

Dynamische Raumerfassung auf Basis einer Time-of-Flight Kamera

Jan Papadoudis; Prof. Anthimos Georgiadis
 Leuphana Universität Lüneburg, Volgershall 1, 21339 Lüneburg
Jan.papadoudis@leuphana.de; georgiadis@uni.leuphana.de

Abstract

Die Raumerfassung ist ein wichtiger Bestandteil um autonome Fahrzeuge realisieren zu können, dabei ist es notwendig Hindernisse unabhängig von deren Art zu erkennen, damit diese Fahrzeuge sich in einer neuen Umgebung bewegen können. Eine schnelle räumliche Darstellung einer Szene, lässt sich mit Hilfe einer Time-of Flight (ToF) Kamera realisieren, da sie für jeden Pixel eine Distanzinformation liefert und sich daher ideal für die Raumerfassung eignet. Traditionelle Methoden der Bildverarbeitung um Objekte zu erkennen, lassen sich jedoch kaum auf die Daten einer ToF-Kamera anwenden. Gradientenverläufe der Distanz lassen sich nur selten sicher verwenden, da ein grundsätzlicher Verlauf durch die Tiefe des Raumes immer gegeben ist und z.B. Hindernisse auf dem Boden keine signifikante Änderung der Distanz an dem Übergang, bedingen. Es wurde daher ein Verfahren entwickelt, welches zuverlässig die grundlegenden Ausmaße und Gegebenheiten eines Raumes erkennt und Hindernisse detektiert. Abbildung 1 zeigt eine Szene aufgenommen mit einer ToF-Kamera

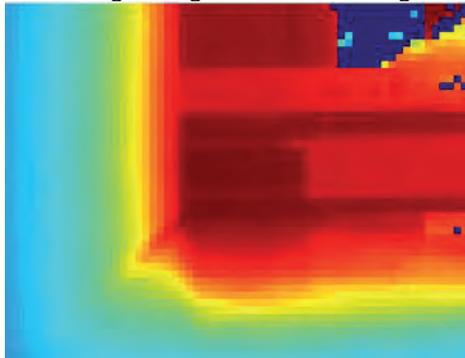


Abbildung 1: Rohdaten der ToF-Kamera.

Es ist deutlich zu erkennen, dass eine exakte Abgrenzung der Objekte nur schwer zu erkennen ist, insbesondere der Übergang von der Wand zum Boden. Durch die Distanzinformation ist es ohne großen Aufwand möglich die Pixelkoordinaten in ein Weltkoordinatensystem relativ zur Kameraposition umzurechnen. Dies ermöglicht die Verwendung der Vektorrechnung. Für jeden Pixel wird eine Ebene ermittelt, die durch die, ihn unmittelbar umgebenden Pixel definiert ist. Für jede der drei, das Kamerakoordinatensystem aufspannenden Achsen wird ein Schnittwinkel zur ermittelten Ebene bestimmt.

Die Schnittwinkel zu allen Achsen geben einen Anhaltspunkt zur Abgrenzung verschiedener Objekte. Anschließend werden noch zusammenhängende Flächen ähnlicher Schnittwinkel untersucht um kleinere Objekte zu identifizieren und das Ergebnis zu verbessern. Dies kann mit einer herkömmlichen Gradientenuntersuchung realisiert werden. Die bearbeitete Aufnahme ist in der nachfolgenden Abbildung dargestellt.

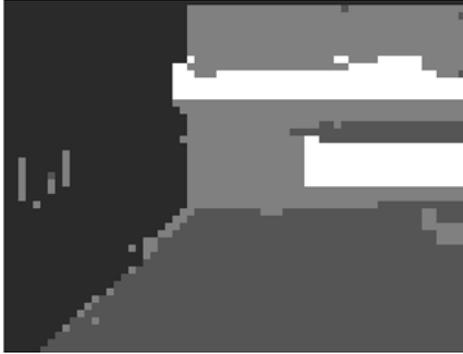


Abbildung 2: Bearbeitetes Bild mit abgegrenzten Objekten.

Es ist deutlich zu erkennen, dass die einzelnen Objekte klar voneinander abgegrenzt sind. Mit diesem Verfahren kann eine grundlegende dynamische Raumerfassung mit Hinderniserkennung realisiert werden, die es z.B. autonomen Systemen ermöglicht im Raum zu navigieren. In dieser Arbeit werden das mathematische Modell, das zugehörige Programm und die ersten Ergebnisse vorgestellt.

Einleitung

Bei autonomen Fahrzeugen, im speziellen bei fahrerlosen Transportsystemen, gibt es verschiedene Ansätze der Navigation im Raum. Am häufigsten vertreten ist die Navigation mittels Laser und Reflexmarken und die Navigation mit Hilfe von Magnetpunkten. Für diese Arten der Navigation sind jedoch Änderungen an der Lagerhalle nötig, in der das fahrerlose Transportsystem fahren soll. Somit lassen sich diese Methoden nicht ohne Probleme auf andere Anwendungsgebiete übertragen.

Eine kamerabasierte Navigation würde keinerlei Veränderungen an der Umgebung erfordern und wäre somit überall auf jedem System einsatzbereit. Die Entwicklung, des in diesem Artikel vorgestellten Systems, basiert auf der Anwendung bei fahrerlosen Transportsystemen, die in jeder Halle ohne Umbauten eingesetzt werden können. Die Verwendung einer Kamera hat bei dieser Anwendung noch weitere Vorteile. In vielen Fahrzeugen ist bereits eine Kamera integriert, die für andere Aufgaben genutzt wird, wie z.B. das Erkennen der Position einer Palette. Somit würde zur Benutzung keine zusätzliche Sensorik gebraucht, was die Kosten senkt. Ebenfalls wäre das vorgestellte System dazu in der Lage Sicherheitsaspekte zu erfüllen, wie den Personenschutz. Zusätzliche Sensorik wäre dafür dann ebenfalls nicht notwendig.

Eine kamerabasierte Navigation würde also im Bereich der fahrerlosen Transportsysteme einen Mehrwert an Flexibilität bringen und die Anschaffungskosten senken.

Das Bildverarbeitungssystem der Kamera muss dazu in der Lage sein den befahrbaren Bereich einer Lagerhalle von Objekten zu unterscheiden. Für die grundsätzliche Funktionalität ist es dabei nicht notwendig, dass die erkannten Objekte exakt klassifiziert werden.

Für die sichere Navigation muss die Größe und Position der Objekte genau erfasst werden können, damit das Fahrzeug sicher um das Objekt navigieren kann bzw. es zu keiner Kollision kommt. Das System muss dementsprechend in der Lage dazu sein Größen in einer unbekanntem Umgebung zu bestimmen bzw. Größen von unbekanntem Objekten. Das verwendete Kamerasystem ist eine Time-of-Flight (ToF) Kamera, die in der Lage ist Distanzen zu erfassen.

Verwendete Sensorik

Das Messprinzip der ToF Kamera basiert auf dem Prinzip der Lichtlaufzeitmessung. Die Kamera besteht prinzipiell aus einem CMOS/CCD - Chip und einem LED-Array mit LED's, die Licht im infraroten Bereich (z.B. 850 nm) aussenden. Das ausgesandte Signal der LED's, welches z.B. sinusförmig ist mit einer definierten Amplitude, beleuchtet die Szene und wird von dieser reflektiert. Das reflektierte Signal wird von dem CMOS/CCD – Chip gemessen. Aufgrund der zurückgelegten Distanz des Signals ist die reflektierte Welle, relativ zur ausgesandten, verschoben. Diese Verschiebung ist eine Phasenverschiebung und ist proportional zur zurückgelegten Distanz. Es gilt für die Phasenverschiebung φ :

$$\varphi = \frac{4\pi f}{c} * d$$

Wobei f die Frequenz des ausgesandten Signals ist, c die Lichtgeschwindigkeit und d die Distanz.

Das empfangene Signal kann ebenfalls hinsichtlich der Amplitude untersucht werden. In Abhängigkeit der Distanz und vor allem abhängig vom Reflektions – und Absorptionsvermögen der Objekte der Szene ist die Amplitude im Verhältnis zur ausgesandten Amplitude abgeschwächt. Die Kamera liefert also mit einer Aufnahme zwei Bilder zurück. Das Phasenbild, also das Distanzbild und ein Amplitudenbild, was einem Intensitätsbild einer herkömmlichen Kamera ähnelt.

Da die Kamera für die Distanz (Phase) und die Amplitude für jeden Pixel einzeln bestimmen kann bei jeder Aufnahme eignet sich die Kamera sehr gut für dynamische Anwendungen bei denen die Distanz von Objekten ermittelt werden muss.

Als alternative Systeme zur Distanzermittlung bieten sich die Triangulation mit Hilfe von zwei 2D-Kameras an und ein Laserscanner. Das Verfahren der Triangulation ist sehr störanfällig, da die verwendeten Kameras sehr genau justiert werden müssen und auch robust gegenüber Vibrationen sein müssen, sodass die beiden Kameras immer in der selben relativen Position sind. Ebenfalls ist durch die Zusammenführung der Daten beider Kameras mit einem erhöhten Rechenaufwand zu rechnen. Je nach Auslegung des verarbeitenden Systems kann das Bildverarbeitungssystem dann nicht mehr für Sicherheitskritische Anwendungen genutzt werden. Ein System basierend auf einem Laserscanner bietet sehr genaue Distanzdaten ist jedoch sehr langsam, da die Szene zeilenweise abgescannt wird und daraus das gesamte Bild generiert wird. Es ist also keine Momentanaufnahme wie bei einer ToF-Kamera möglich.

Daher bietet eine ToF-Kamera eine ideale Basis für das Bildverarbeitungssystem zur Bewältigung der Messaufgabe. In der folgenden Abbildung ist die verwendete Kamera dargestellt, die für die Versuche benutzt worden ist.



Abbildung 3: O3D201 PMD 3D Kamera von IFM.

Die Kamera besitzt eine Auflösung von 64 x 50 Pixeln und kann mit einer Frequenz von bis zu 20 Hz Bilder aufnehmen.

Messprinzip

Die dynamische Raumerfassung soll dazu dienen, dass auf der Strecke der Boden von den Wänden separiert werden kann, um dem Weg folgen zu können, sowie Hindernisse zu erkennen, die sich auf der Fahrbahn befinden, um einen Zusammenstoß zu vermeiden. Dabei soll auf das Distanzbild der ToF-Kamera zurückgegriffen werden. Durch die Distanzinformationen aus den Daten ist es möglich, Objekten eine genaue Größe und Position im Raum zuzuordnen. Bei herkömmlichen Kameras ist dies nur sehr eingeschränkt möglich. Abbildung 4 zeigt das Distanzbild einer Ecke eines Raumes. Zur besseren Orientierung wurde die selbe Szene in Abbildung 5 noch einmal mit einer herkömmlichen 2D-Kamera aufgenommen.

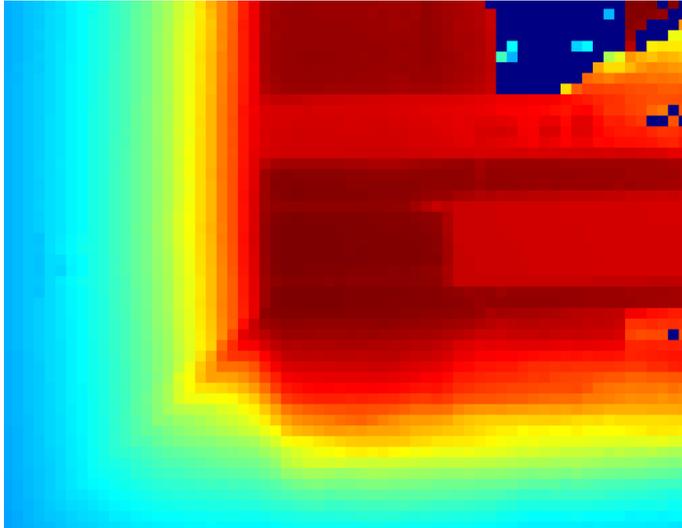


Abbildung 4: Dieses Bild zeigt das Distanzbild eines Raumes. Die Szene ist der Blick in die Ecke eines Raumes mit Sicht auf eine Fensterbank und einen Heizkörper.



Abbildung 5: Die Szene aus der vorherigen Abbildung aufgenommen mit einer herkömmlichen 2D-Kamera.

Diese Szene eignet sich gut als Anschauungsbeispiel für das Verfahren der Raumerfassung, da in dieser Szene der befahrbare Boden sowohl von einer Seitenwand begrenzt wird, als auch von der Wand unter dem Fenster, die gleichzeitig als Hindernis dient. Die Standardmethode, die angewandt wird, um Objekte zu erkennen und vom Hintergrund abzutrennen ist eine Kantendetektion basierend auf Grauwertänderungen bzw. im Falle eines Distanzbildes die Änderung der Distanz. Dazu werden die Gradienten in x – bzw- y – Richtung ermittelt und miteinander verrechnet. Dadurch ergibt sich das Gradientenbild f durch:

$$f(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2}$$

Abbildung 4.10 zeigt das berechnete Gradientenbild. Die Abgrenzungen sind schemenhaft zu erkennen, jedoch ist es sehr schwer einen geeigneten Grenzwert zu finden, sodass der Boden von der Wand getrennt wird ohne, dass es innerhalb des Bodens zu Abgrenzungen kommt. In Abbildung 4.10 ist der Boden grob gelb umrandet und die Wand auf der linken Seite in rot. Der Rest des Bildes ist die Rückwand des Raumes mit Fensterbank und Heizkörper. Es ist deutlich zu erkennen, dass es innerhalb der Fläche des Bodens und der Wand einen Verlauf des Gradienten gibt, dies ist dadurch begründet, dass nicht jeder Pixel dieselbe Flächengröße abbildet. Außerdem ist zu erkennen, dass beim Übergang des Bodens zu der Seitenwand nur ein sehr geringer Unterschied in den Gradienten auszumachen ist. Dies führt dazu, dass ein Schwellwert zur Bestimmung der Kante sehr genau bestimmt werden muss, was bei sich ändernden Bedingungen und Szenen nicht zu realisieren ist.

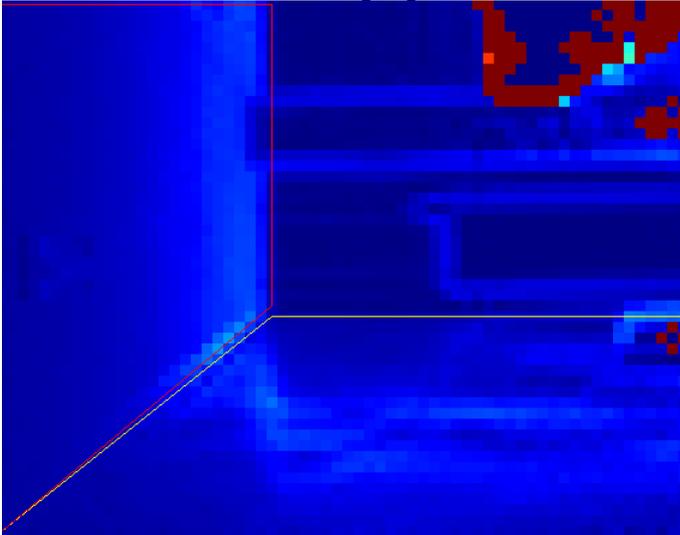


Abbildung 6: Gradientenbild der ausgewählten Szene.

Der Heizkörper, in der rechten Mitte des Bildes, sorgt darüber hinaus noch für Streulicht, welches die Distanz vor dem Objekt etwas verfälscht. Diese, durch Streulicht bedingte, Veränderung der Distanz ist zwar nur gering, führt aber dazu, dass auf dem Boden vor dem Objekt ebenfalls Gradientensprünge zu erkennen sind, die vom Wert her sogar über dem Unterschied beim Übergang vom Boden zur Seitenwand liegen. Somit ist ein globaler Schwellwert für den Gradienten oder auch die Änderung des Gradienten, kein geeignetes Maß für die Selektion der Objekte, sondern kann wenn überhaupt unterstützend wirken.

Es muss also eine Möglichkeit gefunden werden, mindestens den Boden von den Wänden zu separieren ohne dabei nur auf den Gradienten zurückzugreifen und relativ unabhängig vom Streulicht zu sein.

Das Distanzbild ermöglicht es für jeden Pixel genau die abgebildete Fläche zu berechnen.

und somit auch die Ausdehnung eines Objektes. Sei α_x der Öffnungswinkel in x-Richtung und P_x die Anzahl der Pixel, dann ist die Größe eines Pixels in x-Richtung $SP_x(d)$ in Abhängigkeit der gemessenen Distanz d definiert durch:

$$SP_x(d) = \tan\left(\frac{\alpha_x}{2}\right) * \frac{2d}{P_x}$$

Somit lässt sich für jeden Pixel eine genaue Position im Raum mit den Koordinaten (x, y, z) erstellen. Auf dieser Basis kann ein Algorithmus entwickelt werden, der, weitestgehend unabhängig vom Gradienten, Objekte erkennt. Der gesamte Algorithmus, der im Folgenden genauer erläutert wird.

Im ersten Schritt wird das Distanzbild vom Rauschen befreit. Dies geschieht mit einem „Gauß – Filter“, der vergleichbar einem Tiefpassfilter in der Elektrotechnik wirkt. Der Gauß – Filter basiert auf der Glockenkurve und gewichtet die Pixel in der Umgebung nicht gleichmäßig, sodass ein „weicherer“ Verlauf entsteht. Zur Filterung des Bildes wird die Filtermaske mit dem Bild gefaltet. Das Ergebnis entspricht einem geglätteten Bild. Eine einfache Möglichkeit die Einträge der Filtermaske zu berechnen bietet das Pascalsche Dreieck, in dem z.B. die 4. Zeile des Dreiecks als erste Zeile und Spalte der Maske dient, die restlichen Werte errechnen sich aus dem Produkt aus der jeweiligen Wertes in der 1. Zeile und 1. Spalte.

Da bei diesem Verfahren die Werte der Bildmatrix erhöht werden, müssen sie anschließend zur Normierung wieder durch die Summe der gesamten Filtermaske geteilt werden.

Zur späteren Verifizierung wird anschließend ein das Gradientenbild, wie beschrieben, erstellt.

Als Hauptkriterium für das Separieren von Objekten soll auf den Winkel zwischen der Ebene bzw. dem Normalenvektor der Ebene, der durch den Pixel und seine nähere Umgebung bestimmt wird, und den verschiedenen Achsen, des durch die Kameraausrichtung definierten Koordinatensystems, zurückgegriffen werden. Als Ebene wird jeweils die 3×3 – Matrix M_e um einen Pixel p_{mn} mit diesem als Mittelpunkt, behandelt. Zur Bestimmung einer Ebene werden drei Punkte benötigt, diese seien $p_1 = M_e(1,1)$, $p_2 = M_e(3,1)$ und $p_3 = M_e(2,3)$. Um eine Ebene berechnen zu können, ist es notwendig die Pixelkoordinaten in Weltkoordinaten (x,y,z) umzurechnen. In dem Distanzbild D liegen für jeden Pixel Distanzinformationen vor, diese können als z-Koordinaten direkt übernommen werden. Die Spalte eines Pixels soll als x – und die Zeile als y -Koordinate dienen. Womit das Kamera-Koordinatensystem hinreichend definiert ist. Die Pixelkoordinaten, also Zeilen – und Spaltenindex in D , müssen noch in die entsprechenden Weltkoordinaten umgerechnet werden. Somit sind die Ortsvektoren der zu untersuchenden Punkte p_1 , p_2 und p_3 , wenn $p_{mn} = D(m,n)$ der Referenzpixel ist:

$$\begin{aligned}\vec{p}_1 &= \begin{pmatrix} (n-1) * SP_x \\ (m-1) * SP_y \\ D(m-1, n-1) \end{pmatrix} \\ \vec{p}_2 &= \begin{pmatrix} (n-1) * SP_x \\ (m+1) * SP_y \\ D(m+1, n-1) \end{pmatrix} \\ \vec{p}_3 &= \begin{pmatrix} (n+1) * SP_x \\ m * SP_y \\ D(m, n+1) \end{pmatrix}\end{aligned}$$

Zur Berechnung der Ebene wird \vec{p}_1 als Stützvektor verwendet und die Spannvektoren ergeben sich dann durch:

$$\begin{aligned}\vec{u} &= \vec{p}_2 - \vec{p}_1 \\ \vec{v} &= \vec{p}_3 - \vec{p}_1\end{aligned}$$

Der Normalenvektor \vec{n} errechnet sich dann mit $\vec{n} = \vec{u} \times \vec{v}$. Durch den Vergleich mit den kanonischen Einheitsvektoren, ist es möglich eine Aussage über den Winkel der Ebene im Raum zu tätigen. Sei \vec{e} ein beliebiger kanonischer Einheitsvektor, dann lässt sich der Winkel φ zwischen dem Normalenvektor und dem Einheitsvektor durch

$$\cos \varphi = \frac{\vec{n} * \vec{e}}{|\vec{n}| * |\vec{e}|}$$

berechnen.

Der Wert φ wird dann für jede der drei Achsen dem Pixel p_{mn} zugeordnet. Somit entstehen drei neue Matrizen, die als Werte die Winkel zu der entsprechenden Achse beinhalten. Diese Matrizen dienen nun dazu, zusammenhängende Flächen zu ermitteln. Sofern die Oberfläche des Objektes nicht stark gewölbt ist, was beim Boden und einer Wand in der Lagerhalle ausgeschlossen ist, ändern sich die Winkel innerhalb eines Objektes nur geringfügig. Es gilt jedoch zu beachten, dass im \mathbb{R}^3 mehrere Vektoren zueinander z.B. einen 90° Winkel besitzen. Der Normalenvektor der linken Wand, sowie der Normalenvektor des Bodens haben eine Differenz von 90° zu dem Vektor, der die Z-Achse darstellt. Es ist also erforderlich eine Drehrichtung der Vektoren definieren zu können. Dies ist am einfachsten zu realisieren, wenn man die Winkel des Normalenvektors mit den Winkeln zu allen drei Achsen untereinander vergleicht.

In Abbildung 7 ist das verarbeitete Bild aus Abbildung 4 dargestellt. Die einzelnen Flächen heben sich deutlich voneinander ab, sodass die Objekte genau bestimmt werden können. Zur Bestimmung der Objekte können verschiedene Alternativen gewählt werden.



Abbildung 7: Dieses Bild zeigt die extrahierten Winkel. Die Flächen sind deutlich voneinander abgegrenzt.

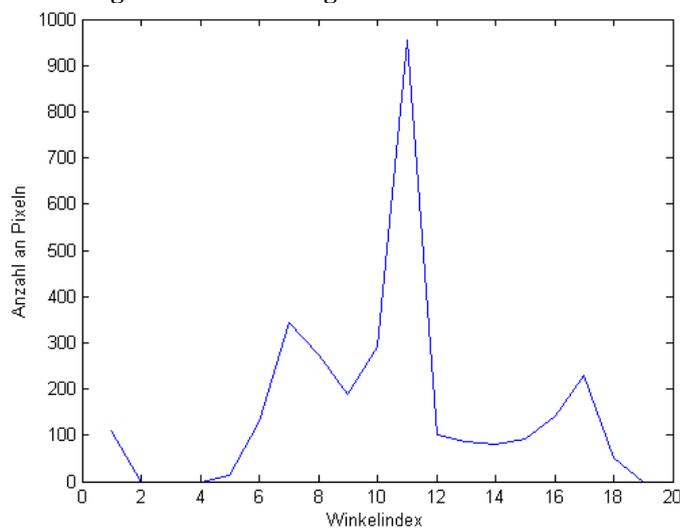


Abbildung 8: Histogramm der Winkelindizes.

Die größte Methode wäre es, auf Basis der Winkel zu den verschiedenen Achsen die verschiedenen Flächen direkt zu labeln. Eine Methode dies zu realisieren ist es, ein Histogramm der Werte aus Abbildung 7 zu erstellen. Das Histogramm zeigt auf der X-Achse den Wert, in diesem Fall einen Winkelindex, und auf der Y-Achse die Anzahl der Pixel die diesen Wert haben. Abbildung 8 ist das Histogramm dargestellt. Es ist zu erkennen, dass es drei lokale Maxima gibt, was zu den größten Flächen, der linken Wand, dem Boden und der hinteren Wand, in dem Bild passt. Auf Basis dieses Histogramms können nun die Maxima ermittelt werden. Alle Indizes in der näheren Umgebung eines Maximums werden dann dem Index mit dem entsprechenden Maximum zugeordnet. Auf diese Weise kann das Labeln realisiert werden. Kleinere Objekte, die parallel zu einer Wand stehen werden aber in diesem Fall nicht immer erkannt. Abbildung 9 zeigt das gelabelte Bild. Die Heizung und die Fensterbank, die etwas vor der Wand angebracht sind, werden bei dieser Methode nicht beachtet. Daher eignet sich diese nur in begrenzten Anwendungsfällen, wenn ausgeschlossen werden kann, dass Objekte parallel zu einer Wand stehen, auf die senkrecht geschaut wird.



Abbildung 9: Gelabeltes Bild auf Basis der Winkel der Ebenen zu ihren Achsen.

Für eine genauere Identifizierung kann nun auf den Gradienten bzw. auf eine einfache Kantendetektion zurückgegriffen werden. Hier können jedoch die bereits angesprochenen Probleme auftreten. Eine weitere Methode ist die Differenzierung auf Basis der Distanzen. Hindernisse vor dem Fahrzeug haben immer einen Winkel von ca. 180° zur Kameraebene bzw. einen Anteil, der parallel zur Kameraebene ist. Dementsprechend wird ein Zwischenbild generiert in dem alle Distanzen aufgeführt sind an denen der Pixel einen Winkel von ca. 180° aufweist. Im Folgenden werden dann die Distanzen der Abschnitte untersucht. Dabei werden große und abrupte Änderung der Distanz als Indiz dafür genommen, dass es sich um ein anderes Objekt handelt. Durch diese Methode können Objekte, deren Ausrichtung parallel zur Kameraebene ist, relativ einfach erkannt werden. In Abbildung 10 sind die erkannten Objekte voneinander abgegrenzt worden. Die weißen Flächen spiegeln hierbei die Objekte wieder, die innerhalb der Rückwand gefunden worden sind. Mit diesem Verfahren ist es möglich neben der Heizung ebenfalls die Fensterbank eindeutig zu identifizieren. Somit kann mit dieser Methode eine noch genauere Objekterkennung realisiert werden und das Fahrzeug so navigiert werden, dass Kollisionen vermieden werden.



Abbildung 10: Verarbeitetes Bild. Es wurden ebenfalls Fensterbank und Heizung erkannt.

Ergebnisse

Die Raumerfassung bedient sich der Vektorrechnung und der Distanzinformationen der Kamera. Auf Basis dieser Informationen kann gesichert Auskunft über die Position und Größe eines jeden Pixels gegeben werden. Damit können grundsätzliche Strukturen eines jeden Raumes erkannt werden. Da für die Berechnung des Winkels eine Ebene um den zu untersuchenden Pixel konstruiert wird, werden runde Objekte nicht optimal mit den Winkeln erfasst. Die geringe Auflösung der Kamera erschwert es zusätzlich kleine und runde Objekte zu erkennen. Bei voranschreitender Entwicklung der ToF-Kameras ist eine Verbesserung des Algorithmus zu erwarten. Abbildung 11 zeigt eine andere Szene, die mit Hilfe des Algorithmus erkannt werden soll. Dabei handelt es sich um den Blick auf eine Tür und eine Wand, sowie einen Mülleimer, der vor der Wand positioniert ist. Die Wand und die Tür haben dieselbe Ausrichtung zur Kamera und müssen durch die Distanzüberprüfung voneinander unterschieden werden. Gleiches gilt für den Mülleimer und die Wand. Abbildung 12 ist das Ergebnis des Algorithmus abgebildet. Es ist deutlich zu erkennen, dass die einzelnen Objekte voneinander getrennt worden sind. Das Bildverarbeitungssystem bietet somit eine ideale Grundlage zur dynamischen Raumerfassung.

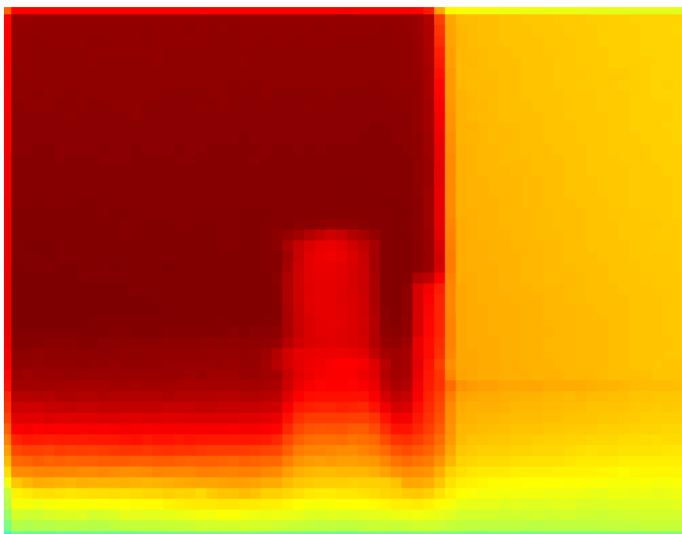


Abbildung 11: Aufnahme mit Blick auf eine Tür und einen Mülleimer.



Abbildung 12: Das vorgestellte System hat die Objekte grundlegend voneinander separiert.