

Implementierung von Skelett-Algorithmen mit dem Kinect-Sensor

Arbeitsbereich Kognitive Systeme (KOGS)
Fachbereich Informatik
Universität Hamburg

Projektarbeit
Projekt Bildverarbeitung
Johannes Böhler, Christopher Kroll, Sandra Schröder

Sommersemester 2012 bis Wintersemester 2012/2013

1 Einleitung

2 Aufgabenstellung

2.1 Die Kinect-Funktionsweise

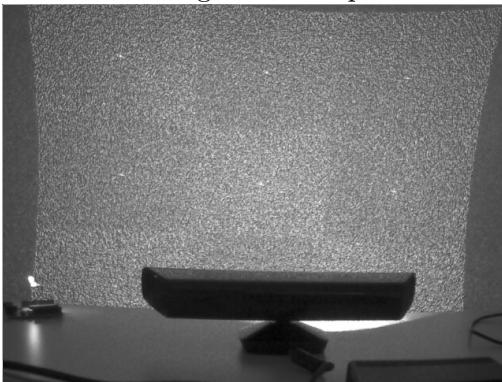
2.1.1 Funktionsweise

Die drei Haupt Hardware-Komponenten der Kinect sind ein Infrarot Projektor, eine Infrarot Kamera sowie eine Infrarot Kamera.



Die Kombination aus Infrarot Strahler und Infrarot Kamera 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.



2 Aufgabenstellung

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.

Wird die Tiefeninformation dazu genutzt um die Pixel mit Hilfe der durch die RGB Kamera erfassten Farbwerte der XY Koordinaten im dreidimensionalen Raum anzugeben, erhält man eine sogenannte (farbige) Punktwolke. Die in der 2d -Betrachtung benachbarte Pixel müssen in der 3d Repräsentation nicht miteinander verbunden da die Z-Koordinate unterschiedliche Werte aufweisen kann.

2.1.2 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. Die Firma PrimeSense (welche von Microsoft aufgekauft wurde) implementiert ein besonderes Verfahren zur Generierung dieses Musters. 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 angewandte 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. . Trotzdem ist die Reichweite des IR Projektors eingeschränkt, da zu hohe Intensitäten der IR Strahlen Augenschäden verursachen könnten.

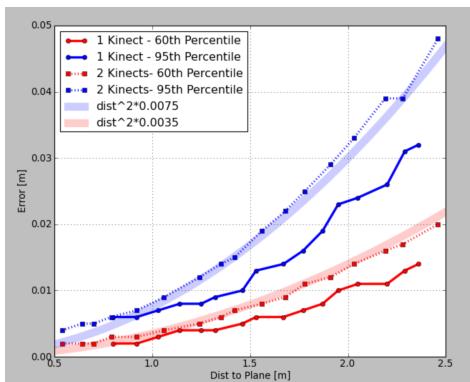


Es ist wichtig dass die Wellenlänge der Infrarot strahlen konstant bleibt, was durch eine konstante Temperatur der Laserdiode und eine konstante Stromleistung (60 mW) gewährleistet wird. Problem hierbei sind variierende Außentemperaturen(die empfohlene Betriebstemperatur liegt zwischen 5 und 35 °C). Um diese Konstanz zu gewährleisten wird

ein sogenanntes Peltier-Element verwendet, welches sowohl kühlen als auch wärmen kann.

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 übertragbarer Datenraten 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 3mm, in der Z Ebene 1cm. 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. Die wird durch einen Filter welcher auf dem IR Kameraobjektiv angebracht ist realisiert. Trotz des Filters sollte die Kinect eher in abgedunkelten Innenräumen verwendet werden, da im Sonnenlicht auch elektromagnetische Wellen im Infrarotbereich enthalten sind.

RGB Kamera

Die RGB Kamera nimmt bei einer Wiederholrate von 30 Hz ebenfalls mit einer Auflösung von 640x480 Pixeln auf, könnte aber alternativ 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

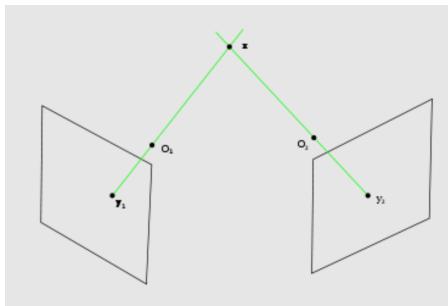
Die Kinect beinhaltet vier Mikrofone welche verteilt verbaut sind. Sie dienen sowohl der Erfassung von Ton, als auch der Lokalisierung und Unterscheidung von Soundquellen. Jedes der Mikrophone tastet mit einer Quantisierungsauflösung von (65536) Bit und einer Abtastrate von 16 KHz ab.

2.1.3 Tiefenberechnung

Die Berechnung der Objektentfernungen (Tiefenwerte) innerhalb der Kinect erfolgt durch Aussendung von strukturiertem Licht durch die Projektionseinheit und durch Weiterverwertung der durch die IR 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 bei dem von der Kinect angewendeten Verfahren eine Punktematrix verwendet, welche fest definiert ist. Die IR Kamera sowie der Projektor müssen sich hierzu in einem fest definiertem, gleich großem Abstand zueinander befinden. Die Umgebungsreflektionen der projizierten Punktematrix werden von der Infrarotkamera erfasst. Um aus diesem Gitter von Infrarotpunkten Tiefeinformation extrahieren zu können wird das Verfahren der aktiven Stereotriangulation verwendet.

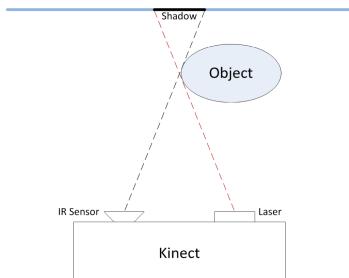
Die Triangulation ist ein Verfahren zur optischen Abstandsmessung welches sich hierzu trigonometrischer Funktionen innerhalb von aufgespannten Dreiecken bedient. Es wird allgemein zwischen aktiver und passiver Triangulation unterschieden. Aktive Triangulation bedeutet dass mindestens eine strukturierte 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 einzurufen. 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 Tiefeinformationen. Hierzu 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.



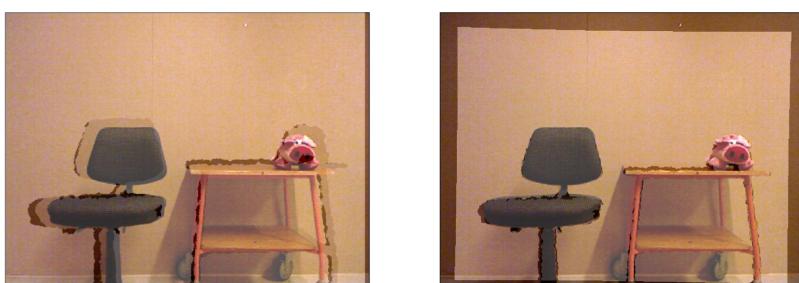
Der Grund warum die Pixel der „Schablone“ im Projektor zufällig angeordnet sind liegt darin, dass die unterschiedlichen lokalen Nachbarschaftsbedingungen die Pixelzuordnung zwischen dem statischen und dem dynamisch veränderten Bild erleichtern.

2.1.4 Das Schattenproblem

Aufgrund der Entfernung der verbauten RGB Kamera zur Infrarot Kamera weisen die Bilder beider Kameras einen kleinen Versatz auf. Schatten im Tiefenbild entstehen aufgrund der Entfernung des Infrarotprojektors zur Infrarotkamera. Der Schatten im Muster macht es für den Sensor unmöglich die Tiefe festzustellen. Daher werden die Pixel in diesen Regionen 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.



2 Aufgabenstellung

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

2.2 Anforderungen

3 Die Skelettierung

3.1 Thinning

3.2 Distanztransformation

3.3 Weitere Verfahren

3.4 Verwandte Arbeiten

3.4.1 Extracting Skeletons From Distance Maps

3.4.2 A Fast Parallel Algorithm for Thinning Digital Patterns

4 Implementierung der Algorithmen

4.1 Wahl der Algorithmen

4.2 Die Arbeitsumgebung

4.3 Spieler-Segmentierung

4.4 Thinning

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 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.

4.4.1 Fast Parallel Thinning Algorithm

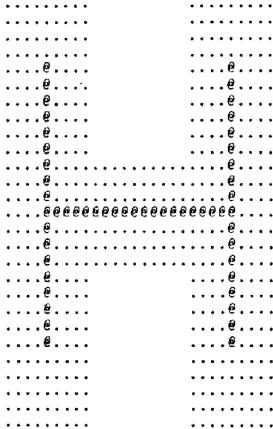
Der Algorithmus ist insgesamt in mehrere Iterationen unterteilt. Die Randpixel des Musters werden Schicht für Schicht abgetragen. Die Iterationen sind selbst wiederum in mehrere Subiterationen unterteilt. Das Abtragen der „Schichten“ wird somit in zwei unterschiedliche Phasen aufgespalten. Mit Hilfe der ersten Subiteration werden sowohl Süd- und Ostgrenzpunkte als auch Nordwest Eckpunkte entfernt.

4 Implementierung der Algorithmen

Das entfernen von Nord- und Westgrenzpunkten sowie von Südosteckpunkten erfolgt in der zweiten Subiteration.

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 zu Beginn ein binär digitalisiertes Bild vorliegen zu haben. 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 nun in Abhängigkeit von den 8 Nachbarpixeln, Punkt für Punkt iterative Transformationen auf die Matrix 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 P_1 über dessen Transformation entschieden wird, ist mit allen acht Nachbarn 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 $(i + 1, j + 1)$

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 P_1 aus dem Muster gelöscht wenn er folgende Bedingungen erfüllt:

- a) $2 \leq B(P_1) \leq 6$ B à Anzahl der Nachbarn von $P_1 \neq 0$ Die Anzahl der Nachbarn von P_1 welche den Wert 1 haben, muss somit zwischen 2 und 6 liegen.

- b) $A(P1)=1$ AàAnzahl 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.

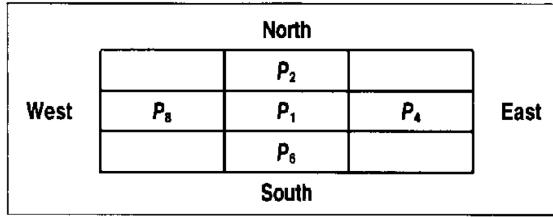
0	0	1
1	P_1	0
1	0	0

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

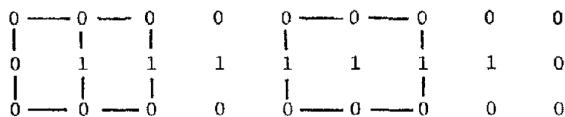
- a) $2 <= B(P1) <= 6$
- b) $A(P1)=1$
- c) $P2*P4*P8=0$
- d) $P2*P6*P8=0$

In der zweiten Subiteration haben sich nur die Bedingungen c und d geändert.

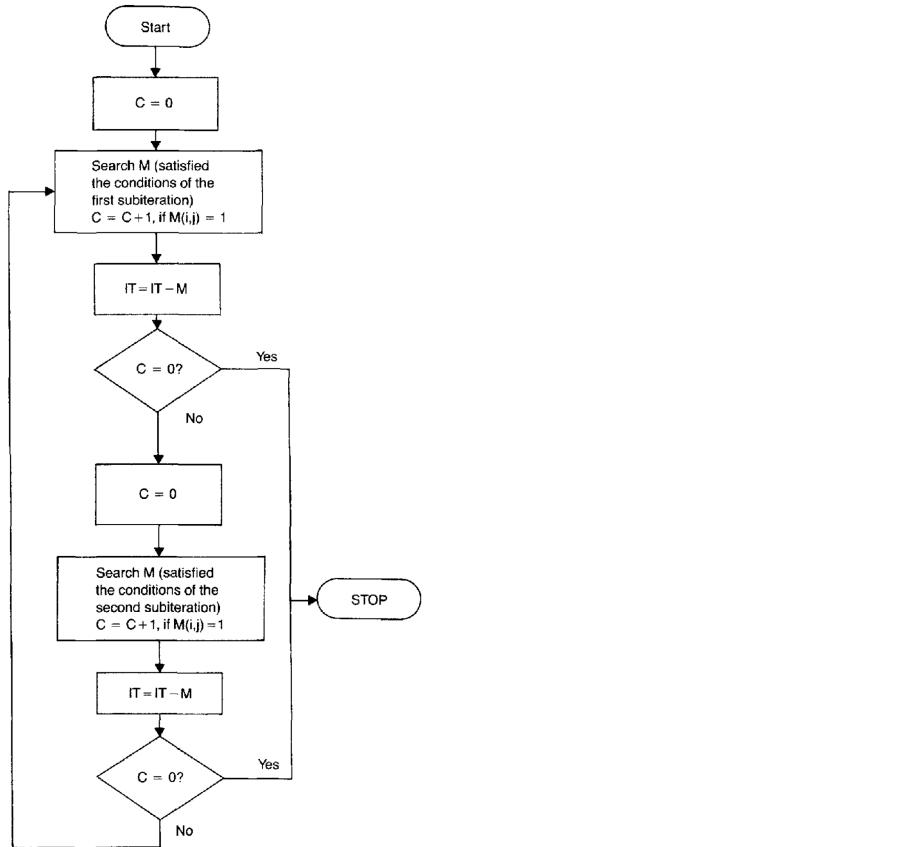
Um die Bedingungen der ersten Subiteration zu erfüllen muss $P4=0$ oder $P6=0$ oder ($P2=0$ und $P8=0$) erfüllt sein. Dies impliziert dass P1 entweder Süd- oder Ost-Grenzpunkt, oder Nordwesteckpunkt ist. Um die Bedingungen der zweiten Subiteration zu erfüllen muss $P2=0$ oder $P8=0$ oder ($P4=0$ und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



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.



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 anscheinend keine zu löschenen 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 nochmals statt. Ist auch nach dem Durchlaufen der zweiten Subiteration der Counter ungleich null so wird der Vorgang iterativ fortgeführt.



Der Algorithmus erzielt sehr gute Ergebnisse im Bezug auf Verbundenheit, und Rausch-sicherheit bei Randpunkten. Die Bedingungen welche zum Auffinden der zu löschen 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.

4.5 Distanztransformation

5 Ergebnisse

- Vergleich der Algorithmen
- Anhand der Kriterien
 - Erhaltung der Topologie
 - Pixelkonnektivität
 - Zentriert
 - 1 Pixel breit
- Echtzeitfähigkeit -> Messungen machen -> Vergleich
- Verbesserung des Skeletts (Distanztransformation) mit Breitensuche um Pixelkonnektivität zu erreichen -> Weitere Verbesserungen? -> Ohne Features sondern anhand der weißen Pixel
- Anwendung: Vergleich von Posen -> Features bestimmen. Vllt sowas wie SSpannweite" der Pose in x und in y Richtung (Abstand des "linkestenßum "rechtesten"Pixel).

5 Ergebnisse

6 Zusammenfassung

6 Zusammenfassung

7 Fazit und Ausblick

7.1 Fazit

7.2 Ausblick

8 TODO Anhang

Tabelle: Wer hat was geschrieben?

Vollständiger Quellcode