

Philipps-Universität Marburg

Fachbereich 12 - Mathematik und Informatik



Masterarbeit

Analyse von Videodaten eines Smart Home Systems zur Erkennung von außergewöhnlichen Situationen

von
Anh Duc Nguyen
April 2018

Betreuer:
Prof. Dr. Thorsten Thormählen
Dr. Tobias Kühnl
Dr. Marcel Schneider

Arbeitsgruppe Grafik und Multimedia Programmierung

Erklärung

Ich, Anh Duc Nguyen (Informatikstudent an der Philipps-Universität Marburg, Matrikelnummer xxxxxx), versichere an Eides statt, dass ich die vorliegende Masterarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die hier vorliegende Masterarbeit wurde weder in ihrer jetzigen noch in einer ähnlichen Form einer Prüfungskommission vorgelegt.

Marburg, 01. April 2018

Anh Duc Nguyen

Zusammenfassung

Bei meiner Masterarbeit geht es um eine Echtzeitanwendung, wo man zur Erkennung von außergewöhnlichen Situationen für alte Alleinlebende verwenden kann. Die Eingaben des Programms sind Frames von einer Kamera, und im nächsten Schritt werden Informationen erzeugt, die uns sagen ob da eine Bewegung gibt. Aus die Bewegungsmeldungen wird ein Vordergrund von dem Hintergrund mit Hilfe von Hintergrundsubtraktion und -Update extrahieren. Nach dem Extrahieren des Vordergrundes wird eine Silhouette von Schwarzweiß erstellt. Mit Hilfe von einer Histogrammanalyse auf Silhouette wird eine Körperhaltung vorhersagen. Das Programm kann sagen ob das bewegende Objekt steht, sitzt, beugt oder liegt. Anhand von Positionen wo das Objekt oft bewegt und wann das Objekt steht, sitzt oder liegt... kann außergewöhnlichen Situationen erkannt werden.

Abstract

... (exakte englische Übersetzung der deutschen Kurzfassung)

Inhaltsverzeichnis

Inhaltsverzeichnis	I
1. Einleitung	1
1.1. Motivation	1
1.2. Ziele	1
1.3. Aufbau der Arbeit	2
1.4. Verwandte Arbeiten	2
1.4.1. Hintergrundsubtraktion	2
1.4.2. Körperhaltungserkennung basiert auf Histogrammanalyse	3
2. Grundlagen	5
2.1. Hintergrundsubtraktion	5
2.1.1. Gaußschen Mixture Modell	6
2.1.2. Adaptive Gaußschen Mixture Modell	7
2.1.3. Kernel Density Estimation	7
2.1.4. Vibe	8
2.2. Histogrammanalyse	9
2.3. Fuzzylogik	10
2.4. OpenCV Framework	10
2.5. Bosch Innenkamera	12
3. Eigenes Verfahren	14
3.1. Hintergrundsubtraktion	14
3.1.1. Vergleichen verschiedener Methoden	14
3.1.2. Eigene Methode	18
3.1.3. Ergebnisse der eigene Hintergrundsubtraktion	18
3.2. Schätzung der Körperhalterung mit Histogrammanalyse	18
3.3. Erkennung außergewöhnlicher Situation	18
4. Ergebnisse und Evaluation	20
5. Zusammenfassung und Ausblick	22
Literaturverzeichnis	23
Abkürzungsverzeichnis	26

Abbildungsverzeichnis	28
Tabellenverzeichnis	29
Liste der Algorithmen	31
Listings	33
A. Anhang	35
Danksagung	36

1. Einleitung

Diese Masterarbeit beschäftigt sich mit der Hintergrundsubtraktion um eine Bewegung zu erkennen, Segmentieren eine ROI um eine Silhouette zu verbessern, Histogrammanalyse um Körperhaltung vorherzusagen, Wahrscheinlichkeitsberechnungen um die abnormale Situationen zu schätzen. OpenCV ist hier angewendet, das Framework vereinfachte den Weg zur Implementierung sowie erhöht die Geschwindigkeit des Programms.

1.1. Motivation

In Deutschland werden die Menschen immer älter und die Meisten wohnen oft allein in einer Wohnung oder Haus. Manche Familien haben nur ein Kind und nicht selten beide Elternteile arbeiten gehen. Außerdem verlangt die Arbeitswelt immer mehr Flexibilität, deshalb wohnen die Kinder oder Verwandten von Senioren oft in anderen Städten und sind nur schwer zu erreichen. Im Besonderen je älter die Senioren sind, desto größere Angst haben die vor Altenheime. Deutschland stellt vor neue Herausforderung. Heutzutage leben immer mehr und mehr Senioren allein. Beispielsweise wohnt jede zweite Seniorin allein und 20 Prozent sind über 65 Jahre alt. Wenn die Alte einsam sind, erhöht auch das Risiko, dass die unter Depressionen, Herzerkrankung oder Unfall leiden. Das ist aber kein Ende des Tunnels, mit die Entwicklung von Technologien kann das Problem mit Assisted Ambient Living ALL schon gelöscht werden. Derzeit springt der IT-Bereich besonders im Computer Vision einen großartigen Schritt. Durch Anwendung einer Kamera können die Verwandter oder Kinder von Senioren einfach im real-time sehen. Aber nicht jeder hat Zeit den ganzen Tag vor Bildschirm beobachten, deshalb das Programm hier ist eine Hilfe zur Erkennung der Situationen, wenn irgendwas mit allein wohnende Alte passiert.

1.2. Ziele

Das Ziel dieser Arbeit ist die Entwicklung eines Programms, das mit Hilfe von Kamera Bewegungen erkennt, und dadurch Vordergrund von Hintergrund extrahieren analysiert. Am Ende soll das Programm die Situationen abschätzen, ob alles noch gut ist oder irgendwas schief läuft.

Echtzeit

Da es um eine Unfallmeldung geht, deshalb muss die Anwendung schon in Echtzeit laufen. Die durchschnittliche Laufzeit des Programms soll höchstens 3 Frames pro Sekunde benötigen, um die Situation in der Echtzeit zu erkennen.

Kosten

Das Programm sollte mit der normalen, kostengünstigen Privatanwender-Hardware laufen, sodass für die Ausführung keine Spezialrechner benötigt werden. Die einzige Voraussetzung an das System ist eine Kamera. Diese ist aber in den meisten Desktoprechnern vorhanden oder kann kostengünstig nachgerüstet werden.

1.3. Aufbau der Arbeit

Im ersten Abschnitt des Kapitels 2 soll zunächst die angewendete Hintergrund Subtraktionen verdeutlicht werden. Dann soll im zweiten Abschnitt die Histogrammanalyse beschrieben werden, bevor im Abschnitt 2.4 ein Überblick über das OpenCV Framework geben wird. An dieser Stelle werden auch die Vorteile der Anwendung von OpenCV bei dieser Arbeit genauer erläutern. Kapitel 3 beschäftigt sich mit der Theorie des im Rahmen dieser Masterarbeit entwickelten Algorithmus. Dabei sollen die einzelnen Schritte des Algorithmus genauer von erläutert werden. Die Implementierung des Algorithmus wird im diesen Kapitel besprochen, dabei sollen die benutzten Bibliotheken sowie die Umsetzung des Programms beschrieben werden. Anschließend wird im Kapitel 4 die durchgeführten Tests und deren Ergebnisse dokumentiert und diskutiert, sowie im Kapitel 5 ein Ausblick auf weiterführende Arbeiten gegeben.

1.4. Verwandte Arbeiten

1.4.1. Hintergrundsubtraktion

Bewegungserkennung spielt heutzutage eine wichtige Rolle in Überwachungssystemen. Durch die Subtraktion von Hintergrund wird die Bewegung von dem Hintergrund extrahiert. Es gibt verschiedene wissenschaftliche Arbeiten für diese Methode. Ein der Verfahren ist "Adaptive background mixture models" [SG99], das ist eine gängige Methode zur Echtzeit-Segmentierung bewegter Regionen in Bildsequenz. Die Methode beinhaltet eine "Hintergrundsubtraktion" oder eine Schwellenwertbindung des Fehlers zwischen einer Schätzung des Bild ohne sich bewegende Objekte und dem aktuellen Bild. In [SG99] wird das Modellieren jedes Pixels als eine Mischung von Gauß-Werten und das Verwenden einer Online-Approximation zum Aktualisieren des Modells erläutert. Die Gaußschen Verteilungen des adaptiven Mischungsmodells werden dann ausgewertet, um zu bestimmen, welche am wahrscheinlichsten aus einem Hintergrundprozess resultieren.

Eine verbesserte Version von Gauß'schen Mischungsmodells ist in [ZVDH06] als "KNN" Methode beschrieben. Rekursive Gleichungen werden präsentiert, die dazu dienen, die Parameter eines Gauß'schen Mischungsmodells ständig zu aktualisieren und gleichzeitig die geeignete Anzahl von Komponenten für jeden Pixel auszuwählen.

In [EHD00] wird eine Methode, die "kernel density estimation" (KDE) heißt, besprochen. KDE ist ein neuartiges nicht-parametrisches Hintergrundsubtraktion. Das Modell kann mit Situationen umgehen, in denen der Hintergrund der Szene überladen und nicht vollständig statisch ist, sondern kleine Bewegungen wie Äste und Büsche enthält.

Das Modell schätzt die Wahrscheinlichkeit der Intensitäten von Pixel auf der Grundlage einer Stichprobe von Intensitäten für jeden Pixel. Das Modell passt sich schnell an Veränderungen in der Szene an, was eine sehr empfindliche Erkennung bewegter Ziele ermöglicht.

In [BVD09] und [BVD11] wird “Vibe”, eine leistungsfähige Methode zur Hintergrundextraktion vorgestellt. “Vibe” verbessert die Genauigkeit und reduziert die Rechenlast. Die Hauptinnovation betrifft die Verwendung einer zufälligen Strategie zur Auswahl von Werten, um eine stichprobenartige Schätzung des Hintergrunds zu erstellen. Es war das erste Mal, dass eine zufällige Wahrscheinlichkeit im Bereich der Hintergrundextraktion verwendet wird.

Für diese Arbeit werden alle oben genannten Methoden verglichen und die Vergleichen aller oben genannten Methoden werden im nächsten Kapitel genauer beschrieben.

1.4.2. Körperhaltungserkennung basiert auf Histogrammanalyse

Histogrammanalyse ist eine einfache und schnelle Methode zur Erkennung der Körperhaltung. Es gibt viele wissenschaftliche Arbeiten, die über Histogrammanalyse studiert und präsentiert werden. In [GM06] wird das Projektionshistogramm verwendet, um unterschiedliche Körperhaltungen zu unterscheiden. Ein neues Verfahren wird vorgeschlagen, dass ein Projektionshistogramm zur statischen Erkennung der menschlichen Haltung verwendet. Diese Arbeit in [GM06] besteht aus drei Schlüsselmodulen: Hintergrundsubtraktion, Projektionshistogrammberechnung und Vorlagenabgleich.

Eine andere Methode ist in [HHD98] beschrieben, die als “Ghost” System benannt. “Ghost” ist ein Echtzeitsystem zum Schätzen der menschlichen Körperhaltung und zum Detektieren von Körperteilen. Es erstellt ein Silhouetten-basiertes Körpermodell, um die Position der Körperteile zu bestimmen, während sich Personen in generischen Positionen befinden. Es kombiniert eine hierarchische Schätzung der Körperposen mit Histogrammanalyse, eine konvexe Rumpfanalyse der Silhouette und eine teilweise Abbildung von den Körperteilen zu den Silhouetten-Segmenten. Die Histogrammanalyse in dieser Masterarbeit basiert auf die Methode, die in “Ghost” beschrieben wird.

2. Grundlagen

Im diesen Kapitel werden die theoretischen Grundlagen der Masterarbeit beschrieben, die zum Verständnis der nachfolgenden Kapitel notwendig sind. Zu Beginn wird ein Überblick der Hintergrundsubtraktion gegeben und aufgezeigt, welche Besonderheiten diese Arbeit auszeichnen. Am schließend wird die Histogrammanalyse zur Erkennung der Körperhaltung erläutert. Zuletzt befasst der letzte Teil sich mit einem Überblick über das wichtiges Framework OpenCV, das für moderne Computer Vision entwickelt wird.

2.1. Hintergrundsubtraktion

Die Hintergrundsubtraktion ist eine gebräuchliche und weit verwendete Technik unter Verwendung von statischen Kameras zum Erzeugen eines Binärbildes, das die Pixel enthält, die zu sich bewegenden Objekten in der Szene gehören. Wie der Name andeutet, berechnet Hintergrundsubtraktion die Vordergrundmaske, die eine Subtraktion zwischen dem aktuellen Bild und einem Hintergrundmodell durchführt, wobei der statische Teil der Szene oder allgemeiner alles, was angesichts der Merkmale der beobachtete Szene als Hintergrund betrachtet werden kann, enthalten ist.

In der Arbeit [BJEL10] wird eine vergleichende Studie verschiedener Hintergrundsubtraktionsverfahren nach dem Stand der Technik präsentiert. Dieses Verfahren wurde seit den 1990er umfassend untersucht und hauptsächlich für Videoüberwachungsanwendungen, da sie zuerst Personen, Fahrzeuge, Tiere usw. erkennen müssen, bevor komplexere Prozesse zur Einbruchserkennung, Verfolgung, Personenzählung [AMFT11] usw. ausgeführt werden. Viele Algorithmen wurden entworfen, um die Vordergrundobjekte vom Hintergrund einer Sequenz zu segmentieren und teilen im Allgemeinen das gleiche Schema:

- Initialisierung des Hintergrundes: Ein Hintergrundmodell wird zuerst dank einer festen Anzahl von Frames zu bauen. Dieses Modell kann auf verschiedene Arten entworfen werden (statistisch, Fuzzy...)
- Erkennung des Vordergrundes: In den nächsten Frames wird ein Vergleich zwischen dem aktuellen Frame und dem Hintergrundmodell durchgeführt. Diese Subtraktion führt zu Berechnung des Vordergrundes der Szene.
- Initialisierung des Hintergrundes: Während dieses Erfassungsprozesses werden auch Bilder analysiert, um das im Initialisierungsschritt gelernte Hintergrundmodell in Bezug auf eine Lernrate zu aktualisieren. Ein Objekt, das sich nicht lange bewegt, sollte im Hintergrund integriert sein.

Im nächsten Teil werden verschiedene state-of-art Verfahren beschrieben, um dies Verfahren genauer zu verstehen, wie die Hintergrundsubtraktion funktionieren soll.

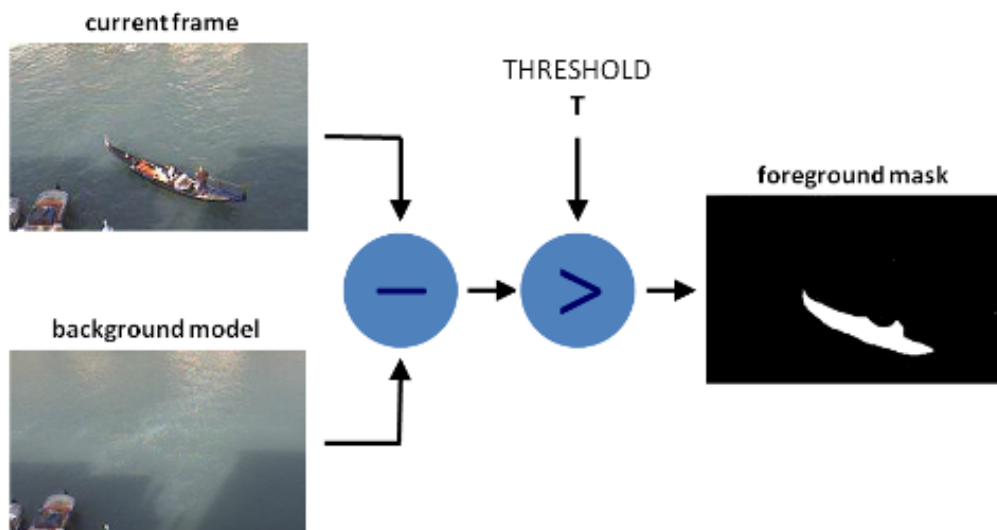


Abbildung 2.1.: Ein Beispiel für Hintergrundsubtraktion

2.1.1. Gaußschen Mixture Modell

In [KB02] ist das entwickelte Hintergrundmodell auf der Grundlage der Gauß'schen Mischung präsentiert, das ein meist gebräuchlicher Weg ist. Es verwendet eine Methode, um jedes Hintergrundpixel durch eine Mischung von K -Gaußschen Verteilung (normalerweise $K = 3$ bis 5) zu modellieren. Die Gewichte der Mischung stellen die Zeitanteile dar, die diese Farbe in der Szene verbleiben. Die wahrscheinlichen Hintergrundfarben sind diejenigen, die länger und statischer bleiben [KB02].

Jeder Pixel in der Szene wird durch eine Mischung von K -Gaußschen Verteilung modelliert. Die Wahrscheinlichkeit, dass ein bestimmter Pixel zum Zeitpunkt N einen Wert von x_n hat, kann wie folgt geschrieben werden:

$$p(x_n) = \sum_{j=1}^K w_j \eta(x_n; \theta_j) \quad (2.1)$$

wobei w_k ist Gewichtsparameter der k -ten Gauß-Komponente ist. $\eta(x_n; \theta_j)$ ist die Normalverteilung der k -ten Komponente, die wie folgt dargestellt wird [KB02]:

$$\eta(x; \theta_j) = \eta(x; \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)} \quad (2.2)$$

wobei μ_k der Durchschnitt ist und $\Sigma_k = \sigma_k^2 I$ ist die Kovarianz der k -ten Komponente [KB02]. Die K -Verteilungen sind auf der Grundlage des Fitnesswerts $\frac{w_k}{\sigma_k}$ geordnet und die ersten B -Verteilungen werden als ein Modell des Hintergrundes der Szene verwendet, wo B wie folgt beschrieben werden:

$$B = \arg \min_b \left(\sum_{j=1}^b w_j > T \right) \quad (2.3)$$

Und das Schwellwert T ist Mindestanteil des Hintergrundmodells, nämlich ist T das minimale Wahrscheinlichkeit, dass der Hintergrund in der Szene ist. Die Hintergrundsubtraktion wird durchgeführt, indem ein Vordergrundpixel jeden Pixel markiert wird, das mehr als 2,5 Standardabweichungen von irgendeiner der B-Verteilung entfernt ist. Die erste Gauß-Komponente, die die oben genannte Bedingung erfüllt, wird durch die folgenden Aktualisierungsgleichungen aktualisiert [KB02]:

$$\hat{w}_k^{N+1} = (1 - \alpha) \hat{w}_k^N + \alpha \hat{p}(w_k | x_{N+1}) \quad (2.4)$$

$$\hat{\mu}_k^{N+1} = (1 - \alpha) \hat{\mu}_k^N + \rho x_{N+1} \quad (2.5)$$

$$\hat{\Sigma}_k^{N+1} = (1 - \alpha) \hat{\Sigma}_k^N + \rho (x_{N+1} - \hat{\mu}_k^{N+1})(x_{N+1} - \hat{\mu}_k^{N+1})^T \quad (2.6)$$

$$\rho = \alpha \eta(x_{N+1}; \hat{\mu}_k^N; \hat{\Sigma}_k^N) \quad (2.7)$$

$$X = \begin{cases} 0, & \text{wenn } w_k \text{ erste Gauß-Komponente ist} \\ 1, & \text{sonst} \end{cases} \quad (2.8)$$

wobei w_k die k-ten Gaußschen Komponente und $\frac{1}{\alpha}$ definiert die Zeitkonstante, die die Änderung bestimmt. Wenn keine der K-Verteilung mit diesem Pixelwert übereinstimmt, wird die unwahrscheinliche Komponente durch eine Verteilung mit dem aktuellen Wert als Mittelwert ersetzt.

2.1.2. Adaptive Gaußschen Mixture Modell

In [ZVDH06] wurde eine interessante Erweiterung von GMM vorgeschlagen. Normales GMM wird mit eine bestimmte Anzahl von Gaußschen Modell, aber AGMM wird automatisch die Anzahl der Gaußschen Variablen angepasst, die zur Modellierung eines gegebenen Pixels verwendet werden. Die Erweiterung reduziert die Speicheranforderung des Algorithmus, erhöht die Recheneffizienz und kann die Leistung verbessern, wenn der Hintergrund stark sich ändert.

2.1.3. Kernel Density Estimation

Ein Nachteil von GMM ist, dieses Modell kann keine empfindliche Detektion erreichen, wenn der Hintergrund sehr hohe Frequenzvariation aufweist. Dieser Nachteil kann mit KDE Modell, das in [EHD00] beschrieben wurde, gelöst werden.

Sei x_1, x_2, \dots, x_n eine aktuelle Stichprobe von Intensitätswerten für einen Pixel. Unter Verwendung dieses Stichprobe kann die Dichtefunktion, dass dieser Pixel einen Intensitätswert x_t zur Zeit t haben wird, unter Verwendung des Kernschätzers K als nicht-parametrisch geschätzt werden [EHD00].

$$P(x_t) = \frac{1}{n} \sum_{i=1}^N K(x_t - x_i) \quad (2.9)$$

Wenn wir unsere Kernschätzfunktion K als Normalfunktion $N(0; \sigma)$ wählen, wobei σ die Kernfunktionsbandbreite darstellt. Wenn verschiedene Farbkanäle mit unterschiedlichen Kernel-Bandbreite σ_j^2 für den j -ten Farbkanal:

$$\sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} \quad (2.10)$$

dann wird die Dichtefunktion wie folgt geschrieben:

$$P(x_t) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^d \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2} \frac{(x_{tj} - x_{ij})^2}{\sigma_j^2}} \quad (2.11)$$

Unter Verwendung dieser Wahrscheinlichkeitsschätzung wird der Pixel als ein Vordergrundpixel betrachtet, wenn $P(x_t) < T$, wobei T ein globaler Schwellenwert über das gesamte Bild ist, der eingestellt werden kann, um einen gewünschten Prozentsatz von Falsch-Positiven zu erreichen.

Angenommen m ist Median von $|x_i - x_{i+1}|$ für jedes Paar (x_i, x_{i+1}) in der Stichprobe. Nach [EHD00] wird die Standardabweichung der erste Verteilung wie folgt geschätzt:

$$\sigma = \frac{m}{0,68\sqrt{2}} \quad (2.12)$$

Bei KDE gibt es zwei Alternative für Update des Hintergrundes, nämlich “Selective Update” und “Blind Update”. Die erste Alternative fügt neue Stichprobe in Modell hinzu, genau dann wenn es als Hintergrund klassifiziert ist. Die zweite Alternative fügt einfach neue Stichprobe in Modell hinzu. Im Allgemeinen funktioniert KDE draußen besser als GMM Methode. Für meine Masterarbeit wurden mehrere Methode ausprobiert, verglichen und ausgewertet, damit eine bessere Anwendung benutzt werden kann.

2.1.4. Vibe

In [BVD09] wurde ein Verfahren beschrieben, das ersten Ansatz von zufälliger Aggregation in Hintergrundsubtraktion ist. Das Verfahren ist “ViBe” genannt. Sei $p_t(x)$ ein Pixelwert x in Zeit t . Bei GMM oder KDE Modell wird ein Pixelwert $p_t(x)$ als Hintergrund oder Vordergrund klassifiziert, abhängig davon, wie der Pixel mit Dichtefunktion des Modells passt. In “ViBe” wird aber eine Menge von Stichprobewerte als Pixelmodell benutzt. Um einen Wert $p_t(x)$ zu klassifizieren, wird der Wert mit seinen nächsten Werten in der Menge der Stichproben, in dem wie eine Kugel $S_R(p_t(x))$ mit Radius R und Punkt $p_t(x)$ definiert wird. Ein Pixel ist genau dann als Background klassifiziert, wenn die Überschneidung \sharp von Kugel $S_R(p_t(x))$ und Menge von Punkte p_1, p_2, \dots, p_n mehr als Schwellwert \sharp_m ist (siehe Abbildung 2.2).

Im nächsten Kapitel wird ein Vergleich von dem vier oben genannte Verfahren und meine eigene Methode beschrieben.

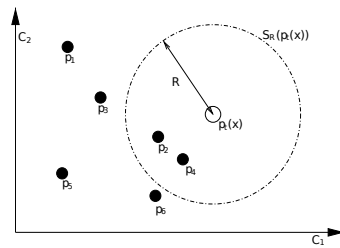


Abbildung 2.2.: Klassifizierung von $p_t(x)$ basiert auf Überschneidung von Kugel $S_R(p_t(x))$ mit Menge der Stichproben [BVD09].

2.2. Histogrammanalyse

Ein Histogramm eines Bildes ist graphische Darstellung der Tonwertverteilung in einem digitalen Bild. Ein Histogramm zeichnet die Anzahl der Pixel für jeden Tonwert auf. Auf dem Gebiet der Computervision können Bildhistogramme nützliche Werkzeuge für *Thresholding* sein. Da die in dem Graphen enthaltene Information eine Darstellung der Pixelverteilung als eine Funktion der Tonvariation ist, können sich Bildhistogramme auf Spitze und Täler analysieren lassen.

Da eine unterschiedliche Körperhaltung ein unterschiedliches Muster von Projektionshistogramm aufweist, kann das Projektionshistogramm als eines der Merkmale verwendet werden, um unterschiedliche Körperhaltungen zu unterscheiden. Nach der Hintergrundsubtraktion wird eine Silhouette des Vordergrundes als ein Binärbild erstellt. Eine Körperhaltungsanalyse wird auf die Silhouette angewendet, um die Ähnlichkeiten der horizontalen und vertikalen Projektionshistogramme der erkannten Silhouette und der Haupthaltungen zu berechnen. Die Normalisierung des durchschnittlichen Histogramm erfolgt durch Skalierung der Silhouette in eine vertikale Länge unter Beibehaltung des ursprünglichen Seitenverhältnisses [HHD98]. Normalisierte horizontale und vertikale Projektionsvorlagen für jede Körperhaltung (nämlich: Stehen, Krabbeln, Legen und Sitzen) wurden experimentell unter Verwendung von 4500 Silhouette von 7 verschiedenen Personen berechnet (siehe Abbildung 2.3).

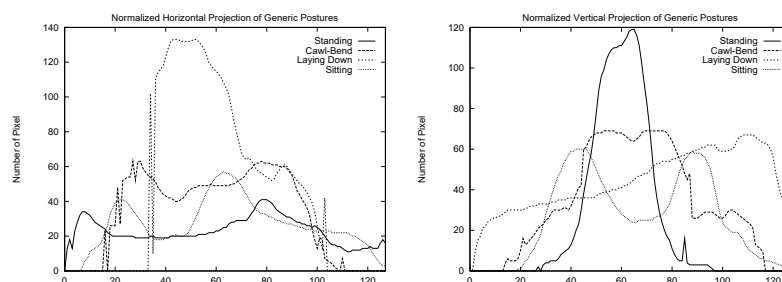


Abbildung 2.3.: Normalisierte horizontale und vertikale Projektionsvorlagen für jede Körperhaltung [HHD98].

Diese Methode werden verwendet, um die Ähnlichkeit der gegebenen Körperhaltung mit

einer der vier Haupthaltungen zu bestimmen. Die erstellte Silhouette durch Hintergrundsubtraktion wird mit den durchschnittlichen Projektionshistogramme verglichen, wobei die Summe der absoluten Differenz verwendet wird, um die ähnlichste Körperhaltung zu schätzen. Angenommen ist S_i die Ähnlichkeit zwischen der erkannten Silhouette und i -ten Hauptkörper. Seien H_i und V_i die horizontalen und vertikalen durchschnittlichen Projektionshistogramme, P und R die horizontalen und vertikalen Histogramme der erkannten Silhouette. S_i wird wie folgt berechnet:

$$S_i = -\log\left(\sum_h \sum_v^{128} |H_h^i - P_h| + |V_v^i - R_v|\right) \quad (2.13)$$

Die Körperhaltung, die das höchste Ähnlichkeitsmaß ergibt, wird als geschätzte Haltung genommen.

2.3. Fuzzylogik

Mit Hilfe von Hintergrundsubtraktion und Histogrammanalyse wird erstmal nur eine sich bewegte Person erkannt, aber zum Unterschied zwischen täglichen Bewegungen und ungewöhnlichen Situationen kommt Fuzzylogik zum Einsatz. Im Jahr 1971 wurde die erste Forschung über Fuzzy-Algebra beschrieben [Ros71]. Fuzzylogik geht um eine unscharfe logische Menge, wobei eine Fuzzymenge seinen Mitglieder ermöglicht, Mitgliedschaftsgrade zu haben. Wenn der Wert 1 an Objekten zugewiesen ist, die vollständig innerhalb der Menge liegen, und Objekten außerhalb der Menge eine 0 zugewiesen wird, hat jedes Objekt, das teilweise in der Menge ist, einen Wert zwischen 0 und 1. Der Prozess der Fuzzylogik wird wie folgt erläutert [Din11]:

- Zuerst wird ein scharfer Menge von Eingabedaten gesammelt und unter Verwendung von Zugehörigkeitsfunktionen in eine Fuzzymenge umgewandelt.
- Eine Schlussfolgerung wird basierend auf ein oder mehrere (*IF – THEN*) Regeln getroffen.
- Schließlich wird die Fuzzy-Ausgabe unter Verwendung der Zugehörigkeitsfunktionen in dem Defuzzifizierungsschritt auf eine scharfe Ausgabe abgebildet.

Beispielweise gibt es ein Smart-Thermostat in einem Raum und der Thermostat soll sich an der Raumtemperatur anpassen. Eine Fuzzylogik wird erstellt, wobei die Temperatur Eingabe ist und Einstellung des Thermostates ist Ausgabe (siehe Abbildung 2.4) und der Regel wird wie in Algorithmus 1 definiert:

2.4. OpenCV Framework

OpenCV ist eine Open-Source-Computer-Vision-Bibliothek, die in C und C++ programmiert wird und unter Linux, Windows, MacOS läuft. OpenCV ist auch auf Schnittstelle

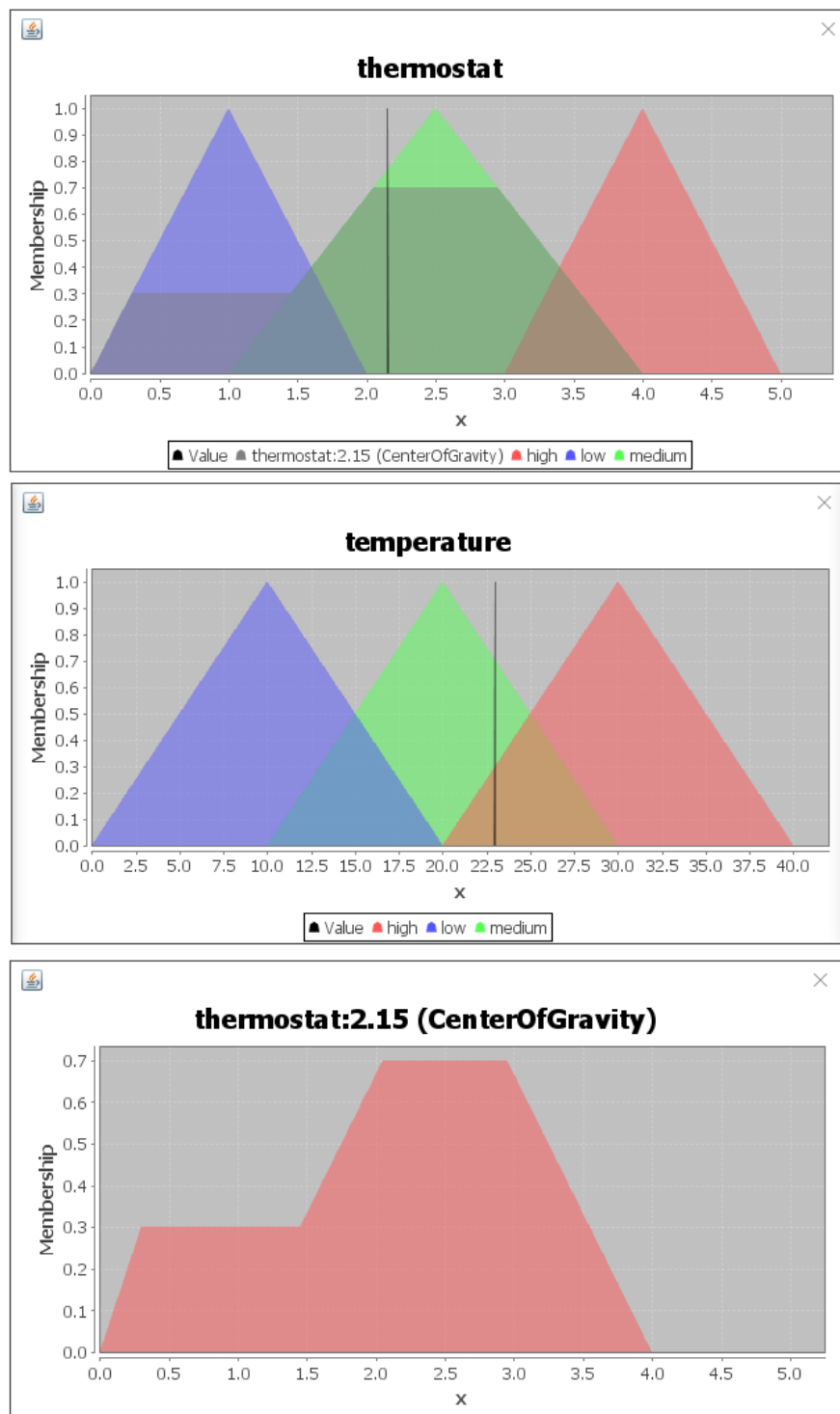


Abbildung 2.4.: Ein Beispiel für Fuzzylogik

Algorithm 1: Regel für Raumtemperatur und Einstellung des Thermostates.

```
if temperature IS low then  
  | thermostat IS high;  
end  
if temperature IS medium then  
  | thermostat IS medium;  
end  
if temperature IS high then  
  | thermostat IS low;  
end
```

für Java, Python, Ruby, Mathlab... entwickelt. OpenCV wurde für Recheneffizienz und mit einem starken Fokus auf Echtzeitanwendungen entwickelt. Diese Bibliothek enthält über 500 Funktionen, die viele Bereiche in der Vision einschließlich Bildverarbeitung, maschinelles Lernen, neuronale Netze... umfassen [BK08]. Da die Arbeit um ein Echtzeitanwendung zur Erkennung der abnormalen Situationen geht, deswegen kommt OpenCV hier zum Einsatz.

2.5. Bosch Innenkamera

Für diese Masterarbeit wurden die Videos mit eine Bosch Innenkamera aufgenommen. Die Bosch 360 Innenkamera kann in jede Richtung drehen und schauen. Die im Innenbereich angeschlossene Kamera umfasst Bewegungssensoren, eine Gegensprechanlage für Zwei-Wege-Audio und Infrarot-Nachtsicht, so dass Sie zu jeder Zeit über Ihr Haus schauen können. Wenn es etwas wahrnimmt, zeichnet es HD-Material auf dem lokalen Speicher auf und Sie können mit Ihrem Telefon Clips oder Live-Filmmaterial ansehen.



Abbildung 2.5.: Bosch 360 Grad Kamera (Quelle: www.bosch-smarthome.com)

3. Eigenes Verfahren

Diese Arbeit besteht aus drei große Hauptschritte, nämlich Hintergrundsubtraktion, Schätzung der Körperhaltung mit Histogrammanalyse und Erkennung außergewöhnlicher Situation mit Fuzzylogik

3.1. Hintergrundsubtraktion

3.1.1. Vergleichen verschiedener Methoden

In dieser Masterarbeit ist Hintergrundsubtraktion ein wichtiger Baustein. Ein großes Problem ist wie ein korrekter Hintergrund erstellt werden kann, damit man sich bewegtes Objekt robust erkennen kann. Für diese Arbeit wurden verschiedenen moderne Methoden (Gaussian Mixture Model, Kern Density Estimation und Vibe) ausprobiert und verglichen. Die alle genannte Methoden wurden schon in 2.1 theoretisch beschrieben. In diesem Abschnitt geht es um Bewertungen von Hintergrundsubtraktionsverfahren. Bei GMM wird eine Mischung von 5-Gaußschen Verteilung modelliert und eine Anzahl von 100 letzte Frames wird als "History" angewendet. Mit dem Lernrate 0.01 funktioniert das GMM Verfahren gut in unserem Test. Für das AGMM nehmen wir auch die gleichen Parameters, um den Vergleich der Hintergrundsubtraktionsverfahren objektiv zu bewerten. Bei KDE handelt es sich um eine Berechnung von Intensitätswerte für einen Pixel, deshalb wird nur ein Parameter von 100 als "History" wie bei AGMM in diesem Fall eingesetzt. Wie in [BVD09] schon gemeint, Vibe ist ein nicht-parametrisches Verfahren und weshalb kein Parameter wird hier gebraucht.

Aus dieser Grafik 3.1 wird deutlich, dass AGMM und Vibe ziemlich besser als die zweit andere. Hier ist kritisch anzumerken, dass Vibe und KDE längere Verarbeitungszeit als den Rest brauchen. Das erste Ziel der Arbeit ist die Untersuchung der Hintergrundsubtraktion, damit eine beste Silhouette erkannt werden kann. Anhand des Schaubildes kann man erkennen, dass AGMM ein guter Kandidat für Hintergrundsubtraktion ist. Viele Videos in verschieden Orte mit verschiedene Körperhaltung wurden für diese Arbeit aufgenommen, damit die Qualität der Software gut geprüft wird.

Die oben genannte Hintergrundsubtraktionsverfahren basieren auf die Änderung des Intensitätswerten von jedem Pixel, deshalb liefern die Verfahren ein Schwarzweißbild mit viel Rausch zurück. Um die Rausch zu entfernen, wird Erosion und Dilatation zum Einsatz gebracht, die zwei Verfahren sind zwei grundlegende Operationen bei der morphologischen Bildverarbeitung, auf der alle anderen morphologischen Operationen basieren. Unter Erosion kann man verstehen ein Verfahren, das ein Bild mit einer einfachen, vordefinierten Form untersucht und daraus Schlussfolgerungen zieht, wie diese Form in die Formen im Bild verfehlt. Und Dilatation ist ein umgekehrtes Verfahren von Erosion, es

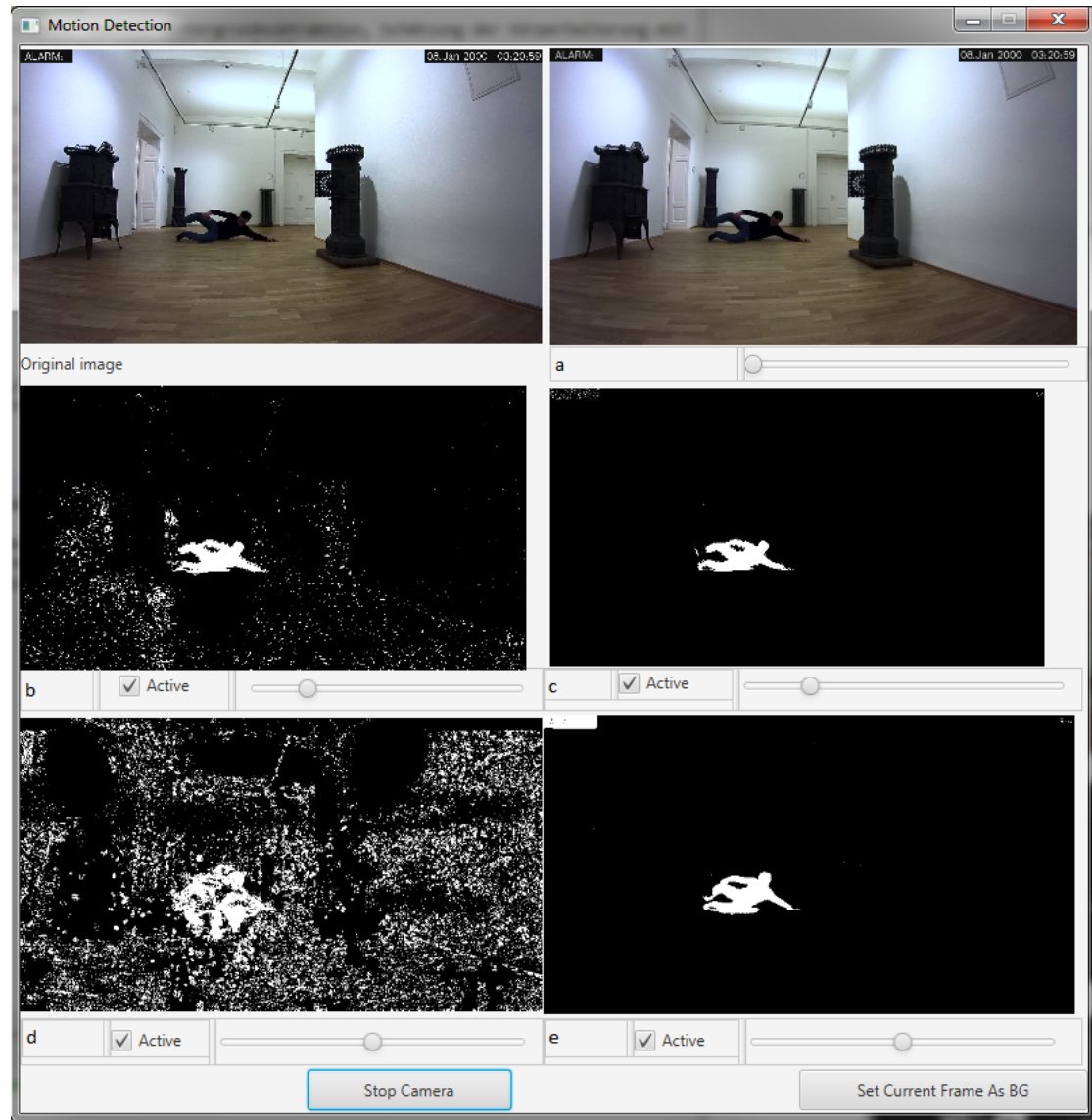


Abbildung 3.1.: Vergleichen von verschiedene Techniken an Hintergrundsubtraktion (a) Weichzeichnen (b) GMM (c) AGMM. (d) KDE (e) Vibe

versucht ein Bild zu erweitern mit passender Form von originalem Bild. Eingaben für die zwei Verfahren sind ein originales Bild A und einen Filter B . Wenn das strukturierende Element B ein Zentrum hat, kann die Erosion von A durch B als der Ort von Punkten verstanden werden, die durch das Zentrum von B erreicht werden, wenn B sich in A bewegt. Angenommen, dass der Ursprung B in seiner Mitte ist, überlagert für jedes Pixel in A der Ursprung von B , wenn B vollständig in A enthalten ist, wird das Pixel bei Erosion beibehalten, ansonsten gelöscht. Aber bei Dilatation überlagert jeder Pixel in A der Ursprung von B und der Pixel wird in der Erweiterung von A und B enthalten. Um die zwei Operationen besser zu verstehen, wurden zwei Beispiele für Erosion und Dilatation in Abbildungen 3.2 und 3.3 dargestellt.

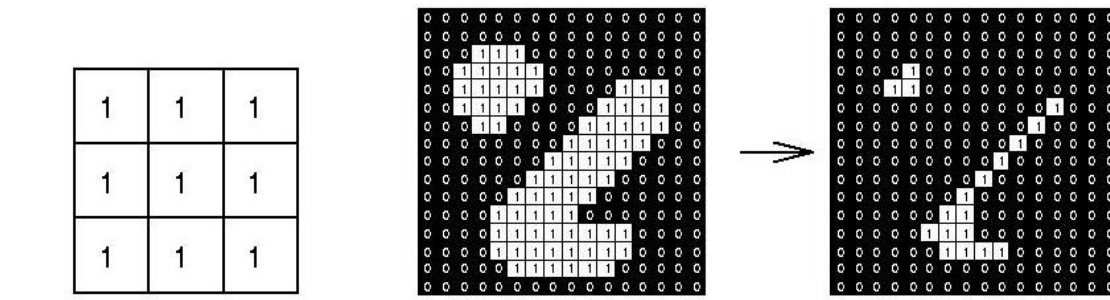


Abbildung 3.2.: Erosion

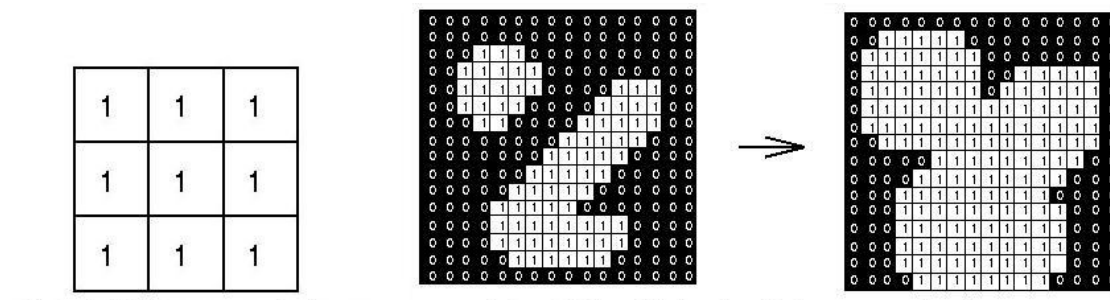


Abbildung 3.3.: Dilatation

Erosion wird in diesem Fall angewendet, um Rausch nach dem Hintergrundsubtraktion zu entfernen. Es folgt auch eine kleine Änderung an der Silhouette. Um die genannte Änderung zu vermeiden wird Dilatation hier genutzt. Die Abbildung 3.4 zeigt, dass nach Erosion unerwartete Rausch außer Silhouette entfernt ist und auch die Silhouette leicht geschädigt ist. Aber nach der Dilatation wird die Silhouette besser vervollständigt.

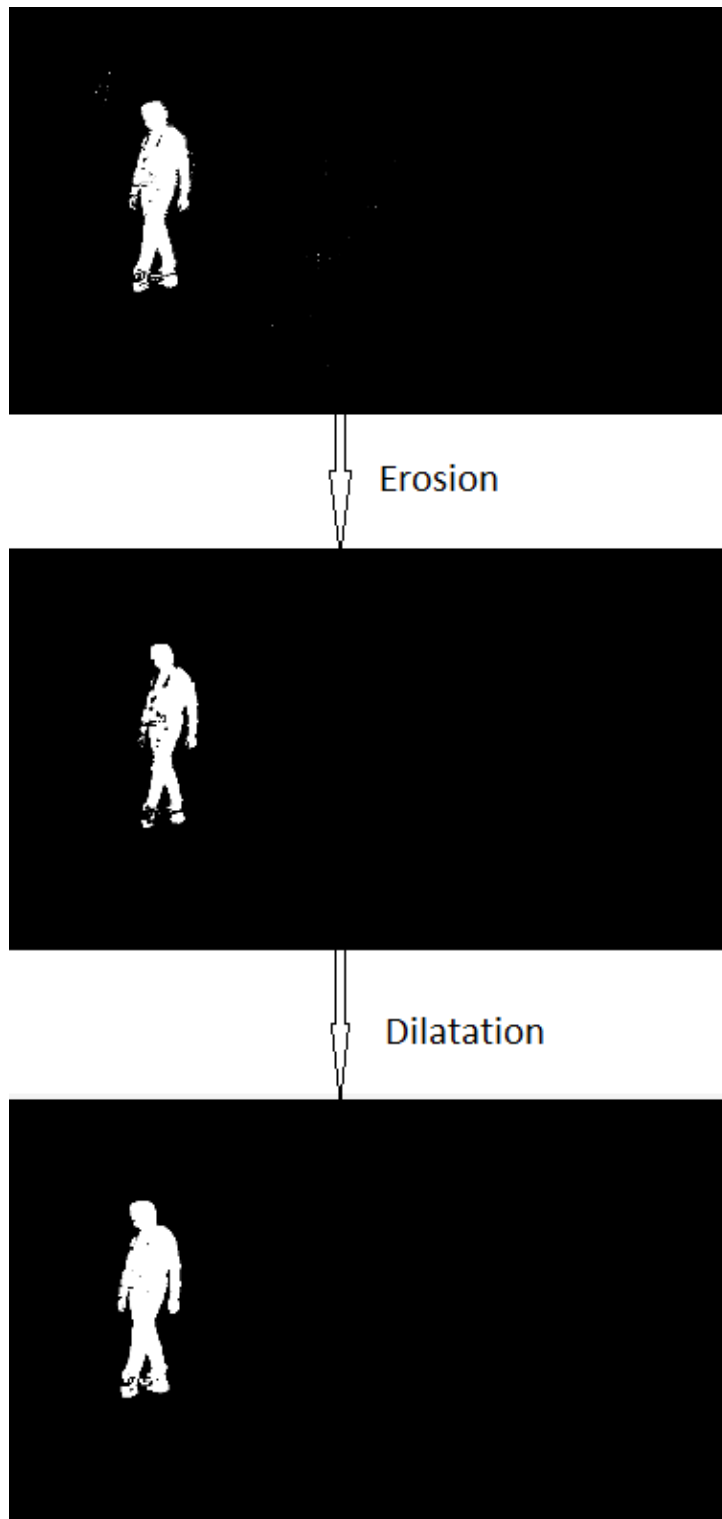


Abbildung 3.4.: Ergebnis nach Erosion und Dilatation

3.1.2. Eigene Methode

3.1.3. Ergebnisse der eigene Hintergrundsubtraktion

**3.2. Schätzung der Körperhalterung mit
Histogrammanalyse**

Im Abschnitt 3.1

3.3. Erkennung außergewöhnlicher Situation

4. Ergebnisse und Evaluation

In diesem Kapitel sollen die Ergebnisse dieser Arbeit diskutiert werden.

5. Zusammenfassung und Ausblick

In diesem Kapitel sollen zunächst die erreichten Ziele diskutiert und abschließend ein Ausblick auf mögliche, weiterführende Arbeiten gegeben werden.

Literaturverzeichnis

- [AMFT11] AZIZ K.-E., MERAD D., FERTIL B., THOME N.: Pedestrian head detection and tracking using skeleton graph for people counting in crowded environments. In *MVA* (2011), pp. 516–519.
- [BJEL10] BENEZETH Y., JODOIN P.-M., EMILE B., LAURENT ROSENBERGER C.: Comparative study of background subtraction algorithms. *Journal of Electronic Imaging* 19, 3 (2010), 033003–033003.
- [BK08] BRADSKI G., KAEHLER A.: *Learning OpenCV: Computer vision with the OpenCV library*. Ö'Reilly Media, Inc., 2008.
- [BP12] Bhole C., PAL C.: Automated person segmentation in videos. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (2012), IEEE, pp. 3672–3675.
- [BTVG06] BAY H., TUYTELAARS T., VAN GOOL L.: Surf: Speeded up robust features. *Computer vision–ECCV 2006* (2006), 404–417.
- [BVD09] BARNICH O., VAN DROOGENBROECK M.: Vibe: a powerful random technique to estimate the background in video sequences. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (2009), IEEE, pp. 945–948.
- [BVD11] BARNICH O., VAN DROOGENBROECK M.: Vibe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image processing* 20, 6 (2011), 1709–1724.
- [DG13] DUFFNER S., GARCIA C.: Pixeltrack: a fast adaptive algorithm for tracking non-rigid objects. In *Proceedings of the IEEE international conference on computer vision* (2013), pp. 2480–2487.
- [Din11] DINGLE N.: Artificial intelligence: Fuzzy logic explained. *Retrieved October 6* (2011), 2014.
- [EHD00] ELGAMMAL A., HARWOOD D., DAVIS L.: Non-parametric model for background subtraction. *Computer Vision-ECCV 2000* (2000), 751–767.
- [FW06] FLEET D., WEISS Y.: Optical flow estimation. In *Handbook of mathematical models in computer vision*. Springer, 2006, pp. 237–257.

- [GM06] GUO P., MIAO Z.: Projection histogram based human posture recognition. In *Signal Processing, 2006 8th International Conference on* (2006), vol. 2, IEEE.
- [GRB13] GODEC M., ROTH P. M., BISCHOF H.: Hough-based tracking of non-rigid objects. *Computer Vision and Image Understanding* 117, 10 (2013), 1245–1256.
- [HHD98] HARITAOGLU I., HARWOOD D., DAVIS L. S.: Ghost: A human body part labeling system using silhouettes. In *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on* (1998), vol. 1, IEEE, pp. 77–82.
- [KB02] KAEWTRAKULPONG P., BOWDEN R.: An improved adaptive background mixture model for real-time tracking with shadow detection. *Video-based surveillance systems 1* (2002), 135–144.
- [KHD05] KIM K., HARWOOD D., DAVIS L. S.: Background updating for visual surveillance. In *International Symposium on Visual Computing* (2005), Springer, pp. 337–346.
- [MPWSH16] MÄRKI N., PERAZZI F., WANG O., SORKINE-HORNUNG A.: Bilateral space video segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 743–751.
- [Pic04] PICCARDI M.: Background subtraction techniques: a review. In *Systems, man and cybernetics, 2004 IEEE international conference on* (2004), vol. 4, IEEE, pp. 3099–3104.
- [PS02] POWER P. W., SCHOONEES J. A.: Understanding background mixture models for foreground segmentation. In *Proceedings image and vision computing New Zealand* (2002), vol. 2002, pp. 10–11.
- [Ros71] ROSENFELD A.: Fuzzy groups. *Journal of mathematical analysis and applications* 35, 3 (1971), 512–517.
- [SG99] STAUFFER C., GRIMSON W. E. L.: Adaptive background mixture models for real-time tracking. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)* (1999), vol. 2, p. 252 Vol. 2.
- [ZVDH06] ZIVKOVIC Z., VAN DER HEIJDEN F.: Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters* 27, 7 (2006), 773–780.

Abkürzungsverzeichnis

KDE kernel density estimation

ROI region of interest

ALL assisted ambient living

GMM Gaußschen Mixture Modell

AGMM Adaptive Gaußschen Mixture Modell

Abbildungsverzeichnis

2.1. Ein Beispiel für Hintergrundsubtraktion	6
2.2. Klassifizierung von $p_t(x)$ basiert auf Überschneidung von Kugel $S_R(p_t(x))$ mit Menge der Stichproben [BVD09].	9
2.3. Normalisierte horizontale und vertikale Projektionsvorlagen für jede Körperhalterung [HHD98].	9
2.4. Ein Beispiel für Fuzzylogik	11
2.5. Bosch 360 Grad Kamera (Quelle: www.bosch-smarthome.com)	13
3.1. Vergleichen von verschiedene Techniken an Hintergrundsubtraktion (a) Weichzeichnen (b) GMM (c) AGMM. (d) KDE (e) Vibe	15
3.2. Erosion	16
3.3. Dilatation	16
3.4. Ergebnis nach Erosion und Dilatation	17

Tabellenverzeichnis

List of Algorithms

1. Regel für Raumtemperatur und Einstellung des Thermostates. 12

Listings

A. Anhang

Thema 1

Beispiel für einen Anhang

Thema 2

Danksagung

Hiermit möchte ich mich besonders bei Prof. Dr. XXXX, Prof. Dr. xxxx, Dipl. Inf. xxxx und Dipl. Inf. xxx für die Betreuung meiner Arbeit, hilfreiche Diskussionen und viel Geduld bei zahlreichen Fragen bedanken.