

# Abschätzung des Disparitätsraums für das Stereo-Matching mithilfe von Bildmerkmalen

Christoph Drexler

FernUniversität in Hagen  
Lehrgebiet Mensch-Computer-Interaktion  
christoph@drexler.eu

Art der Arbeit: Masterarbeit im Studiengang Master of Science (Praktische Informatik)  
Betreuer/in der Arbeit: Dr. Klaus Häming, Prof. Dr. Gabriele Peters

**Abstract:** Beim Stereo-Matching zwischen zwei Bildern einer Szene werden Disparitäten zwischen einzelnen Bildpunkten ermittelt. Dabei wird die Größe des Disparitätsraumes, in dem die Disparitäten gesucht werden, üblicherweise global vorgegeben. Diese Vorgangsweise stößt an ihre Grenzen, wenn die tatsächliche Größe des Suchraumes unbekannt ist, wie beispielsweise bei der Freihanderfassung mit unkalibrierten Kameras. Um die Anzahl der zu berechnenden Disparitätsstufen zu reduzieren, wird hier ein Algorithmus vorgeschlagen, der den Disparitätsbereich unter Zuhilfenahme von ohnehin vorhandenen Merkmalskorrespondenzen iterativ und für lokale Bildbereiche adaptiv abschätzt. In einem Benchmarking mit bekannten Datensätzen hat sich der Algorithmus als effizient und genau erwiesen.

## 1 Problemstellung

Die meisten einschlägigen Forschungsarbeiten zum Stereo-Matching<sup>1</sup> gehen davon aus, dass die beiden verwendeten Bilder von kalibrierten Kamerapositionen in exakt paralleler Position, aus demselben Abstand und unter kontrollierten Beleuchtungsbedingungen erstellt wurden. Aufgrund der parallelen Ausrichtung kann die Differenz dann durch einen einzelnen Wert („Disparität“), der die horizontale Verschiebung der beiden Bildpunkte in Pixeln angibt, beschrieben und mittels sog. Disparitätskarten optisch dargestellt werden. Die Anzahl der in Frage kommenden Disparitätsstufen hängt dabei nur von den Abständen der beiden Kamerapositionen zueinander und zum fotografierten Objekt ab und wird beispielsweise bei den oft zum Testen verwendeten Middlebury-Datensätzen (vgl. <http://vision.middlebury.edu/stereo/submit/>) vorgegeben.

Bei der Freihanderfassung von 3D-Objekten treffen einige dieser idealisierten Bedingungen nicht zu. Vor allem ist es äußerst unwahrscheinlich, dass die Kameraachsen der beiden Bilder parallel zueinander verlaufen, sie werden vielmehr fast immer windschief

---

<sup>1</sup> Beim Stereo-Matching geht es darum, in zwei Bildaufnahmen derselben Szene jene Pixelpaare zu identifizieren, die denselben Punkt der Szene abbilden, um aus der Positionsdimension der beiden Punkte auf Tiefeninformationen bzgl. der abgebildeten Objekte zu schließen, vgl. [Sz11].

zueinander stehen. Während sich das Problem, dass die zusammengehörigen Bildpunkte dann meistens nicht mehr auf derselben Bildzeile liegen, relativ einfach durch eine sog. Rektifizierung lösen lässt, bleibt ein weiterer Nebeneffekt bestehen: Bei parallelen Kameraachsen kommen nur relativ wenige Disparitätswerte in Betracht, sofern von einem bestimmten Mindestabstand der Kamera zur aufgenommenen Objektszene ausgegangen werden kann (was meistens der Fall ist). Bei windschiefer Position der Kameraachsen sind die in Frage kommenden Disparitätswerte vorerst nur durch die horizontale Pixelanzahl der Bilder begrenzt. Aber nicht nur theoretisch, sondern auch praktisch führt der unterschiedliche Betrachtungswinkel zu einer deutlich höheren Zahl an Disparitätsstufen. Da bei fast allen Verfahren zum Stereo-Matching die Berechnungen für eine bestimmte Anzahl möglicher Disparitätsstufen durchgeführt werden und die Rechenkomplexität bei der Freihandfassung von 3D-Objekten möglichst gering gehalten werden soll, stellt sich an dieser Stelle die Frage, ob und wie die Anzahl der zu berechnenden Disparitätsstufen effizient reduziert werden kann.

## 2 Algorithmus zur Abschätzung des Disparitätsraums

Während das genannte Problem u. U. auch gelöst werden könnte, indem die Disparitätswerte für das Bildpaar zuerst in geringerer Auflösung berechnet und dann durch Hochskalieren verfeinert werden, wird hier ein Lösungsweg vorgeschlagen, der bei der Überlegung ansetzt, dass ja bereits vor dem eigentlichen Stereo-Matching, nämlich bei der Rektifizierung der Bilder, eine Reihe von Bildpunkten einander zugeordnet wurden, deren Disparitätswerte sich sehr einfach berechnen lassen. Um ein Bildpaar überhaupt rektifizieren zu können, werden nämlich üblicherweise zuvor mittels eines sog. Feature-Detektors besonders auffällige Merkmale der Objektszene ermittelt. Erst wenn zumindest sieben „sichere“ Merkmalspaare identifiziert worden sind, die einander mit hoher Wahrscheinlichkeit zugeordnet werden können, ist eine Rektifizierung der Bilder möglich.

Allerdings werden bei dieser Suche meist nicht nur sieben Merkmalspaare ermittelt, sondern sehr viel mehr. Die Idee des hier vorgeschlagenen Algorithmus ist es, von diesen Disparitätswerten auszugehen und auf deren Grundlage einen Disparitätsraum für die übrigen Bildbereiche festzulegen, der die korrekten Disparitätswerte mit hoher Wahrscheinlichkeit enthält, aber trotzdem deutlich kleiner ist als der theoretisch mögliche Suchraum, sodass der Rechenaufwand – bei annähernd gleich guten Ergebnissen – deutlich verringert werden kann.

Dazu werden in einem ersten Schritt zu jedem Merkmal, dessen Zuordnung als korrekt angenommen werden kann, fünf Nachbarmerkmale gesucht. Als Kriterium der „Nähe“ wird die maximale Farbdistanz<sup>2</sup> zwischen je zwei Pixeln, die auf einer Linie zwischen den beiden Merkmalen liegen, verwendet. Diese Vorgangsweise soll sicherstellen, dass Punkte, die zur selben Fläche gehören, gegenüber Punkten, die an verschiedenen Seiten von Diskontinuitäten liegen, bevorzugt werden. An der Position des Merkmals wird nun

---

<sup>2</sup> Die Farbdistanz wird als  $D(p, q) = \max(|r_p - r_q|, |g_p - g_q|, |b_p - b_q|)$  bestimmt, wobei es sich bei  $r_p, r_q, g_p, g_q, b_p, b_q$  um die Farbwerte der beiden Bildpunkte handelt.

die jeweilige Disparität als untere und obere Schranke für den Suchraum eingetragen. Die Disparitätswerte für die Bildpunkte auf den Geraden zwischen je zwei benachbarten Merkmalspaaren werden linear interpoliert<sup>3</sup>. Diese Ersteinschätzung der Disparität wird nun im nächsten Schritt zu einem Suchraum erweitert, indem die untere Schranke um einen fixen Wert erniedrigt, die obere Schranke um denselben Wert erhöht wird.

Als nächstes werden die ermittelten Disparitätssuchbereiche auf benachbarte Bildpunkte übertragen (Abb. 1a). Dabei sollen die Suchbereiche möglichst nur an jene Nachbarpixel weitergegeben werden, die ungefähr die gleiche Disparität aufweisen. Dabei möchte man den Suchraum nicht über Diskontinuitäten hinaus ausweiten. Da man bei der für lokale Stereo-Matching-Algorithmen wichtigen Kostenaggregation auch bemüht ist, die Kosten nicht über Diskontinuitäten hinaus aufzuaddieren, ist es sinnvoll, dort Anleihen zu nehmen. Daher werden gemäß der Logik der „Cross-based Aggregation“ (vgl. [Me11]) Unterstützungsregionen gebildet, die allerdings nicht wie im ursprünglichen Verfahren zur Kostenaggregation verwendet werden, sondern zur Ausweitung des Disparitätsraums. Dabei wird der Suchraum zuerst in horizontaler und dann in vertikaler Richtung so lange ausgedehnt, wie die Farbdistanz zweier benachbarter Pixel sowie die Farbdistanz zum ursprünglichen Pixel, dessen Suchbereich ausgedehnt wird, einen festgelegten Schwellenwert nicht überschreitet. Der Schwellenwert wird umso niedriger gewählt, je weiter das in Frage kommende Pixel vom Ausgangspixel entfernt ist, sodass die Ausbreitung des Suchraums in stärker strukturierten Bildregionen schneller stoppt als in texturlosen. Falls der Bildpunkt, auf den hin der Suchraum bei diesem Verfahren ausgedehnt wird, bereits eine Einschätzung für eine untere und obere Schranke besitzt, wird diese nicht überschrieben, sondern entsprechend erweitert. Um bereits in der ersten Iterationsstufe Disparitätswerte für möglichst viele Pixel berechnen zu können, wird dieser Schritt dreimal – immer zuerst horizontal und dann vertikal – durchgeführt.

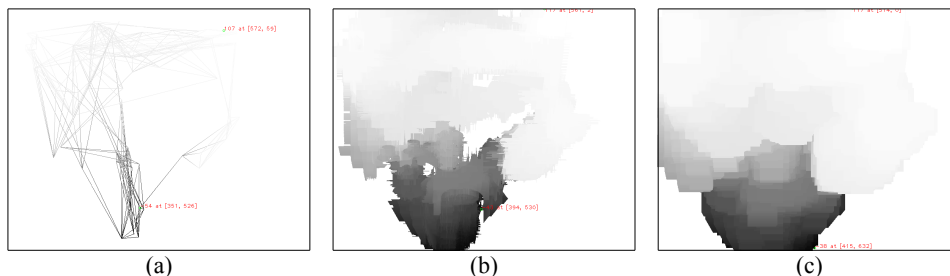


Abbildung 1: Schrittweise Einschätzung des Disparitätsraums (obere Suchraumgrenze) anhand des Bildpaars „Wohnraum“ (Abb. 2): (a) Initialisierung mittels Interpolation zwischen benachbarten Merkmalen; (b) Ausdehnung des Suchraums mittels der Unterstützungsregionen; (c) Ausdehnung um eine fixe Distanz in allen vier Richtungen. Dunkel = niedrige, hell = hohe Disparitätswerte.

Auf diese Weise entsteht – abhängig von der Anzahl der zur Verfügung stehenden Merkmale sowie von den Eigenschaften der Szene und der Fotografien – eine zusammenhängende Teilmenge an Pixeln, für die eine erste Einschätzung bzgl. des Disparitätsraums getroffen wurde (Abb. 1b). Diese Teilmenge weist allerdings an den Rändern viele Fransen auf, was sich für die nachfolgende Kostenaggregation nachteilig auswirkt.

<sup>3</sup> Die Verwendung der Farbdistanz als Kriterium trägt an dieser Stelle dazu bei, dass auch weiter entfernte Punkte miteinander verbunden werden und so eine größere Zahl an Disparitätsvermutungen zustande kommt.

Deshalb werden in einem abschließenden Schritt (Abb. 1c) die Disparitätsräume noch einmal auf eine fixe Anzahl von Nachbarpixeln in allen vier Hauptrichtungen ausgedehnt, wiederum zuerst horizontal und dann vertikal.

Für jene Pixel, für die bereits ein Suchraum festgelegt worden ist, wird nun das Stereo-Matching durchgeführt. Die dabei ermittelten Disparitäten werden verwendet, um damit den nächsten Iterationsschritt der Suchraumbestimmung zu initialisieren. Der Prozess wird bis zur Erfüllung eines definierten Abbruchkriteriums (z. B. bis eine bestimmte Prozentzahl an Pixeln berechnet wurde) wiederholt.

### 3 Evaluation

Der o. g. Algorithmus wurde in einer Referenzimplementierung<sup>4</sup> erprobt und evaluiert, und zwar mittels der Bildpaare „Venus“, „Teddy“ und „Cones“ der Middlebury-Datensätze sowie mit Bilddaten des Lehrgebiets Mensch-Computer-Interaktion der Fernuniversität in Hagen<sup>5</sup>. Die Ergebnisse zeigen, dass der Suchraum durch den beschriebenen Algorithmus effizient eingeschätzt und eingeschränkt werden kann, weil die untere und obere Disparitätsschranke nicht nur für das Gesamtbild, sondern für jedes einzelne Pixel bestimmt werden. Probleme ergeben sich, wenn nur wenige Merkmalspaare gefunden werden, allerdings würde das Stereo-Matching dann wohl schon zuvor an der Rektifizierung scheitern.

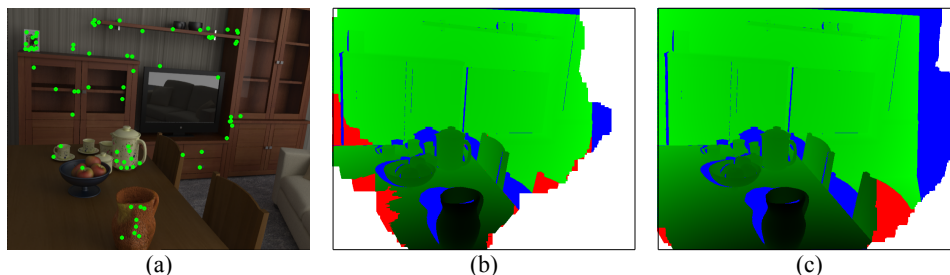


Abbildung 2: Ergebnisse für das Beispiel „Wohnraum“: (a) verwendete Merkmale; (b)–(c) geschätzte Disparitätsräume im ersten und zweiten Iterationsschritt (grün: korrekte Disparität liegt innerhalb des Suchraums; rot: falsche Schätzung; blau: verdeckte/ungültige Regionen. Fehler des ersten Schritts werden im zweiten korrigiert. Schwierigkeiten hat der Algorithmus hier v. a. mit kontrastarmen, weil schattigen Bildbereichen.

### Literaturverzeichnis

- [Sz11] Szeliski, R.: Computer Vision – Algorithms and Applications. Springer, London, 2011.
- [Me11] Mei, X. et al.: On Building an Accurate Stereo Matching System on Graphics Hardware. In *GPUCV'11: ICCV Workshop on GPU in Computer Vision Applications*, 2011.

<sup>4</sup> Die Implementierung ist unter <https://bitbucket.org/ChristophDrexler/masterarbeit> frei verfügbar.

<sup>5</sup> Die Bilddaten wurden von Dr. Klaus Häming erstellt. Sie berücksichtigen spezifische Besonderheiten der Freihandfassung von 3D-Objekten: Beleuchtungsdifferenzen, windschiefe Kameraachsen, Bildrauschen.