



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Ausarbeitung Anwendungen 1

WiSe 2011/12

Torben Woggan

Personenerkennung in Videoaufnahmen

Inhaltsverzeichnis

1	Einleitung	3
1.1	Evakuierungssimulationen	3
1.2	WALK-Projekt	4
1.3	Idee für die Masterarbeit	4
1.4	Ziele von Anwendungen 1.....	5
2	Personenerkennung in Videoaufnahmen	5
2.1	Anwendungsbereiche	5
2.2	Anforderungen der Idee.....	6
2.3	Probleme bei der Erkennung von Personen	6
2.4	Generelles Vorgehen zur Objekterkennung	7
2.5	Ansätze zur besseren Erkennung bei teilweisen Verdeckungen	7
2.6	Vorstellung bestimmter Verfahren bei teilweisen Verdeckungen	7
3	Bewertung des Forschungsstandes	10
3.1	Risiken	10
3.2	Zuverlässigkeit von Personenerkennung	10
4	Fazit	12
5	Literaturverzeichnis	13
6	Abbildungsverzeichnis	15

1 Einleitung

Diese Ausarbeitung befasst sich mit der Erkennung von Personen in Videoaufnahmen durch Algorithmen. Die durch die Personenerkennung gewonnenen Informationen sollen als Basis für eine automatische Validierung von Evakuierungssimulationen dienen.

Im ersten Kapitel wird zuerst der Begriff „Evakuierungssimulation“ verdeutlicht. Danach wird das Projekt vorgestellt, in dessen Rahmen die Masterarbeit und die auf die Masterarbeit hinarbeitenden Kurse, wie auch „Anwendungen 1“, durchgeführt werden. Die Idee für die Masterarbeit wird als nächstes vorgestellt und es wird deutlich gemacht, welche Ziele mit „Anwendungen 1“ verfolgt wurden.

Das zweite Kapitel befasst sich mit der Personenerkennung in Videoaufnahmen durch Algorithmen. Es wird darüber aufgeklärt, in welchen Bereichen sie Anwendung findet und welche Anforderungen die Idee der Masterarbeit an die Personenerkennung stellt. Danach wird das generelle Vorgehen bei der Objekterkennung beschrieben, bevor spezielle Ansätze und Verfahren zur besseren Erkennung von Personen bei leichten bis starken Verdeckungen, die bei Aufnahmen von Evakuierungen häufig auftreten, vorgestellt werden.

Das dritte Kapitel dient der Bewertung des Forschungsstandes. Zuerst wird über die Risiken der vorgestellten Idee aufgeklärt und im Folgenden geklärt, wie verlässlich die aktuellen Personenerkennungsverfahren funktionieren und wie sich dies auf die geplante automatische Validierung mit den durch diese Verfahren gewonnenen Informationen auswirkt.

Im vierten Kapitel wird schließlich ein Fazit zum Forschungsstand und der Machbarkeit der Idee gezogen.

1.1 Evakuierungssimulationen

Evakuierungssimulationen sind eine spezielle Form von Gruppensimulationen, bei denen das Entfluchten, das „In-Sicherheit-Bringen“ von Personen aus einem gefährlichen Bereich, simuliert wird. Sie dienen unter anderem der Bestimmung von Entfluchtungszeiten und Gefahrstellen von baulichen Anlagen und Freiflächen, auf denen der freie und ungehinderte Personenstrom beeinflusst wird (Meyer-König et al. 2009).

Bei Gruppensimulationen handelt es sich um Simulationen, die das kollektive Verhalten von Individuen simulieren. Unterschieden wird vor allem zwischen makroskopischen und mikroskopischen Gruppensimulationen bzw. Evakuierungssimulationen (Thalmann und Musse 2007).

Der makroskopische Ansatz abstrahiert von den Individuen und betrachtet stattdessen die Menschenmenge als eine Flüssigkeit in einer Fluss-Simulation. Individuelles Verhalten der einzelnen Personen wird somit nicht betrachtet, es lassen sich jedoch trotzdem Entfluchtungszeiten abschätzen und Stautellen erkennen.

Eine komplexere und realitätsnähere Betrachtung erlaubt der mikroskopische Ansatz. In diesem Ansatz werden alle simulierten Personen durch Software-Agenten dargestellt. Jeder dieser Agenten hat eine eigene Intelligenz, ein eigenes Verhalten und kann selbstständig Entscheidungen treffen.

Das Verhalten der Agenten in der Simulation lässt sich mit dem Verhalten von Personen in realen Situationen vergleichen. Somit können im Gegensatz zum makroskopischen Ansatz weitere Informationen gewonnen werden, wie das Verhalten der individuellen Agenten in bestimmten Situationen.

1.2 WALK-Projekt

Das WALK-Projekt ist ein Projekt an der HAW Hamburg, welches sich mit der Entwicklung einer mikroskopischen Evakuierungssimulation beschäftigt. Geleitet wird das Projekt von Prof. Dr. Thomas Thiel-Clemen und Prof. Dr. Stefan Sarstedt. Die Besonderheit von WALK ist, dass unter anderem der Einfluss von Emotionen auf das Verhalten der Agenten berücksichtigt wird. Außerdem wird das System so entwickelt, dass es möglich ist, dass dynamische Ereignisse wie Feuer, Rauch und Wassereintrich während der Simulation auftreten können.

Ein wichtiger Punkt bei der Entwicklung von WALK ist die Validität der Simulation. Die Simulation muss die gewünschten Aspekte des simulierten Systems darstellen. Im Speziellen heißt dies in diesem Fall, dass in der Simulation sowohl das gleiche Verhalten auf der makroskopischen Ebene (z.B. Entfluchtungszeiten und Personendichten), als auch auf der mikroskopischen Ebene (z.B. individuelles Verhalten der Agenten bzw. Personen) auftreten muss, das auch in der Realität auftritt. Dies ist Voraussetzung dafür, dass die mit der Simulation gewonnen Erkenntnisse auf die Realität übertragen werden können.

1.3 Idee für die Masterarbeit

Meine Bachelorarbeit umfasste die Erstellung eines Konzeptes zur Validierung von Gruppensimulationen aus einer Kombination von verschiedenen Verfahren und Techniken. Die Idee für die Masterarbeit setzt auf der vorherigen Arbeit auf. Ziel der Masterarbeit ist die Erstellung eines automatischen Verfahrens zur Validierung von mikroskopischen Evakuierungssimulationen mit Hilfe von Personenerkennung in Videoaufnahmen. Dieses Verfahren soll auf die im WALK-Projekt erstellte Simulation angewendet werden.

Aus vorhandenen Videoaufnahmen echter Evakuierungen, also z.B. Aufnahmen von Überwachungskameras, sollen von einem Algorithmus Informationen zu den sich auf dem Bild befindenden Personen extrahiert werden (z.B. Ausrichtungen, Geschwindigkeiten und Positionen der Personen). Hierfür müssen zuerst die Personen auf den Einzelbildern der Aufnahme als Personen erkannt werden. Danach wäre z.B. ein Vergleich der durchschnittlichen Ausrichtungen, Geschwindigkeiten und Positionen der Personen auf der Aufnahme mit den Agenten in einem Ausschnitt der Simulation möglich. Prinzipiell wäre dies auch ohne einen Algorithmus zur Personenerkennung möglich, jedoch würde das Markieren aller Personen „von Hand“ in jedem Einzelbild der Aufnahme einen sehr hohen Aufwand bedeuten.

Ein solcher Vergleich zwischen Aufnahme und Simulation erlaubt eine objektive Validierung. Im Bereich der Gruppensimulationen wird sich oft auf subjektive Experteneinschätzungen der Validität der Simulation verlassen. Eine objektive Validierung erlaubt verlässlichere und vergleichbarere Ergebnisse. Ein Algorithmus, der die Vergleiche durchführt, hätte außerdem noch den Vorteil, dass er z.B. nach jeder Änderung des Simulationscodes oder in einem Nightly Build durchgeführt werden könnte und die Effekte der Änderungen auf die Validität der Simulation gut verfolgbar sind.

Die Entwicklung eines besseren Verfahrens zur Personenerkennung soll nicht Teil der Masterarbeit sein, sondern es sollen soweit möglich bereits vorhandene Verfahren als Werkzeug für die Validierung eingesetzt werden.

1.4 Ziele von Anwendungen 1

„Anwendungen 1“ stellt den ersten Schritt auf dem Weg zur Masterarbeit dar. Im Rahmen des Kurses fand daher eine Einarbeitung in die Grundlagen der Personenerkennung statt. Da gerade bei Videoaufnahmen von Evakuierungen oft ein großes Gedränge herrscht, wurde der Fokus auf die Personenerkennung in Situationen mit leichten bis starken Verdeckungen der Personen gelegt. Des Weiteren sollten mögliche Probleme der Idee der Masterarbeit und die Machbarkeit dieser geklärt werden.

2 Personenerkennung in Videoaufnahmen

Dieses Kapitel gibt eine Einführung in den Bereich der Personenerkennung in Videoaufnahmen. Dazu werden zuerst die bisher am häufigsten vertretenen Anwendungsbereiche der Personenerkennung genannt. Außerdem wird erläutert, welche Anforderungen die automatische Validierung mit Hilfe der Personenerkennung in Videoaufnahmen an die Erkennungsalgorithmen stellt. Danach wird auf die speziellen Probleme, die bei der Erkennung von Personen auftreten, eingegangen. Hiernach wird zuerst das generelle Vorgehen bei der Erkennung von Objekten skizziert und anschließend spezielle Ansätze für die Erkennung von teilweise verdeckten Personen erläutert. Zuletzt erfolgt eine Übersicht über verschiedene Verfahren, die diese Ansätze umgesetzt haben.

2.1 Anwendungsbereiche

Die Erkennung von Personen in Videoaufnahmen ist in sehr unterschiedlichen Bereichen auffindbar. Diese Bereiche umfassen unter anderem den elektronischen Unterhaltungsbereich, die Videoüberwachung und die Fahrzeugsicherheit (Dollár et al. 2011).

Im Bereich der elektronischen Unterhaltung findet sich Personenerkennung unter anderem bei Zusatzhardware („Kameras“) für Spielekonsolen wieder. So ermöglichte „EyeToy“ von Sony bereits im Jahre 2003 die Steuerung von PlayStation-2-Spielen durch den eigenen Körper (Wikipedia 2011). Eine ähnliche Steuerung erlaubt „Kinect“ von Microsoft aus dem Jahre 2010 für die XBOX 360. Dieses bietet nicht nur eine höhere Auflösung als „EyeToy“, sondern ersetzt die Kamera unter anderem durch eine Tiefenkamera, wodurch eine genauere Personenerkennung und Steuerung ermöglicht wird (Wikipedia 2012).

In der Überwachungstechnik kann die Personenerkennung z.B. bei der Überwachung gesperrter Bereiche eingesetzt werden. Wird eine Person in einem solchen Bereich erkannt, könnte ein Alarm ausgelöst werden oder ein Wachmann verständigt werden. Personenerkennung kann auch bei der Altenpflege eingesetzt werden. Hier können unter anderem Situationen erkannt werden, in denen die überwachte Person Hilfe benötigt, wie bei einem Sturz.

Viel Forschung findet im Bereich der Fahrzeugsicherheit statt. In diesem Bereich dient die Personenerkennung dem Erkennen von Fußgängern. Hierdurch kann der Fahrer vor diesen gewarnt werden, wenn er sie nicht rechtzeitig bemerkt oder ein Bremsvorgang kann vorbereitet werden.

2.2 Anforderungen der Idee

Damit eine automatische Validierung der Simulation mit Videoaufnahmen erfolgen kann, muss ein Mindestmaß an Anforderungen durch die Personenerkennung erfüllt werden.

Die allgemeinste Anforderung ist, dass Personen, die auf dem aufgenommenen Bild sichtbar sind, auch wirklich erkannt werden. Des Weiteren ist es erforderlich, dass die Positionen der erkannten Personen bestimmbar sind. Hierdurch lässt sich bereits eine einfache Validierung auf der Makroebene durchführen, indem die Anzahl an Agenten in einem bestimmten Bereich mit der Anzahl an Personen in dem korrespondierenden Bereich der Aufnahme verglichen wird.

Eine genauere Validierung wäre möglich, wenn Personen auch zwischen verschiedenen Einzelbildern der Aufnahme wiedererkannt werden können. Hierdurch ließen sich die Geschwindigkeiten der Personen bestimmen. Die Wiedererkennung von Personen zwischen verschiedenen Kameras könnte jedoch ein Problem darstellen, da sich die Aufnahmen der Kameras unter anderem in der Perspektive unterscheiden können. Eine weitere optionale Anforderung wäre die Bestimmung der Ausrichtung der Personen. Durch diese beiden weiteren Anforderungen ließe sich eine genauere Validierung auf der Makroebene durchführen, indem unter anderem die durchschnittlichen Ausrichtungen, Geschwindigkeiten und Positionen der Agenten der Simulation in bestimmten Bereichen mit den jeweiligen Werten der Personen der Aufnahme im korrespondierenden Bereich verglichen werden würden.

Die Personenerkennung sollte möglichst hohe Erkennungsraten und niedrige Falsch-Positiv-Raten haben. Dies muss auch bei leichten bis schweren Verdeckungen der Personen in der Aufnahme gelten. Nur dadurch ist eine möglichst genaue Validierung möglich.

Unwichtig ist die Geschwindigkeit des Algorithmus, da die Personenerkennung, die Bestimmung von Positionen und die Berechnung von Werten wie Geschwindigkeiten nicht in Echtzeit erfolgen müssen.

Zu beachten ist, dass nur 2D-Aufnahmen von Überwachungskameras zur Verfügung stehen. Viele moderne Verfahren setzen auf Tiefeninformationen z.B. durch Nutzung von Stereokameras. Diese Verfahren sind jedoch in diesem Fall nicht anwendbar. Außerdem ist zu beachten, dass die Qualität der Aufnahmen schlecht sein kann. So können sie unter anderem eine niedrige Auflösung und schlechte Kontraste besitzen und nur aus Graustufenbildern bestehen.

2.3 Probleme bei der Erkennung von Personen

Bei der Erkennung von Personen auf Videoaufnahmen gibt es eine Reihe von Problemen, die die Erkennung erschweren. Eines der Probleme ist die große Vielfalt im Aussehen von Personen. Personen können sich z.B. in der Größe und Statur unterscheiden. Ihr Aussehen kann außerdem unter anderem durch Kleidung, Taschen und Rucksäcke verändert werden. Auch die Körperhaltung der Person, die Rotation der Person zur Kamera und der Kamerawinkel haben einen Einfluss auf das Aussehen der Person in der zweidimensionalen Videoaufnahme. Beispielsweise unterscheidet sich der Körperumriss einer laufenden oder liegenden Person stark von dem Körperumriss einer stehenden Person.

Ein weiteres Problem ist durch Verdeckungen gegeben. Da es sich um Aufnahmen handelt, auf denen große Menschenmengen zu sehen sein können, kann es zu viele Verdeckungen von Personen durch andere Personen kommen. Personen können teilweise oder sogar komplett verdeckt werden. Auch Objekte der Umgebung können Personen auf der Aufnahme verdecken. Sogar die Person selbst kann die Personenerkennung erschweren, wenn sie sich selbst verdeckt. Dies könnte z.B. durch das Verdecken des Kopfes durch den Arm erfolgen.

Es gibt also eine Vielzahl an Problemen, die die Personenerkennung erschweren. Da die Verdeckung von Personen in dem behandelten Szenario ein häufig auftretendes Phänomen ist und einen hohen Einfluss auf die Erkennungsrate der Personen hat, wird der Schwerpunkt bei den Verfahren zur Erkennung von Personen auf jene gelegt, die hierfür spezialisiert sind.

2.4 Generelles Vorgehen zur Objekterkennung

Das generelle Vorgehen zur Objekterkennung oder auch in diesem Fall der Personenerkennung, läuft meistens nach dem gleichen Schema ab. Zuerst findet eine Trainingsphase mit positiven und negativen Beispielbildern statt. Mit diesen Bildern wird maschinelles Lernen auf Basis eines neuronalen Netzwerkes, einer Support Vector Machine oder AdaBoost (adaptive boosting, Algorithmus für maschinelles Lernen) durchgeführt, wodurch ein Klassifikator entsteht. Dieser Klassifikator kann nun zum Suchen von Treffern in Einzelbildern des Videos verwendet werden. Hierzu wird ein Fenster (sliding window) über das Bild geschoben und jeweils mit Hilfe des Klassifikators entschieden, ob es sich um das gesuchte Objekt handelt oder nicht. Dies passiert für verschiedenen Größen und Ausrichtungen des Fensters (Leibe 2008).

2.5 Ansätze zur besseren Erkennung bei teilweisen Verdeckungen

Das Problem der leichten bis starken Verdeckung von Personen bedarf eines speziellen Ansatzes zur Erkennung der Personen. Personen sollen auch dann erkannt werden, wenn diese nur teilweise sichtbar sind. Die Grundidee der verschiedenen Verfahren ist dabei meistens die gleiche.

Da bei der Verdeckung des Körpers dieser möglicherweise nicht mehr als Körper eines Menschen erkannt wird, muss der Algorithmus verschiedene kleinere Teile des Körpers erkennen können. Aus erkannten Körperteilen oder Formen lässt sich somit weiterhin auf eine Person schließen, auch wenn diese teilweise verdeckt ist.

2.6 Vorstellung bestimmter Verfahren bei teilweisen Verdeckungen

Ein Verfahren, das den körperteilbasierten Ansatz zur Erkennung von Personen nutzt, ist das Verfahren von Rao et al. (2008). Bei diesem Verfahren wurden für einzelne explizit definierte Körperteile Klassifikatoren gelernt. Insgesamt wurden sechs verschiedene Klassifikatoren trainiert. Ein Klassifikator diente der Erkennung von Kopf und Schultern, einer der Erkennung des Torsos mit den Armen. Jeweils ein Klassifikator diente der Erkennung des linken Oberkörpers inklusive des linken Arms und des rechten Oberkörpers inklusive des rechten Arms. Ein weiterer Klassifikator diente der Erkennung des gesamten Körpers (Abb. 1).



Abbildung 1 - Aufteilung des Körpers (Rao et al. 2008)

Für jeden dieser Klassifikatoren wurden nun acht verschiedene Ausrichtungen gelernt, so dass die Erkennung unabhängig von der Ausrichtung der Person zur Kamera möglich war (Abb. 2).



Abbildung 2 - Verschiedene gelernte Ausrichtungen (Rao et al. 2008)

Das Aussehen der Beine unterliegt beim Gehen starken Änderungen. So können die Beine unter anderem geschlossen sein oder weit voneinander entfernt sein. Um dieses Problem gesondert zu berücksichtigen, wurden des Weiteren noch drei verschiedene Beinhaltungen gelernt (Abb. 3).

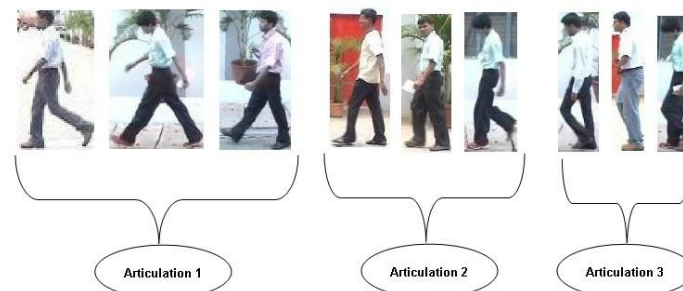


Abbildung 3 - Verschiedene gelernte Beinhaltungen (Rao et al. 2008)

Der Ablauf der Erkennung einer Person beginnt nun mit der Erkennung von Körperteilen durch die erstellten Klassifikatoren. Da die räumliche Verteilung der einzelnen Körperteile im menschlichen Körper bekannt ist, kann das Zentrum der zugehörigen Person bestimmt werden. Hierzu stimmen die erkannten Körperteile über das Zentrum der Person ab. Die Position mit den meisten Stimmen wird zum Körperzentrum gewählt. Als nächstes wird die Bounding Box der Person bestimmt. Die Bounding Box der Person ist durch die Maxima der Bounding Boxes der erkannten Körperteile begrenzt (Abb. 4).

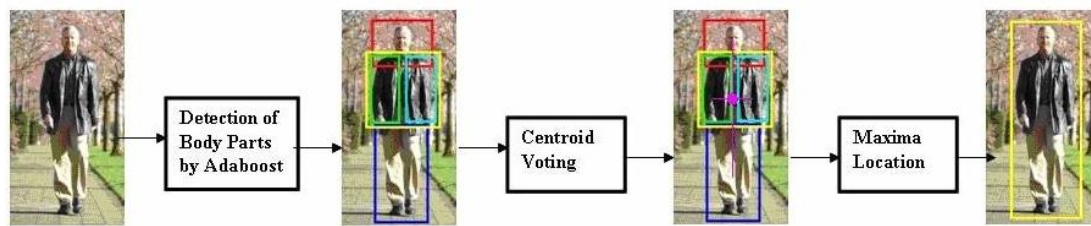


Abbildung 4 - Ablauf der Personenerkennung (Rao et al. 2008)

Ein ähnliches Verfahren ist das Verfahren von Xing et al. (2009). Auch hier fand eine Erkennung von Teilen des Körpers durch Klassifikatoren statt. Jedoch wurde eine andere Aufteilung verwendet. Der Körper wurde in drei Teile aufgeteilt: Kopf und Schulter, Kopf und Torso und ganzer Körper. Dies wurde jeweils für die frontale Ansicht sowie die Ansicht des Körpers von links und von rechts durchgeführt. Die Begründung für diese Aufteilung lag darin, dass alle Klassifikatoren für die jeweiligen Ansichten jeweils gleichzeitig und mit den gleichen Daten trainiert werden können, wodurch Zeit und Aufwand gespart wird (Abb. 5).

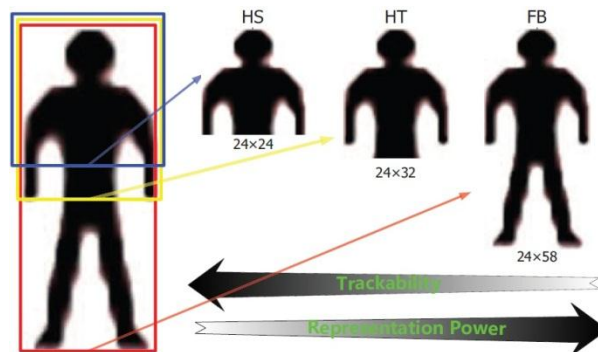


Abbildung 5 - Aufteilung des Körpers (Xing et al. 2009)

Das Verfahren von Rodriguez und Shah (2007) setzt dagegen auf die Erkennung von Formen im menschlichen Umriss. Hierfür wurden zuerst verschiedene Körperhaltungen zur Segmentierung gelernt. Für jede dieser Formen im Körperumriss einer Person wurde außerdem die Position im Bezug zum Zentrum des Körpers gelernt sowie der wahrscheinlichste Körperumriss, zu dem diese Form passen würde. Die erkannten Formen stimmen über einen Mehrheitsentscheid über den Körperumriss und das Zentrum des Körpers ab (Abb. 6).

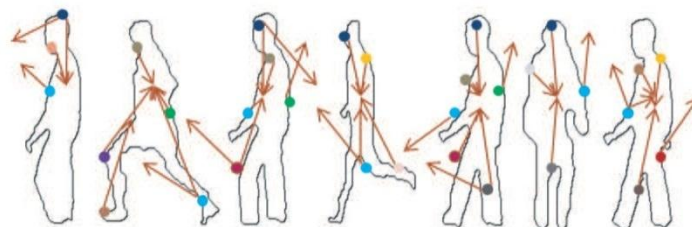


Abbildung 6 - Bestimmung des Zentrums durch Mehrheitsentscheid (Rodriguez und Shah 2007)

Eine verbesserte Erkennung von Personen in Videoaufnahmen ist auch durch Tracking möglich. Das Verfahren von Sherrah (2010) setzt auf das Tracking von sogenannten Interest Points. Dies sind für den menschlichen Körper spezifische Punkte. Ansammlungen solcher Punkte lassen auf eine Person schließen. Wird die Person nun teilweise verdeckt, so kann sie weiterhin erkannt werden, wenn noch zu ihr gehörige Interest Points nicht verdeckt sind (Abb. 7).



Abbildung 7 - Tracking von Interest Points (Sherrah 2010)

3 Bewertung des Forschungsstandes

Dieses Kapitel soll klären, inwieweit eine automatische Validierung mit Hilfe von Personenerkennung in Videoaufnahmen mit dem derzeitigen Forschungsstand machbar ist. Hierzu werden zuerst die Risiken der Idee erläutert. Danach wird auf die Zuverlässigkeit verschiedener aktueller Verfahren zur Personenerkennung unter verschiedenen Rahmenbedingungen eingegangen.

3.1 Risiken

Eine automatische Validierung von Evakuierungssimulationen mit Hilfe von Videoaufnahmen und Personenerkennungsalgorithmen birgt verschiedene Risiken. Das Hauptproblem ist die Zuverlässigkeit der Personenerkennung. Problematisch wären sowohl eine Nicht-Erkennung von zu vielen Personen als auch zu viele Falsch-Positiv-Erkennungen. In beiden Fällen würde die Simulation mit zu stark von der Realität abweichenden Daten verglichen bzw. validiert werden. So ergeben sich durch die beiden Probleme unter anderem falsche Anzahlen von Personen in bestimmten Bereichen und als Folge dessen auch falsche Personendichten. Auch falsche Positionen, Ausrichtungen und Geschwindigkeiten von Personen wären die Folge. Somit würde die Simulation an fehlerhafte Daten angepasst werden. Hierdurch würde trotz oder sogar wegen der Validierung das Verhalten der Simulation von der Realität abweichen. Der Vorteil der höheren Genauigkeit gegenüber einer subjektiven Bewertung der Simulation wäre dadurch nicht mehr gegeben.

Leichte Ungenauigkeiten bei der Personenerkennung können noch tolerierbar sein, da sich durch sie keine großen Unterschiede z.B. in den Anzahlen der Personen in bestimmten Bereichen ergeben. Bei zu großen Ungenauigkeiten wäre das Validierungsverfahren jedoch wegen der genannten Folgen nicht brauchbar. Die Zuverlässigkeit der aktuellen Verfahren zur Personenerkennung muss also zuerst auf ihre Eignung für das hier vorgeschlagene Validierungsverfahren überprüft werden.

3.2 Zuverlässigkeit von Personenerkennung

Eine große Studie zum Stand der Personenerkennung wurde von Dollár et al. (2011) durchgeführt. Sie verglichen die Zuverlässigkeit, genauer gesagt die Erkennungsraten und Fehlerraten, von 16 repräsentativen aktuellen Verfahren zur Personenerkennung. Als Basis für ihre Tests diente ein von ihnen für diesen Zweck erstellter Datensatz, das Caltech Pedestrian Dataset.

Dieser Datensatz wurde aus Videoaufnahmen, die von einem mit einer Kamera ausgestatteten Auto in Bereichen von Los Angeles mit hohen Fußgängerdichten aufgezeichnet wurden, erstellt. Diese wurden nachträglich stabilisiert. Aus diesen Aufnahmen wurden 250.000 Einzelbilder mit einer Auflösung von jeweils 640x480 Pixel ausgesucht, so dass der Datensatz aus 137 Minuten Videoaufzeichnungen besteht. Diese Einzelbilder wurden auf Personen und Menschengruppen untersucht, welche danach mit Bounding Boxes markiert wurden. Insgesamt enthält der Datensatz 2.300 verschiedene Personen und 350.000 Bounding Boxes und ist damit sehr umfangreich.

Durch die Erstellung des Datensatzes konnte auch die Wichtigkeit des Beachtens von Verdeckungen von Personen bekräftigt werden. So ergab sich eine Wahrscheinlichkeit von 29%, dass eine Person auf keinem Einzelbild verdeckt ist. 53% der Personen waren manchmal verdeckt und 19% der Personen waren in jedem Einzelbild verdeckt. 70% der Personen waren in mindestens einem Einzelbild verdeckt.

Die Schwierigkeit des Datensatzes wurde verringert, indem keine Personen mit Bounding Boxes unter 20 Pixel Höhe erkannt werden mussten. Auch von den Rändern der Aufnahme beschnittene Bounding Boxes mussten nicht erkannt werden. Des Weiteren mussten Gruppen von Personen, bei denen es nicht möglich war, Bounding Boxes für die einzelnen Personen zu erstellen, nicht erkannt werden. Es wurde weder als Fehler angesehen, wenn ein Algorithmus diese genannten Bounding Boxes nicht erkannte, noch wenn er sie erkannte.

Betrachtet wurde zuerst für alle 16 Verfahren die Nicht-Erkennungsrate von Personen bei nur einer erlaubten falsch-positiven Erkennung alle zehn Einzelbilder. Durch Inkaufnahme von höheren Falsch-Positiv-Raten lassen sich zwar bessere Erkennungsraten erreichen, jedoch wurde eine höhere Falsch-Positiv-Rate als nicht annehmbar bezeichnet.

Zuerst fand eine Betrachtung der Nicht-Erkennungsrate in Abhängigkeit zur Entfernung und somit der Größe der Personen auf der Aufnahme statt. Auf den betrachteten Aufnahmen gab es keine Verdeckungen der Personen. Selbst bei sehr guten Voraussetzungen, d.h. dichten Personen mit mindestens 80 Pixel Höhe (Abb. 8, links), ergab sich für das beste Verfahren eine Nicht-Erkennungsrate von 22%. Bei mittleren Entfernungen der Personen, d.h. die Personen hatten eine Höhe von 30-80 Pixel, erkannte das beste Verfahren bereits 77% der Personen nicht mehr (Abb. 8, rechts). Bei weiten Entfernungen bei denen die Höhe der Personen unter 30 Pixel betrug, erkannte das beste Verfahren sogar ganze 95% der Personen nicht mehr.

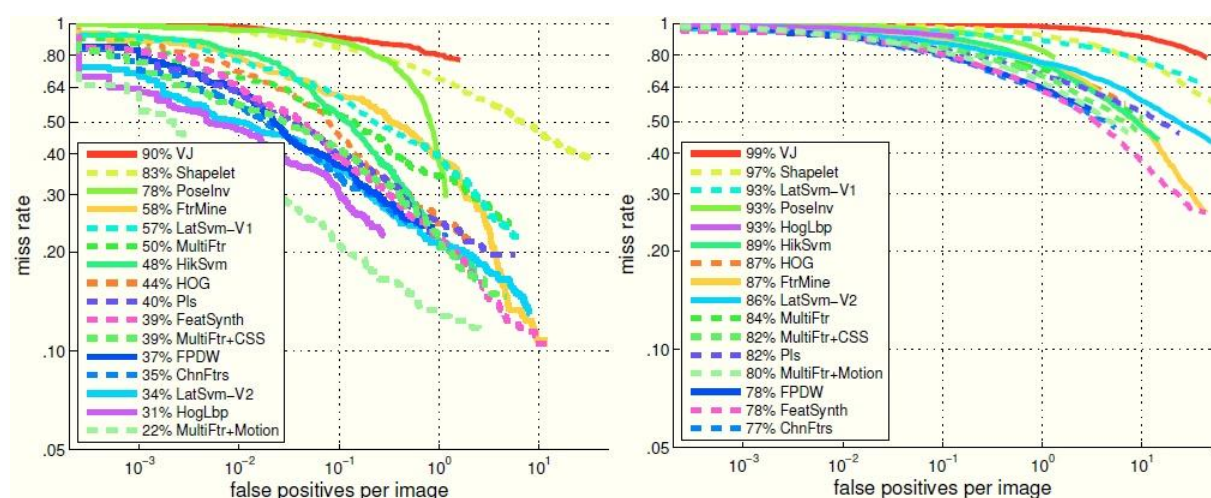


Abbildung 8 - Graph der Nicht-Erkennungsraten in Abhängigkeit zur Falsch-Positiv-Rate der verschiedenen Verfahren.
Links: dichte Entfernung, rechts: mittlere Entfernung (Dollár et al. 2011)

Danach wurde die Abhängigkeit der Zuverlässigkeit der Erkennung von der Schwere der Verdeckungen der Personen bestimmt. Es wurden jeweils Personen in mittlerer bis dichter Entfernung (mindestens 50 Pixel Höhe) betrachtet. Ohne Verdeckungen wurden hierbei 48% der Personen vom besten Verfahren nicht erkannt (Abb. 9, links). Waren die Personen nur leicht verdeckt, d.h. 1% bis 35% des Körpers waren verdeckt, so wurden bereits 73% vom besten Verfahren nicht mehr erkannt (Abb. 9, rechts). Waren 35% bis 80% des Körpers verdeckt, so erkannte das beste Verfahren ganze 93% der Personen nicht mehr.

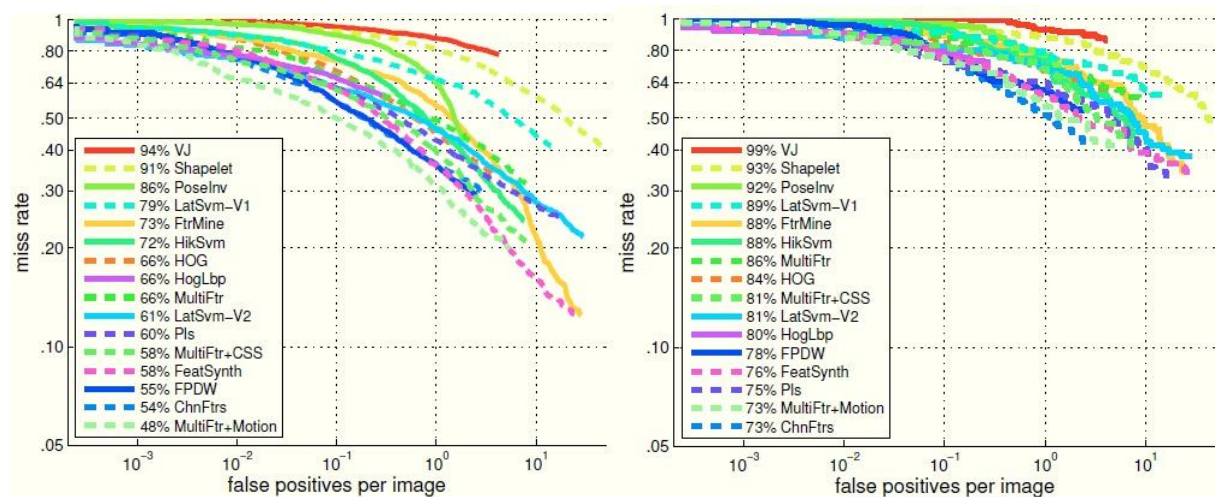


Abbildung 9 - Graph der Nicht-Erkennungsraten in Abhängigkeit zur Falsch-Positiv-Rate der verschiedenen Verfahren.
Links: keine Verdeckungen, rechts: leichte Verdeckungen (Dollár et al. 2011)

4 Fazit

Es wurde gezeigt, dass sich die Personenerkennung noch in einem frühen Entwicklungsstadium befindet. Selbst ohne Verdeckungen, die die Erkennung zusätzlich erschweren, kann das beste Verfahren keine überzeugenden Erkennungsraten liefern. Bessere Erkennungsraten lassen sich nur durch Inkaufnahme höherer Falsch-Positiv-Raten erreichen, welche aber ebenfalls problematisch sind. Es müsste hierfür geklärt werden, bei welcher Nicht-Erkennungsrate und Falsch-Positiv-Rate im betrachteten Szenario die geringsten Abweichungen von der Realität zu verzeichnen sind. Der Algorithmus müsste also richtig auf das Szenario eingestellt werden, so dass er die bestmöglichen Ergebnisse liefert. Ohne Arbeit am Algorithmus wird das Ziel der Validierung mit Hilfe der Personenerkennung in Videoaufnahmen also nicht möglich sein.

Das Szenario, für das die Validierung eingesetzt werden soll, erschwert die Personenerkennung zusätzlich. Überwacht eine Kamera einen großen Bereich, so sind die Personen auf den Bildern der Aufnahme klein und Verdeckungen der Personen sind sehr wahrscheinlich. Es wurde gezeigt, dass die Personenerkennung noch sehr große Probleme mit dieser Kombination hat.

Die Machbarkeit und Sinnhaftigkeit der Idee unter den derzeitigen Bedingungen ist fragwürdig, da sie sehr stark von den Fortschritten in der Personenerkennung abhängig ist. Ohne große Fortschritte in den Erkennungsraten der Personenerkennungsalgorithmen ist die gewünschte Art der Validierung zu ungenau und kann somit keine valide Simulation garantieren.

5 Literaturverzeichnis

Dollár et al. 2011

DOLLÁR, P. ; WOJEK, C. ; SCHIELE, B. ; PERONA, P.: Pedestrian Detection: An Evaluation of the State of the Art. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2011), Nr. 99, S. 1-20. – Online verfügbar unter:

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5975165&isnumber=4359286>

Abruf: 2012-02-28

Leibe 2008

LEIBE, Bastian: *Computer Vision – Lecture 10, Sliding-Window based Object Detection*. Vorlesungsskript, RWTH Aachen, 2008. –

Online verfügbar unter:

<http://www.mmp.rwth-aachen.de/teaching/cvws08/lecture/cv08-part10-categorization1.pdf>

Abruf: 2012-02-28

Meyer-König et al. 2009

MEYER-KÖNIG, Tim ; MOROGE, Christian ; SCHRECKENBERG, Michael ; SCHWENDIMANN, Martin ; WALDAU, Nathalie: *Richtlinie für Mikroskopische Entfluchtungsanalysen : Version: 2.2.1 08. Juni 2009*. RiMEA-Projekt, 2009. –

Online verfügbar unter: <http://www.rimea.de/downloads/richtlinien/r2.2.1.pdf>

Abruf: 2012-02-28

Rao et al. 2008

RAO, Supriya ; PRAMOD, N. C. ; PATURU, Chaitanya Krishna: People detection in image and video data. In: *Proceeding of the 1st ACM workshop on Vision networks for behavior analysis (VNBA '08)*. New York, NY, USA : ACM. S. 85-91. –

Online verfügbar unter: <http://doi.acm.org/10.1145/1461893.1461909>

Abruf: 2012-02-28

Rodriguez und Shah 2007

RODRIGUEZ, Mikel D. ; SHAH, Mubarak: Detecting and segmenting humans in crowded scenes. In: *Proceedings of the 15th international conference on Multimedia (MULTIMEDIA '07)*. New York, NY, USA : ACM, 2007. S. 353-356. –

Online verfügbar unter: <http://doi.acm.org/10.1145/1291233.1291310>

Abruf: 2012-02-28

Sherrah 2010

SHERRAH, Jamie: Occluded Pedestrian Tracking Using Body-Part Tracklets. In: *2010 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. Sydney, NSW, Australien : DICTA 2010 Organisation Committee, 2010. S. 314-319. –

Online verfügbar unter:

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5692582&isnumber=5692215>

Abruf: 2012-02-28

Thalmann und Musse 2007

THALMANN, Daniel ; MUSSE, Soraia R.: *Crowd Simulation*. 1. Aufl. London: Springer-Verlag London Limited, 2007. – ISBN 978-1-84628-824-1

Wikipedia 2011

WIKIPEDIA (Hrsg.): *EyeToy* Stand: 2011-07-03 <http://de.wikipedia.org/wiki/EyeToy>

Abruf: 2012-02-28

Wikipedia 2012

WIKIPEDIA (Hrsg.): *Kinect* Stand: 2012-02-21 <http://de.wikipedia.org/wiki/Kinect>

Abruf: 2012-02-28

Xing et al. 2009

XING, Junliang ; AI, Haizhou ; LAO, Shihong: Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2009*. Miami, FL, USA : IEEE Computer Society, 2009. S. 1200-1207. –

Online verfügbar unter:

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5206745&isnumber=5206488>

Abruf: 2012-02-28

6 Abbildungsverzeichnis

Abbildung 1 - Aufteilung des Körpers (Rao et al. 2008)	8
Abbildung 2 - Verschiedene gelernte Ausrichtungen (Rao et al. 2008)	8
Abbildung 3 - Verschiedene gelernte Beinhaltungen (Rao et al. 2008)	8
Abbildung 4 - Ablauf der Personenerkennung (Rao et al. 2008)	9
Abbildung 5 - Aufteilung des Körpers (Xing et al. 2009)	9
Abbildung 6 - Bestimmung des Zentrums durch Mehrheitsentscheid (Rodriguez und Shah 2007)	9
Abbildung 7 - Tracking von Interest Points (Sherrah 2010)	10
Abbildung 8 - Graph der Nicht-Erkennungsraten in Abhängigkeit zur Falsch-Positiv-Rate der verschiedenen Verfahren. Links: dichte Entfernung, rechts: mittlere Entfernung (Dollár et al. 2011).....	11
Abbildung 9 - Graph der Nicht-Erkennungsraten in Abhängigkeit zur Falsch-Positiv-Rate der verschiedenen Verfahren. Links: keine Verdeckungen, rechts: leichte Verdeckungen (Dollár et al. 2011).....	12