

FREIE UNIVERSITÄT BERLIN

Hintergrundentfernung unter Zuhilfenahme von Hidden Markov Models

Softwareprojekt Mobilkommunikation
Im Rahmen des Projekts SAFEST

Abschlussbericht

Vorgelegt am 3. März 2014 von

Marcin Nawrocki

Patrick Mertes

Hinnerk van Bruinehsen

Betreut durch

Prof. Dr. Katinka Wolter | **Dipl.-Inform. Alexandra Danilkina**



Fachbereich Mathematik und Informatik

Inhaltsverzeichnis

1	Einleitung	2
1.1	Kontext und Motivation	2
1.2	Aufgabenbeschreibung	3
2	Entwurf eines Vordergrund-Extractors	4
2.1	Anforderungsanalyse	4
2.2	Hidden Markov Models	4
2.3	Standardalgorithmen der HMM	5
2.4	Diskrete Kosinus Transformation	6
2.5	Modellierung eines Eingabealphabets	6
2.5.1	DC-Wert	7
2.5.2	AC-Wert	7
2.6	Verfahren zur Hintergrundentfernung	8
3	Implementierung und praktische Details	10
3.1	Videomaterial	10
3.2	C++	11
3.3	OpenCV	11
3.4	CvHMM	11
4	Evaluation	12
4.1	Vergleich mit Histogramm-basierter Implementierung: Innenhof	12
4.2	Vergleich mit Histogramm-basierter Implementierung: Eingang	12
4.3	Vergleich mit Histogramm-basierter Implementierung: Tegel	12
4.4	Diskussion	12
5	Fazit und Ausblick	14

Kapitel 1

Einleitung

Dieses Kapitel soll eine Einführung in das Projekt SAFEST geben und die vorliegende Problemstellung erläutern, welches im Rahmen des Softwareprojektes Mobilkommunikation im Wintersemester 13/14 bearbeitet wurde.

1.1 Kontext und Motivation

SAFEST ist ein deutsch-französisches Projekt mit der Zielsetzung, die Sicherheit an öffentlichen Plätzen und kritischen Infrastrukturen durch Realisierung eines Sensornetzwerkes mit Infrarotkameras zu erhöhen. Um dieses Ziel zu erreichen, bewacht das Sensornetzwerk ein vorgegebenes Areal und verarbeitet die gewonnenen Mittelinfrarotbilder mit Hilfe von Algorithmen in Hinblick auf verschiedene Parameter. Der eigentliche Gewinn an Sicherheit soll dadurch erfolgen, dass automatisiert auf Grundlage der ermittelten Parameter die Dichte an sich in dem Areal aufhaltenden Personen ermittelt wird, woraus wiederum ein Rückschluss auf die Wahrscheinlichkeit des Auftretens einer Massenpanik gezogen werden soll.

Der erste wichtige Schritt bei dieser Analyse ist die automatisierte Erkennung und Zählung von sich im Bildbereich befindlichen Personen. Um diesen Schritt wiederum zu ermöglichen, muss zunächst der Hintergrund vom Vordergrund des Bildes getrennt werden. Hintergrund ist in diesem Zusammenhang als der statische Teil des Bildes zu verstehen, wohingegen der Vordergrund die sich bewegendenden Personen sind. Durch die Verwendung einer Wärmebildkamera wird diese Trennung erst einmal erleichtert, da die relativ warmen Personen heller gefärbt sind als der relativ kalte Boden. Da der Hintergrund aber nicht total statisch ist und der Unterschied der Helligkeit stark von der Temperaturdifferenz zwischen Boden und Personen abhängt, wird die Trennung allerdings erschwert. Ein weiteres Problem ist die ständige Neukalibrierung der verwendeten Kamera. Diese findet statt, wenn sich das Wärmespektrum des beobachteten

Areals stark ndert, so zum Beispiel wenn viele Personen in das Bild ein- oder austreten.

Der nchste Schritt in dieser Analyse ist die automatisierte Bestimmung der auf dem Bild sich befindlichen Personen. Wird diese Analyse erfolgreich ausgefhr, so muss nur noch dieser Wert an eine Kontrolleinheit bertragen werden, nicht aber das eigentliche Bild, was im Hinblick auf den Datenschutz sehr wichtig ist.

Eigentlich fr Objekterkennung bewhrte Algorithmen wie Histogram of oriented gradients und von open-CV vorimplementierte Klassen wie BackgroundSubtractorMOG2 , knnen sich auf die Heterogenitt des Hintergrundes, insbesondere die durch die Kamerakalibrierung entstehende Dynamik, nicht akkurat einstellen. Daher soll ein neuer Algorithmus evaluiert werden, der ber die Fhigkeit verfgt, sich der Dynamik anzupassen, um das Problem zu lsen.

1.2 Aufgabenbeschreibung

Das Ziel unseres Softwareprojektes ist es, ein Verfahren zu entwickeln, welches das Bild einer Infrarotkamera in Vorder- und Hintergrund trennt. Die genauen Anforderungen fr diesen Prozess werden in Kapitel 2. 1 beschrieben. Hierbei soll die allgemeine Verwendbarkeit von Hidden Markov Models(HMM) und Diskreten Kosinus Transformation(DCT). Die theoretischen Grundlagen werden hierzu im Kapitel 2. 2 ff. erlutert. Schlussendlich soll eine vergleichende Evaluation durchgefhr werden, in dem das entwickelte Verfahren gegen ein bereits vorliegendes Histogramm-basiertes Verfahren antreten muss, vergleiche hierzu Kapitel 4.

Ein Algorithmus zum Zhlen von Personen liegt bereits vor und muss innerhalb dieses Projektes nicht mehr entworfen beziehungsweise implementiert werden.

Kapitel 2

Entwurf eines Vordergrund-Extractors

Dieses Kapitel soll den ntigen theoretischen Hintergrund herausarbeiten und gem den Anforderungen unseren Entwurf vorstellen. Es wird das Konzept der hidden markov models und der diskreten Kosinus Transformation erlutert und deren Verknpfung eingefhrt.

2.1 Anforderungsanalyse

Betrachtet man die funktionalen Anforderungen, so ist bereits in der Aufgabenbeschreibung angedeutet, was der Vordergrund-Extractor knnen muss. Es soll demonstriert werden, dass hidden markov models zur Trennung von Vordergrund und Hintergrund bei stark heterogenen Hintergrnden geeignet sind, wie in folgender Abbildung dargestellt:

Die nichtfunktionalen Anforderungen fordern vor allem eine einfache Benutzbarkeit durch Automatisierung. Der Ressourcenbedarf darf nicht zu gro sein - ein Video wird fr gewhnlich mit 25 Frames pro Sekunde aufgezeichnet, im Idealfall ist eine optimale Analyse jedes einzelnen Frames in Echtzeit mglich. Hierbei ist zu beachten, dass eine hohe Korrektheit zu erzielen ist, das heit, dass die Ergebnisse fehlerfrei sein mssen. Eine Flexibilitt bezglich unterschiedlicher Videotypen ist gewnscht, also die Untersttzung von variierenden Videoauflsungen et cetera.

2.2 Hidden Markov Models

Die Markov Modelle stammen aus der Wahrscheinlichkeitstheorie und entsprechen einem stochastischem Zustandsautomaten, bei dem die Zustandswechsel gem einer Wahrscheinlichkeit statt finden und nicht abhngig von der Vergangenheit sind, sondern nur von dem aktuellen Zustand. Bei den hidden markov (HMM) models knnen jene Zustnde nicht beobachtet werden, sondern nur die Beobachtungen/Ausgaben,

welche während dieses Zustandes auftreten, sie werden Emissionen genannt. Die Zustandsbergangswahrscheinlichkeiten sind in den HMM somit nicht die einzigen Parameter, die Emissionswahrscheinlichkeiten bilden die zweiten Parameter.

Formal definiert ist ein HMM ein 5er-Tupel $\lambda = (S, V, A, B, \pi)$ mit:

$S = s_1, \dots, s_n$ sei die Menge aller Zustände $V = v_1, \dots, v_m$ das Alphabet der möglichen Emissionen $A = a_1, \dots, a_n$ sei eine bergangsmatrix zwischen den Zuständen, a_{ij} entspricht der Wahrscheinlichkeit des bergangs von Zustand s_i in Zustand s_j $B = b_1, \dots, b_m$ sei eine Beobachtungsmatrix, wobei $b_i(v_j)$ die Wahrscheinlichkeit angibt, im Zustand s_i die Beobachtung (Emission) v_j zu sehen π die Initialverteilung, π_i sei die Wahrscheinlichkeit, dass s_i der Startzustand ist

Ein HMM heißt zeitinvariant, wenn sich die bergangs- und Emissionswahrscheinlichkeiten nicht mit der Zeit ändern. HMM werden zunehmend in der Literatur zur Sprach-, Schrift und Mustererkennung verwendet, da sie mit den probabilistischen bergangen die Prozesse der echten Welt besser widerspiegeln als deterministische Definitionen, zu dem können die Parameter durch Lernalgorithmen automatisiert bestimmt werden. siehe hierzu Kapitel 2.5.

Die vorliegende Abbildung demonstriert ein HMM mit 3 Zuständen und 4 Emissionen, wobei der bergang aus jedem Zustand zu jedem Zustand und zusätzlich in jedem Zustand jede Emission möglich ist. Es liegt eine Gleichverteilung vor, bergänge zwischen Zuständen haben stets die Wahrscheinlichkeit $\frac{1}{3}$, die Emissionen haben stets eine Wahrscheinlichkeit von $\frac{1}{4}$. Beschriftung der Kanten aus bersichtsründen nicht dargestellt.

2.3 Standardalgorithmen der HMM

Für die HMM wurden bereits viele Algorithmen entwickelt, welche Standardprobleme lösen. Die zwei häufigsten Problemstellungen bei HMMs sind das Lernproblem und das Evaluations-/Decodingproblem.

Das Lernproblem wird durch den Baum-Welch Algorithmus gelöst, welcher dazu verwendet wird, unbekannte Parameter, genauer die bergangs- und Emissionswahrscheinlichkeiten eines HMM, zu bestimmen. Hierbei handelt es sich um einen erwartungsmaximierenden Algorithmus, welcher anhand von gegebenen Trainingssequenzen die Maximum-Likelihood-Schätzwerte berechnet. Die initialen Werte eines HMMs müssen geschützt werden, hier wird für gewöhnlich eine Gleichverteilung angenommen, das heißt das jeder bergang und jede Emission in einem Zustand gleich wahrscheinlich sind.

Das Evaluations-/Decodingproblem wird durch den Forward-Algorithmus beziehungsweise den eng verwandten Viterbi-Algorithmus gelöst. Gegeben sei eine Chronik an k -letzten Emissionen, was ist die Wahrscheinlichkeit für eine bestimmte Emission? Und ferner, was ist die wahrscheinlichste Emission? Der Viterbi-Algorithmus berechnet zur Beantwortung dieser Frage die wahrscheinlichste Zustandssequenz, also eine Sequenz die die Wahrscheinlichkeit der gegebenen Emissionssequenz maximiert, und kann anhand des letzten Zustandes dieser Sequenz das Problem lösen. Der Forward-Algorithmus ist ein Algorithmus aus der dynamischen Programmierung und optimiert im Gegensatz zum Viterbi nicht rückwirkend

die gesamte Zustandssequenz neu, sondern berechnet den aktuell wahrscheinlichsten Zustand auf Basis der zuvor berechneten Zustände und hängt diesen an. Somit ist der Forward Algorithmus zur Laufzeit grundstzlich schneller, jedoch nicht so genau wie der Viterbi-Algorithmus.

2.4 Diskrete Kosinus Transformation

Die diskrete Kosinus Transformation (DCT) ist ein Verfahren, welches zur verlustbehafteten Kompression von Daten verwendet wird, wobei die bekannteste Anwendung das Dateiformat JPEG ist. hnlich zu der diskreten Fourier Transformation wird eine Information mittels Frequenzen reprsentiert, wobei wie der Name bereits nahe legt nur Kosinusfunktionen verwendet werden.

Es existieren mehrere Varianten der DCT. Wird jedoch die DCT auf Bildinformationen angewandt, speziell bei der JPEG-Kompression, so wird die zwei-dimensionale Typ II DCT verwendet, welche die gngigste Variante ist. Die DCT erhlt eine $n \times n$ Matrix und gibt eine Matrix der selben Gröe zurck, wobei im JPEG-Standard 8×8 Pixel Matrizen betrachtet werden. Hierbei wird das Element $[1,1]$ als DC-Wert bezeichnet und bildet den Durchschnittswert der Farben der betrachteten Matrix; die restlichen 63 Werte der Matrix sind ein Offset zu dem DC-Wert und kodieren somit den Unterschied zum DC-Wert innerhalb der Matrix. Diese Werte werden als AC-Werte bezeichnet.

Formal definiert ist die DCT eine lineare, invertierbare Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$, welche N reellwertige Werte aus $x[n]$ in N reellwertige Werte nach $X[k]$ berfhrt:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cos \left[\left(N + \frac{1}{2} \right) k \right] \text{ mit } k = 0, \dots, N-1$$

Die folgende Abbildung zeigt das Ergebnis nach der Anwendung der DCT mit Einfhren der gesamten Blcke in die jeweils berechneten DC-Werte.

2.5 Modellierung eines Eingabealphabets

Das Kapitel 2.2 hat das Konzept der HMM vorgestellt, jedoch nicht nher definiert wie eine Emission aussieht. Das Alphabet der Emissionen wird als Eingabealphabet bezeichnet, da die Symbole des Alphabets dem HMM bergeben werden. Hierbei ist zu beachten, dass das Eingabealphabet endlich und diskret sein muss. Zustzlich sollte das Eingabealphabet aus mglichst wenigen, aussagekrftigen Symbolen bestehen. Die Kodierung eines Eingabealphabets geht daher in unserem Fall zu der Kernleistung bei der Verwendung von HMM.

Die im Kapitel 2.3 vorgestellte DCT wird von uns verwendet, um ein Eingabealphabet zu erzeugen, wobei wir die DCT auf fest definierte Bereiche anwenden. Wir bernehmen die aus der Bildkompression blichen 8×8 Pixelblcke zur Einteilung des vorliegenden Bildes, da wir dieses als optimalen Kompromiss zwischen einer zu hohen und niedrigen Granularitt ansehen: kleinere Betrachtungen, so zum Beispiel

pixelbasierte Verfahren, wren deutlich rechenintensiver und anflliger auf Bildrauschen und wrden daher keine akkuraten Rckschlssse bei Vernderungen des Pixelwertes ermgligen. Bei greren Betrachtungen wren relevante Vernderungen deutlich schwerer wahrzunehmen, speziell im DC-Wert, da dieser stets den Mittelwert aller Subpixel der Matrix darstellt.

2.5.1 DC-Wert

Da Bildaufnahmen von Mittelinfrarot-Kameras stets schwarzwei sind (und warme Bereiche heller dargestellt werden als khle) besitzt ein Pixel genau einen Informationskanal mit einem Grauwert zwischen 0 (schwarz) bis 255 (wei). Der DC-Wert liegt somit als Mittelwert ebenso in diesem Intervall. Dieser Bereich ist zu gro um als Eingabealphabet fr das HMM dienen zu knnen, da zum Beispiel der bergang von einem Grauwert von 230 auf 235 nicht aussagekrftig ist, da es sich hierbei um Rauschen oder aber eine Neukalibrierung der Kamera, die sich ja stets auf die gesamte 8 x 8 Matrix auswirkt, halten kann.

Wir nehmen jedoch eine Dreiteilung des Wrmespektrums an, einen khlen Hintergrund, warmen Vordergrund und ein mittelwarmen Bereich. Histogramm-basierte Analysen der DC-Werte ber mehrere unterschiedliche Videosequenzen ber alle Pixelblcke genormt besttigen diese Annahme und manifestieren 3 Cluster, die wir vereinfacht Cluster schwarz, Cluster grau und Cluster wei nennen - der Darstellungsfarbe des Wrmespektrums entsprechend. Die Grenzen dieser Cluster sind abhngig von jedem Video und dem betrachteten Umfeld, denn die insbesondere die Klte (und damit die Farbe) des Hintergrundes kann sich bei den jeweiligen Videos stark unterscheiden.

Wir bilden auf Basis dieser Erkenntnis die ersten 3 Symbole, wobei, wenn der DC-Wert eines Blockes in das Intervall eines Clusters fllt, wir die entsprechende Beobachtung erstellen:

DC-Wert im Intervall Cluster schwarz Beobachtung: DC_BLACK DC-Wert im Intervall Cluster grau Beobachtung: DC_GREY DC-Wert im Intervall Cluster wei Beobachtung DC_WHITE

Das folgenden Bild zeigt eine exemplarische Einteilung in die drei Cluster. Die X-Achse beschreibt den Grauwert, die Y-Achse das Vorkommen.

2.5.2 AC-Wert

Die bei der Anwendung der DCT entstehende 8 x 8 Matrix enthlt 63 AC-Werte, diese bilden demnach den Groteil der in der Matrix kodierten Informationen; die Integration dieser Werte in das Eingabealphabet ist aufgrund der hohen Informationsdichte essenziell.

Eine Histogramm-basierte Betrachtung der AC-Werte ist nicht zielfhrend, da es sich hierbei um Offsets handelt, welche demnach im Histogramm ein Maximum um den Nullwert bilden. Wir erkennen jedoch, das Menschen aufgrund ihrer warmen Austrahlung eine starke Kante zu dem Hintergrund bilden, luft demnach ein Mensch durch einen Block, msste eine deutliche Abweichung einiger AC-Werte zum DC-Wert entstehen. Eine Block-basierte Analyse besttigt diese Annahme, die Standardabweichung (STD)

der AC-Werte bezüglich des DC-Wertes steigt deutlich an, falls eine Person sich im Block befindet beziehungsweise durch diesen hindurch läuft.

An folgender Abbildung sieht man auf der X-Achse den Zeitverlauf in Frame und auf der Y-Achse den Wert der Standardabweichung für den betrachteten Block; es lässt sich feststellen, dass die Kurve der Standardabweichung stark ausschlägt, falls eine Person durch den Block läuft.

Anhand von empirischen Tests lässt sich ein Threshold ermitteln, falls die Standardabweichung diesen übersteigt, befindet sich mit hoher Wahrscheinlichkeit eine Person im Block. In dem vorgestellten Beispiel liegt dieser bei 25.

Zusätzlich ist zu beachten, dass die Standardabweichung von wenigen, aber extremen statistischen Ausreißern (statistical outliers) kaum beeinflusst wird, diese Situation liegt jedoch während der Erkennung von Menschen vor, insbesondere bei Personen, die Kleidung tragen, welche die Wärmeabgabe der Person dominiert, und dennoch einige freiliegende Körperteile sichtbar sind. In diesem Fall ist der Großteil der AC-Werte relativ klein und der DC-Wert des betrachteten Blockes für gewöhnlich innerhalb des grauen Intervalls. Um in diesen Situationen den Vordergrund besser zu erkennen, prüfen wir mithilfe eines selbstdefinierten Verfahrens outliers() ob stark-positive (warme) Ausreißer vorliegen, welches den Wahrheitswert true zurückliefert, falls eine gewisse Anzahl von Ausreißern erkannt wird.

Vor allem bei Kleidung tragenden Personen führt dies dazu, dass nicht nur ihre Körperkanten durch die STD erkannt werden, sondern jedoch auch die Körpermitte mathematisch erfasst wird.

Infolge dieser Erkenntnisse bilden wir die zwei weiteren Beobachtungen:

(STD von AC1-63 > Threshold) → Outliers Beobachtung AC_LOW (STD von AC1-63 < Threshold) → Outliers Beobachtung AC_HIGH

Auf Basis unserer Beobachtungen können nun anhand aller möglichen Permutationen von DC- und AC-Beobachtungen die Symbole unseres Eingabealphabets gebildet werden:

Symbol 1 DC_BLACK AC_LOW Symbol 2 DC_BLACK AC_HIGH Symbol 3 DC_GREY AC_LOW
Symbol 4 DC_GREY AC_HIGH Symbol 5 DC_WHITE AC_LOW Symbol 6 DC_WHITE AC_HIGH

2.6 Verfahren zur Hintergrundentfernung

Unser Verfahren zur Vordergrund-Extraktion basiert auf den zuvor genannten Algorithmen, der Bildpartitionierung in Blöcke und dem eingeführten Eingabealphabet.

Jeder Block erhält ein eigenes, individuelles HMM, da Blöcke unterschiedliche Übergangs-/Emissionswahrscheinlichkeiten aufgrund ihrer Lokalität aufweisen: ein Block am Rand neigt eher dazu hell und dunkel zu sein, ein Block an einer Tür hingegen ist eher für Veränderungen ausgesetzt, da erwartungsgemäß häufiger Personen in diesem Bereich hindurchlaufen.

In der ersten Phase ist ein Trainingsvideo zu verwenden, welches dazu dient, (individuelle) Trainingssequenzen für die Blöcke zu erstellen und im Folgenden diese dem Baum-Welch-Algorithmus zu übergeben, welcher die HMMs blocklokal trainiert. Da in einem Block hauptsächlich der Hintergrund zu beobachten ist, wird jeder Block nach der Lernphase seinen eigenen Hintergrund kennen lernen und somit die korrespondierenden Emissionen als wahrscheinlicher bewerten.

In der zweiten Phase kann bereits ein Evaluationsvideo eingesetzt werden. Für jeden Block wird die aktuelle Beobachtung gebildet. Die HMMs werden nun zur Vorhersage und Verifikation verwendet. Unter Verwendung des Forward- oder Viterbi-Algorithmus kann nun die nächste Emission prognostiziert werden. Diese wird mit der tatsächlich vorliegenden Emission (eine der 6 definierten Symbole unseres Eingabealphabetes) verglichen. Bei Gleichheit handelt es sich offensichtlich um den erlernten Hintergrund, bei Abweichungen ist ein Vordergrundobjekt in den Block eingetreten. Die aktuelle Emission wird an eine Warteschlange der k -letzten Emissionen angehängt und die Operationen wiederholt.

TODO: PSEUDOCODE?

Erklärung warum eine HMM-TOPOLOGIE NICHT so wichtig ist bzw. wir keine mit 2 Zuständen genommen haben? Vorschlag:

Bei der Konzeption des von uns verwendeten HMMs muss der Kontext betrachtet werden, in dem diese eingesetzt werden sollen: mit Hilfe der HMMs sollen die einzelnen Frames eines Videos in Hintergrund, d.h. statische Elemente wie zum Beispiel Wände oder Fußböden, und Vordergrund, d.h. dynamische Elemente, wie vor allem Personen, aufgeteilt werden. Zum Vordergrund gehörende Blöcke werden später auf Grundlage statistischer Verfahren auf Personen aufgeteilt oder zusammengefasst, zum Hintergrund gehörende Blöcke werden nicht weiter betrachtet. Aufgrund dessen ist es konsequent als Zustände für das verwendete HMM auch genau zwei Zustände zu verwenden, die wir als Vordergrund und Hintergrund bezeichnen. Diese Aufteilung ist insofern konsequent, als dass weitere Zustände für unser Problem keine weiterführenden Informationen bieten würden. Falls wir uns für ein Modell mit mehr als zwei Zuständen entschieden hätten, müsste man für jeden einzelnen zusätzlichen Zustand eine Entscheidung treffen, ob wir ihn zum Vordergrund oder zum Hintergrund hinzuzählen. Diese Entscheidung müsste ähnlich wie die Übergänge im HMM probabilistisch erfolgen, da wir nicht sicher sagen können, ob ein Block zum Vordergrund oder Hintergrund gehört. Anstatt also eine weitere probabilistische Auswertung vorzunehmen, erfolgt die Aufteilung aufgrund der Übergangswahrscheinlichkeiten des HMMs.

Kapitel 3

Implementierung und praktische Details

Dieses Kapitel soll einen kurzen Überblick zu den von uns verwendeten Technologien bieten, die wir zur Realisierung des theoretischen Entwurfs aus Kapitel 2 verwendet haben.

3.1 Videomaterial

Das uns zur Verfügung gestellte Bildmaterial besteht aus einzelnen Videosequenzen, wir können also nicht auf Live-Material arbeiten. In der Praxis spielt dies jedoch keine große Rolle, da unter Linux der Zugriff auf eine Videodatei sich nicht wesentlich von dem Zugriff auf eine Kamera unterscheidet. Insgesamt standen uns sieben unterschiedliche Videosequenzen zur Verfügung, von denen sechs paarweise entstandene Aufnahmen sind. Das bedeutet, dass Kameraposition sowie Winkel zwischen den beiden Videosequenzen eines Paares nicht differiert, lediglich die aufgenommenen Szenen sind unterschiedlich. Diese Tatsache ist insofern hilfreich, als dass wir die Möglichkeit haben, auf einer Videosequenz zu lernen und die erlernten Parameter später auf der anderen Videosequenz anzuwenden. Der größte Nachteil der Verwendung von Videosequenzen ist, dass die einzelnen Sequenzen relativ kurz sind (Dauer beträgt nicht 10 Minuten) und es somit nicht möglich war, wirklich lange Lernphasen von zum Beispiel einigen Stunden unter realistischen Bedingungen zu testen. Sämtliche Videosequenzen laufen mit 25 Frames pro Sekunde, was auf jeden Fall genug Daten zur Auswertung liefert. In der Praxis werden vermutlich auch schon weniger Frames genügen, da zu erwarten ist, dass sich die Personen im Bild nicht so schnell bewegen, dass sich innerhalb eines anderen Bruchteils einer Sekunde sehr viel verändert.

3.2 C++

Die Implementierung sowohl unseres Testbeds als auch die finale Implementierung als Komponente zur Foreground extraction in einem vorgegebenen Framework fand in C++ statt. Für die Wahl von C++ gab es verschiedene Gründe, von denen vor allem die Performance, die Portabilität sowie das gute Angebot an zur Verfügung stehenden Bibliotheken im Vordergrund standen.

3.3 OpenCV

OpenCV ist eine Bibliothek, die eine Vielzahl unterschiedlicher Algorithmen für die Bildbearbeitung und somit letztlich für die Videobearbeitung zur Verfügung stellt. Hinzu kommt, dass auch Funktionen für das Lesen, Schreiben und Abspielen von Videodateien angeboten werden. OpenCV bietet Interfaces für C, Python, Java und C++ und lässt sich auf vielen unterschiedlichen Plattformen einsetzen. Für unser Projekt relevant waren dabei vor allem die Funktionen zur Wiedergabe von Videos sowie die in OpenCV enthaltene DCT Implementierung.

3.4 CvHMM

Die Wahl einer geeigneten Bibliothek für HMMs ist nicht trivial. Es existieren nicht sehr viele effiziente und aktiv entwickelte Bibliotheken für HMMs, die für unser Projekt in Frage kämen. Letztlich fiel unserer Wahl auf CvHMM, vor allem da diese Bibliothek direkt auf dem in OpenCV enthaltenen Standard-Videodatentyp Mat arbeitet und daher die Vermutung nahe liegt, dass weniger Performanceeinbußen vorliegen, da keine Konvertierung von Daten in andere Formate erfolgen muss. Leider hat sich im Laufe des Projekts gezeigt, dass CvHMM leider sehr ineffizient auf einzelne Daten zugreift. Ein weiterer großer Nachteil von CvHMM ist, dass nur der Baum-Welch-, der Viterbi- und der Decode-Algorithmus angeboten werden. Gerade hier ist wiederum ein Manko zu sehen, da wir später den Forward-Algorithmus gebraucht hätten und nun stattdessen den Viterbi-Algorithmus verwenden müssen, der jedoch ineffizienter als der Forward-Algorithmus ist. Zusammenfassend lässt sich zur Wahl der HMM Bibliothek sagen, dass es wohl das Ergebnis des Projekts erheblich hätte verbessern können, wenn unsere Wahl auf eine andere Bibliothek gefallen wäre. Mögliche Kandidaten zum Testen wären dabei HMMlib oder vielleicht sogar die, wie wir später herausgefunden haben, in OpenCV enthaltenen HMM-Funktionen, die allerdings innerhalb des OpenCV-Projekts nicht weiterentwickelt werden und derzeit des Status deprecated haben.

Kapitel 4

Evaluation

Wir haben mit Hilfe des gegebenen tracking-Frameworks verschiedene Konstellationen mit einander verglichen:

- a) die vorimplementierte Lsung, die mit Histogrammen arbeitet
- b) unsere HMM-basierte Lsung mit generischen Werten fr Initialwerte, Emission und Transition ohne Lernen
- c) unsere HMM-basierte Lsung mit generischen Werten fr Initialwerte, Emission und Transition mit Lernen
- d) unsere HMM-basierte Lsung mit vorgelernten Werten und Lernen

4.1 Vergleich mit Histogramm-basierter Implementierung: Innenhof

4.2 Vergleich mit Histogramm-basierter Implementierung: Eingang

4.3 Vergleich mit Histogramm-basierter Implementierung: Tegel

4.4 Diskussion

bezug auf anforderungen nehmen, punkte erflft? hmm topologie? andere libs? forward vs. viterbi dynamische histogramm/cluster grenzen fr dc wert in zukunft? viterbi/forawrd gibt zustand, man kann entweder

gewichtetes wrfeln oder aber deterministisch die wahrscheinlichste beobachtung nehmen, was ist besser?
test? performance gewinn durch threading zustzlich

Kapitel 5

Fazit und Ausblick