

Stabilität von Regionen im Skalenraum

MARTIN DRAUSCHKE¹, HANNS-FLORIAN SCHUSTER¹ & WOLFGANG FÖRSTNER¹

Zusammenfassung: Für die automatische Erfassung von Gebäuden aus Luftbildern ist es nützlich, Bildstrukturen im Skalenraum, d. h. über mehrere Auflösungsstufen zu beobachten, um für die Objekterkennung hinderliche Details ausblenden zu können. Große Bedeutung messen wir dabei den homogenen Regionen sowie deren Nachbarschaften zu.

Regionen betrachten wir als stabil, wenn sie über einen mehrere Skalenstufen invariant bleiben. Sie haben spezielle Eigenschaften: Beim Vergrößern der Skala verschmelzen benachbarte Regionen, wobei eine Region immer vollständig in der anderen aufgeht. Zudem zerteilen sich stabile Regionen nicht, wenn die Skala steigt. Diese speziellen Eigenschaften erleichtern das Bestimmen der Nachbarschaften in einer vorgegebenen Skala, denn der Regionennachbarschaftsgraph (RNG) muss nur einmal auf der untersten Ebene des Skalenraums berechnet werden. Die RNGs in den anderen Ebenen können leicht aus der untersten Ebene berechnet werden.

1 Einleitung

Gebäude und Gebäudeteile stellen wegen ihrer großen Variabilität in Form und Ansicht eine große Herausforderung für die automatische Erfassung aus Luftbildern dar. Für das Erkennen von komplexeren Strukturen, die sich z.B. im Vorhandensein von Dachgauben, Erkern oder Anbauten zeigen, wollen wir einen generischen Ansatz entwickeln. Dafür benötigen wir aus dem Bild extrahierte Merkmale sowie ihre topologischen Beziehungen untereinander.

In Luftbildern sind von Gebäuden vor allem Dachflächen sichtbar. Diese weisen nur eine schwache Textur auf, weil Dächer in der Regel gleichmäßig abgedeckt werden. Eine Differenzierung der Dachdeckung ist zudem durch die zu grobe Bodenauflösung von Luftbildern (5 bis 30 cm) kaum möglich. Bilder unterschiedlicher Auflösung zeigen Objekte verschiedener Skala. Viele kleine Dachaufbauten wie Antennen und Schornsteine sind für die maschinelle Gebäudeerkennung hinderlich. Auf einer niedrigeren Auflösungsstufe sind diese für die Problemstellung entbehrlichen kleinskaligen Gebäudeteile häufig nicht mehr erkennbar, dagegen sind die Kanten zwischen großskaligen Objekten wie planaren Dachflächen oftmals auch bei niedrigerer Auflösung noch detektierbar. Aus diesem Grund untersuchen wir, inwieweit sich die Analyse der Veränderung von Bildstrukturen im Skalenraum für die Gebäudeerkennung eignet.

Wir wollen die Stabilität der detektierten Merkmale im Skalenraum untersuchen. Als Stabilitätskriterium verwenden wir Position und Ausdehnung von Regionen, sie sollen sich über größere Skalenbereiche nur geringfügig ändern. Damit ist unser Ansatz dem von MATAS ET AL. (2002) ähnlich, der Regionen als stabil betrachtet, die sich über mehrere Intensitätsniveaus nicht ändern.

Die Merkmalsextraktion im Skalenraum ist ein intensiv untersuchtes Gebiet. Eine gute Einführung gibt LINDEBERG (1994). Für die Detektion von Regionen verwendet er einen normalisierten Laplace-Filter. Kanten werden durch die Vorzeichenwechsel der zweiten Ableitun-

¹ Martin Drauschke, Hanns-Florian Schuster und Prof. Dr. Wolfgang Förstner, Institut für Photogrammetrie, Rheinische Friedrich-Wilhelms-Universität Bonn, Nussallee 15, 53115 Bonn, e-mail: martin.drauschke@uni-bonn.de, schuster@ipb.uni-bonn.de bzw. wf@ipb.uni-bonn.de

gen erfasst. Die Lage dieser Merkmale ist uns jedoch über den Skalenraum zu instabil. Skaleninvariante Punktdetektoren, wie z. B. der von LOWE (2004) vorgestellte Operator, sind für die Gebäudeerfassung aus Luftbildern wegen der schwachen Texturierung der Objektflächen ebenfalls nur bedingt einsetzbar.

Wir wollen untersuchen, ob sich extrahierten Merkmale künstlicher Objekte wie Gebäude, Autos oder Straßen in Bezug auf ihre Stabilität im Skalenraum von den Merkmalen der natürlichen Objekten wie Wiesen und Bäume unterscheiden. Zudem haben künstliche Objekte im Allgemeinen eine einfach beschreibbare geometrische Form.

Die Fläche von Regionen ist deutlich stabiler ihr Rand. Aus diesem Grund kann man erwarten, dass im Skalenraum Regionen stabiler sind als Kanten.

Dieser Artikel ist wie folgt gegliedert. Die Wahl des Skalenraums und die geeigneten Verfahren für die Bildsegmentierung diskutieren wir im folgenden Kapitel. Auf die Stabilität der Regionen gehen wir im dritten Kapitel ein. Da sich für diese Regionen die skalenübergreifende Nachbarschaftsbestimmung stark vereinfacht, thematisieren wir diesen Sachverhalt in Kapitel 4.

2 Methodische Ansätze: Skalenraum und Bildsegmentierung

Ein mehrkanaliges Bild g ist ein c -Vektor, d. h. bei Farbbildern ein 3-Vektor, von einkanaligen Bildfunktionen g_c mit

$$g_c : R^2 \rightarrow F, \quad g_c(x,y) = \text{farbwert.}$$

Betrachtet man ein Bild im Skalenraum, dann wird der Definitionsbereich dieser Bildfunktionen um einen dritten, nicht negativen Wert, die Skala σ , erweitert:

$$g_c : R^2 \times R^+ \rightarrow F, \quad g_c(x,y;\sigma) = \text{farbwert.}$$

Die Ebene $g_c(x,y;0)$ enthält das Originalsignal eines Bildkanals, entsprechend wird das transformierte Bild in Skala σ als Layer im „Bildraum“ dargestellt (siehe Abb. 1, das untere Layer zeigt das Originalbild, das obere Layer zeigt das im Gauß'schen Skalenraum transformierte Bild bei $\sigma = 8$).

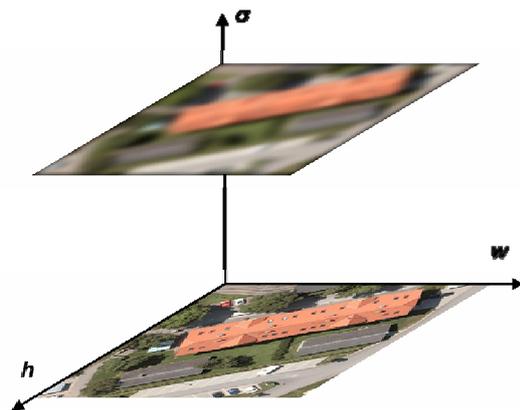


Abb. 1: Bild-Layer im Skalenraum

Für unsere Anwendung haben wir zwei verschiedene theoretische Ansätze des Skalenraums in Betracht gezogen und genauer untersucht: den Gauß'schen sowie den Morphologischen Skalenraum.

2.1 Gauß'scher Skalenraum

Der Gauß'sche Skalenraum wird seit KOENDERINK (1984) in der Literatur ausgiebig diskutiert. Er zeigte, dass die Transformation im Skalenraum die Wärmeleitungsgleichung

$$\partial_\sigma L = \frac{1}{2} \nabla^2 L$$

erfüllen muss, damit willkürliche Veränderungen des Bildsignals wie das Auftreten neuer Bildstrukturen bei einer größeren Auflösung unterbunden werden. Die Lösung dieser Differentialgleichung wird durch die Faltung des Bildsignals mit dem Gaußfilter beschrieben:

$$g_c(x, y; \sigma) = g_c(x, y) * \frac{1}{2\pi\sigma} e^{-(x^2+y^2)/2\sigma}.$$

Bei digitalen Bildern wird diese Faltung approximiert. Auf diese Weise wird ebenfalls die Gauß'sche Bildpyramide berechnet (CROWLEY ET AL., 2002). Zwischen den Pyramidenstufen λ und der Skala $\sigma > 0$ besteht folgender Zusammenhang: $\sigma = 2^{\lambda-1}$.

Die Layer des diskreten Skalenraums haben wir gleichmäßig verteilt, d.h. zwischen den Pyramidenstufen $\lambda = 1$ und 2 sowie zwischen 2 und 3 gibt es gleichviele Layer. Sind $\sigma_0 = 1$ und σ_N die größte Skala des von uns betrachteten Skalenraums sowie n die Anzahl der Bild-Layer zwischen zwei Pyramidenstufen entsprechenden Skalen, dann wird der Skalenraum durch die Skalen $\sigma = 2^{i/n} \sigma_0$, $i = 0, \dots, N$ dargestellt. Die Skalen zwischen den Pyramidenstufen 0 und 1 können durch negative Exponenten i berechnet werden.

2.2 Morphologischer Skalenraum

Die Glättung des Bildes mit einem großen Gaußfilter $G_\sigma(x, y)$ hat starke Auswirkungen auf die Segmentierung. Durch Manipulation des Bildes mit morphologischen Filtern

$$g_c(x, y; \sigma) = g_c(x, y) * M^\sigma(x, y)$$

bleiben die Konturen großer Bildstrukturen eher bestehen. Deshalb bildet der von BANGHAM (1996) vorgestellte Morphologische Skalenraum eine Alternative zum Gauß'schen Skalenraum.

Im Morphologischen Skalenraum wird die Skala nicht mehr über die Gauß-Funktion definiert. Stattdessen kennzeichnet σ nun die Größe, z. B. den Durchmesser des bei den morphologischen Operationen verwendeten Strukturelements, etwa eines Kreises oder eines Quadrats. Mit $\psi_\sigma(g_c(x, y))$ wird ein Opening und mit $\gamma_\sigma(g_c(x, y))$ ein Closing notiert. Die morphologischen Skalenraum-Filter sind Kompositionen dieser beiden Operationen, sie sind wie folgt definiert:

$$M^\sigma = \gamma_\sigma \psi_\sigma, \quad N^\sigma = \psi_\sigma \gamma_\sigma.$$

Die Abbildungen 2(a-c) zeigen die Transformation eines Bildes im morphologischen Skalenraum. Links ist das Originalbild zu sehen, daneben das Bild in Skala 3 bzw. in Skala 10. Bei kleinen Skalen (wie Abb. 2b) sind die Bildstrukturen noch deutlich erkennbar, kleine Details verschwinden. Bei großen Skalen dagegen sind die Konturen der im Bild dargestellten Objekte nur bei großen Kontrasten noch vorhanden. Zudem werden die Bildstrukturen durch die Form des Strukturelements verändert.

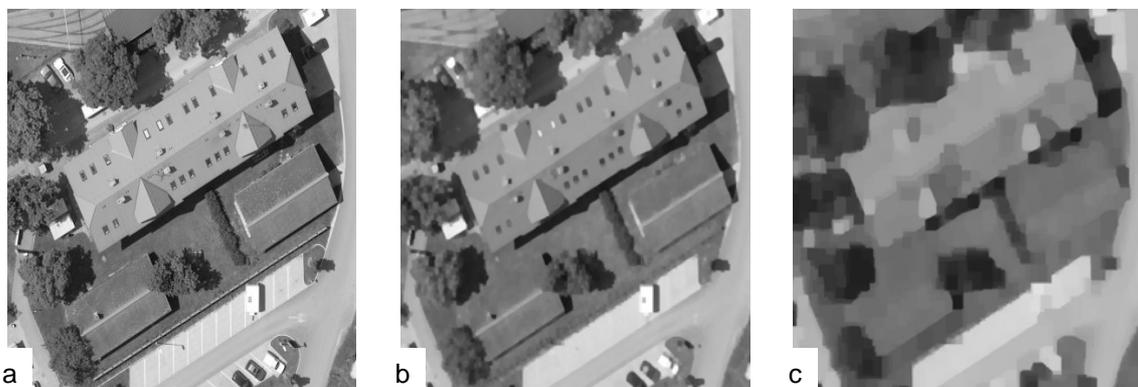


Abb. 2: Transformation im Morphologischen Skalenraum mit M^σ -Filter (v.l.n.r. $\sigma = 0, 3$ bzw. 10).

Somit lässt sich feststellen, dass der Morphologische Skalenraum nur mit kleinen Strukturelementen, also in kleinen Skalen verwendet werden sollte. Zudem lassen sich keine Zwischenstufen wegen der Atomarität der Pixel des Strukturelements definieren. Für unsere Untersuchung kommt er somit nicht mehr in Betracht, ein Segmentierungsergebnis wird dennoch in Abbildung 5 präsentiert.

2.3 Bildsegmentierung

In der Einleitung haben wir bereits die Merkmalsextraktion diskutiert. Punktoperatoren sind wegen der schwachen Textur der von uns betrachteten Objekte nur bedingt einsetzbar. Wir benötigen Regionen und ihre Nachbarschaften für die Erfassung von Dachflächen. Es gibt eine Vielzahl von Möglichkeiten für die Detektion von Regionen. LINDEBERG (1994) charakterisiert Blobs über normalisierte Laplace-Filter und detektiert somit den Rand dieser Regionen. FÖRSTNER (1994) prüft mit einem Schwellwertverfahren auf Homogenität in der Umgebung von Pixeln und erfasst somit Regionen als homogener Bereiche. Die auf diese Weise erfassten Regionen haben sehr komplizierte Ränder und eignen sich nicht sehr gut zum Erkennen von Bildstrukturen.

Neben den Regionen werden auch die Regionennachbarschaften für die Interpretation der Bildstrukturen benötigt. Bei den eben vorgestellten Operatoren können die Nachbarschaften zwischen den Regionen beispielsweise durch ein Voronoi-Diagramm berechnet werden. Bei Segmentierungsalgorithmen, die eine vollständige Partition des Bildes erzeugen, können diese Nachbarschaften direkt aus der generierten Merkmalskarte abgelesen werden. Zwei solche Ansätze haben wir für unser Problem untersucht, das Multiple Thresholding nach Gonzalez et al. (zitiert in OLSEN, 1996) sowie die Segmentierung mit Wasserscheiden nach VINCENT ET AL. (1991).

Das Multiple Thresholding wird für jeden Bildkanal ausgeführt. Nach Analyse des Histogramms erfolgt eine Quantisierung des Bildes, indem jedem Farbintervall ein einheitlicher Farbwert zugewiesen wird. Auf diese Weise werden benachbarte Pixel mit ähnlichen Farbwerten gruppiert und zu Region miteinander verbunden.

Die Intervallgrenzen werden durch die dem Verfahren gebenden Schwellwerte bestimmt. Wir haben dazu die lokalen Minima des Histogramms verwendet. Die Ergebnisse dieses Segmentierungsverfahren hängen stark von der Größe der Umgebung ab, in denen die Minimalstellen gesucht werden.



Abb. 3: Farb-Segmentierung des Bildes mit Multiple Thresholding

Die Abbildungen 3(a-c) zeigen die Veränderungen des Bildes durch Anwendung des Multiple Thresholding. Die lokalen Minimalstellen wurden links in einer Umgebung von 9 Farbwerten gesucht, in der Mitte in einer Umgebung von 21 und rechts von 39.

Kleine Umgebungen führen demzufolge zu Partitionen mit sehr kleinen Farbregionen (siehe Abb. 3a), bei großen Umgebungen verstärken sich die Kontraste Bild, was wegen der Schatten hinderlich für die Gebäudeerkennung bei schlechten Beleuchtungsverhältnissen ist.

Das andere von uns untersuchte Verfahren für die Bildsegmentierung nach VINCENT ET AL. (1991) verwendet Wasserscheiden auf dem Quadratischen-Gradienten-Bild. Die Funktionsweise des Wasserscheidenalgorithmus kann man sehr gut mit sich füllenden Wasserbecken veranschaulichen. Für jeden Bildkanal wird der quadratische Gradient

$$\|\nabla g_c(x, y; \sigma)\|^2$$

berechnet. Diese Werte werden als Gebirge interpretiert, von den Tälern aus wachsen die Regionen. Die Grenzen der Regionen sind durch die „Gebirgskämme“ festgelegt, wo die Regionen aneinander stoßen. Weitere Veranschaulichungen des Algorithmus findet man bei OLSEN (1996, Seite 37ff.). Ferner zeigt die Abbildung 4 einen Querschnitt durch ein Gradientenbild mit den Grenzen der Regionen.

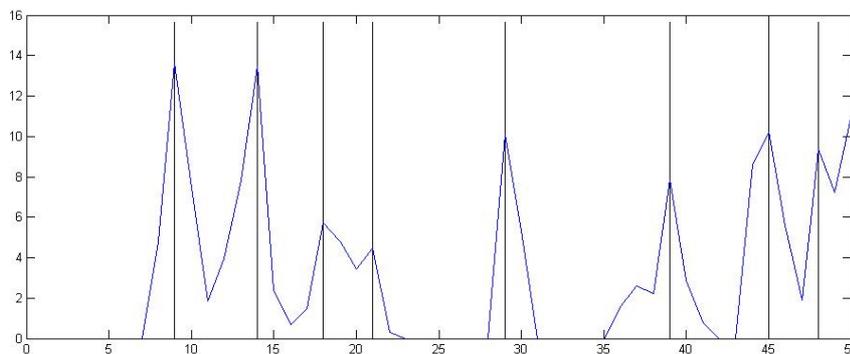


Abb. 4: Wasserscheiden im manipulierten Gradientenbild

Für die Segmentierung muss das unterschiedliche Rauschverhalten der einzelnen Bildkanäle berücksichtigt werden. Aus diesem Grund gewichten wir die quadratischen Gradienten mit dem Reziproken der geschätzten Varianz des Rauschens n_c und erhalten somit als „Gradientengebirgsfunktion“ h :

$$h(x, y; \sigma) = \sqrt{\sum_{c=1}^3 \frac{\|\nabla g_c(x, y; \sigma)\|^2}{\text{var}(n_c)}}.$$

Für diese Abschätzung der Rauschvarianz verwenden wir eine robuste Schätzung mit Hilfe des Median (vgl. BRÜGELMANN ET AL. (1992):

$$\text{var}(n_c) = \max \{1, \text{med} (\|\nabla g_c(x, y; \sigma)\|^2 / (1 - \ln 2))\}.$$

Eine zusätzliche Rauschunterdrückung erschwert eine Übersegmentierung des Bildes, zu der der Wasserscheidenalgorithmus sehr häufig neigt. Durch folgende Transformation des gewichteten Quadratischen-Gradienten-Bildes erhalten wir unsere Eingabefunktion h' , wie sie in Abbildung 4 bereits dargestellt wird und mit der wir die Wasserscheiden bestimmen:

$$h'(x, y; \sigma) = \max \{0, h(x, y; \sigma) - k * \text{var}(n_c)\}.$$

Wir verwenden in den folgenden Untersuchungen $k = 0.3$. Die Abbildung 5(a) zeigt das Originalbild und die Abbildungen 5(b,c) zeigen die jeweilige Bildsegmentierung mit dem Wasserscheidenalgorithmus für ein Bild im Skalenraum. In der Mitte wurde der Gauß'sche Skalenraum, rechts der Morphologische Skalenraum mit einem Quadrat als Strukturelement ver-

wendet. In beiden Fällen wurde ein 7x7-Filter verwendet, so dass sich um die jeweiligen Skalen $\sigma_{gauss}=1.1329$ und $\sigma_{morph}=7$ handelt.

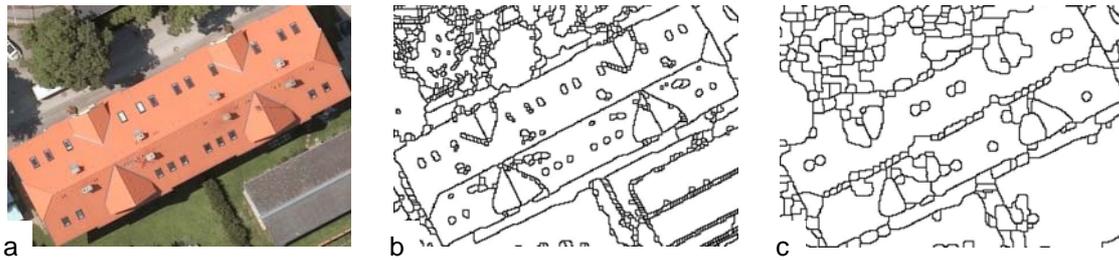


Abb. 5: Wasserscheidensegmentierung im Skalenraum

Durch diese Gegenüberstellung wird noch einmal deutlich, dass der Morphologische Skalenraum für unsere Untersuchung inakzeptabel ist. Die geometrische Form der Dachgauben im segmentierten Bild wird durch das Strukturelement stark verändert, zudem gibt es deutlich größere Probleme beim Übergang zwischen dem Haus und seiner Umgebung.

3 Stabilität und Unsicherheit von Regionen

Die Ausgabe des Wasserscheidenalgorithmus ist ein gelabeltes Bild, das wir mit $l(x,y;\sigma)$ bezeichnen. Pixel mit gleichem Label, l , bilden die Region $R(l,\sigma)$. Für eine feste Skala σ gilt: Die Vereinigung aller Regionen $\bigcup_l R(l,\sigma)$ ergibt das gesamte Bild sowie zwei Regionen schneiden sich nicht: $R(l,\sigma) \cap R(l',\sigma), l \neq l'$.

Die Zuordnung von Regionen über mehrere Skalen ist dagegen nur sinnvoll, wenn sich diese nur in wenigen Pixeln unterscheiden. Analog zu MSER-Verfahren von MATAS ET AL. (2002) lösen wir das Zuordnungsproblem über die Stabilität von Regionen.

Eine Region $R(l,\sigma_i)$ ist im Intervall $[\sigma_1,\sigma_2]$ genau dann stabil, wenn für einen Schwellwert t gilt:

$$\forall \sigma_i, \sigma_j \in [\sigma_1, \sigma_2]: |[R(l, \sigma_i) \setminus R(l, \sigma_j)] \cup [R(l, \sigma_j) \setminus R(l, \sigma_i)]| \leq t$$

Die vom Wasserscheidenalgorithmus detektierten Regionen verändern ihre Lage während der Transformation im Gauß'schen Skalenraum, allerdings sind diese Veränderungen erst bei größeren Skalenunterschieden gravierend. Aus diesem Grund kann die Veränderung der Region pro Pixel zwischen zwei dicht aufeinander folgenden Skalenstufen durch die Veränderung in der Fläche gemessen werden.

Durch die Glättung im Gauß'schen Skalenraum werden auch die Beträge des Quadratischen-Gradienten-Bilds einander angeglichen, entsprechend verschmelzen die vom Wasserscheidenalgorithmus detektierten Regionen miteinander. Wenn sich zwei Regionen mit einer genügend großen Fläche vereinen, gibt es einen gravierenden Unterschied in der Fläche. Diese Events im Skalenraum bestimmen den Anfang bzw. das Ende von stabilen Regionen.

Die Grenzen der Regionen schwanken ein wenig, was durch den Wasserscheidenalgorithmus bedingt wird. Aus diesem Grund wollen wir auch die Unsicherheit der Flächeninhalte der Regionen diskutieren. Sie hat direkte Auswirkung auf die Beurteilung der Stabilität.

Wir können eine Region auch als Polygon mit n Stützpunkten p_j darstellen, das die Kantelängen d_j zwischen zwei aufeinander folgenden Punkten hat. Zudem nehmen wir an, dass alle Punkte dieselbe Unsicherheit von $\sigma_p = 0.5$ [Pixel] haben. Dann wird die Unsicherheit der Fläche σ_A einer Region durch folgende Formel berechnet: $\sigma_A^2 = \frac{1}{4} \sum_j d_j^2 \sigma_p^2$.

Bei dichten Punkten und einer glatten Kontur der Region, d.h. $n \rightarrow \infty$, ergibt sich für die Standardabweichung der Fläche: $\sigma_A = U/\sqrt{n} \sigma_p$, d.h. sie hängt ausschließlich vom Umfang der Region ab.

Der Flächeninhalt wird für jede Region mit einem $3 \sigma_A$ -breiten Fehlerband versehen, wodurch die Unsicherheit der Fläche ausgedrückt wird. Dieses Fehlerband bewirkt einen größeren Toleranzbereich für das Stabilitätskriterium bzw. einen größeren Schwellwert t .

Die Abbildung 6 zeigt drei Regionen um einen manuell ausgesuchten Bildpunkt (v.l.n.r. in den Skalen $\sigma = 1, 4$ und 8) sowie die Entwicklung des Flächeninhalts. Waagrecht verlaufende Fehlerbänder markieren eine stabile Region, die starken Sprünge im Diagramm kennzeichnen die Events im Skalenraum, wo zwei Regionen miteinander verschmelzen.

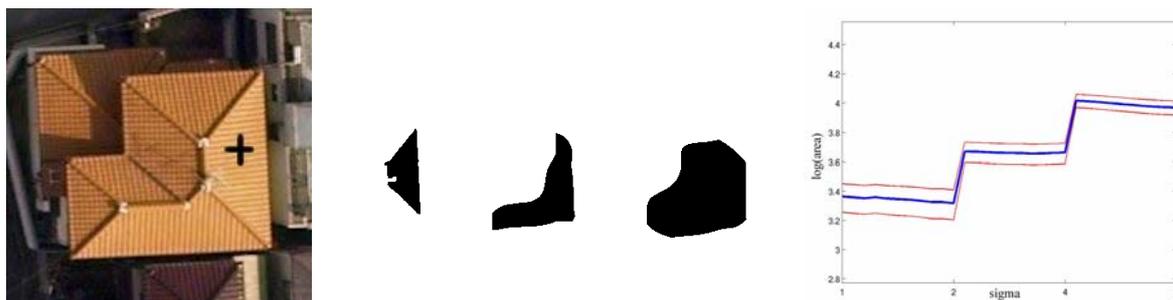


Abb. 6: Beispiel für Entwicklung des Flächeninhalts von Regionen im Skalenraum

4 Nachbarschaftsbestimmung der stabilen Regionen

Stabile Regionen haben im Skalenraum spezielle Eigenschaften, die mit der Implementierung des Wasserscheidenalgorithmus erklärt werden können. Bei kleinen, instabilen Regionen haben Rundungen erkennbare Auswirkungen auf Erscheinen und Form der Regionen. Bei stabilen Regionen gilt dagegen: Beim Verschmelzen benachbarter, stabiler Regionen geht immer eine Region vollständig in der anderen auf. Anderenfalls gäbe es eine Wasserscheide, die sich erst in höheren Skalen bemerkbar macht, d.h. der quadratische Gradient schrumpft bei zunehmender Glättung umgekehrt proportional zu seiner Größe. Das ist aber nicht der Fall.

Für den generischen Ansatz zur Gebäudererkennung benötigen wir auch die Nachbarschaften der stabilen Regionen. Dazu interpretieren wir alle Regionen als Hintergrund, die im gesamten Skalenraum instabil bleiben. Eine Nachbarschaftsbestimmung zuzüglich dieser instabilen Regionen ist nicht erforderlich, zudem werden so die Berechnungen beschleunigt.

Auf diese Weise muss der Regionennachbarschaftsgraph (RNG) nur einmal auf der untersten Ebene des Skalenraums berechnet werden. Die RNGs in den anderen Ebenen können leicht aus der untersten Ebene berechnet werden, weil die Nachbarschaftsverhältnisse beim Verschmelzen von zwei Regionen übertragen werden.

5 Literaturverzeichnis

BANGHAM, J.A., HARVEY, R., LING, P.D., AND R.V. ALDRIDGE, 1996: Morphological scale-space preserving transforms in many dimensions. *Journal of Electronic Imaging* 5 (3), S. 283-299.

- BRÜGELMANN, R., AND W. FÖRSTNER, 1992: Noise Estimation for Color Edge Extraction. - Förstner/Ruwiedel (Ed.): Robust Computer Vision, Wichmann, Karlsruhe, S. 90-106.
- CROWLEY, J.L., RIFF, O., AND J.H. PIATER, 2002: Fast Computation of Characteristic Scale Using a Half-Octave Pyramid. CogVis 2002, International Workshop on Cognitive Computing, Zürich.
- FÖRSTNER, W., 1994: A Framework for Low Level Feature Extraction. LNCS 802, Springer, S. 383-394.
- KOENDERINK, J., 1984: The Structure of Images. - Biological Cybernetic 50, S. 363-370.
- LINDBERG, T., 1994: Scale-Space in Computer Vision, 423 S., Kluwer, Dordrecht.
- LOWE, D.G., 2004: Distinctive Image Features from Scale-Invariant Keypoints. - International Journal of Computer Vision 60 (2), S. 91-110.
- MATAS, J., CHUM, O., URBAN, M. AND T. PAJDLA, 2002: Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. - Image and Vision Computing 22 (10), S. 761-767.
- OLSEN, O.F., 1996: Multi-Scale Segmentation of Grey-Scale Images, Master Thesis, University of Copenhagen, 89 S.
- VINCENT, L., AND P. SOILLE, 1991: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (6), S. 583-598.