

Seminararbeit  
zum Thema

Verdeckungsanalyse in Stereobildern  
unter der Verwendung von  
Disparity-Space-Images  
und  
Dynamischer Programmierung

Christian Thiel, Januar 2003.

Verfasst im Rahmen des Seminars „3-D-Rekonstruktion“ im WS 2002/03 am Lehrstuhl  
Praktische Informatik IV der Universität Mannheim.

Betreuer: Dirk Farin.

## Inhalt:

1. Übersicht	3
2. Einleitung	3
2. Disparity-Space-Images (Verschiebungs-Raum-Bilder)	4
3.1 Verdeckungsanalyse und DSI Pfadbedingungen	5
3.2 Wie findet man den besten Pfad?	6
3.3 Ground Control Points (GCP)	7
4. Kanten im DSI	9
5. Verwendung der Kanten im DSI-Ansatz	11
5.1 Keine Kosten für Verdeckungskanten: Ein degenerierter Fall	11
5.2 Konzentration auf verdeckte Bereiche	12
5.3 Nicht-zyklopische Ansichten: Stereoebenen	14
6. Zusammenfassung	15
7. Referenzen	16

## 1. Übersicht

Dieser Arbeit liegen die Paper von Stephen Intille und Aaron Bobick [9, 15] zu Grunde. Vorgestellt wird eine Methode, die Informationen über „intensity edges“, d.h. Kanten die durch Intensitätsübergänge in Bildern entstehen, zur Gewinnung von Informationen über Verdeckungsregionen auf Basis eines pixelbasierten Stereoalgorithmus nutzt. Dazu werden die Konstruktion und der Nutzen von „Disparity-Space-Images“ bei der Lösung des Stereoverdeckungsproblems gezeigt. Des Weiteren wird auf den Einfluss von „intensity edges“ auf Disparity-Space-Images eingegangen und ein erweiterter Stereoalgorithmus wird entwickelt, um das Wissen über Intensitätsschwankungen an verdeckten Kanten darin zu nutzen. In diesem Ansatz wird abschließend vorgeschlagen, für die Analyse von Verdeckungsregionen zwei separate Berechnungen und Ergebnisse – eines für die rechten Verdeckungsregionen und eines für die linken Verdeckungsregionen – einem konventionellen zyklischen Prozess vorzuziehen.

## 2. Einleitung

Eine genaue Verdeckungserkennung ist eine der wertvollsten Aufgaben, die ein Stereoalgorithmus durchführen kann. Verdeckungen trennen die Welt in zwei Ebenen – dreidimensional isolierte Regionen – welche die Basis für eine natürliche Auflösung einer Szene sein können. Unglücklicherweise hat der Augenmerk der Stereoforschung auf der Gewinnung von dichten, genauen Tiefenkarten von Szenen geruht – in der die Verdeckungsanalyse generell nur ein Nachgedanke war, verschoben bis das Stereomatching fertig war.

Neuere Arbeiten über Stereoverdeckungen haben aber gezeigt, dass das Berechnen von Verdeckungen direkt in den Stereomatchingalgorithmus mit einbezogen werden kann [1, 7, 5, 9]. Stereobildmaterial enthält sowohl verdeckte Kanten als auch verdeckte Regionen [7]. Verdeckte Regionen sind räumlich zusammenhängende Pixelgruppen die in einem Bild erscheinen und im anderen nicht. Diese verdeckten Regionen sind durch verdeckte Oberflächen verursacht und können direkt für Stereo- und Verdeckungsschlussfolgerungen verwendet werden.

Auch monookulare Bilder beinhalten Informationen über Verdeckungen. Verschiedene Objekte haben unterschiedliche Oberflächenstrukturen, Farben und Ausleuchtungen. Deswegen generieren Sprünge zwischen solchen Objekten fast immer „intensity edges“, d.h. Kanten an Intensitätsübergängen, in einer monookularen Aufnahme. In dieser Ausarbeitung wird die Entwicklung einer Stereomatchingmethode und einer Verdeckungsanalyse beschrieben, die den Zusammenhang zwischen Verdeckungssprüngen und „intensity edges“ vereint.

Wir beginnen mit einer kurzen Beschreibung eines Disparity-Space-Algorithmus, der in [9] im Detail beschrieben wird. Nachdem der Einfluss der „intensity edges“ auf die Disparity-Space-Images gezeigt wurde, werden wir die Stereo-Technik erweitern, so dass sie Informationen über Intensitätssprünge und Verdeckungskanten nutzt, während die Verdeckungsbedingung, definiert durch die verdeckten Regionen im pixelbasierten Matchingprozess, weiter verstärkt wird. Diese Erweiterung erlaubt eine genaue Untersuchung von Verdeckungsregionen in beiden Bildern des Stereopaars.

## 2. Disparity-Space-Images (Verschiebungs-Raum-Bilder)

Ein Disparity-Space-Image (DSI) ist eine explizite Wiedergabe des Übereinstimmungsraumes; es ist verwandt mit Bildern aus früheren Arbeiten [4, 7, 14, 11]. Die DSI-Darstellung für die  $i$ -te Scanlinie ist wie folgt: Nimm die  $i$ -te Scanlinie aus dem linken und dem rechten Bild, nennen wir diese Scanlinien  $s_i^L$  bzw.  $s_i^R$ , und schiebe diese Pixel für Pixel übereinander. In jedem dieser Schritte werden die Scanlinien von einander subtrahiert und das Ergebnis wird in die nächste Zeile des DSI geschrieben. Die DSI-Darstellung speichert somit das Ergebnis der Subtraktion jedes Pixels in  $s_i^L$  von jedem Pixel in  $s_i^R$  und behält dabei die räumlichen Beziehungen zwischen Punkten. Es kann also als ein  $(x, \text{Verschiebung})$  Übereinstimmungsraum angesehen werden, mit den  $X$ -Werten entlang der horizontalen Achse und der Verschiebung entlang der Vertikalen. Gegeben zwei Bilder  $I_L$  und  $I_R$ , so berechnet sich der Wert des DSI durch:

$$DSI_i^R(x, d) = I_R(x, i) - I_L(x + d, i) \quad (1)$$

für  $0 \leq (x + d) < N$ ,  $0 \leq d < D$ , and  $0 \leq x < N$ . Das hochgestellte  $R$  in  $DSI^R$  zeigt an, dass es sich um das rechte DSI handelt. Das  $DSI^L$  dieser Scanlinie ist einfach die verschoben, gescherte Version des  $DSI^R$ . Diese zwei Bilder werden verwendet, entgegen einer zyklischen Repräsentation wie sie in [9] angewendet wurde, um explizit den Unterschied zwischen verdeckten Regionen (Pixel die nur in einem Bild sichtbar sind) und Verdeckungssprüngen (plötzliche Sprünge der Verschiebung einem Bild, im Zusammenhang mit einer verdeckten Region im anderen Bild) deutlich zu machen.

Um das DSI robuster gegen Rauschen zu machen, kann die Subtraktion auch durch Korrelationen ersetzt werden, indem eine vereinfachte Version der adaptiven Fenster [10] dafür genutzt wird, die bei Verdeckungssprüngen scharfe Grenzen im DSI erhält [9]<sup>1</sup>.

Bild 1a ist ein Testbildpaar einer schräg ansteigenden „Hochzeitstorte“, deren Tiefenprofil in Bild 1b gezeigt wird. Abbildung 1c zeigt das Korrelations-DSI für eine Scanlinie durch die Mitte des Testbildpaares; Werte, die nahe Null lagen, wurden für das menschliche Auge hervorgehoben. Zu beachten sind die charakteristischen Streifenmuster, die aus dem festhalten einer Scanlinie und dem „Darüber-her-ziehen“ der anderen resultieren. Wenn eine Region mit ausgeprägter Oberflächenstruktur auf der linken Scanlinie über die korrespondierende Region der rechten Scanlinie gezogen wird, so kann man im DSI eine Linie mit Matches sehen (dunkle Pixel). Werden zwei oberflächenstrukturarme korrespondierende Regionen übereinander gezogen, so wird eine diamantenförmige Region von „Nahe-Null-Matches“ sichtbar. Der „Null-Pfad“ durch das DSI, der zu der Verschiebung in Beziehung steht und die Höhenlinie widerspiegelt, kann als schwarze Linie gefunden werden, die die blockartigen Segmente im DSI verbindet.

---

<sup>1</sup> Wenn wir die korrekte Korrelation benutzen, dann wird implizit von einer gewissen Konsistenz über Scanlinien hinweg ausgegangen (z.B. [11, 5]), da sich die dafür benutzten Fenster vertikal über mehrere Scanlinien hinweg ausdehnen. Das ist aber nicht in jedem Fall gewünscht, da dies eine nicht gewollte Unschärfe des Ergebnisses nach sich zieht.

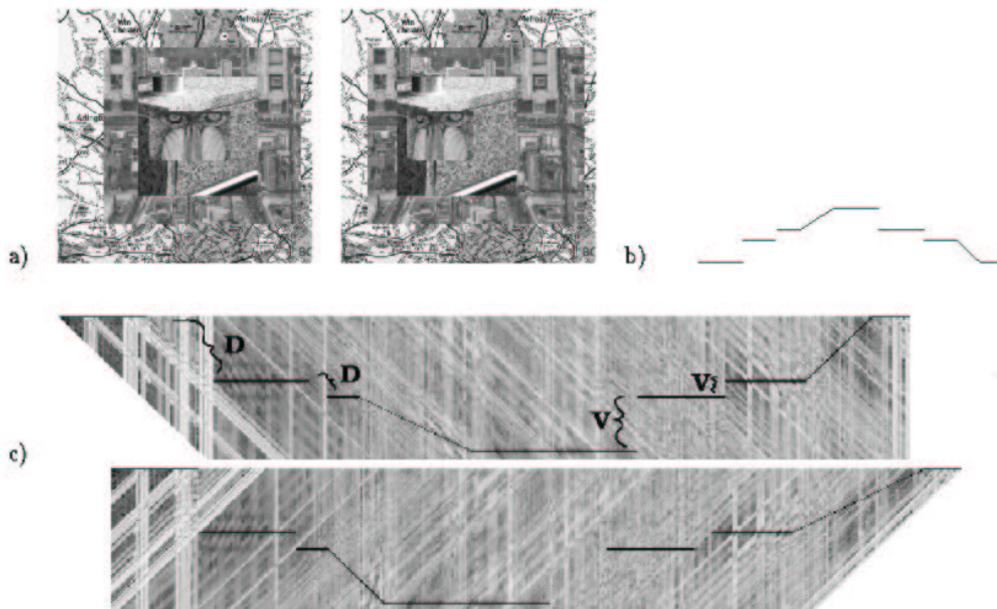


Abbildung 1: (a) Ein simuliertes Bildpaar in Form einer schräg ansteigenden "Hochzeitstorte", deren Tiefenprofil in (b) abgebildet ist. (c) das verbesserte, beschnittene  $DSI^L_i$  und  $DSI^R_i$  für eine Scanlinie durch die Mitte der Torte.

### 3.1 Verdeckungsanalyse und DSI Pfadbedingungen

In einer diskreten Formulierung des Stereomatchingproblems hat jede Region mit nicht konstanter Verschiebung dazugehörige nicht gematchte Pixel. Wegen diesen Verdeckungsregionen kann der „Null-Match-Pfad“ durch das DSI nicht kontinuierlich verlaufen. Die mit „D“ gekennzeichneten Regionen in Bild 1c markieren horizontale Lücken in der verstärkten Nulllinie in den  $DSI^L_i$  und  $DSI^R_i$ . Diese Spalten korrespondieren jeweils mit verdeckten Regionen im linken und rechten Bild. Die mit „V“ gekennzeichneten Spalten im DSI markieren vertikale Sprünge von Verschiebung zu Verschiebung. Diese Sprünge entstehen durch rechte und linke verdeckte Kanten. Wir benutzen diese „Verdeckungsbedingung“ [7] um den Match-Pfad, der im DSI gefunden werden kann, einzuschränken: Jedes mal, wenn eine verdeckte Region oder Kante erkannt wird, muss der gefundene Pfad den entsprechenden diagonalen oder vertikalen Spalt überbrücken.

Fast alle Stereoszenen unterliegen der „Ordnungsbedingung“ (oder auch Monotoniebedingung [7] genannt): Ist Objekt a links von Objekt b im linken Bild, so wird es auch im rechten Bild links von Objekt b liegen. Dünne Objekte, bei gleichzeitig großer dazu gehörender Verschiebung (d.h. Sie liegen nahe an der Kamera) können diese Regel verletzen, aber solche Anordnungen sind selten. Setzt man die Ordnungsbedingung als erfüllt voraus, so kann man eine zweite Bedingung für den „Null-Match-Pfad“ durch das DSI angeben, die die Komplexität des „Finde-den-Pfad-Problems“ signifikant reduziert. Bewegt man sich im  $DSI^L_i$  von links nach rechts, so können diagonale Sprünge nur nach vorne (schräg nach unten) verlaufen und vertikale Sprünge können nur zurück (nach oben) springen. Im  $DSI^R_i$  ist diese Beziehungen umgekehrt: Bewegt man sich von links nach rechts, so können diagonale Sprünge nur nach schräg hinten und vertikale Sprünge können nur nach vorne (unten) springen.

### 3.2 Wie findet man den besten Pfad?

Die Überdeckungs- und Ordnungsbedingungen schränken die möglichen Lösungspfade durch den Verschiebungsraum stark ein. Von jedem Punkt im DSI aus gibt es nur drei Richtungen, die der Pfad nehmen kann – ein horizontaler Match, eine diagonale Verdeckung oder eine vertikale Verdeckung. Diese Beobachtung erlaubt es, einen Stereoalgorithmus zu entwickeln, der die Matchanalyse und die Verdeckungsanalyse in einem Prozess vereinigt.

Dieser Algorithmus, der den besten Pfad durch das DSI finden soll, ist als dynamische Programmierung (DP) Finde-Pfad-Problem im  $(x, \text{Verdeckung})$ -Raum formuliert. Wir wollen, unter der Annahme der Verdeckungs- und Ordnungsbedingung einen Weg mit minimalen Kosten durch das DSI finden. Diese Bedingungen schränken die Richtung, die ein Pfad vom aktuellen Endpunkt aus nehmen kann, stark ein. Wenn die Entscheidung, welcher Pfad (von jedem Pixel aus) zu wählen ist, nur auf den Kosten jeden möglichen Pfades der von dort aus genommen werden kann, treffen (und nicht aufgrund von Wegen, denen bis jetzt gefolgt wurde), so sind damit die DP-Vorraussetzungen erfüllt und DP kann benutzt werden um den optimalen Pfad zu finden.

Durch die Analyse der DSI-Struktur kann man das Verdeckungsproblem in verschiedene mögliche Zustände unterteilen: Wenn das DSI auf der Suche nach dem optimalen Pfad durchquert wird, so kann man sich immer nur in einem der drei folgenden Zustände befinden: Match(M), vertikale Verdeckung (V) oder diagonale Verdeckung(D). Bild 2 zeigt symbolisch die möglichen Zustandsübergänge. Der Pfad ist darüber hinaus an den Kanten des DSI weiteren Einschränkungen unterworfen, da an diesen Stellen einige dieser Zustandsübergänge nicht zugelassen sind (der Pfad darf das DSI nicht verlassen).

Kosten werden jedem Pixel des Pfades abhängig vom aktuell gültigen Zustand zugewiesen. Der DP-Algorithmus ist so gestaltet, dass er die Kosten des Pfades minimiert, wobei die Kosten eines Matches dem Absolutwert des DSI-Pixels an dieser Stelle entsprechen. Je besser der Match, desto kleiner werden die dafür veranschlagten Kosten. Der Algorithmus wird versuchen, die Anzahl der „guten“ Matches im endgültigen Pfad zu maximieren. Da der Algorithmus aber auch nicht matchende Pixel für den Pfad aussuchen kann – die verdeckten Regionen – so müssen nicht passenden Pixel in den vertikalen und diagonalen Sprüngen ebenfalls Werte zugewiesen werden. Anderenfalls würde der „beste Pfad“ fast keine Matches enthalten.

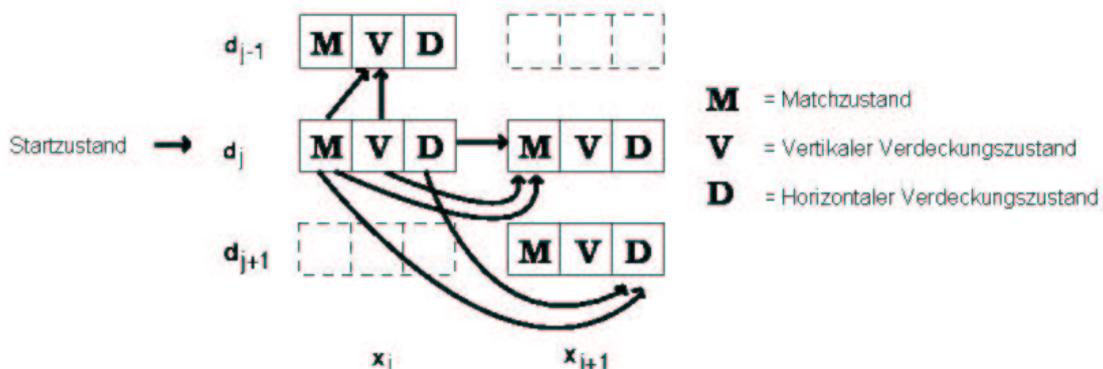


Abbildung 2: Zustandsdiagramm der möglichen Zustandsübergänge, die beim Durchlaufen des DP-Algorithmus auf dem Weg durch das DSI auftreten können.

### 3.3 Ground Control Points (GCP)

Unglücklicherweise verändern schon leichte Variationen bei den, diesen verdeckten Pixeln zugewiesenen, Kosten den kompletten Minimum-Match-Pfad durch das DSI, insbesondere bei verrauschtem Ausgangsmaterial [5]. Da diese Kosten für jedes im Pfad enthaltene, verdeckte Pixel angenommen werden, verhalten sich die Kosten eines verdeckten Bereiches linear zur Breite dieses Bereiches.

Um diese Anfälligkeit gegenüber den zugewiesenen Verdeckungskosten zu durchbrechen, muss eine weitere Nebenbedingung zusätzlich zu den Verdeckungs- und Ordnungsbedingungen eingeführt werden. Trotzdem soll hier nicht, wie in früheren Ansätzen geschehen, die Lösung anhand irgendwelchen, willkürlich ausgedachten Werten wie z.B. „smoothness“ [7], Konsistenz über Scanlinien hinweg [11, 5] oder „Intra-Scanlinien-Güte“ [5], nachgebessert werden.

Stattdessen wird in [15] vorgeschlagen, Punkte mit einer „sehr vertrauenswürdigen Matching Vermutungen“, sog. *Ground Control Points (GCP)*, einzuführen. Durch diese Punkte wird der Pfad gezwungen und macht dadurch große Sprünge über Verschiebungen hinweg, die ansonsten wegen der hohen Verdeckungskosten umgangen worden wären.

Bild 3 illustriert diese Idee. Zu sehen sind zwei GCP's und eine Anzahl möglicher Pfade von einem zum anderen. Beachten Sie, dass unabhängig davon welcher dieser Pfade durch die Verschiebung gewählt wird, die diskrete Anordnung erzwingt, dass jeder Pfad - ob A, B, oder C - sechs verdeckte Pixel enthält. Deswegen verursacht jeder dieser Pfade dieselben Verdeckungskosten. Der Algorithmus wird denjenigen Pfad wählen, der die Kosten der beteiligten Pixel minimiert, unabhängig davon, wo die Verdeckungssprünge anfallen und unabhängig davon, welche Kosten bei diesem Sprung entstehen. Bei einer einzigen verdeckten Region zwischen den GCP's, wie im originalen Bild, ist der beste Pfad ähnlich wie A oder B. Wenn aber der Bereich zwischen zwei GCP's leicht ansteigend verläuft, dann wird ein Pfad wie C, mit kleinen unterbrochenen Verdeckungssprüngen, vorgezogen. Der Pfad durch den (x, Unähnlichkeit)-Raum ist darum nur durch die Verdeckungs- und Ordnungsbedingung und die Güte der Matches zwischen den GCP's beschränkt. Eine Ausnahme von dieser Situation entsteht, wenn der Algorithmus zusätzliche verdeckte Regionen vorschlägt wie in Pfad D; solche falsche Lösungen haben in der Regel sehr viel höherer Kosten als die korrekte Lösung.

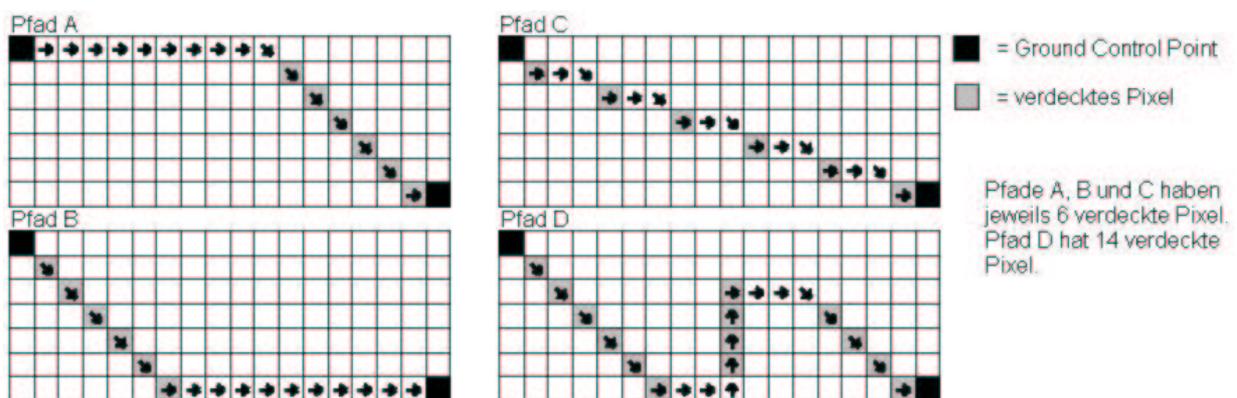


Abbildung 3: Vier mögliche Wege durch zwei GCP's.

Bild 4-a zeigt das „Birken-Bild“ aus der JISCT Stereotestmenge [2]. Die verdeckten Bereiche in diesem Bild sind schwer richtig zu erkennen, wegen der schlanken Bäume, den vielen oberflächenstrukturarmen Regionen und einem fünfzehnprozentigen Helligkeitsunterschied zwischen beiden Bildern. Wenn GCP's nicht benutzt werden, machen die schlanken Bäume

die Überdeckungserkennung teilweise sehr sensitiv gegenüber den Verdeckungskosten. Der Grund dafür ist, dass es an diesen Stellen relativ wenig gute Matches auf jedem der dünnen Bäume gibt, verglichen mit der Größe der Verdeckungssprünge von Baum zu Baum. Bild 4b zeigt das Ergebnis des Algorithmus ohne die Verwendung von GCP's. Die Verdeckungskosten hinderten den Pfad bei den meisten Scanlinien daran zu einigen der Bäume zu springen. Bild 4c zeigt den Algorithmus bei gleichen Verdeckungskosten unter der Verwendung von GCP's. Die meisten der verdeckten Regionen rund um die Bäume sind gut erkannt worden, weil die GCP's auf den Baumoberflächen die Abhängigkeit von den Verdeckungskosten eliminierten. Es sind trotzdem noch einige Fehler in diesem Bild. Einige schattige Bereiche des Birken-Bildes sind komplett mit Nullwerten ausgewaschen. Als Folge führten einige dieser Bereiche zu falschen GCP's, die wiederum inkorrekte Verdeckungssprünge im Ergebnis zur Folge hatten. Dieses Problem kann durch eine Änderung im GCP-Selektionsalgorithmus minimiert werden. An als GCP's vorgeschlagenen Stellen sollte dann dieser verbesserte GCP-Selektionsalgorithmus die Oberflächenstruktur vergleichen [8]. Auf einigen Scanlinien wurden keine GCP's entdeckt, was zu Lücken in den Scanlinien bei einigen Bäumen führte<sup>2</sup>.

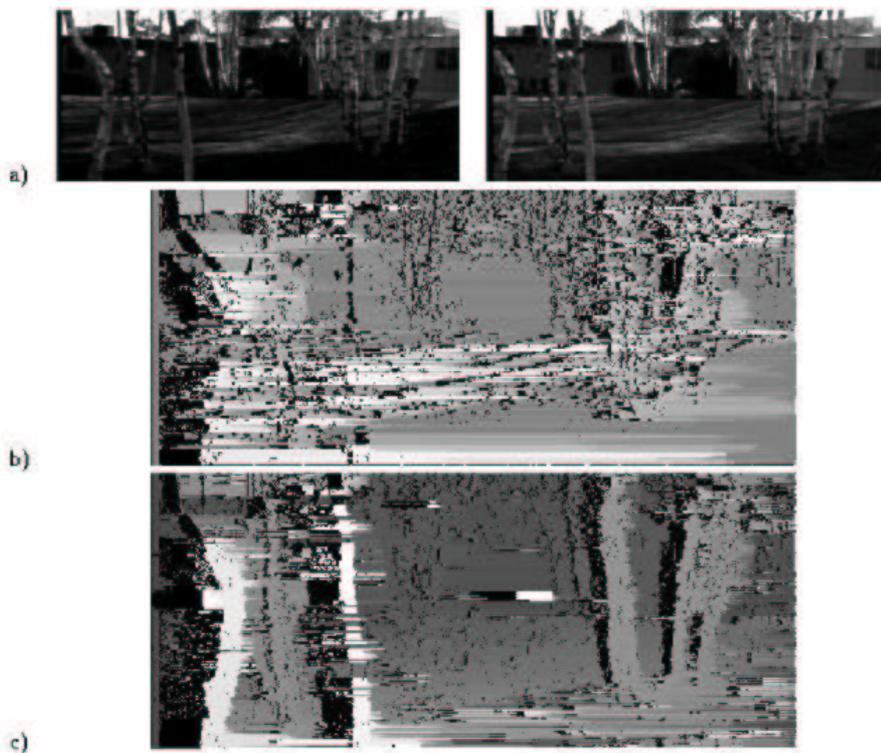


Abbildung 4: (a) Das "Birken"-Stereobildpaar, welches Teil der JISCT Stereotestmenge ist [2], (b) Ergebnis unseres Stereoalgorithmus ohne GCP's und (c) Ergebnis unseres Algorithmus unter der Verwendung von GCP's.

Der auf Verdeckungen basierende und auf GCP's arbeitende Stereoalgorithmus führte zu mindestens gleichwertigen, meist sogar deutlich besseren Ergebnissen wie andere ähnliche Methoden, wobei die Komplexität des Algorithmus und die Sensitivität gegenüber den Verdeckungskosten reduziert wurde. Des Weiteren wurde hier vermieden, zusätzliche künstliche Kriterien – wie z.B. „smoothness“ - einzuführen, die das berechnete Ergebnis nachträglich ausgleichen. Aber bis jetzt wurden noch nicht die wichtigen Kanteninformationen ausgenutzt. Indem das getan wird, kann diese Methode weiter verbessert werden.

<sup>2</sup> Bei der Verarbeitung der Birken-Datenmenge wurde Subtraktion, nicht Korrelation, benutzt um das DSI zu erstellen. Deswegen wurde darin keine künstliche Konsistenz über Scanlinien hinweg verursacht.

## 4. Kanten im DSI

Abbildung 5a zeigt ein Stereobildpaar eines Mannes und zweier Kinder. In Bild 5b ist das dazugehörige  $DSI^L_i$  für eine Scanlinie durch die Beine des Mannes wiedergegeben. Es sind sowohl vertikale als auch diagonale Streifen im DSI zu erkennen. Diese linienartigen Streifen entstehen immer bei großen Änderungen in der Intensität (z.B. an „Kanten“) in der linken oder der rechten Scanlinie. Im  $DSI^L_i$  gehören die vertikalen Streifen zu großen Änderungen der Intensität im linken Bild, die diagonalen Streifen gehören zu großen Änderungen in der Intensität im rechten Bild. Das Umgekehrte gilt entsprechend für das  $DSI^R_i$ . Da die inneren Regionen von Objekten dazu tendieren, kleinere Intensitätsänderungen zu zeigen, als Objektkanten, neigt auch die Subtraktion einer inneren Region einer Scanlinie von einer Intensitätskante der anderen Scanlinie dazu, die Kantenstruktur intakt zu lassen. Das Verbleiben der Kanten führt zu der linearen Struktur in DSI's. In [15] werden diese Linien in einem DSI „Kantenlinien“ genannt.

Wie schon erwähnt, neigen die Verdeckungsgrenzen dazu, an Intensitätskanten aufzutreten. Verdeckungen in DSI's zeigen sich durch diagonale oder vertikale Sprünge im Stereo-Lösungs-Pfad. Wenn eine Verdeckungskante mit einer Intensitätskante zusammenfällt, *dann wird die Verdeckungsspalte des DSI-Stereo-Ergebnisses auch mit einer Linie im DSI, definiert durch die dazugehörige Intensitätskante, zusammenfallen*. Bild 6a und 6b zeigen ein DSI und das dazugehörige „Kantenlinien“-Bild, das den darin enthaltenen, linienartigen Streifen entspricht. Abbildung 6c zeigt die Lösung, die für diese Scanlinie gefunden wurde, Abbildung 6d blendet die gefundene Lösung über das Linienbild. Die vertikalen und diagonalen Verdeckungen im DSI verlaufen entlang den Linien, die im DSI-Kantenbild auftreten.

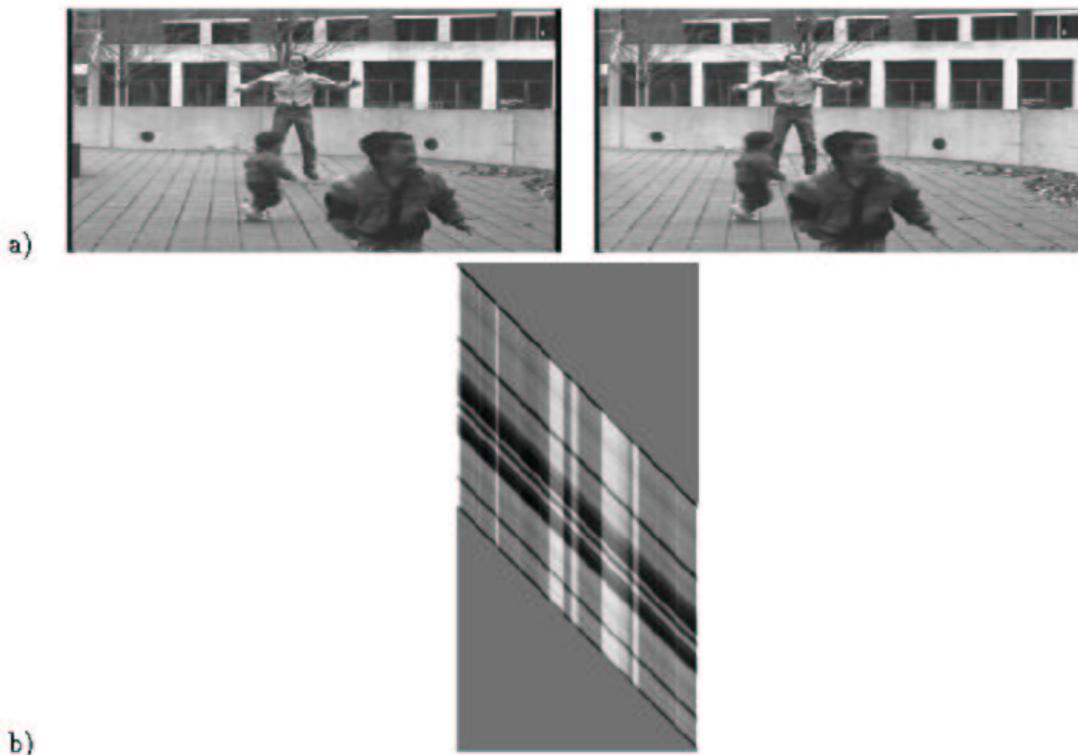


Abbildung 5: (a) Ein Stereobildpaar eines Mannes und zweier Kinder. (b) Ein Subtraktions-DSI des Stereobildes für eine Scanlinie durch den Rumpf des Mannes. Zu Beachten sind die diagonalen und vertikalen Streifen, die durch Intensitätsschwankungen in den Ausgangsbildern hervorgerufen werden. Diese Kantenlinien treten an den Kanten verdeckter Bereiche auf.



Abbildung 6: (a) Ein beschnittenes, durch Subtraktion entstandenes DSI. (b) Diese Linien entsprechen den linienartigen Streifen in (a). (c) Der gefundene Pfad. (d) Der Pfad und das Bild aus (b) übereinandergelegt. Der Pfad entlang den Verdeckungen entspricht dem Pfad entlang der Linien.

Im nächsten Abschnitt wird die Entwicklung einer Technik beschrieben, die diese Linien in das „Dynamische Programmierung“-Verfahren der vorherigen Kapitel integriert. Das Ziel ist es, das Ergebnis dahingehend anzugleichen, dass fast alle wesentlichen Verdeckungen entsprechend korrespondierende Intensitätskanten im DSI aufweisen.

Bevor der Algorithmus die Kanteninformationen verwenden kann, müssen diese Kanten im DSI gefunden werden. Die Linienerkennung im DSI ist eine relativ leichte Aufgabe, im Prinzip muss ein Kantenerkennungsalgorithmus nur nach diagonalen und vertikalen Linien suchen. Für erste Experimente wurde in [15] ein solcher Kantensfinder implementiert. Aber der ineffektive Weg, Kanten in den DSI's für jede Scanlinie zu suchen, führte dazu, einen Kantendetektionsalgorithmus zu suchen, der in einem einzigen Durchlauf Kanten in allen DSI's erkennt.

Die dafür in [15] vorgeschlagene Heuristik ist, einen Standard-Kantenerkennungsalgorithmus auf jedes Bild des Ausgangsmaterials anzuwenden und aus den dabei gefundenen Kanten ein Kantenlinienbild für jedes DSI zu generieren. Verwendet wurde dazu ein vereinfachter Canny-Kantenerkennungsalgorithmus, der mögliche Kanten im linken und rechten Bild [3] findet. Danach wurden die vertikalen Komponenten dieser Kanten kombiniert, um Kantenlinien herauszufinden.

Die Verwendung eines standardisierten Kantenoperators führt eine weitere Bedingung in die Stereolösung ein, die bis jetzt absichtlich darin vermieden wurde: Konsistenz über Scanlinien hinweg. Weil jeder räumliche Operator dazu neigt, kohärente Kanten zu finden, wird das Ergebnis der Verarbeitung einer Scanlinie nicht länger unabhängig von den ihr benachbarten Scanlinien sein. Da aber die Konsistenz über Scanlinien hinweg nur durch deren Bezug auf Kanten und Verdeckungen ausgelöst wird, wird in [15] diese Angleichung hingenommen und im Gegenzug die effiziente maschinelle Verarbeitung in einem einzigen Durchlauf erreicht.

## 5. Verwendung der Kanten im DSI-Ansatz

Das Ziel ist es, die Kanteninformationen derart in den „Dynamische Programmierung“-Ansatz einzuarbeiten, das 1.) die Lösung in korrekter Weise angeglichen wird, so dass Verdeckungen an Intensitätskanten vorgeschlagen werden; 2.) die vorher eingeführte Verdeckungsordnung nicht verletzt wird; und 3.) die Verarbeitungskosten des gesamten Algorithmus nicht signifikant erhöht wird.

Wie gezeigt treten die Segmente des Lösungspfades durch das DSI, die Verdeckungen anzeigen, normalerweise an Kantenlinien im DSI auf. Deswegen ist es eine einfache Strategie den Verdeckungs-Such-Algorithmus unter Einhaltung der drei oben genannten Kriterien zu verbessern, die Kosten einer Verdeckung entlang der im DSI gefunden Kanten zu reduzieren.

Abbildung 7 illustriert diese Kostenreduzierung. Angenommen, ein GCP oder ein Bereich guter Matches liegt auf beiden Seiten des Verdeckungssprunges. Die zu den Intensitätskanten korrespondierenden Kantenlinien im DSI sind im Diagramm als gepunktete Linien zu erkennen. Die feinen durchgezogenen Linien geben einige mögliche Pfade, die im Einklang mit der Kantenbedingung stehen, wieder. Werden die Kosten einer Verdeckung entlang einer Kantenlinie signifikant reduziert, so wird der durch die dunkle durchgezogene Linie markierte Pfad am „preiswertesten“ und dieser Pfad wird den verdeckten Bereich an der richtigen Stelle wiedergeben.

Durch die Kostenreduzierung entlang der Linien wird die Verdeckungserkennung verbessert, ohne das dadurch weitere Verarbeitungskosten bei dem Algorithmus entstehen, außer einer Vorverarbeitung der Kanten im originalen Bildpaar. Das Matchen wird immer noch anhand der Pixelwerte durchgeführt, wird aber dort, wo es angebracht ist, durch Kanteninformationen beeinflusst. Außerdem verhindern die GCP's, das Intensitätskanten, die nicht durch Verdeckungen hervorgerufen wurden, unbegründete Verdeckungen in der Lösung mit den minimalen Kosten verursachen. Der einzig verbleibende Punkt ist, auf welche Art und Weise die Verdeckungskosten entlang der Kantenlinien zu reduzieren sind.

### 5.1 Keine Kosten für Verdeckungskanten: Ein degenerierter Fall

Eine simple Methode die Verdeckungskosten entlang der Kantenlinien zu verringern, ist einfach die Kosten der verdeckten Pixel zu reduzieren, unabhängig davon, ob sie auf vertikalen oder diagonalen Kantenlinien liegen. Sicherlich wird eine Reduzierung der Kosten um jede beliebige Menge den Algorithmus dazu verleiten, Verdeckungen die im Zusammenhang mit Intensitätskanten auftreten, vorzuschlagen. Solange jedoch die Kosten einer Verdeckung entlang einer Linie frei wählbar sind, besteht die Möglichkeit, dass irgendwo entlang des Verdeckungspfades zufällig ein falscher, aber trotzdem guter Match liegt, der die Verdeckungsregion durchbrechen kann. In Abbildung 7 liegt der durch den Algorithmus vorgeschlagene Pfad nahe an der gepunkteten Linie, jedoch kann er immer noch zwischen Verdeckungszustand und Match-Zustand hin und her springen, abhängig von den vorliegenden Daten. Wichtiger noch, die einfache Reduzierung der Verdeckungskosten wie oben beschrieben führt wieder eine Überempfindlichkeit gegenüber der Höhe dieser Kosten ein – das Ziel der Einführung von GCP's war es, diese zu eliminieren.

Wenn der gepunktete Pfad in Abbildung 7 kostenfrei wäre, würden zufällige, gute Matches die gefundene Verdeckungsregion nicht beeinflussen. Es kann ein Algorithmus definiert werden, bei dem alle vertikalen und diagonalen Verdeckungssprünge, die mit einer Kantenlinie korrespondieren, keine Kosten haben. Diese Methode würde sicherlich die Erkennung von Verdeckungen entlang der Linien fördern.

Unglücklicherweise ist diese Methode ein degenerierter Fall. Der DP-Algorithmus wird eine Lösung vorschlagen, die die Anzahl der Verdeckungssprünge durch das DSI maximiert und die Anzahl der Matches minimiert, unabhängig davon, wie gut diese Matches sein mögen. Abbildung 8a illustriert wie Null Kosten für vertikale, wie auch diagonale

Verdeckungssprünge zu annähernd keinen richtigen Matches führt. Abbildung 8b zeigt, dass dieser degenerierte Fall möglicherweise mit echten Kamera- und Objektkonfigurationen zusammenfallen kann. Der Algorithmus hat eine mögliche Lösung gefunden. Das Problem jedoch ist, dass er dabei große Mengen gut matchender Daten ignoriert, da er die Verdeckungen an anderer Stelle vermutet.

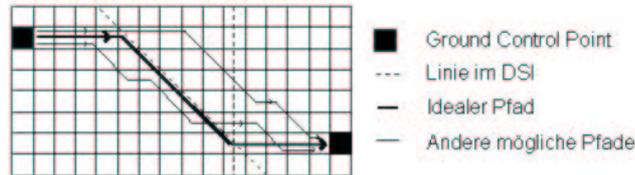


Abbildung 7: Diese Abbildung illustriert, wie die Kostenreduzierung entlang der Linien, die im DSI auftreten (durch gepunktete Linien dargestellt), die Verdeckungserkennung verbessern kann. Angenommen, der Bereich zwischen den zwei GCP's ist vertauscht, so stellen die dünnen, durchgezogenen Linien mögliche Pfade dar, die der Algorithmus aussuchen könnte. Werden die Kosten, die Verdeckungen zugewiesen werden, reduziert, so wird höchstwahrscheinlich der hervorgehobene Pfad ausgesucht. Dieser Pfad wird die Verdeckung sauber lokalisieren, Start- und Endpunkte liegen an den richtigen Stellen.

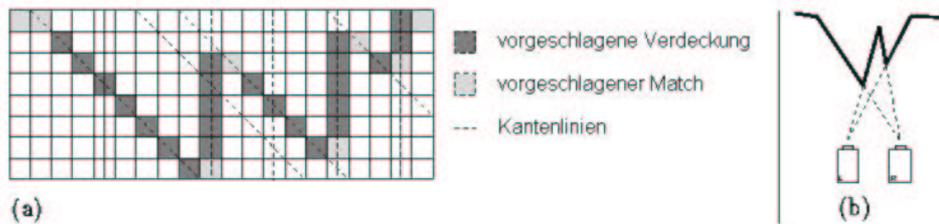


Abbildung 8: (a) Werden die Verdeckungskosten an vertikalen und diagonalen Kantenlinien auf Null gesetzt, so wird der Pfad die Anzahl der möglichen Verdeckungen maximieren und die Anzahl der Matches minimieren. Auch wenn solche Fälle in realen Daten existieren, wie (b) gezeigt, so generiert die Kostenfreiheit von vertikalen und diagonalen Verdeckungen doch diese Lösung sogar dann, wenn genug matchende Daten existieren, die eine andere, wahrscheinlichere Lösung zu unterstützen.

## 5.2 Konzentration auf verdeckte Bereiche

In den vorhergehenden Abschnitten wurde demonstriert, dass man das Verfolgen von vertikalen und diagonalen Linien im DSI nicht gleichzeitig kostenlos erlauben darf. Auch der Kompromiss, einfach die Verdeckungskosten entlang beider Linienarten zu verringern, führt trotzdem zu unerwünschten Abhängigkeiten von diesen Kosten. Da es eines der Ziele des vorgestellten Ansatzes ist, verdeckte Bereiche gut zu finden, wählen wir für die diagonalen Segmente Kosten Null, während wir für die vertikalen Segmente die normalen Verdeckungskosten beibehalten. Das erwartete Resultat ist, dass die verdeckten Bereiche, die mit den diagonalen Spalten im DSI zusammenfallen, gut herausgefunden werden sollten, während die Verdeckungskanten (die vertikalen Sprünge) nicht verändert werden. Des Weiteren erwarten wir keine erhöhte Sensibilität gegenüber den Verdeckungskosten.

Abbildung 9a zeigt ein künstlich erzeugtes Stereobildpaar aus der JISCT Testmenge [2] das einige Bäume in einer Felsenlandschaft wiedergibt. Abbildung 9b zeigt die Verdeckungsregionen, die durch unseren Algorithmus ohne Verwendung von GCP's und Kanteninformationen gefunden werden. Das Ergebnis ist vergleichbar mit früheren Ansätzen (z.B. aus [5]). Die schwarzen Verdeckungsgebiete um die Bäume und Felsen werden zumeist gefunden, aber die Grenzen dieser Bereiche sind oft nicht gut definiert und es verbleiben einige größere Fehler. Abbildung 9c zeigt das Ergebnis unter Verwendung von GCP's, aber ohne Kanteninformationen. Die dramatische Verbesserung zeigt die Mächtigkeit des GCP-Ansatzes. Abbildung 9d zeigt das Ergebnis unter Anwendung von GCP's und Kanteninformationen. Obwohl die Verbesserung nicht annähernd so groß ist wie zwischen

Abb. 9b und 9c, so ist die Lösung doch besser. Zum Beispiel wurden die Streifen an der linken Kante der Krone des rechten Baumes reduziert. Generell wurden die Verdeckungsregionen fast perfekt gefunden, mit nur wenigen oder gar keinen Streifen innerhalb dieser Regionen. Wenn auch der Effekt durch Verwendung der Kanteninformationen alles in Allem gering ist, so ist es doch wichtig, dass die durch Verdeckungen hervorgerufenen Unterbrechungen so abgestimmt wurden, dass sie an genau den richtigen Stellen auftreten.

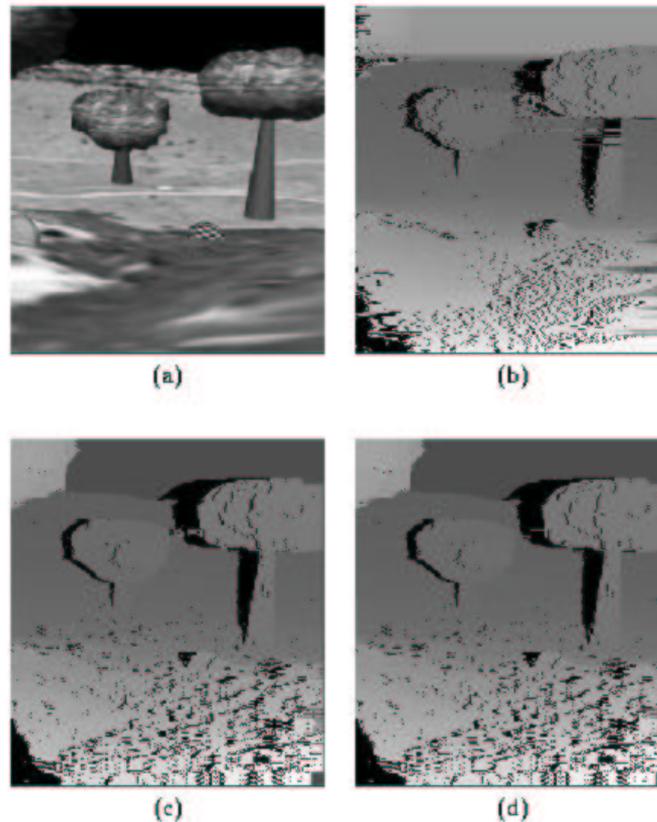


Abbildung 9: (a) Künstlich erzeugtes Bild zweier Bäume (linker Teil eines Stereobildes). (b) Verdeckungsergebnis ohne GCP's oder Kantenlinien. (c) Ergebnis unter der Verwendung von GCP's (d). Resultat unter der Verwendung von GCP's und Kantenlinien

### 5.3 Nicht-zyklopische Ansichten: Stereoebenen

Verdeckte Regionen können dazu benutzt werden, Objekten Grenzen und Objektebenen zuzuordnen. Für viele, wenn nicht die meisten, Applikationen sind qualitative Informationen die die Szene in Ebenen aufteilen nützlicher als präzise, quantitative Tiefangaben. Objektsegmentierungs-, Bewegungsanalyse-, Objekterkennungs-, und Bildkodierungsalgorithmen würden von Tiefenebenen profitieren, die die scharfen, scherenschnittartigen Objektgrenzen beibehalten [13].

Die meisten früheren Arbeiten, die sich mit der Erkennung von Tiefensprüngen in Stereobildern beschäftigten (z.B. in [6]) schlagen irgendwelche Operationen auf vorher berechneten Tiefkarten vor. Terzopolous, zum Beispiel, schlägt eine Interpolationstechnik vor [12], bei der eine dünne Metallplatte an Stellen mit hoher Dehnung durchbrochen wird. Aber eine sehr viel direktere Methode der Tiefensprungerkennung ist es, einen expliziten Nutzen aus den gewonnenen verdeckten Bereichen zu ziehen.

Die Abbildungen 10a und 10c zeigen ein Stereobild eines Mannes und zweier Kinder, diese Art Bild hat sehr große Verdeckungsgebiete. Für die meisten Anwendungen im Bereich der Computervisualisierung, wie z.B. Objektverfolgung oder -Modellierung, ist eine Segmentierung der Arme und Beine des Mannes eine wertvollere Information als der exakte Verlauf der Beine. Der äußere Umriss des Mannes und seine Zuordnung zu einer Tiefenebene in Bezug auf andere Objekte in der dargestellten Szene (der Hintergrund und die zwei Kinder im Vordergrund) ist eine kritische Information, die zuverlässig von Stereoalgorithmen mit eingebauter Verdeckungsanalyse geliefert werden kann.

Die hier vorgestellte Methode nutzt Kanteninformationen asymmetrisch, da sie sich auf die Erkennung der verdeckten Regionen (im Gegensatz zur Kantenerkennung) in den Bildern des Stereopaars konzentriert. Deswegen generiert dieser Algorithmus zwei Verdeckungskarten, jeweils eine darauf optimiert, eine bestimmte Art der Verdeckung zu finden. Abbildungen 10b und 10d zeigen die Stereoverschiebungen und die verdeckten Bereiche, die entdeckt werden, wenn man das  $DSI^L$  bzw. das  $DSI^R$  als Grundlage für den Algorithmus nimmt. Studiert man die rechten und linken Verdeckungsgebiete in diesen Abbildungen, so wird der Vorteil einer Zwei-Bild-Lösung deutlich. Im linken Ergebnis (10c) sind die linken Verdeckungsgebiete korrekt, die rechten Verdeckungskanten haben schlechtere Qualität, wie es gut an der rechten Kante des linken Beines der Mannes (markiert als Bereich A) zu erkennen ist. Im zweiten Bild der Lösung (10d) sind jedoch die rechten Verdeckungsgebiete desselben Beines präziser erkannt worden (markiert als Bereich B). Außerdem werden auch andere Stellen der Verdeckungserkennung durch die Zweifach-Lösung verbessert, wie z.B. die linke Kante des Kopfes des rechten Kindes. Das rechte Verdeckungsergebnis ergab eine „ausgefranzte“ linke Verdeckungskante (markiert als Bereich C) aber das linke Verdeckungsergebnis findet die linke verdeckte Region und die Grenzen des Kopfes des Kindes sauber (markiert als Bereich D). Anstatt einer einzigen, gemittelten, zyklischen Rekonstruktion der verdeckten Bereiche und Verdeckungskanten wird in [15] zwei vorgeschlagen: Eine mit nahezu perfekten linken Verdeckungsgebieten und die andere mit nahezu perfekten rechten Verdeckungsgebieten.

In [15] wurde gerade erst damit begonnen, die Idee der Stereoebenen zu erforschen und es wurde noch keine Methode konstruiert, um aus den Verdeckungskarten Tiefenebenen zu definieren. Da die Idee der Ebenen auch nützlich sein kann, wenn man Bildsequenzen bedenkt [13], müssen in Zukunft auch Interaktionen zwischen Ebenen, definiert durch Bewegung, und solche definiert durch Stereobilder, in diese Überlegungen mit einbezogen werden.

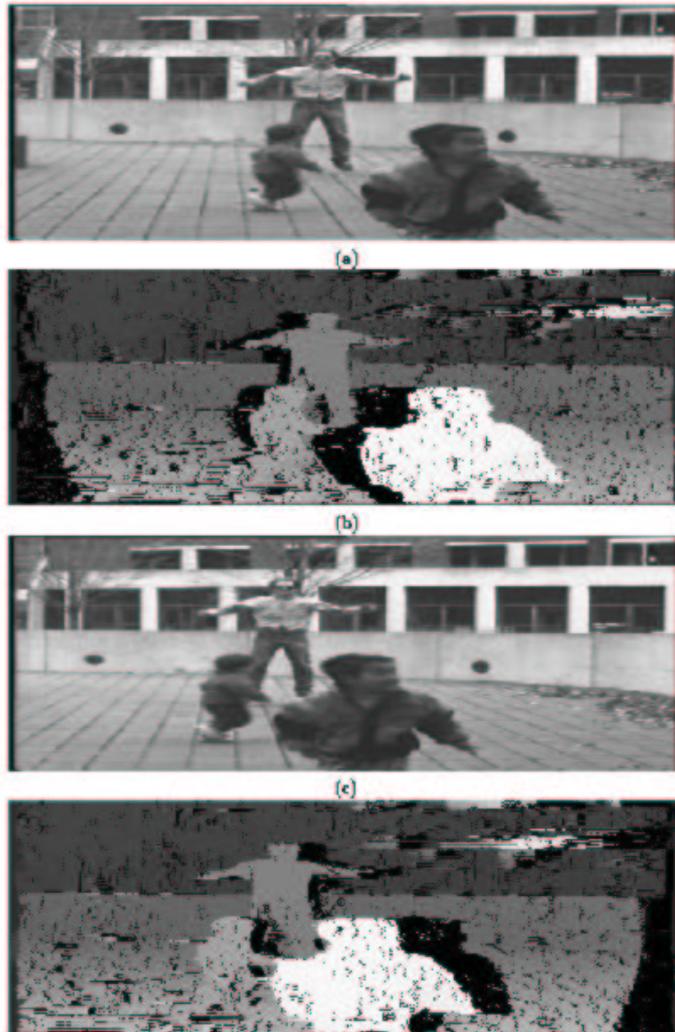


Abbildung 10: (a), (c) Stereobild eines Mannes und zweier Kinder. (b), (d) Das Verdeckungsergebnis unter Verwendung von GCP's und Kantenlinien.

## 6. Zusammenfassung

Es wurde eine Technik zum Auffinden von Stereomatches und verdeckten Bereichen, die auf einem pixelbasiertem Ansatz beruht, vorgestellt. Dieser Ansatz basiert darauf, den besten Pfad durch ein Disparity-Space-Image zu finden. In der modifizierten Formulierung wurde auf den Zusammenhang zwischen Verdeckungssprüngen und Intensitätskanten eingegangen. Kantenlinien wurden benutzt, um die Kosten vorgeschlagener Verdeckungskanten, die mit Intensitätskanten zusammenfielen, zu verringern. Das Ergebnis ist ein Algorithmus, der große verdeckte Bereiche akkurat extrahiert, ohne Kriterien wie Konsistenz über Scanlinien hinweg oder smoothness zu benutzen. Eine direkte Anwendung dieses Ansatzes ist die Entdeckung von Stereoebenen – dreidimensional isolierten Ebenen.

## 7. Referenzen

- [1] P. Belhumeur und D. Mumford. A bayesian treatment of stereo correspondece problem using half-occluded regions. In *Proc. Vis. and Pattern Rec.*, 1992
- [2] R. Bolles, H. Bakera und M. Hannah. The JISCT stereo evaluaion. In *Proc. Image Understanding Workshop*, Seite 263 – 274. 1993.
- [3] J. Canny. A computational approach to edge detection. *IEEE Trans. Patt.Analy. and Match. Intell.*, 8(6): Seite 679 – 698, 1986
- [4] S.D. Cochran and G. Medioni. 3-D surface description from binocular stereo. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 14(10): 981 – 994, 1992
- [5] I. J. Cox, S. Hingorani, B. Maggs, and S. Rao. Stereo without regularization. NEC Research Institute Report, NEC Research Institute, October 1992.
- [6] P. Fua. Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities. In *Proc. Int. Joint Conf. Art. Intel.*, Seite 1292 – 1298, 1991.
- [7] D. Geiger, B. Ladendorf und A. Yuille. Occlusion and binocular stereo. In *Proc. European Conf. Comp. Vis.*, Seite 425 – 433, 1992.
- [8] M. J. Hannah. A system for digital stereo image matching. *Photogrammetric Eng. And Remote Sensing*, 55(12): 1765 – 1770, 1989.
- [9] S. S. Intille und A. F. Bobick. Disparity-space images and large occlusion stereo. *MIT Media Lab Perceptual Computing Group Technical Report No. 220*, Massachusetts Institute of Technology, Oktober 1993.
- [10] T. Kanade und M. Okutomi. A stereo matching algorithm with an adaptive window: theory and experiment. In *Proc. Image Understanding Workshop*, Seite 383 . 389, 1990.
- [11] Y. Ohata und T. Kanade. Stereo by intra- and interscanlinie search using dynamic programming. *IEEE Trans. Patt. Analy. and Mach, Intell.*, 7:139 – 154, 1985.
- [12] D. Terzopoulos. Regularization of inverse problems involving discontinuities. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 8(4): 413 – 424, 1986.
- [13] J. Y. A. Wang und E. H. Adelson. Layered representation for motion analysis. In *Proc. Comp. Vis. And Pattern Rec.*, Seite 361 – 366, New York City, Juni 1993.
- [14] Y. Yang, Ayuille und J. Lu. Local, global and multilevel stereo matching. In *Proc. Comp. Vis. And Pattern Rec.*, 1993.
- [15] S. S. Intille und A. F. Bobick. Incoorporating Intensity Edges in Recovery of Occlusion Regions. *MIT Media Lab Perceptual Computing Group Technical Report No. 246*, Massachusetts Institute of Technology, 1994.