



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Seminar-Ausarbeitung

Nikolaus Rusitska

nikolaus.rusitska [at] haw-hamburg.de

Human Motion Tracking

Verfolgung menschlicher Bewegung zur Steuerung

Inhalt

1	Einleitung	3
1.1	Ziel der Arbeit	3
2	Bisherige Arbeiten	4
2.1	Anwendungen 2.....	4
2.2	Projekt 1	4
3	Vorhandene Systeme.....	6
3.1	Software	6
3.1.1	Camera Calibration Toolbox	6
3.1.2	OpenCV.....	6
3.2	Hardware	6
3.2.1	Monokamera.....	6
3.2.2	Stereokamera.....	6
3.2.3	Multikamera	7
3.2.4	Time of Flight.....	7
3.2.5	Kinect	7
4	Vorgehen	9
4.1	Aufgabenbereiche.....	9
4.1.1	Tiefenbild	9
4.1.2	Person extrahieren	9
4.1.3	3D-Modell	9
4.1.4	Gelenke und Verbindungen.....	10
4.2	Komponentenvergleich	10
4.3	Bearbeitung und Erstellung der Komponenten	11
5	Abgrenzung und Bewertungskriterien	12
6	Risiken.....	13
6.1	Aufnahme und Bild	13
6.2	Rekonstruktion	13
6.3	Geschwindigkeit.....	14
	Referenzen.....	15
	Abbildungsverzeichnis	16

1 Einleitung

Beim Human Motion Tracking werden die Bewegungen von Menschen digital erfasst. Diese Arbeit beschäftigt sich mit der Bewegungsverfolgung durch digitale Bildverarbeitung. Es wird dabei anhand von Bildaufnahmen von Menschen die aktuelle Pose ihres Körpers interpretiert.

Die Position des Körpers wird dann im Verlauf der Zeit betrachtet um zu sehen welche Körperteile wie bewegt wurden.

1.1 Ziel der Arbeit

Das Ziel der Bewegungsverfolgung soll die Steuerung eines Systems sein. Dabei soll es möglich sein, entweder ein interaktives Programm oder auch einen Roboter (z.B. Knickarmroboter) zu steuern. Dabei ist es auch wichtig, dass die Bewegungserkennung möglichst wenig Verzögerung erzeugt.

Es wird von einer Situation ausgegangen, in der eine einzelne Person vom Aufnahmesystem gesehen wird. Es soll dann die Person von ihrer Umgebung unterschieden werden können und die Bewegungen der Person erkannt und verfolgt werden.

Bei der Steuerung von Programmen sollen die Bewegungen so verfolgt werden, dass es danach möglich ist, die Bewegungen einzelner Körperteile zu unterscheiden, sodass es später möglich ist, Gesten mit einzelnen oder mehreren zu erkennen.

Soll beispielsweise ein Knickarmroboter gesteuert werden, muss die Bewegungserkennung einzelne Gelenke erkennen können, so dass später die Bewegungen auf den Roboter übertragen werden können, der möglicherweise eine andere Anzahl von Gelenken oder auch andersartige Gelenke hat.

Es ist dafür also wichtig, die Bewegung so zu analysieren, dass die Modifikation einzelner Gelenke der aufgenommenen Person abgeschätzt werden kann.

Es wird davon ausgegangen, dass bei beiden Einsatzszenarien ist die Bewegung der Arme am wichtigsten ist.

Wenn eine Fortbewegung als Steuerungskommando verwendet würde, so wäre es möglich dass das gefilmte Subjekt den Aufnahmebereich verlässt.

Mit Händen und Armen gibt es mehr Möglichkeiten Gesten zu formulieren, als z.B. mit den Beinen. Einfache Roboter, wie ein Knickarmroboter, die Werkzeuge führen oder Aktionen wie Greifen durchführen sind häufig Interpretationen des Menschlichen Arms.

Aus diesen Gründen wird bei der Bewegungserkennung falls nötig der Erkennung von Armen und Händen eine größere Bedeutung zugestanden, als anderen Körperteilen.

Das System, das erstellt wird, soll möglichst einfach aufzubauen sein und kostengünstig sein. Grund hierfür ist, dass vor allem Möglichkeiten betrachtet werden sollen, die auch wirtschaftlich umsetzbar wären. Es soll hier zwar kein System mit Marktreife entwickelt werden, aber im Zuge der Praxisorientierung soll auch eine Wirtschaftliche Praktikabilität berücksichtigt werden.

2 Bisherige Arbeiten

Die vorherigen Arbeiten beschäftigten sich mit der Rekonstruktion von zweidimensional gefilmten Szenen. Ziel dabei war es, aus den Filmaufnahmen ein sich bewegendes Kamera ein dreidimensionales Modell der aufgenommenen Szene zu berechnen.

2.1 Anwendungen 2

In der Ausarbeitung zu der Vorlesung Anwendungen 2 (zweites Mastersemester, Sommersemester 2010) wurden verschiedene Techniken und Algorithmen vorgestellt, die bei den verschiedenen Schritten einer dreidimensionalen Rekonstruktion aus zweidimensionalen Aufnahmen von Nutzen sind.

In der Arbeit wurden verschiedene Algorithmen vorgestellt, mit denen besondere Merkmale in Bildern gefunden werden, die dann später in weiteren Bildern wiedergefunden werden können. Dadurch ist es möglich übereinstimmende Punkte zwischen Bildern entdecken, anhand derer die Positionen der aufnehmenden Kameras ermittelt werden kann. Dadurch kann die Geometrie der aufgenommenen Szene rekonstruiert werden.

Weiterhin wurden Algorithmen und Bibliotheken diskutiert, mit denen dreidimensionale Koordinaten von Merkmalen berechnet und visualisiert werden können. Durch diese kann eine dreidimensionale Punktwolke der Szene hergestellt werden.

Desweiteren wurden Bibliotheken besprochen, die in der Lage sind eine dreidimensionale Punktwolke zu triangulieren und so in ein nutzbares dreidimensionales Gittermodell zu überführen.

Für die aktuelle Arbeit sind Merkmalsfindungen teilweise interessant. Sie können bei der Berechnung eines Tiefenbildes aus dem Bildpaar einer Stereokamera verwendet werden.

Die Triangulation von dreidimensionalen Punkten bleibt für diese Arbeit ebenfalls aktuell.

2.2 Projekt 1

Im Projekt 1 (zweites Mastersemester, Sommersemester 2010) wurde Rekonstruktionen von verschiedenen Szenen durchgeführt. Dabei wurden Bilder verwendet, die eine Szene aus zwei verschiedenen Ansichten zeigen. Beide Bilder wurden dabei von der gleichen Kamera aufgenommen. Es wurden auch zweidimensionale Ansichten dreidimensionaler Modelle verwendet, da so die geometrischen Relationen leichter zu überprüfen waren.

In den Bildern wurden markante Punkte (Features) ermittelt und miteinander verglichen um inhaltliche Übereinstimmungen in beiden Bildern zu finden. Anhand derer wurde eine Geometrie zwischen beiden Bildern erstellt, die möglichst viele Übereinstimmungen zwischen beiden Bildern berücksichtigt.

Anhand des ermittelten Verhältnisses beider Bilder zueinander wurden die in den Bildern gefundenen Punkte in dreidimensionale Punkte überführt.

Da mit unkalibrierten Kameras gearbeitet wurde, wurde die endgültige Skalierung der Szene nicht berücksichtigt. Der Maßstab der Szene kann so nicht berechnet werden, wurde in diesem Projekt allerdings auch nicht gesucht.

Für die aktuelle Arbeit ist die Berechnung von dreidimensionalen Punkten anhand von zwei Bildern noch interessant. Allerdings ist hier gewünscht, dass ein Maßstab für die Szene (bzw. das aufgenommene Subjekt) vorliegt, damit bei Bewegungen des verfolgten Subjektes auch bekannt ist wie weit sich das Subjekt bewegt hat.

3 Vorhandene Systeme

3.1 Software

In der vorherigen Arbeit (Anwendungen 2) wurden bereits Bibliotheken zur dreidimensionalen Bildverarbeitung vorgestellt. An dieser Stelle sind nur noch einige Ergänzungen erwähnt, die auf Grund der geänderten Zielsetzung zwischen dieser Arbeit und der vorherigen nötig geworden sind.

3.1.1 Camera Calibration Toolbox

Mithilfe der Camera Calibration Toolbox für Matlab können Stereokameras mithilfe eines Schachbrettmusters kalibriert werden. Hierdurch werden innere Parameter der Kameras, wie die Verzerrung durch die Linse berechnet. Sind diese bekannt können sie danach wieder ausgeglichen werden.

3.1.2 OpenCV

OpenCV ist eine Open-Source Bibliothek in C/C++ deren Entwicklung von Intel angestoßen wurde. Ein wichtiger Anspruch dieser Bibliothek an sich selbst ist die Geschwindigkeit der enthaltenen Algorithmen.

Es sind stets auch viele aktuelle Algorithmen in der Bibliothek enthalten. In der aktuellen Version (V2.2 Dezember 2010) enthält die Bibliothek mehr als 500 verschiedene Algorithmen. Darunter sind auch Algorithmen zur dreidimensionalen Bildverarbeitung und z.B. zur Triangulation enthalten.

3.2 Hardware

3.2.1 Monokamera

Eine einzelne Kamera ist als Aufnahmesystem leicht zu installieren und verhältnismäßig kostengünstig. Allerdings ist dabei die Rekonstruktion der aufgenommenen Objekte aufwändiger. Die verfolgten Bewegungen werden ungenauer beobachtet, als bei Systemen mit mehreren Kameras, da weniger Bildinformationen verfügbar sind.

Ist die Kamera fest positioniert, (was im Anwendungsfall sehr wahrscheinlich ist) so lässt sich zwar die Bewegung des Subjektes verfolgen, allerdings ist eine dreidimensionale Verfolgung der Bewegungen nur mit hohem Aufwand möglich.

3.2.2 Stereokamera

Bei einer Stereokamera handelt es sich um zwei Kameras, die in einer bekannten Lage zueinander stehen. Es wird dabei die menschliche Wahrnehmung mit zwei Augen imitiert. Es wird ein dreidimensionaler Eindruck gewonnen, dadurch dass aus zwei leicht unterschiedlichen Blickwinkeln aufgenommen wird. Da die Lage der Kameras zueinander durch eine initiale Kalibrierung bekannt ist, vereinfacht sich die Rekonstruktion und die dreidimensionale Verfolgung.

Die einzelnen Kameras einer Stereokamera sind in der Regel nahe bei einander, wodurch das aufgenommene Subjekt zwar mit einem dreidimensionalen Eindruck aufgenommen wird, allerdings nur von einer Seite. Es gibt Stellen, die auf keiner der Aufnahmen erscheinen, wodurch es

nicht möglich ist, ein vollständiges Modell zu erzeugen.

Auch kann es durch die Bewegung der Person zu Verdeckungen auf der zugewandten Seite kommen, beispielsweise wenn ein Arm Teile des Körpers verdeckt. Um dieses zu vermeiden, könnte zu Beginn der Interaktion ein Modell erstellt werden, auf das die verfolgten Bewegungen angewendet werden, um danach entstandene Verdeckungen zu erkennen.

3.2.3 Multikamera

Bei einem Multikamerasystem werden mehrere Kameras so ausgerichtet, dass das Subjekt von mehreren Seiten aufgenommen wird und zwar so, dass möglichst alle Seiten des Subjekts aufgenommen werden. Es werden mindestens drei Kameras, so im Raum verteilt, dass das Subjekt in seinem Aktionsbereich¹ möglichst vollständig (von allen Seiten) aufgenommen wird. Dadurch ist es möglich ein möglichst vollständiges Modell der Oberfläche des Ziels zu erstellen, weil (fast) nichts verdeckt wird.

Bei einem solchen System kann durch verwenden weiterer Kameras die Genauigkeit der Rekonstruktion erhöht werden. Allerdings erhöhen sich dabei die Materialkosten, sowie der Rechenaufwand, da auch ein höheres Datenvolumen verarbeitet werden muss.

Da die Kameras in einem Multikamerasystem um das Subjekt herum aufgebaut werden müssen, benötigt es mehr Platz. Während andere Systeme in sich geschlossen und leicht transportabel sein können (z.B. ein Aufnahmesystem, das an einem Bildschirm angebracht ist), muss ein Multikamerasystem an genau dem Ort aufgebaut werden, an dem es auch eingesetzt wird.

3.2.4 Time of Flight

Eine Time of Flight –Kamera liefert dreidimensionale Informationen, indem die Laufzeit des Lichtes von der aufgenommenen Szene zurück zur Kamera misst (bzw. hin und wieder zurück). So ist für jeden Bildpunkt auch seine Entfernung zur Kamera bekannt.

Die Weiterverarbeitung der Daten wird vereinfacht, da die meisten Modelle dieser Kameras direkt Tiefenbilder der aufgenommenen Szene anbieten.



Abbildung 1: TOF-Kamera

3.2.5 Kinect

Kinect ist eine Steuerungserweiterung für die Spielekonsole Xbox 360 von Microsoft. In dem Open-Source Projekt OpenKinect wurden Treiber entwickelt um Kinect auch ohne die Xbox 360 Konsole verwenden zu können.

Kinect stellt (unter anderem) Farbbilder aus einer zweidimensionalen Kamera und Tiefenbilder (ermittelt durch IR-Projektion und Aufnahme durch schwarz-weiß Kamera) zur Verfügung.

¹ Als Aktionsbereich wird hier der Bereich verstanden, in dem das Subjekt von allen beteiligten Aufnahmegeschichten aufgenommen werden kann. In diesem Bereich sollen die zu verfolgenden Bewegungen des Subjekts durchgeführt werden.



Abbildung 2: Ein Kinect-Sensor

4 Vorgehen

In diesem Abschnitt wird das geplante Vorgehen für die Erstellung des Systems beschrieben. Es werden die verschiedenen geplanten Teile und Abschnitte der Arbeit vorgestellt.

4.1 Aufgabenbereiche

Die Erkennung der Bewegung kann in verschiedene Aufgaben unterteilt werden. Mit einigen Themen dazu wurde sich bereits in vorherigen Ausarbeitungen beschäftigt. (Vgl. Kapitel 2)

Zu den einzelnen Aufgaben gehören:

- Erzeugung eines Tiefenbildes
- Person extrahieren
- 3D-Modell generieren
- Gelenke und Verbindungen unterteilen

4.1.1 Tiefenbild

An dieser Stelle geht es darum aus den aufgenommenen Daten ein Tiefenbild zu erzeugen. Je nach verwendetem System müssen erst aus mehreren Bildströmen Tiefenbilder oder 3D-Punkte berechnet werden.

Sollte ein für die Aufnahme ein System verwendet werden, das direkt ein Tiefenbild der betrachteten Szene zur Verfügung stellt, muss in diesem Schritt außer einer eventuellen Vorverarbeitung für den nächsten Schritt nichts weiter getan werden.



Abbildung 3: Tiefenbild einer Person

4.1.2 Person extrahieren

In diesem Schritt soll die handelnde Person von der restlichen Umgebung getrennt werden, damit im folgenden Teil nur ein Modell der Person erstellt wird.

Dabei ist eine möglichst genaue Trennung wichtig, damit nicht Punkte verloren gehen, die zum verfolgten Subjekt gehören. Das Zurücklassen von Punkten, die tatsächlich zur Umgebung gehören ist genauso zu vermeiden. Beides kann große Auswirkungen auf die Qualität des zu erzeugenden Modells haben.

4.1.3 3D-Modell

Nachdem das Tiefenbild nur noch Punkte enthält, die zu der Person gehören, soll aus dem Tiefenbild ein 3D-Modell erzeugt werden. Dazu werden mittels Triangulation aus den einzelnen Punkten Dreiecke erzeugt.

In Abhängigkeit von der Aufnahme und der nachfolgenden Unterteilung in Gelenke und Verbindungen kann es nötig sein das erzeugte Modell zu erweitern. Wenn die Person nur von einer Seite aufgenommen wird, dann entspricht das Modell nur der Oberfläche dieser Seite der Person und ist sozusagen auf der Rückseite offen. Es kann dann nötig sein, dieses Modell – beispielsweise durch Spiegelung – zu einem geschlossenen Modell zu erweitern.

4.1.4 Gelenke und Verbindungen

Wurde ein dreidimensionales Modell erzeugt, müssen Gelenke und Gelenkverbinder (Knochen) hinzugefügt werden. Zu der Oberfläche soll ein Skelett hinzugefügt werden. Dieses sollte möglichst auch der tatsächlichen Beschaffenheit des menschlichen Körpers entsprechen.

An dem Skelett kann dann die Pose der Person erkannt werden und Haltung und Position einzelner Körperteile bestimmt werden.

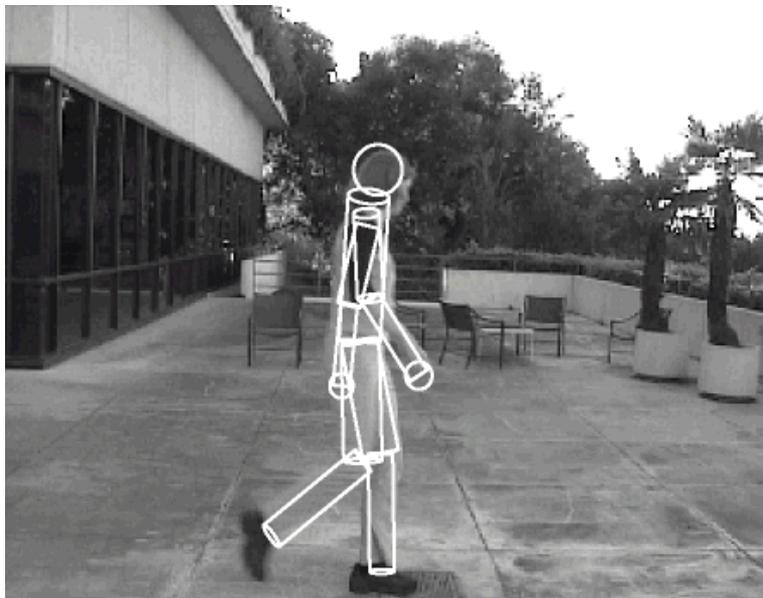


Abbildung 4: Einfaches Menschen-Modell mit Segmenten

4.2 Komponentenvergleich

Für die einzelnen Aufgaben, die durchzuführen sind, gibt es zum Teil schon bestehende Lösungen. Diese müssen dann evaluiert werden.

Für die verfügbaren Lösungen für eine Aufgabe müssen passende Vergleichskriterien gefunden und untersucht werden.

Gibt es für eine der Aufgaben mehrere Lösungen, so sind wichtige Entscheidungskriterien für die Auswahl geringe Komplexität der Komponente und hohe Geschwindigkeit. Die Qualität der Ergebnisse wird hier unter die Geschwindigkeit gestellt. Dieses soll allerdings nicht beliebig so verfolgt werden. Es muss ein annehmbarer Kompromiss aus Geschwindigkeit und Qualität erreicht werden.

4.3 Bearbeitung und Erstellung der Komponenten

Bibliotheken und Algorithmen, die für die Erstellung des Systems ausgewählt werden, müssen möglicherweise noch angepasst werden.

Es müssen Schnittstellen geschaffen werden, um die einzelnen Komponenten miteinander zu verbinden. Falls nötig müssen dabei auch noch Weiterverarbeitungen durchgeführt werden um die gewonnenen Daten für die folgende Komponente vorzubereiten.

Verwendete Komponenten müssen möglicherweise auch erst für die Verwendung angepasst werden. Auf Grund der Vorannahmen können einige der Verfahren eventuell auch noch optimiert werden. Einige Details über die Beschaffenheit der Szene sind bekannt, da davon ausgegangen wird, dass die Person ungefähr in der Mitte des Bildbereiches stehen wird und dort auch das naheste Objekt ist.

Da die Proportionen eines Menschen zu einem gewissen Grad ähnlich sind, können bei der Bildung des Modells einige Vorannahmen gemacht werden. Zudem kann das Modell anhand von idealisierten Menschen-Modellen überprüft werden.

Beim Erzeugen des Skeletts kann es nötig werden weite Teile der Komponente selbst zu erstellen. Die Algorithmen dazu können entweder speziell auf ihre ursprüngliche Anwendung angepasst sein, oder sie verhalten sich zu ungenau bzw. zu allgemein. (Auch bei einem ausgestreckten Arm muss ein Ellenbogen-Gelenk vorhanden sein, etc.)

5 Abgrenzung und Bewertungskriterien

Grundlagen für die Bewertung des Systems sollen Genauigkeit der einzelnen gefundenen Posen, Robustheit der Bewegungsverfolgung über die Zeit und die Geschwindigkeit der Erkennung sein.

Bei den einzelnen Posen ist es wichtig, dass die Gelenke und Verbindungen auch möglichst den Knochen und Gelenken der aufgenommenen Person entsprechen. Dabei ist allerdings zu beachten, dass eine vollständige Rekonstruktion des menschlichen Skeletts hier nicht gesucht ist. Nur die interessanten Körperstellen müssen möglichst genau mit Knochen versehen werden.

Hier liegt dementsprechend das Augenmerk auf Oberkörper, Armen und Händen. Dabei ist bei den Händen nicht unbedingt eine vollständige Erkennung der Fingerhaltung und aller Fingerknochen vonnöten. Es geht nur darum die Haltung der Hand ausreichend gut interpretieren zu können.

Erkannt werden sollen zumindest die Lage und die Ausrichtung der Schulter-, Ellenbogen- und Handgelenke. Die Ausrichtung von Kopf und Oberkörper sollten auch erkannt werden. Unterleib, Beine und Füße müssen nicht zwingend vollständig erfasst werden.

Die Bewegungsverfolgung muss auch im Verlauf der Zeit Robust und sinnvoll sein. Gerade bei Verdeckungen ist kann es sonst zu Fehlern kommen. Das System soll in der Lage sein, damit umgehen zu können, falls ein Körperteil vorübergehend nicht sichtbar ist und später wieder auftaucht.

Wenn beispielsweise Hände sich gegenseitig verdecken, müssen die Hände danach wieder erkannt und vor allem auch richtig zugeordnet werden.

Es wäre wünschenswert, dass bei verdeckten Körperteilen anhand der sichtbaren Körperteile eine Schätzung erstellt wird, wo sich der verdeckte Körperteil momentan befinden müsste. Dieses ist allerdings kein notwendiges Kriterium der Arbeit.

Die Geschwindigkeit der Erkennung ist ein wichtiges Kriterium, da es für eine Steuerung wichtig ist, dass sie auch eine schnelle Antwort bietet.

Es ist wünschenswert, dass das System nahezu in Echtzeit arbeitet. Die Verfahren sollen schnell genug sein, dass keine besonders aufwändige Hardware zur zeitnahen Berechnung nötig ist.

Eine gewisse Verzögerung ist allein durch die Digitalisierung der aufzunehmenden Szene gegeben, allerdings sollte das restliche System schnell genug sein, dass die Verzögerung den Anwender nicht bei der Steuerung behindert.

Eine Laufzeit von über einer Sekunde ist nicht wünschenswert.

6 Risiken

Hier sollen einige Risiken und Probleme betrachtet werden, die bei der Erstellung des Systems auftreten können. Es werden auch mögliche Lösungsansätze angegeben. Dabei können selbstverständlich nicht alle Probleme vorhergesehen werden, es geht lediglich um eine Übersicht möglicher Schwachstellen.

6.1 Aufnahme und Bild

Bereits die Aufnahme der Szene kann Probleme verursachen. Die Beleuchtungssituation und die Oberflächenbeschaffenheit sowie die Farbe von Objekten in der Szene können die Erzeugung eines dreidimensionalen Bildes negativ beeinflussen.

Time of Flight-Kameras können auf Grund Spiegelungen fehlerhafte Daten erzeugen. Auch Sonnenlicht kann bei einer Infrarot-Ausleuchtung Probleme verursachen.

Die Farbe von Objekten kann die Erzeugung eines Tiefenbildes von einer Stereokamera stören. Besonders Glatte Flächen gleicher Farbe können dabei ein Problem darstellen.

Falls derartige Probleme auftreten, muss die Aufnahmesituation eventuell dementsprechend angepasst werden, um die Probleme zu beseitigen. Dazu könnte dafür gesorgt werden, dass kein Tageslicht einfällt und keine problematischen Objekte im Bild zu sehen sind.

Würde die Person durch andere Gegenstände verdeckt werden können sich hieraus ebenfalls Probleme ergeben. Es müsste daher sichergestellt sein, dass dieses nicht der Fall ist.

Eine Verdeckung durch eigene Körperteile kann nicht vermieden werden. Dieses muss dann dementsprechend in dem System berücksichtigt werden.

Falls es nicht möglich sein sollte dieses auszugleichen müssen die erlaubten Bewegungen für den Benutzer des Systems dementsprechend eingeschränkt werden.

6.2 Rekonstruktion

Es ist möglich dass bei der Rekonstruktion des Skeletts die einzelnen Gelenke nicht korrekt erkannt werden können.

Eine mögliche Abhilfe dafür könnte sein, zu Beginn der Interaktion Grundposen einzunehmen, in der die Gelenke ermittelt werden können. Deren Lage könnte dann später wieder berücksichtigt werden um die Gelenke zu positionieren.

Aufnahmesystem können nicht immer große Auflösungen bieten. Aktuelle Time of Flight-Kameras erreichen oft nur Auflösungen von ca. 200 x 200 Pixeln. Sollte dieses die Ursache für eine schlechte Rekonstruktion sein, die sich nicht mehr verbessern lässt, müsste angedacht werden stattdessen eine andere Bildquelle zu verwenden, die eine bessere Auflösung bietet.

6.3 Geschwindigkeit

Ein weiteres Risiko besteht darin, dass das System nicht ausreichend schnell arbeiten kann (und keine Optimierungen greifbar sind, die es beschleunigen können). Eine Möglichkeit ist dann die Auflösung und die Genauigkeit zu senken, um wieder eine annehmbare Geschwindigkeit zu erreichen.

Es wäre auch möglich leistungsfähigere Hardware zu verwenden. Dieses sollte aber dann in einem akzeptablen Rahmen bleiben.

Lässt sich dennoch keine annehmbare Geschwindigkeit erreichen, sollte darüber nachgedacht werden Komponenten des Systems, die zu aufwändig sind, gegen andere, einfachere Komponenten auszutauschen.

Referenzen

- Rusitska, N. (2010, August). Scene Reconstruction with Multiple View Geometry. Hamburg.
- Bandouch, J., & Beetz, M. (2009). Tracking Humans Interacting with the Environment Using Efficient Hierarchical Sampling and Layered Observation Models. In *IEEE Int. Workshop on Human-Computer Interaction (HCI). In conjunction with ICCV2009*. München, Deutschland.
- Fossati, A., Salzmann, M., & Fua, P. (2009). Observable Subspaces for 3D Human Motion Recovery. In *cvpr, 2009 IEEE Conference on Computer Vision and Pattern Recognition* (S. 1137-1144).
- Gaßmann, F., & Franke, K.-H. (27. Oktober 2005). Verfahren zur Triangulation ungeordneter 3D-Punktmengen. Ilmenau.
- Goshen, L., & Shimshoni, I. (2006). Balanced Exploration and Exploitation Model Search for Efficient Epipolar Geometry Estimations (BEEM). In *9th European Conference on Computer Vision (ECCV 2006)* (S. 151-164).
- Hartley, R., & Zisserman, A. (2003). *Multiple View Geometry in Computer Vision* (2. Ausg.). Cambridge: Cambridge University Press.
- Heyden, A., & Pollefeys, M. (2004). Multiple View Geometry. In G. Medioni, & S. B. Kang, *Emerging Topics in Computer Vision* (S. 45-107). Prentice Hall.
- Košecká, J. (2004). Multiple-View Geometry for Image-Based Modeling. *Siggraph Course #23*. Los Angeles.
- Mikolajczyk, K., & Schmid, C. (23. Februar 2005). A performance evaluation of local descriptors. Oxford, Montbonnot.
- Rosenhahn, B., Kersting, U., He, L., Smith, A., Brox, T., Klette, R., et al. (August 2005). A Silhouette Based Human Motion Tracking System. New Zealand: Centre for Imaging Technology and Robotics, University of Auckland.
- Sarfraz, M. S., & Hellwich, O. (2008). Head Pose Estimation In Face Recognition Across Pose Scenarios. *International Conference on Computer Vision Theory and Applications*, (S. 235-242). Berlin.
- Thormählen, T., & Seidel, H.-P. (2008). 3D-Modeling by Ortho-Image Generation from Image Sequences. In *ACM Transactions on Graphics, Vol. 27, No. 3, Article 86*.
- Wan, C., & Sato, J. (2008). Computing Multiple View Geometry in Space-Time from Mutual Projections of Multiple Cameras. *19th International Conference on Pattern Recognition (ICPR) 2008*, (S. 1-4).

Abbildungsverzeichnis

Abbildung 1: TOF-Kamera	7
Quelle: http://commons.wikimedia.org/wiki/File:TOF_Kamera.jpg	
Abbildung 2: Ein Kinect-Sensor	8
Quelle: http://commons.wikimedia.org/wiki/File:KinectSensor.png	
Abbildung 3: Tiefenbild einer Person	9
Quelle: http://commons.wikimedia.org/wiki/File:TOF_Kamera_3D_Gesicht.jpg	
Abbildung 4: Einfaches Menschen-Modell mit Segmenten.....	10
Quelle: http://www.cs.brown.edu/~black/3Dtracking.html	