matching wird nun für alle möglichen Bildpaare durchgeführt. Mithilfe der Disparitäten aus dem Disparitätsbild können homologe Punkte einander zugeordnet werden. Durch den räumlichen Vorwärtsschnitt entsteht für jedes Punktpaar eine 3D-Koordinate, und diese bilden zusammen eine 3D-Punktwolke für jedes Bildpaar. Aus ihnen wird mit der nearest neighborhood Methode ein DOM-Raster interpoliert. Die so entstandenen rohen DOM können noch größere Teilflächen besitzen, die als leer markiert wurden. Aus diesem Grund werden die rohen DOM aus allen Bildpaaren kombiniert, indem in jedem DOM-Pixel der Median aus den vorhandenen Höhenwerten gebildet wird. Die noch bestehenden Löcher in dem kombinierten DOM, also Pixel, die nach wie vor keinen Höhenwert besitzen, werden geschlossen. Besitzen die umliegenden Pixel einander ähnliche Höhen, können die Höhen der Pixel innerhalb des Loches linear interpoliert werden. Andernfalls wird das Loch durch nichtlineare Diffusion unter alleiniger Berücksichtigung der niedrig gelegenen Teile des Randes des Loches geschlossen (Kosov et al., 2012).

² Region Growing ist ein Bildsegmentierungsverfahren. Beginnend an einem Startpunkt werden angrenzenden Nachbarpixel von bereits bestehenden Regionen auf ein Homogenitätskriterium überprüft und der Region hinzugefügt falls das Kriterium erfüllt ist. Existiert kein Nachbar, der das Kriterium erfüllt, wird ein neuer Startpunkt für eine weitere Region aus den noch nicht zugeordneten Pixeln gewählt. Sind alle Bildelemente einer Region zugeordnet, endet das Verfahren.

3.1.2. True Orthophoto

Ein Orthophoto ist ein entzerrtes und maßstabsgetreues digitales Bild, welches fast vollständig einer Orthogonalprojektion des Objekts in ein orthogonales Koordinatensystem entspricht (Baldenhofer, 2012). Der Vorteil eines Orthophotos ist, dass es den Informationsgehalt eines Luftbildes aufweist und die geometrischen Eigenschaften einer Karte (Kraus, 1990). Ein True Orthophoto (TOP) stellt zusätzlich Objekte, wie Gebäude und Bäume, lagerichtig dar und benötigt zu dessen Herstellung ein DOM (Wiggenhagen, Wißmann, 2003).

Die Luftbilder werden, unter Berücksichtigung der Sichtbarkeit, auf das DOM projiziert, sodass aus jedem digitalen Eingangsbild ein rohes TOP entsteht. Für die Sichtbarkeitsanalyse werden für jeden Punkt des DOM die Abbildsstrahlen zu jedem Bild rekonstruiert (Abb. 3.3). Schneidet diese Gerade das DOM, wie in Abbildung 3.2 der Abbildungsstrahl von p_1' , befindet sich dieser Punkt in einem sichttoten Raum und wird in dem rohen TOP als solcher markiert.

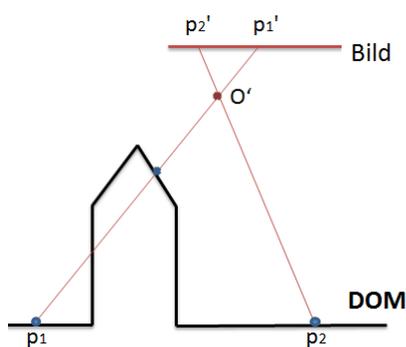


Abbildung 3.2

Projektion eines Bildpunktes
auf ein DOM

Aus den Bildpunkten der rohen TOP wird ein kombiniertes TOP gebildet, wobei nur die Grauwerte berücksichtigt werden, die nicht als sichttot markiert wurden. Bei Pixeln, die in mehreren Luftbildern abgebildet wurden, wird der Median aus den verschiedenen Grauwerten verwendet. Der Vorteil des Median ist, dass bewegte Objekte wie Autos herausfallen, wenn mindestens drei Bilder diesen Bereich abdecken.

Das True Orthophoto dient als Grundlage für die Berechnung radiometrischer Merkmale bei der Klassifikation. Für jedes zu klassifizierende Pixel kann nun die Reflexionseigenschaft des Objektpunktes im grünen (G), roten (R) und infrarotem (IR) Bereich als Merkmal mit einfließen. Auch Funktionen dieser Eigenschaften können als Merkmal genutzt werden. Ein Beispiel dafür ist der Normalized Difference Vegetation Index $NDVI = \frac{NIR-R}{NIR+R}$, welcher als Merkmal zur Identifikation von Vegetation dienen kann.

3.1.3. Normalisiertes digitales Oberflächenmodell

Aus dem DOM wurde mit einer Methode nach Rottensteiner et al. (2005) ein grobes digitales Geländemodell (DGM) erzeugt. Allerdings wird anstelle des morphologischen Openings ein Rangfilter verwendet, um den Einfluss von Ausreißern im DOM zu eliminieren.

Dabei werden anfangs alle Höhenwerte z aus dem DOM in einem $N_1 * N_1$ großen Fenster, zentriert in dem Pixel (x,y) , genutzt. Der Parameter N_1 beschreibt die maximale Ausdehnung des größten Gebäudes in dem Interessengebiet. Die Höhenwerte z werden nun der Größe nach sortiert, das s -quantil der Höhen wird bestimmt und als genäherte Geländehöhe genutzt. Der Parameter s liegt zwischen 1 und 100 und zeigt an, dass s % der Pixel in dem Fenster einen kleineren Wert als der gewählte Wert aufweisen. Durch dieses Verfahren werden nun die Geländehöhen für jeden Punkt des DOM bestimmt und so ein genähertes DGM erstellt.

In Gebieten mit unterschiedlich großen Gebäuden kann es vorkommen, dass durch die Filterung nicht nur die Gebäude eliminiert wird, sondern auch Geländestrukturen (Abb. 3.3 a, b). Bei einer Fenstergröße von $N_n * N_n$, wobei der Parameter N_n die typische Ausdehnung kleinerer Gebäude beschreibt, tritt dieses Problem zwar nicht auf, jedoch werden größeren Gebäude so nicht eliminiert (Abb. 3.3 c). Aus diesem Grund erfolgt die DGM Erstellung iterativ mit vordefinierten N_j mit $j \in \{1, \dots, n\}$, wobei $N_{j+1} < N_j < N_{j-1}$. Dabei wird, wie oben erläutert, aus dem digitalen Oberflächenmodell in dem Iterationsschritt j ein digitales Geländemodell des Iterationsschrittes j (DGM_j) mit einem Fenster der Größe $N_j * N_j$ erstellt. Durch eine vereinfachte Klassifikationsmethode aus (Rottensteiner et al., 2005), welches mit dem NDVI, der Oberflächenstruktur des DOM und der Höhe aus dem DOM und DGM_j arbeitet, werden die Gebäude mit einer Ausdehnung von etwa N_j erkannt und die entsprechenden Pixel markiert. Beim nächsten Iterationsschritt werden die markierten Stellen des Ergebnisses des Rangfilters durch die Höhen des DGM_{j-1} ersetzt, sodass ein neues Höhenmodell (DGM_j) entsteht.

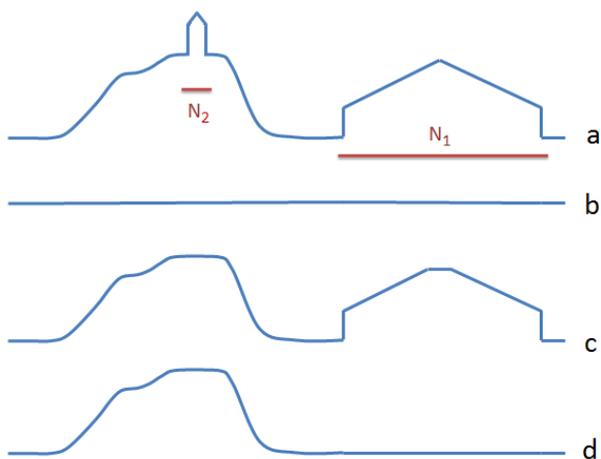


Abbildung 3.3

Beispiel zur Erstellung eines groben DGM aus einem DOM mit den Fenstergrößen N_1 und N_2

- a) DOM
- b) DGM mit N_1
- c) DGM mit N_2
- d) $Kombiniertes DGM$

Durch die Subtraktion des DGM vom DOM wird das normalisierte digitale Oberflächenmodell (nDOM) erzeugt. Dieses repräsentiert die Höhe von Objekten über der Geländeoberfläche und stellt somit ein wichtiges Merkmal zur Unterscheidung solcher Objekte (z.B. Bäume oder Bebauung) von Objekten am Gelände dar.

3.2. Maximum Likelihood Klassifikation

Ein weit verbreitetes Standardverfahren zur überwachten Klassifikation ist das Maximum-Likelihood-Verfahren (ML-Verfahren), welches z.B. in Kraus (1990) erläutert wird. Dabei wird für jedes Bildprimitiv mit dem Merkmalsvektor x die Wahrscheinlichkeit $p(x|C)$ berechnet. Diese zeigt an, wie wahrscheinlich es ist, dass der Merkmalsvektor x in der Klasse C auftritt. Das Objekt wird dann der Klasse mit der größten Wahrscheinlichkeit zugeordnet.

Bei dem ML-Verfahren wird davon ausgegangen, dass $p(x|C)$ für jede Klasse durch eine k -dimensionale Normalverteilung dargestellt werden kann. k ist hierbei die Dimension des Merkmalsvektors. Da es sich um ein überwachtetes Klassifikationsverfahren handelt, benötigt das Verfahren für jede Klasse mindestens ein Trainingsgebiet. Aus diesen werden dann für jede Klasse Mittelwert-Vektor μ_k und Kovarianzmatrix Σ_k ermittelt.

Mit diesen statistischen Kenngrößen einer Klasse C_k kann für einen Merkmalsvektor x für jede Klasse C_k die Wahrscheinlichkeit $p(x|C_k)$ berechnet werden (Sörgel, 2012):

$$p(x|C_k) = \frac{1}{(2\pi)^{\frac{k}{2}} * |\Sigma_k|^{\frac{1}{2}}} * e^{-\frac{1}{2}[(x-\mu_k)^T * \Sigma_k^{-1} * (x-\mu_k)]}$$

Das Objekt wird schließlich der Klasse mit der größten Wahrscheinlichkeitsdichte zugeordnet. In Abbildung 3.4 ist das Maximum-Likelihood-Verfahren im 1D-Fall abgebildet.

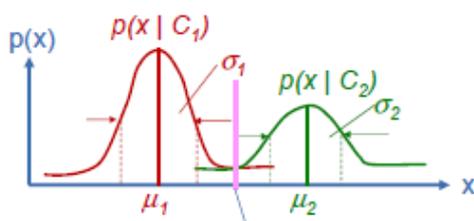


Abbildung 3.4

Wahrscheinlichkeitsverteilung zweier Klassen bei einer Maximum-Likelihood-Klassifizierung mit einem Merkmal
[Sörgel, 2012]

Das ML-Verfahren benötigt keine zusätzlichen Parameter (Albertz, 1991). Mit unterschiedlichen Merkmalsausprägungen innerhalb einer Klasse kann das Verfahren aufgrund der Mittelbildung nicht gut umgehen (Helmholz, 2012). Ein weiterer Nachteil ist, dass die Annahme getroffen wird, dass die Merkmale der Klassen normalverteilt sind, was jedoch nicht gegeben sein muss (Albertz, 1991).

3.3. Support Vector Machines

Das überwachte Klassifikationsverfahren Support Vector Machines (SVM) wurde 1995 von Vapnik eingeführt (Vapnik, 1995) und findet seitdem in vielen Bereichen Verwendung (Helmholz, 2012).

Bei der SVM handelt es sich um ein überwachtes, binäres Klassifikationsverfahren, mit dem Objekte anhand von Merkmalen in eine von zwei Klassen zugeordnet werden können. Die Trennung erfolgt mit einer Hyperebene im Merkmalsraum.

Die Funktionsweise des SVM-Verfahrens wird im Folgenden, wie in Heinert (2010) beschrieben, erläutert, wobei anfangs von einem linear trennbaren Datensatz ausgegangen wird.

(1.) Prinzip

Die Trainingsdaten beinhalten l Pixel mit jeweils einem Merkmalsvektor x_i und einer Klassenzugehörigkeit $y_i \in \{+1, -1\}$ mit $i \in \{1, \dots, l\}$. Die Trennung der Klassen erfolgt bei SVM durch eine Hyperebene H im Merkmalsraum, die so gewählt wird, dass der Abstand d zu den nächsten Punkten in den Trainingsdaten maximal wird (Prinzip des *maximum margin*). Die Punkte mit genau diesem Abstand d heißen Stützvektoren (*Support Vectors*). Durch die Stützvektoren werden 2 Ebenen E_1 und E_2 definiert, die beide parallel zu der Hyperebene laufen und den Abstand d zu ihr aufweisen. Der Bereich zwischen den Ebenen beinhaltet keine Merkmalsvektoren und wird Trennbereich (*margin*) genannt. Sind die Stützvektoren aus dem Training bekannt, lässt sich die Klassenzugehörigkeit auch für neue Pixel bestimmen.

(2.) Geometrischer Ansatz

In Abbildung 3.5 ist die Hyperebene H mit deren Normalenvektor w dargestellt, die die 2 Klassen $y=+1$ und $y=-1$ voneinander trennt. Die eingekreisten Punkte sind die Stützvektoren.

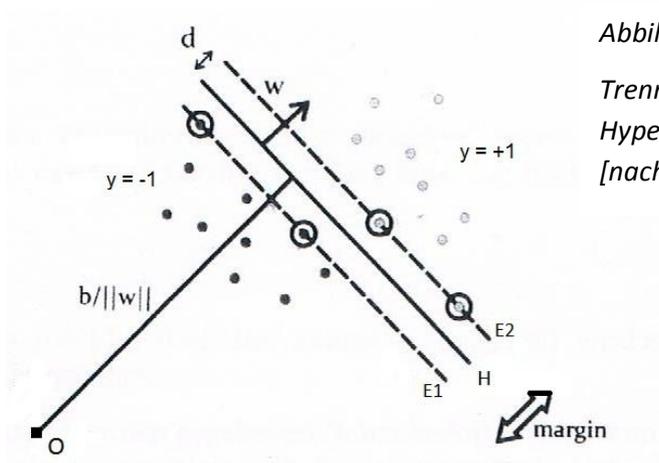


Abbildung 3.5
Trennung zweier Klassen durch die
Hyperebene H
[nach Helmholz, 2012]

Die Hyperebene kann durch die Gleichung $wx + b = 0$ dargestellt werden. w ist dabei der Normalenvektor der Hyperebene und $b/\|w\|$ der Abstand der Hyperebene zum Ursprung O , wobei $\|w\|$ die euklidische Norm von w ist.

Da die Trainingspunkte $[x_i, y_i]$ außerhalb des Trennbereichs liegen, müssen sie die folgende Ungleichung erfüllen: $y_i(wx_i + b) \geq 1$.

Für Stützvektoren gilt die Gleichheit. Die beiden Hyperebenen der Stützvektoren E_1 und E_2 werden daher durch $wx + b = \pm 1$ repräsentiert.

Der Abstand von E_1 bzw. E_2 zum Ursprung ist demnach $(b - 1)/\|w\|$ bzw. $(b + 1)/\|w\|$. Der Trennbereich von $2d$, also der Abstand von E_1 und E_2 , ist demzufolge $\frac{b-1}{\|w\|} - \frac{b+1}{\|w\|} = \frac{2}{\|w\|}$. Daraus folgt:

$$2d = \frac{2}{\|w\|}$$

Die gewünschte Maximierung des Trennbereichs entspricht der Maximierung von d , dies ist somit äquivalent zur Minimierung von $\|w\|$. Zur Vereinfachung der Berechnung wird die Funktion $z = \frac{1}{2}\|w\|^2$ minimiert, und zwar unter den Nebenbedingungen $y_i(wx_i + b) \geq 1$ für alle Trainingspunkte. Zur Lösung wird für jede Nebenbedingung ein Lagrange-Multiplikator $\alpha_i \geq 0$ eingeführt. Die optimalen Lagrange-Multiplikatoren können mittels quadratischer Optimierung bestimmt werden (Helmholz, 2012).

Bei dem Lagrange-Multiplikator gilt bei allen Stützvektoren $\alpha_i > 0$ und bei allen anderen $\alpha_i = 0$. Mit Hilfe der Stützvektoren kann der optimale Gewichtsvektor $w_{opt} = \sum_{i=1}^l \alpha_{opt,i} y_i x_i$ bestimmt werden. Dieser wird jedoch nie direkt berechnet, sondern wird genutzt um b mit den folgenden Formeln zu ermitteln:

$$b = \frac{1}{|S|} \sum_{i \in S} (y_i - w_{opt} x_i) = \frac{1}{|S|} \sum_{i \in S} \left(y_i - \sum_{i=1}^l \alpha_{opt,i} y_i x_i x_i \right)$$

Dabei gilt $S = \{i: \alpha_i > 0\}$, da hier nur noch die Stützvektoren mit einfließen.

Durch das Einsetzen des optimalen Gewichtsvektors in die Gleichung der Hyperebene $H = wx + b = 0$ entsteht die Entscheidungsfunktion $F(x)$ abhängig von dem Merkmalsvektor x des Bildprimitivs (Abe, 2010):

$$F(x) = \text{sgn} \left(\sum_{i \in S} \alpha_{opt,i} y_i x_i x + b \right)$$

(3.) Feature Space Mapping und Kernel-Trick

Bei Datensätzen, die in dem gegebenen Raum \mathbb{R}^k nicht linear trennbar sind, werden die Daten durch eine geeignete nichtlineare Abbildungsfunktion $\Phi(x) = (\Phi_1(x), \dots, \Phi_m(x))^T$ in einen höherdimensionalen Raum \mathbb{R}^m mit $m > k$ überführt. In dem höherdimensionalen Raum ist eine

Separierung durch die Hyperebene möglich. Die Gleichungen (s. o.) verlieren dadurch nicht ihre Gültigkeit, es muss lediglich x durch $\Phi(x)$ ersetzt werden. Die Entscheidungsfunktion lautet dann:

$$F(x) = \text{sgn} \left(\sum_{i \in S} \alpha_{opt,i} y_i \Phi(x_i) \Phi(x) + b \right)$$

Diese Abbildung in einen höher dimensionalen Raum wird Feature Space Mapping genannt.

Abbildung 3.6 zeigt den Vorteil des Feature Space Mapping am Beispiel eines Datensatzes, welcher im 2D-Raum nicht linear zu trennen ist. Durch eine Transformation in den 3D-Raum wird die lineare Trennung möglich. In der Abbildung ist ebenfalls zu sehen, dass die Trennebene im 3D-Raum in dem Eingangsraum zu einer nichtlinearen Trennfunktion führt.

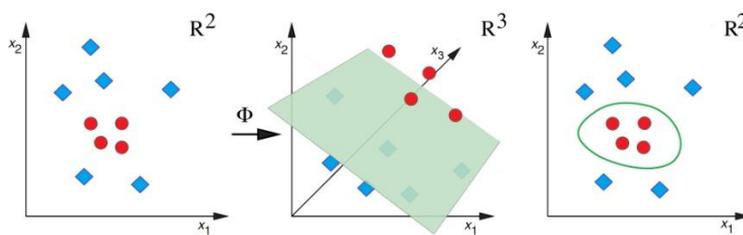


Abbildung 3.6

Lineare Trennung eines Datensatzes durch Überführung in einen höher dimensionalen Raum.
[Graf, 2009]

Diese Lösung hat jedoch auch Nachteile. Eine Transformation von großen Datenmengen in einen höherdimensionalen Raum kann sehr aufwendig sein und die Ermittlung einer geeigneten Abbildungsfunktion ist ebenfalls ein Problem. Aus diesem Grund wird der sogenannte Kernel-Trick (Graf, 2009) angewandt. Ist eine Kernfunktion $K(x, x_i) = \Phi(x_i) \Phi(x)$ vorhanden, ist eine explizite Berechnung von Φ nicht nötig, da $\Phi(x)$ im Training und in der Klassifikation ausschließlich in Form eines Skalarproduktes $\Phi(x_i) \Phi(x)$ auftritt und so durch K ersetzt werden kann. Die Eigenschaften, die eine Funktion haben muss, damit sie als Kernfunktion dienen kann, sind in Graf (2009) beschrieben. Der Trick liegt in der Ersetzung der Transformation und des Skalarproduktes durch die Kernfunktion. Diese ist im Vergleich einfacher als die ursprüngliche Berechnung. Die Entscheidungsfunktion $F(x)$ wird so zu

$$F(x) = \text{sgn} \left(\sum_{i \in S} \alpha_{opt,i} y_i K(x_i, x) + b \right)$$

vereinfacht.

Die Auswahl der Kernfunktion und der in ihr enthaltenen Parameter müssen entsprechend des Klassifikationsproblems gewählt werden. Zwei häufig genutzte Kernfunktionen werden im Folgenden genannt.

Der Lineare Kern wird bei linear trennbaren Trainingsdaten verwendet und ist als $K(x, z) = xz$ definiert. Bei nichtlinear trennbaren Datensätzen wird oft die Gauß-Kernfunktion $K(x, z) = e^{(-\gamma \|x-z\|^2)}$ genutzt, welche auch radiale Basisfunktion genannt wird. Die Trennebenen müssen hier

nicht zusammenhängen. Der Parameter $\frac{1}{2\gamma}$ beschreibt die Breite der Gaußfunktion und muss vor der Klassifikation festgelegt werden.

(4.) Datensätze mit Fehlern in den Trainingsdaten

Durch den Kernel-Trick mit einem Gauß-Kern ist es im Prinzip möglich, jeden Datensatz zu trennen. Bei überlappenden Clustern in den Trainingsdaten ist es notwendig, falsche Klassifizierungen von Trainingspunkten zuzulassen, um eine Überanpassung, wie in Abbildung 3.7 abgebildet, zu vermeiden.

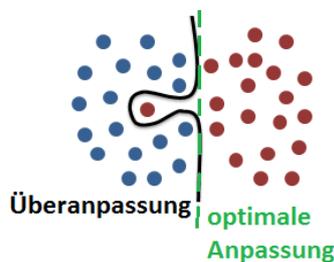


Abbildung 3.7

Überanpassung bei der Trennung von Klasse Blau und Rot

Zu diesem Zweck werden nicht negative Schlupfvariable (*slack variables*) ξ_i , $i = 1, \dots, l$ mit $\xi_i \geq 0$ eingeführt, die die Abweichung jedes Vektors zur Trennfläche angeben. Liegt der Vektor x_i auf der Seite der richtigen Klasse oder auf der entsprechenden Hyperebene E1 bzw. E2 der Stützvektoren, ist $\xi_i = 0$. Gilt $1 > \xi_i > 0$ befindet sich x_i in dem Trennbereich, aber auf der richtigen Seite der Trennebene, und bei $\xi_i = 1$ liegt x auf der Hyperebene H. Bei $\xi_i > 1$ liegt x auf der falschen Seite der Trennebene.

Die Erweiterte Bedingungsungleichung mit den Schlupfvariablen lautet: $y_i(wx_i + b) \geq 1 - \xi_i$.

Die Hyperebene und der Trennbereich soll jetzt so gewählt werden, dass möglichst wenig falsche Klassifizierungen in den Trainingsdaten vorliegen und dass der Trennbereich möglichst groß wird. Um das Ziel zu erreichen, muss die Funktion $\frac{\|w\|^2}{2} + C(\sum_i \xi_i)$ minimiert werden. Der Parameter C beschreibt die „Kosten“ für Punkte mit $\xi_i > 0$. Er reguliert einerseits die Größe des Trennbereichs und andererseits die Anzahl der Trainingsfehler und muss von dem Training bestimmt werden.

Zur Bestimmung der Entscheidungsfunktion $F(x)$ kann wie zuvor eine Optimierung unter Nebenbedingungen durchgeführt werden. Der einzige Unterschied in der Berechnung ist, dass nun die Lagrange-Multiplikatoren durch C begrenzt sind: $0 \leq \alpha_i \leq C$

(5.) Mehrklassenprobleme

Sind mehr als zwei Klassen vorhanden, muss die Klassifikation auf mehrere binäre Probleme aufgeteilt werden. Beispielsweise wird bei der One-Versus-All Methode (z.B. in Helmholtz, 2012)

jeweils eine Klasse isoliert und die anderen zu einer großen Klasse zusammengefasst, sodass wieder zwei Klassen entstehen. Jede Klasse wird einmal isoliert und es erfolgt jeweils eine SVM-Klassifikation. Für jedes Pixel wird dann gezählt, wie oft es der jeweiligen Klasse zugeordnet wurde und dann der Klasse zugewiesen, in die es mit SVM am häufigsten klassifiziert wurde.

Eine weitere Möglichkeit Mehrklassenprobleme mit dem SVM-Verfahren zu lösen ist die One-Versus-One-Methode (z.B. in Helmholz, 2012), bei der, im Gegensatz zur One-Versus-All Methode, alle möglichen Kombinationen der Klassen bei dem Training gegeneinander getestet werden.

(6.) Vor- und Nachteile

Das SVM-Verfahren erzielte in vielen Applikationen bereits gute Ergebnisse und dient daher häufig als Referenz beim Vergleich mit anderen Verfahren. Selbst bei wenigen Trainingsdaten entstehe ein gutes Ergebnis (Helmholz, 2012). Ein weiterer Vorteil der SVM ist laut Heinert et al. (2010) die Trennung nicht linear trennbarer Klassen durch die Verwendung des Kernel-Tricks. Durch den Strafterm C kann die SVM auch mit Ausreißern³ gut umgehen, obwohl zu beachten ist, dass die Ausreißer nicht explizit, durch die Bestimmung der Entfernung zu den anderen Vektoren, erkannt werden. Zur Ermittlung wird lediglich die Lage des Merkmalsvektors gegenüber der Trennfläche genutzt. Die Variabilität des Strafterms, sowie weitere Parameter in der genutzten Kernfunktion, ermöglicht außerdem die individuelle Anpassung des Verfahrens an das gegebene Problem. Die Parameter sind in den meisten Fällen nicht einfach zu bestimmen, weswegen sie oft mit Hilfe der Kreuzvalidierung aus den Testdaten bestimmt werden.

Das SVM-Verfahren besitzt auch verschiedene Nachteile wie z.B. der Rechenaufwand zur Bestimmung der Stützvektoren. Besonders gegenüber der ML-Methode ist das Training bei der SVM sehr rechenintensiv und durch die Kreuzvalidierung zur Bestimmung der Parameter verstärkt sich der Rechenaufwand zusätzlich. Ein weiteres Problem stellt der Übergang auf Mehrklassenprobleme dar, da die SVM-Methode nur für binäre Probleme konzipiert ist. Darüber hinaus besteht der Nachteil, dass kein direktes Qualitätsmaß abgeleitet werden kann (Rottensteiner, 2011).

3.4. Gebäudeerkennung mit der Dempster-Shafer-Theorie

Die Theorie von Dempster kombiniert Informationen aus unterschiedlichen Quellen zu einer Gesamtaussage. Sie wurde später von Shafer zu der Dempster-Shafer theory of evidence erweitert (Liu, 2012). Eine Besonderheit diese Theorie ist, dass auch Daten mit unvollständigem Wissen über

³ Als Ausreißer werden Merkmalsvektoren bezeichnet, die sich im Merkmalsraum weit entfernt von den anderen Merkmalsvektoren befinden.

deren statistische Verteilung genutzt werden können. Die Theorie, sowie die Methode zur Gebäudeerkennung nach Dempster-Shafer, ist in (Rottensteiner et al., 2005, 2007) ausführlich erläutert.

(1.) Prinzip der Dempster-Shafer-Theorie

Die Bildprimitive sollen in n Klassen C_n eingeteilt werden. Der Hypothesenraum θ beinhaltet alle Klassen C_n . Die Potenzmenge 2^θ enthält alle Teilmengen von θ , außer der leeren Menge \emptyset . Den einzelnen Elementen A von 2^θ wird eine Wahrscheinlichkeitsmasse $m(A)$ zugewiesen, sodass $0 \leq m(A) \leq 1$, $m(\emptyset) = 0$, und $\sum_{A \in 2^\theta} m(A) = 1$ ist. Wahrscheinlichkeitsmassen sind dabei keine Wahrscheinlichkeiten, da die Gleichung $m(C_i \cup C_j) = m(C_i) + m(C_j)$ für unabhängige $C_i, C_j, i \neq j$ nicht gelten muss. $m(A)$ drückt eher das Maß der Unterstützung der Daten für die Hypothese A aus. Es gibt verschiedene Möglichkeiten, um die Wahrscheinlichkeitsmassen festzulegen. Eine Möglichkeit ist z.B., diese aus den Wahrscheinlichkeiten vorheriger Klassifizierungen abzuleiten. Diese zuvor ermittelten Wahrscheinlichkeitsmassen können dabei durchaus zusammengesetzte Klassen in der eigentlichen Klassifikation zugewiesen werden. Dies hat den Vorteil, dass in den Daten einzelner Quellen nicht alle Klassen des Klassifizierungsproblems unterschieden werden müssen. Auch a posteriori Wahrscheinlichkeiten, die durch heuristische Methoden approximiert wurden, können genutzt werden.

Mit den Wahrscheinlichkeitsmassen können für jedes $A \in 2^\theta$ die Parameter Support $Sup(A)$ und Plausibility $Pls(A)$ berechnet werden. $Sup(A)$ ist die Summe aller Wahrscheinlichkeitsmassen, die direkt A zugewiesen werden: $Sup(A) = \sum_{B \subseteq A} m(B)$. $Pls(A)$ bezeichnet die Summe aller Wahrscheinlichkeitsmassen, die nicht der komplementären Hypothese \bar{A} von A zugewiesen werden: $Pls(A) = 1 - Sup(\bar{A}) = \sum_{B_{PL} \cap A \neq \emptyset} m(B_{PL})$. Die beiden Größen legen ein Unsicherheitsintervall fest, bei dem Support als Unter- und Plausibility als Obergrenze gesehen wird. Die Breite dieses Intervalls zeigt die Ungenauigkeit der Eintrittswahrscheinlichkeit von A .

Die Theorie nach Dempster-Shafer ermöglicht die Kombination der Wahrscheinlichkeitsmassen verschiedener Datenquellen. Dabei muss bei p Datenquellen für jede Datenquelle i mit $1 \leq i \leq p$ und für jede Klasse $B_j \in 2^\theta$ die Wahrscheinlichkeitsmasse $m_i(B_j)$ bestimmt sein. Die kombinierte Wahrscheinlichkeitsmasse kann wie folgt berechnet werden:

$$m(A) = \frac{\sum_{B_1 \cap B_2 \dots \cap B_p = A} (\prod_{1 \leq i \leq p} m_i(B_j))}{1 - \sum_{B_1 \cap B_2 \dots \cap B_p = \emptyset} (\prod_{1 \leq i \leq p} m_i(B_j))}$$

Die Summe im Nenner wird Konflikt genannt und spiegelt die Widersprüche in den Daten wieder.

Letztendlich muss noch eine Entscheidungsregel festgelegt werden zur Bestimmung von C_{max} als eine der einfachen Klassen $A \in \theta$. Häufig verwendet Regeln sind

- a) Maximum Support,
- b) Maximum Plausibility und
- c) Maximum Support, bei dem bei Überlappung der Unsicherheitsintervalle das Objekt als „nicht klassifizierbar“ gekennzeichnet wird.

(2.) Gebäudeerkennung mit der Dempster-Shafer-Theorie

Die Dempster-Shafer theory of evidence kann als ein Klassifikationsverfahren zur Gebäudeextraktion genutzt werden. Wie bei den zuvor beschriebenen Verfahren werden auch hier die beobachteten Merkmalen jedes Pixels genutzt um sie als Gebäude zu identifizieren. Zur Gebäudeerkennung in einem Bild müssen mindestens ein normalisiertes digitales Oberflächenmodell (nDOM) sowie der Normalized Difference Vegetation Index (NDVI) für das zu untersuchende Gebiet vorliegen. Außerdem können aus dem DOM Merkmale, welche mit der Oberflächenrauigkeit zusammenhängen, gewonnen werden. Dies ist jedoch nicht zielführend, wenn das DOM, z.B. aufgrund von Glättung oder einer schlechten Auflösung, keine Unterschiede der Oberflächenstruktur aufweist. Die Höhe aus dem nDOM und der NDVI bilden hier zusammen die Merkmale, anhand derer die Klassifikation erfolgt.

In einem ersten Klassifikationsprozess wird jedes Pixel des DOM unabhängig von den anderen einer der Klassen „Gebäude“(B), „Baum“(T), „Grasland“(G) und „vegetationsloser Boden“(S) zugeordnet. Jedes Merkmal alleine könnte die Klassen nicht unterscheiden, da das nDOM lediglich eine Höheninformation liefert (und daher $B \cup T$ von $G \cup S$ trennt) und der NDVI ausschließlich zur Identifikation von Vegetation dient (d.h. $B \cup S$ von $G \cup T$ trennt). Durch Kombination der Merkmale ist eine vollständige Klassifikation jedoch möglich. Die Dempster-Shafer-Theorie bietet ein einfaches Modell um die Merkmale zu verbinden.

Für die Klassifikation werden nun zwei Wahrscheinlichkeitsmassen, ein Wahrscheinlichkeitsmasse pro Merkmal, benötigt. Hier sind die Wahrscheinlichkeitsmassen $m_{B \cup T}(H)$ und $m_{G \cup T}(V)$, wobei H der Höhe aus dem nDOM entspricht und V dem NDVI-Wert. $m_{B \cup T}(H)$ zeigt die Unterstützung der Hypothese, dass ein Pixel mit der Höhe H der Klasse Gebäude oder Baum zugeordnet wird. $m_{G \cup T}(V)$ unterstützt die Hypothese, dass ein Pixel mit dem NDVI-Wert V einer Klasse mit hohem NDVI, also Grasland oder Baum, angehört.

Für die Modellierung der Wahrscheinlichkeitsmassen kann ein heuristisches Verfahren, wie in Abbildung 3.8 genutzt werden. Dabei gibt eine Wahrscheinlichkeitsmassefunktion die Wahrscheinlichkeitsmasse $m_i(x)$ für eine Gruppe von Klassen abhängig von der Merkmalsausprägung x an. Die dargestellte Wahrscheinlichkeitsmassefunktion wird durch eine

kubische Parabel zwischen zwei Wahrscheinlichkeitszustände P_1 und P_2 beschrieben. Die Wahrscheinlichkeitsmasse wird dann wie folgt bestimmt:

$$m_i(x) = \left\{ \begin{array}{ll} P_1 & \forall x | x \leq x_1 \\ P_1 + (P_2 - P_1) * \left[3 * \left(\frac{x - x_1}{x_2 - x_1} \right)^2 - 2 * \left(\frac{x - x_1}{x_2 - x_1} \right)^3 \right] & \forall x | (x > x_1) \wedge (x < x_2) \\ P_2 & \forall x | x \geq x_2 \end{array} \right\}$$

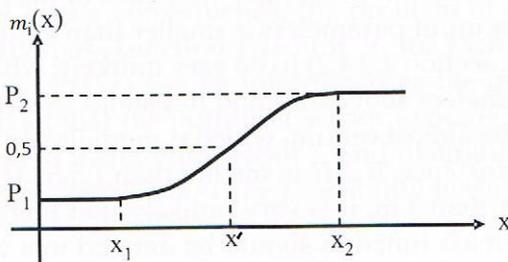


Abbildung 3.8

Bestimmung der Wahrscheinlichkeitsmassen $m_i(x)$ aus der der Merkmalsausprägung x

Alternativ ist z.B. auch eine lineare Verbindung möglich.

Rottensteiner et al. (2005) hat gezeigt, dass die Massenwahrscheinlichkeit allein von der Angabe des Parameters x' abhängt und so eine konkrete Ermittlung der Parameter P_1, P_2, x_1 und x_2 nicht nötig ist.

Bei der Gebäudeerkennung werden als x' , die geschätzte Gebäudehöhe für die Berechnung von $m_{BUT}(H)$ und eine Grenze für den NDVI-Wert zur Bestimmung von $m_{GUT}(V)$ genutzt.

Die Klassifikation erfolgt daraufhin durch die vordefinierte Entscheidungsregel, hier durch Maximum Support. Die Pixel der Klasse „Gebäude“ bilden dann ein binäres Bild, welches mit morphologischen Filtern geglättet wird. Dabei werden kleine Löcher in den Daten geschlossen (closing) und kleine Ausbuchtungen bzw. zu kleine Gebäude gelöscht (opening).

Anschließend erfolgt ein zweiter Fusionsprozess auf Basis von zusammenhängenden Segmenten von Gebäudepixel. Segmente der Klasse Gebäude, die jedoch eine zu kleine Fläche für ein Gebäude aufweisen, werden eliminiert.

In den binären Bildern werden abschließend die Randpixel von zusammenhängenden Segmenten bestimmt und mit dem Douglas-Peucker-Algorithmus durch Polygone approximiert. Zur Verbesserung der Approximation werden die Geraden am Rand der Polygone durch eine Ausgleichung bestimmt und die Schnittpunkte der Geraden ermittelt um ein vollständiges Polygon zu erhalten.

Ein Vorteil der Gebäudeerkennung nach Dempster-Shafer ist, dass nicht jeder Klasse eine Wahrscheinlichkeitsmasse zugewiesen werden muss, sondern auch zusammengesetzte Klassen genutzt werden können. So kann unvollständiges Wissen über die Verteilung der Daten modelliert werden.

Ein Nachteil ist, dass die Parameter zur Berechnung der Wahrscheinlichkeitsmassen bekannt sein müssen und nicht z.B. durch Trainingsdaten, erlernt werden.

4. Experimentelle Untersuchungen

4.1. Testdaten und Evaluierungsmethoden

Gegeben sind mehrere pan-geschärfte Farbinfrarot-Luftbilder aus zwei Befliegungen der Stadt Vaihingen in Baden-Württemberg. Der eine Datensatz besitzt eine Objektpixelgröße (GSD) von 20 cm, während die Daten der zweiten Befliegung eine GSD von 8 cm aufweisen. Die einzelnen Bilder besitzen eine radiometrischen Auflösung von 11 bit und wurden als 16-bit RGB TIF-files zur Verfügung gestellt. Die Aufnahmen entstanden im Sommer 2008 mit der digitalen Großformat-Luftbildkamera Integrgraph/ZI DMC mit einer Kammerkonstante von 120 mm durch die Firma RWE Power. Die Daten gehören zu einem größeren Block von Luftbildern, die von der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation (DGPF) zur Erprobung verschiedener digitaler Luftbildkameras genutzt wurden.

Die Bilder mit 20 cm GSD gehören zu einem Block, der aus drei überlappenden Streifen mit zwei zusätzlichen Querstreifen an den Enden besteht. Der zweite Block (8 cm GSD) beinhaltet fünf überlappende Streifen sowie ebenfalls zwei Querstreifen. Die Längs- und Querüberlappungen betragen in beiden Fällen 60%. Weiter Informationen zu den DGPF-Daten sind in (Cramer et al., 2010) zu finden.

In dem beflogenen Bereich befinden sich drei unterschiedlich charakterisierte Gebiete, die in dieser Arbeit als Testgebiete genutzt werden. Gebiet 1 zeigt die Stadtmitte mit einigen Bäumen, wenig Gras und vielen verschieden geformten Gebäuden. Gebiet 2 ist durch wenige, hohe Gebäude charakterisiert, die von vielen Bäumen und Grasland umgeben sind. Gebiet 3 zeigt eine Wohngegend mit freistehenden Einfamilienhäusern.

Für jeden Datensatz und jedes Gebiet wurden ein True Orthophoto und ein normalisiertes digitales Oberflächenmodell erstellt. In Tabelle 4.1 ist die Größe jedes Testgebietes zu sehen, und in Abbildung 4.1 sind die Gebiete innerhalb der Orthophotos gekennzeichnet.

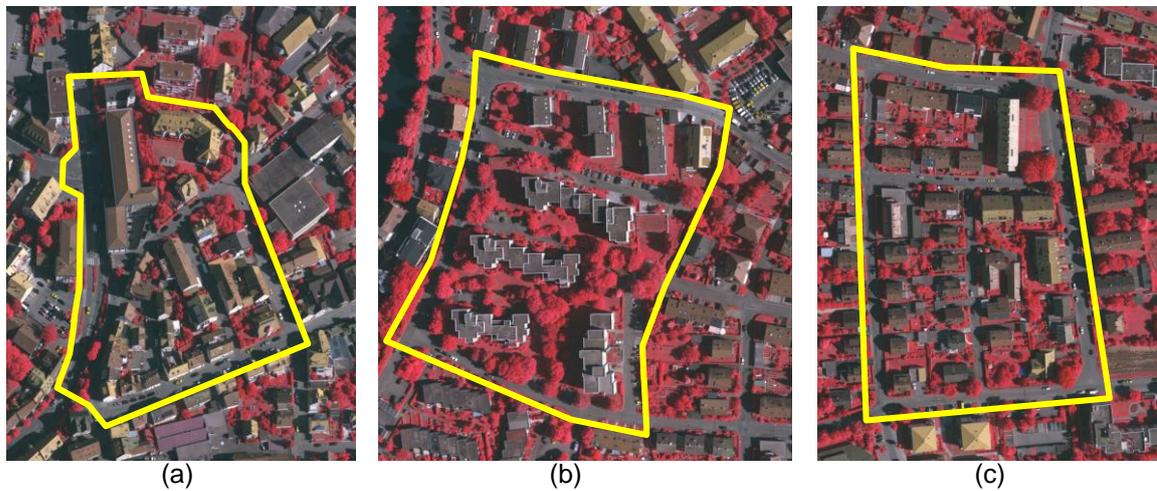
Tabelle 4.1 Größe der Testgebiete

<i>Gebiet</i>	<i>1</i>	<i>2</i>	<i>3</i>
<i>Größe[m*m]</i>	160 * 220	300 * 280	220 * 280

Abbildung 4.1 Testgebiete mit ihrer Begrenzung (gelbe Polygone)

(a) Gebiet 1 (b) Gebiet 2 (c) Gebiet 3

[ISPRS Projekt, 2011]



Diese Bilder wurden mit der Maximum-Likelihood- Klassifikationsmethode, dem Support-Vector-Machines-Verfahren und der Gebäudeerkennung nach Dempster-Shafer klassifiziert.

Für die Evaluierung wurden nur die Gebiete innerhalb der gelben Polygone in Abb. 4.1 berücksichtigt. Die Methode für die Evaluierung, sowie Referenzdaten, stammen aus dem ISPRS Test-Projekt zur Objekterkennung in urbanen Gebieten (Rottensteiner, 2011). Die Evaluierungsmethode vergleicht dabei die Klassifikationsergebnisse mit den Referenzdaten, die als fehlerfrei angesehen werden.

Die Pixel der klassifizierten Bilder sowie die der Referenzdaten dürfen für die Evaluierung lediglich zwei Zustände, Gebäude und Hintergrund, aufweisen. Die Evaluierungsmethode bestimmt dann True Positive (TP), False Negative (FN), False Positive (FP) und True Negative Pixel. Ein Pixel mit korrekter Zuweisung als Gebäude wird True Positive genannt. Ein False Negative Pixel wurde als Hintergrund klassifiziert, obwohl es in der Referenz als Gebäude gekennzeichnet ist und ein False Positive Pixel entspricht laut der Referenz dem Hintergrund, wurde jedoch als Gebäude klassifiziert. Ein True Negative Pixel ist sowohl im Referenzbild als auch bei der Klassifikation dem Hintergrund zugewiesen worden.

Diese Einteilung in TP, FN und FP wird ebenfalls für ganze Objekte durchgeführt. Dafür werden die einzelnen Pixel zu größeren Segmenten zusammengefasst und daraufhin topologisch bereinigt, sodass die extrahierten Gebäude die gleiche Topologie wie jene der Referenz aufweisen. Ein Objekt ist ein TP, wenn ein vorgegebener minimaler Prozentsatz seiner Fläche einem Gebäude in den Referenzdaten entspricht. Ein FN ist dementsprechend ein Objekt der Referenz, von dem weniger als dieser minimale Prozentsatz der Fläche durch die Klassifikation erkannt wurde, und ein FP ist ein

detektiertes Objekt, von dem weniger als ein bestimmter Prozentsatz der Fläche in der Referenz zu finden ist. Der vorgegebene Prozentsatz wird in dieser Arbeit auf 50 % festgesetzt.

Die TP, FN und TP Objekte/Pixel dienen zur Berechnung der Completeness, Correctness, und Quality (Rutzinger et al., 2009). Completeness, auch Producer's-Accuracy genannt, ist der prozentuale Anteil von Objekten/Pixel der Referenz, die durch die Klassifikation korrekt erkannt werden konnten. Correctness oder auch User's-Accuracy beschreibt, wie viel Prozent der Objekte/Pixel, die detektiert wurden, auch wirklich in diese Klasse gehören.

Quality verbindet beide Maße. Nur bei einem hohen Wert der beiden Maße ist auch die Quality hoch.

Die drei Qualitätsmaße werden wie folgt bestimmt:

$$Completeness = \frac{Anz. TP}{Anz. TP + Anz. FN}$$

$$Correctness = \frac{Anz. TP}{Anz. TP + Anz. FP}$$

$$Quality = \frac{1}{\frac{1}{Comp.} + \frac{1}{Corr.} - 1} = \frac{Comp.* Corr.}{Comp. + Corr. - Corr.* Comp.}$$

Die Qualitätsmaße werden sowohl auf Basis korrekt klassifizierter Pixel als auch auf Basis von Objekten angegeben. Bei der Evaluierung auf Basis konkreter Pixel werden die einzelnen Pixel gezählt, die bei dem Vergleich als TP, FN und FP klassifiziert wurden und in die obigen Gleichungen eingesetzt. Bei der Evaluierung auf Basis von Objekten werden die einzelnen Objekte als TP, FN und FP klassifiziert und gezählt. Die Qualitätsmaße auf Basis von Objekten werden zusätzlich als Funktion der Fläche ausgegeben.

Während des Evaluierungsprozesses werden auch die geometrischen Genauigkeitsmaße bestimmt. Dabei wird von jedem Randpunkt des klassifizierten Gebäudes der nächste Randpunkt eines Referenzgebäudes gesucht. Ist die Euklidische Distanz d zwischen den beiden Punkten größer als ein gegebener Schwellenwert von 3 m, wird der Punkt verworfen.

Mit den restlichen Euklidischen Distanzen d wird der RMS-Fehler der Distanzen (RMS_d) berechnet:

$$RMS_d = \sqrt{\frac{\sum d^2}{N}} \quad \text{mit } N = \text{Anzahl der nicht verworfenen Punkte.}$$

Der RMS_d -Fehler sowie die Anzahl der Randpunkte (RP) und die Anzahl der Randpunkte, bei denen ein Gegenstück gefunden wurde (RPG), werden von der Evaluierungsmethode ermittelt.

4.2. Vorgangsweise

Für jedes Gebiet wurde pro Auflösung ein true Orthophoto (TOP), der Normalized Difference Vegetation Index (NDVI) für jedes Pixel des TOP und ein normalisiertes digitales Oberflächenmodell (nDOM) erstellt. Die Klassifikation erfolgte bei allen Methoden pixelbasiert, wenn nicht anders erläutert.

Zur Generierung des DOM wurde eine Erweiterung der OpenCV Implementierung von „semiglobal matching“ für mehrfach überlappende Bilder genutzt, wie in Abschnitt 3.1.1 beschrieben. Für die Bestimmung des Disparitätsbildes D wird die Kostenfunktion von Birchfield und Tomasi (1998) genutzt und die Strafterme P_1 und P_2 des Glattheitskriteriums E_{glatt} wie in Tabelle 4.2 dargestellt gewählt. Die Bestimmung von D aus der Minimierung der Energiefunktion $E(D)$ erfolgt approximativ, indem die Kosten entlang eines 1D-Pfades aus fünf verschiedenen Richtungen akkumuliert werden. Am Ende wurde eine Segmentierung durch Region Growing vorgenommen. Die verwendeten Parameter S_{diff} , S_{seg} und Q_{max} (s. Abschnitt 3.1) werden, abhängig von der Auflösung, in Tabelle 4.3 gezeigt. Alle weiteren Parameter sind in OpenCV (2011) beschrieben.

Gebiet	P1	P2
1	36	56
2	12	36
3	12	45

Tabelle 4.2
Strafterme der Gebiete

Auflösung	8 cm	20 cm
S_{diff}	16	16
S_{seg}	300 Pixel	300 Pixel
Q_{max}	15%	25%

Tabelle 4.3
Weitere Parameter beim SGM

Zur Bestimmung des DGM wird ein Rangfilter verwendet. Die Berechnung des DGM erfolgte iterativ mit 2 Gebäudemaßen N_1 und N_2 als Größe des Filterfensters, wobei N_1 die maximale Ausdehnung des größten und N_2 die typische Ausdehnung eines kleineren Gebäudes repräsentiert. Die Werte sind in Tabelle 4.4 gegeben. Das Quantil für den Rangfilter s wurde mit 5 % festgesetzt. Alle Parameter bei der Berechnung des DGMs wurden unabhängig von der Auflösung der Bilder gewählt.

	Gebiet 1	Gebiet 2	Gebiet 3
N_1	65 m	65 m	65 m
N_2	20 m	10 m	20 m

Tabelle 4.4
Gebäudemaße der Gebiete

Mit dem nDOM und dem TOP als Grundlage wurden drei verschiedene Klassifikationsverfahren unabhängig voneinander durchgeführt. Es wurde versucht, dass alle Klassifikationsmethoden die gleichen Informationen zur Verfügung haben. Aufgrund der verschiedenen Ansätze und der damit einhergehenden unterschiedlichen Variablen und Parameter war dies nicht immer möglich.

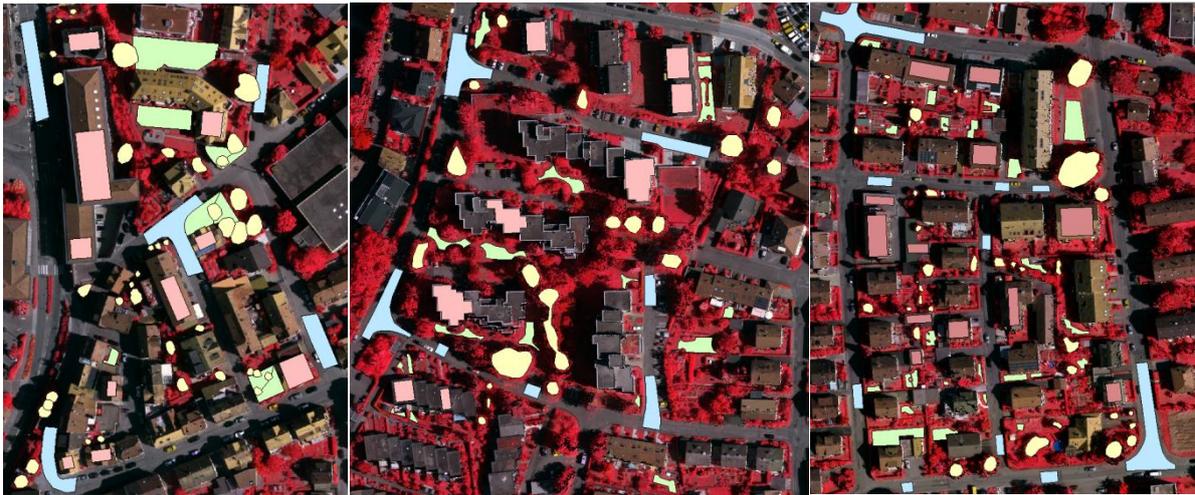
Die Maximum-Likelihood-Klassifikation (ML) wurde mit der Software ENVI durchgeführt.

Als Trainingsgebiete wurden in jedem der drei gegebenen Gebiete mit ArcGIS einige Objekte den Klassen „Gebäude“(B), „Straße“(S), „Grasland“(G) und „Baum“(T) manuell zugeordnet (s. Abb. 4.2).

Abbildung 4.2 Trainingsgebiete

Rot = Gebäude, Blau = Straße, Grün = Vegetation, Gelb = Baum

Reihenfolge: links Gebiet1, Mitte Gebiet2, rechts Gebiet3



Das ML-Verfahren wurde, wie in Abschnitt 3.2 beschrieben, durchgeführt. Als Merkmale dienten die Höhen aus dem nDOM, die drei Kanäle des TOP und NDVI, welcher aus dem roten und infraroten Kanal des TOP berechnet wurde. Dementsprechend wurden jedem Pixel fünf Merkmale zugeordnet, anhand derer die Klassifikation durchgeführt wurde. Die Ergebnisse wurden als Labelbilder gespeichert, bei dem der Grauwert dem Klassenlabel entspricht. Aus diesem Labelbild wurde ein binäres Bild abgeleitet, in dem Gebäudepixel weiß und Hintergrundpixel schwarz dargestellt sind. Dieses wurde durch morphologische Filter geglättet.

Die Support-Vector-Machines-Klassifikation (SVM) erfolgte mit einer flexiblen SVM-Implementierung des Instituts für Informationsverarbeitung der Leibniz Universität Hannover (Büschendorf, 2012).

Die Trainingsgebiete waren die gleichen wie bei der ML-Klassifikation.

Die Menge an Trainingspixel wurde, um die Trainingszeit zu verkürzen, verringert, indem bei den Bildern mit 8 cm Auflösung nur jedes dritte Pixel der Trainingsgebiete genommen wird. Außerdem wurde bei den Trainingsregionen ein fünf Pixel breiter Puffer nach innen gelegt. Dieser Bereich wird nicht zum Training verwendet, um Mischpixel oder Pixel von falschen Klassen in den Trainingsdaten zu vermeiden.

Bei den Bildern mit der GSD von 20 cm wurde aufgrund der ohnehin geringeren Datenmenge auf eine Reduzierung der Trainingsdaten verzichtet. Lediglich der Rand der Trainingsdaten wurde wie bei

den hochaufgelösten Daten nicht genutzt, um auch hier die Verwendung von Mischpixeln zu vermeiden. Allerdings wurde die Breite der Pufferzone am Rand der Trainingsdaten, in der die Daten nicht verwendet werden, auf drei Pixel reduziert.

Als Grundlage für die Merkmalsbestimmung dienten das nDOM, das TOP und ein NDVI-Bild. Vor dem Training wurde bei der höheren Auflösung zuvor eine Glättung aller Daten durch einen Gaußfilter mit der Standardabweichung $\sigma = 10$ durchgeführt. Bei der geringeren Auflösung wurden lediglich das TOP und das NDVI-Bild geglättet.

Der Merkmalsvektor jedes Pixels beinhaltet neun verschiedene Werte. Aus dem nDOM bzw. dem geglätteten nDOM wurde ein Höhenwert als der Median in einer Nachbarschaft von 5x5 Pixeln ermittelt.

Aus den drei Kanälen des TOP und dem NDVI-Bild wurden ebenfalls aus einer 5x5 Nachbarschaft der Median und die Standardabweichung berechnet, sodass acht weitere Merkmale entstehen.

Der Strafterm C wurde mit 10^3 manuell festgelegt. Zusätzlich muss noch ein Parameter der Kernfunktion angegeben werden. Bei der genutzten Gauß-Kernfunktion ist dieser Parameter γ , welcher aus der Dimension k des Merkmalsvektors durch die Gleichung $\gamma = \frac{1}{k}$ bestimmt wird. Die Werte der Parameter wurden vom Institut für Informationsverarbeitung für eine normale Anwendung empfohlen.

Die SVM Klassifikation erfolgte wie in Abschnitt 3.3 beschrieben. Die Ergebnisse wurden als Labelbilder, bei dem der Grauwert dem Klassenlabel entspricht, gespeichert. Aus ihnen wurde ein binäres Bild abgeleitet, in dem Gebäudepixel weiß und Hintergrundpixel schwarz dargestellt sind.

Die Gebäudeerkennung nach Dempster-Shafer wurde mit der Software Barista durchgeführt. Die Rauigkeit und Isotropie des DOM zeigten bei Versuchen keinen großen Einfluss bei der Klassifikation der Testbilder, da bei dem Matching zur Erstellung des DOM eine Glättung stattfand. Besonders Bäume weisen daher keine starke Rauigkeit auf, im Gegensatz zu hoch aufgelösten Laserscanningdaten. Aus diesem Grund wurde die Rauigkeit der Oberfläche nicht als Merkmal genutzt. Zur Regulierung des Einflusses des NDVI wurde dessen Standardabweichung mit einbezogen (Rottensteiner et al, 2007).

Im Gegensatz zu den anderen Verfahren werden bei dieser Methode die Informationen aus dem Orthophoto nur indirekt über den NDVI mit einbezogen. Die Merkmale sind daher lediglich die Höhe aus dem nDOM sowie der NDVI.

Mit einem heuristischen Verfahren wurden aus den geschätzten minimalen Gebäudehöhen $x' = \Delta H_B$ und einer Grenze für den NDVI-Wert $x' = NDVI_G$ die Wahrscheinlichkeitsmassen bestimmt. Die verwendeten Parameter sind in Tabelle 4.5 aufgelistet. P1 ist mit 0,05 und P2 mit 0,95 festgelegt. Der Abstand von x_1 und x_2 zu x' ist von der Software festgelegt worden.

Auflösung	8 cm			20 cm		
Gebiet	1	2	3	1	2	3
ΔH_B [m]	2,5	3,5	2,2	2,5	4,5	2,2
$NDVI_G$ [%]	20	15	15	15	15	15

Tabelle 4.5

Parameter zur Bestimmung der Wahrscheinlichkeitsmassen

Die Ergebnisse wurden durch morphologisches Opening, gefolgt von einem morphologischen Closing, geglättet. Für beide wurden quadratische Strukturelemente mit einer Seitenlänge von ca. 0,5 m, entsprechend drei Pixel bei den Bildern mit 20 cm GSD und 7 Pixel bei einer GSD von 8 cm, verwendet.

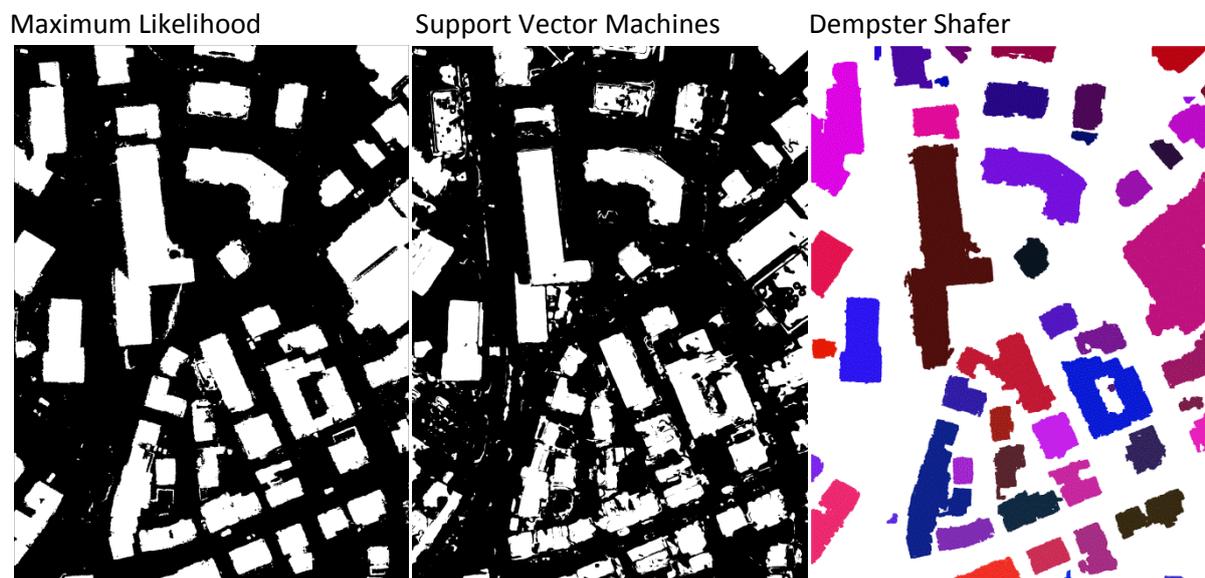
Die Ergebnisse wurden als Labelbilder abgespeichert, bei denen der Hintergrund weiß dargestellt ist und den Gebäuden zufällig Farbwerte zugeordnet wurden.

Am Ende wurden alle Ergebnisse mit der Evaluierungsmethode aus (Rutzinger et al., 2009) ausgewertet.

4.3. Ergebnisse und Vergleich

Im Anhang sind das Orthophoto, nDOM und NDVI für jedes Gebiet in beiden Auflösungen abgebildet, sowie die Ergebnisse der Klassifikation. Bei den Ergebnissen der SVM- und ML- Methode sind die Gebäude weiß dargestellt und der Hintergrund schwarz. Bei der DS-Methode sind die Gebäudepixel farbig und der Hintergrund weiß. In Abbildung 4.3 sind die Ergebnisse beispielhaft für Gebiet 1 mit 8 cm GSD abgebildet.

Abbildung 4.3 Labelbilder der Klassifikationen für Gebiet 1 mit 8 cm GSD



Der Vergleich erfolgt unter verschiedenen Gesichtspunkten. Als erstes wird die Dauer der einzelnen Klassifikationen verglichen, gefolgt von der Interpretation der Evaluierungsergebnisse. Dabei werden mögliche Gründe für die Fehlklassifikation einzelner Gebäude genannt und auf Bereiche, die fälschlicherweise als Gebäude klassifiziert wurden, eingegangen. Im Anschluss werden Correctness, Completeness und Quality sowie die geometrischen Fehler analysiert.

4.3.1. Bearbeitungszeit der Klassifikationen

Die benötigte Zeit für eine Klassifikation ist sehr verschieden, abhängig von dem Verfahren, der Bildauflösung, sowie der Größe der Trainings- und Testgebiete. Bei der SVM spielt auch die Trennbarkeit der einzelnen Klassen im Merkmalsraum eine große Rolle, da dies die Anzahl der Stützvektoren beeinflusst.

Die Durchführung der Maximum-Likelihood-Methode benötigte ein paar Sekunden pro Bild und daher die geringste Zeit für die Klassifikation und das Training.

Die Klassifikation eines Bildes mit dem SVM-Verfahren bedarf deutlich mehr Zeit. Die Klassifikations- und Trainingszeiten sind in Tabelle 4.6 dargestellt.

<i>GSD</i>	<i>8 cm</i>			<i>20 cm</i>		
<i>Gebiet</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>1</i>	<i>2</i>	<i>3</i>
<i>Training</i>	7	5	2	7	10	2
<i>Klassifikation</i>	79	159	34	10	31	5
<i>Summe</i>	86	164	36	17	41	7

Tabelle 4.6

Benötigte Zeiten für die SVM-Klassifikation in Minuten

Bei dem Vergleich der Klassifikationszeiten der beiden Auflösungen fällt auf, dass die Klassifikation bei den Bildern mit hoher Auflösung deutlich länger braucht. Dies liegt daran, dass bei den Bildern mit 8 cm GSD mehr als viermal so viele Pixel klassifiziert wurden wie bei den Bildern mit 20 cm GSD.

Bei konstanter GSD und steigender Größe des abgebildeten Gebietes (Tabelle 3.1) erhöht sich ebenfalls die Anzahl der Pixel. Daraus folgt eine Korrelation zwischen der Klassifikationszeit und der Größe des Gebietes. Bei dem Vergleich fällt jedoch auf, dass Gebiet 1 kleiner ist als Gebiet 3, obwohl die Zeit der Klassifikation von Gebiet 1 deutlich größer ist. Der Zusammenhang zwischen der Klassifikationszeit und der Anzahl der Stützvektoren, die bei der Klassifikation genutzt wurden (Tabelle 4.7), kann diese Unstimmigkeit erklären. In beiden Auflösungen erfolgte die Klassifikation von Gebiet 3 mit deutlich weniger Stützvektoren als die Klassifikation der beiden anderen Gebiete, die mit etwa gleich vielen Stützvektoren klassifiziert wurden.

Tabelle 4.7 Anzahl der Stützvektoren (gerundet)

8 cm GSD			20 cm GSD		
Gebiet 1	Gebiet 2	Gebiet 3	Gebiet 1	Gebiet 2	Gebiet 3
11000	9000	2500	9000	11500	2500

Die Anzahl der Pixel in den Trainingsgebieten hat auch eine erhebliche Wirkung auf die Länge des Trainingsprozesses. Aus diesem Grund wurde das Trainingsgebiet bei 8 cm GSD ausgedünnt (vgl. Abschnitt 4.2). Ohne diese Verkleinerung hätte z.B. das Training von Gebiet 1 mit 8 cm GSD ungefähr 15,5 Stunden benötigt.

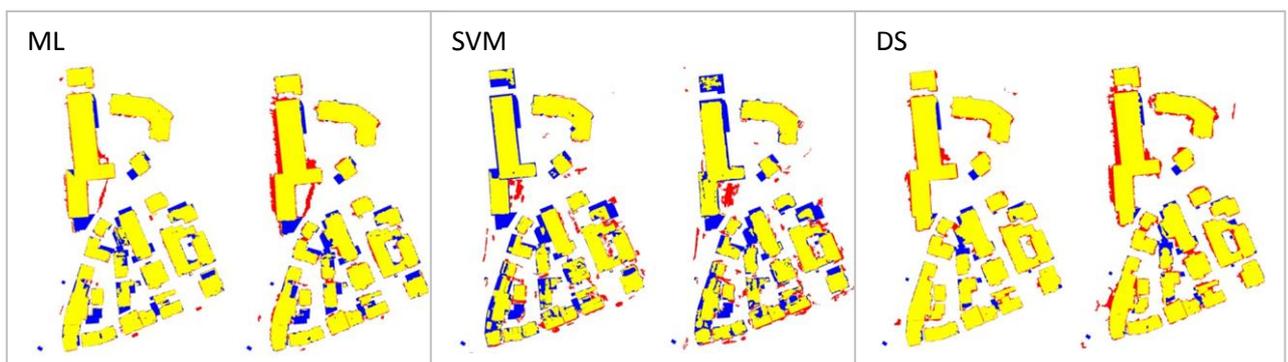
Die Gebäudeerkennung mit Dempster-Shafer verlief im Gegensatz zur SVM-Klassifikation schnell. Bei den Bildern mit der GSD von 8 cm vergingen mehrere Minuten, während ein Bild mit der GSD von 20 cm auch hier nach einigen Sekunden klassifiziert war.

4.3.2. Visueller Vergleich der Evaluierungsergebnisse

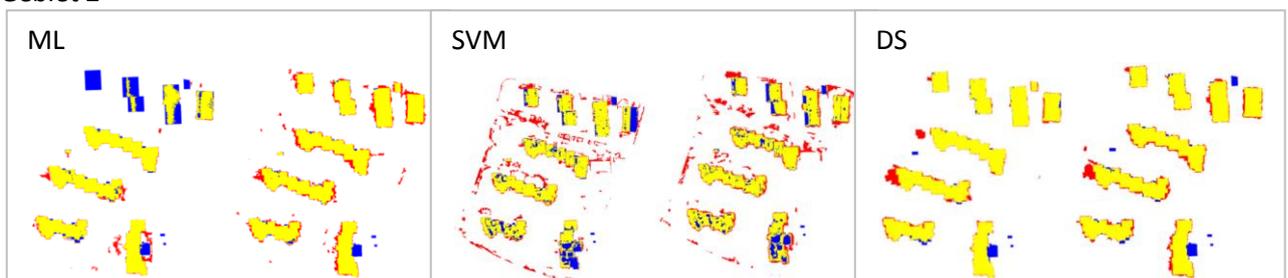
In den Evaluierungsbildern (Abb. 4.4) sind TP, FP und FN Pixel (vgl. Abschnitt 4.1) gekennzeichnet. Gelb entspricht einem TP, blau einem FN und rot einem FP. Falsch klassifizierte Pixel befinden sich gehäuft an den Rändern der Klassen. Dies lässt sich z.T. auf das Matching zurückführen, welches eine leichte Glättung des DOMs verursacht und so zu Fehlern im nDOM und true Orthophoto führt. Auf größere Ballungen von falsch klassifizierten Pixeln wird im Folgenden genauer eingegangen.

Abbildung 4.4 Visualisierung der Ergebnisse der Evaluierung: GSD: links 8 cm, rechts 20 cm

Gebiet 1



Gebiet 2



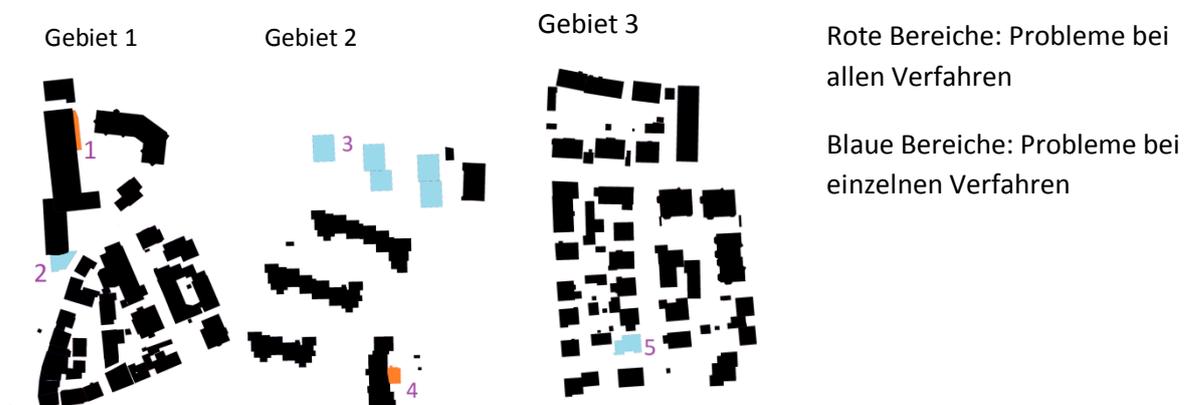
Gebiet 3



a) FN Pixel

False negative Pixel (Abb. 4.4, blau) sind Pixel, die in der Referenz als Gebäude klassifiziert wurden, aber von dem Klassifikationsverfahren nicht als solche erkannt werden. Einige ausgewählte Gebäude und Gebäudeabschnitte, die größtenteils nicht erfasst wurden, sind in Abbildung 4.5 zu sehen und mit Zahlen gekennzeichnet. Sie zeigen jeweils ein Problem bei der Klassifikation.

Abbildung 4.5 Schwer zu klassifizierende Gebäude



Ein großes Problem bei allen Klassifikationsverfahren scheinen bewachsene Gebäude zu sein. Sind Gebäude mit Vegetation bedeckt, weisen die einzelnen Pixel die Merkmale eines Baumes auf, also einen hohen NDVI und eine große Höhe. Um diese Gebäude von Bäumen unterscheiden zu können müsste z.B. die Umgebung mit einbezogen werden oder die Rauigkeit mit einfließen, da die meisten Gebäude mit Vegetation flach sind, sodass die umliegenden Pixel eine ähnliche Höhe aufweisen, und die Rauigkeit in hochaufgelösten Daten nicht so stark ist wie bei Bäumen. In Abbildung 4.6 sind das TOP und das NDVI-Bild mit 8 cm Auflösung beispielhaft für Gebäude 4 (in Abb. 4.5) gezeigt. Es ist deutlich zu sehen, dass der NDVI den umliegenden Bäumen gleicht und nicht den anderen Gebäuden.

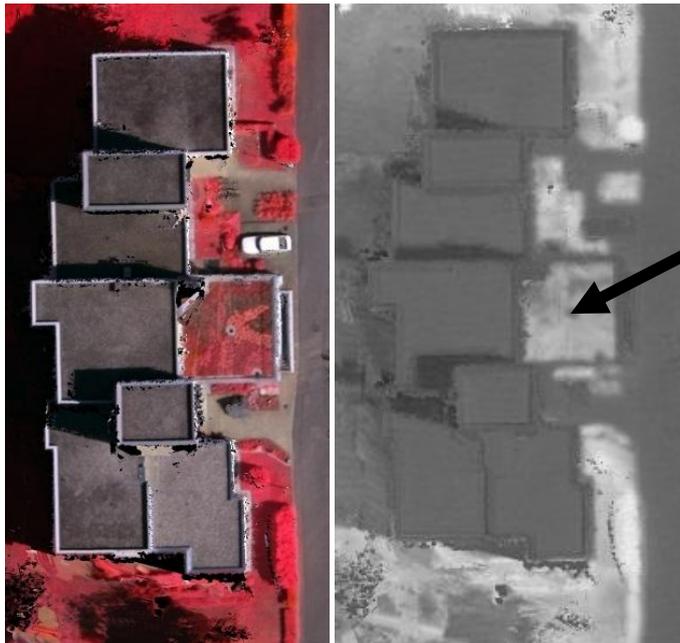


Abbildung 4.6

TOP und NDVI eines Gebäudes, welches teilweise mit Vegetation bedeckt ist. Beide Bilder besitzen eine Auflösung von 8 cm. (Gebäude 4 in Abb. 4.5)

Ein weiteres Problem bereiten kleinere Gebäude, wie z.B. das Gebäude links unten in Gebiet 1. Besonders bei groben Auflösungen kann es vorkommen, dass nicht genug Pixel vorhanden sind, um ein Gebäude zu repräsentieren. Wenn Teile des Gebäudes verdeckt sind, wird die Anzahl der Gebäudepixel weiter verkleinert. Um kleine Klassifikationsfehler zu eliminieren, werden nach dem ML-Verfahren und der Methode nach Demster-Shafer die Klassifikationsergebnisse mit morphologischen Filtern bearbeitet. Bei der SVM-Klassifikation wurden die Eingangsbilder schon vor der eigentlichen Klassifikation geglättet. Dies hat jedoch zur Folge, dass nicht nur kleine Fehlklassifikationen, sondern auch kleine Gebäude herausgefiltert werden.

Schattenbereiche sind ebenfalls problematisch, da in diesen Bereichen weniger Licht auf den Fotosensor trifft, sodass diese Bereiche kaum Informationen enthalten. Aus diesem Grund können Fehler im Matching entstehen, welche Fehler in den nDOM und TOP zur Folge haben. Ein Beispiel zeigt das Gebäude 1 (in Abb. 4.5), welches auf der rechten Seite einen Schatten wirft. An der Stelle des Schattens weist das nDOM fehlerhafte Höheninformationen auf, was sich auch auf das TOP auswirkt (Abb.4.7). Selbst bei korrektem Matching sind Schattenbereiche problematisch, da dort z.B. der NDVI nicht präzise bestimmt werden kann.

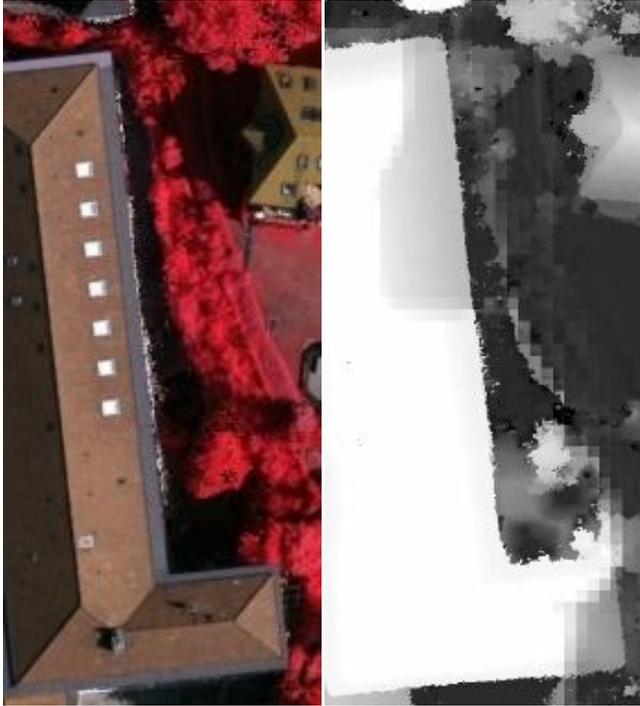


Abbildung 4.7

Matchingfehler im Schattenbereich, gezeigt an einem TOP und nDOM mit einer Auflösung von 8 cm. (Gebäude 1 in Abb. 4.5)

Das ML- und SVM-Verfahren scheinen in Gebiet 1 Probleme mit niedrigen Flachdächern zu haben, wie an Gebäude 2 (in Abb. 4.5) zu sehen ist. Da bei den überwachten Klassifikationsverfahren, im Gegensatz zur Methode nach Dempster-Shafer, Trainingsdaten genutzt werden, kann vermutet werden, dass die Fehlklassifikation aufgrund der Wahl der Trainingsdaten auftreten. Die Merkmale der meisten Trainingsdaten unterscheiden sich stark von denen der Pixel der Flachdächer. Viele der Gebäude der Trainingsdaten weisen eine andere Reflexionseigenschaft der Dächer auf, was in dem Orthophoto (Abb. 4.8) zu sehen ist.



Abbildung 4.8

TOP, nDOM und NDVI (alle 8 cm GSD) mit zwei Gebäuden. (Gebäude 2 in Abb. 4.5)

Oben: kein Flachdach

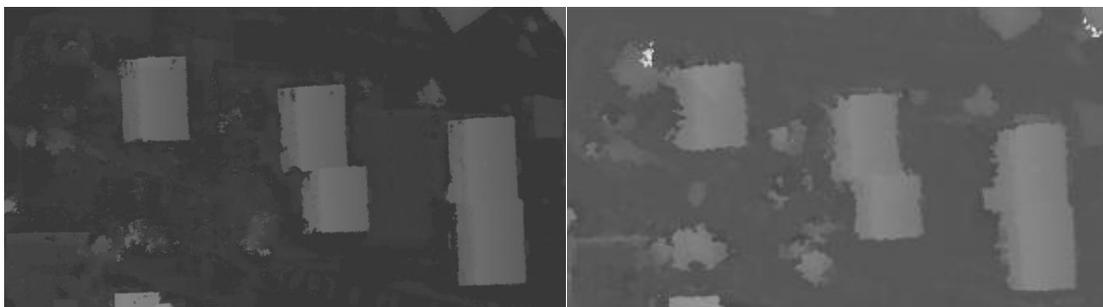
Unten: Flachdach

Die drei Gebäude im oberen Bereich von Gebiet 2 (Nr. 3 in Abb. 4.5) konnten besonders mit dem ML-Verfahren bei der 8 cm Auflösung nicht oder kaum erfasst werden. Da bei dem rechten der drei Gebäude nur die tieferliegenden Ränder nicht erfasst werden konnten, liegt das Problem vermutlich an der Höhe. Die meisten Gebäude in dem Gebiet sind sehr viel höher, und obwohl das kleinste der drei Gebäude (links) in den Trainingsdaten vorhanden ist (Abb.4.2), ist der Mittelwert der Gebäudehöhen, welcher aus dem Training des ML-Verfahren ermittelt wird, größer. Bei diesem Gebiet mit zwei stark verschiedenen Gebäudehöhen müsste für das ML-Verfahren vermutlich die Klasse „Gebäude“ in die zwei Klassen „niedrige Gebäude“ und „hohe Gebäude“ aufgeteilt werden, um auch die kleinen Gebäude zu erfassen. Diese Hypothese wird dadurch unterstützt, dass das SVM-Verfahren, welches mehrere Cluster im Merkmalsraum innerhalb einer Klasse modellieren kann, auch diese niedrigen Gebäude fast genauso vollständig detektiert hat wie die hohen. Die gleichen Gebäude wurden in den Daten mit 20 cm Auflösung richtig klassifiziert. Es fällt jedoch auf, dass das nDOM mit 20 cm Auflösung gehäuft Höhenfehler in der Region aufweist, wie in Abb. 4.9 zu sehen ist.

Abbildung 4.9

nDOM mit 8 cm Auflösung (links) und 20 cm Auflösung (rechts)

(Gebäudegruppe 3 in Abb. 4.5)



Das Gebäude 5 (in Abb. 4.5) wurde ausschließlich mit dem Dempster-Shafer-Verfahren im 8 cm Bild nicht erkannt. Im 20 cm Bild wurde das Gebäude richtig klassifiziert. Bei dem Vergleich der Merkmale der unterschiedlichen Auflösungen fällt auf, dass in dem Orthophoto und nDOM mit 8 cm Auflösung Fehler sind (Abb. 4.10). Dieser Fehler wirkt sich auf die Erstellung des Orthophoto aus, sodass in dem Orthophoto Pixel entstehen, in dem keine Informationen gespeichert sind. Diese Fehler sind bei der 20 cm Auflösung nicht zu finden. Der Grund, warum das gesamte Gebäude nicht als solches erkannt wurde, liegt in der Größe der Teilstücke, die korrekt klassifiziert werden könnten. Die meisten dieser Stücke sind so klein, dass sie vermutlich durch das Postprocessing herausgefiltert wurden. Das einzige größere Stück, welches weder im Orthophoto noch im nDOM beschädigt ist, liegt in der oberen rechten Ecke. Diese ist aber teilweise im Schatten, da sich in der Mitte der Dachfläche eine

Gaube befindet, sodass dort der NDVI nur unsicher bestimmbar ist, weshalb es in diesem Bereich vermutlich zu einer Fehlklassifikation gekommen ist.



Abbildung 4.10

Fehlerhaftes TOP und nDOM mit 8 cm Auflösung.
(Gebäude 5 in Abb. 4.5)

a) FP Pixel

False positive Pixel sind Pixel, die fälschlicherweise als Gebäude klassifiziert wurden. In Gebiet 1 treten nur um das große Gebäude in der linken oberen Ecke (Abb. 4.11) bei allen Verfahren vermehrt FPs auf.

Abbildung 4.11

Bereiche mit FP Pixel in Gebiet 1 und Orthophoto-, nDOM- und NDVI-Ausschnitt dieses Bereichs mit 20 cm GSD.



Bei den Ergebnisbildern aus der ML- und Dempster-Shafer Klassifikation der Bilder mit 20 cm GSD befinden sich am linken Rand des Gebäudes viele Fehlklassifikationen (Abb. 4.11, Nr.1). Diese zusätzlichen Pixel sind vermutlich aufgrund des Schattens, den das Haus wirft, entstanden (Abb.4.11, Orthophoto). Wie bereits oben erläutert entstehen in Schattenbereichen oft Matchingfehler und somit Fehler im nDOM. Die Merkmale in Schattenbereichen sind oft unsicher oder werden aus den

Merkmale der Umgebungspixel berechnet (vgl. 2.2.1). In diesem Fall sind die berechneten Werte eine große Höhe und ein geringer NDVI (Abb.4.11, nDOM, NDVI), was auf ein Gebäude hinweist. Auf der rechten Seite des Gebäudes (Abb. 4.11, Nr.2) sind es die Schatten der Bäume, die die Bestimmung des NDVI und der Höhe beeinflussen und zu einer Fehlklassifikation mit dem ML- und Dempster-Shafer-Verfahren führen. Bei dem SVM-Verfahren sind die Merkmale von Randbereichen nicht beim Training genutzt worden, da die Ränder der Trainingsgebiete zur Vermeidung von Mischpixeln entfernt wurden. Dadurch werden fehlerhafte Randbereiche nicht als Gebäude erkannt. Der Nachteil ist jedoch, dass auch niedrige Teile eines Gebäudes ggf. nicht erkannt werden. Dass auch das SVM-Verfahren Probleme mit Schattenbereichen hat, zeigt das Gebiet im Schatten der Mauer (Abb. 4.11, Nr.4). Dies ist in beiden Auflösungen der Klasse Gebäude zugeordnet worden. Auch Mauern (Abb.4.11, Nr.3) können ein Problem darstellen, da sie, wie Gebäude, eine große Höhe und einen geringen NDVI aufweisen. Das ML-Verfahren hat daher diesen Bereich als Gebäude klassifiziert. Bei der Gebäudeerkennung nach Dempster-Shafer ist die Mauer aufgrund ihrer kleinen Breite durch die Filter morphologisch eliminiert worden. Auch das SVM-Verfahren hat die Mauern im 8 cm GSD Bild korrekterweise nicht als Gebäude klassifiziert. Da die Trainingsgebiete der anderen Klassen alle eine geringe Höhe aufweisen, die Mauer in dem nDOM jedoch einen großen Wert, kann vermutet werden, dass die Mauer durch die Glättung des Gaußfilters (vgl. 4.2) der Umgebung angepasst wurde. Diese Hypothese wird bestätigt durch die Klassifikation des Bildes mit 20 cm GSD, wo ein Großteil der Mauer als Gebäude klassifiziert wurde. Dort fand keine Glättung des nDOM statt.

In Gebiet 2 gibt es ebenfalls Probleme mit Schattenbereichen, besonders im Schatten der Hochhäuser (Bsp. Abb. 4.12). In diesem Fall befinden sich dort Bäume, die aber aufgrund des Schattens einen ungenauen NDVI besitzen und daher der Klasse Gebäude zugeordnet werden.

Abbildung 4.12

Bereiche mit vermehrten false positive Pixel in Gebiet 2 und Orthophoto-, nDOM- und NDVI-Ausschnitt dieses Bereichs mit der GDS 20 cm.



In Gebiet 3 gibt es bei der Dempster-Shafer-Gebäudeerkennung nur wenige FP Pixel, die meistens an Rand- und Schattengebieten liegen.

Auffällig ist, dass besonders in Gebiet 3 oft Büsche am Straßenrand von der SVM- und ML-Klassifikation als Gebäude klassifiziert wurden. Ein möglicher Grund könnte die Wahl der Trainingsdaten sein. Da sich in ihnen keine Büsche befinden, konnte das Klassifikationsverfahren dies nicht beim Training erlernen. Es könnte demnach sinnvoll sein, die Trainingsdaten und bzw. oder die Menge der Klassen zu erweitern. Hinzu kommt das Problem des Datenverlustes bei der Abtastung und Quantisierung der Höhen. Da die Implementierungen der beiden überwachten Methoden nur mit 8-bit-Grauwertbildern arbeiten, mussten die Höhen des nDOMs auf 256 Höhenwerte reduziert werden. Dies erfolgte durch eine lineare Skalierung von H_{min} bis H_{max} auf die Grauwerte von 0 bis 255. Da das nDOM auch einige Ausreißer beinhaltet, wurde ein Teil der Grauwerte nicht genutzt. Dies hat zur Folge, dass die größeren Büsche an den Rändern der Straße in etwa den gleichen Höhenwert besitzen wie ein niedriges Gebäude (Abb. 4.13).

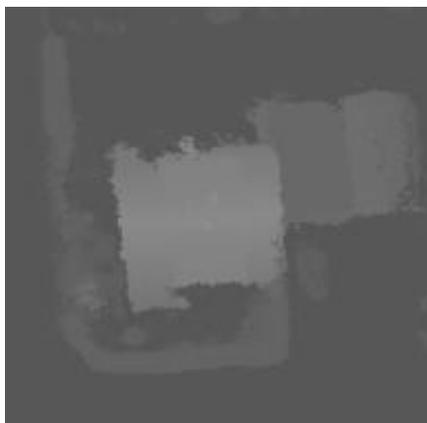


Abbildung 4.13

Folgen der Quantisierung am Beispiel eines nDOM-Ausschnittes mit 8 cm Auflösung

	Straße	Büsche	niedriges Haus	hohes Haus
Durchschnittlicher Grauwert der Abbildung	87	100	104	132

4.3.3. Completeness, Correctness, Quality

Die drei Qualitätsmaße Completeness, Correctness und Quality, werden aus der Anzahl der true positive (TP), false negative (FN) und false positive (FP) Objekte bzw. Pixel, wie in Abschnitt 4.1 beschrieben, berechnet.

Als erstes werden die Qualitätsmaße betrachtet, die sich auf die korrekt klassifizierte Fläche beziehen (Abb.4.14). Bei den Ergebnissen für 8 cm GSD liegt die Completeness im Durchschnitt bei 86 %. Die Correctness ist mit einem Mittelwert von 83 % etwas schlechter, das Mittel der Quality liegt bei 74 %. Die Completeness der klassifizierten Bilder mit 20 cm GSD ist mit einem Durchschnitt von 89 % sogar etwas besser. Die Correctness und Quality sind mit einem Durchschnitt von 80 % und 7 % etwas schlechter.

Die Methode nach Dempster-Shafer besitzt meist die größten Werte. Die Completeness zeigt hier ein Minimum von 92 % und ein Maximum von 95 %. Correctness und Quality ist mit einem Minimum

von 82 % und 78 % etwas schlechter. Die kleinsten Werte weist das SVM-Verfahren auf mit einer minimalen Completeness von 75 %, einer minimalen Correctness von 66% und einer minimalen Quality von 58 %. Die Ergebnisse der ML-Methode sind nicht viel besser als die der SVM, mit einer minimalen Completeness von 77 %, Correctness von 70 % und Quality von 65 %. Die etwas schlechteren Werte der ML- und SVM-Methode können, wie in Abschnitt 4.3.2 erläutert, z.B. an den gewählten Trainingsdaten liegen. Ein Grund für die schlechteren Werte des ML-Verfahrens ist außerdem die vermutlich nicht zutreffende Voraussetzung, die Klassen seien im Merkmalsraum durch eine Normalverteilung darstellbar.

Auffällig in allen Gebieten ist, dass das SVM-Verfahren deutlich geringere Qualitätsmaße aufweist als die anderen Verfahren. Eine Ursache für die Fehler könnte die Wahl von Strafterm C und dem Parameter der Kernelfunktion γ sein. Da die Parameter vom Nutzer gewählt wurden, ist es möglich, dass die Werte nicht zu dem gewünschten Ergebnis führen. Eine Möglichkeit dem entgegenzuwirken ist, dass die beiden Parameter durch Kreuzvalidierung ebenfalls aus den Trainingsdaten bestimmt werden. Dies würde jedoch die Trainingszeit weiter vergrößern.

Bei der Betrachtung der Mittelwerte im Bezug auf die Auflösung der klassifizierten Bilder (Tabelle 4.8), fällt auf, dass sich die Quality bei der ML- und SVM-Methode nur um maximal 3 % unterscheidet. Bei der DS-Methode ist jedoch deutlich eine Verschlechterung der Quality von 6 % durch die Reduktion der Auflösung von 8 cm auf 20 cm zu sehen.

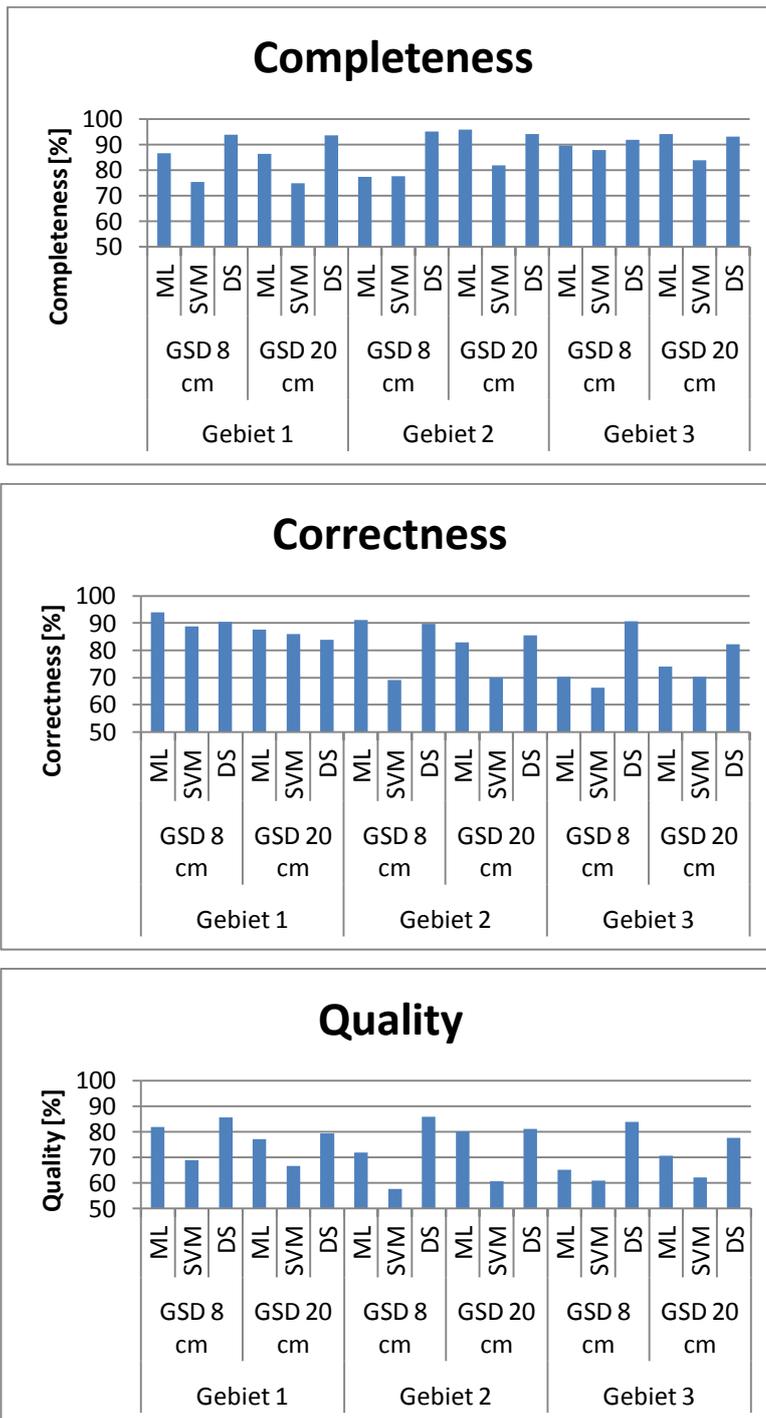
Tabelle 4.8 flächenbasierte Qualitätsmaße: Mittelwerte aus den drei Testgebieten

Verfahren	ML		SVM		DS	
	8 cm	20 cm	8 cm	20 cm	8 cm	20 cm
Completeness [%]	84	92	80	80	94	94
Correctness [%]	85	82	75	76	90	84
Quality [%]	73	76	62	63	85	79

Die grafische Darstellung der Qualitätsmaße (Abb.4.14) zeigt, dass die Completeness in Gebiet 3 etwas besser ist als in den anderen Gebieten. Dementsprechend sind dort die Gebäude am besten zu erkennen gewesen. Bei den Qualitätsmaßen der ML- und SVM- Methode fällt jedoch auf, dass die Correctness in dem Gebiet schlechter ist als in den anderen, sodass der hohe Completeness-Wert daher stammen kann, dass allgemein mehr Pixel als Gebäude klassifiziert wurden. Die niedrige Correctness unterstützt die visuelle Interpretation, dass viele Büsche fälschlicherweise als Gebäude klassifiziert wurden.

Gebiet 1 besitzt im Mittel eine etwas bessere Correctness als die beiden anderen Gebiete und demnach weniger FP. Eine mögliche Erklärung ist, dass sich in Gebiet 1 weniger kleine Büsche an den Straßenrändern befinden, die, besonders in Gebiet 3, meist als Gebäude fehlklassifiziert wurden.

Abbildung 4.14 flächenbasierte Qualitätsmaße



Die objektbasierten Qualitätsmaße (Abb. 4.15) sind deutlich schlechter. Die Dempster-Shafer Methode besitzt auch hier meist die besten Werte und die SVM die schlechtesten. Besonders in Gebiet 2 ist die Quality der SVM aufgrund einer Correctness von nur 12 % sehr gering. Im Vergleich zu den anderen Gebieten ist die Quality jedoch bei allen Verfahren in Gebiet 2 am geringsten, was bei den ML- und SVM- Verfahren fast ausschließlich an der niedrigen Correctness in diesem Gebiet liegt. Eine ebenfalls schlechte Correctness weist auch Gebiet 3 auf. Beide Gebiete zeigten auch in der

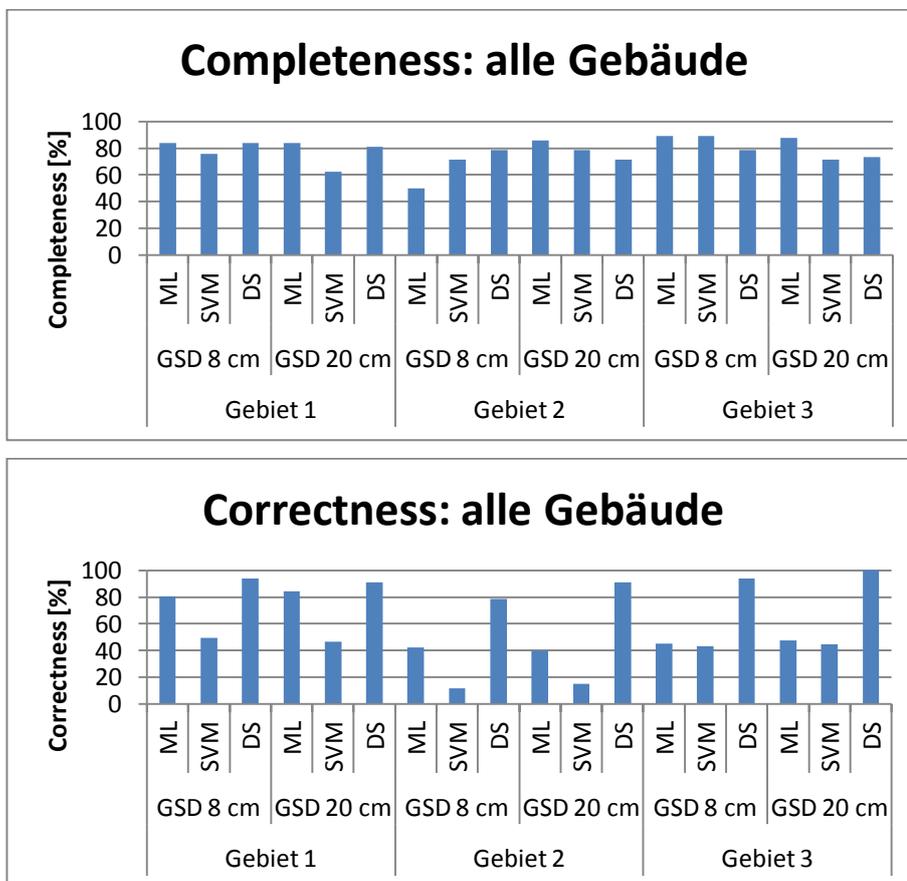
visuellen Interpretation viele FP Pixel. Die vielen FP Pixel führen zu vielen kleinen zusammenhängenden Komponenten, die von der Evaluierungsmethode als Gebäude interpretiert werden und so zu der schlechten Correctness führen könnte.

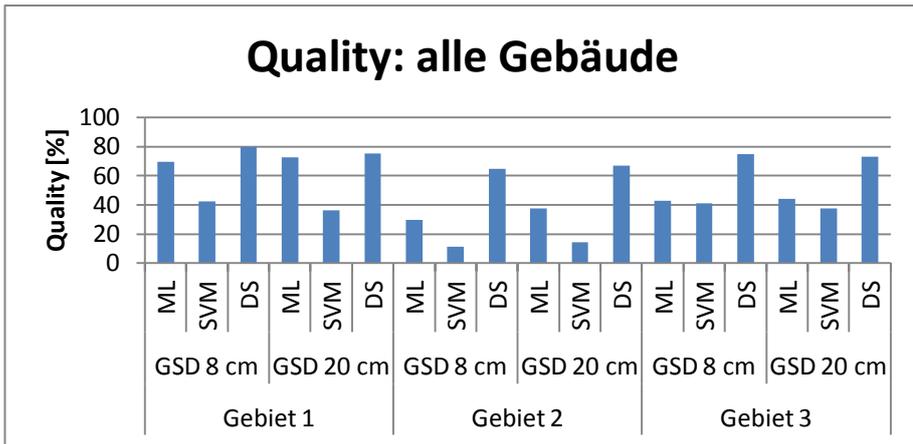
Die Completeness liegt im Gegensatz zu der Correctness immer über 62 %, abgesehen von dem Ergebnis der ML-Methode in Gebiet 2 (8 cm GSD). Dieser schlechte Completeness-Wert von 50 % liegt an den kleineren Gebäuden im oberen Bereich (Abb. 4.5, Nr.3) wie in der visuellen Interpretation beschrieben.

Ein Vergleich der Qualitätsmaße bei den verschiedenen Auflösungen zeigt nur einen geringen Unterschied zwischen der hohen und der niedrigen Auflösung (Tabelle 4.9). Während das Mittel der Quality identisch ist, zeigen Completeness und Correctness nur 1 bzw. 2 % Unterschied.

	GSD	8 cm	20 cm	
Completeness [%]		78	77	<i>Tabelle 4.9: Mittelwerte der objektbasierten Qualitätsmaße aller Ergebnisse der gleichen Auflösung</i>
Correctness [%]		60	62	
Quality [%]		51	51	

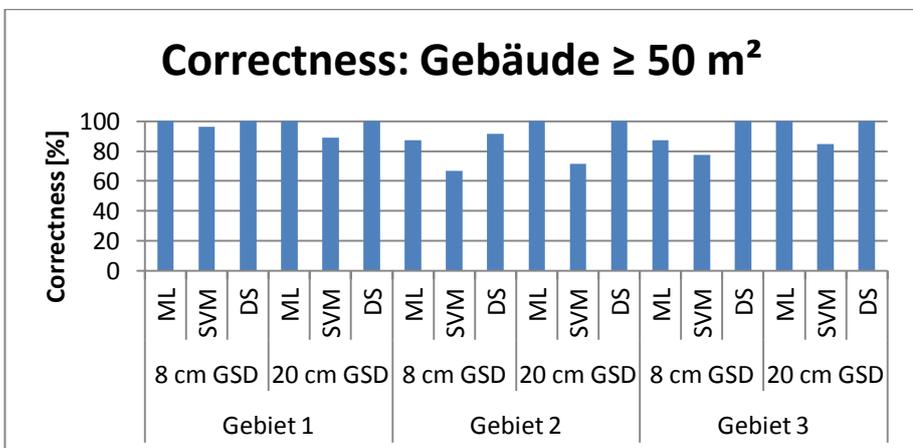
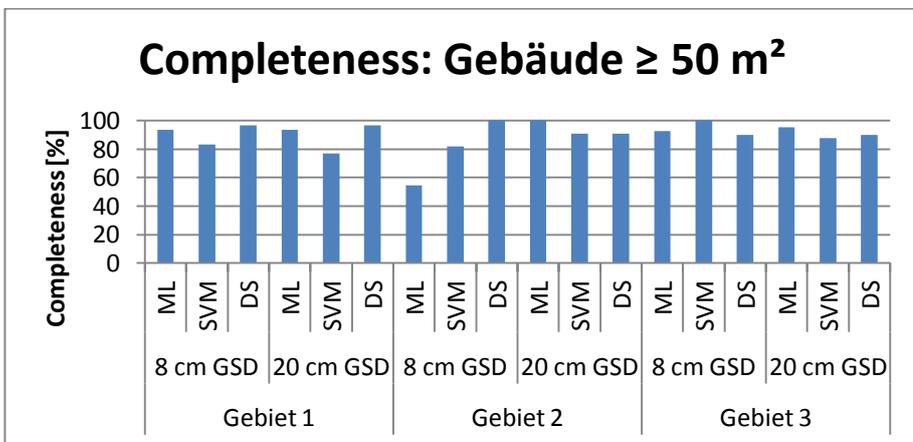
Abbildung 4.15 objektbasierte Qualitätsmaße (alle Gebäude)

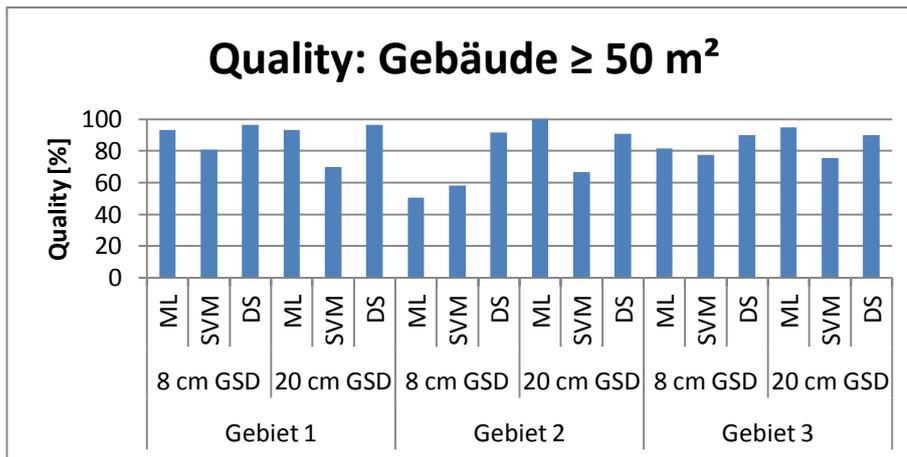




Bei der Betrachtung der Qualitätsmaße für Gebäude die größer als 50 m² sind (Abb. 4.16), fällt auf, dass die Werte erheblich besser sind als die objektbasierten Qualitätsmaße unter Berücksichtigung aller Gebäude. Besonders die Correctness, die bei der Betrachtung aller Gebäude sehr schlecht war ist stark gestiegen, was die zuvor gestellte Hypothese, die schlechte Correctness stamme von den vielen kleinen FP, bestätigt.

Abbildung 4.16 objektbasierte Qualitätsmaße (Gebäude $\geq 50 \text{ m}^2$)





Die besseren Qualitätsmaße bei der ausschließlichen Betrachtung größerer Gebäude führt zu der Annahme, dass die Qualitätsmaße stark mit der Gebäudegröße korrelieren. Aus diesem Grund werden in Abbildung 4.17 (links) die Qualitätsmaße von Gebiet 1 (8 cm GSD) als Funktion der Objektgröße dargestellt. Ebenfalls in der Abbildung 4.17 enthalten ist die kumulative Darstellung dieser Funktion. Dort werden die Qualitätsmaße für alle Gebäude, die größer als eine Größe $x \text{ [m}^2\text{]}$ sind, dargestellt.

Die Darstellungen der restlichen Gebiete und der Bilder mit 20 cm GSD befinden sich im Anhang. Im Wesentlichen steigen die Werte bei allen Darstellungen anfangs stark an und nähern sich mit steigender Größe und sinkender Steigung immer weiter dem Maximalwert von 100 %. Das zeigt, dass größere Gebäude besser erkannt werden als kleine. Die Gebäudefläche, bei der Completeness, Correctness und Quality 100 % erreichen, variiert.

Bei dem Vergleich der Methoden ist auch hier zu sehen, dass die Erkennung von 100% der Gebäude bei der Methode nach Dempster-Shafer bereits bei verhältnismäßig geringen Gebäudegrößen auftritt. Beispielhaft ist dies gut an Area 1 zu sehen. Bei den Bildern mit 8 cm Auflösung entstand mit der DS-Methode ab einer Größe von $87,5 \text{ m}^2$ eine Quality von 100%. Bei der ML-Methode wurden die 100% erst ab einer Größe von $112,5 \text{ m}^2$ erreicht und bei der SVM-Methode ab einer Größe von $237,5 \text{ m}^2$.

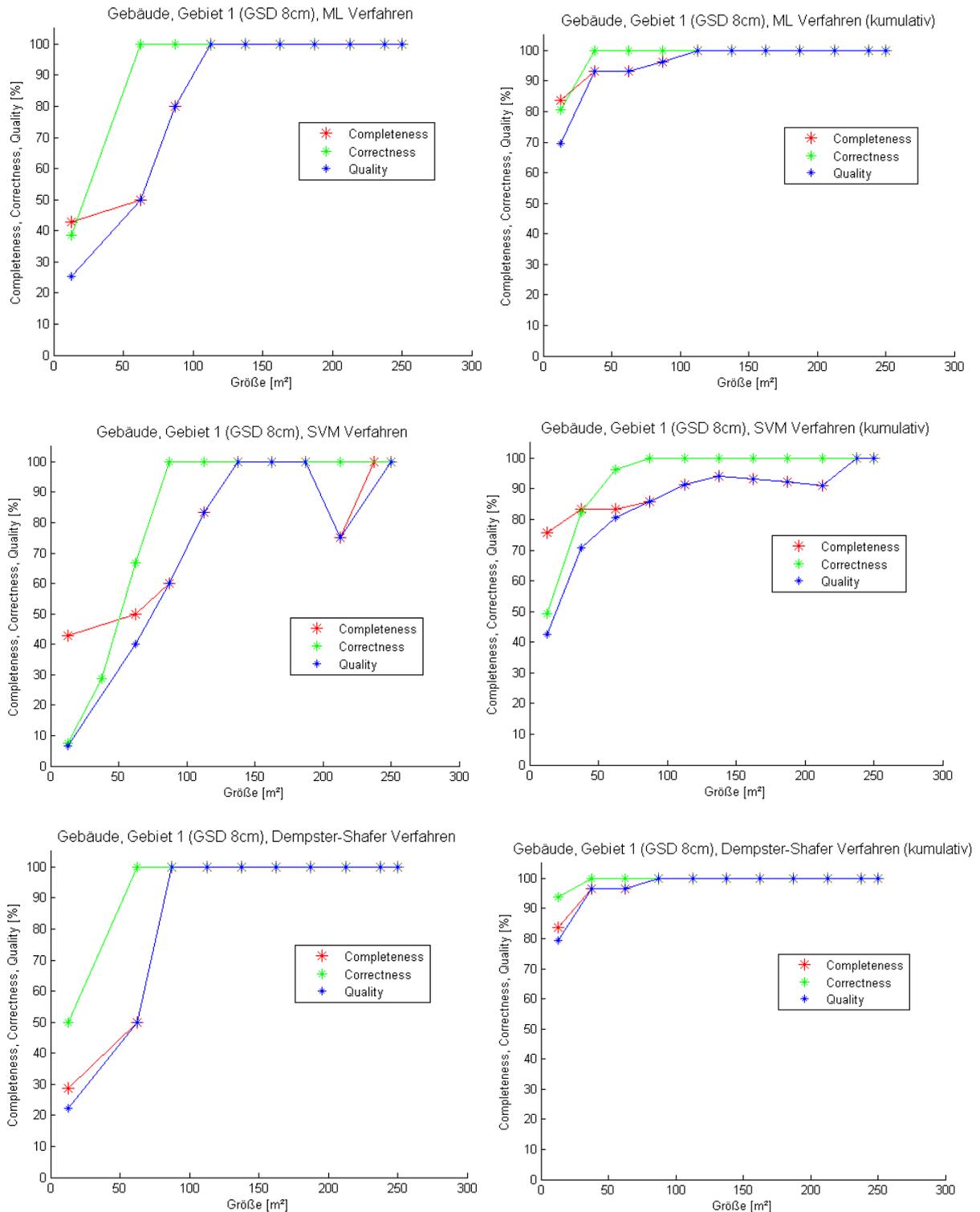
Liegt die Gegenüberstellung der drei Gebiete im Vordergrund, fällt auf, dass die Quality bei Gebiet 3 bei der ML- und SVM- Methode nie 100% erreicht. Auch bei der Methode nach Dempster-Shafer wird dieser Wert erst bei sehr großen Gebäudeausmaßen, $212,5 \text{ m}^2$ bzw. $137,5 \text{ m}^2$, erreicht. Der Grund könnte ein Gebäude (Nr.5 in Abb.4.5) sein, welches aufgrund von Fehlern im nDOM nicht erkannt wird und so zu einer Completeness von unter 100 % führt.

Werden die Ergebnisse der beiden Auflösungen verglichen, ist zu erkennen, dass die geringere Auflösung nicht unbedingt schlechtere Ergebnisse erzeugt. Bei der ML-Methode sind die Qualitätsmaße der geringen Auflösung sogar etwas besser als die der hohen Auflösung. Bei den

Ergebnissen der DS-Methode zeigte sich kaum ein Unterschied hinsichtlich der verschiedenen Auflösungen und die Gebäudegröße, bei der alle Gebäude dieser Größe und größer erkannt werden, sind bei Gebiet 1 und 2 sogar identisch. Nur die Ergebnisse der SVM-Methode zeigten teilweise eine schlechtere Qualität bei sinkender Auflösung.

Abbildung 4.17

Normale und kumulative Darstellung der Qualitätsmaße als Funktion der Objektgröße.



4.3.4. Geometrische Genauigkeit

Für jedes Gebiet und jeder Auflösung wurde der RMS-Fehler der Distanzen, wie in 4.1 beschrieben, berechnet (Tabelle 4.10).

Ein Vergleich des RMS der drei Methoden zeigt, dass die DS-Methode die kleinsten Werte aufweist und somit, wie bei den in Abschnitt 3.4.4 behandelten Qualitätsmaßen, bessere Ergebnisse als die beiden anderen Verfahren erzeugt. Das SVM-Verfahren weist auch hier die größten Fehler auf.

Bei einem Vergleich der Genauigkeiten im Bezug auf die verwendete Auflösung fällt auf, dass die Werte bei den Bildern mit 20 cm GSD größer sind als bei denen mit einer hohen Auflösung. Bei den Ergebnissen, basierend auf einer 20 cm Auflösung, kommt es sogar vor, dass für einige Randpunkte der klassifizierten Objekte keine Randpunkt der Referenz in einem Abstand von 3 m gefunden wurde, sodass kein RMS berechnet werden konnte. Eine geringere Auflösung führt demnach zu einer ungenaueren Bestimmung der Objektränder. Auf die Objekterkennung scheint die geringere Auflösung jedoch keine großen Auswirkungen zu haben (vgl. Abschnitt 3.4.4). Ebenfalls zu beachten ist, dass die RMS nicht linear mit der GSD ansteigt. Beispielhaft wird dies an den Werten der DS-Methode verdeutlicht. Die Bilder mit einer GSD von 8 cm sind um das 2,5-Fache besser aufgelöst als die Bilder mit einer GSD von 20 cm. Der RMS der höheren Auflösung ist aber nur um einen Faktor von etwa 1,45 besser als der RMS der niedrigen Auflösung.

Tabelle 4.10 RMS-Fehler der Distanzen

Methode	ML					
GSD [cm]	8			20		
Gebiet	1	2	3	1	2	3
RMS [m]	0,76	0,68	0,76	0,84	/	2,89
RPG / RP	454 / 513	223 / 275	516 / 570	449 / 534	0 / 297	2 / 563
Methode	SVM					
GSD [cm]	8			20		
Gebiet	1	2	3	1	2	3
RMS [m]	0,96	0,86	0,93	1,54	/	/
RPG / RP	468 / 627	158 / 324	409 / 615	2 / 507	0 / 291	0 / 553

Methode	DS					
GSD [cm]	8			20		
Gebiet	1	2	3	1	2	3
RMS [m]	0,77	0,71	0,56	1	0,99	0,92
RPG / RP	436 / 490	238 / 246	484 / 492	422 / 486	230 / 242	477 / 502

4.3.5. Potential der Methoden und Daten

Die Genauigkeit der Methoden hängt stark von dem Trainings- und Testgebieten ab. In dieser Arbeit lieferte die DS-Methode meist die besten Ergebnisse. In einigen Bildern entstanden durch das ML-Verfahren ebenfalls akzeptable Ergebnisse. In anderen Testgebieten könnte das Ergebnis der ML-Methode sehr viel schlechter sein, z.B. in Gebieten mit vielen unterschiedlichen Merkmalsausprägungen innerhalb einer Klasse.

Die Ergebnisse der DS-Methode besaßen bei beiden Auflösungen die beste Quality und wurden daher für die Analyse des Potentials der Daten verwendet. Bei der Betrachtung unterschiedlicher Qualitätsmaße der DS-Methode (Tabelle 4.11) ist zu sehen, dass die Gebäudeerkennung allein auf Basis von hochaufgelösten Infrarot-Luftbildern mit bekannten Klassifikationsmethoden recht gute Ergebnisse hervorbringen kann. Die Completeness, bezogen auf komplette Objekte, erreicht 81 % bei einer Auflösung von 20 cm (Gebiet 1), dementsprechend wird weniger als $\frac{2}{10}$ der Gebäude nicht erkannt. Bei einer Auflösung von 8 cm ist die Completeness bei demselben Gebiet etwas besser. Werden allein die Gebäude, deren Fläche größer als 50 m² ist, betrachtet, liegt die Completeness in allen Gebieten und Auflösungen über 90 %.

Die objektbasierte Correctness ist, außer bei Gebiet 2 (8 cm GSD), über 90 %, was zeigt, dass nicht einmal 10 % der detektierten Gebäude in den Referenzdaten keine Gebäude darstellen. Bei den Gebäuden mit einer Fläche ≥ 50 m² beträgt die Correctness meist 100 %.

Bei der Betrachtung der RMS fällt auf, dass bei beiden Auflösungen das Genauigkeitspotential, welches die Daten besitzen durch die Klassifikationsmethoden noch nicht komplett genutzt wird. Es ist jedoch zu erwähnen, dass die verwendeten Klassifikationsverfahren auch keine Elemente zur Verbesserung der Genauigkeit besitzen.

Tabelle 4.11 Genauigkeitswerte des DS-Verfahrens

Gebiet	GSD	$Com_{pixel} /$ $Cor_{pixel} [\%]$	$Com_{objekt} /$ $Cor_{objekt} [\%]$	$Com_{obj \geq 50m^2} /$ $Cor_{obj \geq 50m^2} [\%]$	RMS [m]
1	8 cm	94 / 91	84 / 94	97 / 100	0,77
1	20 cm	94 / 84	81 / 91	97 / 100	1,00
2	8 cm	95 / 90	79 / 79	100 / 92	0,71
2	20 cm	94 / 86	71 / 91	91 / 100	0,99
3	8 cm	92 / 91	79 / 94	90 / 100	0,56
3	20 cm	93 / 82	73 / 100	90 / 100	0,92

5. Zusammenfassung und Ausblick

Die vorliegende Arbeit beschäftigt sich mit der Gebäudeerkennung in hochaufgelösten Infrarot-Luftbildern in städtischen Gebieten. Zur Realisierung wurden drei Klassifikationsmethoden an drei unterschiedlich charakterisierten Gebieten getestet.

Die Verfahren sollten in allen Gebieten die Klassen „Gebäude“ von der Umgebung in zwei verschiedenen Bildauflösungen unterscheiden. Als Merkmale zur Klassifikation dienten bei allen Verfahren die Höhe über dem Gelände und der NDVI. Das Maximum-Likelihood- und Support-Vector-Machines Verfahren nutzten zusätzlich noch die radiometrischen Informationen des True Orthophotos als Merkmal. Nach der Klassifikation wurden die Gebäude in einem Labelbild gespeichert. Im Anschluss folgten die Evaluierung der Labelbilder und eine Analyse der Ergebnisse.

Die Analyse zeigte, dass besonders Schattenbereiche und kleine Gebäude zu Problemen führen, da sie bereits in dem TOP und nDOM fehlerhaft dargestellt werden. Bei einem Vergleich der Ergebnisse der Klassifikationsmethoden wies die Gebäudeerkennung nach Dempster-Shafer am wenigsten Fehler auf. Probleme entstanden besonders bei bewachsenen Gebäuden. Die ML- und SVM-Methode zeigten besonders Probleme in Bereichen, die in den Trainingsdaten nicht gut repräsentiert waren, wie z.B. bewachsene oder asphaltierte Gebäudedächer und niedrige Vegetation. Die SVM-Methode benötigte zusätzlich mehrere Stunden für die Klassifikation, während die beiden anderen Methoden höchstens einige Minuten benötigten.

Die flächenbasierten Qualitätsmaße zeigten bei der DS-Methode sehr gute Ergebnisse während die Qualitätsmaße der SVM- und ML- Methode etwas schlechter ausfielen. Die objektbasierten Qualitätsmaße waren bei allen Methoden etwas schlechter als die flächenbasierten, da kleine Gebäude nicht gut detektiert werden konnten. Gebäuden mit einer Fläche $\geq 50 \text{ m}^2$ werden jedoch fast alle erkannt, was an den erheblich besseren Qualitätsmaßen zu sehen ist.

Ein Vergleich der flächenbasierten Qualitätsmaße, basierend auf unterschiedlichen Genauigkeiten, zeigte bei der ML- und SVM- Methode keine größeren Unterschiede. Bei der DS-Methode sind die Qualitätsmaße der höheren Auflösung jedoch deutlich besser. Dies unterstützte der RMS-Fehler, der bei den Daten mit 20 cm GSD etwas größer ist als bei den Daten mit 8 cm GSD. Der Unterschied der RMS-Fehler ist jedoch im Verhältnis zur Auflösung gering.

Weiterführend könnten die Klassifikationsmethoden erneut mit präziseren Höhendaten durchgeführt werden, da die Fehler des nDOM einige Fehlklassifikationen verursachten. Eine Möglichkeit wäre die Verbesserung des Matchingverfahrens oder es werden andere bzw. zusätzliche Daten verwendet, wie z.B. Laserscanningdaten.

Auch bei der SVM Methode könnte mit den gleichen Daten noch weitergeforscht werden. Die Vereinfachungen, die in dieser Arbeit zur Verringerung der Laufzeit angebracht wurden, könnten nicht verwendet werden, sodass zwar deutlich mehr Zeit für eine Klassifikation benötigt wird, aber ggf. auch bessere Ergebnisse entstehen. Außerdem könnte untersucht werden, ob die Ergebnisse der SVM besser ausfallen, wenn der Strafterm C und der Parameter des Gaußkernels γ durch Kreuzvalidierung bestimmt werden. Auch die Veränderungen der Ergebnisse aufgrund einer Erweiterung oder Umgestaltung der Trainingsdaten könnte weiter untersucht werden.

6. Literaturverzeichnis

- Abe, Shigeo (2010), *Support Vector Machines for Pattern Classification*, Springer, London.
- Albertz, Jörg (1991), *Grundlagen der Interpretation von Luft- und Satellitenbildern*, Wissenschaftliche Buchgesellschaft, Darmstadt.
- Baldenhofer, Kurt G. (2012), *Lexikon der Fernerkundung*, ErdWare Geographische Software, <http://www.fe-lexikon.info/index.htm>, 22.9.2012.
- Cramer, M. (2010), *The DGPF-Test on Digital Airborne Camera Evaluation-Overview and Test Design*. In: Photogrammetrie Fernerkundung Geoinformation, Heft 2, E. Schweizerbart'sche Verlagsbuchhandlung Stuttgart.
- de Lange, Norbert (2006), *Geoinformatik in Theorie und Praxis*, Springer, 2.Auflage, Osnabrück.
- Graf, R.J. (2009), *Objektklassifizierung Support Vector Machines*, Masterarbeit, Paris-London-Universität Salzburg.
- Heinert, Michael (2010), *Support Vector Machines – Teil 1: Ein theoretischer Überblick*, Zeitschrift für Geodäsie, Geoinformation und Landmanagement, 135. Jg.
- Heinert, Michael, Riedel, Björn (2010) *Support Vector Machines – Teil 2: Praktische Beispiele und Anwendungen*, Zeitschrift für Geodäsie, Geoinformation und Landmanagement, 5/2010 135. Jg.
- Heipke, Christian (2010), Kapitel 2.6. *Digitale Bildzuordnung*, Photogrammetrie & Fernerkundung II, Institut für Photogrammetrie und GeoInformation, Leibniz Universität Hannover.
- Helmholz, Petra (2012), *Verifikation von Ackerland- und Grünlandobjekten eines topographischen Datensatzes mit monotemporalen Bildern*, Wissenschaftliche Arbeit der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover.
- Hirschmüller, Heiko (2008), *Stereo Processing by Semiglobal Matching and Mutual Information*, IEEE Transactions on pattern analysis and machine intelligence, Vol. 30 No. 2.
- Kosov, S., Rottensteiner, F., Heipke, C., Leitloff, J., Hinz, S. (2012), *3D classification of crossroads from multiple aerial images using Markov random fields*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXIX-B3, S. 479-484.
- Kraus, Karl (1990), *Fernerkundung Band 2, Auswertung photographischer und digitaler Bilder*, Dümmler, Bonn.
- Liu, Feng (2012), *Objektverfolgung durch Fusion von Radar- und Monokameradaten auf merkmalsebene für zukünftige Fahrerassistenzsysteme*, KIT Scientific Publishing, Karlsruhe.
- OpenCV (2012), *Camera Calibration and 3d Reconstruction*, http://opencv.willowgarage.com/documentation/cpp/calib3d_camera_calibration_and_3d_reconstruction.html#stereosgmbm, 24.10.2012.
- Prinz, T. (2007), *Digitale Fernerkundungsmethodik in den Geowissenschaften*, http://ivvgeo.uni-muenster.de/Vorlesung/FE_Script/3_3.html, 24.10.2012.
- Professur für Geodäsie und Geoinformatik der Universität Rostock (2001), *GI-Lexikon*, Geoinformatik-Service der Universität Rostock, <http://www.geoinformatik.uni-rostock.de/az.asp?letter=a> , 22.9.2012.

Rottensteiner (2011), *Statistische Methoden II*, Bildanalyse II, Institut für Photogrammetrie und GeoInformation, Leibniz Universität Hannover.

Rottensteiner, F., Baillard, C., Sohn, G., Gerke, M. (2011), *ISPRS Test Project on Urban Classification and 3D Building Reconstruction*, ISPRS – Commission III – Photogrammetric Computer Vision and Image Analysis, Working Group III/4 – Complex Scene Analysis and 3D Reconstruction.

Rottensteiner, Franz, Trinder, John, Clode, Simon, Kubik, Kurt (2005), *Using the Dempster Shafer method for the fusion of LIDAR data and multispectral images for building detection*, Information Fusion 6 (4) pp. 283-300.

Rutzinger, M., Rottensteiner, F., Pfeifer, N. (2009), *A Comparison of Evaluation Techniques for Building Extraction From Airborne Laser Scanning*, IEEE Journal of selected topics in applied earth observations and remote sensing, Vol. 2 No. 1.

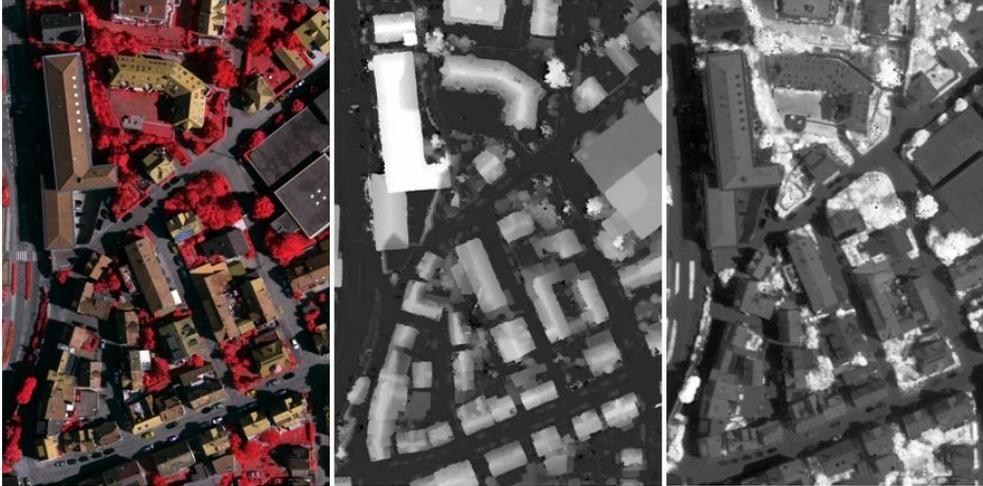
Sörgel, Uwe (2012), *Klassifikation der Landbedeckung*, Photogrammetrie & Fernerkundung III, Institut für Photogrammetrie und GeoInformation, Leibniz Universität Hannover.

Wiggenhagen, M., Wißmann, U.(2003), *Untersuchungen zur Prüfung digitaler Photogrammetrischer Produkte nach DIN 18740*. In: Eckhard Seyfert (Hrsg.): Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation. Band 12. Bochum, S. 163-170.

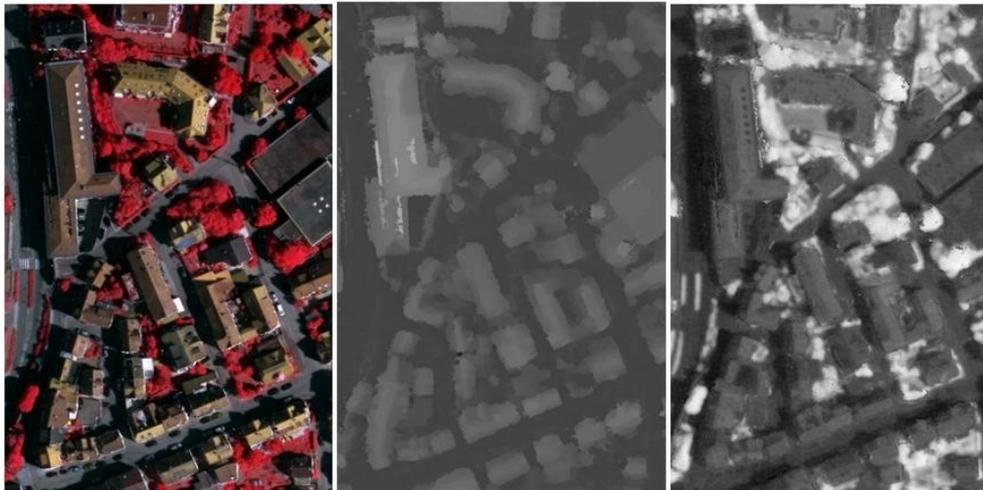
7. Anhang

von links nach rechts: true Orthophoto, nDOM, NDVI

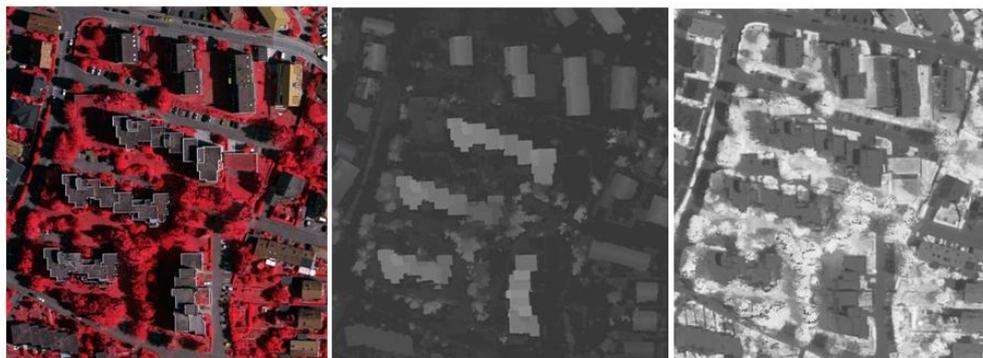
Gebiet 1, 8 cm GSD



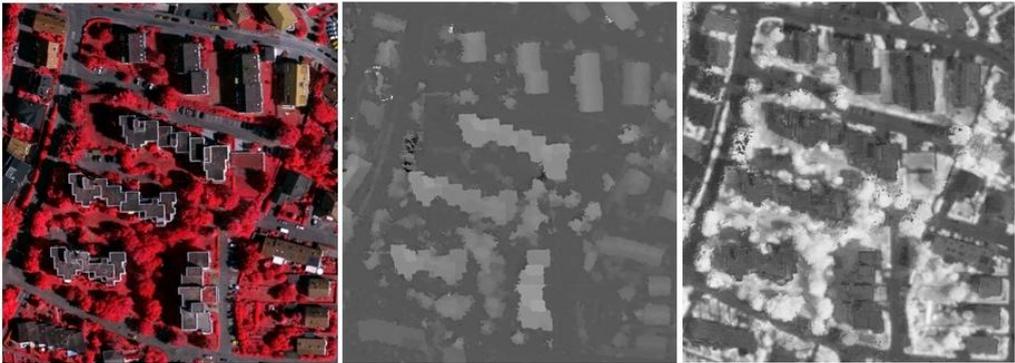
Gebiet 1, 20 cm GSD



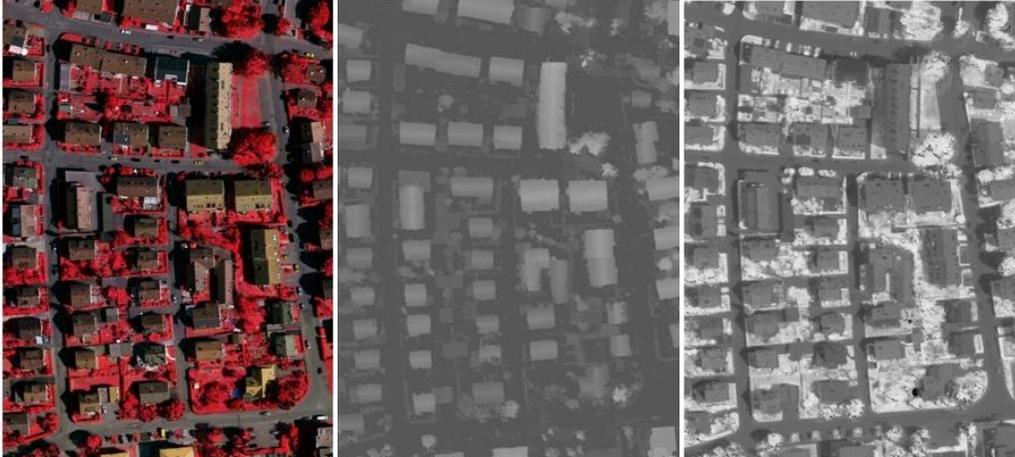
Gebiet 2, 8 cm GSD



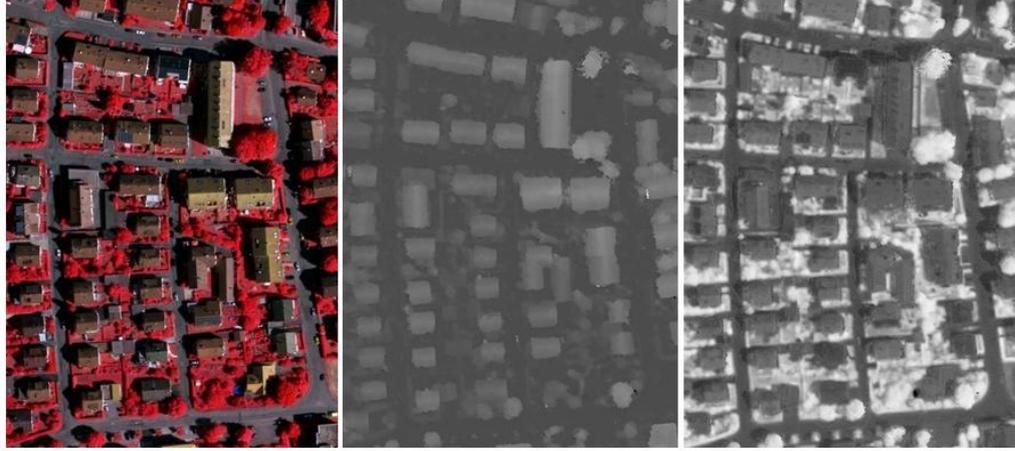
Gebiet 2, 20 cm GSD



Gebiet 3, 8 cm GSD



Gebiet 3, 20 cm GSD



Ergebnisse der Klassifikationen

Maximum Likelihood Klassifikation

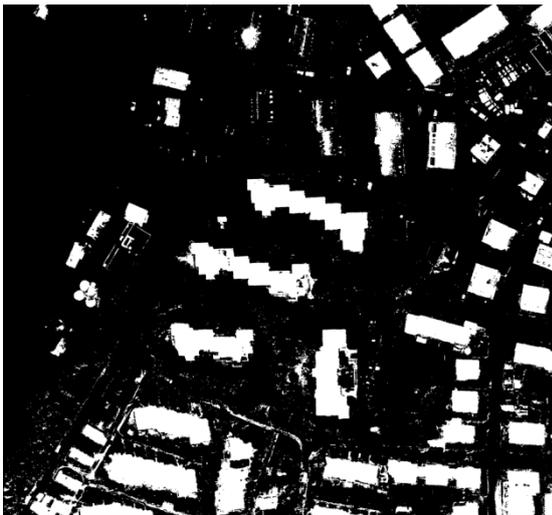
Gebiet 1, 8cm GSD



Gebiet 1, 20cm GSD



Gebiet 2, 8cm GSD



Gebiet 2, 20cm GSD



Gebiet 3, 8cm GSD



Gebiet 3, 20cm GSD



Support Vector Machines Klassifikation

Gebiet 1, 8cm GSD



Gebiet 1, 20cm GSD



Gebiet 2, 8cm GSD



Gebiet 2, 20cm GSD



Gebiet 3, 8cm GSD

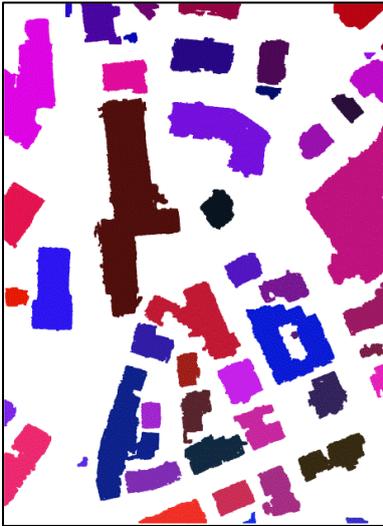


Gebiet 3, 20cm GSD



Dempster-Shafer Gebäudeextraktion

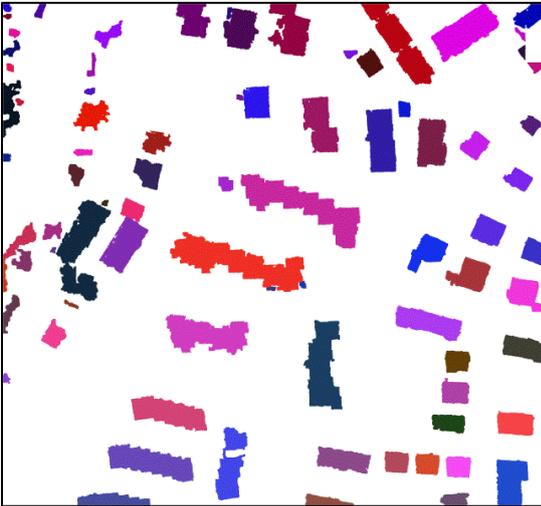
Gebiet 1, 8cm GSD



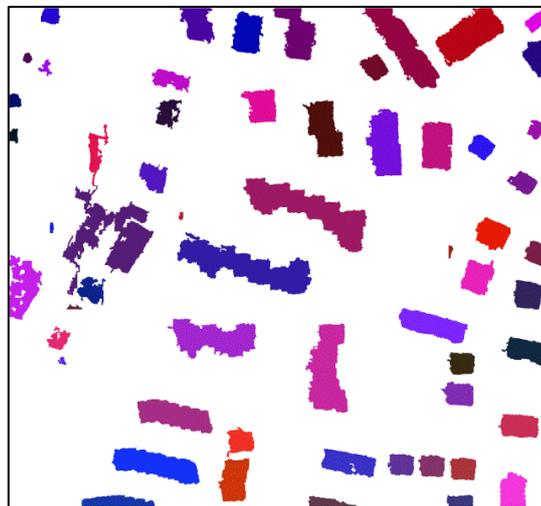
Gebiet 1, 20cm GSD



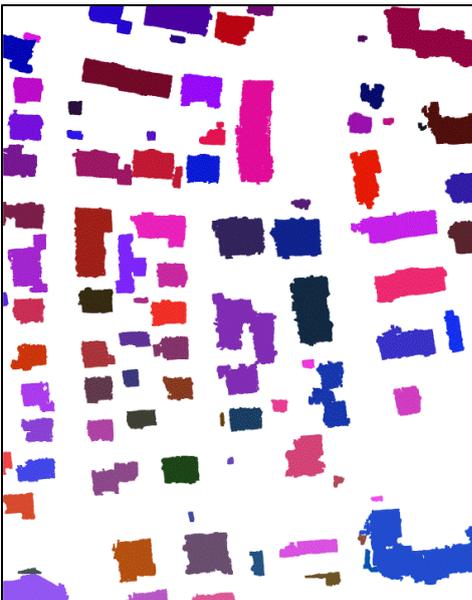
Gebiet 2, 8cm GSD



Gebiet 2, 20cm GSD



Gebiet 3, 8cm GSD



Gebiet 3, 20cm GSD



Qualitätsmaße

