

# **Implementierung von Skelettierungs-Algorithmen mit dem Kinect-Sensor**

## **— Projektbericht —**

Arbeitsbereich Kognitive Systeme

Fachbereich Informatik

Fakultät für Mathematik, Informatik und Naturwissenschaften

Universität Hamburg

Vorgelegt von:

Johannes Böhler,  
Christopher Kroll (6367171),  
Sandra Schröder (6060939)

Hamburg, den 10.03.2013

# Inhaltsverzeichnis

<b>1 Einleitung</b>	<b>3</b>
1.1 Aufgabenstellung . . . . .	4
1.2 Aufbau der Projektarbeit . . . . .	4
<b>2 Die Kinect</b>	<b>5</b>
<b>3 Die Skelettierung</b>	<b>10</b>
3.1 Skelettierung mittels Thinning . . . . .	10
3.2 Skelettierung mittels Distanztransformation . . . . .	16
3.3 Weitere Verfahren . . . . .	21
<b>4 Implementierung der Algorithmen</b>	<b>23</b>
4.1 Technische Umsetzung . . . . .	23
4.2 Spieler-Segmentierung . . . . .	24
4.3 Skelettierung mittels Thinning . . . . .	24
4.4 Skelettierung mittels Distanztransformation . . . . .	25
<b>5 Evaluation</b>	<b>27</b>
5.1 Skelettqualität . . . . .	27
5.2 Laufzeiten . . . . .	29
5.3 Distanztransformation - Verbesserung der Skelettqualität . . . . .	30
5.4 Fazit . . . . .	36
<b>6 Zusammenfassung und Ausblick</b>	<b>38</b>
6.1 Zusammenfassung . . . . .	38
6.2 Ausblick . . . . .	38
<b>Literaturverzeichnis</b>	<b>39</b>

# 1 Einleitung

*Autor: Christopher Kroll*

Im Master-Studiengang Informatik an der Universität Hamburg ist ein zweisemestriges Projekt vorgesehen. Die Autoren dieser Arbeit belegten im Sommersemester 2012 und Wintersemester 2012/2013 das Projekt 'Bildverarbeitung' unter der Leitung von Prof. Dr. Leonie Dreschler-Fischer. Der Verlauf und das Ergebnis dieses Projektes wird in dem vorliegenden Dokument dargestellt.

Die Studenten sollten in diesem Projekt die Daten der Microsoft Kinect-Kamera nutzen und diese für ein Thema ihrer Wahl verarbeiten. Nach einer Vorstellung der Kinect und der Programmiersprache Python wurden die Gruppen eingeteilt und das Thema gewählt (siehe Kapitel 1.1 ).

Die digitale Bildverarbeitung hilft unter anderem bei der Erkennung von Objekten, wie zum Beispiel bei der vollautomatisierten Qualitätskontrolle. Um auch bei großen Datenmengen noch performant arbeiten zu können, bedient man sich oft des Mittels der Abstraktion. Das Projektthema dieser Arbeit dreht sich um eine Möglichkeit der Abstraktion, nämlich der Skelettierung.

Der Begriff 'Skelett' taucht in vielen Fachbereichen, wie zum Beispiel der Anatomie, Biologie oder Architektur auf. In der digitalen Bildverarbeitung hat er einen ähnlichen Sinn wie in diesen Bereichen: Es stellt von einem Objekt das Grundgerüst dar. Dieses Gerüst ist der zentrale Hauptbestandteil, mit dessen Hilfe sich Rückschlüsse auf das gesamte Objekt (siehe Dinosaurierskelette) ergeben.

Aus Sicht der Bildverarbeitung kann ein Skelett folgendermaßen definiert werden: Ein Skelett ist ein nützlicher Deskriptor, um Informationen über die Region und den Rand eines Objektes kompakt und effizient zu kodieren und gibt die wesentlichen Grundzüge eines Objektes wieder. In diesem Fachbereich kommen noch weitere Eigenschaften des Skeletts hinzu. Das Ziel ist die Information eines Objektes zu reduzieren ohne dabei die Grundstruktur zu verletzen. So ist eine Anforderung, dass die 'Pixelkonnektivität' gewährleistet ist, also alle späteren Skelettpixel mindestens einen benachbarten Skelettpixel besitzen und das Skelett nicht unterbrochen ist.

Bei der Skelettierung muss weiterhin beachtet werden, dass die topologische Struktur des Originalbildes nicht verändert wird. Trotz eventueller Verformung bei der Skelettbildung, wie in Abbildung 1.1 dargestellt, muss die strukturelle Eigenschaft erhalten bleiben; so ist bei einer automatisierten Zeichenerkennung die Strichstärke unerheblich, lediglich der generelle Aufbau des Buchstaben ist relevant.



Abbildung 1.1: Skelettierung des Buchstaben A TODO cite [?]

TODO geschichte und 3 Gruppen der skelettierung

## 1.1 Aufgabenstellung

Unsere Gruppe entschied sich für die Datenauswertung einer Person ('Spieler'). Dabei war zunächst das Ziel den Bewegungsablauf bei Sportübungen zu analysieren und eine Rückmeldung zu geben, ob diese richtig ausgeführt wurden. So soll zum Beispiel bei Kniebeugen durch eine Messung des Winkels zwischen Ober- und Unterschenkel der Person eine Hilfestellung gegeben werden.

Um dieses Ziel zu erreichen muss zunächst der Spieler von anderen Gegenständen im Raum getrennt, also herausgefiltert werden (Segmentierung). Für die Bewegungsanalyse ist es hilfreich nicht den gesamten Menschen, womöglich noch mit störender Kleidung, zu betrachten, sondern nur sein Skelett. Um das Skelett zu erhalten, bot sich die Wahl zwischen schon implementierten Skelettierungsalgorithmen zu benutzen oder dies selbst zu implementieren. Da das Thema des Projektes die Bildverarbeitung und nicht eine Anwendungsprogrammierung ist, fiel die Entscheidung auf die Konzentration auf die Skelettierung. Das Ziel war nun verschiedene Ansätze zu implementieren und hinsichtlich Qualität und Leistung zu vergleichen.

Dabei mussten einige Kriterien beachtet werden. Zuallererst muss auf die Qualität des erzeugten Skeletts geachtet werden. Es sollte als eine Skelettrepräsentation des Quellbildes erkannt werden, also ein zentrales Grundgerüst mit vorhandener Pixelkonnektivität. Des Weiteren ist die Leistungsfähigkeit relevant. Die Algorithmen sollen echtzeitfähig sein, damit der Spieler eine unmittelbare Rückmeldung seiner Bewegungen erhält.

## 1.2 Aufbau der Projektarbeit

TODO am ende wenn struktur steht

## 2 Die Kinect

*Autor: Johannes Böhler*

Die drei Haupt Hardware-Komponenten der Kinect sind ein Infrarotprojektor, eine RGB-Kamera sowie eine Infrarotkamera. Die Kombination aus Infrarotstrahler und In-



Abbildung 2.1: *Links:* Infrarotprojektor *Mitte::* RGB-Kamera *Rechts::* Infrarotkamera

frarotkamera ermöglicht die Gewinnung von Tiefeninformationen aus der Umgebung. Im Gegensatz zu gewöhnlichen Kameras welche einem Pixel Farbinformation (z.B. über RGB-Farbkanäle) zuordnen, wird dem Pixel mit Hilfe der Infrarot Kamera eine Entfernung zugeordnet.

Diese ergibt sich aus der Art und Weise wie der Infrarotstrahl von dem durch den Pixel repräsentierten Bereich eines Objektes reflektiert wurde.

Die Tiefenbilder welche man von der IR Kamera erhält, sehen aus wie komplett verrauschte Graustufenbilder. Hierbei steht jeder Grau Wert eines Pixels für die entsprechende Entfernung des korrespondierenden Objektausschnittes zur Kinect.

Hohe Grauwerte (helle Pixel) repräsentieren nahe Objekte, während niedrige Grauwerte (dunkle Pixel) weiter entfernte Objekte beschreiben. In einem Tiefenbild ist die gesamte Information über die Entfernung der im Bildausschnitt erfassten Objekte zur Kinect enthalten.

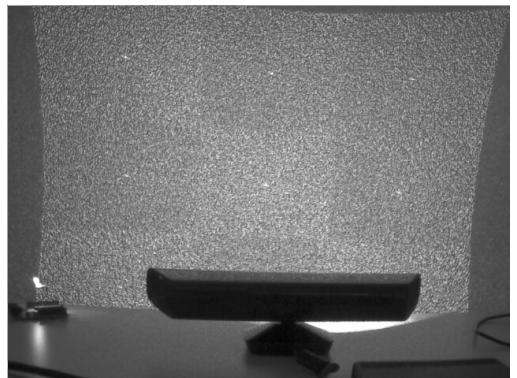


Abbildung 2.2: Vom Infrarotprojektor ausgestrahltes strukturiertes Licht

Wird die Tiefeninformation dazu genutzt um die Pixel im dreidimensionalen Raum anzugeordnen, erhält man eine Punktwolke. Die in der 2d -Betrachtung benachbarte Pixel müssen in der 3d Repräsentation nicht miteinander verbunden da die Z-Koordinate unterschiedliche Werte aufweisen kann.

## 2.0.1 Funktionale Komponenten

### Infrarotprojektor

Der Infrarotprojektor emittiert elektromagnetische Strahlen, deren Wellenlänge (830nm) außerhalb des für den Menschen sichtbaren Bereichs liegt (380nm-780nm). Der Projektor strahlt zur Tiefenbestimmung ein Gitter von Infrarotpunkten (strukturiertes Licht) auf die Objekte in seiner Umgebung ab. Zur Generierung dieses Musters wird ein besonderes Verfahren implementiert. Normalerweise produzieren Filter die diese Art von Muster generieren einen sehr hellen Punkt in der Mitte des Bildes, welcher die Leistungsfähigkeit des IR Projektors limitiert. Durch das hier verwendete Verfahren entstehen statt einem einzigen sehr hellen, neun helle Punkte, welche durch unvollständige Lichtfilterung zur Mustererstellung bedingt sind. Das hier eingesetzte Verfahren produziert weniger starke Artefakte und ermöglicht die Verwendung einer leistungsfähigeren Diode was eine erhöhte Genauigkeit sowie eine größere Reichweite ermöglicht. Die Reichweite des IR Projektors ist dennoch eingeschränkt, da zu hohe Intensitäten der IR Strahlen Augenschäden verursachen könnten.

Es ist wichtig dass die Wellenlänge der Infrarot strahlen konstant bleibt, was durch eine gleichbleibende Temperatur der Laserdiode und eine konstante Stromleistung (60 mW) gewährleistet wird. Problematisch sind variierende Außentemperaturen (die empfohlene Betriebstemperatur liegt zwischen 5 und 35 °C). Um diese Konstanz zu gewährleisten wird ein sogenanntes Peltier-Element verwendet, welches sowohl kühlen als auch wärmen kann.

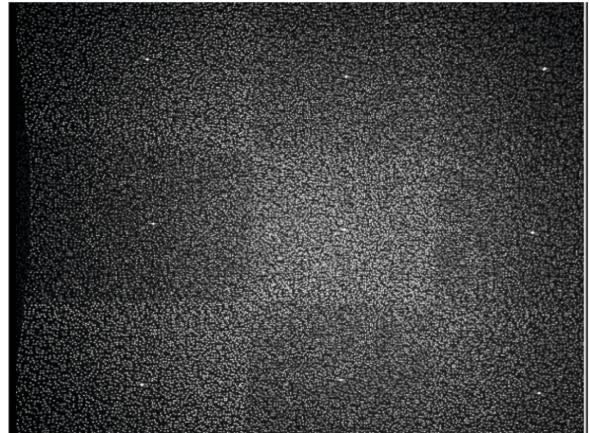


Abbildung 2.3: Generierung des Infrarot Random-Patterns mit Hilfe von 9 Bereichen

### Infrarotkamera

Die IR Kamera nimmt Bilder mit einer nativen Auflösung von 1280x1024 Pixeln, bei einer Bildwiederholrate von 30 Hz auf. Weitergeleitet werden allerdings nur Bilder mit einer Auflösung von 640x480 Pixeln da der USB Datenbus eine Limitierung bezüglich der Datenübertragungsrate darstellt. Das Blickfeld der IR Kamera beträgt in der Horizontalen 58 °, in der Vertikalen 45 °.

Damit sich die Mustererkennung funktioniert ist eine Mindestabstand von 0,8 Metern erforderlich, ab einer Entfernung von 3,5 Metern wird die Intensität der reflektierten Strahlen zu gering um Tiefeninformationen mit ausreichender Präzision zu erhalten. Bei einem optimalen Abstand von 2 Metern zum Objekt beträgt die Auflösung in der XY Ebene 3 mm, in der Z Ebene 1 cm.

Die Quantisierungsauflösung liegt bei (2048) Bit. Es muss sichergestellt werden dass die IR Kamera nur die erwünschte elektromagnetische Strahlung im 830 nm Bereich aufnimmt und nicht von Strahlungen anderer Wellenlänge gestört wird . Dies wird durch einen Filter, welcher auf dem IR Kameraobjektiv angebracht ist realisiert. Trotz des Filters sollte die Kinect in abgedunkelten Innenräumen verwendet werden, da Sonnenlicht auch elektromagnetische Wellen im Infrarotbereich enthält.

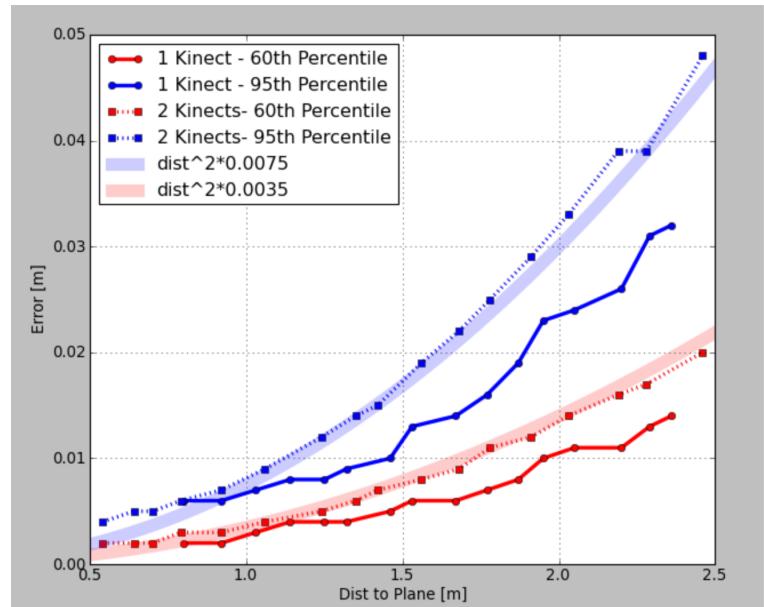


Abbildung 2.4: Generierung des Infrarot Random-Patterns mit Hilfe von 9 Bereichen

## RGB Kamera

Die RGB Kamera nimmt bei einer Wiederholrate von 30 Hz ebenfalls mit einer Auflösung von 640x480 Pixeln auf, könnte jedoch auch mit einer Auflösung von 1280x1024 Pixeln bei einer reduzierten Wiederholrate von 15 Bildern pro Sekunde angesteuert werden. Die Quantisierungsauflösung liegt hier bei (256)Bit.

## Mikrofon Array

Die Kinect beinhaltet vier Mikrofone, welche verteilt verbaut sind. Sie dienen sowohl der Erfassung von Ton, als auch der Lokalisierung und Unterscheidung von Soundquellen. Jedes der Mikrophone tastet mit einer Quantisierungsauflösung von (65536) Bit und einer Abtastrate von 16 KHz ab.

### 2.0.2 Tiefenberechnung

Die Berechnung der Objektentfernungen (Tiefenwerte) innerhalb der Kinect erfolgt durch Aussendung von strukturiertem Licht durch die Projektionseinheit und durch Weiterverwertung der durch die Infrarotkamera empfangen reflektierten Strahlen auf einem Chip

der Firma PrimeSense (PS1080-A2-Chip). Das Muster für das strukturierte Licht wird mit Hilfe eines Diffusors(einer Lochplatte mit fest definiertem Muster) erzeugt.

Das Prinzip des Strukturierten Lichts ist an ein Verfahren angelehnt, welches sich Streifenprojektion nennt und in dieser Anwendung abgewandelt wurde um die Erfassung von beweglichen Objekten zu ermöglichen. Statt Lichtstreifen wird eine Punktematrix verwendet, welche zufällig und fest definiert ist. Die IR Kamera sowie der Projektor müssen sich hierzu in einem definierten, gleich großen Abstand zueinander befinden. Die Umgebungsreflektionen der projizierten Punktematrix werden von der Infrarotkamera erfasst. Um aus diesem Gitter von Infrarotpunkten Tiefeninformation zu extrahieren, wird das Verfahren der aktiven Stereotriangulation verwendet.

Die Triangulation ist ein Verfahren zur optischen Abstandsmessung, welches sich hierzu trigonometrischer Funktionen innerhalb von aufgespannten Dreiecken bedient. Es wird allgemein zwischen aktiver und passiver Triangulation unterschieden. Aktive Triangulation bedeutet, dass mindestens eine strukturierte Lichtquelle zur Abstandsberechnung erforderlich ist, während dies bei passiver Triangulation nicht der Fall ist. Da der IR Projektor ein statisches Pseudozufallsmuster emittiert, ist dieser als strukturierte Lichtquelle einzuordnen. Stereotriangulation bedeutet dass zwei unterschiedliche Bildquellen benötigt werden um die Tiefe jedes Pixels eines Bildausschnittes berechnen zu können. Eine Bildquelle ist der Diffusor (das „Lochmuster“) welcher die vom Projektor emittierten Strahlen statisch definiert. Die andere Bildquelle ist die IR Kamera. Das erste Bild ist immer identisch und statisch, das zweite hingegen variiert je nach Umgebung. Diese beiden Bilder sind die Grundlage für die trigonometrischen Operationen zur Berechnung der Tiefeninformationen.

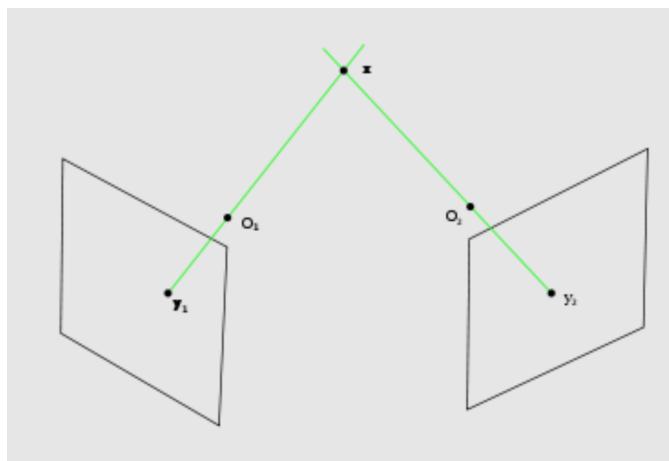


Abbildung 2.5: aktive Stereotriangulation

Zur Gewinnung der Tiefeninformation wird die horizontale Differenz des Punktes Y1 des von der IR Kamera aufgenommenen Bildes zum korrespondierenden Punkt Y2 des virtuellen statischen Referenzbildes des Projektors berechnet. Aus dieser Differenz lässt sich die Tiefe des betreffenden Pixels durch aufstellen der beiden Projektionslinien und Schnittpunktbestimmung derselben berechnen.

Der Grund warum die Pixel der „Schablone“ im Projektor zufällig angeordnet sind, liegt

darin, dass die unterschiedlichen lokalen Nachbarschaftsbedingungen die Pixelzuordnung zwischen dem statischen und dem dynamisch veränderten Bild erleichtern.

### 2.0.3 Das Schattenproblem

Aufgrund der Entfernung der verbauten RGB-Kamera zur Infrarotkamera, weisen die Bilder beider Kameras einen kleinen Versatz auf. Schatten im Tiefenbild entstehen aufgrund der Entfernung des Infrarotprojektors zur Infrarotkamera.

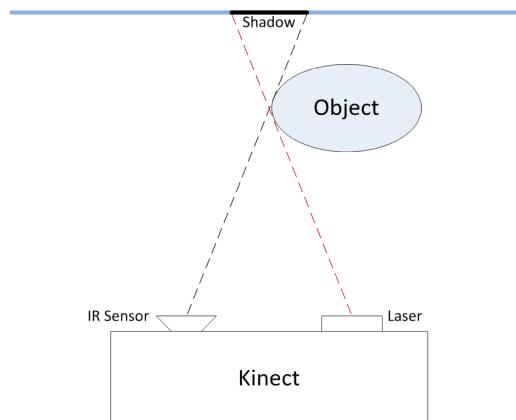


Abbildung 2.6: Blockierte IR Strahlen

Der Schatten im Muster macht es für den Sensor unmöglich die Tiefe festzustellen. Die Pixel in diesen Bereichen werden auf den Wert 0 gesetzt. Das Objekt blockiert die Strahlen des Lasers. Da für die Tiefenberechnung das vom IR-Projektor emittierte Musters benötigt wird, ist es für die Kinect unmöglich die Distanz in Bereichen zu berechnen, welche außerhalb der Erreichbarkeit des IR Strahlenmusters liegen.

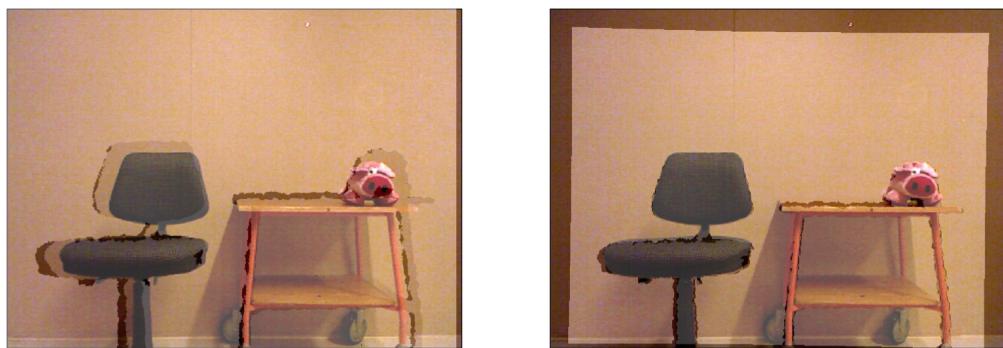


Abbildung 2.7: Da der Infrarot Sensor links des Projektors positioniert ist, treten die Schatten auch linksseitig der Objekte auf.

# 3 Die Skelettierung

*Autor: Sandra Schröder*

Unter einem Skelett versteht man einen Deskriptor, der die topologischen Eigenschaften eines Objekts beschreibt. Das sind Merkmale, die sich nicht explizit auf die konkrete Form einer Region beziehen, sondern auf ihre strukturellen Eigenschaften, die auch bei starken Verformungen erhalten bleiben. Zur Erfassung von bestimmten Eigenschaften von Binärbildern vereinfachen Skelette die Analysen, da durch die Reduzierung der Daten unwesentliche Eigenschaften ausgeblendet werden. Diese Abstraktion von für die Anwendung unwichtigen Eigenschaften ist eine zentrale Eigenschaft des Skelettes beziehungsweise eines Deskriptors im Allgemeinen. Historisch sind im Laufe der Entwicklung diverser Skelettierungsalgorithmen immer mehr konkrete Anforderungen an Skelette entstanden.

Eine wichtige Anforderung ist die *Konnektivität* des Skeletts. Dies bedeutet, dass es keine Lücken und Unterbrechungen im Skelett gibt, denn ein zusammenhängendes Objekt sollte ein zusammenhängendes Skelett besitzen.

Viele Skelettierungsverfahren fordern, dass ein Skelett genau ein Pixel breit ist. Dies ist beispielsweise in der Datenkomprimierung wichtig. Möchte man die Struktur eines Objektes mit möglichst wenig Pixeln speichern, genügt ein Skelett mit der minimal möglichen Pixelbreite. Eine breitere Skelettlinie könnte unter Umständen zu viele unnötige Informationen beinhalten. Des Weiteren sollte das Skelett zentriert im Objekt liegen. Das bedeutet, dass Abstand der Pixel der Skelettlinien nach links und nach rechts möglichst gleich sein sollte.

In wie weit alle Anforderungen an das Skelett erfüllt sein müssen, ist allerdings für jeden Anwendungsfäll neu zu diskutieren. Dieses Kapitel beschreibt bekannte Konzepte und Verfahren zur Skelettierung im Bereich der Bildverarbeitung. Zwei Verfahren wurden im Rahmen der Projektarbeit genauer untersucht: Die Skelettierung mittels *Thinning* und mittels *Distanztransformation*.

Um die Übersicht über die weiteren grundlegenden Verfahren und Konzepte zu vervollständigen, werden diese in einem separaten Abschnitt kurz vorgestellt.

## 3.1 Skelettierung mittels Thinning

*Autor: Johannes Böhler, Christopher Kroll*

Das Thinning bezeichnet eine Kategorie von Methoden zur Skelettierung von 2D sowie 3D Objekten. In dieser Projektarbeit ist der Fokus ausschließlich auf die 2D Skelettierung

gerichtet, da die Kinect kein vollkommenes 3D Modell eines Objektes liefert. Sie erfasst das Objekt lediglich aus einem Blickwinkel, desshalb erhält man nur ein 2,5 dimensionales Modell. Es werden nur Tiefeninformationen bezüglich der Seite des Objektes, welche der Kinect zugewendet ist bereitgestellt. Die Tiefeninformationen der Rückseite bleiben verborgen. Um ein vollständiges 3D Modell zu erhalten müssten mindestens 2 Kinects verwendet werden und die Informationen beider Geräte zusammengeführt und vereinheitlicht werden.

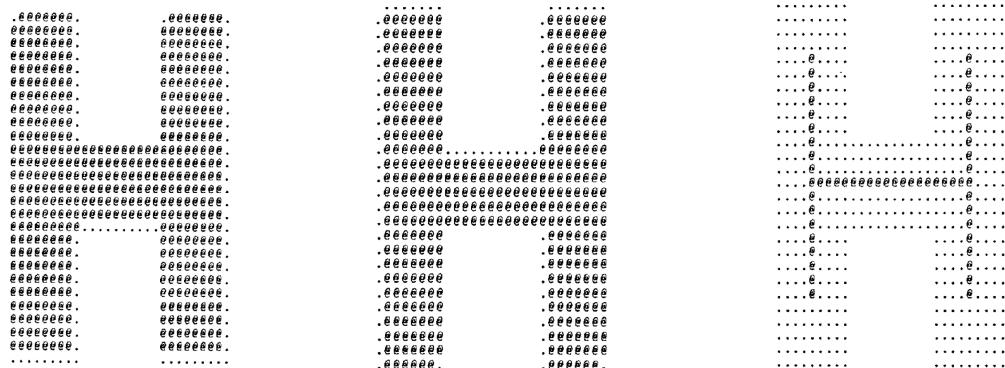
Alle Thinning-Algorithmen verbindet das iterative Abtragen des Musters oder der Oberfläche. Dabei werden nach und nach die Konturpixel, also die Objektpixel, die mit den Hintergrundpixeln benachbart sind, überprüft. Je nach Algorithmus liegen verschiedene Kriterien vor, ob diese gelöscht, also als Hintergrundpixel markiert werden können. Ein Kriterium ist auf jeden Fall, dass die Zusammengehörigkeit der Skelettpixel bewahrt wird. Eine weitere Gemeinsamkeit der unterschiedlichen Ansätze ist, dass das Ausdünnen solange stattfindet, bis am Objekt keine Änderung mehr gemacht wurde.

Beim Thinning wird die Zusammengehörigkeit nicht verändert, wodurch die topologische Struktur des Objektes erhalten bleibt.

### 3.1.1 A Fast Parallel Algorithm for Thinning Digital Patterns

*Autor: Johannes Böhler*

Der gesamte Algorithmus erstreckt sich über mehrere Iterationen. Die Randpixel des Musters werden Schicht für Schicht abgetragen. Die Iterationen selbst, sind wiederum in zwei Subiterationen unterteilt. Das Abtragen der „Schichten“ wird somit in zwei unterschiedliche Phasen aufgespalten. Mit Hilfe der ersten Subiteration werden sowohl Süd- und Ostgrenzpunkte als auch Nordwest Eckpunkte entfernt. Das entfernen von Nord- und Westgrenzpunkten sowie von Südosteckpunkten erfolgt in der zweiten Subiteration.



(a) Nach erster Subiteration (b) Nach zweiter Subiterati- (c) Skelett des Ursprungs-  
(zu Beginn) on (zu Beginn) muster

Abbildung 3.1: Zustände des Algorithmus

## Anforderungen an den Algorithmus

- Das Rauschen, welches der Algorithmus verursacht soll so gering wie möglich gehalten werden.
- Das Skelett des Ursprungsmusters soll die Endpunkt- und Pixelverbundenheit erhalten. Endpunktverbundenheit bedeutet, dass sich zwischen zwei Endpunkten eines Skeletts keine unverbundenen Stellen befinden.
- Das Skelett soll nach Durchlaufen des kompletten Algorithmus in einheitlicher Dicke von einem Pixel vorliegen.
- Der Algorithmus soll möglichst schnell und effizient arbeiten um Echtzeitfähigkeit gewährleisten zu können.

## Ablauf des Algorithmus

Es wird davon ausgegangen, dass zu Beginn ein binär digitalisiertes Bild vorliegt. Die Pixel werden mit Hilfe einer zweidimensionalen Matrix IT durchlaufen, deren Wert an der jeweiligen Stelle  $IT(i,j)$  entweder 0 oder 1 ist. Mit Muster ist die Menge an Pixeln gemeint, welche den Wert eins haben. Es werden in Abhängigkeit von den 8 Nachbarpixeln (siehe Abbildung 4.5), Transformationen auf den betrachteten Pixel P1 angewendet. Dieser Vorgang wird iterativ auf die Matrix IT angewendet.

Der neue Wert eines Pixels während der n-ten Iteration hängt von dem eigenen Wert während der (n-1)ten Iteration und den Werten der acht Nachbarn während der (n-1)ten Iteration ab. Dies ermöglicht paralleles Transformieren mehrerer Bildpunkte. Die Bedingungen, welche zum Ausführen der Transformation erfüllt sein müssen werden über ein 3x3 Pixel Fenster abgefragt. Der Punkt P1 über dessen Transformation entschieden wird, ist mit allen acht Nachbarn direkt verbunden.

$P_9$ $(i - 1, j - 1)$	$P_2$ $(i - 1, j)$	$P_3$ $(i - 1, j + 1)$
$P_8$ $(i, j - 1)$	$P_1$ $(i, j)$	$P_4$ $(i, j + 1)$
$P_7$ $(i + 1, j - 1)$	$P_6$ $(i + 1, j)$	$P_5$ $(i + 1, j + 1)$

Abbildung 3.2: Betrachteter Pixel P1 und Nachbarumgebung

Der Algorithmus entfernt alle Randpunkte des Musters, außer den Pixeln welche Bestandteil des Skeletts sind. Um die Verbundenheit des Skeletts zu gewährleisten wird ein Iterationsschritt in zwei Subiterationen aufgeteilt.

In der ersten Subiteration wird der Punkt P1 aus dem Muster gelöscht, wenn er folgende Bedingungen erfüllt:

- a)  $2 \leq B(P1) \leq 6$  B entspricht der Anzahl der Nachbarn von P1  $\neq 0$ . Die Anzahl der Nachbarn von P1 welche den Wert 1 haben, muss somit zwischen 2 und 6 liegen.
- b)  $A(P1)=1$  Anzahl der „01“-Folgen. Die Anzahl der 01 Folgen in der geordneten Folge P2,P3...P9 muss genau eins betragen.
- c)  $P2 * P4 * P6 = 0$  Mindestens ein Pixel der Pixelmenge P2, P4, P6 muss den Wert Null haben.
- d)  $P4 * P6 * P8 = 0$  Mindestens ein Pixel der Pixelmenge P4, P6, P8 muss den Wert Null haben.

Sind alle Bedingungen a, b ,c und d erfüllt so wird der Wert des Pixels auf 0 gesetzt. Dies bedeutet dass er kein Teil des Skelett-Musters mehr ist. Wird eine der Bedingungen nicht erfüllt, so bleibt der Pixelwert bei 1.

0	0	1
1	$P_1$	0
1	0	0

Abbildung 3.3: Bedingung B: Anzahl 01 folgen in zyklischer Reihenfolge

In der zweiten Subiteration wird P1 gelöscht falls folgende Bedingungen gelten:

- a)  $2 \leq B(P1) \leq 6$
- b)  $A(P1)=1$
- c)  $P2 * P4 * P8=0$
- d)  $P2 * P6 * P8=0$

Nur die Bedingungen c und d haben sich geändert.

Um die Bedingungen der ersten Subiteration zu erfüllen, muss  $P4=0$  oder  $P6=0$  oder ( $P2=0$  und  $P8=0$ ) erfüllt sein. Dies impliziert dass P1 entweder Süd- oder Ost-Grenzpunkt, oder Nordwesteckpunkt ist. Um die Bedingungen der zweiten Subiteration zu erfüllen muss  $P2=0$  oder  $P8=0$  oder ( $P4=0$  und  $P6=0$ ) sein. P1 ist Nord- oder West-Grenzpunkt oder Südosteckpunkt.

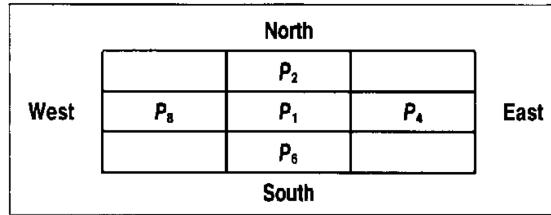


Abbildung 3.4: Betrachteten Nachbarpunkte in den Bedingungen c und d

Während mit Bedingung A ( $2 \leq B(P1) \leq 6$ ) die Endpunkte des Skeletts erhalten werden, so wird mit Bedingung B ( $A(P1)=1$ ) die Auslöschung von Punkten zwischen den Endpunkten der Skelettlinie verhindert.

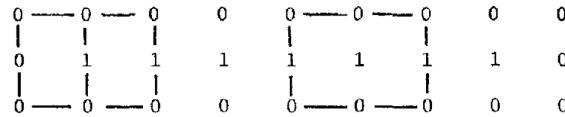


Abbildung 3.5: Gewährleistung der Pixelverbundenheit

In der Matrix Search M befinden sich während der ersten Iteration alle Pixel die gelöscht werden dürfen, da sie den Bedingungen der ersten Iteration genügen. Ist dies nicht der Fall, so ist der Counter=0 und der Algorithmus beendet, da es keine zu löschen Pixel mehr gibt. Die Skelettierung ist somit beendet.

Falls der Counter ungleich null ist, werden die Pixel welche den Bedingungen genügen von der Matrix IT (Skelett-Muster) abgezogen, der Counter wird null gesetzt und es wird zur zweiten Iteration fortgeschritten. Dort findet der Ablauf mit veränderten Bedingungen c und d wiederholt statt. Ist der Counter auch nach dem Durchlaufen der zweiten Subiteration ungleich null so wird der Vorgang iterativ fortgeführt.

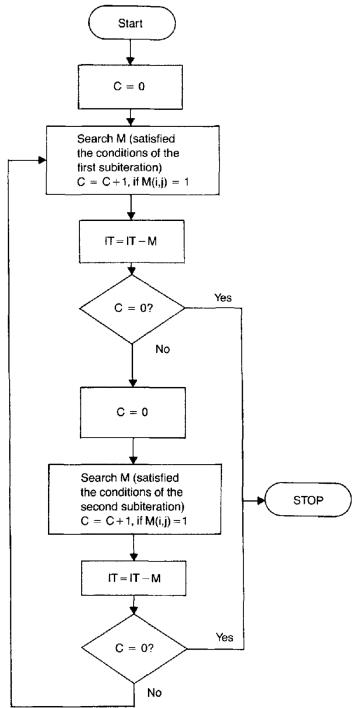


Abbildung 3.6: Gesamter Algorithmus in der Übersicht

## Resultate

Der Algorithmus erzielt sehr gute Ergebnisse im Bezug auf Verbundenheit und Rauschverhalten der Randpunkte. Die Bedingungen welche zum Auffinden der zu löschenenden Randpunkte führen sind sehr simpel. Durch den Bezug auf die n-1te Iteration zur Abfrage der Bedingungen kann der Algorithmus sehr schnell ausgeführt werden, da keine Warteabhängigkeiten bestehen.

## 3.2 Skelettierung mittels Distanztransformation

Autor: Sandra Schröder

Die Distanztransformation eines Binärbildes enthält Informationen über den Abstand der Objektpixel zum Hintergrund. Der Abstand zum Hintergrund wird für jeden Objektpixel bestimmt und in einem weiteren Bild (gleiche Dimension und Größe wie das Binärbild) als Grauwert an der Stelle des Objektpixels gespeichert. Das Ergebnis ist die sogenannte *Distance Map* (Abbildung 3.7).

Zur Bestimmung des Abstands benötigt man Metriken. Eine gängige Metrik - die für den Algorithmus im Rahmen dieser Arbeit auch genutzt wurde - ist die euklidische Metrik  $d_2$ :

$$d_2(p, q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} \quad (3.1)$$

Die Definition einer Nachbarschaft spielt ebenfalls eine Rolle. Wählt man eine 4er-Nachbarschaft, wird der Abstand vom Objektpixel zur linken und zur rechten Seite, sowie nach oben und nach unten berechnet. Man erhält dementsprechend vier Abstände. In der Distance Map wird der minimale Abstand von den vier Werten gespeichert. Bei einer 8er-Nachbarschaft kommen die diagonalen Richtungen dazu. Abhängig von der gewählten Nachbarschaft und der Metrik ergeben sich unterschiedliche Distance Maps.

Die Pixel mit dem größten Abstand zum ersten Pixel des Hintergrunds sind die hellsten Pixel in der Distance Map. Die umgebenden Pixel haben einen kleineren Grauwert und der Grauwert der Pixel verringert sich, je näher die Pixel am Hintergrund liegen. Dementsprechend beschreiben die hellsten Pixel des *Grauwertgebirges* eine skelettförmige Struktur, die im Folgenden aus extrahiert werden soll. Wir verfolgen einen Ansatz, der auf der Bildung des Gradientenbetrages der Distance Map beruht.

Ein Graubild kann als skalare Funktion  $f(x, y)$  beschrieben werden mit  $x, y \in \mathbb{N}$  und



Abbildung 3.7: Beispiel einer Distance Map. Links: Originalbild. Dieses wurde zuerst invertiert, damit das Objekt (Person) weiß markiert ist. Rechts: Resultat der Distanztransformation. Auffällig ist das Maximum in der Mitte.

$0 \leq f(x, y) \leq 255$ . Die Idee ist, den Gradienten beziehungsweise den Gradientenbetrag der Distance Map zu bestimmen. Der Gradient (Gleichung 3.2) ist ein Differentialoperator

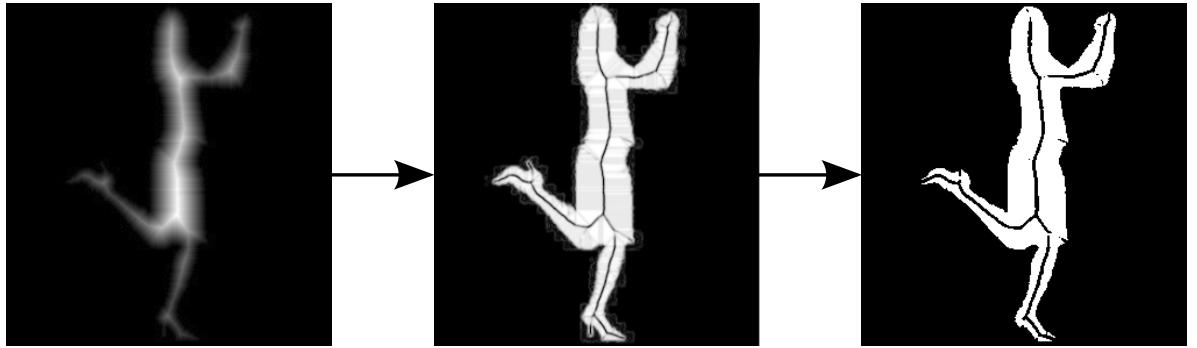


Abbildung 3.8: Segmentierung des Gradientenbetrags aus der Distance Map.

und liefert, angewandt auf ein Skalarfeld, die Richtung des stärksten Anstiegs, sowie die Amplitude des Anstiegs (Gradientenbetrag):

$$grad(f(x, y)) = \nabla f(x, y) = \begin{bmatrix} \frac{df}{dx} \\ \frac{df}{dy} \end{bmatrix} \quad (3.2)$$

Entsprechend der Definition des Gradienten, ist an den lokalen Maxima des Grauwertgebirges der Gradientenbetrag gleich null.

Speichert man den Gradientenbetrag ebenfalls als Grauwertbild mit gleicher Dimension und Größe wie die Distance Map, wird der Gebirgskamm als Linie mit kleinen Grauwerten (fast Null) kodiert. Diese markiert das Skelett. Eine schwellwertbasierte Segmentierung des Gradientenbetragbildes bewirkt, dass die Skelettlinien schwarz sind (Abbildung 3.8). Andere Bereiche, die größer als der Schwellwert sind und somit nicht zum Skelett gehören erhalten den Grauwert 255 (weiß).

Um endgültig nur die Skelettlinie zu erhalten, wird die Differenz zwischen der Distance Map und dem segmentierten Gradientenbetragbild gebildet. Der Teil der Distance Map der nicht auf der höchsten Stelle des Grauwertgebirges liegt hat entweder einen Grauwert kleiner oder gleich 255. Bei der Differenzbildung kann der Grauwert dieser Pixel nur kleiner gleich 0 werden, da von diesen Grauwerten der Grauwert 255 abgezogen wird (weißer Bereich des segmentierten Gradientenbetragbildes). Negative Grauwerte werden auf 0 gesetzt. Als Ergebnis erhält man die Skelettlinie mit Grauwerten ungleich 0. Um ein binärkodiertes Skelett zu erhalten, wird auf dem Differenzbild anschließend eine schwellwertbasierte Segmentierung ausgeführt (Abbildung 3.9). Im Folgenden wird das aus der Distanztransformation bestimmte Skelett zur Abgrenzung zum Thinning-Algorithmus als *Distanzskelett* bezeichnet.

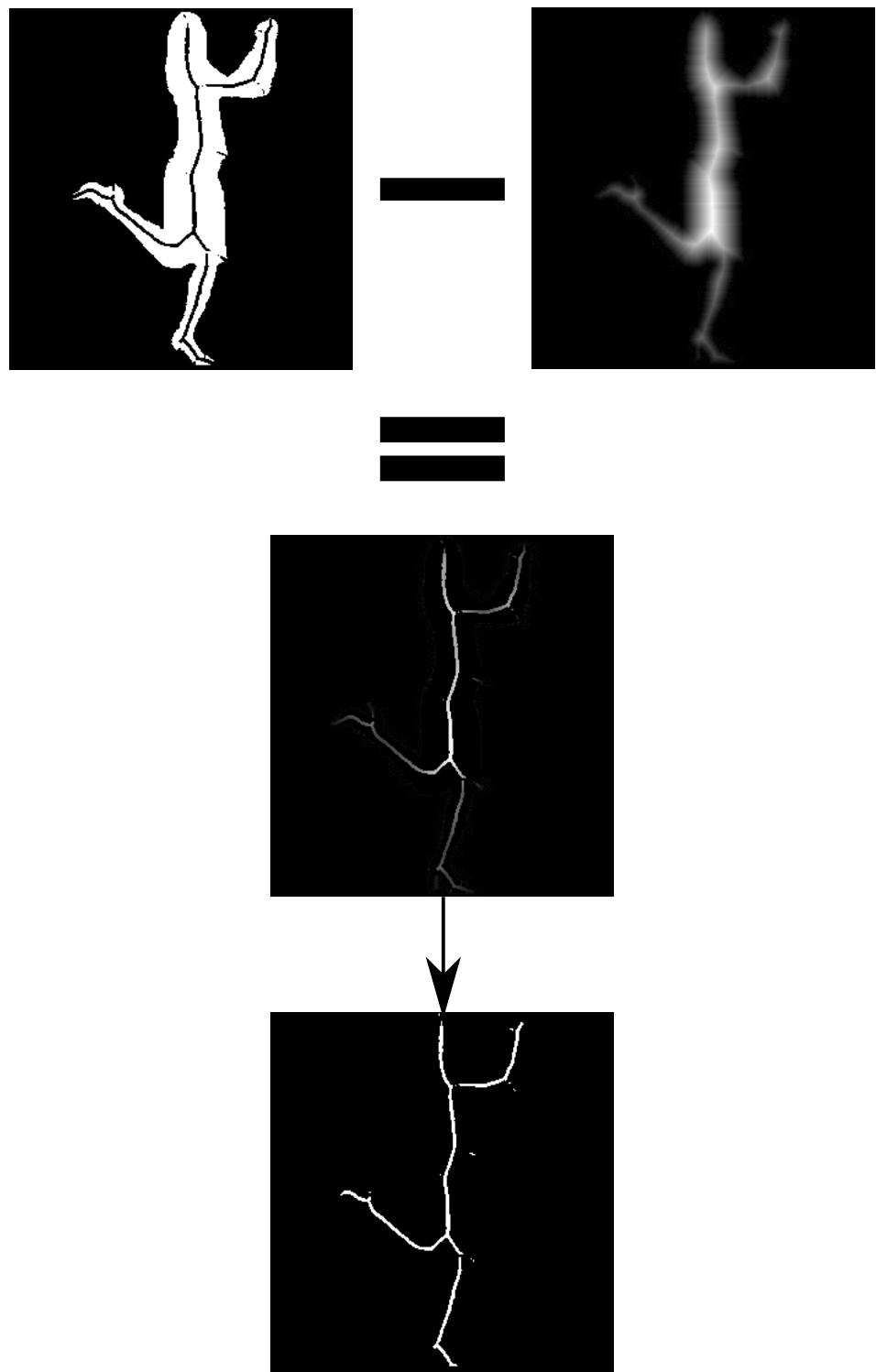


Abbildung 3.9: Differenzbildung zwischen dem segmentieren Gradientenbetragsbild und der Distance Map.

### 3.2.1 Verwandte Arbeit [Cha07]: Extracting Skeletons From Distance Maps

*Autor: Sandra Schröder*

Zur oben beschriebenen Berechnung des Skelettes aus einer Distance Map gibt es verwandte Arbeiten, von denen nun eine vorgestellt wird. In dem Paper *Extracting Skeletons from Distance Maps* beschreibt der Autor einen Algorithmus, der effizient und schnell aus einem distanztransformierten Bild ein Skelett extrahiert [Cha07]. Der Autor legt besonders hohen Wert darauf, dass die Extraktion keine komplizierten Berechnungen beinhaltet. Er möchte vor allem auf die Berechnung von Ableitungen höherer Ordnung und die Auswertung von komplexen Gleichungen verzichten. Das Verfahren, welches der Autor zur Extraktion der Skelettlinien eines Objekt vorstellt, ist die sogenannte *Ridge Point Detection* (deutsch: *Gebirgskammdetektion*). Dabei nutzt der Autor eine grundlegende Eigenschaft der Distance Map. Wie in Abschnitt 3.2 beschrieben, ist die Distance Map ein Grauwertgebirge, wobei der Gebirgskamm zentriert im Objekt liegt. Betrachtet man nur diesen Teil der Distance Map und projiziert ihn auf das Originalbild, lässt sich eine skelett-artige Beschreibung des Objekts erkennen.

Die Gebirgskammdetektion ist ein gradientenbasiertes Verfahren. Liegt ein Punkt auf den Gebirgskamm, so hat er in seiner unmittelbaren Umgebung den größten Abstand zum Objektrand und den größten Grauwert relativ zu seinen Nachbarn. Dieser Punkt ist somit ein lokales Maximum. Aufgrund dieser Tatsache eignet sich der Gradient am besten, um solch einen Punkt zu detektieren. Der Gradient zeigt nach Definition in die Richtung des stärksten Anstiegs. Dies bedeutet, dass der Gradient eines Punktes, der nicht auf dem Kamm liegt, in die Richtung des Kammes zeigt. Wählt man nun Punkte näher am Kamm, so wird der Anstieg geringer, da die Differenz zwischen dem Grauwert des betrachteten Gebirgspunktes und dem aktuell gewählten Punkt kleiner wird. Überquert man den Kamm, kehrt sich die Richtung des Gradienten um und zeigt wieder zum Kamm. Hier findet ein Vorzeichenwechsel des Gradientenbetrags statt. Der Punkt auf dem Gebirgskamm bildet eine *sign barrier* zwischen den Gradientenbeträgen der sich gegenüberliegenden Punkte, wobei sich der Gebirgspunkt zwischen diesen beiden Punkten befindet. Diese Beobachtung ist in Abbildung 3.10 dargestellt. Es ist ein Querschnitt eines Grauwertgebirges abgebildet. Legt man nun eine Projektionslinie genau durch das lokale Maximum (roter Punkt), und projiziert die Gradienten (violette Pfeile) der sich gegenüberliegenden Punkte (blau) auf diese Linie, so kann man erkennen, dass die Projektionen der Gradienten (orangene Pfeile) genau in die entgegengesetzte Richtung zeigen.

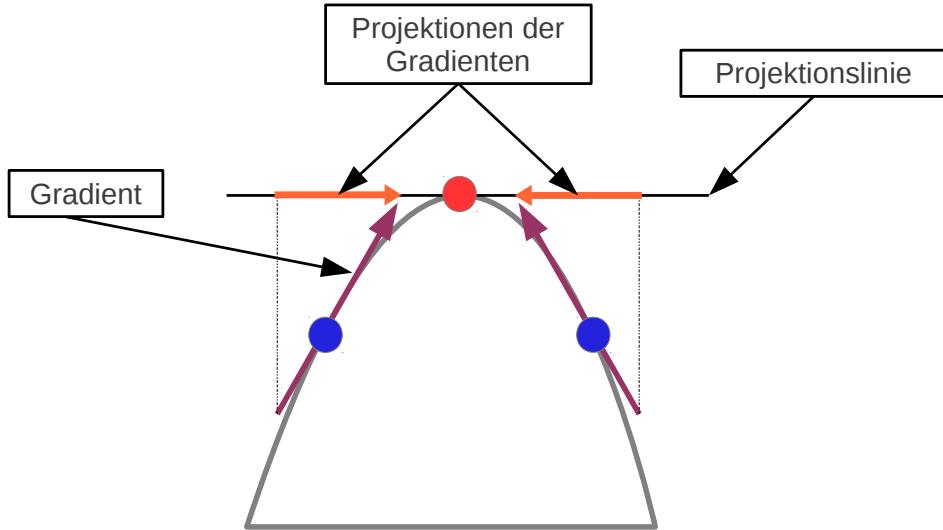


Abbildung 3.10: Ridge Point Detection. Der rote Punkt ist ein lokales Maximum auf dem Gebirgskamm. Die blauen Punkte liegen sich gegenüber und umschließen den roten Punkt. Ihre auf die Projektionlinie projizierten Gradientenrichtungen zeigen in entgegengesetzte Richtungen. Der rote Punkt bildet somit eine *sign barrier* für die beiden Richtungen.

Die Idee des Algorithmus ist, Projektionslinien durch die Distance Map zu legen und das Verhalten der Gradientenbeträge auf diesen Linien zu beobachten. Man stellt fest, dass sich dabei mehrere Muster von Vorzeichenwechsel der Gradientenbeträge erkennen lassen. Diese Muster können dabei ein Indiz für einen Gebirgspunkt und somit für einen Punkt des Skeletts sein.

Dabei stellt sich die Frage, wieviele Richtungen mit diesen Linien untersucht werden sollen. Man kann beobachten, dass es einen Vorzeichenwechsel in den Gradientenbeträgen gibt, wenn die Projektionslinie den Gebirgskamm schneidet. Dementsprechend gibt es keinen Vorzeichenwechsel, wenn die Projektionslinie parallel zum Kamm verläuft. Findet man also in einer Richtung keinen Gebirgskamm, so muss einer in der orthogonalen Richtung liegen. Deshalb genügt es, zwei zueinander senkrechte Richtungen zu untersuchen. Dabei wählt man die eine Richtung parallel zur x-Achse und die andere Richtung parallel zur y-Achse des untersuchten Bildes. Es existieren insgesamt vier Muster, die auf einen Gebirgskamm deuten. Zwei davon sind in Abbildung 3.11 zu sehen. Die Muster beschreiben, in welcher Weise Vorzeichenwechsel zwischen zwei benachbarten Pixeln auftreten können. Die Symbole + und – die Richtungen der Gradienten. + ist eine positive Richtung (bergauf), – eine negative Richtung (bergab) auf einer Projektionslinie. Das Symbol  $\circ$  besagt, dass sich in diesem Bereich der Gradient nicht ändert, da der Grauwert im

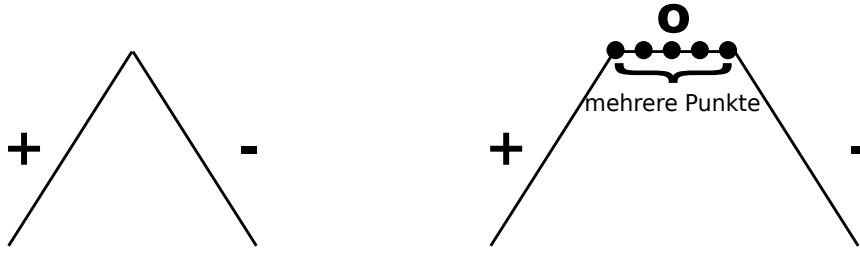


Abbildung 3.11: Muster für Hinweise auf einen Gebirgskamm. Diese beiden Muster sind ein starkes (*strong*, linke Abbildung) und ein gutes (*good*, rechte Abbildung) Indiz für einen Punkt auf dem Gebirgskamm.

nächsten Nachbarpunkt gleich ist. Die linke Abbildung entspricht genau der Beobachtung, wie sie anhand Abbildung 3.10 beschrieben wurde. Schneidet die Projektionslinie genau einen Punkt auf dem Gebirgskamm, erzeugt dieser Punkt einen Vorzeichwechsel zwischen den beiden Punkten, die den Punkt auf dem Kamm umschließen. Die rechte Abbildung zeigt, dass es mehrere Punkte hintereinander in der Distance Map mit dem gleichen Grauwert geben kann, aber auch ein Hinweis für einen Gebirgskamm sind. Dies entspricht im Grauwertgebirge einem Plateau.

Der Algorithmus sucht nun in x -und in y Richtung - von oben nach unten und von links nach rechts - in der Distance Map nach diesen Mustern und markiert die Punkte nach den Eigenschaften *strong*, *good*, *weak* und *none*. Diese Markierung gibt die Stärke der Sicherheit des Punktes wieder, ein Punkt auf dem Gebirgskamm zu sein.

Wurden alle Punkte in beide Richtungen untersucht, hat jeder Punkt ein passendes Label. Diese Labels werden weiter benutzt, um eine Graphenrepräsentation des Skeletts zu erstellen.

Der Algorithmus findet zu jedem Punkt das richtige Label und alle Punkte, die zu einem Gebirgskamm gehören [Cha07].

### 3.3 Weitere Verfahren

*Autor: Christopher Kroll*

Neben Thinning und Distanztransformation bilden Algorithmen, die auf kritische Punkte basieren, die dritte Kategorie. Hier werden kritische, bzw. wichtige Punkte detektiert, die dann zu einem Skelett verbunden werden. Zwei Vertreter dieser Kategorie sind der einfache Kritische-Punkte-Ansatz und die Triangulations-Technik.

Beim einfachen Kritische-Punkte-Ansatz wird das Quellbild spalten- oder zeilenweise durchgegangen. Werden Objektpixel gefunden, so wird die Mitte dieses 'Objektpixel-blocks' markiert. Dies sind die kritischen Punkte, die nach der Markierung verbunden werden und somit das Skelett bilden. Ein Nachteil dieses Ansatzes ist, dass abhängig von der Lage des Objektes und der Wahl, ob das Quellbild zeilen- oder spaltenweise

durchlaufen wird unterschiedliche Ergebnisse erzielt werden, das Verfahren ist somit nicht isotrop. TODO naeher beschreiben, drauf eingehen

Zu erwähnen ist außerdem, dass es neben diesen drei vorgestellten Skelettierungskategorien noch weitere Algorithmen gibt, die jedoch sehr speziell und sich dadurch schwer kategorisieren lassen (zum Beispiel das Hamilton-Jacobi-Skelett).

# 4 Implementierung der Algorithmen

*Autor: Sandra Schröder*

Im Rahmen des Projekts wurden zwei Algorithmen für die Skelettierung implementiert. Dies ist zum einen die Skelettierung nach Thinning nach dem Algorithmus der in Abschnitt 3.1.1 beschrieben wurde. Die Skelettierung mittels Distanztransformation wurde nach einer eigenen Idee entwickelt und umgesetzt.

Die Skelettierung läuft in zwei Schritten ab. Erst wird der Spieler segmentiert. Man erhält als Ergebnis ein Binärbild, welches im zweiten Schritt weiterverarbeitet wird, um ein Skelett zu extrahieren. Die Segmentierung des Spielers ist bei beiden Ansätzen identisch. In diesem Kapitel wird die Umsetzung der Algorithmen anhand signifikanten Codeausschnitten der Implementierungen vorgestellt. Ein Überblick über die technische Umsetzung gibt einen Eindruck über die verwendeten Programmiersprachen und Arbeitsumgebungen.

## 4.1 Technische Umsetzung

*Autor: Christopher Kroll*

Für die Implementierung wurden die bereitgestellten iMacs am Informatikum in Stellingen gewählt. Der Vorteil war neben der Performanz die schon eingerichtete Arbeitsumgebung. Um die Verbindung zu der Kinect herzustellen, wurde das quelloffene Framework libfreenect der OpenKinect-Gruppe benutzt. Libfreenect bietet Treiber und Bibliotheken, um zum Beispiel die Bilder der Kamera anzuzeigen oder den eingebauten Motor zu steuern.

Um die empfangenen Bilder verarbeiten zu können wurde die Bibliothek OpenCV eingesetzt. Sie unterstützt unter anderem beim Speichern und Laden von Bildern und bei der Segmentierung.

Der Großteil der Programmierung erfolgte in der Sprache Python. Hierbei handelt es sich um eine leicht zu erlernende Skriptsprache. Allerdings traten im Laufe des Projektes Performanzprobleme bei der Implementierung des Thinning-Algorithmus auf. Aus diesem Grund wurde über einen Python-Wrapper der in C++ implementierte Algorithmus eingebaut. Näheres dazu wird in Kapitel 4.3 beschrieben. Als Entwicklungsumgebung wurde Spyder für die Python-Programmierung und Apples Xcode für die C++-Programmierung benutzt. TODO numpy

## 4.2 Spieler-Segmentierung

*Autor: Sandra Schröder*

Es wird anhand der Tiefeninformation segmentiert, die die Kinect liefert. Der Open-Source-Treiber für die Kinect - *Freenect* - bietet Funktionen für den Zugriff auf die Tiefenwerte. Die Funktion `pretty_depth` des Moduls `frame_convert` normiert die Tiefenwerte auf das Intervall [0, ..., 255].

Listing 4.1: Tiefenwerte zurückgeben

```
1 def get_depth():
2     return frame_convert.pretty_depth
3     (freenect.sync_get_depth()[0])
```

Um nicht für jeden einzelnen Pixel die Bedingung zu überprüfen, ob er den Schwellwert für die Segmentierung überschreitet beziehungsweise unterschreitet, wird die effiziente Numpy-Funktion `logical_and` benutzt, die global auf dem Bild arbeitet und für das gesamte Bild die Schwellwertbedingung prüft. Eine pixelweise Überprüfung wäre mit Python eine ineffiziente Lösung.

Für den Schwellwert werden zwei Werte definiert, um ein Intervall festzulegen, in dem sich das Objekt befinden darf. In Listing 4.2 legen die Variablen `current_depth` und `threshold` das Intervall fest.

Listing 4.2: Spielersegmentierung in Python

```
1 #Segmentierung des Tiefebildes
2 depth = 255 * np.logical_and
3     (depth >= current_depth - threshold,
4      depth <= current_depth + threshold)
```

Da die Funktion auf Numpy-Arrays arbeitet, muss das Bildobjekt zuvor in ein Array umgewandelt werden. Es wurden vorgefertigte Funktionen von OpenCV benutzt, die diese Konvertierung vornehmen.

## 4.3 Skelettierung mittels Thinning

*Autor: Christopher Kroll*

Da man sich schon durch den Seminarteil mit dem Algorithmus 'A Fast Parallel Algorithm for Thinning Digital Patterns' beschäftigte, fiel die Wahl des Thinning-Algorithmus zunächst auf diesen.

Der Algorithmus arbeitet mit Binärbildern, also alle Objektpixel sind auf 1, bzw. 255 gesetzt und alle Hintergrundpixel auf 0. Die in Kapitel 4.2 vorgestellte Spieler-Segmentierung liefert jedoch ein Grauwertbild. So war nun der erste Schritt dieses Bild in ein Binärbild umzuwandeln. Die Idee war das zweidimensionale Grauwertbildarray mit zwei Schleifen zu durchlaufen und jeden Wert, der größer als 128 ist auf 255 zu setzen. Jeder Wert der kleiner oder gleich 128 ist, wird auf 0 gesetzt (siehe Listing 4.3).

Listing 4.3: Binarisierung des Grauwertbildes.

```
1 for(int y=0; y < height; y++) {
2     for(int x=0; x < width; x++) {
3         if(img(x,y)>128){
4             img(x,y)=255;
5         } else{
6             img(x,y)=0;
7         }
8     }
9 }
```

So erhalten wir ein Array, das nur aus den Werten 0 und 255 besteht. Das Problem dabei war allerdings die nicht ausreichende Leistungsfähigkeit von Python. Schon bei diesen zwei ineinander verschachtelten Schleifen bestand keine Echtzeitfähigkeit mehr. Ohne auch nur den iterativen - und somit rechenintensiven - Thinning-Algorithmus eingesetzt zu haben stockte das Ergebnisbild zu stark. Aus diesem Grund musste eine Alternative zu Python gefunden werden. Die Wahl fiel auf die leistungsstärkere Compilersprache C++.

TODO quelle heraussuchen vergleich python-c++?

Nun sollten jedoch nicht zwei unterschiedliche Programme entstehen - Distanztransformation mit Python und Thinning mit C++ und die Spieler-Segmentierung in Python sollte auch für die Thinning-Skelettierung benutzt werden. Die Lösung war die Möglichkeit der Anbindung von Python an eine andere Programmiersprache, in diesem Fall C++.

## 4.4 Skelettierung mittels Distanztransformation

*Autor: Sandra Schröder*

Der theoretische Ablauf der Skelettierung wurde bereits im Kapitel 3 beschrieben. Zur Rekapitulation werden die Schritte kurz aufgezählt.

Die Skelettierung anhand der Distanztransformation läuft folgendermaßen ab:

- Bestimmen der Distanztransformation des Binärbildes
- Berechne den Gradientenbetrag auf der Distance Map
- Differenz zwischen dem Gradientenbild und der Distance Map bilden

Die Distance Map kann mit einer Funktion aus der Bildverarbeitungsbibliothek *OpenCV* einfach berechnet werden. Die Funktion (Listing 4.4) erwartet als Eingabe das Originalbild (`img`) und ein Bild (`dist_img`), um das Ergebnis speichern zu können (gleiche Größe und Dimension wie das Originalbild). Eine weitere Möglichkeit, die die Funktion bietet, ist die Angabe einer Metrik, nach der der Abstand eines Pixels zum Hintergrund bestimmt wird. Es wurde die euklidische Metrik benutzt.

Listing 4.4: Berechnen der Distance Map des Spielers.

```
1 cv.DistTransform(img, dist_img, distance_type=cv.CV_DIST_L2)
```

Zur Bestimmung des Gradientenbetrages der Distance Map wurde die Numpy-Funktion `gaussian_gradient_magnitude` genutzt. Die Funktion berechnet den Gradientenbetrag mit Ableitungen der Gaussfunktion. Die Variable `sigma` ist die Standardabweichung des Gaussfilters. Das Ergebnis dieser Funktion wird in ein Bildobjekt konvertiert und entsprechend festgelegter Schwellwerte (`lowerbound`, `upperbound`) segmentiert.

Listing 4.5: Gradientenbetrag der Distance Map und Segmentierung des Gradientenbeitragsbildes.

```
1 #Gradienten-Berechnung
2 ndimage.gaussian_gradient_magnitude(distancemap ,
3     sigma,gradient_image)
4 #Vor -und Nachbearbeitungen
5 #Segmentierung
6 cv.InRangeS(gradient_image,lowerbound,upperbound,
7     threshed_gradient_image)
```

Zur Differenzbildung und endgültigen Berechnung des Distanzskelettes werden die Bildobjekte in Arrays umgewandelt. Diese Arrays können einfach voneinander abgezogen werden.

Listing 4.6: Differenz zwischen Distance Map und segmentiertem Gradientenbetrag

```
1 #Berechnung Differenzbild
2 diff = distancemap_array - threshed_gradient_array
3 diff = diff * 1.0
```

Bei der Implementierung wurden in keinem Fall Operationen ausgeführt, die auf einzelne Pixel zugreifen. Situationen, in denen pixelweise Operationen durchgeführt werden könnten, wurden umgangen, indem Arrayoperationen oder Funktionen aus der OpenCV-Bibliothek benutzt wurden. Ein pixelweiser Zugriff könnte bei einer Interpretersprache wie Python zu einer sehr langsamen Ausführung der Skelettberechnung führen.

# 5 Evaluation

*Autor: Sandra Schröder*

In diesem Kapitel werden die beiden entwickelten Algorithmen gegenüber gestellt. Zwei zentrale Kriterien zur Evaluation der Algorithmen werden thematisiert.

- Skelettqualität
- Echtzeitfähigkeit

Zur Bewertung der Skelettqualität greifen wir auf die gewünschten Eigenschaften eines Skeletts zurück (Kapitel 3) und überprüfen, ob die Algorithmen sie realisieren. Auch die Echtzeitfähigkeit ist von Bedeutung. Grundlage für eine verzögerungsfreie Interaktion des Spielers mit der Anwendung (dem Spiel) ist eine schnelle Berechnung des Skeletts. Denn trotz nachfolgender Verarbeitungsschritte soll eine hohe Anzahl von Bildern schnell (Größenordnung 20 Bilder pro Sekunde) prozessiert werden.

## 5.1 Skelettqualität

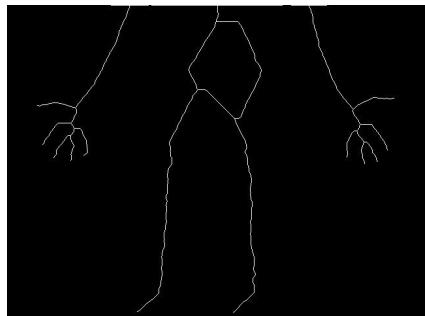
Zum Vergleich der Algorithmen wurden Posen des Spielers aufgenommen (Abbildung 5.1). Bei gleichen Eingabedaten weisen die beiden Algorithmen Resultate von unterschiedlicher Qualität auf. Pose 1 zeigt den Unterkörper des Spielers. Beide Skelette beschreiben die Grundstruktur des Spieler sehr gut. Arme und Beine werden jeweils durch eine im ursprünglichen Objekt mittig liegende Linie beschrieben. Die Hände des Spielers sind deutlich vom Körper gestreckt und die Finger gespreizt. Der Thinning-Algorithmus ist in der Lage die Topologie der Hände und Finger wiederzugeben, während bei der Distanztransformationen das Skelett der Hände schlechter zu erkennen ist. Die zweite getestete Pose erzeugt ebenfalls Skelette von unterschiedlicher Qualität. Auffallend sind die Lücken in den Skelettlinien des Skeletts, das aus der Distanztransformation entstanden ist. Der Thinning-Algorithmus hat stets ein Skelett mit zusammenhängenden Skelettkomponenten als Ergebnis. Jedoch ist beim Skelett des Thinning-Algorithmus auf Bauchhöhe ein horizontaler Ausläufer zu erkennen, der nicht die eigentliche Form des Objektes beschreibt. Das aus der Distanztransformation resultierende Skelett gibt den Oberkörper bis zu den Schultern als eine Linie ohne Ausläufer wieder.

Eine weitere Eigenschaft einer Skelettlinie ist ihre Breite, die sich idealerweise auf einen Pixel beschränkt. Dies ist beim Thinning der Fall, bei der Distanztransformation fällt die Skelettlinie breiter aus. Wird das Skelett zur Datenkomprimierung genutzt, enthalten unnötige Pixel mehr Informationen, als eigentlich benötigt wird. Für eine konkrete

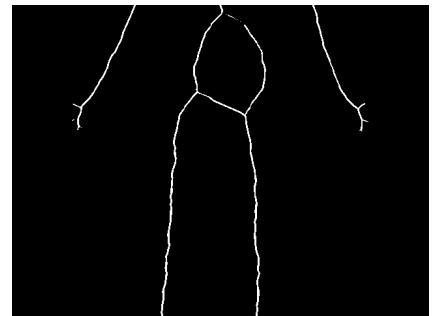
Pose 1



Spieler (segmentiert)



Thinning-Skelett

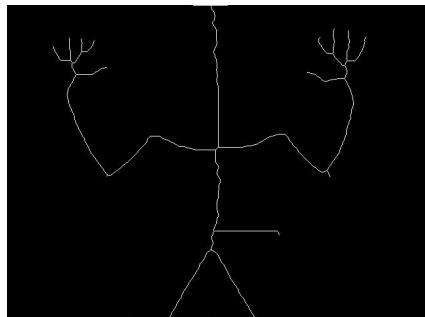


Distanzskelett

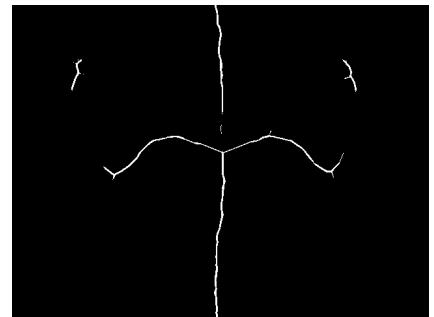
Pose 2



Spieler (segmentiert)



Thinning-Skelett



Distanzskelett

Abbildung 5.1: Vergleich zweier Posen. Der Thinning-Algorithmus und die Skelettierung mittels Distanztransformation weisen jeweils Resultate von unterschiedlicher Qualität auf.

Anwendung muss diskutiert werden, ob eine 1-Pixel-Breite unbedingt nötig ist. Bezuglich der Skelettqualität liefert der Thinning-Algorithmus sehr gute Skelette. Hier sind keine weiteren Verbesserungen mehr nötig. Bei dem Distanzskelett sind Verbesserungen, besonders bei der Konnektivität des Skelettes, möglich. In Abschnitt 5.3 werden zwei Verbesserungsmöglichkeiten vorgestellt.

## 5.2 Laufzeiten

Die Laufzeit wurde mit dem Profiling-Tool *CProfile* gemessen. Es erstellt für Python-Programme eine Statistik über die Laufzeiten eines Programms und seine einzelnen Funktionen. Es wurde jeweils eine Statistik für die Implementierung des Thinning-Algorithmus und des Algorithmus der Distanztransformation erstellt. Die wichtigsten Funktionen der Programme wurden aus der Statistik entnommen. Die Tabellen 5.1 und 5.2 geben einen Überblick über die Laufzeiten.

Obwohl der Thinning-Algorithmus bezüglich der geforderten Eigenschaften von Skeletten die besten Ergebnisse liefert, ist er deutlich langsamer als die Skelettierung mittels Distanztransformation. Zur Initialisierung der Algorithmen wird einmalig zu Beginn eine Startfunktion ausgeführt. Diese beansprucht weniger als 30 Sekunden und kann gegenüber der Gesamlaufzeit der Anwendung (Minuten bis Stunden) vernachlässigt werden. Die Spielersegmentierung und die Skelettberechnung wird pro Bild einmalig ausgeführt. Die Segmentierung unterscheidet sich bei den beiden Ansätzen nicht. Die Abweichung zwischen den Laufzeiten ist auf Messungenauigkeiten zurück zu führen. Die Skelettierung nimmt bei beiden Ansätzen deutlich mehr CPU-Zeit in Anspruch als die Segmentierung. Pro Bild ergibt sich eine Verhältnis der Laufzeiten von:

$$\frac{\text{Laufzeit Skelett aus Thinning}}{\text{Laufzeit Skelett aus Distanztransformation}} = \frac{0.003 \text{ s} + 0.904 \text{ s}}{0.004 \text{ s} + 0.076 \text{ s}} \approx 11 \quad (5.1)$$

Die Berechnung des Distanzskeletts ist also um mehr als eine Größenordnung schneller als die Berechnung mittels Thinning. Der Thinning-Algorithmus ist maschinennah in C++ programmiert, und die Distanztransformation ist in der Interpretersprache Python umgesetzt. Es ist also anzunehmen, dass die Distanztransformation durch die Verwendung einer Compiler-Sprache deutlich beschleunigt werden kann.

46371 Funktionsaufrufe in 16.723 CPU-Sekunden			
Funktion	Anzahl Aufrufe	Gesamtzeit [s]	Pro Aufruf [s]
Startfunktion	1	16.723	16.723
Spielersegmentierung	126	0.453	0.004
Berechnung des Distanzskeletts	126	9.594	0.076

Tabelle 5.1: Laufzeiten Skelettierung mittels Distanztransformation

1424 Funktionsaufrufe in 28.159 CPU-Sekunden			
Funktion	Anzahl Aufrufe	Gesamtzeit [s]	Pro Aufruf [s]
Startfunktion	1	28.159	28.159
Spielersegmentierung	29	0.088	0.003
Skelettierung nach Thinning	29	26.224	0.904

Tabelle 5.2: Laufzeiten Skelettierung mittels Thinning

## 5.3 Distanztransformation - Verbesserung der Skelettqualität

Im Folgenden wird ein Ansatz zur Verbesserung der Skelettqualität vorgestellt.

Das Skelett, welches mit der Methode der Distanztransformation bestimmt wurde, weist Lücken zwischen den Skelettteilen auf. Um die Topologie und geometrische Eigenschaften des Objekts gut wiederzugeben, ist ein lückenloses Skelett wünschenswert.

Die Segmentierung des Gradientenbetrages der Distance Map erzeugt Lücken im Distanz-skeletts. Dies wird anhand von Abbildung 5.2 deutlich. Im Gradientenbild (Teilabbildung b) sind noch durchgehende Skelettlinien zu erkennen. Per Schwellwert-Filter wird das Skelett aus dem Gradientenbild extrahiert. Für das gezeigte Beispielbild konnte allerdings kein geeigneter Schwellwert gefunden werden, der sowohl Pixelkonnektivität gewährleistet als auch Artefakte verhindert. Teilabbildung c zeigt die Segmentierung des Gradientenbildes. In diesem Schritt ist die Pixelkonnektivität erstmals nicht gewährleistet. Auch in den folgenden Schritten wird die Pixelkonnektivität nicht wieder hergestellt.

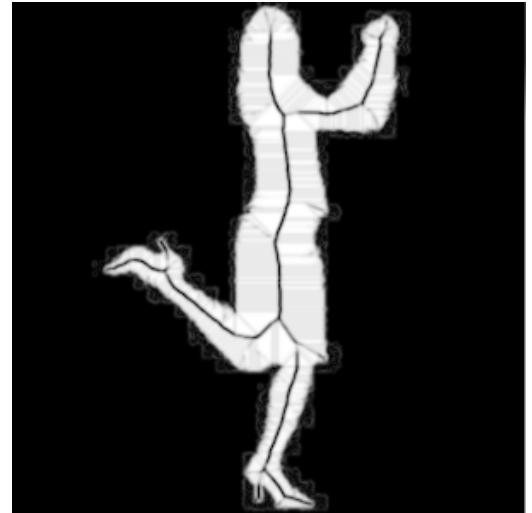
In den nächsten Abschnitten werden Methoden zur Verbesserung des Skeletts vorgestellt. Dies umfasst zum einen die Herstellung Konnektivität der Skelettkomponenten beziehungsweise dem Füllen der Lücken zwischen den Skelettlinien. Zum anderen ist es wünschenswert, dass die Skelettlinie des Objektes nicht breiter als ein Pixel ist. Zuerst wurden markante Punkte auf dem Skelett bestimmt. Diese wurden genutzt, um mittels Breitensuche oder Tiefensuche Pfade zwischen ihnen zu finden und sie zu verbinden. Punkte werden miteinander verbunden, wenn sie nah genug beieinander sind. Das Ergebnis ist ein Skelett mit zusammenhängenden Skelettlinien.

### 5.3.1 Berechnung von Features auf dem Skelett

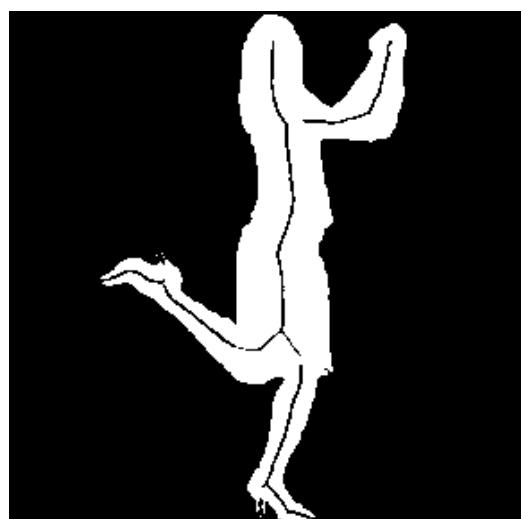
Die Bildverarbeitungsbibliothek *OpenCV* bietet Funktionen zur Bestimmung von markanten Punkten, sogenannten *Features*, in einem Bild. Die Funktion `goodFeaturesToTrack` [ST94] bestimmt die stärksten Ecken in einem Bild oder Bildausschnitt. Mittels eines Qualitätsmaßes wird entschieden, ob die Stärke einer Ecke an einem bestimmten Pixel ausreicht, um in die Featuremenge aufgenommen zu werden. Für die Funktion können der Wert des Qualitätsmaßes, den eine Ecke erfüllen muss, die Anzahl der Ecken, die gefunden werden sollen und die minimale Distanz zwischen den stärksten Ecken übergeben werden. Die Qualität einer Ecke wird anhand von Eigenwerten bestimmt. Die Eigenwerte beziehen



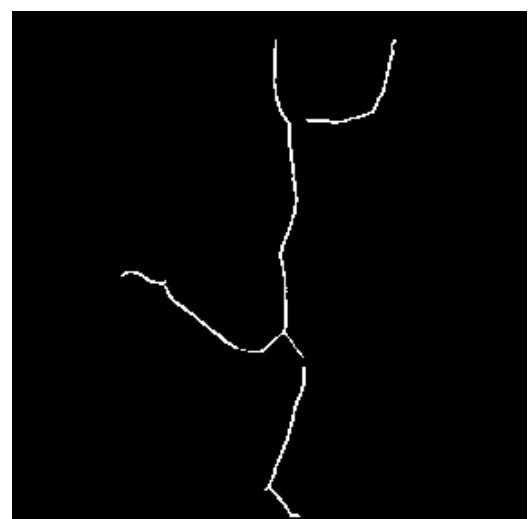
(a)



(b)



(c)



(d)

Abbildung 5.2: Bestimmung des Skelett mittels Distanztransformation. (a) Originalbild  
(b) Gradientenbild (c) Gradientenbild segmentiert(d) Skelett

sich dabei auf die Kovarianzmatrix von Ableitungen einer festgelegten Umgebung eines Pixels. Es wird minimale Eigenwert für die Eckendetektion weiterverwendet. Ist der minimale Eigenwert einer Ecke kleiner als das gewünschte Qualitätsmaß, wird diese Ecke verworfen. Die verbleibenden Ecken werden nach ihrer Qualität absteigend sortiert. Anschließend wird überprüft, ob es in der spezifizierten Distanz Ecken gibt, die stärker sind. Abbildung 5.3 zeigt das Ergebnis der Berechnung. Die Kreise markieren die

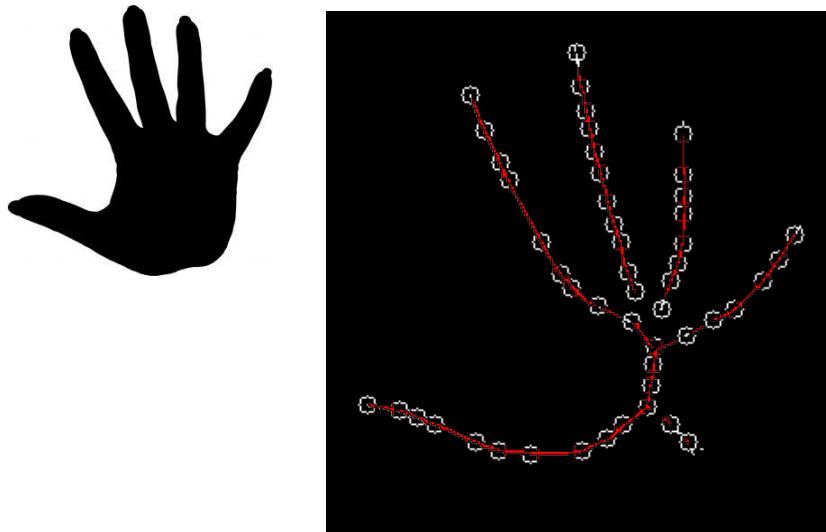


Abbildung 5.3: Ergebnis der Funktion `goodFeaturesToTrack`. Links: Originalbild. Rechts: Berechnetes Distanzskelett mit Features. Die roten Linien markieren das Rohskelett, die weißen Kreise markieren die Features, die auf dem Skelett gefunden wurden.

Feature-Punkte. Wie man erkennen kann, befinden sich die Features auf der Skelettlinie. Dies ist eine Grundlage für die weiteren Verbesserungen des Skeletts. Befänden sich Features außerhalb der Skelettlinien, könnte die ursprüngliche Form des Skeletts und die Topologie des Objekts verfälscht werden. Diese Problematik wurde bei dem gewählten Algorithmus [Cha07] nicht beobachtet.

### 5.3.2 Breitensuche

Der Breitensuche-Algorithmus durchsucht einen Graphen ausgehend von einem Startpunkt nach weiteren Knoten. Der Algorithmus sucht zunächst nur nach direkt nachfolgenden Knoten und somit in die Breite des Graphen. Knoten, die bereits besucht wurden, werden markiert. Wurde ein Knoten noch nicht besucht, wird er in eine *Queue* (Warteschlange) aufgenommen.

Mittels Breitensuche sollen Pfade zwischen den markanten Punkten auf dem Skelett gefunden werden. Die markanten Punkte entsprechen Knoten in einem Graphen. Für die Nachbarschaftsbeziehung zwischen zwei Knoten wird ein Nachbarschaftsmaß definiert. Abbildung 5.4 zeigt, wie überprüft wird, ob ein Punkt in einem festgelegten Intervall

des Punktes  $(x, y)$  liegt. Es wird eine Suchdistanz für beide Richtungen festgelegt. Sie wird entsprechend der minimalen Distanz, die zwischen markanten Punkten erlaubt ist (Abschnitt 5.3.1), gewählt.

Erst wird in x-Richtung gesucht, dann in y-Richtung. Fällt der Punkt in das Intervall, wird er als besucht markiert und mit dem Punkt  $(x, y)$  verbunden. In Anhang ?? befindet sich der Quellcode zur Breitensuche (Listing ??, Zeile 78/79).

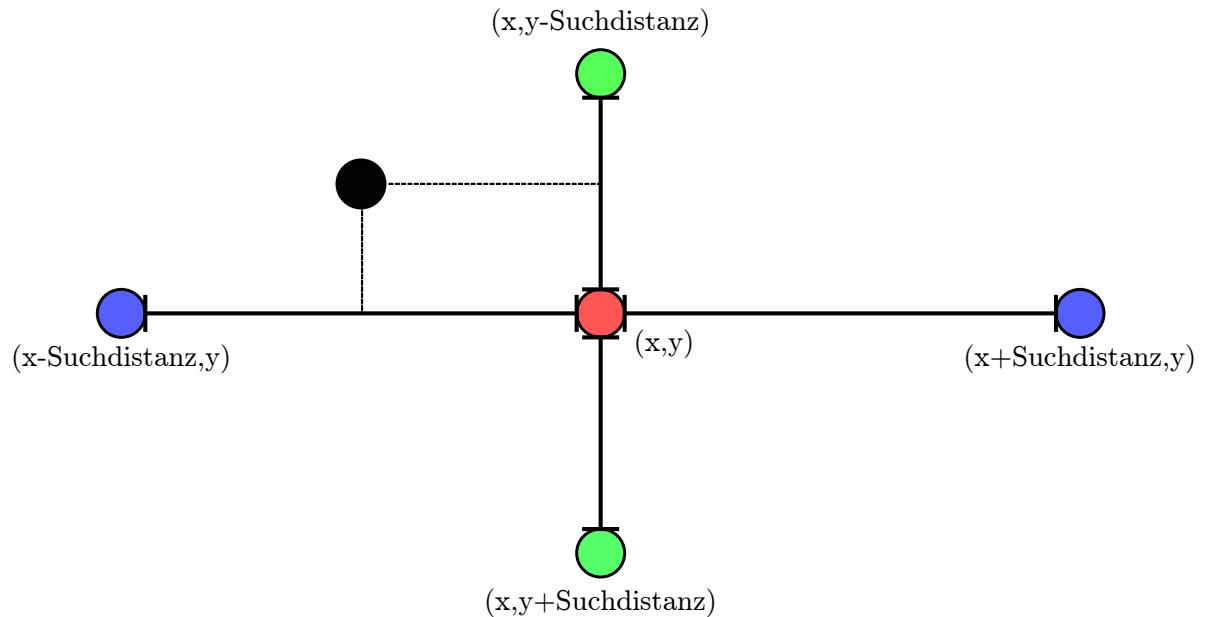


Abbildung 5.4: Das aufgespannte Suchkreuz vom Punkt  $(x, y)$  (rot) aus. Für den schwarzen Punkt wird überprüft, ob er in den festgelegten Suchintervallen des Punktes  $(x, y)$  liegt.

Das Ergebnis des Algorithmus ist in Abbildung 5.5 im Vergleich zum vorigen Skelett zu sehen. Das Skelett besitzt nun zusammenhängende Komponenten. Auffällig sind die Ausläufer, die auf die Vorgehensweise des Breitensuchealgoritmus zurückzuführen sind. Wird die Suchdistanz zu groß gewählt, findet der Algorithmus mehrere Nachfolger, die er dann mit dem aktuellen Punkt verbindet. Das Ergebnis sind fächerartige Ausläufer.



Abbildung 5.5: Ergebnis der Breitensuche.

### 5.3.3 Tiefensuche

Wie bei der Breitensuche wird ausgehend von einem Startknoten ein Graph nach weiteren Knoten durchsucht. Im Gegensatz zur Breitensuche erfolgt das Traversieren des Graphens in die Tiefe. Wurde ein Nachfolgeknoten gefunden, wird für diesen weiter geprüft, ob er ebenfalls einen Nachfolgeknoten besitzt. Dies wird solange wiederholt (rekursiv) bis kein Nachfolgeknoten mehr gefunden werden kann. Mittels *Backtracking* wird der Pfad zurückverfolgt. Anschließend wird ein neuer Knoten als Startpunkte gesucht, der noch nicht besucht wurde und das Durchsuchen nach nächsten Nachbarn wird für diesen Knoten wiederholt.

Die Suche nach dem nächsten Nachbarn funktioniert wie bei der Breitensuche (Abbildung 5.4).

Das Ergebnis ist in Abbildung 5.6 zu sehen. Es fällt auf, dass es noch Lücken im Skelett gibt, was auf die Suchdistanz zurückzuführen ist. Ist sie zu klein, können keine weiteren Punkte im Umkreis des aktuellen Punktes gefunden werden und der Algorithmus endet für diesen Pfad.



Abbildung 5.6: Ergebnis der Tiefensuche.

Um die Konnektivität des Skeletts zu gewährleisten, muss das Skelett nachbearbeitet werden.

Das Skelett, welches aus der ersten Stufe der Verbesserung entstanden ist, besitzt Zusammenhangskomponenten. Innerhalb der Zusammenhangskomponenten ist die Pixelkonnektivität gewährleistet, die globale Konnektivität zwischen den Komponenten jedoch nicht. Die Idee des nächsten Verbesserungsschrittes ist es, die Zusammenhangskomponenten zu verbinden, die am nächsten beieinander liegen. Dazu werden die Punkte aus den Zusammenhangskomponenten extrahiert, die die Endpunkte der Komponenten sind. In Graphenrepräsentation entsprechen diese Punkte den Wurzeln und Blättern des Waldes. Das sind entweder Punkte ohne Vorgänger oder ohne Nachfolger. Für  $n$  Komponenten müssen  $n - 1$  Verbindungen gefunden werden, um vollständige Konnektivität zu gewährleisten. Dafür werden die kürzesten Verbindungen zwischen Endpunkten verschiedener Komponenten bestimmt: Es wird komponentenweise vorgegangen. Für die Endpunkte der aktuellen Komponente wird der nächste Nachbarpunkt bestimmt, der ein Endpunkt einer anderen Komponente ist. Für die Bestimmung des Abstandes wird die euklidische Metrik  $d_2$  (Gleichung 3.1) verwendet. Welcher Punkt auf welcher Komponente liegt, wird bereits bei der Tiefensuche bestimmt. Der Quellcode befindet sich in Anhang ??.

In Abbildung 5.7 wurden die Endpunkte farbig markiert. Dies sind alles Kandidaten zum Verbinden der Komponenten. Das Skelett besitzt in dieser Abbildung fünf Zusammenhangskomponenten, wobei eine Zusammenhangskomponente aus genau einem Punkt besteht (auf dem Ringfinger der Hand). Dieser Punkt besitzt weder einen Vorgänger noch einen Nachfolger.

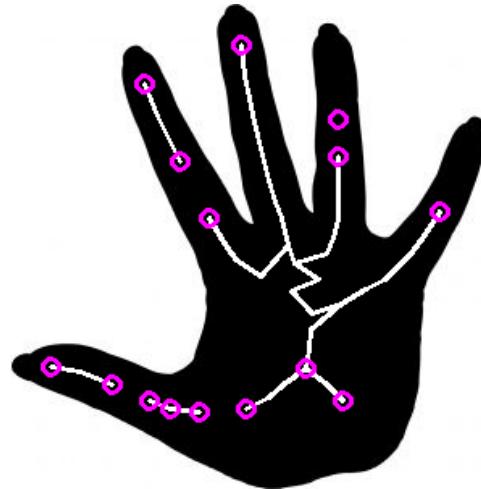


Abbildung 5.7: Zusammenhangskomponenten und ihre Endpunkte.

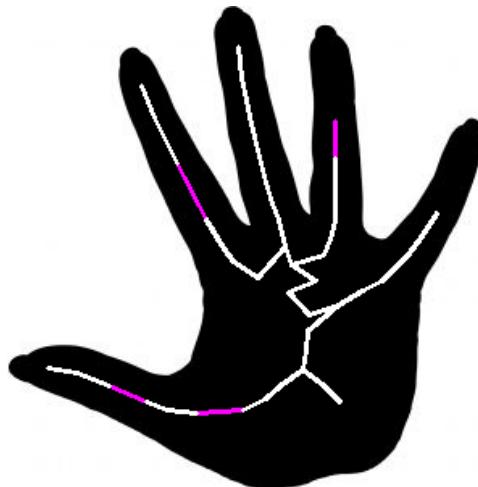


Abbildung 5.8: Ergebnis des zweiten Verarbeitungsschrittes der Tiefensuche. Die weiße Linie ist das Skelett, das aus dem ersten Verarbeitungsschritt entstanden ist. Die grünen Linien sind die Verbindungen zwischen den Zusammenhangskomponenten.

Abbildung 5.8 zeigt, wie die Zusammenhangskomponenten miteinander verbunden wurden. Es wurden immer die Zusammenhangskomponenten miteinander verbunden, die sich am nächsten liegen. Nun ist die Skelettkonnektivität gegeben.

## 5.4 Fazit

Beide Algorithmen (Thinning und Distanztransformation) liefern Skelette. Diese realisieren die gewünschten Eigenschaften eines Skelettes. Dabei ist die Skelettqualität unterschiedlich. Der Thinning-Algorithmus liefert bezüglich Pixelkonnektivität, 1-Pixel-Breite und Wiedergabe der Topologie hervorragende Ergebnisse. Bei der Skelettierung mittels Distanztransformation ist vor allem auffällig, dass die Skelettkonnektivität nicht vollständig gegeben ist.

Es wurden Verbesserungsmöglichkeiten diskutiert, die zur globalen Skelettkonnektivität bei einem Distanzskelett führen. Dabei liefert die Tiefensuche nicht nur bei der Konnektivität sehr gute Ergebnisse. Auch die topologischen Eigenschaften des Objektes werden hervorragend beschrieben. Bei der Breitensuche fielen vor allem die fächerartigen Ausläufer auf. Diese sind nicht Teil der Topologie des ursprünglichen Objekts. Es muss auch hier abhängig vom konkreten Anwendungsfall diskutiert werden, ob die Ausläufer stören. Das durch die Tiefensuche verbesserte Skelett realisiert am besten die in Kapitel 3 genannten Qualitätsanforderungen.

Da diese Verbesserungen nach der offiziellen Projektarbeitszeit entstanden sind, konnten keine detaillierten Laufzeit-Messungen mit der Kinect vorgenommen werden. Die Laufzeiten der beiden Algorithmen wurden deshalb mit dem Tool *CProfile* auf statischen Bildern gemessen. Die Laufzeiten fallen bei beiden Algorithmen für einen einzigen

Durchlauf sehr gering aus. Deshalb wurden beide Algorithmen mehrmals ausgeführt (bis zu 500 Mal, entspricht einer Bildfolge von etwa 20 Sekunden). Dabei viel auf, dass der Breitensuche-Algorithmus bei mehreren Durchläufen deutlich langsamer wurde als der Tiefensuche-Algorithmus, der selbst bei einer hohen Anzahl von Wiederholungen schnell ein Ergebnis lieferte. Bei einer Integration in die Kinect-Umgebung ist der Tiefensuche-Algorithmus aufgrund der geringen Laufzeit und der überzeugenden Skelettqualität zu bevorzugen.

# **6 Zusammenfassung und Ausblick**

## **6.1 Zusammenfassung**

## **6.2 Ausblick**

- Der Vergleich hat gezeigt, dass Algorithmen unterschiedliche Ergebnisse liefern
- in qualität und laufzeit
- weiternutzen der skelette
- beide eignen sich
- vor allem: anwenden der verbesserten version des distanzskelettes: performant mit der kinect oder nicht?
- wenn nicht: kann man das verbessern?
- grundstein dafür wurde gelegt.

# Literaturverzeichnis

- [Cha07] Sukmoon Chang. Extracting Skeletons from Distance Maps. *International Journal of Computer Science and Network Security*, 7, 2007.
- [ST94] Jianbo Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 593 –600, jun 1994.