

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

<b>1</b>	<b>Einführung</b>	<b>3</b>
1.1	Motivation . . . . .	3
1.2	Setup . . . . .	4
1.3	Ziel der Arbeit . . . . .	4
<b>2</b>	<b>State of the Art Algorithmen</b>	<b>6</b>
2.1	Aktiv optische Algorithmen . . . . .	6
2.2	Passiv optische Algorithmen . . . . .	7
<b>3</b>	<b>Zugrunde liegende Konzepte und Algorithmen</b>	<b>10</b>
3.1	Epipolargeometrie . . . . .	10
3.2	Kamerakalibrierung und Bild-Rektifizierung . . . . .	11
3.3	Stereo Matching . . . . .	13
3.3.1	Klassifikation . . . . .	14
3.3.2	Block Matching . . . . .	15
3.3.3	Semi Global Block Matching . . . . .	15
3.4	mvStereoVision Framework . . . . .	18
<b>4</b>	<b>Entwickelte Hinderniserkennungen</b>	<b>21</b>
4.1	Mean Disparity Detection . . . . .	21
4.2	Samplepoint Detection . . . . .	24
<b>5</b>	<b>Limitierungen und Lösungsansätze</b>	<b>27</b>
5.1	Bestehende Konflikte und Lösungsansätze . . . . .	27
5.2	Diskussion . . . . .	29
<b>6</b>	<b>Evaluierung</b>	<b>32</b>
6.1	Testsetup . . . . .	32
6.2	Evaluierung Subimage Detection . . . . .	33
6.2.1	Robustheit . . . . .	33
6.2.2	Performanz . . . . .	33
6.3	Evaluierung Samplepoint Detection . . . . .	33

6.3.1	Robustheit . . . . .	33
6.3.2	Performanz . . . . .	33
<b>7</b>	<b>Diskussion</b>	<b>34</b>
7.1	Anforderungsevaluierung . . . . .	34
7.2	Gegenüberstellung beider Algorithmen . . . . .	34
<b>8</b>	<b>Fazit und zukünftige Arbeiten</b>	<b>35</b>
	<b>Literaturverzeichnis</b>	<b>36</b>

# 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

Die aktive Entwicklung der Algorithmen erfolgte im Hinblick auf eine Verwendung dieser durch das von Ascending Technologies [Technologies, 2015] entwickelte UAS (Unmanned Aerial System) Pelican (Abbildung ??). Es handelt sich dabei um einen Quattrocopter welcher für den Forschungsbereich entwickelt wurde. Dieser ist mit einem Bordcomputer ausgestattet welche die nötige Leistung für die Entwicklung der Algorithmen bereitstellt (3rd Generation Intel Core i7). Weiterhin wurden zwei MatrixVision BlueFOX mv-MLC200wC Industriekameras als visuelles System verwendet. Die maximale Auflösung beider Kameras beträgt  $752 \times 480$  bei 60 möglichen Bildern pro Sekunde, in Abhängigkeit verschiedener Parameter (verwendet Verschlusszeit, aufgenommene Bitrate, u.a). Für den Echtzeit Aspekt des Systems werden beide Kameras im horizontalen und vertikalen Binning verwendet. Dies reduziert die Anzahl der Bildpunkte in beiden Dimensionen auf  $376 \times 240$ , wodurch der Berechnungsaufwand verringert und die die Aufnahmerate der Kameras auf bis zu 170 Einzelbilder pro Sekunde maximiert wird. Für Entwicklung der Algorithmen wurde die freie Computer Vision Bibliothek OpenCV [OpenCV, 2015b] verwendet. Diese stellt benötigte Algebraische Grundoperationen sowie bestimmte Algorithmen welche im Rahmen dieser Arbeit genutzt wurden zur Verfügung.

### 1.3 Ziel der Arbeit

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.

Vor diesem Hintergrund werden zunächst einige State of the Art Algorithmen beschrieben wobei dabei zwischen aktiven und passiven optischen Systemen unterschieden wird. Diese Einteilung dient einerseits dafür einen Überblick über bereits bestehende Techniken sowie implementierte System zu erhalten, andererseits um auch die Vor und Nachteile der jeweiligen Technik herauszu-



**Abbildung 1:** AscTec Pelican

arbeiten. Kapitel 3 beleuchtet zugrundeliegende Konzepte und Algorithmen, dabei liegt der Fokus primär auf essenziell notwendigen Techniken für die im Rahmen dieser Arbeit entwickelten Systeme. Des Weiteren wird das genutzte Framework sowie dessen Funktionsweise erläutert. Im Anschluss daran beschreibt Kapitel 4 die beiden entwickelten Algorithmen zur Hinderniserkennung detailliert. Anschließend daran erfolgt die Evaluation beider Systeme unter den Gesichtspunkten der Robustheit sowie der Performance jedes einzelnen Algorithmus. Weitere Tests zur zukünftigen Verbesserung der Algorithmen weisen auf welche aktuellen Limitierungen durch einerseits genutzte Konzepte vorliegen. Kapitel 5 erläutert eben diese Limitierungen und gibt Ansätze zur Lösung aus der Fachliteratur sowie eigene Konzepte zur Bewältigung dieser. Die anschließende Diskussion wertet die im Rahmen der Evaluation in Kapitel 6 erlangten Ergebnisse weitergehend aus und stellt beide Algorithmen aktiv gegenüber. Kapitel 8 zieht ein Résumé aus den Ergebnissen der Arbeit und gibt einen Ausblick auf mögliche zukünftige Arbeiten in diesem Bereich.



# Kapitel 2

## 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 verschiedene aktive optische Systeme näher beschrieben, bei denen die betrachtete Szene aktiv ausgeleuchtet 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.

### 2.1 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 Tie-

fendaten 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 werden, werden diese unter dem Ansatz die Objekte so leichter segmentieren zu können entfernt. 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. 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 wird angenommen, dass sich das System vor einem Hindernis befindet. Ausgehend von der Menge an Sektoren mit gefunden Hindernissen wird die neue Bewegungsrichtung angepasst.

## **2.2 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 werden nach Kostavelis et al, [Kostavelis et al., 2010] die errechneten Disparity Maps in 3 horizontale Bildbereiche aufgeteilt, welche die möglichen Richtungen des Sy-

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. Angesichts der Tatsache, dass die Entscheidung darüber, welcher Weg als der sicherste angesehen werden muß, teilweise 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 Werte sich über einem vordefinierten Threshold befinden markiert und gezählt. Wenn die Anzahl der als Hindernis definierten Pixel über einem ebenfalls vordefinierten 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 diese 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 dass sich das Hindernis sehr nah vor dem System befindet.

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 Hinder-

nisse) 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. Trotzdem ist die Benutzung zweier Kameras nicht immer möglich, sei es durch die Limitierung durch das System selber aus Platzgründen oder, im Falle unbemannter Flugsysteme, durch 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. setzten 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 mit Hilfe 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 3

## 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.

### 3.1 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 2 visualisiert diesen Prozess. Gegeben sind die beiden Projektionszentren  $C_L$  und  $C_R$  der beiden Bildebenen sowie die Bildpunkte  $P_L$  und  $P_R$ .



**Abbildung 2:** Darstellung der Epipolargeometrie.

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 definiert<sup>1</sup>. 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.

## 3.2 Kamerakalibrierung und Bild-Rektifizierung

Um den Prozess des Stereo Matchings zu vereinfachen und beschleunigen bietet die Rektifizierung der Referenzbilder einen simplen Einstiegspunkt zur Lösung des Stereo-Korrespondenzproblems. Dafür 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. Wei-

<sup>1</sup>Im 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.

terhin 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 (Objektive, 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. Mithilfe der zuvor berechneten intrinsischen und extrinsischen Parameter kann nun die Transformation der beiden Kamerabilder zueinander berechnet werden. Dies geschieht unter dem Aspekt der Epipolargeometrie. Nach der Rektifizierung und dem Anwenden der ROIs auf beiden Kamerabildern entspricht die jeweilige Pixelreihe im linken Bild der Pixelreihe mit dem selben Index im rechten Bild. Abbildung 3 verdeutlicht den Ablauf der Rektifizierung, so ist in Abbildung 3 (c) erkennbar das sich die korrespondierenden Bildpunkte auf einer Epipolarlinie befinden.



(a) Korrespondierende Punkte innerhalb der entzerrten Bilder



(b) Korrespondierende Punkte in den rektifizierten Bilder



(c) Korrespondierende Punkte in den geschnittenen und skalierten rektifizierten Bildern

**Abbildung 3:** Resultat der Rektifizierung

### 3.3 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 verbunden. Im Laufe dieses Prozesses kommt es zu zwei wesent-



lichen Problemstellungen: die Berechnung der Disparität (Stereo Korrespondenz) sowie die Invertierung der projektiven Geometrie, um dreidimensionale Informationen aus der errechneten Disparität zu erhalten. Sofern eine Lösung beider Probleme vorhanden ist, können diese Informationen durch einfache Triangulierung errechnet werden.

### 3.3.1 Klassifikation

#### Lokale Methoden:

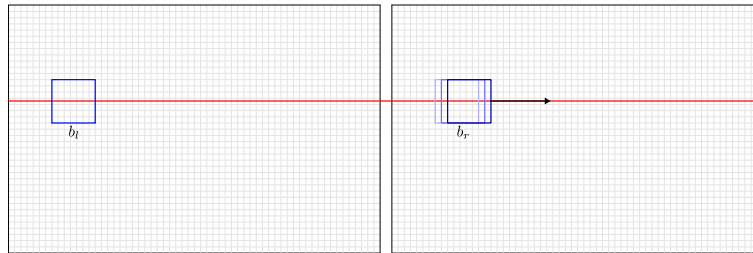
Zur Berechnung der Disparität in lokalen Stereo Matching Algorithmen gilt grundlegendes Prinzip: "Finde Pixel  $P_2$  korrespondierend zu  $P_1$  im Referenzbild". Dabei wird die Korrelation von  $P_1$  und  $P_2$  durch deren lokale Umgebung bestimmt. Der Referenzpunkt ist dabei das Zentrum eines Bereiches 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.

### 3.3.2 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 (NCC, SAD, etc.) auf Korrespondenz untersucht. Dabei reduziert die Epipolargeometrie 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.



**Abbildung 4:** Visualisierung des Block Matching Algorithmus

Diesen Prozess veranschaulicht Abbildung 4. Durch den Fokus auf einzelne Pixelreihen und deren unmittelbare Umgebung ist dieses Verfahren wesentlich schneller, aber auch ungenauer als globale Matching Algorithmen.

### 3.3.3 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 [Hirschmüller, 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 <sup>2</sup>. Die globale zweidimensionale Smoothness Einschränkung wird dabei mithilfe mehrerer eindimensionaler Einschränkungen erzeugt. Ausgangsvoraussetzung dafür sind verschiedene Bilder derselben Szene mit vorhandener und bekannter Epipolargeometrie. Zunächst 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 <sup>3</sup> REF berechnet.

$$MI_{I_1, I_2} = \sum_p mi_{I_1, I_2}(I_{1p}, I_{2p}) \quad (3.1)$$

$$mi_{I_1, I_2}(i, k) = h_{I_1}(i) + h_{I_2}(k) - h_{I_1, I_2}(i, k) \quad (3.2)$$

Zur Berechnung der MI mithilfe der in 3.1 und 3.2 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 3.3 mit einbezogen.

$$foo = foo \quad (3.3)$$

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 5).

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 Vi-

---

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

<sup>3</sup>muss noch kommen



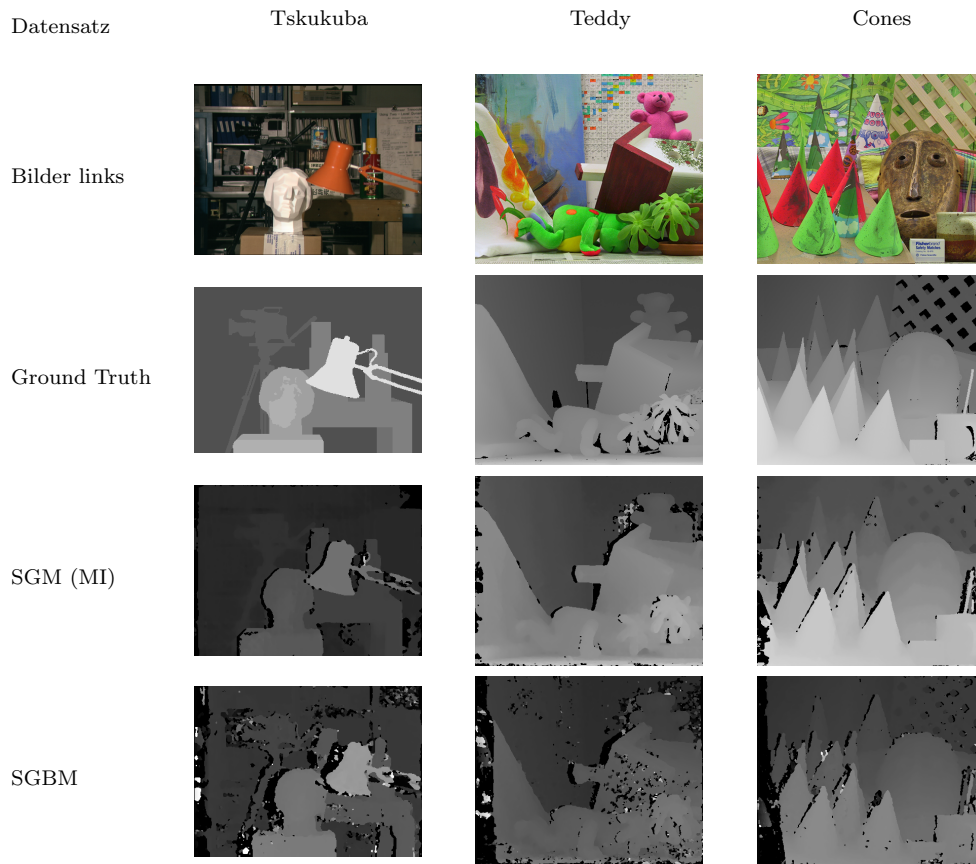
**Abbildung 5:** Darstellung der verschiedenen Richtungen

sion Library OpenCV implementiertes Verfahren zur schnellen Berechnung von Disparity Maps. Die Grundlage dafür bietet Hirschmüller et al.'s SGM [Hirschmüller, 2008] mit den folgenden grundlegenden Änderungen:

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

Abbildung 6 stellt die originalen Bildern der Szene ([Middlebury Univeristy, 2015]), die mithilfe von strukturiertem Licht aufgenommenen Ground Truth Bilder sowie die jeweiligen berechneten Tiefenkarten durch den SGM und den SGBM dar. Aus dieser ist der Qualitaetsverlust des SGBM beim Tsukuba- sowie beim Teddy Datensatz zu erkennen. Trotz dessen sind die Grundformen sowie die berechnete Tiefe klar erkennbar.



**Abbildung 6:** Vergleich zwischen Ground Truth Bildern sowie dem SGM(2005) und OpenCV's SGBM

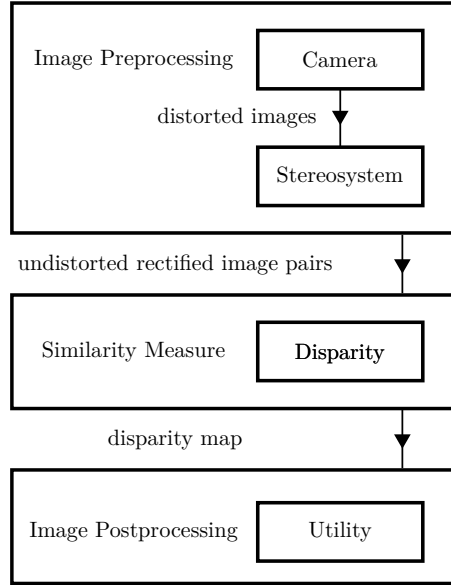
### 3.4 mvStereoVision Framework

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.

#### **Bildaufnahme:**

Die Aufnahme der einzelnen Bilder beginnt bei einer Anfrage an die Kamera nach den jeweiligen Rohdaten. Diese werden anschliessend mithilfe der durch die Kalibrierung ermittelten Parameter entzerrt und rektifiziert und liegen so zur weiteren Verarbeitung bereit. Die Aufnahme der Bilder erfolgt dabei in

separaten Threads um unabhängig von weiteren Berechnungen Bilder aufzunehmen. Die Berechnung der Disparity Map erfolgt ebenfalls in einem eigenen Thread. Sofern eine neue Disparity Map vorliegt wird innerhalb des Hauptprogramms mit Hilfe eines Wahrheitswertes darüber informiert das diese vorliegt. Der gesamte Prozess wird in Abbildung 7 visualisiert.



**Abbildung 7:** Pipeline zur Bildaufnahme, Disparity Map Berechnung sowie weiteren Verarbeitung

### Distanzberechnung:

Zugrundeliegend bei der Distanzberechnung ist die Reprojektionsmatrix  $Q$  welche im Rahmen der Rektifizierung berechnet wird und wie folgt aufgebaut ist.

$$Q = \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 0 & f \\ 0 & 0 & \frac{-1}{T_x} & \frac{C_x - C'_x}{T_x} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 0 & f \\ 0 & 0 & a & b \end{bmatrix} \quad (3.4)$$

Dabei beschreiben  $-C_x$  und  $-C_y$  die beiden negativen Koordinaten des Bildhauptpunktes der linken Kamera sofern dies als Referenz vorlag, der Parameter  $f$  die Brennweite der linken Kamera.  $T_x$  ist die horizontale Translation zwischen beiden Kameras. Die in der  $Q$ -Matrix enthaltenen Parameter sind mit Ausnahme von  $C'_x$  alle die der linken Kamera. Es wird angenommen das sich die beiden Strahlen der Bildhauptpunkte in der Unendlichkeit schneiden woraus

sich der Parameter 0 in der unteren rechten Ecke berechnet. Zur Einfacheren Darstellung und Erläuterung werden die Baseline der Kamera  $\frac{-1}{T_x}$  im folgenden als  $a$  sowie die Darstellung beider Bildhauptpunkte  $\frac{C_x - C'_x}{T_x}$  als  $b$  notiert.

Eine Eigenschaft der Reprojektionsmatrix ist die einfache Berechnung von dreidimensionalen Weltkoordinaten. Diese Berechnung ist in Formel 3.5 abgebildet.

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = Q \cdot \begin{bmatrix} I_x \\ I_y \\ d(x, y) \\ 1 \end{bmatrix} \quad (3.5)$$

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

Die eigentliche Berechnung der einzelnen Parameter ergibt sich somit aus der Multiplikation des transponierten Vektors, welcher sowohl die Koordinaten des Pixels als auch die Disparität dessen enthält, und der Reprojektions Matrix. Die dabei benötigten Operationen werden im folgenden genauer beleuchtet.

$$\left. \begin{array}{l} X = I_x - C_x \\ Y = I_y - C_y \\ Z = f \\ W = d \cdot a + b \end{array} \right\} \begin{array}{l} X' = \frac{I_x - C_x}{d \cdot a + b} \\ Y' = \frac{I_y - C_y}{d \cdot a + b} \\ X' = \frac{f}{d \cdot a + b} \end{array} \quad (3.6)$$

# Kapitel 4

## 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 ??). 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 4.1 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 4.2.

### 4.1 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.

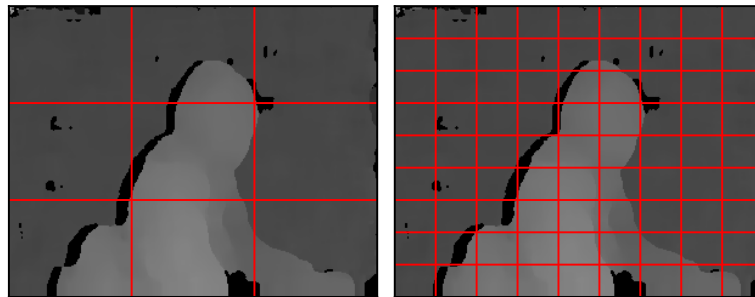


Subimages selber halten lediglich die Positionsinformationen (obere linke und untere rechte Ecke des definierten Rechtecks), Mittelwert der ROI, die ROI selbst sowie den Mittelpunkt des Rechtecks zur späteren Pointcloud Generierung, wie Abbildung 8 aufzeigt.

Subimage
+tl: Point
+br: Point
+roi_center: Point
+value: float
+Subimage(tl:Point,br:Point): Subimage
+calculateSamplepointValue(d_Map:Mat): float

**Abbildung 8:** Subimage Klasse

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 9).



**Abbildung 9:** Intiale Segmentierung der Disparity Map sowie die weitere Unterteilung zur genaueren Bestimmung der Werte der Subimages

Um die Hindernisse innerhalb der Segmente zu Erkennen wurde auf die Berechnung des Mittelwertes dieser 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.

---

**Algorithm 1** Berechnung des Disparity Medians

---

```

1: procedure CALCMEANDISPARIITY(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

```

---

Es werden demnach nur positive Disparitäten betrachtet was auch dazu führt das sich der Gesamtwert aller Werte aus der Menge dieser ergibt. Mit Hilfe dieser Berechnung ist es möglich Hindernisse auch in Bereichen zu erkennen in denen 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. Die Gefahrenzone wird zur Initialisierung der Erkennung berechnet. Da die Disparitäten in Abhängigkeit der Bildgröße sowie der Q-Matrix berechnet werden ergibt kann kein fest definierter Disparitätenwert für eine Distanz bestimmt werden. Demnach müssen zu Beginn die minimale sowie maximale Distanz der Erkennung auf einen Tiefenwert zurückgerechnet werden (siehe Formel 4.1). Dies erfolgt

als Erweiterung der in Abschnitt 3.4 beschriebenen ursprünglichen Distanzberechnung.

$$\begin{aligned} d &= \frac{f - Z' \cdot b}{Z' \cdot a} \\ I_x &= X' \cdot (d \cdot a + b) + C_x \\ I_y &= Y' \cdot (d \cdot a + b) + C_y \end{aligned} \tag{4.1}$$

Diese Rückrechnung ist ausschlaggebend für die schnelle Erkennung der Tiefe. Im Falle einer weiteren Segmentierung müsste die Berechnung der Distanz für jedes Subimage einzeln erfolgen, was gerade bei der Samplepoint Detection zum Tragen kommt, sollen die Distanzen für ca. 6000 Punkte (bei voller Auflösung in Abhängigkeit der verwendeten Parameter des SGBM) berechnet werden.

---

**Algorithm 2** Ablauf der Hinderniserkennung

---

- 1: **procedure** DETECTOBSTACLES(*submatrix*)
  - 2: **end procedure**
- 

Die in Algorithmus 2 berechnete Pointcloud ist in diesem Fall ein Matrix mit der Größe der ursprünglichen Tiefenkarte. Dies ist für die Berechnung der dreidimensionalen Koordinate von Nöten. Mithilfe des Mittelpunktes der Subimages wird für jedes dieser die jeweilige 3D-Koordinate (siehe Abschnitt 3.4) berechnet und an der Position des Mittelwertes in die Punktwolke geschrieben. Innerhalb der Pointcloud sind nun die Weltkoordinaten jedes Samplepoints relativ zur Kameraposition gespeichert. Der Vektor zwischen dem Koordinatenursprung und der Koordinate entspricht dabei der Distanz zum Hindernis.

## 4.2 Samplepoint Detection

In Hinblick auf aktive optische Algorithmen zur 3D-Rekonstruktion und damit verbunden Techniken wie Laser Scanning oder Time of Flight wurde die Samplepoint Detection entwickelt. Das Ziel war dabei eine ressourcensparende Alternative zur Mean Disparity Detection zu schaffen indem nicht zwangsläufig alle Pixel der Disparity Map betrachtet werden müssen.

Ebenso wie in der in Abschnitt 1 vorgestellten *Mean Disparity Detection* bedient sich die Samplepoint Detection einer vorverarbeiteten Disparity Map zur Hinderniserkennung. Bei der Initialisierung wird die Anzahl der zu berechnenden Samplepoints festgelegt, dabei richtet sich der Wert an der Anzahl der

Reihen und Spalten der Disparity Map Matrix um eine äquidistante Verteilung gewährleisten zu können. Der hier verwendete Standardfaktor ist 8 (siehe Formel REF). Dies sorgt unter anderem für eine gleiche Anzahl von Samplepoints im gebinnten und ungebinnten Aufnahmemodus, da die jeweilige Menge relativ zur Bildgröße ist. Auf eine Verteilung der Messpunkte am direkten Rand der Disparity Map wurde bewusst verzichtet, da dieser aufgrund der Rektifizierung eher Fehler aufweist als Bereiche in der Mitte des Bildes.

Ein Samplepoint umfasst im entwickelten System entweder einen Pixel oder einen Zusammenschluss mehrerer Pixel. Die dabei festgelegte Ausgangskordinate entspricht dabei entweder dem eigentlichen Punkt, oder dem Mittelpunkt einer festgelegten Region of Interest. Da ein Pixel gerade bei Bildern mit hoher Auflösung weniger Aussagekraft besitzt als der Pixel inklusive der lokalen Umgebung wurde ein Samplepoint auf diese Weise definiert. Der Wert eines einzelnen Samplepoints  $S$  mit dem Zentrum  $C$  ergibt sich im Falle eines Radius  $r = 3$  aus dem Mittelwert des Quadrats  $Q$  mit den Eckpunkten  $Q_{tl} = (C_x - r, C_y - r)$  und  $Q_{br} = (C_x + r, C_y + r)$ . Dies erhöht zwar die Anzahl der zu betrachtenden Pixel welche jedoch, in Abhängigkeit vom gewählten Radius, geringer ist als die Gesamtzahl aller Bildpunkte. Dieser Prozess wird in Abbildung 10 verdeutlicht.



**Abbildung 10:** Darstellung der Samplepoints mit verschiedenen Radien

Nachdem zunächst alle Samplepoints initialisiert wurden, werden die jeweiligen Werte pro Frame aktualisiert. Die eigentliche Erkennung erfolgt dabei

einerseits durch die Erstellung einer Pointcloud welche ebenfalls die selben Dimensionen wie die zugrunde liegende Disparity Map besitzt. Erweiternd dazu lässt sich mithilfe der jeweiligen Punkte das als aktuell am gefährlichsten einzustufende Hindernis sowie dessen Position bestimmen. Dabei werden Samplepoints innerhalb der Gefahrenzone zunächst nur auf Bildebene betrachtet. Zunächst wird die Liste an Samplepoints der errechneten Disparität sortiert. Da generell Hindernisse direkt vor dem UAV als am kritischsten zu betrachten sind, berechnet der Algorithmus den zweidimensionalen Abstand zwischen Bildhauptpunkt und den Samplepoints mit der geringsten Disparität.

# Kapitel 5

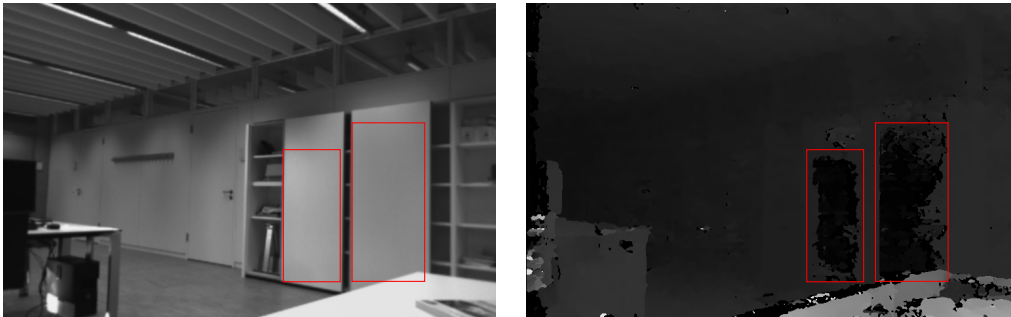
## Limitierungen und Lösungsansätze

In der Erkennung von Hindernissen mithilfe von passiv optischen Systemen können verschiedenste Faktoren der Grund für eine fehlerhafte Erkennung sein. Sei es die Berechnung einer Disparity Map von Bereichen mit einer Vielzahl homogener oder reflektierender Flächen oder die Veränderung der Lichtverhältnisse in einem der beiden Kamerabilder. In diesem Kapitel werden einige der bestehenden, sowie einige potentiell mögliche Limitierungen erläutert sowie dazugehörige Lösungsansätze entwickelt.

### 5.1 Bestehende Konflikte und Lösungsansätze

Die Validierung der erkannten Hindernisse ist ein kompliziertes Problem in der autonomen Hinderniserkennung. Durch etwaige äußere Einflüsse wie die Veränderung der Lichtverhältnisse kann die Berechnung der Tiefenkarte fehlerhaft sein. Dies kann im Fall eines autonomen Flugs dazu führen, dass das UAV ein Hindernis innerhalb eines Korridors erkennt und aufgrund dessen versucht diesem imaginären Hindernis auszuweichen. Gerade in engen Umgebungen ohne viel verfügbaren Platz kann dies zum totalen Systemausfall führen. Daher sollte ausgewertet werden ob es sich bei erkannten Hindernissen auch um solche handelt. Ein Ansatz zur Vermeidung solcher Falscherkennungen wäre, die Objekte insofern zu verfolgen, dass für jeden Frame (in Abhängigkeit der zugrundeliegenden Framerate) überprüft wird, ob im vorherigen Frame bereits ein Objekt gefunden wurde. Erst nachdem dies sichergestellt wurde, wird daraufhin eine Warnung ausgegeben. Zeitgleich wäre es möglich, dass Informationen wie die Eigenbewegung der Drohne sowie die Information, ob sich das erkannte Objekt selber im Raum bewegt, verloren gehen.

Eine andere Problematik stellt die Erkennung von homogenen, spiegelnden sowie durchsichtigen Flächen dar. Aufgrund fehlender Texturen sowie vorhandener Pixeldifferenzen kann die Korrespondenz zweier Pixel unter Umständen nicht bestimmt werden. Diese homogenen Bereiche haben demnach fehlerhafte Informationen zur Folge, so kann einerseits einem Pixel an einer weißen Wand kein zugehöriger Pixel zugeordnet werden, andererseits ist es auch möglich, dass falsche Bildpunkte miteinander gematcht werden, welche wiederum eine hohe Disparität zur Folge haben. In diesem Fall wird ein Hindernis erkannt, obwohl sich keines an dieser Position befunden hat. Abbildung 11 zeigt deutlich, in welchen Bereichen der *SGBM* aufgrund fehlender Textur keine Korrespondenzen finden konnte.



**Abbildung 11:** Konflikt in der Berechnung der Disparität bei nicht texturierten Flächen. Dabei ist (a) das linke aufgenommene Kamerabild und (b) die dazu gehörige Disparity Map. Die in rot markierten Flächen sind das Resultat homogener Texturen.

Auch spiegelnde Flächen sind eine optische Problemstellung. Aufgrund der Projektion eines anderen Bereiches kommt es zu falschen Informationen innerhalb der Disparity Map. Hindernisse können zwar erkannt werden sofern sich die spiegelnde Fläche in einer günstigen Position befindet. Dies ist der Fall, wenn beide Kameras exakt den selben Bereich erfassen. Jedoch wird auch in diesem Fall eine weiter entfernte Distanz wahrgenommen anstelle der des eigentlichen Objektes. Ist dies nicht der Fall so kann auch im Fall spiegelnder Flächen keine Disparität und folglich keine Tiefe wahrgenommen werden.

Eine weitere Fehlerquelle sind durchsichtige Bereiche. Diese vereinen zum einen Fehlerquellen, welche auch bei spiegelnden Arealen auftreten, zum anderen werden Objekte, die sich hinter einer Glasscheibe befinden zwar erkannt (sofern keine Reflexion vorliegt), jedoch das Glas als eigentliches Hindernis nicht. Weiterhin besteht die Möglichkeit, dass Reflexionen aus Perspektive der beiden

Kameras betrachtet einen anderen Tiefeneindruck vermitteln. Objekte werden so entweder weiter entfernt oder auch in kürzerer Distanz erkannt als sie sich wirklich befinden.

Aufgrund des Aspektes, dass die Erkennung des Systems auf den Gebrauch in Innenbereichen konzipiert ist, gibt es verschiedenste Probleme bei der Anwendung in Außenbereichen. Zum einen ist die Verschlusszeit der Kameras fixiert. Die Verwendung einer automatischen Belichtungszeit ist nur für jede Kamera einzeln möglich. Um mit verschiedenen Lichtverhältnissen umgehen zu können, ohne dabei die Möglichkeit der Korrespondenzanalyse zu verlieren, muss die Verschlusszeit daher bei beiden Kameras immer gleich sein. Weiterhin ist die Erkennung des Bodens ein zu betrachtender Aspekt. Dieser ist bei aktueller Berechnung ein Teil der zu betrachtenden Hindernisse, was einerseits seine Berechtigung hat, andererseits auch bei niedrigen Flughöhen zu einer fehlerhaften Erkennung als Hindernis führt.

## 5.2 Diskussion

Ein Ansatz, erkannte Hindernisse zu validieren ist, sich auf die letzte bekannte Position des Objektes zu berufen und im nächsten Einzelbild zu überprüfen, ob es sich noch immer an derselben Position befindet. Dabei würde für jedes Subimage/ jeden Samplepoint ausgewertet werden, ob sich das Hindernis noch innerhalb dessen befindet. Jedoch schränkt dies die Erkennung bewegter Objekte stark ein. Ist die Geschwindigkeit eines Hindernisses so hoch, das in mindestens zwei aufeinanderfolgenden Frames kein Objekt im selben Subimage/Samplepoint erkannt werden kann so würde das System kein Hindernis erkennen können. Diese Methode ist daher prinzipiell möglich, jedoch stark von der Framerate sowie der eigenen bzw. Bewegung des Objektes abhängig. Grundlegend ist das System aufgrund der Beschaffenheit der Subimages für die Mean Disparity Detection geeignet, da die Größe der Subimages eine solche Validierung zulassen würde. Im Falle der Samplepoint Detection müssten auch die unmittelbar benachbarten Messpunkte mit einbezogen werden sodass die Bewegung auch verfolgt werden kann. Eine Kombination von Feature Tracking wäre dabei hilfreich, jedoch sehr rechenaufwändig.

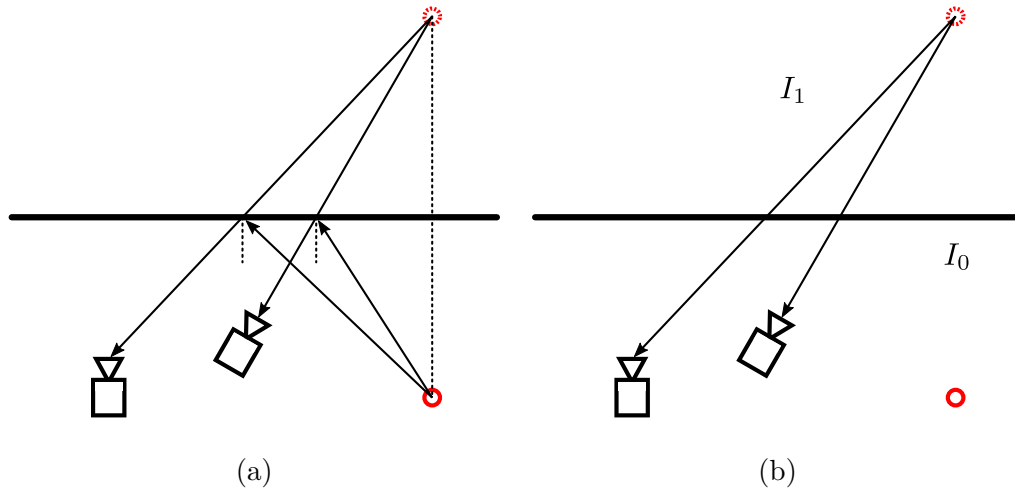
Um auch die Erkennung homogener Flächen gewährleisten zu können, besteht die Möglichkeit, die Szene mithilfe externer Lichtquellen auszuleuchten um eine Texturierung zu erzwingen. Mit Hilfe eines Lasers würde dabei ein zufälliges Punktmuster auf die Szene projiziert werden, durch welches es dem Matching Algorithmus möglich ist, korrespondierende Punkte innerhalb eines



nicht texturierten Bereiches zu finden. Die Zufälligkeit der Punktwolke ist dabei ein wichtiges Kriterium, da die Gleichmässigkeit eines Rasters zu falschen Korrespondenzen führen könnte. Die Anwendung dieses Verfahrens ist jedoch hauptsächlich in Innenbereichen möglich, da die zu erkennenden Entfernungen unter Betrachtung der Einzelbild-Auflösung eine Detektion dieser zulassen würden. In Aussenbereichen ist es prinzipiell schwer, Hindernisse aufgrund anderer Lichtverhältnisse ausfindig zu machen. Für diesen Anwendungsfall würde sich eine Berechnung homogener Flächen auf Matching Ebene anbieten.

In vielen Fällen sind gefundene Stereo Korrespondenzen nicht eindeutig, so beschreiben Geiger et al. [Geiger et al., 2011] die Berechnung solcher im ELAS (Efficient Large-scale Stereo) Algorithmus. Zu Beginn des Matching Verfahrens wird eine karge Menge an Hilfspunkten berechnet. Diese sind dabei als Pixel definiert, welche aufgrund der gegebenen Textur oder ihrer Einzigartigkeit robust gematcht werden können. Dabei werden solche Support Points als eine Konkatenation der Koordinaten des Pixels  $(x_m, y_m)$  sowie der zugehörigen Disparität  $d_m$  definiert. Die Verteilung der Hilfspunkte auf den Referenzbildern erfolgt in gleichem Abstand zueinander. Unklare Matches werden währenddessen anhand eines Thresholds aussortiert. Anschliessend wird unter Zuhilfenahme der erstellten Hilfspunkte nach sogenannten *Observations* gesucht, welche aus den Bildkoordinaten  $(x_n, y_n)$  sowie einem Feature Vector  $f_n$  bestehen. Der Feature Vektor ist dabei entweder die Intensität des Pixels oder ein berechneter Deskriptor, der aus den Intensitäten der benachbarten Pixel berechnet wird. Mithilfe dieser beiden Informationen wird ein Gitternetz generiert. Die eigentliche Berechnung der Disparität erfolgt mit Hilfe der Observations auf deren jeweiliger Epipolarlinie unter Benutzung des Maximum a posteriori (MAP) Verfahrens und wird für jedes Bild separat angewandt. Zuletzt werden beide Disparity Maps auf ihre Konsistenz geprüft.

Der von Tsin et al. [Tsin et al., 2003] entwickelte Stereo Matching Algorithmus kann sowohl zwischen Reflexion als auch Transparenz unterscheiden. Dabei wird das grundlegende Problem wie folgt beschrieben: Aufgrund des Reflexionsmodells nach Hero (Abbildung 12(a)) welches besagt das sich Einfalls- sowie Ausfallswinkel, gemessen an der Oberflächennormale gleichen, ist es trivial, dass die Position einer Reflexion eines Szenepunktes unabhängig vom aktuellen Sichtfenster ist, da sich die Reflexion an einer festen Position hinter dem Reflektor befindet. Daher kann die Abhängigkeit eines Szenepunktes von seiner Reflexion ignoriert werden. Tsin et al. beschreiben ein geschichtetes Modell in dem das reflektierende Objekt als vordere Ebene  $I_0$  und die Reflexion selber als hintere Ebene  $I_1$  dargestellt ist (Abbildung 12(b)).



**Abbildung 12:** (a) beschreibt die physikalische Reflexion nach Hero, (b) das von Tsin et al. verwendete geschichtete Modell

Die aufgenommenen Bilder sind demnach eine Zusammensetzung beider Ebenen. Selbiges gilt für nicht planare Flächen sowie Durchsichtigkeit solange die Translation zwischen den einzelnen Bildern gering ist. Während des Algorithmus werden die relativen Bewegungen der Ebenen zwischen aufeinanderfolgenden Frames anhand der verschiedenen Tiefen analysiert.

Ein weiterer, bereits in Abschnitt 5.1 angeschnittener Punkt ist die Erkennung des Bodens als Hindernis. Dabei stellt sich die Frage, ob eine Erkennung als Hindernis gewünscht ist oder nicht. Im Falle einer geringen Flughöhe wird der Boden ebenfalls als Hindernis erkannt. Dies kann einerseits eine Absicherung dafür sein, dass die durch interne Sensoren ermittelte Flughöhe korrekt ist, andererseits kann es dazu führen, dass der Algorithmus zur Vermeidung von Hindernissen versucht, dieses zu umgehen. Dies könnte unter Umständen dazu führen, dass das UAV eine Flughöhe annimmt, welche über der durch die Umgebung gegebenen maximalen Flughöhe liegt. Eine mögliche Lösung wäre die Erkennung des Bodens in Abhängigkeit der aktuellen Höhe zu gestalten. Sollte sich das UAV unterhalb der zur Erkennung des Bodens nötigen Flughöhe befinden, so wird der Boden nicht weiter betrachtet. Sofern die Flughöhe die Erkennung des Bodens ausschließt wird dieser entfernt.

# Kapitel 6

## Evaluierung

### 6.1 Testsetup

Bei der Durchführung der Tests befinden sich die Kameras statisch im Raum. Die zu erkennenden Hindernisse werden innerhalb und ausserhalb der zu erkennenden Reichweite platziert, wobei die Ausrichtung der Hindernisse teils zufällig, teils bewusst an kritischen Positionen erfolgt, um ein reales Anwendungsszenarien passend zu simulieren. Ein aufgenommenes Testset besteht dabei aus beiden Bildern der Kamera, der normalisierten Disparity Map sowie eine komplette Pointcloud dieser um etwaige Fehler der Algorithmen leichter erkennen zu können, sowie den geloggtten Pointclouds der Hinderniserkennung. Weiterhin werden diverse Parameter gespeichert, wie die Anzahl der erkannten Hinderniselemente, sowie deren Disparitäten.

Der zu erkennende Bereich wurde auf 0,2 bis 1,5 Meter definiert. Dies entspricht einem Szenario in welchem das System auch aufgrund der hohen Framerate der Erkennung angewendet werden kann. Eine Erweiterung dessen auf beispielsweise 2.0 Meter wurde nicht durchgeführt, da der Algorithmus auch bei großen Entfernungen robuste Werte in der Distanzberechnung liefert [Al-Hallak and Hiller, 2015].

Ein aufgenommenes Testset besteht aus jeweils 12 Testbildern. Für jede Methode wurden drei verschiedene Hindernisgrößen getestet, groß, klein und winzige Hindernisse. Anhand dieser wird ausgewertet welche minimale, maximale sowie mittlere Disparität, und daraus resultierende Distanz erkannt wird.

Weitere Tests beinhalten die Erkennung winziger Hindernisse unter Veränderung der *SGBM* Parameter. Dabei wird unter anderem untersucht ob beispielsweise eine verringerte Blockgröße Einfluss auf die Erkennung kleiner Bereiche nimmt. Des Weiteren wird die Zeit für die Hinderniserkennung eines Frames

untersucht um eine durchschnittliche Zeit für die Erkennung sowie die daraus resultierende Framerate zu ermitteln. Dies geschieht einerseits durch die Erkennung eines Hindernisses, welches sich über das gesamte Bild ausbreitet, andererseits für nur ein Teilelement jeder Erkennungsmethode (Subimage, Samplepoint).

Zudem wird geprüft inwiefern die Algorithmen mit Limitierungen des *SGBM* umgehen können. Dazu zählen die Erkennung bei spiegelnden, reflektierenden und durchsichtigen Flächen, sowie die Erkennung schwach texturierter Hindernisse.

## **6.2 Evaluierung Subimage Detection**

### **6.2.1 Robustheit**

### **6.2.2 Performanz**

## **6.3 Evaluierung Samplepoint Detection**

### **6.3.1 Robustheit**

### **6.3.2 Performanz**

# Kapitel 7

## Diskussion

7.1 Anforderungsevaluierung

7.2 Gegenüberstellung beider Algorithmen

## Kapitel 8

### Fazit und zukünftige Arbeiten

# Literaturverzeichnis

- [Al-Hallak and Hiller, 2015] Al-Hallak, M. and Hiller, H. (2015). Fast depth map estimation from stereo camera systems.
- [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.
- [Bradski and Kaehler, 2008] Bradski, D. G. R. and Kaehler, A. (2008). *Learning OpenCV, 1st Edition*. O’Reilly Media, Inc., first edition.
- [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.
- [Cyganek and Siebert, 2011] Cyganek, B. and Siebert, J. P. (2011). *An introduction to 3D computer vision techniques and algorithms*. John Wiley & Sons.
- [Geiger et al., 2011] Geiger, A., Roser, M., and Urtasun, R. (2011). Efficient large-scale stereo matching. In *Computer Vision–ACCV 2010*, pages 25–38. Springer.
- [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.
- [Middlebury Univeristy, 2015] Middlebury Univeristy (2015). Middlebury stereo datasets.
- [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.
- [Technologies, 2015] Technologies, A. (2015).
- [Tsin et al., 2003] Tsin, Y., Kang, S. B., and Szeliski, R. (2003). Stereo matching with reflections and translucency. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–702. IEEE.
- [Zureiki et al., 2008] Zureiki, A., Devy, M., and Chatila, R. (2008). *Stereo Matching and Graph Cuts*. INTECH Open Access Publisher.