

Bauhaus-Universität Weimar
Fakultät Medien
Studiengang Medieninformatik

Echtzeit Hinderniserkennung für unbemannte Flugsysteme unter Benutzung eines Stereo Kamera Systems

Bachelorarbeit

Hagen Hiller
Geboren am 04.06.1992 in Berlin

Matrikelnummer 110514

1. Gutachter: Prof. Dr. Volker Rodehorst
2. Gutachter: TBA

Datum der Abgabe: 25.01.2016

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Weimar, den 25.01.2016

.....
Hagen Hiller

Zusammenfassung

Lorem ipsum dolor sit Amet.

Inhaltsverzeichnis

1	Einführung	3
1.1	Motivation	3
1.2	Setup	4
1.3	Ziel der Arbeit	4
2	Anforderungsanalyse	5
2.1	Bildaufnahme und Preprocessing	5
2.2	Erfassung von Hindernissen	6
2.3	Performanz	6
2.4	Positionsinformationen	6
3	State of the Art Algorithmen	7
3.1	Aktiv optische Algorithmen	7
3.2	Passiv optische Algorithmen	8
4	Zugrunde liegende Konzepte und Algorithmen	11
4.1	Epipolargeometrie	11
4.2	Stereo Matching	12
4.2.1	Klassifikation	13
4.2.2	Block Matching	13
4.2.3	Semi Global Block Matching	14
4.3	mvStereoVision Framework	16
5	Entwickelte Hinderniserkennungen	17
5.1	Mean Disparity Detection	17
5.2	Samplepoint Detection	20
6	Konflikte in der Erkennung und Lösungsansätze	21
6.1	Validierung der Hindernisse	21
6.2	Hindernisgröße	21
6.3	Position Mapping	21

7	Evaluierung	22
7.1	Aufbau des Testsetups	22
7.2	Evaluierung Subimage Detection	22
7.2.1	Robustheit	22
7.2.2	Performanz	22
7.3	Evaluierung Samplepoint Detection	22
7.3.1	Robustheit	22
7.3.2	Performanz	22
8	Diskussion	23
8.1	Anforderungsevaluierung	23
8.2	Gegenüberstellung beider Algorithmen	23
9	Fazit und zukünftige Arbeiten	24
	Literaturverzeichnis	25

Kapitel 1

Einführung

1.1 Motivation

Die Forschung im Bereich der Robotik ist ein immer weiter fortschreitender Prozess. Eine besondere Domäne innerhalb dessen ist die autonome Navigation sowie Steuerung von unbemannten Flugsystemen. Dies findet Anwendung in der Erkundung von unbekannten oder gefahrenträchtigen Gebieten. Dabei müssen nicht nur die physikalischen Eigenschaften der Drohne betrachtet werden sondern auch die Fusion verschiedenster Sensoren.

Ein wichtiges Kriterium in der Entwicklung autonomer Roboter ist die Erkennung von Hindernissen in Echtzeit. Dabei muss jedoch zuerst definiert werden was vom System als potentiell Hindernis erkannt werden soll. Einerseits sind alle Objekte welche sich in der unmittelbaren Nähe des Systems befinden eine Gefahrenquelle, andererseits wird im Falle Kamera basierter Navigation von einer Bewegung in Blickrichtung ausgegangen. Dies schließt Objekte welche sich außerhalb des Sichtfeldes, also neben oder hinter der Kamera befinden, für die Erfassung durch die Algorithmen aus. Auch sehr weit entfernte Objekte sind prinzipiell nicht als Hindernis anzusehen, wobei sich die maximale Gefahrendistanz relativ zur aktuellen Geschwindigkeit verhält. Unter Betrachtung dieser Gesichtspunkte wird ein Hindernis innerhalb dieser Arbeit als ein Objekt definiert welches sich innerhalb eines definierten Distanzbereichs und innerhalb des Sichtfeldes der Kamera befindet.

Die hauptsächliche Anwendung des im Rahmen dieser Arbeit entwickelten Systems zielt auf die autonome Navigation von Unbemannten Flugsystemen ab. Weitere Anwendungsbereiche können sowohl komplexerer als auch einfacherer Natur sein. Prinzipiell ist es möglich die entwickelten Algorithmen im Automobil Bereich zu verwenden um beispielsweise Objekte vor oder Hinter dem

Kraftfahrzeug zu erkennen und die Distanz zu berechnen. Weiterhin ist es vorstellbar grobe kartographische Höhen Klassifikationen vorzunehmen, um Bild basiert eine Höhenkarte geographischer Areale zu erstellen. Das jeweilige Anwendungsszenario richtet sich jedoch dabei nach der verwendeten Hardware.

1.2 Setup

1.3 Ziel der Arbeit

Die Verwendung Stereobild basierter Daten ermöglicht eine simple Berechnung räumlicher Tiefe.

Mithilfe photogrammetrischer Verfahren ist es mögliche räumliche Tiefe aus zweidimensionalen Daten zu berechnen. Dabei gilt die Verwendung Stereobild basierter Daten als ein einfachste Möglichkeit Disparitäten zwischen korrespondierenden Pixeln zu bestimmen um damit einen räumlichen Eindruck der abgebildeten Szenerie zu erhalten. Unter Betrachtung dieser Techniken wurde im Rahmen dieser Arbeit

Vor diesem Hintergrund werden diverse Anforderungen definiert, welche in Kapitel 2 näher beschrieben werden. Die zugrunde liegende Berechnung von *Disparity Maps* basiert auf dem Semi Global Block Matching Algorithmus [[Hirschmuller, 2005]] sowie dem Konzept der Epipolarometrie welche in Kapitel 4 näher erläutert werden. Kapitel 3 gibt einen Einblick in State of the Art Algorithmen zur Hinderniserkennung. Dabei wird sowohl auf Kamera basierte als auch auf Sensor basierte Algorithmen eingegangen. In den Kapiteln 5 und 6 werden die beiden im Rahmen dieser Arbeit entwickelten Algorithmen zur Hinderniserkennung vorgestellt. Um die Echtheit gefundener Hindernisse zu validieren wurde das Verfahren *Frame Skipping* implementiert welches in Kapitel 7 erläutert wird. Ebenfalls werden bestehende Konflikte in der Erkennung sowie praktische Lösungen als auch hypothetische Ansätze zur Bewältigung dieser erläutert. Die Evaluierung der beiden implementierten Methoden erfolgt in Kapitel 8. Hierbei werden diese hinsichtlich ihrer Robustheit und Performanz statistisch ausgewertet. Kapitel 9 diskutiert die Erfüllung der durch die Anforderungsanalyse definierten Vorgaben. Des Weiteren erfolgt eine Gegenüberstellung der beiden entwickelten Algorithmen. Abschließend fasst Kapitel 10 die Beiträge dieser Abschlussarbeit zusammen und gibt einen Ausblick auf mögliche zukünftige Arbeiten in diesem Bereich.

Kapitel 2

Anforderungsanalyse

Lorem ipsum dolor sit amet.

2.1 Bildaufnahme und Preprocessing

Anforderung 1 *Die zur Erkennung von Hindernissen verwendeten Bilder müssen kalibriert, entzerrt sowie stereo-rektifiziert vorliegen.*

Um Anforderung 1 zu erfüllen müssen zunächst die intrinsischen und extrinsischen Parameter der Kamera berechnet werden. Dieser Prozess wird als Kalibrierung der Kamera bezeichnet. Dafür werden Bilder eines Kalibrierobjektes aufgenommen. Diese sind meistens einfache Schachbrettmuster mit ungleicher Anzahl an Quadraten, oder auch radiale Objekte mit einem spezifischerem Aufbau. Der Prozess der Kalibrierung zielt darauf ab die 'inneren' Parameter der Kamera zu bestimmen. Dies beinhaltet einerseits die Kameramatrix in welcher Parameter wie der Fokus Punkt(Focal Point), der Bildhauptpunkt(Principal Point) sowie die Brennweite der Kamera. Weiterhin werden geometrische Verzerrungen des Bildes, parametrisiert und somit nutzbar für die Entzerrung der Bilder in weiteren Schritten. Weiterhin werden bei der Kalibrierung die extrinsischen (äußeren) Parameter der Kamera bestimmt. Zu diesen zählt die Translation (x,y und z) sowie die Rotation der Kamera um die jeweiligen Achsen des Weltkoordinatensystems. Ein spezieller Aspekt bei der Kalibrierung von Stereo-Systemen ist die Berechnung der extrinsischen Parameter von einer Kamera zur anderen. Die dabei berechnete Translation einer Kamera zur Referenzkamera wird als Baseline bezeichnet. Diese ist für Prozeduren wie Stereo-Matching unabdingbar. Der Prozess der Parameterfindung ist in der Regel ein Prozess welcher einmalig nach dem Aufbau ausgeführt werden muss und solange währt wie sich nichts am Aufbau der Kameras ändert (Linsen,

Translation, Rotation).

Folgend aus der Kalibrierung der Kameras beginnt der Prozess der Rektifizierung der Bilder. Dieser kann dabei in zwei Schritte unterteilt werden, einerseits die Berechnung der Region of Interest, andererseits die Rektifizierung jedes aufgenommenen Einzelbildes.

2.2 Erfassung von Hindernissen

Anforderung 2 *Die Erfassung von Hindernissen sollte eine signifikant hohe Wahrscheinlichkeit haben um etwaige Fehlentscheidungen zu vermeiden.*

2.3 Performanz

Anforderung 3 *Das System sollte Hindernisse in Echtzeit erfassen und erkennen können.*

2.4 Positionsinformationen

Anforderung 4 *Relative Positionen der gefundenen Objekte müssen in jedem Frame vorliegen.*

Kapitel 3

State of the Art Algorithmen

Eines der wichtigsten und am meisten erforschten Felder in der Entwicklung autonomer Systeme ist die Vermeidung von Hindernissen in Echtzeit. Die wesentliche Grundlage dafür besteht in der Erkennung der Hindernisse. Dies geschieht oft mit Hilfe optischer Messtechniken.

Im folgenden Kapitel werden einige State of the Art Algorithmen erläutert. Dabei werden zunächst einige aktive optische Systeme näher beschrieben, bei denen die betrachtete Szene aktiv ausgestrahlt wird. Anschließend erfolgt die Analyse weiterer passiver optischer Algorithmen welche auf der Berechnung von *Disparity Maps* unter Zuhilfenahme stereo optischer Systeme basiert. Die Kalkulation dieser ist generell sehr rechenaufwändig, liefert aber nach der Auswertung zuverlässige Informationen über die Entfernung sowie Position des aktuellen Objektes.

3.1 Aktiv optische Algorithmen

Aktiv optische Algorithmen beziehen sich auf die aktive Emission von Licht auf die zu rekonstruierende Szene. Dies geschieht indem beispielsweise Muster auf die Szene projiziert werden (Gitter, Streifen, Farben, etc.), mit deren Hilfe es möglich ist korrespondierende Pixel innerhalb des Bildes zu finden. Jedoch können bei uniformen beziehungsweise einfachen Mustern falsche Bereiche des Musters erkannt werden und somit eine falsche räumliche Repräsentation der Szene. Ein alternativer Ansatz ist die Projektion zufälliger Muster um etwaige falsche Korrespondenzen durch bewusste Zufälligkeit zu vermeiden. Ein weiteres Feld in der Betrachtung aktiver visueller Techniken ist Time of Flight. Hierbei wird die Szene mit einem Lichtimpuls (meist infrarot) ausgeleuchtet und für jeden Pixel die Zeit gemessen die jener benötigt um wieder auf dem Sensor aufzutreffen. Damit ist es möglich in geringer Auflösung exakte Tiefendaten für jedes Einzelbild zu erhalten.

Die von Lee et al. [Lee et al., 2012] verfolgte Methodik bedient sich einer Time of Flight Kamera. Die aufgenommenen Tiefenbilder werden in verschiedene Segmente eingeteilt welche die verschiedenen Objekte repräsentieren. Sobald etwaige Kanten gefunden wurden werden diese entfernt unter dem Ansatz die Objekte so leichter segmentieren zu können. Anschließend erfolgt eine Tiefenanalyse der nun vorliegenden einzelnen Segmente. Zum Finden der Hindernisse wird nun die Standardabweichung jedes Segmentes berechnet. Dabei gelten Hindernisse als sich bewegende oder statische Objekte welche sich innerhalb eines definierten Gefahrenbereichs befinden. Lee et al. definieren Ihre Gefahrenzone dabei mit 1 – 2 Metern. Die erkannten Hindernisse werden nun mit Ihrer Distanz markiert. Ein häufig auftretendes Problem in diesem Ansatz spiegelt der Boden wieder, welcher oft als Hindernis erkannt wird. Ein Mittel zur Unterscheidung ist hierbei ebenfalls die Standardabweichung.

Bei der Entwicklung eines autonomen Roboters zur Indoor-Überwachung verschiedener Areale bedienen sich Correa et al. [Correa et al., 2012] eines Microsoft Kinect Sensors zur Erstellung von *Disparity Maps*. Dabei werden diese ebenfalls wie bereits in diversen passiven Algorithmen ([Pire et al., 2012], [Kostavelis et al., 2010]) in mehrere horizontale Segmente eingeteilt. Dabei besteht die finale Karte aus 5 Bereichen, von denen 3 als mögliche Richtungen betrachtet werden. Sobald die die minimalste Distanz eines Sektors geringer als 60cm ist, so wird angenommen, dass sich das System vor einem Hindernis befindet. Ausgehend von der Menge an Sektoren mit gefunden Hindernissen wird die neue Bewegungsrichtung angepasst.

3.2 Passiv optische Algorithmen

Passiv optische Algorithmen beziehen sich auf die Erfassung dreidimensionaler Informationen aus der Szene ohne eigene Lichtquelle. Dabei sind viele dieser Methoden an das menschliche oder auch tierische Sehen angelehnt. Ein großer Teil der Methodik ist das Prinzip Stereo Vision welches durch die Differenz zweier Bilder derselben Szene einen Eindruck von Tiefe verschafft. Weiterhin kann eine dreidimensionale Rekonstruktion der Umgebung durch Lichteinflüsse, Schattierung oder durch Veränderung des Kamerafokus vorgenommen werden.

Bei der Entwicklung eines autonomen Roboters auf Bodenebene wird nach Kostavelis et al, [Kostavelis et al., 2010] die errechneten *Disparity Maps* in 3 horizontale Bildbereiche aufgeteilt, welche die möglichen Richtungen des Systems

beschreiben. Für jedes dieser einzelnen Unterbilder wird nun der Durchschnitt der Disparität berechnet, wobei der Bildteil mit dem geringsten Wert auf Hindernisse hinweist, welche sich näher am System befinden. Aufgrund dessen, dass die Entscheidung welcher Weg nun der als am sichersten zu betrachten ist, zu Teilen sehr schwer zu treffen ist wurde die *Threshold Estimation* Methode entwickelt. Die Unterteilung der Bilder wird in diesem Algorithmus ebenfalls in drei Regionen vorgenommen. Zunächst werden alle Pixel deren Wert sich über einem vordefinierten Threshold befinden markiert, und gezählt. Wenn die Anzahl der als Hindernis definierten Pixel über einem ebenfalls vor definiertem Prozentsatz liegen, so wird ein Hindernis als gefunden markiert (in der jeweiligen Region). Die letzte vorgestellte Methode von Kostavelis et al. orientiert sich in ihrer Funktionsweise an der soeben beschriebenen *Threshold Estimation* und erweitert den Algorithmus um die Betrachtung aller 3 Bildteile. Bei der *Multi Threshold Estimation* wird jedes Drittel des Bildes betrachtet, wobei in allen Regionen nach Pixeln innerhalb des Thresholds gesucht und markiert werden. Anschließend werden die Ergebnisse untereinander verglichen, und das Drittel mit dem geringsten Wert als Hindernis ausgewählt. Sollten die prozentualen Werte aller Drittel größer sein als die gegebene Grenze so wird angenommen das sich das Hindernis sehr nah vor dem System befindet.

Im Rahmen dieser Arbeit wurde ebenfalls eine Methode entwickelt welche sich grob am ersten Ansatz orientiert. Diese wird in Kapitel 5 erläutert.

Einen anderen Ansatz zur Erkennung von Hindernissen geben Richards et al. [Richards et al., 2014]. Der Algorithmus wurde als Hindernisvermeidung für ein UAV entwickelt. Optischer Fluss sowie Feature Tracking bieten dabei die Grundlage für die Erkennung, Vermeidung und Voraussage welcher Bereich der sicherste ist. Dabei gehen die beiden Ausgangstechniken jeweils einer anderen Aufgabe nach. Der Optische Fluss dient zum Erkennen und verfolgen von Objekten sowie für die Voraussage der Position des Objektes im nächsten Einzelbild, wohingegen das Feature Tracking [Shi and Tomasi, 1994] für die Erkennung von markanten Punkten innerhalb des Objektes genutzt wird. Bei diesem wird in jedem folgenden Frame verglichen, ob bereits bekannte Punkte innerhalb einer bestimmten Distanz mit denen des aktuellen Frames übereinstimmen. Zur Abschätzung der nächsten Position des Features wird mit Hilfe des Optische Fluss der Verschiebungsvektor zwischen beiden Frames berechnet. Aufgrund dessen, dass der zugrunde liegende Algorithmus von Lukas-Kanade [Lucas et al., 1981] nur bei kleinen Verschiebungen valide Ergebnisse liefert, wird ein pyramidaler Ansatz [Bouguet, 2001] für das Matching verwendet bei welchem die jeweiligen Bilder des Frames herunter skaliert werden. Durch die Verbindung dieser beiden Techniken, lassen sich Objekte erkennen und verfolgen. Zur Bestimmung der nächstbesten Position für das UAV wird ausgehend

von den berechneten Resultaten (gefundene und erkannte Hindernisse) eine stochastische Matrix erstellt welche zur weiteren Planung des Fluges verwendet wird.

Bei der Entwicklung von Systemen, welche auf der Erkennung räumlicher Tiefe basieren, bieten stereo optische Systeme einen großen Vorteil. Durch die Verschiebung der Kameras an der Basislinie wird die Szene aus zwei minimal verschiedenen Blickwinkeln aufgenommen. Dies ermöglicht die Berechnung dreidimensionaler Informationen auf einfache Weise. Trotz dessen ist die Benutzung zweier Kameras nicht immer möglich, sei es die Limitierung durch das System selber aus Platzgründen oder, im Falle unbemannter Flugsysteme, die begrenzte maximale Traglast. Aufgrund letzterer entschieden sich Mori et al. [Mori and Scherer, 2013] für die Nutzung eines monokularen Setups. Durch die Verwendung des optischen Flusses ist es auch mit nur einer Kamera möglich Hindernisse zu erkennen, jedoch fehlt die eigentliche Erkennung von Tiefe. Sofern sich ein Objekt direkt auf das System zu bewegt, kann es nur schwer erkannt werden, da kaum perspektivische Veränderungen vorhanden sind. Um diese wahrzunehmen muss der Algorithmus dazu in der Lage sein, die Veränderung der relativen Größe eines Objektes in aufeinander folgenden Bildern abzuschätzen. Mori et al. setzen bei der Erkennung von Features auf den SURF Algorithmus [Bay et al., 2006] welcher gefundene Features auch nach Veränderung der Größe wieder erkennen kann. Im Anschluss daran wird die Veränderung der Größe der benachbarten Umgebung untersucht um eine Aussage darüber treffen zu können ob sich das Hindernis auf das System zu bewegt. Jene Features welche nicht skaliert wurden, werden nicht weiter betrachtet. In jedem folgenden Einzelbild wird nun mithilfe von Template Matching verglichen wie sich die Skalierung der Umgebung eines Features im Vergleich zum vorherigen Frame verändert hat. Sollte die Distanz eines Features zu nah am System sein, so wird dieses als potentiell Hindernis für die Vermeidung verwendet.

Kapitel 4

Zugrunde liegende Konzepte und Algorithmen

Dieses Kapitel beleuchtet dieser Arbeit zugrunde liegende Konzepte und Algorithmen. Zunächst wird das Prinzip der Epipolargeometrie beschrieben welches ein wesentlicher Bestandteil photogrammetrischer Verfahren ist. Daraufhin folgt eine Erläuterung des Terms Stereo Matching sowie eine Klassifizierung in lokale und globale Algorithmen. Im Anschluss daran wird das Prinzip des Block Matching Algorithmus grundlegend beschrieben, welcher die Grundlage für den im Rahmen dieser Arbeit verwendeten Semi Global Block Matching Algorithmus bildet. Anschließend erfolgt eine Erläuterung des verwendeten Frameworks sowie Details zur Implementierung dessen.

4.1 Epipolargeometrie

Bei der Verwendung der meisten photogrammetrischen Verfahren spielt die Epipolargeometrie eine entscheidende Rolle. Das mathematische Modell beschreibt die geometrische Beziehung zwischen verschiedenen Kamerabildern desselben Objektes, sowie die Beziehung korrespondierender Punkte. Generell gesehen ist die Epipolargeometrie durch das Lochkamera-Modell beschrieben. Dabei liegt jeder Punkt des aufgenommenen Objektes mit dem Projektionszentrum sowie dem Bildpunkt auf einer Geraden. Unter Zuhilfenahme der extrinsischen Orientierung sowie der intrinsischen Parameter der Kamera, ist es möglich den Schnittpunkt zweier Geraden zu berechnen um die dreidimensionalen Koordinaten des Objektpunktes zu erhalten. Dabei gilt generell, wenn ein Punkt P im linken Bild gegeben ist, so wird die Suche des korrespondierenden Punktes P' auf die Epipolarlinie des rechten Bildes reduziert.

Abbildung 1 visualisiert diesen Prozess. Gegeben sind die beiden Projektionszentren C_L und C_R der beiden Bildebenen sowie die Bildpunkte P_L und

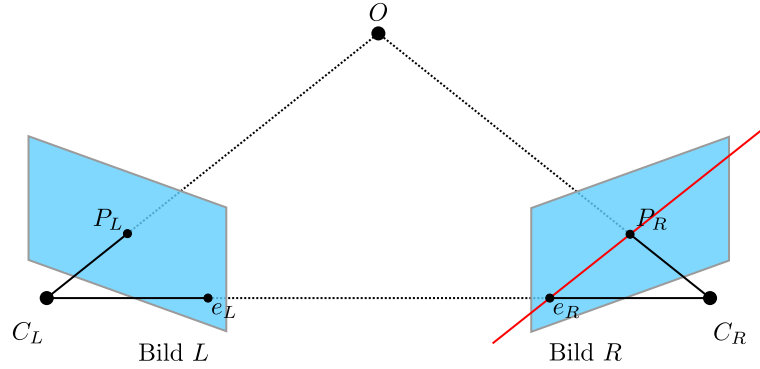


Abbildung 1: Darstellung der Epipolargeometrie.

P_R . Der Objektpunkt O bildet im linken Kamerabild auf P_{x_1, y_1} ab. Zunächst ist es nur möglich den Strahl auf dem sich O befindet zu bestimmen. Aufgrund des Wissens der extrinsischen Orientierung der Kameras können mithilfe der Basislinie zwischen den beiden Kameras die beiden Epipole in L und R ermittelt werden. Dabei sind diese durch den Schnittpunkt der Basislinie mit den beiden Bildebenen definiert¹. Somit ist es möglich mithilfe der vorhandenen Epipole e_L, e_R die Epipolarlinie e_L, P_{x_2, y_2} zu bestimmen. Anhand dieser kann im Anschluss die dreidimensionale Position des Objektpunktes O bestimmt werden.

4.2 Stereo Matching

Das grundlegende Konzept des Stereo Matchings beschreibt das Finden korrespondierender Punkte in zwei simultan aufgenommenen Bildern. Die Position der beiden Kameras ist dabei leicht versetzt, um einen jeweils anderen Blickwinkel auf die Szene zu erhalten. Mithilfe der verschiedenen Perspektiven können Disparitäten (Differenz der Projektion eines Objektes vom linken zum rechten Bild) zwischen korrespondierenden Pixeln berechnet werden. Die dabei erhaltenen Tiefeninformationen eines Objektes sind mit der errechneten Disparität sowie der relativen Position der Kameras und deren intrinsischen Parametern REF verbunden. Im Laufe dieses Prozesses kommt es zu zwei wesentlichen Problemstellungen: die Berechnung der Disparität (Stereo Korrespondenz) sowie die Invertierung der projektiven Geometrie, um dreidi-

¹Im Falle eines Stereonormalfalles ist ein Schnittpunkt der Basislinie mit den Bildebenen nicht möglich, sodass die Epipole in der Unendlichkeit und parallel zur x-Achse liegen. Dies hat unter anderem den Vorteil, dass die Epipolargeometrie bereits bekannt ist, und korrespondierende Bildpunkte nur noch innerhalb einer Pixelreihe gesucht werden müssen.

mensionale Informationen aus der errechneten Disparität zu erhalten. Sofern eine Lösung beider Probleme vorhanden ist, können diese Informationen durch einfache Triangulierung errechnet werden.

4.2.1 Klassifikation

Lokale Methoden:

Zur Berechnung der Disparität in lokalen Stereo Matching Algorithmen gilt grundlegendes Prinzip: “Finde Pixel P_2 korrespondierend zu P_1 im Referenzbild”. Dabei wird die Korrelation von P_1 und P_2 durch deren lokale Umgebung bestimmt. Der Referenzpunkt ist dabei das Zentrum eines Bereiches in welchem das Matching ausgeführt wird. Geläufige Methoden dafür sind Sum of Absolute Differences (SAD) (Hirschmüller 2011), Zero-mean Normalized Cross-Correlation (ZNCC) (Chen & Medioni, 1999), (Sára, 2002), Sum of Squared Differences (SSD) (Cox et al., 1996), etc.. Jedoch können aufgrund fehlender Beschränkungen verzerrte Berechnungen der räumlichen Tiefe auftreten, da benachbarte Pixel verschiedene Disparitäten aufweisen können (beispielsweise an horizontalen Kanten wie Türrahmen etc.). Strukturell gesehen sind lokale Methoden einfacher gehalten als globale Methoden, wodurch ein hoher Grad an Optimierung in der Implementierung möglich ist.

Globale Methoden:

Bei globalen Stereo Matching Algorithmen wird ein globales Modell der zu betrachtenden Szene erstellt sowie eine ebenfalls globale Kostenfunktion definiert, welche minimal gehalten werden soll. Dabei werden im Gegensatz zu lokalen Methoden die Matches in einer Pixelreihe und die des gesamten Bildes miteinander verglichen. Zusätzlich zu den benachbarten Pixeln werden hier ebenfalls die Matches der angrenzenden Pixel betrachtet. Zur Vereinfachung dieses Vorgangs betrachten einige Algorithmen nur die Epipolargeometrie, wobei ein zweidimensionales Problem auf ein eindimensionales reduziert wird. Resultierend daraus liegt die Stärke globaler Methoden in der Bewältigung schwacher Texturen sowie auftretender Okklusionen und unterschiedlicher Lichteinfälle, was, aufgrund der höheren Komplexität in einem höheren Rechen- und Speicheraufwand mündet. Weitere Verbesserungen der Ergebnisse können durch Techniken wie Dynamische Programmierung und Graph Cut erreicht werden.

4.2.2 Block Matching

Einen der einfachsten Ansätze zur Berechnung von Korrespondenz zwischen zwei Bildern bietet der Block Matching Algorithmus. Bei dieser lokalen Methode werden Blöcke bestimmter Größe mithilfe mathematischer Berechnungen

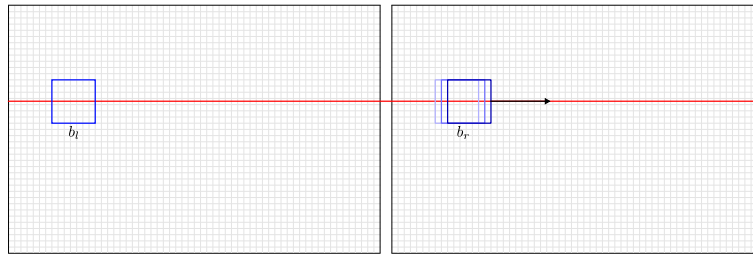


Abbildung 2: Visualisierung des Block Matching Algorithmus

(NCC, SAD, etc.) auf Korrespondenz untersucht. Dabei reduziert die Epipolar-geometrie diesen Prozess auf ein eindimensionales Problem indem ein Block im linken Bild mit einem Block im rechten Bild auf der Epipolarlinie des rechten Bildes pixelweise verglichen wird. Das eigentliche Matching erfolgt über die Berechnung eines Blockes mithilfe einer Energiefunktion.

Diesen Prozess veranschaulicht Abbildung 2. Durch den Fokus auf einzelne Pixelreihen und deren unmittelbare Umgebung ist dieses Verfahren wesentlich schneller, aber auch ungenauer als globale Matching Algorithmen.

4.2.3 Semi Global Block Matching

Das im Rahmen dieser Arbeit verwendete Verfahren zur Berechnung von Disparity Maps ist der in der freien Computer Vision Library implementierte Semi Global Block Matching Algorithmus. Diese leicht abgewandelte Implementierung des Semi Global Matching Algorithmus von Hirschmüller et. al [Hirschmuller, 2005] zeichnet sich sowohl durch seine Genauigkeit, als auch durch seine performante Berechnung aus. Unter Zuhilfenahme dieses Verfahrens ist es möglich, Disparity Maps in Echtzeit mit sehr guter Framerate zu berechnen. Im Folgenden wird zunächst die grundlegende Funktionsweise des SGM erläutert. Im Anschluss daran werden die Unterschiede zum SGBM herausgearbeitet sowie eine kurze Erklärung der Parameter dessen vorgenommen.

SGM:

Die grundlegende Idee des Semi-Global Matching Algorithmus besteht in dem pixelweisen Matching von Mutual Information ². Die globale zweidimensionale Smoothness Einschränkung wird dabei mithilfe mehrerer eindimensionaler Einschränkungen erzeugt. Ausgangsvoraussetzung dafür sind verschiedene Bilder derselben Szene mit vorhandener und bekannter Epipolar-geometrie. Zunächst

²Mutual Information(MI), zu deutsch Transinformationen, beschreiben den statistischen Zusammenhang zweier Zufallsgrößen...

werden für jeden Pixel P aus der Intensität I_{bp} sowie der vermuteten Korrespondenz I_{mq} auf der Epipolarlinie $q = e_{bm}(pd)$ die Matching Kosten berechnet. Die eigentliche Berechnung dieser erfolgt dann mit Hilfe von Mutual Information ³ REF berechnet.

$$MI_{I_1, I_2} = \sum_p mi_{I_1, I_2}(I_{1p}, I_{2p}) \quad (4.1)$$

$$mi_{I_1, I_2}(i, k) = h_{I_1}(i) + h_{I_2}(k) - h_{I_1, I_2}(i, k) \quad (4.2)$$

Zur Berechnung der MI mithilfe der in 4.1 und 4.2 dargestellten Formeln wird zunächst eine Initial Disparity Map benötigt. Diese wird nach Kim et al.'s [Zureiki et al., 2008] Ansatz zufällig gewählt, um die Kosten berechnen zu können. Dies ist insbesondere für iterative Verfahren geeignet, da andernfalls die Berechnung verlangsamt werden könnte. Zur Steigerung der Performanz wird die Disparity Map zunächst nur auf halber Auflösung berechnet, wodurch der Rechenaufwand um den Faktor 2·3 reduziert wird. Zur Vermeidung falscher Kostenberechnungen durch auftretende Noise innerhalb des Bildes nimmt der Algorithmus weitere Einschränkungen vor. Dabei werden die benachbarten Disparitäten in die Berechnung 4.3 mit einbezogen.

$$foo = foo \quad (4.3)$$

Das letzte Problem besteht in der Berechnung der Korrespondenz sowie der daraus resultierenden Disparitäten. Dabei wird nach der Disparität D mit der geringsten berechneten Energiefunktion gesucht. Anstatt nun einfach den minimalsten Pfad der Kosten zu summieren, werden alle Richtungen zur aktuellen Disparität mit einbezogen (siehe Abbildung 3).

Zur Berechnung valider Werte sollten dabei mindestens 8 Richtungen vorliegen; - eine vermehrte Anzahl an Richtungen, etwa 16, ist dabei vorteilhaft. Die berechnete Disparität ergibt sich aus den minimalen Kosten anhand dieser Pfade. Dabei gilt: "Je mehr Übereinstimmungen in den Kosten, desto wahrscheinlicher ist es, dass D zur Intensität I gehört".

SGBM:

Der Semi Global Block Matching Algorithmus ist ein in der Computer Vision Library OpenCV implementiertes Verfahren zur schnellen Berechnung von Disparity Maps. Die Grundlage dafür bietet Hirschmüller et al.'s SGM [Hirschmuller, 2008] mit den folgenden grundlegenden Änderungen:

C.1 Statt den originalen 8 bzw. 16 Richtungen werden nur 5 betrachtet.

³muss noch kommen

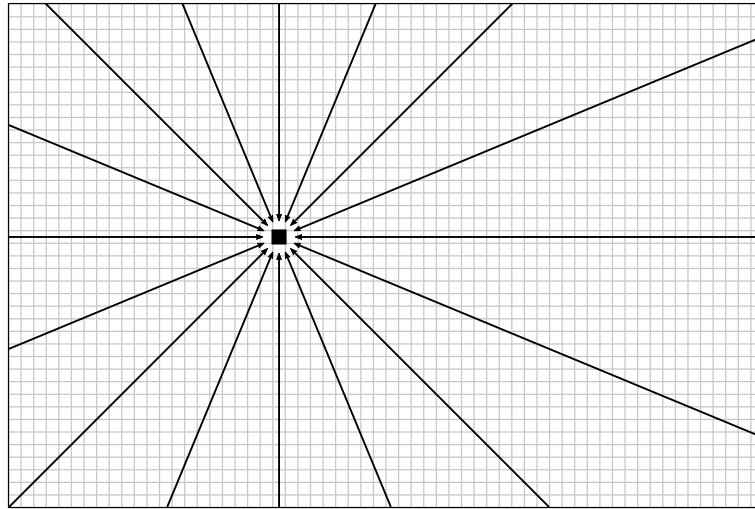


Abbildung 3: Darstellung der verschiedenen Richtungen

- C.2 Es werden standartmässig keine einzelnen Pixel sondern Blöcke verglichen.
- C.3 Anstelle der Mutual Information Kostenfunktion wird das von Birchfield et al. vorgestellte Sub-Pixel Dissimilarity Measurement Verfahren verwendet [Birchfield and Tomasi, 1998].
- C.4 Pre und Post Processing Elemente des *StereoBM* [OpenCV, 2015a] werden verwendet.

Dies erlaubt dem Algorithmus eine schnelle Berechnung der Disparitäten auf einem qualitativ hochwertigen Niveau. Der geringe Verlust an Qualität kann in Anbetracht der Berechnung in Echtzeit vernachlässigt werden. In Abbildung ?? findet sich eine Gegenüberstellung beider Algorithmen.

4.3 mvStereoVision Framework

Das verwendete Framework zur Bildaufnahme und Disparity Map Berechnung wurde im Rahmen des Projektes “SLAM for UAV” entwickelt. Dieses bedient sich der von Matrix Vision zur Verfügung gestellte Library [GmbH, 2015] zur Kommunikation mit den Kameras, sowie OpenCV [OpenCV, 2015b] zur Verarbeitung der Bilder. Die wesentlichen Funktionen werden im folgenden näher beleuchtet.

Kapitel 5

Entwickelte Hinderniserkennungen

Im Rahmen dieser Arbeit wurden zwei Systeme zur Erkennung von Hindernissen in Echtzeit entwickelt. Diese richten sich nach den in Kapitel REF erläuterten Algorithmen und Konzepten. Anhand dieser ist es möglich aus den beiden Bildern des Stereo Systems für jeden Frame die Disparity Map zu berechnen, welche im Anschluss daran in mehreren Schritten zunächst so angepasst wird, dass nicht verwertbare Bereiche der Tiefenkarte entfernt werden (siehe REF preprocessing). Dies reduziert die Anzahl der zur Erkennung zu verarbeitenden Punkte und somit die Rechenzeit.

Im folgenden Kapitel werden beide Methoden detailliert beleuchtet. Zu Beginn wird die zugrunde liegende Klassenstruktur beschrieben. In Abschnitt 5.1 die *Mean Disparity Detection* erläutert, wobei auf das grundlegende Konzept sowie den Algorithmus zur Erkennung selber eingegangen wird. Selbiges gilt für die *Samplepoint Detection* in Abschnitt 5.2.

5.1 Mean Disparity Detection

Das grundlegende Funktionsprinzip der *Mean Disparity Detection* ist grob an die von Kostavelis et al. vorgestellten Algorithmen angelehnt. Auch bei diesen werden die berechneten Disparity Maps in Segmente eingeteilt von welchen in jedem Frame der Mittelwert errechnet wird.

Zu Beginn des Algorithmus, nach der vorherigen Bearbeitung der eigentlichen *region of interest*, wird während der Initialisierung die Segmentierung der Disparity Map festgelegt. Dabei wird für jedes Segment ein eigenes Subimage erzeugt. Diese definieren eine *region of interest* (ROI) innerhalb einer Matrix.

Subimages selber halten lediglich die Positionsinformationen (obere linke und untere rechte Ecke des definierten Rechtecks), Mittelwert der ROI, die ROI selbst sowie den Mittelpunkt des Rechtecks zur späteren Pointcloud Generierung, wie Abbildung REF aufzeigt.

Für die Unterteilung der Disparity Map wurden initial nur 9 Segmente vorgesehen wobei jeder Bereich eine der möglichen Flugrichtungen repräsentieren sollte. Aufgrund der Größe der Submatrizen konnte jedoch keine valide Erkennung kleiner Hindernisse gewährleistet werden, da solche in der Berechnung des Mittelwertes untergegangen sind. Aufgrund dessen wurde die Anzahl der Segmente auf 81 erhöht, welches eine wesentlich genauere Erkennung ermöglicht. Zudem ist somit eine weitaus genauere Einteilung der Flugrichtungen möglich, da jede dieser noch einmal genauer betrachtet wird (siehe Abbildung REF).

Um die Hindernisse innerhalb der Segmente zu Erkennen wurde auf die Berechnung des Mittelwertes gesetzt. Einerseits, da die Berechnung des Mittelwertes unter Betrachtung des Echtzeit-Aspektes eine ressourcensparende und schnelle Operation ist, andererseits weil jeder Pixel des Bildes in das Endresultat mit einfließt. Jedoch musste die Berechnung auf das Szenario angepasst werden, da Berechnung des Medians aller Werte zu Verzerrungen geführt hätte. Bei der Kalkulation von Disparity Maps werden Bereiche welche nicht gematcht werden können, sei es aufgrund von fehlenden Informationen im Referenzbild oder homogener Texturen in der Szene, als negativer Wert ausgedrückt. Berechnet man nun den Mittelwert unter Betrachtung positiver und negativer Werte, so entspricht dies zwar dem definierten Term Median, verfälscht allerdings das in diesem Anwendungsbereich erwünschte Ergebnis. Im schlechtesten Fall enthält eine Submatrix mehr negative als positive Werte, was dazu führen kann das Hindernisse vor homogenen Flächen nicht erkannt werden. Daraus ergibt sich die in Algorithmus 1 dargestellte Berechnung des Mittelwertes.

Es werden also nur positive Disparitäten betrachtet was auch dazu führt, dass sich der Gesamtwert aller Werte aus der Menge dieser ergibt. Mit Hilfe dieser Berechnung ist es möglich Hindernisse auch in Bereichen zu erkennen wo die Mehrzahl der Pixel keine Matches aufweisen.

Die eigentliche Hinderniserkennung erfolgt in jedem Frame. Wobei der Term Frame hierbei nicht nach einem von der Kamera aufgenommenen Bild, sondern eine neue Disparity Map definiert. Da die Bildgröße pro Einzelbild stetig ist besteht keine Nötigkeit die einzelnen Segmente erneut zu generieren. Es genügt also die Mittelwerte eines jeden Subimages anhand der neuen Disparity Map zu aktualisieren. Daraufhin wird geprüft ob sich einer oder mehrere der Subimage Mittelwerte innerhalb der Gefahrenzone befinden. Anschließend wird

Algorithm 1 Berechnung des Disparity Medians

```

1: procedure CALCMEANDISPARITY(submatrix)
2:    $elements_{number} \leftarrow 0$ 
3:    $elements_{sum} \leftarrow 0$ 
4:   for value in submatrix do
5:     if value > 0 then
6:        $elements_{sum} \leftarrow elements_{sum} + value$ 
7:        $elements_{number} \leftarrow elements_{number} + 1$ 
8:     end if
9:   end for
10:  return  $elements_{sum}/elements_{number}$ 
11: end procedure

```

unter Zuhilfenahme der neuen Werte eine Pointcloud erstellt und geloggt. Der gesamte Ablauf des der Hinderniserkennung wird in Algorithmus 2 dargestellt.

Algorithm 2 Ablauf der Hinderniserkennung

```

1: procedure DETECTOBSTACLES(submatrix)
2:    $elements_{number} \leftarrow 0$ 
3:    $elements_{sum} \leftarrow 0$ 
4:   for value in submatrix do
5:     if value > 0 then
6:        $elements_{sum} \leftarrow elements_{sum} + value$ 
7:        $elements_{number} \leftarrow elements_{number} + 1$ 
8:     end if
9:   end for
10:  return  $elements_{sum}/elements_{number}$ 
11: end procedure

```

Die in Algorithmus 2 berechnete Pointcloud ist in diesem Fall ein Matrix mit der Größe der ursprünglichen Tiefenkarte. Mithilfe der Mittelpunkte der Sub-images wird für jedes dieser der jeweilige Tiefenwert an der Position des Mittelwertes abgelegt, dabei ist die Beibehaltung der ursprünglichen Größe für die Berechnung der Distanz (Formel 5.1) zum Hindernis von Nöten.

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = Q \cdot \begin{bmatrix} image_x \\ image_y \\ disparity(x, y) \\ 1 \end{bmatrix} \quad (5.1)$$

$$pointcloud(x, y) = [\frac{X}{W}, \frac{Y}{W}, \frac{Z}{W}]$$

5.2 Samplepoint Detection

Kapitel 6

Konflikte in der Erkennung und Lösungsansätze

6.1 Validierung der Hindernisse

6.2 Hindernisgröße

6.3 Position Mapping

Kapitel 7

Evaluierung

7.1 Aufbau des Testsetups

7.2 Evaluierung Subimage Detection

7.2.1 Robustheit

7.2.2 Performanz

7.3 Evaluierung Samplepoint Detection

7.3.1 Robustheit

7.3.2 Performanz

Kapitel 8

Diskussion

8.1 Anforderungsevaluierung

8.2 Gegenüberstellung beider Algorithmen

Kapitel 9

Fazit und zukünftige Arbeiten

Literaturverzeichnis

- [Bay et al., 2006] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *Computer vision–ECCV 2006*, pages 404–417. Springer.
- [Birchfield and Tomasi, 1998] Birchfield, S. and Tomasi, C. (1998). A pixel dissimilarity measure that is insensitive to image sampling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(4):401–406.
- [Bouguet, 2001] Bouguet, J.-Y. (2001). Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. *Intel Corporation*, 5:1–10.
- [Correa et al., 2012] Correa, D. S. O., Sciotti, D. F., Prado, M. G., Sales, D. O., Wolf, D. F., and Osório, F. S. (2012). Mobile robots navigation in indoor environments using kinect sensor. In *Critical Embedded Systems (CBSEC), 2012 Second Brazilian Conference on*, pages 36–41. IEEE.
- [GmbH, 2015] GmbH, M. V. (2015). MATRIX VISION GmbH - industrial image processing.
- [Hirschmuller, 2005] Hirschmuller, H. (2005). Accurate and efficient stereo processing by semi-global matching and mutual information. 2:807–814 vol. 2.
- [Hirschmuller, 2008] Hirschmuller, H. (2008). Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):328–341.
- [Kostavelis et al., 2010] Kostavelis, I., Nalpantidis, L., and Gasteratos, A. (2010). Comparative presentation of real-time obstacle avoidance algorithms using solely stereo vision. In *IARP/EURON International Workshop on Robotics for risky interventions and Environmental Surveillance-Maintenance, Sheffield, UK*.

- [Lee et al., 2012] Lee, C.-H., Su, Y.-C., and Chen, L.-G. (2012). An intelligent depth-based obstacle detection system for visually-impaired aid applications. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2012 13th International Workshop on*, pages 1–4. IEEE.
- [Lucas et al., 1981] Lucas, B. D., Kanade, T., et al. (1981). An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679.
- [Mori and Scherer, 2013] Mori, T. and Scherer, S. (2013). First results in detecting and avoiding frontal obstacles from a monocular camera for micro unmanned aerial vehicles. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 1750–1757. IEEE.
- [OpenCV, 2015a] OpenCV (2015a). OpenCV - camera calibration and 3d reconstruction.
- [OpenCV, 2015b] OpenCV (2015b). OpenCV - open source computer vision.
- [Pire et al., 2012] Pire, T., de Cristóforis, P., Nitsche, M., and Berlles, J. J. (2012). Stereo vision obstacle avoidance using depth and elevation maps. *IEEE VI RAS Summer School on “Robot Vision and Applications”, Santiago, Chile*.
- [Richards et al., 2014] Richards, B., Gan, M., Dayton, J., Quintana, J., Liu, J., and Enriquez, M. (2014). Obstacle avoidance system for uavs using computer vision.
- [Shi and Tomasi, 1994] Shi, J. and Tomasi, C. (1994). Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR’94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE.
- [Zureiki et al., 2008] Zureiki, A., Devy, M., and Chatila, R. (2008). *Stereo Matching and Graph Cuts*. INTECH Open Access Publisher.