

Analyse von Strategien und Kostenfunktionen beim Semi-Global-Matching in Luftbildpaaren

ROBERT MEICHSNER *

Zusammenfassung: Der Beitrag diskutiert eine Re-Implementation des Semi-Global-Matching-Ansatzes und analysiert die Qualität der Resultate, die mit unterschiedlichen Kostenfunktionen erzielt wurden. Bei der Auswertung von Stereoluftbildpaaren zeigte sich, dass bereits einfache Kostenfunktionen wie absolute Differenzen und Census hohe Korrelationen der Bildpixel ermöglichen. Durch bilaterale Filterung und Diskontinuitätsprüfungen ließen sich diese Ergebnisse noch deutlich verbessern.

Um nicht nur ganzzahlige Verschiebungsvektoren bestimmen zu können, wurden zusätzlich Subpixelalgorithmus eingeführt. Ein solches Verfahren ist die Subpixelinterpolation durch kubische Pfadeneinpassung. Dabei werden ausgehend von Diskontinuitätsgrenzen und charakteristischen Punkten der einzelnen Disparitätsstufen Kurven eingepasst, die die Subpixelanteile der einzelnen Bildpixel abschätzen.

Zusätzlich soll der Speicherbedarf des Semi-Global-Matching betrachtet werden. Um eine Berechnung auf herkömmlichen Rechnern zu ermöglichen, wurde eine Blockdefinition eingeführt, welche ein nahezu identisches Ergebnis wie Semi-Global-Matching selbst liefern kann. Der implementierte Ansatz wurde sowohl mit Stereoluftbildpaaren (Normalbilder) wie auch mit Bildpaaren in einer Nahbereichsanwendung getestet.

1 Einleitung

Ein Kernproblem der Computer Vision und ein wichtiger Schwerpunkt der Photogrammetrie ist das Finden und Herstellen von Punkt- bzw. Bildkorrespondenzen. Durch das Matching von charakteristischen Punkten in mehreren Bildern können relative Bewegungen detektiert und Rückschlüsse auf 3D-Punkte gezogen werden. Allerdings repräsentieren diese Punkte nur einen geringen Teil der jeweils in den Bildern vorhandenen Pixel. Ziel ist jedoch die flächendeckende Zuordnung von Punktpaaren. Eine Lösung dessen ist mit gängigen Verfahren der Photogrammetrie möglich.

Least-Square-Matching liefert bereits gute Ergebnisse, braucht hingegen aufgrund seiner iterativen Herangehensweise wesentlich mehr Rechenzeit als aktuellere Verfahren. Kreuzkorrelation stellt ein deutlich schnelleres Verfahren dar, jedoch wird keine globale Berechnung durchgeführt. Ein alternatives Verfahren ist das Semi-Global-Matching (SGM), welches durch seinen semi-globalen Ansatz lokale Punktkorrespondenzen abschätzt und global aufeinander bezieht [HIRSCHMÜLLER, 2005]. Es hat vor Allem im Nahbereich Stärken, wobei dort aufgrund der hohen Variabilität der Entfernung der Objekte in der jeweiligen Szene sowie das Schwanken von Bildmerkmalen zu aufwendigeren Analyse-Methoden wie beispielsweise Fuzzy-Methoden gegriffen wird [Hintzmann & Richardt, 2010].

* Technische Universität Dresden, Institut für Photogrammetrie und Fernerkundung, D-01062 Dresden, robert.meichsner@tu-dresden.de

Eine Anwendung in Luftbildern wird bereits in [HIRSCHMÜLLER, 2005] gezeigt. Die Genauigkeiten und Ergebnisse von Stereopaaren waren kein Gegenstand der Arbeit. Eine Analyse erfolgte in [HAALA, 2011], jedoch stellt der Einsatz aufwendiger Kostenfunktionen wie Mutual Information eine erhebliche Mehrbelastung der Rechenzeit dar.

In diesem Beitrag soll auf Grundlage von Stereoluftbildpaaren eine Einschätzung über geeignete einfache Kostenfunktionen getroffen werden, und durch Diskontinuitätsprüfung sowie durch eine eigens entwickelte Subpixelinterpolation eine Verfeinerung der Ergebnisse erreicht werden. Der Beitrag gliedert sich im Wesentlichen in die Berechnung eines Disparitätsraums (Abschnitt 2.1) sowie die Kostenaggregation (Abschnitt 2.2). Hierbei wird näher auf die Eigenschaften des SGM eingegangen. Die Ergebnisse des SGM sind dabei nicht fehlerfrei und sollen im Rahmen der Diskontinuitätsprüfung (Abschnitt 2.3) und Subpixelinterpolation (Abschnitt 2.4) verbessert werden. Um die Berechnung für Luftbildpaare zu ermöglichen, wurde eine Blockdefinition eingeführt (Abschnitt 2.5), die auf Grundlage der vorhandenen Rechnerressourcen Blockgrößen der Bilder abschätzt und das SGM blockweise durchführt. In Abschnitt 3 werden die Ergebnisse des SGM anhand eines Luftbildpaares und einer Nahbereichsaufnahme vorgestellt und hinsichtlich seiner Kostenfunktionen sowie Subpixelinterpolation gewertet.

2 SGM

Das Semi-Global-Matching (SGM) ist ein Kompromiss zwischen globalen und lokalen Stereo-Matching-Ansätzen und versucht, auf Grundlage von Stereobildern die Disparitäten aller Bildpixel zu optimieren. Disparitäten beschreiben hierbei die Verschiebungsvektoren korrespondierender Bildpunkte. Um einen Vergleich von Bildpixeln ermöglichen zu können, werden Kostenfunktionen eingeführt, die durch Wertung aller Pixel zueinander den Disparitätsraum aufspannen (Abschnitt 2.1). Der Algorithmus durchläuft dabei den Disparitätsraum sternförmig in bis zu acht Pfadrichtungen und versucht, auf dessen Grundlage eine globale Abschätzung der Disparitätswerte vorzunehmen. Dies wird im Rahmen der Kostenaggregation realisiert (Abschnitt 2.2).

2.1 Disparitätsraum und Kostenfunktionen

Bei der Berechnung des Disparitätsraums muss vorab über die Verschiebung der Bildebenen zueinander entschieden werden. Die Distanz zwischen korrespondierenden Pixeln der einzelnen Bildebenen wird in Folge dessen als Disparität D bezeichnet. Für rektifizierte Bildpaare (Normalbilder) entspricht die Disparität der Differenz der x -Koordinaten der Pixel. Wird dieser Gedanke weitergeführt und auf das gesamte Bild übertragen, resultiert eine Matrix, in welcher alle sich ergebenden Disparitätswerte aufgeführt werden können. Diese entspricht im Wesentlichen einem einfachen relativen Tiefenbild, indem der Abstand zwischen realen Punkten und Kamera den Verschiebungsvektoren der korrespondierenden Bildpunkte gegenübergestellt wird. Da jedoch zum derzeitigen Zeitpunkt noch keine Kenntnis über Korrespondenz einzelner Pixel existiert, wird ein Vergleich aller Pixel entlang der Epipolarlinie definiert. Die sich ergebende Disparitätsmatrix enthält nun alle Vergleichswerte aller Pixel entlang der jeweiligen Epipolarlinie, und umfasst somit eine Matrix mit den Abmessungen des Bildes sowie in der Tiefe alle möglichen Disparitätsschritte. Es ist erkennbar, dass mit zunehmender Größe des Disparitätsumfangs auch

die Größe der Matrix steigt. Eine Lösung dieses Problems wird in Abschnitt 2.5 (Blockdefinition) vorgestellt.

Die Art des Vergleichs der Pixel miteinander stellt zudem ein wesentliches Problem dar. Zum einen soll dieser wenige Ressourcen fordern und zum anderen dennoch sehr robust sein. Ein solcher Vergleich lässt sich mittels Kostenfunktionen herstellen. Diese berechnen auf Grundlage eines Vergleichskriteriums den Unterschied zwischen zwei Elementen. Bei den Absoluten Differenzen entspricht dies beispielsweise der Differenz der Intensitätswerte I und I' zweier Bildpixel. Jedoch muss in diesem Rahmen auch auf die zu verwendeten Bilddaten eingegangen werden. Die Wahl der Kostenfunktion C ist oft grundlegend verschieden zwischen Nahbereichs- und Luftbildphotogrammetrie. In Nahbereichsanwendungen existieren häufig untexturierte Bereiche, die das Aufspannen passender Disparitätsräume erschweren. Dieses Problem besteht in der Regel bei Luftbildaufnahmen von städtischen Bereichen nicht. Hier kann von einer hohen Variabilität in der gegebenen Situation ausgegangen werden. Das bedeutet, dass falsch zugewiesene Pixel innerhalb kleiner Regionen relativ unwahrscheinlich sind. Gerade aus diesen Gründen heraus lassen sich bereits mit sehr einfachen Kostenfunktionen gute Ergebnisse erreichen. Angewendet wurden daher sowohl die Absoluten Differenzen (AD) als auch Census [ZABIH & WOODFILL, 1994]:

$$\begin{aligned} C_{AD} &= \|I - I'\| \\ C_{Census} &= R_T(I) - R_T(I') \end{aligned} \quad (2.1)$$

$$\text{mit: } R_T(I) = \bigotimes_{[i,j] \in N} (I(i,j) > I)$$

Die Berechnung von Census erfolgt dabei über den Vergleich von Bitstrings R , welche durch den Vergleich zwischen dem zentralen Pixel I und den Pixeln in der Umgebung N erzeugt werden. Im Gegensatz zu Census reagiert AD sehr sensibel auf Fehlerpixel. Daher wurde eine Filterung der Ausgangsbilder mittels bilateralen Filters vorgenommen. Der Grundgedanke ist in einem allgemeinen Tiefpassfilter einen geometrischen und einen photometrischen Bezug herzustellen [TOMASI & MANDUCHI, 1998]. Unter dem geometrischen Bezug wird eine Gewichtung aller in Nachbarschaft zum zentralen Pixel gelegenen Bildpunkte verstanden. Diese Gewichtung wird mittels Gaußscher-Verteilungsfunktion realisiert. Der photometrische Bezug ist ein Grenzbereich um den Farbwert des zentralen Pixels. Alle Pixel, die außerhalb dieses Grenzbereichs liegen, werden ausgeschlossen. Es ergeben sich somit wesentliche Vorteile, indem zum einen Objektränder nicht verloren gehen und zum anderen das Rauschen innerhalb von Objekten bearbeitet werden kann.

2.2 Kostenaggregation

Bereits zu diesem Zeitpunkt ist es möglich, ein plausibles Tiefenbild des Ausgangsbildes zu erstellen. Die Minimierung des Disparitätsraums würde somit die theoretisch ähnlichsten Punkt-korrespondenzen aufzeigen. Diese stehen jedoch in keinerlei Nachbarschaftsbezug und weisen somit deutliche Diskontinuitäten auf. Um diesen Umstand zu korrigieren, wird eine globale

Energiefunktion E eingeführt [HIRSCHMÜLLER, 2008], welche die Abweichungen der Disparitätswerte in der Nachbarschaft mit Straftermen P_1 und P_2 belegt:

$$E(D) = \sum_p C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \quad (2.2)$$

mit: $P_2 \geq P_1$

Disparitätssprünge mit der Größe von bis zu einem Pixel werden mit P_1 bestraft, währenddessen alle anderen Sprünge mit P_2 bestraft werden. Um kleinere Disparitätssprünge zu ermöglichen, muss stets P_1 kleiner als P_2 sein.

$$L'_r(p, D) = C(p, D) + \min \left\{ \begin{array}{l} L'_r(p - r, D), \\ L'_r(p - r, D \pm 1) + P_1, \\ \min_i \{L'_r(p - r, i)\} + P_2 \end{array} \right\} - \min_k \{L'_r(p - r, k)\} \quad (2.3)$$

Der eigentliche Vorteil des SGM ist die Abstraktion der globalen Energiefunktion durch lokale Pfadfunktionen L_r (Formel 2.3), welche aufsummiert eine möglichst gute Annäherung an die globale Energiefunktion liefern sollen. Der Berechnungsaufwand kann somit entscheidend verringert werden. Im Wesentlichen entsprechen diese Funktionen dem Prinzip der dynamischen Programmierung [Hadley, 1969]. Es wird in verschiedenen Richtungen der minimale Kostenpfad eines Pixels p durch den Disparitätsraum bestimmt. Diese werden zunächst zu den Gesamtpfadkosten summiert und letztlich minimiert. Durch die Kombination verschiedener Richtungen lassen sich unter anderem auch streifige Artefakte unterbinden [HIRSCHMÜLLER, 2005]. Der sich ergebende minimierte Gesamtkostenraum entspricht den minimierten Disparitäten jedes einzelnen Pixels im Ausgangsbild. Da jedoch nicht alle Pixel korrespondieren, müssen fehlerhaft berechnete nachträglich eliminiert werden. Dies wird im Rahmen des Links-Rechts-Tests durchgeführt. Es wird hierbei geprüft, ob korrespondierende Pixelpaare dieselben Disparitätswerte besitzen. Ist dies nicht der Fall, werden diese als fehlerhaft angenommen und aus dem Ergebnis entfernt.

2.3 Diskontinuitätsprüfung

Der SGM-Ansatz ist besonders anfällig für größere untexturierte Bereiche, die an Objekte anderer Disparität grenzen. Die Gesamtkosten sind aufgrund der lokal definierten Bestrafungsterme für Gebiete gleicher Disparität in der Regel geringer, sofern diese ähnliche Kostenwerte aufzeigen. Es ist somit notwendig, untexturierte Bereiche zu extrahieren. Für diese gilt in der Regel auch, dass sich hier keine Diskontinuitäten befinden können. Folglich werden die Bilder nach Bereichen durchsucht, in denen nur eine sehr schwache Textur auftritt und die Diskontinuitäten beinhalten [HIRSCHMÜLLER, 2008]. Eine schwache Textur lässt sich in der Regel durch Histogrammanalysen der einzelnen Farbkanäle prüfen. Hierbei wird der jeweils entstandene

Graph auf Maxima getestet. Um diese herum wird eine Grenze von 20 Intensitätsstufen festgelegt. Die entstehenden Regionen beschreiben vorerst untexturierte Bereiche des jeweiligen Farbkanales. Existieren diese in allen Farbkänen, werden sie als absolute untexturierte Gebiete übernommen und auf Konnektivität geprüft. Treten darin Diskontinuitäten oder mehrere unabhängige Disparitätsebenen auf, werden die fehlerhaften Bereiche durch eine Nearest-Neighbor-Interpolation neu bestimmt.

2.4 Subpixelinterpolation

Eine wesentliche Einschränkung des SGM ist die ganzzahlige Schätzung der Verschiebungsvektoren. Interessanter sind jedoch Punktkorrespondenzen, die mit Subpixelgenauigkeit bestimmt werden können. Hierzu ist es notwendig, die sich ergebenden Werte des SGM weiter zu bearbeiten. Ein Ansatz, der eine solche Subpixelschätzung ermöglicht, ist die hier eingeführte Subpixelinterpolation durch kubische Pfadeinpassung. Der prinzipielle Gedanke ist, dass auf Grundlage charakteristischer Punkte des Tiefenbildes Kurven dritten Grades streifen- und zeilenweise interpoliert werden. Charakteristische Punkte sind hierbei die Diskontinuitätsgrenzen sowie die mittlere Sprungstellen der einzelnen Disparitätsebenen. Das sich ergebende genäherte Tiefenbild besitzt in der Regel an allen Stellen, an denen Pixel neu berechnet wurden, Subpixelanteile. Diese – für verschiedene Richtungen aufsummiert und gefiltert – ergeben schließlich ein verfeinertes Tiefenbild der gegebenen Situation. Es ist zu beachten, dass dieser Ansatz ausschließlich auf den Ergebnissen des SGM aufsetzt und somit keine Subpixelberechnung aufgrund nachträglich ermittelter Punktkorrespondenzen darstellt.

2.5 Blockdefinition

Das Semi-Global-Matching benötigt aufgrund seiner relativ speicherlastigen Berechnung des Disparitäts- sowie Pfadkostenraumes viele Ressourcen. Diese lassen sich wesentlich durch eine blockweise Berechnung reduzieren [HIRSCHMÜLLER, 2008]. Die Eingangsbilder werden auf Grundlage von deren Größe, des Überlappungsbereichs und der Tiefe des benötigten Disparitätsraums begrenzt und das SGM blockweise bearbeitet (Abbildung 1). Da jeder Block T_i für sich

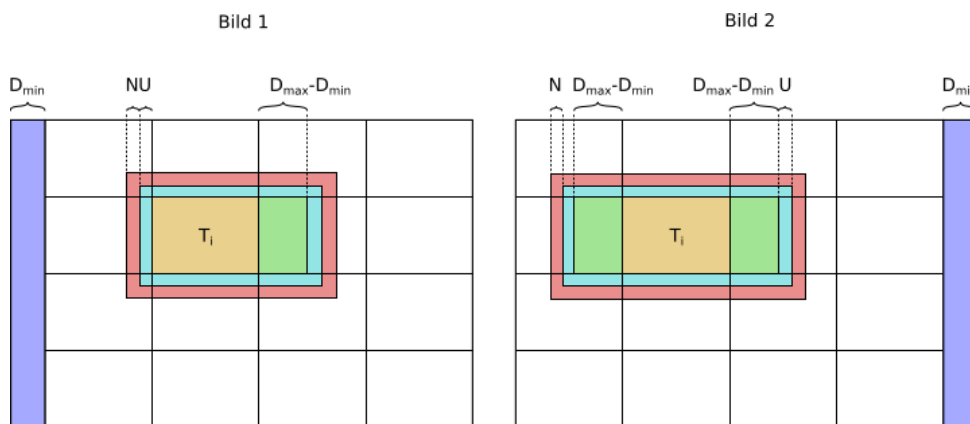


Abbildung 1: Bestimmung der Blöcke (T_i) mit deren Überlappungsbereich (U), Nachbarschaftsbereich (N) und Pufferzone ($D_{max} - D_{min}$)

autark berechnet wird, und somit gewisse Unterschiede zwischen den einzelnen Blöcken an deren Grenzen auftreten können, werden diese bereichsübergreifend definiert. Es kommt somit bei der Überführung in einander zu geringeren Unstimmigkeiten.

3 Ergebnisse

Die im Rahmen der Studie ausgewerteten Datensätze bestehen aus einem Stereoluftbildpaar (Normalbilder) sowie einem Bildpaar im Nahbereich (vorab rektifiziert). Für alle Datensätze wurde SGM sowohl mittels AD als auch mittels Census durchgeführt, wobei die Ausgangsbilder für den AD-Ansatz stets vorab bilateral gefiltert wurden (Abschnitt 2.1). Die Kostenfunktionen zeigen für das Luftbildpaar bereits ohne nachträgliche Korrektur ein gutes Ergebnis. In Abbildung 2a und 2b lassen sich Häusergrundrisse erkennen, sowie zwischen Gebäuden und anderen Objekten unterscheiden. Dennoch weisen die erstellten Tiefenkarten deutliche Fehler wie Ausreißer oder Lücken auf. Durch die Subpixelinterpolation sowie Diskontinuitätsprüfung und Ausreißertest konnte dieses Ergebnis nachträglich verbessert werden (Abbildung 2c, 2d). Es zeigen sich deutliche Unterschiede im Bereich größerer Ebenen (z.B. Straßen und Dächer) als auch an Objektgrenzen (z.B. Häuser, Bäume). Zur besseren Veranschaulichung wurde aus dem verfeinerten SGM-Resultat für Census ein 3D-Modell generiert und mit einer korrespondierenden Ansicht aus [MICROSOFT, 2013] verglichen (Abbildung 3).

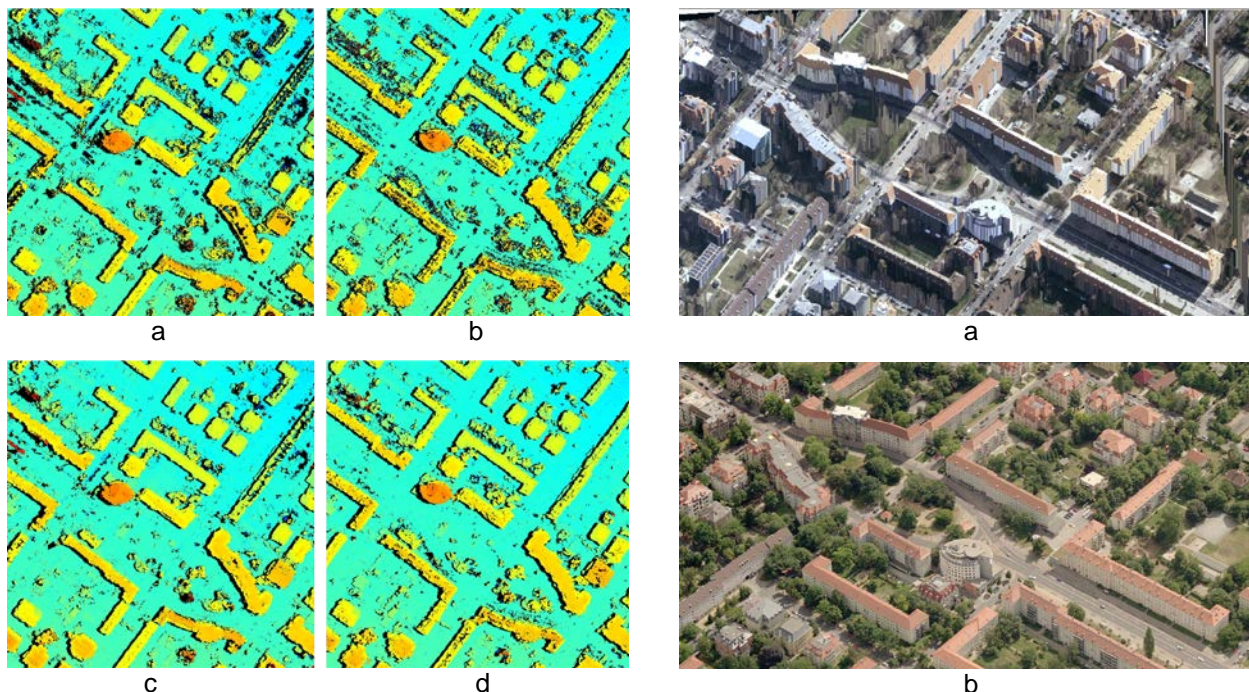


Abbildung 2: SGM-Resultat mittels AD (a) und Census (b), verfeinerte SGM-Resultate für AD (c) und Census (d)

Abbildung 3: generiertes Modell (a), korrespondierende Ansicht aus BING-MAPS (b)

Das entstandene Modell auf Grundlage von Normalbildern repräsentiert die geforderte Situation zuverlässig. Da jedoch für diesen Bereich keine Referenzdaten vorlagen, wurden für den qualitativen Vergleich der Kostenfunktionen beziehungsweise der Verbesserungen eines der in [SCHARSTEIN, 2012] veröffentlichten Bildpaare verwendet (Abbildung 4a, 4b). Für dieses exist-

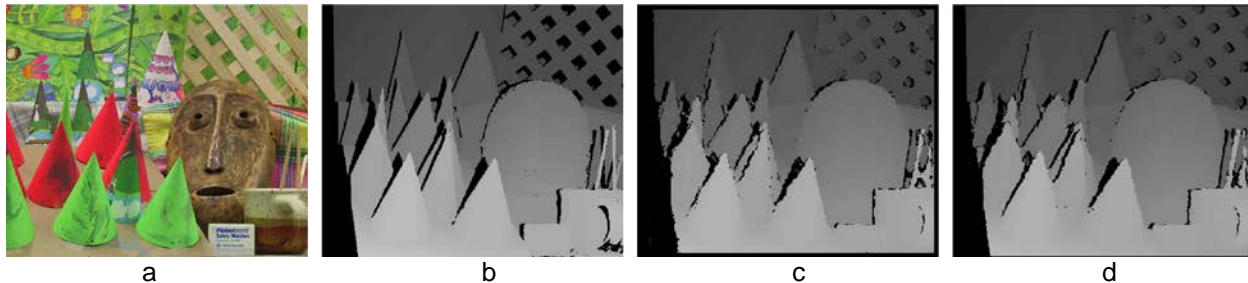


Abbildung 4: Ausgangsbild (a), gegebene Tiefenkarte mit reduziertem Wertebereich (b), berechnete Tiefenkarte mittels AD (c), berechnete Tiefenkarte mittels Census (d)

tiert bereits das berechnete relative Tiefenbild. Zudem ist es überwiegend gut texturiert und beinhaltet viele verschiedene Objekte. Es ermöglicht somit eine Einschätzung der Ergebnisse in Bezug auf das Luftbildpaar. Zum einfachen Vergleich wurden die erstellten Tiefenkarten mittels AD sowie Census der gegebenen gegenübergestellt (Abbildung 4c, 4d). Da die gegebene Tiefenkarte im Gegensatz zu den errechneten einen viermal so großen Wertebereich besitzt, wird für den Vergleich somit auf den berechneten Wertebereich reduziert. In Abbildung 5a lassen sich

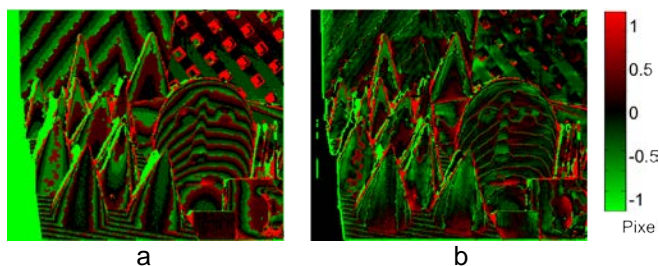


Abbildung 5: Vergleich SGM-Resultat für Census (a) und der Subpixelinterpolation (b)

daher gut die treppenförmigen Unterschiede zwischen der gegebenen Tiefenkarte und den berechneten erkennen.

Es ergab sich dabei für AD ein mittlerer Fehler von 0,26 Pixel und für Census 0,25 Pixel. Durch die Subpixelinterpolation konnte das Ergebnis nur bedingt verändert werden, da die grob fehlerhaften Stellen als Diskontinuitäten angenommen und somit über die Subpixelinterpolation nicht nachträglich berechnet wurden. Dennoch

konnte der mittlere Fehler für AD auf 0,25 Pixel und für Census auf 0,23 Pixel reduziert werden. Um eine bessere Wertung der Ergebnisse vornehmen zu können, wurde der fehlerhafte Anteil an Pixeln bestimmt, welche größer gleich 0,5 Pixel waren. Es ergab sich für AD 23,10 % sowie für Census 21,72 %. Der fehlerhafte Anteil konnte durch die Subpixelinterpolation für AD auf 10,87 % und für Census auf 9,12 % reduziert werden und zeigt somit eine wesentliche Verbesserung der Ausgangsdaten (Abbildung 5a, 5b). Erkennbar ist, dass mittels Subpixelinterpolation durch kubische Pfad einpassung im Raum gelegenen Ebenen geglättet und bereits strukturstarke Regionen weniger stark verfälscht wurden. Wird dieses Ergebnis auf das Luftbildpaar abstrahiert, lassen sich mit Census und Subpixelinterpolation gute Ergebnisse für das Stereomatching im Bereich der Luftbildphotogrammetrie erreichen.

4 Zusammenfassung

In diesem Beitrag wurde eine Re-Implementation des Semi-Global-Matching Ansatzes diskutiert und auf Qualität und Umsetzung verschiedener Kostenfunktionen in Bezug auf Luftbildpaare städtischen Bereichs eingegangen. Es konnte gezeigt werden, dass bereits mit einfachen Kostenfunktionen gute Ergebnisse erreicht werden können. Berechnet wurde hierbei auf Grundlage von Census und Absoluten Differenzen verschiedene Tiefenbilder aus Luftbildern und Bildern im Nahbereich. Der Ansatz kann jedoch nur ganzzahlige Verschiebungsvektoren liefern. Es wurde daher die Subpixelinterpolation über kubische Pfadeinpassung eingeführt, welche auf Grundlage von Diskontinuitätsgrenzen und charakteristischen Punkten der einzelnen Disparitätsstufen kubische Kurven einpasst, die die Subpixelanteile der einzelnen Bildpixel abschätzen. Es konnte gezeigt werden, dass mittels Subpixelinterpolation eine wesentliche Steigerung der Qualität der Ergebnisse erreicht werden kann.

5 Literaturverzeichnis

- HAALA, N., 2011: Multiray Photogrammetry and Dense Image Matching, Photogrammic Week 2011, Wichmann Verlag, Berlin/Offenbach, S. 185-195
- HADLEY, G., 1969: Nichtlineare und Dynamische Programmierung. Verlag der Wirtschaft, Berlin
- HINTZMANN, M., & RICHARDT, J., 2010: Anwendung des SGM-Algorithmus zur Bestimmung der Höhenprofile von Körpern mit Stereo-Vision. In: Tagungsband des 13. Anwendungsbezogener Workshop zur Erfassung, Modellierung, Verarbeitung und Auswertung von 3D-Daten, S. 111-122
- HIRSCHMÜLLER, H., 2005: Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), CVPR '05, Washington, DC, USA. IEEE Computer Society. S. 807–814
- HIRSCHMÜLLER, H., 2008: STEREO Processing by Semiglobal Matching and Mutual Information. IEEE Trans. Pattern Anal. Mach. Intell., 30(2): S. 328–341
- MICROSOFT, 2013: BING-MAPS, Screenshot – Nürnberger Ei, <http://www.bing.com/maps/>, abgerufen am 10. Juni 2013
- SCHARSTEIN, D., 2012: The Middlebury Computer Vision Pages, <http://vision.middlebury.edu/stereo/data/scenes2003/>, abgerufen am 02. September 2012
- TOMASI, C., & MANDUCHI, R., 1998: Bilateral Filtering for Gray and Color Images. In: Proceedings of the Sixth International Conference on Computer Vision, ICCV '98, Washington, DC, USA. IEEE Computer Society. S. 839–846
- ZABIH, R., & WOODFILL, J., 1994: Non-parametric Local Transforms for Computer Visual Correspondence. In: Proceedings of European Conference on Computer Vision, Stockholm, Sweden. S. 151-158