



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Ausarbeitung Anwendungen 2 -
SoSe 2009
Arne Bernin
Segmentierung mit Differenzbildern

Inhaltsverzeichnis

1 Einführung	3
1.1 Einleitung	3
1.2 Zielsetzung des Masterprojektes	3
2 Technologien	4
2.1 Verfahren zur Differenzbildung	4
2.2 Bewegungserkennung	7
2.3 Verwendetes Verfahren	7
2.4 Hardware	8
2.5 Software Frameworks	8
2.6 Vergleich der Performance	10
3 Bewertung und Einordnung	11
3.1 Verfahren	11
3.2 Hardware	12
3.3 Framework	12
3.4 Ausblick	13
3.5 Fazit	13
Abbildungsverzeichnis	14
Literatur	14

1 Einführung

1.1 Einleitung

Spätestens seit der Vorführung des "Projektes Natal"[[Natal \(2009\)](#)] der Firma Microsoft auf der Spielemesse E3 [[E3 \(2009\)](#)] ist das Thema dreidimensionale Gestenerkennung wieder in den Fokus der interessierten Öffentlichkeit gerückt.

Dreidimensionale Gesten- und Bewegungserkennung steht, so die Versprechen der Veranstaltung, vor dem Durchbruch in den Massenmarkt. Sollte das angekündigte Produkt einer Kamera für den Consumerbereich nur annähernd das bieten, was die Präsentation verspricht, ist tatsächlich mit einer ähnlichen Umwälzung des Eingabemediums für Spiele zu rechnen, wie dies schon 2006 durch den Wiimote-Controller für die Spielekonsole Nintendo Wii geschehen ist, die immerhin fast 50% Marktanteil [[vgcharts \(2009\)](#)] erreicht, gegenüber leistungsmäßig deutlich stärkeren Konkurrenten [[spielekonsolen \(2009\)](#)].

Aber die Art der Interaktion mit dem Computerspiel ist ein wichtiger Faktor, der entscheidend zum Spielspaß und Engagement beim Spielen beiträgt [[Bianchi-Berthouze u. a. \(2007\)](#)]. Bleibt die Frage, wie wirkt sich die Veränderung der Eingabemediums im Spielesektor auf die Art aus, wie Menschen mit Computern kommunizieren ?

Die Realisierung einer funktionierenden, natürlicheren Mensch-Maschine-Schnittstelle [[Wexelblat \(1995\)](#)] ist kein neues Forschungsgebiet, sie ist derzeit leider immer noch mit dem Erwerb teurer Spezialhardware verbunden, wie beispielsweise dem System ARTrack [[artrack \(2009\)](#)] im GamecityLab der HAW Hamburg [[GamecityLab \(2009\)](#)].

Die Frage, ob sich dies auch mit preiswerter Hardware realisieren lässt, ist Teil meines Masterprojektes. Diese Arbeit bietet einen Überblick über das verwendete Verfahren zur Bildsegmentierung im Kontext des Living Place Hamburg und einen Vergleich mit anderen Ansätzen.

1.2 Zielsetzung des Masterprojektes

- Segmentierung in Vorder- und Hintergrund
- Erkennen von Bewegung
- Positionsbestimmung in einer zweidimensionalen Ebene (Raumplan).

Ein Weitergehendes Ziel ist die Erkennung von einfachen, dynamischen Gesten durch Bewegungen des Körpers bzw der Extremitäten. Dies ist jedoch nicht Teil des aktuellen Masterprojektes, sondern dieses bildet die Grundlage.

Die folgenden Rahmenbedingungen sind für dieses Projekt gegeben:

- Günstiger Gesamtpreis (Consumerprodukte, keine Spezialhardware)
- Statische Kamerakonfiguration
- Flexible Raumgröße (skalierbar)
- Keine Verwendung von Markern
- Nutzung innerhalb des Living Place Hamburg (Indoor)

2 Technologien

Im Bereich Computer Vision ist eine große Anzahl unterschiedlicher Verfahren veröffentlicht worden, um Bewegungen zu erfassen.

Dabei gibt es unterschiedliche Ansätze. Zum einen, die Erfassung der Objekte (Körper) mit Hilfe von Verfahren wie Mustererkennung, zum anderen die Segmentierung in Vorder- und Hintergrund, also nach Objekten, die den Hintergrund einer Szenerie und solchen die den (bewegten) Vordergrund bilden. Ein Ziel des Masterprojektes ist es, bewegte Objekte im Raum zu erkennen, dazu ist eine Differenzierung nach den Kriterien 'statisch' und 'bewegt' nötig. Die Segmentierung anhand von Bewegung und Vorder-/Hintergrund bietet sich also als Verfahren an.

Dies erfolgt durch Erkennen von Differenzen zwischen verschiedenen Bildern eines Videostroms. Diese Verfahren kommen in den Bereichen Motion Tracking/Detection, Videokomprimierung und Überwachungstechnik zum Einsatz.

Umfassendere Verfahren zur Muster- oder Objekterkennung sind für das angestrebte Endziel einer Erkennung von grober Gestik nicht nötig und würden den Rechenaufwand somit unnötig erhöhen.

2.1 Verfahren zur Differenzbildung

Zur Differenzbildung werden die Videobilder, sofern sie in anderem Format vorliegen, in Graustufenbilder umgewandelt. Im folgenden wird ein Überblick über vorhandene Verfahren zur Bildung dieser Differenzen gegeben.

Fortlaufende Differenz/Temporal Differencing

Bei diesem, in Grundzügen schon lange bekanntem Verfahren [Seyler (1963)] werden je zwei Frames fortlaufend voneinander subtrahiert. Dazu werden die Pixel vom vorherigen Bild (src2) einzeln vom aktuellen Bild (src1) subtrahiert ($dst(I)[c] = abs(src1(I)[c] - src2(I)[c])$), man erhält das korrespondierende Pixel des Differenzbildes (dst). So entsteht ein "Geisterbild" der Differenz, das die Bewegungen beinhaltet.

Als Erweiterung kann eine sofortige Umwandlung in ein Binärbild erfolgen, wenn dabei ein bestimmter (Schwell-) Wert des Pixels erreicht wird, also $dst(I)[c] > Schwellwert$ in diesem Falle wird der Wert das mögliche Maximum des Graustufenbereiches gesetzt.

Eine Erkennung von statischen Gesten (Verharren in der Bewegung) ist nicht möglich, auch werden im Gegensatz zur statischen Hintergrundsubtraktion nur die Konturen erkannt, je nach Framerate zudem unscharf. Das Verfahren der Differenzbildung kam an der HAW im Rahmen des Ambient Awareness Projektes [aawareness (2009)], einem Gemeinschaftliche Kunstprojektes der Departments Design und Informatik zum Einsatz [Pressburger (2009b)], [Pressburger (2009a)].

Dieses Verfahren erzeugt Konturbilder, da die Bewegung hauptsächlich am Rande des Objektes wahrgenommen wird.

Die einfache, fortlaufende Subtraktion ist anfällig gegenüber Störungen in Form von Reflektionen, Schatten, Bewegungen von Gegenständen, größere Bewegungen im Hintergrund, wie Blätter, Erschütterungen der Kamera. Dafür ist sie relativ unempfindlich gegen Helligkeitsveränderungen und langsame Veränderungen des Hintergrundes.

Das Verfahren ist insofern nicht (selbst-) adaptiv, als das aus dem Verfahren keine Parameter zur (Selbst-) Kalibrierung gewonnen werden, durch die Natur des Verfahrens erfolgt jedoch eine ständige Anpassung.

Statische Hintergrundsubtraktion

Bei der statischen Hintergrundsubtraktion [Nagel (1978)] erfolgt eine Differenzbildung zu einem Referenz-Hintergrund. Dabei handelt es sich im einfachsten Falle um ein statisches Hintergrundbild, das zu Beginn ohne Personen aufgenommen wird. Durch Subtraktion auf Pixel-Ebene wie bei der Differenzbildung, ergibt sich ein präzises Bild der Veränderung zur Referenz. Hintergrundsubtraktion bietet die Möglichkeit, Objekte im Vordergrund auch in den Phasen ohne Bewegung zu erfassen.

Die Statische Hintergrundsubtraktion ist, als nicht-adaptives Verfahren, sehr anfällig für Veränderungen des Hintergrundes wie Lichtveränderungen, Verrücken von Gegenständen, Bäumen,

etc. Aus diesem Grund wurden zahlreiche adaptive Verfahren entwickelt, um entweder das Hintergrundbild durch ein Modell zu ersetzen, oder zumindest den Hintergrund adaptiv anzupassen.

Dieses Verfahren ist als Referenz im Bereich der Verfahren zur Hintergrundsubtraktion zu sehen, da es die einfachste Form derselbigen darstellt.

Dynamische Hintergrundsubtraktion

Bei der dynamischen Hintergrundsubtraktion erfolgt ein Update des Hintergrundbildes nach definierten Kriterien. Das Hintergrundbild besteht dabei immer noch aus einem wirklichen Bild, in Form von Pixeln. Für den Algorithmus zum Update des Hintergrundes gibt es eine Reihe von Vorschlägen, beispielsweise die Aufteilung des Hintergrundbildes in Sektoren [Jing u. a. (2005)]. Nach Bildung des Differenzbildes wird für jeden Sektor entschieden, ob es Veränderungen in diesem Sektor gab, und wie viele Pixel geändert wurden. Ist die Anzahl unter einem Schwellwert, werden die geänderten Pixel in das Hintergrundbild übernommen. Dies ermöglicht eine (adaptive) Anpassung an langsame Veränderungen im Hintergrund.

Subtraktion mit Modellbildung

Eine Erweiterung zur klassischen Subtraktion von einem, als Hintergrund definiertem Bild, ist die Verwendung eines Modells. Dies ermöglicht eine größere Berücksichtigung von sich ändernden Pixeln, und ihrer Varianz, beispielsweise bei sich bewegenden Blättern im Bildbereich. Im Laufe der Jahre wurde die Erweiterung des ursprünglichen Ansatzes dadurch weiterentwickelt, bis hin zu Verfahren, die (unterschiedliche) Modelle für den Vorder- sowie Hintergrund zugrunde liegen. Eine Erweiterung ist ebenso die Erstellung eines Modells der Objekte im Vordergrund.

Beispiele für diese Verfahren sind:

- Single Gaussian [Wren u. a. (1997)]
- Mixed Gaussians [Stauffer u. Grimson (1999)]
- Kernel Density Estimator [Elgammal u. a. (2000)]
- Eigenbackgrounds [Rymel u. a. (2004)]

Eine detaillierte Betrachtung der verschiedenen Verfahren würde den Rahmen dieser Arbeit sprengen, es sind allerdings Vergleiche einiger Verfahren verfügbar [El Baf u. a. (2007)], [Piccardi (2004)].

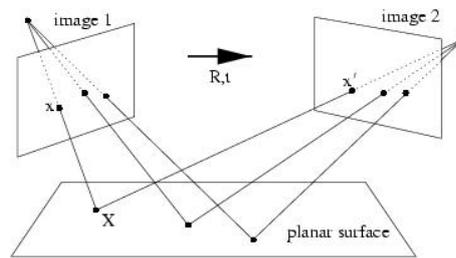


Abbildung 1: 2D Triangulation , Quelle: [Hartley u. Zisserman (2000)]

Allen diesen Beispielen ist gemeinsam das sie die verwendeten Funktionen verwenden, um die Pixel des Hintergrundes zu modellieren. Sie reagieren damit adaptiv auf Veränderungen. Dazu ist ein hoher Aufwand an Berechnungen nötig. Die Verfahren sind bereits seit Ende der 90er Jahre, bzw. Anfang dieses Jahrtausends im Einsatz. Mixed Gaussians ist dabei als Stand der Technik anzusehen, erste theoretische Arbeiten existieren seit 1997. Neuere Entwicklungen in diesem Bereich zielen insbesondere auf die Verbesserung der Performance und der Erkennungsrate[Wang u. Suter (2005)].

2.2 Bewegungserkennung

Zur direkten Erkennung von Bewegung ist nur das Verfahren der Fortlaufenden Differenz geeignet, alle Verfahren der Hintergrundsubstitution benötigen dafür eine weitergehende Beobachtung der Objekte im (erkannten) Vordergrund.

Ein mögliches Verfahren hierfür ist Optical Flow[Horn u. Schunck (1981)] zur Bestimmung von Bewegungsvektoren.

2.3 Verwendetes Verfahren

Der zum Einsatz kommende Ansatz der Bewegungserkennung basiert auf der fortlaufenden Differenzbildung mit Umwandlung in Binärbilder über den Schwellwert.

Die so gewonnenen binären Differenzbilder werden mit Hilfe der morphologischen Verfahren Erosion und Dilatation von Rauschen bereinigt. Danach wird die Konturerkennung nach Suzuki ([Suzuki u. Be (1985)]) verwendet. Über die Berechnung der maximalen Entfernung zwischen den einzelnen Konturpunkten werden diese zusammengefasst und der Mittelpunkt der einzelnen Bereiche bestimmt. Im letzten Schritt erfolgt mit Hilfe von einfacher Triangulation in der X-Ebene der Punkte die Bestimmung von Bewegungs-Bereichen im Raum(siehe 1).

2.4 Hardware

Als Alternative zu den vorgestellten Verfahren in Software-Algorithmen besteht die Möglichkeit, spezielle Hardware zu verwenden, die eine Tiefen-/Raum- Information bietet, und darüber die Erkennung von Vorder- und Hintergrund.

3DV Z-cam bzw. Microsoft Xbox Kamera

Das von Microsoft für die Xbox-Spielekonsole vorgestellte Kamerakzept[Natal (2009)] zur 3D-Bewegungserkennung basiert auf der Z-cam Kamera der Firma 3Dv[3dvsystems (2009)] aus Israel. Das System ist als Hybrid-Kamera ausgeführt, sie enthält sowohl eine normale Webcam als auch eine spezielle Kamera für die Aufnahme von gepulstem Infrarot. Durch diese Kombination wird sowohl eine (2D-)Aufnahme im Normallicht-Bereich als auch eine Tiefenkarte im Infrarot-Bereich erzeugt, die per USB auf den Rechner übertragen werden. Die zugrunde liegende Technik wurde von Yahav, G.J. Iddan , D. Mandelbaum publiziert[Yahav u. a. (2007)].

Stereo Kamera Systeme

Dabei werden zwei Kameras verwendet, um durch die unterschiedliche Perspektive durch geometrische Verfahren die Entfernung eines aufgenommenen Objektes bestimmen zu können. [Yamamoto u. a. (2004)]. Je nach verwendeter Hardware kann diese Berechnung auch im Kamerasystem selber erfolgen. Ein Beispiel für diese Art Kamera ist das Surveyor Stereo Vision System für autonome Roboter [Corporation (2009)].

CMOS-Kamera mit Vordergrunderkennung

Eine weitere Möglichkeit ist die Entwicklung einer Kamera, die die gewünschte Segmentierung anhand von Bewegung selber vornimmt [Verdant u. a. (2007)]. Das vorgeschlagene Konzept kommt allerdings bisher nicht (kommerziell) zum Einsatz.

2.5 Software Frameworks

Anstatt die grundsätzlichen Algorithmen, die für die Verarbeitung nötig sind, selber zu implementieren, ist die Benutzung vorhandener Bibliotheken sinnvoll. Es gibt eine Reihe von qualitativ guten Frameworks im Bereich Bildverarbeitung, die frei verfügbar sind. Dies ist im Hinblick auf die Zielsetzung das System möglichst kostengünstig zu halten, ein Vorteil.

Anforderungen

Das verwendete Framework sollte die folgenden Anforderungen erfüllen:

- Frei verfügbar (Opensource)
- hohe Geschwindigkeit
- große Bandbreite an verfügbaren und relevanten Algorithmen
- Plattformübergreifend
- leichte Austauschbarkeit der Bildquelle
- Umfangreiche Dokumentation

OpenCV

OpenCV [[OpenCV \(2009\)](#)] fungiert im Bereich der Opensource-Bibliotheken zur Bildverarbeitung als Quasi-Standard und wird von zahlreichen Projekten [[Lochmatter u. a. \(2008\)](#)][[stanley \(2009\)](#)],[[vssn \(2005\)](#)] verwendet. OpenCV ist eine ursprünglich von der Firma Intel entwickelte Bibliothek zur Umsetzung zahlreicher Algorithmen aus dem Bereich Bildverarbeitung. Sie ist unter der BSD Lizenz sowohl für Windows, Linux sowie Mac OS X erhältlich. Als umfangreiche Dokumentation zu OpenCV steht ein Tutorial als Buch [[Bradski u. Kaehler \(2008\)](#)] zur Verfügung.

Als weitere Möglichkeiten zur Steigerung der Verarbeitungsgeschwindigkeit bieten sich zwei Wege an, zum einen die Verwendung der kommerziell erhältlichen IPP-Bibliothek der Firma Intel (Erwartete Leistungssteigerung bis zu 50% (siehe 2), als auch die alternative Umsetzung GpuCV.

GpuCV

GpuCV ist eine (partielle) Portierung des OpenCV Frameworks auf die GPU von Grafikkarten. Dieses ermöglicht die Ausführung der Algorithmen auf Systemen mit einer OpenGL-Schnittstelle. [[GpuCV \(2009\)](#)]. Dafür können Grafikkarten der Firmen ATI (AMD) und Nvidia zum Einsatz kommen, und dies auch im Verbund mit mehreren Karten. Eine praktische Beurteilung der Bibliothek für das Projekt steht noch aus, insgesamt sind die Ergebnisse im Vergleich zur nativen OpenCV allerdings vielversprechend [[Farrugia u. a. \(2006\)](#)]. Für die Funktionen, die GpuCV nicht nativ implementiert, wird OpenCV verwendet. Die API der Bibliothek ist vollständig kompatibel zu OpenCV.

IVT

Das Integrating Vision Toolkit (IVT [[ivtlib \(2009\)](#)]) ist ein am Institut für Technische Informatik der Universität Karlsruhe entwickelte Bibliothek für die Sprache C++. Sie ist im großen und ganzen mit der OpenCV Bibliothek vergleichbar, und verwendet diese auch, für Algorithmen, die sie nicht selber implementiert. Die Bibliothek steht unter der GNU General Public License (GPL). Nach eigenen Angaben des Projektes sind die selbst implementierten Algorithmen um den Faktor zwei bis acht schneller verglichen mit OpenCV. Grundlegend unterschiedlich ist die einheitliche Schnittstelle zur Einbindung von Bildquellen, im Gegensatz zur OpenCV ist das Medium (Kamera, Video) transparent. Die empfohlene Dokumentation ist, wie bei der OpenCV Bibliothek, in Buchform erschienen [[Azad u. a. \(2007\)](#)].

LTI

Die LTI-Lib [[ltilib \(2009\)](#)] ist eine am Lehrstuhl für Technische Informatik der RWTH-Aachen entwickelte Bibliothek für C++. Sie steht unter der GNU Lesser General Public License zum Download zur Verfügung und läuft auf den Plattformen Windows und Linux. Der Umfang der Algorithmen ist deutlich geringer als der von OpenCV bzw IVT, auch scheint eine Weiterentwicklung nicht mehr fortgesetzt zu werden (Die letzte Version 1.9 ist von 2005).

VXL

Die VXL Bibliothek [[vxllib \(2009\)](#)] für C++ ist für fast alle derzeit erhältlichen Plattformen (neben Windows, Linux und MacOS-X auch andere UNIX-Derivate) unter einer BSD Lizenz nutzbar. Entwickelt wurde die Bibliothek unter anderem von Robotic Lab der Universität von Oxford. Aktuell ist die Version 1.12 vom Januar 2009. Der Funktionsumfang entspricht in etwa dem von OpenCV, die Dokumentation ist deutlich geringer und unübersichtlicher.

2.6 Vergleich der Performance

Die Geschwindigkeit der Verarbeitung ist ein entscheidender Faktor, um die Reaktionszeit des Gesamtsystems, insbesondere im Hinblick auf eine folgende Gestenerkennung und (merk-bare) Reaktion in einem tolerablen Rahmen zu halten.

Ein Vergleich der Frameworks VLX, LTI und OpenCV (siehe [2](#)) zeigt klare Vorteile im Bezug auf die Geschwindigkeit von OpenCV. Die IVT-Bibliothek war nicht Teil dieses Testes, die auf der IVT Homepage angegebenen Geschwindigkeitsvorteile beziehen sich auch auf andere Algorithmen, ein Vergleich ist aufgrund dieser Daten daher nicht möglich.

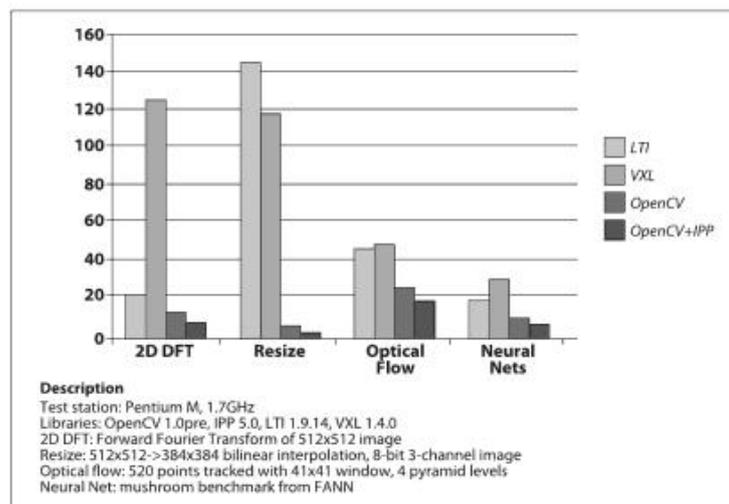


Abbildung 2: Framework Performance , Quelle: [Bradski u. Kaehler (2008)]

3 Bewertung und Einordnung

3.1 Verfahren

Das verwendete Verfahren (2.1) liegt im Bereich zwischen statischer (nicht-adaptiver) Differenzbildung und der dynamischen (adaptiven) Differenzbildung. Im Gegensatz zur statischen Hintergrundsubtraktion werden Veränderungen des Hintergrundes schnell einbezogen, es existiert allerdings kein statisches Modell oder Bild, das eines Updates bedürfte.

Die verwendete Art der Segmentierung per fortlaufender Differenzbildung scheint ein guter Kandidat für die angestrebten Ziele:

- Sie ist performant durch die Einfachheit des Verfahrens.
- Sie ist von der Geschwindigkeit her jeder anderen Form der Differenzbildung mit Modell überlegen, da keine Berechnung der Pixelwerte stattfindet und Bewegungen direkt erkannt werden, ohne das eine weitere Verarbeitung erforderlich ist.
- Im Gegensatz zur statischen Hintergrundsubtraktion ist sie hochdynamisch und wird durch Änderungen des Hintergrundes kaum beeinflusst.
- Es gibt bereits Erfahrungen an der HAW Hamburg mit dieser Technik.
- Sie ist für das geplante Einsatzgebiet, der groben dynamischen Gestenerkennung ausreichend, trotz vorhandener Gefahr der Unschärfe.

- Sie ist in der vorhandenen Umgebung leicht einsetzbar, da es keinen sich überraschend ändernden Hintergrund (zum Beispiel Bäume) gibt.

Ein möglicher Schwachpunkt sind Reflektionen und Schattenwurf, hier ist auch ein möglicher Ansatzpunkt für Erweiterungen.

Ein Vergleich der Verfahren zur Hintergrundsubtraktion [Toyama u. a. (1999)] lässt eine größere Anzahl von falsch-negativen Resultaten, also das "Übersehen" von Bewegungen vermuten. Die Resultate der von mir durchgeführten Versuche legen nahe, dass es sich hierbei um Bewegungen von kleinem Ausschlag oder weiter entfernten Personen handelt. Die Rate der fälschlich erkannten Bewegungen ist gering. Ob dies im Endsystem ein Problem darstellt, wird sich erst endgültig in diesem Kontext zeigen.

Die verwendeten Verfahren der Rauschunterdrückung sowie Konturfindung und Projektion auf einen zweidimensionalen Raumplan sind von einem Wechsel auf ein anderes Verfahren zur Differenzbildung nicht betroffen, sollte sich die Notwendigkeit ergeben.

3.2 Hardware

Die derzeit erhältliche Hardware, die eine (hardwareseitige) räumliche Segmentierung in Vorder- und Hintergrund ermöglicht, sind Stereo-Kameras. Im Consumerbereich sind diese nicht verfügbar. Die Z-cam von 3DV würde dem Einsatzgebiet anscheinend gut entsprechen, sie ist allerdings, ebenso wie die CMOS Variante, nicht im Markt. Somit fallen die Hardware Systeme als Alternativen aus.

3.3 Framework

OpenCV ist gut etabliert, da die Bibliothek für das verwendete Verfahren zudem gut einsetzbar ist, und Möglichkeiten zur weiteren Effizienzsteigerung (GpuGV und IPP) bietet, erscheint mir diese Auswahl richtig. Durch die geschilderten Möglichkeiten ist eine weitere Leistungsreserve im Bereich der Performance gegeben.

Derzeit einziger Kritikpunkt ist die nichtvorhandene Transparenz in der Einbindung unterschiedlicher Quellen (Livekamera vs. Video).

3.4 Ausblick

Mögliche Erweiterungen des Verfahrens sind die Einbeziehung von Farbinformationen (Bewegt sich dort ein Mensch, ein Gegenstand oder ein Schatten ?) sowie die Verwendung eines Modells, das die Historie der Bewegung berücksichtigt. Dies müsste dann allerdings auf höherer Abstraktionsebene erfolgen und nicht auf Hintergrund-/Modellebene, um den Geschwindigkeitsvorteil nicht zu verlieren.

Weitere, alternative Verfahren zur Bewegungserkennung [[Jing u. a. \(2004\)](#)][[Brox u. a. \(5555\)](#)][[Wang u. a. \(2008\)](#)] im dreidimensionalen Raum bedürfen der (praktischen) Evaluierung.

3.5 Fazit

Das derzeit verwendete Verfahren scheint für die Anforderungen geeignet zu sein. Eine endgültige Aussage lässt sich derzeit dazu nicht treffen, da eine letztendliche Evaluierung dieser Eignung nur durch konkrete Messungen und Tests im realen Einsatzgebiet des Living Place möglich sein wird.

Sollte sich herausstellen, dass die gewählte Lösung nicht ausreicht, um die angestrebte Qualität der Segmentierung zu erreichen, stehen einer große Anzahl weiterer Verfahren zur Verfügung. Diese sind allerdings mit höherem Aufwand bei der Berechnung verbunden.

Die verwendete Bibliothek (OpenCV) ist von der Geschwindigkeit der Verarbeitung geeignet, ein genauer, auch praktischer Vergleich mit der IVT-Bibliothek scheint insbesondere auf Hinsicht der einheitlichen Quellen-Schnittstelle und möglicher Performance-Gewinne dennoch sinnvoll.

Das verwendete Verfahren der fortlaufenden Differenz ist, im Kontext der hier erwähnten Verfahren, eines des schnellsten, lediglich die statische Hintergrundsubtraktion ist noch schneller, dieses verlangt aber eine zusätzliche Bewegungserkennung. Alle Modellbasierten Verfahren zur Differenzbildung können nur langsamer sein, da sie zur Bestimmung des Hintergrundpixels erst Berechnungen durchführen müssen.

Abbildungsverzeichnis

1	2D Triangulation , Quelle: [Hartley u. Zisserman (2000)]	7
2	Framework Performance , Quelle: [Bradski u. Kaehler (2008)]	11

Literatur

- [3dvsystems 2009] *3Dvsystems*. Verifiziert am 18.6.2009. <http://www.3dvsystems.com>. Version:2009
- [aawareness 2009] *Ambient Awareness*. Verifiziert am 18.6.2009. <http://www.ambientawareness.org/>. Version:2009
- [artrack 2009] *Advanced Realtime Tracking GmbH*. Verifiziert am 16.6.2009. <http://ar-tracking.eu>Verizifizert:05.01.2008. Version:2009
- [Azad u. a. 2007] AZAD, Pedram ; GOCKEL, Tilo ; DILLMANN, Rüdiger: *Computer Vision Das Praxisbuch*. Elektor, 2007
- [Bianchi-Berthouze u. a. 2007] BIANCHI-BERTHOUCHE, Nadia ; KIM, Whan W. ; PATEL, Darshak: Does Body Movement Engage You More in Digital Game Play? and Why? In: *ACII '07: Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction*. Berlin, Heidelberg : Springer-Verlag, 2007. – ISBN 978–3–540–74888–5, S. 102–113
- [Bradski u. Kaehler 2008] BRADSKI, Gary ; KAEHLER, Adrian: *Learning OpenCV*. O'Reilly, 2008
- [Brox u. a. 5555] BROX, Thomas ; ROSENHAHN, Bodo ; GALL, Juergen ; CREMERS, Daniel: Combined Region- and Motion-Based 3D Tracking of Rigid and Articulated Objects. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 99 (5555), Nr. 1. <http://dx.doi.org/http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.32>. – DOI <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.32>. – ISSN 0162–8828
- [Corporation 2009] CORPORATION, Surveyor: *Surveyor Stereo Vision System*. Verifiziert am 18.6.2009. <http://surveyor-corporation.stores.yahoo.net/srblstca.html>. Version:2009
- [E3 2009] *Project Natal: Microsoft und die Revolution im Wohnzimmer*. Verifiziert am 16.6.2009. <http://www.golem.de/0906/67480.html>. Version:2009
- [El Baf u. a. 2007] EL BAF, F. ; BOUWMANS, T. ; VACHON, B.: Comparison of Background Subtraction Methods for a Multimedia Application, 2007, S. 385–388

- [Elgammal u. a. 2000] ELGAMMAL, Ahmed M. ; HARWOOD, David ; DAVIS, Larry S.: Non-parametric Model for Background Subtraction. In: *ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part II*. London, UK : Springer-Verlag, 2000. – ISBN 3-540-67686-4, S. 751–767
- [Farrugia u. a. 2006] FARRUGIA, J.-P. ; HORAIN, P. ; GUEHENNEUX, E. ; ALUSSE, Y.: GPUVCV: A Framework for Image Processing Acceleration with Graphics Processors. In: *Multimedia and Expo, 2006 IEEE International Conference on*, 2006, S. 585–588
- [GamecityLab 2009] GAMECITYLAB: *Gamecity Lab*. Verifiziert am 18.6.2009. <http://www.gamecitylab.haw-hamburg.de/>. Version: 2009
- [GpuCV 2009] *GpuCV: GPU-accelerated Computer Vision*. Verifiziert am 18.6.2009. <https://picoforge.int-evry.fr/cgi-bin/twiki/view/Gpucv/Web/>. Version: 2009
- [Hartley u. Zisserman 2000] HARTLEY, R. I. ; ZISSERMAN, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000
- [Horn u. Schunck 1981] HORN, Berthold K. ; SCHUNCK, Brian G.: Determining Optical Flow. In: *Artificial Intelligence* (1981), Nr. 17, S. 185–204
- [ivtlib 2009] *Integrating Vision Toolkit*. Verifiziert am 15.6.2009. <http://http://ivt.sourceforge.net/>. Version: 2009
- [Jing u. a. 2005] JING, Guo ; RAJAN, D. ; SIONG, Chng E.: Motion Detection with Adaptive Background and Dynamic Thresholds, 2005, S. 41–45
- [Jing u. a. 2004] JING, Guo ; SIONG, Chng E. ; RAJAN, D.: Foreground motion detection by difference-based spatial temporal entropy image. In: *TENCON 2004. 2004 IEEE Region 10 Conference Bd. A*, 2004, S. 379–382 Vol. 1
- [Lochmatter u. a. 2008] LOCHMATTER, Thomas ; RODUIT, Pierre ; CIANCI, Chris ; CORRELL, Nikolaus ; JACOT, Jacques ; MARTINOLI, Alcherio: SwisTrack - A Flexible Open Source Tracking Software for Multi-Agent Systems. In: *IEEE/RSJ 2008 International Conference on Intelligent Robots and Systems (IROS 2008)*, IEEE, 2008, 4004–4010
- [ltilib 2009] *LTI-Lib*. Verifiziert am 15.6.2009. <http://ltilib.sourceforge.net/doc/homepage/index.shtml>. Version: 2009
- [Nagel 1978] NAGEL, H.H.: Formation of an Object Concept by Analysis of Systematic Time Variations in the Optically Perceptible Environment. 7 (1978), April, Nr. 2, S. 149–194
- [Natal 2009] *Projekt Natal*. Verifiziert am 18.6.2009. <http://www.xbox.com/en-US/live/projectnatal/>. Version: 2009
- [OpenCV 2009] *Welcome - OpenCV Wiki*. Verifiziert am 18.6.2009. <http://opencv.willowgarage.com/wiki/>. Version: 2009

- [Piccardi 2004] PICCARDI, M.: Background subtraction techniques: a review, 2004. – ISSN 1062–922X, S. 3099–3104 vol.4
- [Pressburger 2009a] PRESSBURGER, Julia: *Digital Art Design*. Verifiziert am 18.6.2009. <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master08-09-proj/pressburger/bericht.pdf>. Version:2009
- [Pressburger 2009b] PRESSBURGER, Julia: *Interaktive Kunst*. Verifiziert am 18.6.2009. <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master08-09/pressburger/bericht.pdf>. Version:2009
- [Rymel u. a. 2004] RYMEL, J. ; RENNO, J. ; GREENHILL, D. ; ORWELL, J. ; JONES, G.A.: Adaptive eigen-backgrounds for object detection, 2004. – ISSN 1522–4880, S. 1847–1850 Vol. 3
- [Seyler 1963] SEYLER, A.J.: Real-time recording of television frame difference areas. In: *Proceedings of the IEEE* 51 (1963), March, Nr. 3, S. 478–480. – ISSN 0018–9219
- [spielekonsolen 2009] *Spielekonsolen im Überblick*. Verifiziert am 16.6.2009. <http://www.netzwelt.de/news/74943-spielkonsolen-ueberblick-welche-darf.html>. Version:2009
- [stanley 2009] *Stanford Racing::Home*. Verifiziert am 16.6.2009. <http://cs.stanford.edu/group/roadrunner/old/index.html>. Version:2009
- [Stauffer u. Grimson 1999] STAUFFER, C. ; GRIMSON, W.E.L.: Adaptive background mixture models for real-time tracking, 1999, S. –252 Vol. 2
- [Suzuki u. Be 1985] SUZUKI, S. ; BE, K.: Topological structural analysis of digitized binary images by border following. In: *Computer Vision, Graphics, and Image Processing* 30 (1985), April, Nr. 1, 32–46. [http://dx.doi.org/10.1016/0734-189X\(85\)90016-7](http://dx.doi.org/10.1016/0734-189X(85)90016-7). – DOI 10.1016/0734–189X(85)90016–7. – ISSN 0734189X
- [Toyama u. a. 1999] TOYAMA, K. ; KRUMM, J. ; BRUMITT, B. ; MEYERS, B.: Wallflower: principles and practice of background maintenance. In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* Bd. 1, 1999, S. 255–261 vol.1
- [Verdant u. a. 2007] VERDANT, A. ; DUPRET, A. ; MATHIAS, H. ; VILLARD, P.: Adaptive thresholding for motion detection in a CMOS image sensor. In: *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, 2007. – ISSN 1058–6393, S. 495–499
- [vgcharts 2009] *Hardware Comparison Charts*. Verifiziert am 18.6.2009. www.vgcharts.com. Version:2009
- [vssn 2005] *VSSN - 3rd ACM International Workshop on Video Surveillance and Sensor Networks*. Verifiziert am 18.6.2009. <http://imagelab.ing.unimore.it/vssn05/>. Version:2005
- [vxllib 2009] *C++ Libraries for Computer Vision*. Verifiziert am 15.6.2009. <http://vxlib.sourceforge.net>. Version:2009

- [Wang u. Suter 2005] WANG, Hanzi ; SUTER, D.: A re-evaluation of mixture of Gaussian background modeling [video signal processing applications]. In: *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on* Bd. 2, 2005, ii/1017–ii/1020 Vol. 2
- [Wang u. a. 2008] WANG, J.M. ; CHERNG, S. ; FUH, C.S. ; CHEN, S.W.: Foreground Object Detection Using Two Successive Images, 2008, S. 301–306
- [Wexelblat 1995] WEXELBLAT, Alan: An approach to natural gesture in virtual environments. In: *ACM Trans. Comput.-Hum. Interact.* 2 (1995), Nr. 3, S. 179–200. <http://dx.doi.org/http://doi.acm.org/10.1145/210079.210080>. – DOI <http://doi.acm.org/10.1145/210079.210080>. – ISSN 1073–0516
- [Wren u. a. 1997] WREN, C.R. ; AZARBAYEJANI, A. ; DARRELL, T. ; PENTLAND, A.P.: Pfunder: real-time tracking of the human body. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19 (1997), Jul, Nr. 7, S. 780–785. <http://dx.doi.org/10.1109/34.598236>. – DOI 10.1109/34.598236. – ISSN 0162–8828
- [Yahav u. a. 2007] YAHAV, G. ; IDDAN, G.J. ; MANDELBAUM, D.: 3D imaging Camera for Gaming Application. In: *Consumer Electronics, 2007. ICCE 2007. Digest of Technical Papers. International Conference on* (2007)
- [Yamamoto u. a. 2004] YAMAMOTO, Y. ; YODA, I. ; SAKAUE, K.: Arm-pointing gesture interface using surrounded stereo cameras system, 2004. – ISSN 1051–4651, S. 965–970 Vol.4