

TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Mensch-Maschine-Kommunikation

Echtzeitfähige aktive Stereoskopie für technische und biometrische Anwendungen

Daniel Modrow

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik
der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. R. Kötter, Ph.D.

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. habil. G. Rigoll
2. Univ.-Prof. Dr.-Ing., Dr.-Ing. habil. A. Knoll

Die Dissertation wurde am 17.06.2008 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 27.11.2008 angenommen.

Vorwort

Die vorliegende Arbeit ist das Ergebnis meiner dreieinhalbjährigen Tätigkeit als freier Mitarbeiter der Siemens AG in Zusammenarbeit mit dem Lehrstuhl für Mensch-Maschine-Kommunikation der TU München.

Bei meinem Doktorvater und Leiter des Lehrstuhls, Herrn Professor Gerhard Rigoll, möchte ich mich herzlich für die Betreuung während dieser Arbeit bedanken. Vor allem weiß ich es sehr zu schätzen, dass er trotz seiner oft spürbaren Termindichte stets ein offenes Ohr für mich hatte und als Ansprechpartner für alle aufkommenden Fragen zeitnah zur Verfügung stand.

Ebenso möchte ich mich bei den Mitarbeitern der Abteilung für „Machine Vision & Nondestructive Testing“ der Siemens AG in München-Neuperlach bedanken – allen voran bei meinem Betreuer, Herrn Dr. Claudio Laloni, der mich während der gesamten Zeit begleitete und mit wertvollen Hinweisen und Impulsen maßgeblich zum Erfolg dieser Arbeit beigetragen hat. Auch den übrigen Kollegen der Arbeitsgruppe von Herrn Dr. Frank Forster gebührt mein außerordentlicher Dank für viele wichtige fachliche Diskussionen und besonders die positive Atmosphäre im Arbeitsalltag.

Mein Dank gilt jedoch auch noch weiteren Menschen, ohne die diese Arbeit kaum zu einem erfolgreichen Ende hätte finden können:

Er gilt meinen Freunden, die trotz meiner häufigen Zurückgezogenheit in der Endphase stets mit guter Laune bei mir waren und auch durch fachfremde Betrachtung meiner Arbeit wichtige Denkanstöße geliefert haben. Er gilt meiner Freundin, die mir während der gesamten Zeit zur Seite stand und auf deren Rückhalt ich auch trotz fehlender Zeit immer zählen konnte.

Und er gilt vor allem meiner Familie – meinem verstorbenen Vater, dessen naturwissenschaftliche Gene offenbar an mich fielen, meiner Mutter, ohne deren Unterstützung mein Studium und meine Doktorandentätigkeit nicht durchführbar gewesen wären und allen übrigen Familienmitgliedern, deren Beistand stets zu spüren war – was mir sehr viel bedeutet hat.

Ottobrunn, im Juni 2008

Daniel Modrow

Zusammenfassung

Diese Arbeit befasst sich mit der dreidimensionalen Erfassung generischer Objekte. Der Schwerpunkt liegt dabei auf einer echtzeitfähigen Sensorik für biometrische Anwendungen, von der auch technische Implementierungen profitieren können. Als Basis dient die aktive Stereoskopie, die ideale Voraussetzungen für ein berührungsloses Messverfahren bietet. Es wird ein Überblick über den aktuellen Stand der Technik im Bereich der Tiefen- und Entfernungsmessung gegeben, der auch die Grundlagen für die neu entwickelten Ansätze bereitstellt. Darüber hinaus liefert diese Arbeit einen Ausblick auf mögliche Weiterentwicklungen im Bereich der dreidimensionalen Objekterfassung.

Inhalt

1	EINLEITUNG	1
1.1	Motivation	1
1.2	Aufbau der Arbeit	2
2	GRUNDLAGEN DER DREIDIMENSIONALEN OBJEKTERFASSUNG.....	5
2.1	Grundlagen der zweidimensionalen Bilderfassung.....	5
2.1.1	Kamera-Modelle.....	5
2.1.1.1	Lochkamera	6
2.1.1.2	Linsenkamera	8
2.1.2	Sensoren zur digitalen Bilderfassung	11
2.1.2.1	CCD-Sensoren.....	12
2.1.2.2	CMOS-Sensoren.....	12
2.1.2.3	Farberzeugung	13
2.1.3	Kamera-Kalibrierung.....	15
2.2	Grundlagen der 3D-Messtechnik	19
2.3	Allgemeine Methoden der 3D-Objekterfassung	21
2.3.1	Ultraschall-Verfahren und Laufzeitmessung.....	21
2.3.2	Time-Of-Flight	22
2.3.3	Interferometrie	24
2.4	Passive Stereoskopie.....	25
2.4.1	Depth from Focus	25
2.4.2	Shape from Shading	26
2.4.3	Passives Stereo	28
2.5	Aktive Stereoskopie.....	32
2.5.1	Grundprinzip.....	32
2.5.2	Laservermessung und Lichtschnitt.....	33
2.5.3	Codiertes Licht	36
2.5.3.1	Prinzip.....	36
2.5.3.2	Zeitcodierte Verfahren.....	37
2.5.3.3	Phasenverschiebung	40
2.5.3.4	Direkte Codierung	42

2.5.3.5	Farbcodierte Verfahren	43
3	SENSOREN ZUR DREIDIMENSIONALEN GESICHTSERFASSUNG	47
3.1	Anpassung des Wellenlängenbereichs	47
3.2	Reduktion der Farbkanäle	51
3.2.1	Auswirkungen einer Kanalreduktion auf die Codierung.....	53
3.2.2	Robustheit gegenüber Störeinflüssen	55
3.3	Aktive NIR-Stereoskopie unter Verwendung zweier Wellenlängen	60
3.3.1	Erläuterung des Prinzips.....	60
3.3.2	Erzeugen eines farbcodierten Musters aus zwei Farbkanälen.....	61
3.3.3	Aufbau einer Projektionseinheit zur infraroten Musterdarstellung	62
3.3.4	Kamera-Seite: Aufbau und Funktion	64
3.3.5	Aufbau und Test des Prototypen	67
3.4	Aktive NIR-Stereoskopie unter Verwendung einer Wellenlänge	69
3.4.1	Erläuterung des Prinzips.....	69
3.4.2	Projektions-Seite: Aufbau und Funktion.....	70
3.4.3	Aufbau und technische Daten	73
3.5	Erweiterung und Ergänzung bisheriger Projektionsverfahren.....	76
3.5.1	A-priori-Wissen über das aufzunehmende Objekt	76
3.5.2	Datenrekonstruktion und Fehlerkorrektur	77
3.5.3	Kalibrierung des Systems	78
3.6	Evaluierung der aktiven NIR-Stereoskopie	79
3.6.1	Messfehler eines Stereosystems	79
3.6.2	Statistische Auswertung des Messfehlers	83
3.6.3	Sampling und Bildrekonstruktion	85
3.6.3.1	Testumgebung	85
3.6.3.2	Code-abhängige Rekonstruktionsfehler	86
3.6.3.3	Unregelmäßige Tiefenabweichungen.....	86
3.6.4	Robustheit gegenüber Umgebungslicht	91
3.6.5	Untersuchung verschiedener Kameratypen.....	94
4	ANWENDUNGSGEBIETE DER AKTIVEN (INFRAROTEN) STEREOSKOPIE	97
4.1	Biometrie und 3D-Gesichtserkennung	97

4.2	Biologische Analysen	101
4.3	Multimediale Anwendungen.....	103
5	ERWEITERTE ANSÄTZE ZUR AKTIVEN 3D-OBJEKTERFASSUNG	105
5.1	Dynamische NIR-Stereoskopie.....	105
5.2	Hybride intensitätsbasierte Laufzeitmessung.....	108
5.3	Radiale Lichtcodierung.....	113
5.3.1	Grundsätzliche Überlegungen	113
5.3.2	Mehr-Kamera-System mit kegelförmiger Beleuchtung	115
5.3.3	Mehr-Kamera-System mit telezentrischer Beleuchtung	118
6	ZUSAMMENFASSUNG UND AUSBLICK	121
	Übersicht häufig verwendeter Abkürzungen.....	125
	Übersicht häufig verwendeter Symbole.....	127
	Literaturverzeichnis	129

1 Einleitung

1.1 Motivation

Das Erfassen dreidimensionaler Daten ist ein Problem, das die Wissenschaft und Technik seit dem Bestehen von Methoden zur Aufnahme von Bildern beschäftigt. Die Entwicklung von Verfahren zum Lösen dieser Aufgabe ist dabei getrieben von dem Wunsch, eine für uns Menschen selbstverständliche Fähigkeit auf eine technische Schnittstelle einer Maschine zu übertragen. Denn dies eröffnet die Chance, ein tatsächliches Abbild unserer Umwelt und Realität zu erzeugen, zu nutzen, zu bearbeiten und zu speichern.

Von solchen Techniken profitieren etliche Bereiche der Wissenschaft und Industrie. Dreidimensionale Modelle von Objekten können beleuchtungs- und richtungsunabhängig betrachtet werden und erlauben völlig neue Herangehensweisen der Bildverarbeitung und -nachbearbeitung. Für Verfahren der industriellen Teileinspektion etwa können Fertigungsstücke direkt mit ihren zugrundeliegenden Modellierungen verglichen werden, die Übergabe manuell gefertigter Modelle und Muster an Rapid-Prototyping Systeme kann direkt erfolgen.

Aber nicht nur das Erfassen technischer Gegenstände ist ein Schwerpunkt der dreidimensionalen Objekterfassung, sondern auch die Aufnahme und Modellierung des menschlichen Gesichts, des Körpers, unserer Mimik und Bewegungen. Viele Bereiche der Unterhaltungsindustrie nutzen Studien der menschlichen Physiognomie und versuchen, diese möglichst genau zu erfassen oder nachzuahmen. So werden etwa für Animationen in Spielfilmen menschliche Bewegungen durch Aufbringen von Markern an bestimmten Punkten des Körpers aufgenommen und auf virtuelle Charaktere übertragen. Schwieriger gestaltet sich für diese Zwecke jedoch die Aufnahme des menschlichen Gesichts. Dabei ist dies für einen Betrachter der wichtigste Faktor, um künstlich geschaffene Avatare als Abbilder der realen Welt zu erkennen und anzunehmen. Die Auftrittshäufigkeit solcher virtueller Persönlichkeiten hat vor allem mit der wachsenden Anzahl an multimedialen Anwendungen zugenommen. Es werden Modellierungen von Menschen in Edutainment-Systemen, Computer-Spielen oder in vollständig virtuellen Umgebungen des Internets

verwendet, deren erklärtes Ziel in der Tat die Schaffung einer vollständigen virtuellen Realität ist.

Aus den für Unterhaltungszwecke angewendeten exakten Abbildungen des menschlichen Gesichts und Körpers ziehen aber noch weitere Anwendungen einen Nutzen. Mit dem wachsenden Bedürfnis nach Sicherheit im globalen und persönlichen Umfeld setzen Überwachungssysteme immer häufiger auf biometrische Schnittstellen zur zuverlässigeren Identifikation von Personen. So können beispielsweise bewährte Verfahren, wie etwa die Gesichtserkennung, allein schon durch die bloße Existenz dreidimensionaler Daten von einer erhöhten Betrugssicherheit profitieren. Zudem eröffnet eine weitere Dimension Möglichkeiten für vollkommen neue Algorithmen und Verfahren, um bisherige Standards verbessern oder übertreffen zu können.

Für all die zunächst genannten technischen Verfahren ist eine möglichst exakte und genaue Aufnahme ein wichtiger Punkt. Bei den biometrischen Anwendungen spielt das Aufnahmeobjekt – der Mensch – zusätzlich eine entscheidende Rolle. Denn hierbei müssen etliche Punkte beachtet werden: Die Reaktionen auf Messverfahren, die nicht berührungslos arbeiten können und Haut- oder Körperkontakt erfordern, nicht zu vermeidende Bewegungen, eine eventuelle Schädigungen der Gesundheit durch die eingesetzten Verfahren oder die psychische Auswirkung und Akzeptanz von vollständig abbildenden Systemen. Die meisten bisherigen Ansätze zur dreidimensionalen Objekterfassung haben diese Punkte dabei nicht spezifisch behandelt und sind deshalb nur bedingt als biometrische Schnittstellen einsetzbar.

Das Ziel dieser Arbeit ist die Erforschung von Verfahren, die zum Aufbau eines Systems geeignet sind, das als biometrische Schnittstelle die gezielte dreidimensionale Aufnahme des menschlichen Gesichts und Körpers berührungslos und mit hoher Genauigkeit ermöglicht. Die entwickelten Techniken sollen dabei jedoch nicht einschränkend oder nur spezialisiert einsetzbar, sondern ebenso geeignet und vorteilhaft für viele technische Bereiche und Anwendungen sein. Es werden hierfür bekannte Verfahren der Tiefen- und Entfernungsmessung mit deren grundlegenden Theorien untersucht und daraus neue Methoden entwickelt, die die dreidimensionale Aufnahme von Objekten für ein möglichst breites Spektrum an Anwendungsgebieten ohne große Einschränkungen ermöglichen. Dabei entstehende Prototypen von Sensoren werden bekannten Verfahren gegenübergestellt und hinsichtlich ihrer Leistungsfähigkeit evaluiert.

1.2 Aufbau der Arbeit

Im zweiten Kapitel werden verschiedene Verfahren zur dreidimensionalen Erfassung von Objekten vorgestellt. Dabei werden die Grundlagen passiver und aktiver stereoskopischer Ansätze, aber auch andere Verfahren zur Entfernungs- und Tiefenmessung näher beschrieben. Grundsätzliche Methoden und Vorgehensweisen, die zur Verarbeitung aufge-

nommener Bilddaten nötig sind, werden ebenfalls betrachtet, da diese die Basis für weitere Überlegungen und Entwicklungen bilden. Es wird somit in diesem Kapitel eine Übersicht über die Entwicklungsgeschichte und den Stand der aktuellen Technik dargestellt, um einen geeigneten Einstieg in die dreidimensionale Objekterfassung zu erhalten.

Der Aufbau von Systemen zur dreidimensionalen Gesichtserfassung ist der Schwerpunkt des dritten Kapitels. Hierbei wird das erklärte Hauptziel dieser Arbeit im Wesentlichen behandelt. Es werden die entscheidenden Faktoren betrachtet, die bei der Aufnahme von Gesichtern und lebenden Personen beachtet werden müssen. Entsprechende Grundlagen und die Reaktionen des menschlichen Körpers auf bestimmte Techniken werden ebenfalls untersucht. Neue Ansätze werden anhand der Grundlagen des zweiten Kapitels vorgestellt, deren Entwicklung dabei beispielsweise durch neue Beleuchtungstechniken oder Anpassung der Verarbeitungs- und Berechnungsschritte ergänzt wird. Die Ergebnisse werden schließlich in einer Evaluierung geprüft und die aufgebauten Prototypen mit technisch aktuellen Systemen verglichen.

Im vierten Kapitel werden einige Anwendungsgebiete der beschriebenen Verfahren dargestellt und Forschungsprojekte präsentiert, an denen im Rahmen dieser Arbeit mitgewirkt wurde. Diese finden sich dabei in verschiedenen Bereichen der Wissenschaft, Technik, aber auch der Biologie wieder.

Einen Ausblick auf weitere, neue Ansätze zur dreidimensionalen Objekterfassung bietet das fünfte Kapitel. Hierbei werden Ansätze vorgestellt, die bisher vorgestellte Verfahren der technischen oder biometrischen dreidimensionalen Objekterfassung entscheidend erweitern oder verbessern. Es werden dabei sowohl Alternativen der Musterprojektion, als auch die Kombination verschiedener Verfahren und Methoden betrachtet, um die Messgenauigkeit oder Messfeldgröße zu erhöhen.

Im sechsten Kapitel werden die Ergebnisse dieser Arbeit diskutiert und abschließend zusammengefasst. Des Weiteren wird ein Ausblick auf themenverwandte, zukünftige Erweiterungen gegeben.

2 Grundlagen der dreidimensionalen Objekterfassung

Der Wunsch, unsere dreidimensionale Welt auch digital dreidimensional abzubilden, besteht bereits seit den Anfängen der digitalen Bildverarbeitung und ist in vielen technischen Bereichen sinnvoll und wünschenswert. Besonders in den letzten Jahren wurden, vor allem aufgrund der immer steigenden Rechen- und Speicherkapazitäten, aber auch durch immer günstigere Bildsensoren, einige Verfahren entwickelt, die diese dreidimensionale (3D) Erfassung unserer Umwelt ermöglichen. Dieses Kapitel gibt einen Überblick über die Entwicklung dieser Verfahren sowie den aktuellen Stand der Technik einiger bereits technisch oder industriell eingesetzter Methoden.

2.1 Grundlagen der zweidimensionalen Bilderfassung

Das Aufnehmen von Bildern, also die zweidimensionale Abbildung unserer Umgebung, ist ein unumgänglicher Bestandteil der optischen Messtechnik. Auch die gängigsten Verfahren zur dreidimensionalen Objekterfassung beruhen in den meisten Fällen auf zweidimensionalen Abbildungen. Deshalb ist es zwingend nötig, die Grundlagen der Bilderfassung im Allgemeinen, wie die optischen Eigenschaften der Objektabbildung, und der digitalen Bilderfassung im Speziellen zu verstehen. Erst dadurch wird eine direkte rechnergestützte Bildverarbeitung möglich. Im Folgenden werden diese Grundlagen beschrieben, so dass im Anschluss daran eine Kamera als Messgerät zur 3D-Objekterfassung dienen kann.

2.1.1 Kamera-Modelle

Um die Daten einer Kamera oder eines bilderfassenden Sensors für anschließende Berechnungen zugänglich zu machen, muss die Geometrie der Kameraoptik und damit die Abbildung einer Szene oder eines Objekts auf die Bildebene beschrieben werden. Hierfür

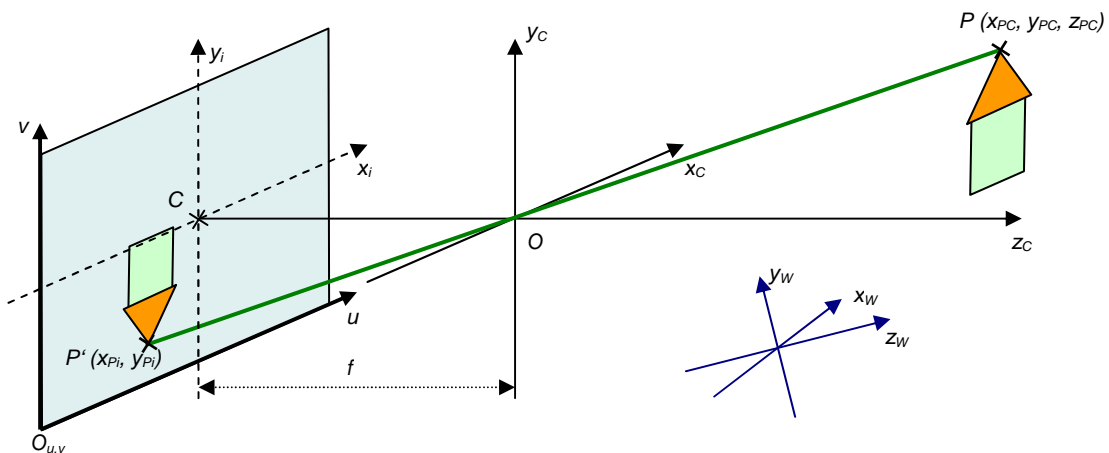


Abbildung 1 Abbildungsmodell einer Lochkamera. Der optische Pfad verläuft durch den Ursprung O des Kamerakoordinatensystems (das Loch) auf die Bildebene, deren Abstand zu O durch die Brennweite f definiert ist.

existieren verschiedene Modellierungen, von denen die beiden gängigsten im Folgenden beschrieben werden: Das Lochkameramodell und die projektive Kamera (Linsenkamera).

2.1.1.1 Lochkamera

Eine der einfachsten geometrischen Beschreibungen einer Kamera ist das Lochkamera-Modell, oder auch die sogenannte „camera obscura“. Abbildung 1 zeigt die schematische Darstellung einer Lochkamera. Die Kamera definiert sich durch ihr optisches Zentrum O , welches ein infinitesimal kleines Loch innerhalb einer lichtundurchlässigen Schicht beschreibt, durch das allein Licht auf die Bildebene trifft. In diesem optischen Zentrum liegt auch das Zentrum des Kamera-Koordinatensystems (x_C, y_C, z_C) , dessen Achsen x_C und y_C parallel und z_C senkrecht zur Bildebene verlaufen. Der Abstand der Bildebene zum optischen Zentrum ist beschrieben durch die Brennweite f . Der Schnittpunkt der optischen Achse z_C mit der Bildebene definiert das Bildzentrum C , welches wiederum das geometrische Zentrum des Sensorkoordinatensystems (x_i, y_i) darstellt. Die Bildkoordinaten der Bildebene werden durch die Achsen u und v beschrieben. Bei digitalen Sensoren (siehe auch 2.1.2) werden diese Koordinaten üblicherweise als Pixel (Picture Element) bezeichnet und beziehen sich auf die diskrete Position eines Bildpunkts innerhalb der Sensorfläche.

Durch das optische Zentrum O erfolgt die Abbildung eines beliebigen Punktes P mit den Koordinaten (x_{PC}, y_{PC}, z_{PC}) auf den Bildpunkt P' mit (x_{Pi}, y_{Pi}) . Dies stellt eine perspektivische Projektion des Punktes P dar und gehorcht folgender Beziehung:

$$\frac{x_{Pi}}{-f} = \frac{x_{PC}}{z_{PC}} \quad \text{und} \quad \frac{y_{Pi}}{-f} = \frac{y_{PC}}{z_{PC}} \quad (1)$$

In homogenen Koordinaten mit Matrixdarstellung lässt sich diese Beziehung durch Anwendung der homogenen Komponente h wie folgt formulieren:

$$\begin{bmatrix} x_{Pi} \cdot h \\ y_{Pi} \cdot h \\ h \end{bmatrix} = \begin{bmatrix} x'_{Pi} \\ y'_{Pi} \\ h \end{bmatrix} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_{PC} \\ y_{PC} \\ z_{PC} \\ 1 \end{bmatrix} \quad (2)$$

Mit der perspektivischen Projektionsmatrix \underline{P}

$$\underline{P} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3)$$

lässt sich somit die Lochkamera in folgender Matrix-Schreibweise darstellen:

$$\underline{p}_i = \underline{P} \cdot \underline{p}_C \quad (4)$$

Diese Darstellung erlaubt es, die Sensorkoordinaten durch die internen Eigenschaften (intrinsische Parameter) der Kamera an die Bildkoordinaten anzupassen. Auch eine Transformation des Weltkoordinatensystems wird dadurch ermöglicht.

Zur Transformation der Bildkoordinaten in Pixelkoordinaten sind Informationen über die Größe der einzelnen Bildpunkte bzw. die Auflösung der Kamera und die Maße des Sensors erforderlich. Daraus und aus den Koordinaten des Bildzentrums C lässt sich eine neue Projektionsmatrix aufstellen, die eine Translation und Skalierung enthält. Es gilt dann:

$k_u = \text{horizontale Auflösung (Sensorbreite)}$

$k_v = \text{vertikale Auflösung (Sensorhöhe)}$

$$C = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \quad (5)$$

$$\underline{P}' = \begin{bmatrix} -f \cdot k_u & 0 & u_0 & 0 \\ 0 & -f \cdot k_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} u_P \\ v_P \\ h' \end{bmatrix} = \underline{p}_{uv} = \underline{P}' \cdot \underline{p}_C \quad (7)$$

Üblicherweise finden Koordinaten-Beschreibungen im Weltkoordinatensystem x_W, y_W, z_W statt. Durch Rotation und Translation lassen sich die verschiedenen Koordinatensysteme transformieren, so dass sich für die Bildkoordinaten aus Gleichung (1) durch die Abbildung der Lochkamera folgende Beschreibung in Weltkoordinaten ergibt:

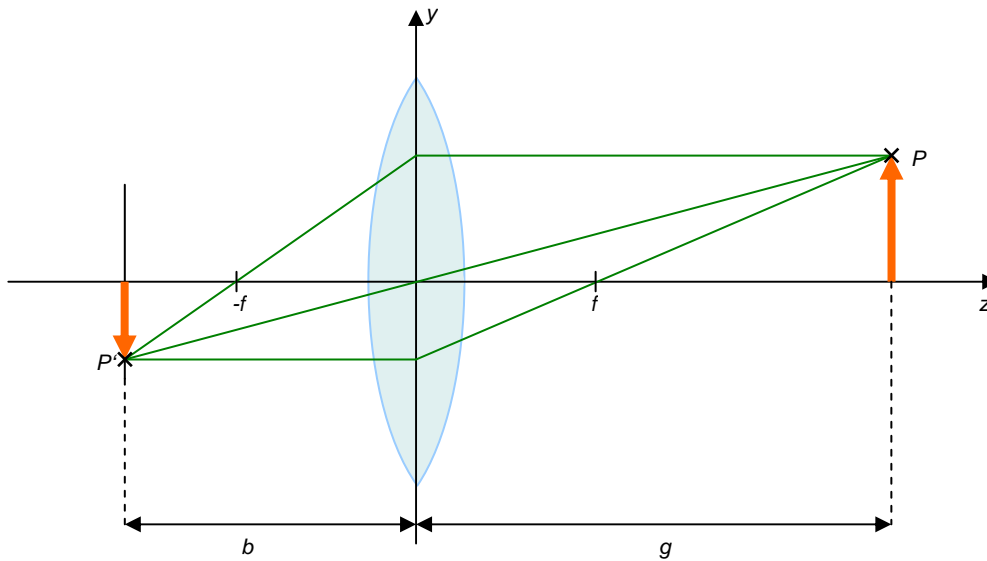


Abbildung 2 Abbildungsmodell einer Linsenkamera. Die Abbildungseigenschaften werden durch den Schliff der Linse und die daraus resultierende Brennweite f bestimmt.

$$x_{Pi} = f \frac{r_{11}x_{PW} + r_{12}y_{PW} + r_{13}z_{PW} + t_x}{r_{31}x_{PW} + r_{32}y_{PW} + r_{33}z_{PW} + t_z} \quad \text{und} \quad y_{Pi} = f \frac{r_{21}x_{PW} + r_{22}y_{PW} + r_{23}z_{PW} + t_y}{r_{31}x_{PW} + r_{32}y_{PW} + r_{33}z_{PW} + t_z} \quad (8)$$

Die Parameter $[r_{11} \dots r_{33}]$ sind hierbei die Elemente der Rotationsmatrix \underline{R} , die die relative Orientierung des Kamerakoordinatensystems zum Weltkoordinatensystem beschreibt. Entsprechend bezeichnet der Translationsvektor $[t_x, t_y, t_z]$ den Versatz der Koordinatenursprünge zueinander.

Eine Lochkamera ist ihrer Theorie nach nicht fokussierend. Das bedeutet, alle Objekte im Sichtfeld der Kamera werden exakt und verzerrungsfrei abgebildet, es gibt keinen Schärfen- und Unschärfebereich. Hierfür wird eine infinitesimal kleine Öffnung innerhalb eines infinitesimal dünnen Trägermaterials zur Abbildung angenommen, was praktisch nicht realisierbar ist. Würde man tatsächlich durch einen möglichst kleinen Durchmesser abbilden, so erfordert dies eine relativ lange Belichtung einer Szene, da nur wenig Licht die Öffnung durchdringen kann. Dies führt bei nicht-statischen Szenen zwangsläufig zu Bildfehlern. Zudem wird in der Praxis der sichtbare Bereich durch das Verhältnis des Lochdurchmessers und der Lochtiefe, also der Dicke des Materials, in welchem sich das Loch befindet, eingeschränkt. Eine alternative Beschreibung einer Kamera mit günstigeren Belichtungseigenschaften liefert dabei das Modell der Linsenkamera, welches im nächsten Punkt beschrieben wird.

2.1.1.2 Linsenkamera

Eine Linsenkamera (oftmals auch projektive Kamera) ist grundsätzlich als eine Erweiterung des Lochkamera-Modells zu verstehen, wobei die, nur theoretisch erreichbare, infinitesimal kleine Lichtöffnung durch eine feste Linse ersetzt wird. Dadurch gelangt mehr

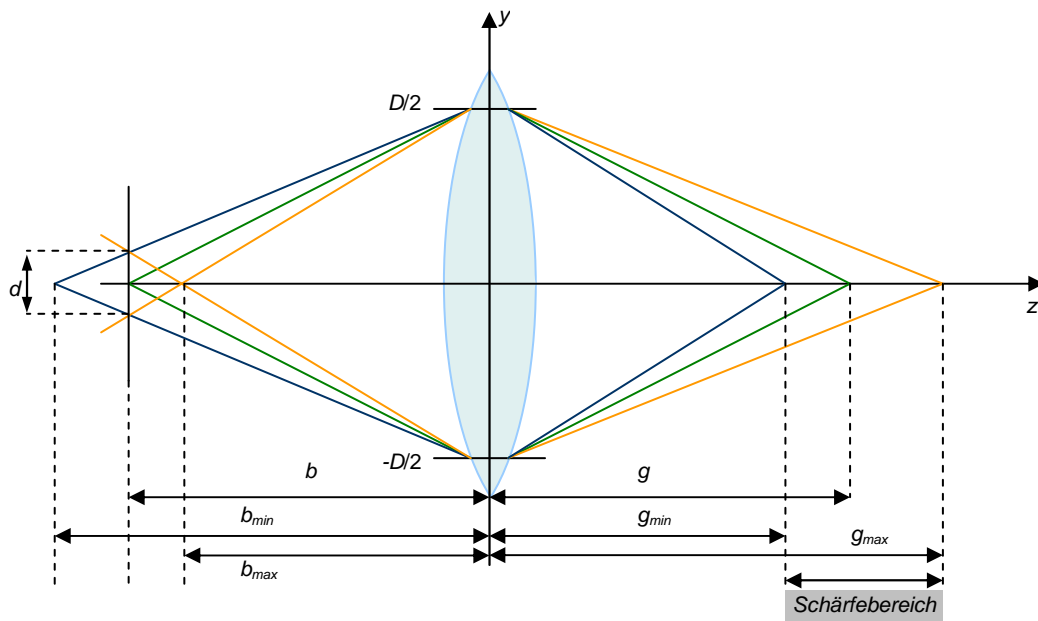


Abbildung 3 Bestimmung des Schärfebereichs einer Linsenkamera. Die minimale (b_{\min}) und maximale (b_{\max}) Bildweite b wird durch den Durchmesser des Unschärfekreises (d) festgelegt, woraus die Grenzen für die Entfernung eines Gegenstandes $g \in [g_{\min}, g_{\max}]$ festgelegt werden, die den Schärfebereich begrenzen.

Licht auf die Bildebene und die Belichtungszeit kann verkürzt werden, allerdings wird nur noch ein bestimmter Bereich scharf abgebildet. Abbildung 2 zeigt die Modellierung einer Linsenkamera.

Der Zusammenhang zwischen der Gegenstandsweite g , der Bildweite b und der Brennweite f wird hierbei durch die Linsengleichung beschrieben:

$$\frac{1}{f} = \frac{1}{b} + \frac{1}{g} \quad (9)$$

Wird ein Punkt unscharf auf die Bildebene abgebildet, so entsteht kein Bildpunkt, sondern ein Bildfleck, der so genannte Unschärfekreis. Der maximal zulässige Durchmesser dieses Kreises bestimmt dabei den Schärfebereich einer Linsen-Kamera. Wie groß dieser in der Praxis sein darf, hängt vom angeschlossenen Bildsensor bzw. der Anwendung ab. Abbildung 3 verdeutlicht den Zusammenhang zwischen der Objektentfernung und der Unschärfe. Neben dem Durchmesser d des Unschärfekreises haben zudem die Blendeneinstellung k mit ihrem Öffnungsdurchmesser D und die Brennweite f einen Einfluss auf den Schärfebereich. Es gilt:

$$k = \frac{f}{D} \quad (10)$$

$$\frac{d}{b - b_{\max}} = \frac{D}{b_{\max}} \quad (11)$$

$$\frac{d}{b_{\min} - b} = \frac{D}{b_{\min}} \quad (12)$$

Hieraus folgt:

$$\frac{d}{D} = \frac{b - b_{\max}}{b_{\max}} = \frac{\frac{f \cdot g}{g - f} - \frac{f \cdot g_{\max}}{g_{\max} - f}}{\frac{f \cdot g_{\max}}{g_{\max} - f}} = \frac{f(g_{\max} - g)}{g_{\max}(g - f)} \quad (13)$$

$$g_{\max} = \frac{f^2 g}{f^2 - d \cdot k \cdot (g - f)} \quad (14)$$

$$\frac{d}{D} = \frac{b_{\min} - b}{b_{\min}} = \frac{\frac{f \cdot g_{\min}}{g_{\min} - f} - \frac{f \cdot g}{g - f}}{\frac{f \cdot g_{\min}}{g_{\min} - f}} = \frac{f(g - g_{\min})}{g_{\min}(g - f)} \quad (15)$$

$$g_{\min} = \frac{f^2 g}{f^2 + d \cdot k \cdot (g - f)} \quad (16)$$

Neben dem definierten Schärfebereich können bei Objektiven für Linsen-Kameras in der Praxis weitere Effekte auftreten, die zu Bildverfälschungen führen können. Diese rühren daher, dass die Linsengleichungen prinzipiell nur für Lichtstrahlen gelten, die dicht und nahezu parallel zur optischen Achse verlaufen. Die auffälligsten Verfälschungen hierbei sind Farbfehler (chromatische Aberrationen) und Verzeichnungsfehler (monochromatische Aberrationen, radiale Verzerrungen). Chromatische Aberrationen treten auf, da Licht unterschiedlicher Wellenlänge ein unterschiedliches Brechungsverhalten zeigt, welches in Objektiven nur unzureichend oder mit erheblichem Aufwand korrigiert werden kann. Eine komplette optische Korrektur ist in der Praxis nicht zwingend notwendig, da der Einfluss chromatischer Aberrationen minimal ist und vom menschlichen Auge kaum wahrgenommen werden kann.

Verzeichnungsfehler hingegen führen zu nichtlinearen Verzerrungen eines Bildes, wodurch sich Objektabstände, Winkel, etc. gegenüber dem Original stark verändern können. Die Hauptursache für Verzeichnungsfehler liegt in der geometrischen Position der Blende [Schröder 1998], also dem Teil eines Objektivs, durch welches der Lichteinfall des optischen Systems geregelt werden kann. Befindet sich die Blende vor dem abbildenden System, so lassen sich tonnenförmige Verzerrungen beobachten. Liegt sie dahinter, so entstehen kissenförmige Verzerrungen.

Der Effekt dieser radialen Verzerrungen lässt sich gemäß Gleichung (17) folgendermaßen beschreiben (u.a. [Slama et al. 1980]):

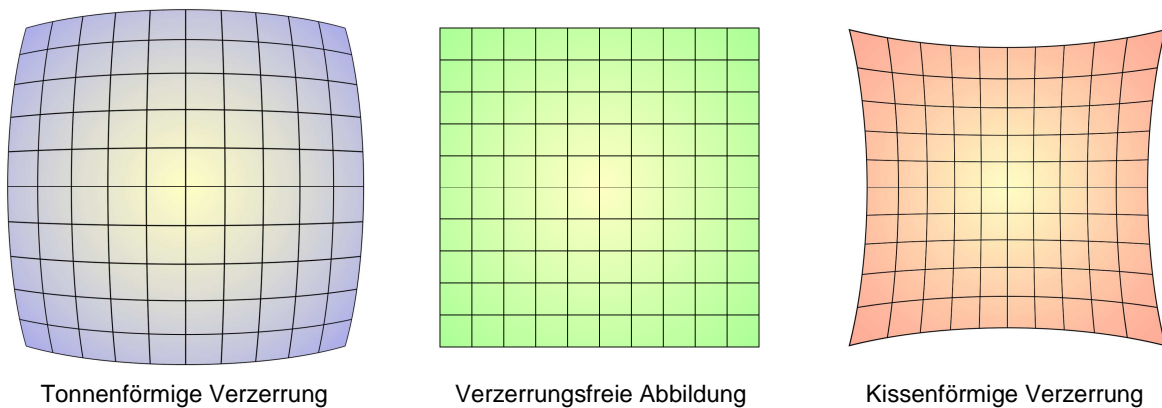


Abbildung 4 Mögliche Abbildungsverzerrungen durch eine Linse. Ein Gitter erscheint so bei einer tonnenförmigen Verzerrung gewölbt, bei einer kissenförmigen Verzerrung gestaucht. Diese Verzerrungen können mittels einer Kalibrierung berücksichtigt und somit zu verzerrungsfreien Bildern rückgerechnet werden.

$$\begin{aligned} x_u &= x_d + x_d \cdot (\kappa_1 r^2 + \kappa_2 r^4 + \dots) \\ y_u &= y_d + y_d \cdot (\kappa_1 r^2 + \kappa_2 r^4 + \dots) \end{aligned} \quad (17)$$

$$r^2 = (x_d^2 + y_d^2) \quad (18)$$

In diesen Gleichungen beschreiben x_u und y_u die verzerrungsfreien (undistorted) Sensorkoordinaten eines Bildpunktes, wie er von einer idealen Linse abgebildet werden würde. Dagegen bezeichnen x_d und y_d die tatsächlich abgebildeten (verzerrten, distorted) Koordinaten des Punktes. Der Koeffizient κ_i definiert hierbei die Art der Verzerrung – ein positiver Verzerrungskoeffizient κ_i steht für eine Tonnenverzerrung, ein negatives κ_i für eine Kissenverzerrung, so wie sie in Abbildung 4 dargestellt sind. Für die meisten Anwendungen genügt eine Beschreibung der Gleichung mit dem ersten Faktor κ_1 für eine hinreichende Entzerrung eines Bildes. Hierfür müssen im Vorfeld mittels einer Kalibrierung der Kamera, wie unter Abschnitt 2.1.3 beschrieben, die einzelnen Faktoren bestimmt werden.

2.1.2 Sensoren zur digitalen Bilderfassung

Möchte man optische Objektabbildungen in irgendeiner Form digital weiterverarbeiten oder speichern, so müssen die auf die Bildebene abgebildeten Informationen erfasst werden. Historisch gesehen beginnt die digitale Bildaufnahme 1923 mit der Erfindung des Ikonoskops, welches den ersten elektronischen Bildabtaster darstellte und die mechanischen Abtastverfahren ablöste. Diese Entwicklung setzte sich über das auf Röhren basierende Vidicon Ende der 50er Jahre fort.

In der heutigen Zeit teilen sich den Massenmarkt Sensortypen, die auf dem inneren photoelektrischen Effekt basieren. In bestimmten Halbleitern bewirkt die Bestrahlung mit Licht, dass Elektronen durch die Energie der Photonen vom Valenzband in das höher gelegene

Leitungsband gehoben werden, wodurch die Leitfähigkeit des Materials zunimmt. Nach diesem Verfahren kann einfallende Lichtintensität gemessen werden. Dennoch unterscheiden sich die zwei am verbreitetsten Sensortypen grundlegend in ihrem inneren Aufbau.

2.1.2.1 CCD-Sensoren

Ein typischer CCD-Sensor (Charge-Coupled Device) lässt sich am anschaulichsten als eine Matrix aus lichtempfindlichen Fotodioden beschreiben, die durch eine elektronische Verschaltung gezielt ausgelesen werden können. Ein einzelner Sensor besteht also aus vielen lichtempfindlichen Elementen, den Pixeln. Jedes dieser Pixel ist in der Lage, gemäß den Grundlagen des photoelektrischen Effekts Photonen „einzufangen“, in elektrische Ladung umzuwandeln und zu speichern. Die Menge der gespeicherten Energie ist dabei direkt proportional zum einfallenden Licht am Ort der Fotodiode. Nach einer bestimmten Belichtungszeit werden diese Ladungen schließlich nach dem Prinzip eines Schieberegisters in einen A/D-Wandler verschoben, ausgewertet und das Ergebnis als digitales Bild übertragen. Jedem Pixel wird dabei ein diskreter Wert zugeordnet, der entsprechend der zugrunde gelegten Farbtiefe in Graustufen interpretiert werden kann. Die Aufnahme von Farbbildern wird unter 2.1.2.3 beschrieben. Die Werte der Graustufen steigen annähernd linear mit der einfallenden Lichtintensität an. Die Sensorbausteine werden üblicherweise aus Silizium gefertigt, welches eine spektrale Empfindlichkeit von 350-1000nm besitzt und somit das Spektrum des für den Menschen sichtbaren Lichts voll abdeckt. Weiterführende Informationen über den Aufbau und die Technik von CCD-Sensoren finden sich unter anderem in [Jähne 2002] oder [Klußmann and Wiegmann 2005].

2.1.2.2 CMOS-Sensoren

Das Funktionsprinzip typischer „active pixel“-Sensoren (APS), allgemein als CMOS-Sensoren (Complementary Metal Oxide Semiconductors) bezeichnet, beruht wie bei den zuvor beschriebenen CCD-Sensoren auf dem Einsatz lichtempfindlicher Halbleiterkomponenten. Ein Sensor besteht in seinem Aufbau also aus vielen einzelnen lichtempfindlichen Pixeln. Der große Vorteil der CMOS-Technik ist dabei die Möglichkeit, zusätzliche Komponenten im gleichen Schaltkreis zu integrieren, in dem sich auch die Bildpunkte befinden. So kann jedes Pixel etwa seinen eigenen Verstärker zur Ladungsübertragung besitzen, was ein Schieberegister obsolet macht und das direkte Auslesen jedes einzelnen Bildpunktes erlaubt. Dieses Vorgehen beispielsweise bietet den Vorteil, dass CCD-typische Bildfehler, wie das Überblenden einzelner Punkte, verhindert werden können. Allerdings benötigt die Integration zusätzlicher Elektronik auch zusätzlichen Platz auf der Sensorfläche, was zu einer Verringerung der effektiven Pixelgröße und damit verbunden einer geringeren Lichtempfindlichkeit führen kann. Eine Möglichkeit, diesen Effekt auszugleichen, bietet hier beispielsweise der Einsatz von Mikrolinsen. Auch bietet die fortlaufende Weiterentwicklung der CMOS-Technik eine beständige Verbesserung der Sensoren und bietet zudem die Möglichkeit der direkten integrierten Datenverarbeitung auf dem Sensor, bei geringerem Gesamtenergieverbrauch und schnelleren Auslesezeiten. Nähere Informa-

tionen und weiterführende Literatur zur Technik von APS und CMOS-Sensoren findet sich unter anderem in [Yadid-Pecht and Etienne-Cummings 2004].

2.1.2.3 Farberzeugung

Nach DIN 5033 ist Farbe diejenige Gesichtsempfindung eines dem Auge des Menschen strukturlos erscheinenden Teiles des Gesichtsfeldes, durch die sich dieser Teil bei einäugiger Betrachtung mit unbewegtem Auge von einem gleichzeitig gesehenen, ebenfalls strukturlosen angrenzenden Bezirk allein unterscheiden kann [DIN 1992].

Farbe ist also keine physikalische Eigenschaft, sondern eine Sinnesempfindung, ein subjektiver Eindruck, der entsteht, wenn Licht bestimmter Wellenlänge auf die Netzhaut des Auges fällt und dort spezielle Sinneszellen anregt. Der für den Menschen sichtbare Wellenlängenbereich liegt zwischen $380nm$ und $780nm$, wobei dieser von Individuum zu Individuum variieren kann. Die lichtempfindlichen Rezeptoren auf der Netzhaut teilen sich dabei in Stäbchen und Zapfen auf. Letztere sind für den Eindruck von Farben verantwortlich, während die Stäbchen nur Hell-Dunkel-Kontraste wahrnehmen können. Es existieren drei spektral unterschiedlich absorbierende Zapfenarten in der Retina, die jeweils für verschiedene Wellenlängenbereiche empfindlich sind [Hauske 1994].

Diese Farbwahrnehmung verschiedener Kanäle, bzw. Wellenlängenbereiche, ist die Ursache für die Einteilung des sichtbaren Lichts in die drei Grundfarben Rot, Grün und Blau gemäß der trichromatischen Farbtheorie. Von der Commission Internationale de l'Eclairage (CIE) wurden den Grundfarben folgende Wellenlängen zugeordnet:

$$\text{Rot} = 700,0nm \quad \text{Grün} = 546,1nm \quad \text{Blau} = 435,8nm$$

Im Gegensatz zum menschlichen Farbempfinden ist es nicht möglich, Farbbilder bzw. verschiedene Wellenlängen direkt mit einem einzelnen lichtempfindlichen Sensor aufzunehmen. Dieser kann lediglich die gesamt einfallende Lichtintensität messen, welche nur bedingt von der einfallenden Wellenlänge abhängt. Es gibt jedoch verschiedene Verfahren, die eine Farberzeugung ermöglichen. Die bekannteste und am weitesten verbreitete Methode ist die Verwendung des Bayer-Filters (Abbildung 5 (1)).

Die Entwicklung des nach Bryce E. Bayer benannten Filters in den 70er Jahren [Bayer 1976] basiert auf dem damaligen Wissen über die menschliche Farbwahrnehmung und orientiert sich am YUV-Farbraum. Dabei stützt sie sich auf die Tatsache, dass sich in einer kleinen räumlichen Umgebung keine großen oder plötzlichen Farbänderungen vollziehen. Folglich wird auf den Bildsensor eine Maske aus Bandpassfiltern gelegt, so dass jeder Bildpunkt nur für eine der drei Grundfarben empfindlich ist. Ein Algorithmus interpoliert daraus für jedes einzelne Pixel nach der Bildaufnahme aus den Nachbarpunkten die beiden fehlenden Grundfarben, so dass durch nachträgliche Mischung der Eindruck eines Farbbildes entsteht. Dieses Verfahren liefert augenscheinlich gute Ergebnisse bei der Abbildung der Umgebung, reduziert aber grundsätzlich die effektiv zur Bilderfassung genutzte Anzahl der Pixel, da 50% der Pixel grüne und jeweils 25% rote und blaue Bildinformatio-

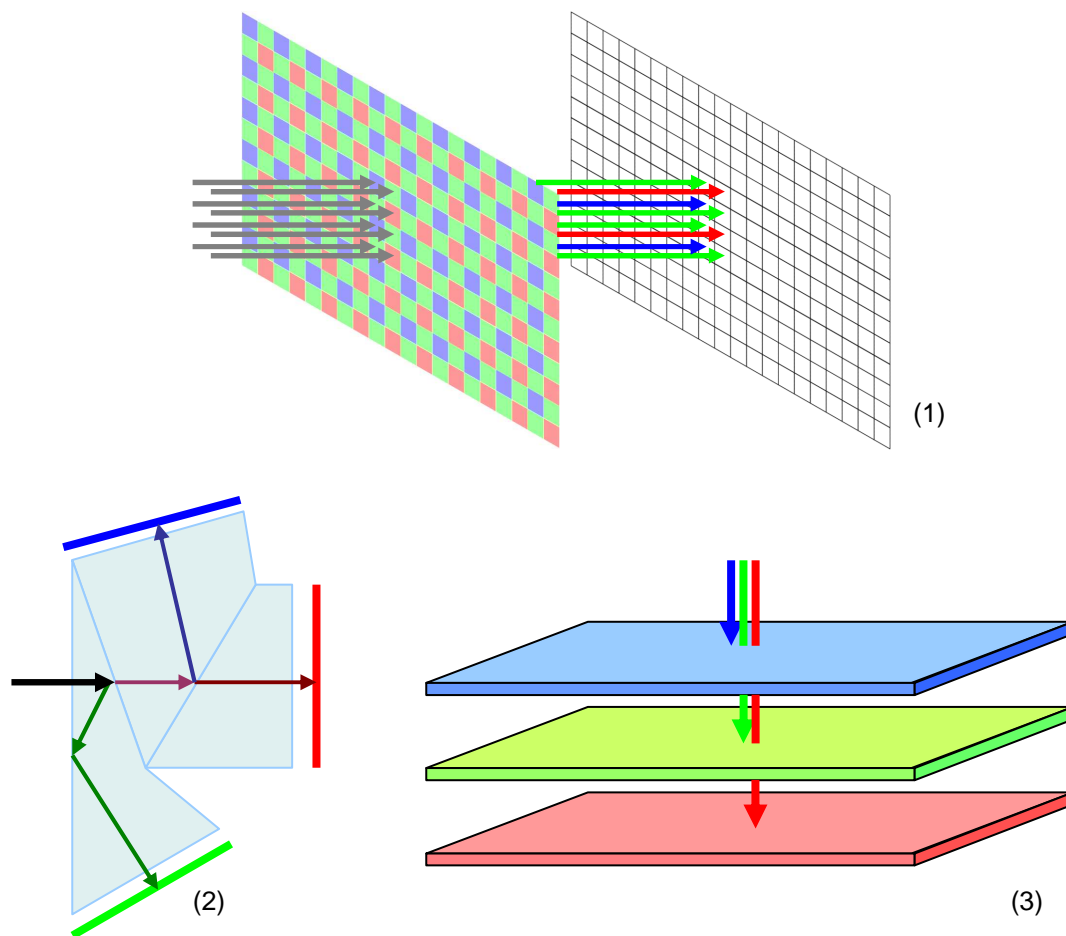


Abbildung 5 Methoden der Farberzeugung auf lichtempfindlichen Sensoren. Ein-Chip-Variante mit Bayer-Filter (1), 3-Chip-Variante mit Prismen zur Farbverteilung (2) und Sensor der Firma Foveon (3) mit hintereinander liegenden, farbempfindlichen Schichten.

nen sammeln. Kein Pixel erhält also durch die Aufnahme seine reale Farbinformation. Dennoch werden durch die geringen Farbunterschiede innerhalb einer räumlich derart kleinen Umgebung im resultierenden Bild kaum Verfälschungen vom menschlichen Auge wahrgenommen.

Anders verhält sich dies bei aktuellen, neueren Methoden der digitalen Bilderfassung. Dies ist zum einen die getrennte Anordnung und nachträgliche Datenkombination von drei Bildsensoren, auf die jeweils nur eine der Grundfarben abgebildet wird (Abbildung 5 (2)). Somit kann jeder der Sensoren seine volle Pixelanzahl nutzen, das endgültige Bild enthält die volle Auflösung jedes der 3 Teilbilder (z.B. [Panasonic 2007]).

Zum anderen ist es möglich, Sensoren zu bauen, die Photonen in verschiedenen Schichttiefen auslesen können – und sich dabei die Eigenschaft zunutze machen, dass Licht verschiedener Wellenlängen verschieden tief in bestimmte Materialien eindringt. Von der Firma Foveon ([Merrill 1999], [Turner et al. 2005]) wurde nach diesem Prinzip ein Sensor

entwickelt, der ebenfalls für jeden Bildpunkt die volle Farbinformation aller drei Kanäle erfassen kann (Abbildung 5 (3)). Eine ideale Lösung für das Problem der Farberfassung bieten jedoch beide Verfahren nicht. Durch die Aufnahmeverfahren können zwar volle Sensorinformationen für jede Grundfarbe ausgelesen werden. Allerdings muss sowohl durch die optischen Pfade des 3-Chip-Designs, als auch durch die verschiedenen Eindringtiefen mit Helligkeitsverlusten der jeweiligen Kanäle gerechnet werden, so dass die Gesamtempfindlichkeit des Sensors abnehmen kann.

2.1.3 Kamera-Kalibrierung

Da alle Methoden der Bildverarbeitung auf der exakten Auswertung und Analyse von Bilddaten basieren, ist es zwingend notwendig, unverzerrte Abbildungen der Umgebung zu erhalten. Dies betrifft vor allem radiale Verzerrungen, die sich am deutlichsten in einer Verfälschung von Seitenverhältnissen, Abständen, Krümmungsverhalten, etc. bemerkbar machen.

Eine Kalibrierung dient dabei dazu, die intrinsischen Parameter einer Kamera, sowie die Verzerrungsparameter ihrer Optik zu bestimmen. Die Kamerakalibrierung nach Tsai gilt als Standard und wird hierfür häufig verwendet [Tsai 1987]. In (vor allem für Projektionseinheiten) erweiterter Form findet sie auch in dieser Arbeit Anwendung, da sie eine gute Balance zwischen Exaktheit und Rechenaufwand herstellt.

Die Kalibrierung basiert auf der perspektivischen Projektion einer Lochkamera und berücksichtigt eine radiale Linsenverzerrung erster Ordnung. Neben den durch die technischen Eigenschaften des Kamerasensors bestimmten Werten

- Anzahl der Sensorelemente in x - bzw. y -Richtung: N_x, N_y
- Maße eines Sensorelements in x - bzw. y -Richtung: d_x, d_y

besteht Tsai's Kameramodell aus 11 zu bestimmenden Parametern, die sich in 5 intrinsische (interne) und 6 extrinsische (externe) Kameraparameter aufteilen. Diese sind:

Interne Parameter:

- Die effektive Brennweite der Kamera: f_k
- Der Verzerrungskoeffizient erster Ordnung: κ_1
- Das optische Bildzentrum: $C = (C_x, C_y)$

C ist definiert als der Schnittpunkt der optischen Achse mit der Bildebene, sowie zugleich als das Zentrum der radialen Verzerrung. Annäherung mittels:

$$C' = \left[c'_x = N_x/2, c'_y = N_y/2 \right]$$

- Der Skalierungsfaktor: s_x
Dieser diente ursprünglich zum Ausgleich von Synchronisationsfehlern im analogen Videosignal – in heutigen Digitalkameras kann dieser Faktor nahezu vernachlässigt werden.

Externe Parameter:

- Die drei Rotationsparameter: R_x, R_y, R_z
Dies sind die drei Eulerwinkel und entsprechen respektive den 9 Einträgen der Rotationsmatrix \underline{R} aus Gleichung (8), beschreiben also die Orientierung der Kamera relativ zum Weltkoordinatensystem.
- Die drei Translationsparameter: $\underline{t} = t_x, t_y, t_z$
Dies sind die drei Verschiebungsparameter aus Gleichung (8), die die Lage der Kamera relativ zum Weltkoordinatensystem beschreiben.

Die internen Parameter beschreiben also die Abbildungseigenschaften der Kamera, während die externen Parameter die Lage und Orientierung der Kamera im Raum definieren.

Zur Durchführung der Kalibrierung wird ein Kalibrierkörper mit Markierungen benötigt, deren exakte Anordnung und Position relativ zueinander bekannt sein müssen. Diese Marken können entweder räumlich angeordnet sein oder in einer Ebene (coplanar), was die praktikablere Lösung ist (siehe auch Abbildung 6). Die Koordinaten der Kalibriermarken definieren dabei das Weltkoordinatensystem und sollten so gewählt werden, dass der Ursprung nicht in der Nähe der Sichtachse der Kamera liegt. Außerdem kann durch die benutzerdefinierte Wahl der Koordinaten und die Lage der Marken in einer Ebene für alle Punkte $z_W = 0$ gesetzt werden, so dass alle Marken in der $x_W y_W$ -Ebene liegen.

Die eigentliche Kalibrierung basiert schließlich auf einer vierstufigen Beschreibung der digitalen Bilderfassung.

Stufe 1: Übergang von Welt- zu Kamerakoordinaten

Ergebnis: $\underline{R}, \underline{t}$

Der Zusammenhang zwischen den Welt- und Kamerakoordinaten wurde bereits in Gleichung (8) beschrieben, so dass sich folgende Beziehung ableiten lässt:

$$\begin{bmatrix} x_C \\ y_C \\ z_C \end{bmatrix} = \underline{R} \cdot \begin{bmatrix} x_W \\ y_W \\ z_W \end{bmatrix} + \underline{t} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} x_W \\ y_W \\ z_W \end{bmatrix} + \begin{bmatrix} t_X \\ t_Y \\ t_Z \end{bmatrix} \quad (19)$$

Stufe 2: Transformation der Kamerakoordinaten in ideale, unverzerrte (undistorted) Sensorkoordinaten

Ergebnis: f_k

Diesem Schritt liegt die bereits beschriebene ideale perspektivische Projektion zugrunde, die folgenden Zusammenhang festlegt:

$$x_u = \frac{f_k x_C}{z_C} \quad \text{und} \quad y_u = \frac{f_k y_C}{z_C} \quad (20)$$

Stufe 3: Radiale Linsenverzerrung

Ergebnis: κ_1

Gemäß Gleichung (17) lässt sich die radiale Linsenverzerrung bestimmen, wobei aufgrund seiner Gewichtung der erste Verzerrungsparameter den größten Einfluss ausübt. Es gilt also:

$$x_d + D_x = x_u \quad \text{und} \quad y_d + D_y = y_u \quad (21)$$

$$\text{mit} \quad D_x = x_d (\kappa_1 r^2 + \kappa_2 r^4 + \dots) \quad \text{und} \quad D_y = y_d (\kappa_1 r^2 + \kappa_2 r^4 + \dots) \quad (22)$$

$$r = \sqrt{x_d^2 + y_d^2} \quad (23)$$

Stufe 4: Transformation der Sensor- in Pixelkoordinaten

Der grundlegende Zusammenhang zwischen Sensor- und Pixelkoordinaten lässt sich laut Tsai beschreiben durch:

$$u = \frac{s_x x_d}{d_x} + C_x \quad \text{und} \quad v = \frac{y_d}{d_y} + C_y \quad (24)$$

Da in unserem Fall aufgrund des Einsatzes eines digitalen Signals $s_x = 1$ angenommen werden darf, vereinfacht sich diese Beschreibung für die Parameter u und v zu:

$$u = \frac{x_d}{d_x} + C_x \quad \text{und} \quad v = \frac{y_d}{d_y} + C_y \quad (25)$$

Zur tatsächlichen Kalibrierung werden die Aufnahmen einer Kalibriertafel betrachtet, wie sie in Abbildung 6 dargestellt sind. Wichtig sind dabei möglichst verschiedene Ansichten, so dass unterschiedliche Orientierungen vorhanden sind. Die Marken des Kalibrierkörpers und ihre Abstände zueinander müssen möglichst exakt bekannt sein, da auf diesen Daten die Kalibrierung der Kamera aufbaut.

Die eigentliche Kalibrierung schließlich folgt den eben beschriebenen Stufen in rückwärtiger Reihenfolge. Sie ist vor allem deshalb möglich, da alle zu lösenden Gleichungssysteme durch die hohe Anzahl der zur Verfügung stehenden Kalibriermarken überbestimmt sind. Der Vorgang basiert auf einer Detektion der Kalibriermarken und beginnt mit deren Positionsbestimmung in Pixelkoordinaten u und v . Die durch die verschiedenen Stufen der Kalibrierung ermittelten Daten dienen als Grundlage zur Formulierung einer relativen Fehlerfunktion des Kalibrierergebnisses in Abhängigkeit der ermittelten Parameter, die durch Standard-Optimierungsverfahren minimiert wird. Dabei sind mehrfache Iterationen

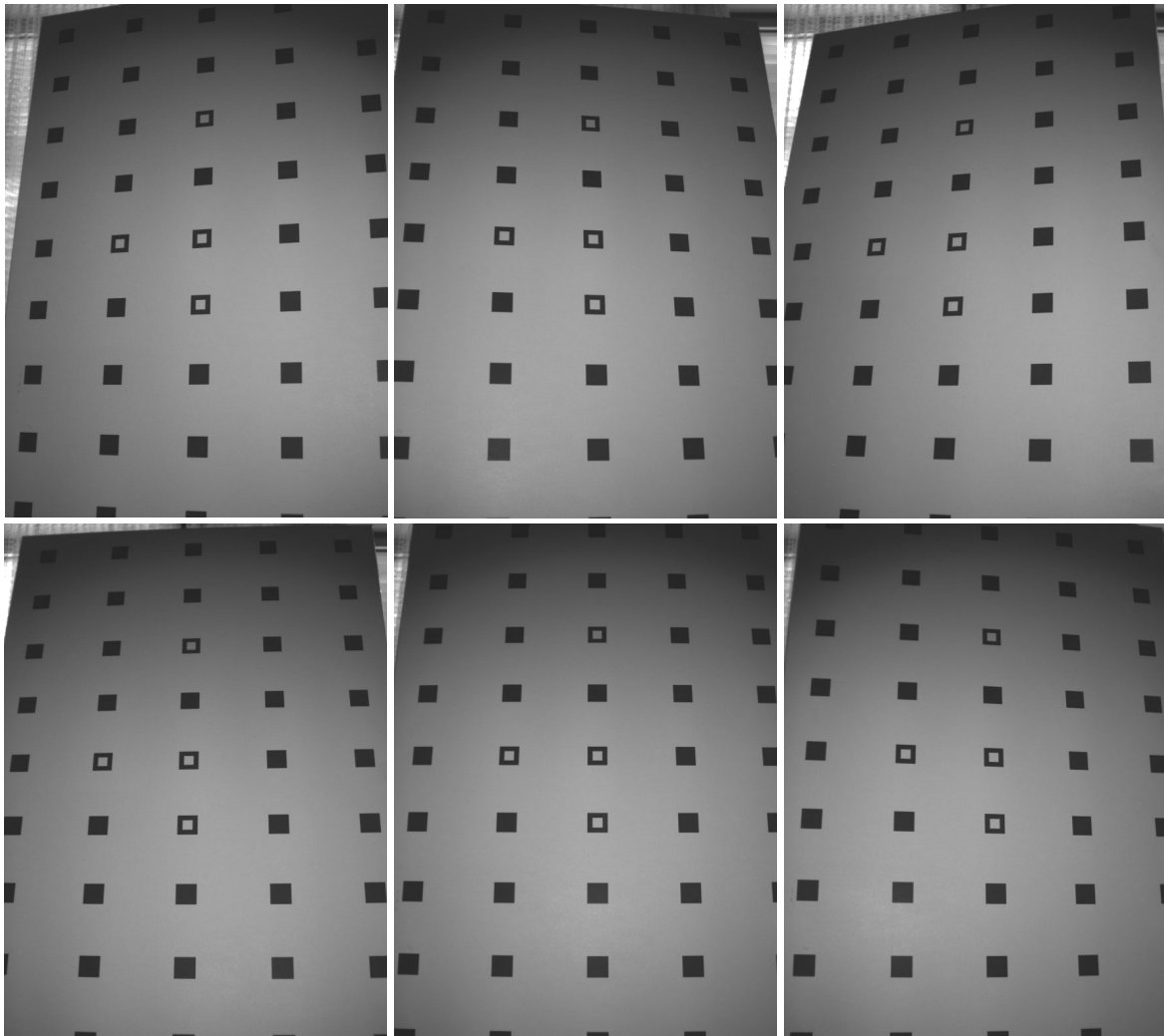


Abbildung 6 Verschiedene Ansichten einer Kalibriertafel. Anhand der Position der Quadrate kann die Lage der Tafel und somit die Lage der aufnehmenden Kamera im Raum bestimmt werden. Die exakten Koordinaten und Abstände auf der Tafel müssen hierfür im Vorfeld bekannt sein. Die weißen Quadrate dienen der Zuordnung der Kalibriermarken zu ihren bekannten Positionen.

des Kalibriervorgangs mit Kombinationen von Aufnahmen verschiedener Perspektiven nötig, um die Ergebnisse der Kalibrierung zu verfeinern und verbessern. Der kleinste Kalibrierfehler liefert sodann die beste Annäherung der Modellparameter (siehe hierzu auch [Faugeras 1993] oder [Hartley and Zisserman 2000]).

Es existieren auch weitere Alternativen der Kamerakalibrierung, die im Verlauf der letzten Jahre entwickelt und verfeinert wurden. Zu nennen ist hier unter anderem die Kalibrierung nach Zhang [Zhang 1998], die durch die freie Verfügbarkeit in der Intel Computer Vision Library (ICV) hohe Popularität erlangt hat. Sie erlaubt, wie die Methode nach Tsai, eine Kalibrierung unter Verwendung einer planaren Kalibrierplatte mit bekannten Markern und kombiniert hierfür mindestens zwei verschiedene Aufnahmen, um die benötigten Kameraparameter zu ermitteln.

2.2 Grundlagen der 3D-Messtechnik

Spricht man von 3D-Messtechnik, so öffnet man ein breites Feld verschiedener Messverfahren für unterschiedlichste Anwendungen. Deshalb ist es zunächst wichtig, genau zu definieren, für welchen Bereich einer 3D-Objekterfassung welche Art von Sensorik geeignet ist. Die in dieser Arbeit besprochenen Verfahren beziehen sich allesamt auf berührungslose Messverfahren mit einem Arbeitsraum im Nahbereich des Sensors und sind somit für Orientierungssensoren, exakte Oberflächenvermessungen kleinerer Volumina, vor allem aber für einen Einsatz als Mensch-Maschine-Schnittstelle geeignet. In der Natur bestehen solche Verfahren seit Jahrtausenden – einige Tiere haben spezielle Rezeptoren zur Erfassung ihrer Umgebung entwickelt, wie beispielsweise Fledermäuse, die Ultraschall nutzen. Andere können sich, wie auch wir Menschen, optisch in unserer dreidimensionalen Welt orientieren, was als räumliches oder auch Stereo-Sehen bezeichnet wird. Der Versuch, diese Verfahren auf technische Systeme zu übertragen, stellt einen Ansatz zur Entwicklung von Systemen zur 3D-Vermessung dar, welcher im weiteren Verlauf dieses Kapitels betrachtet wird.

Einen etwas anderen Ansatz verfolgt die direkte Anwendung der Geometrie, aus der die Triangulation entstanden ist. Dies ist ein Verfahren, das bereits im Mittelalter zur Abschätzung und Vermessung von Entfernungen verwendet wurde. Als ihr Entdecker gilt der niederländische Mathematiker Gemma Frisius [Haasbroek 1968], der als erster eine mathematische Beschreibung der Triangulation als Messtechnik verfasste.

Abbildung 7 zeigt die grundsätzliche geometrische Anordnung dreier Punkte im Raum, die den Aufbau einer Triangulation beschreiben und deren Lagebeziehungen zueinander die Durchführung eben dieser ermöglichen. Um Entfernungen durch eine Triangulation zu bestimmen, muss folglich ein Punkt S von zwei weiteren Punkten P und Q betrachtet werden können, was geometrisch der Verbindung durch eine Sichtgerade entspricht. Die Verbindung zwischen P und Q bildet die Basis b , deren Abmessungen bekannt sein müssen. Durch die beiden Betrachtungswinkel α und β , unter denen S relativ zur Basis erscheint, lassen sich, etwa über den Sinussatz, alle übrigen Größen des Dreiecks bestimmen:

$$\frac{\sin \alpha}{p} = \frac{\sin \beta}{q} = \frac{\sin(180^\circ - \alpha - \beta)}{b} \quad \text{und} \quad z = q \sin \alpha = p \sin \beta \quad (26)$$

Nach diesem Prinzip lassen sich grundsätzlich Entfernungen berechnen, solange man ein Objekt von zwei bekannten Positionen aus betrachten kann. So werden Triangulationsverfahren auch in der Seefahrt, Kartographie und Astronomie eingesetzt.

Genauer betrachtet bildet die Triangulation also auch die Grundlage für das Modell des Stereo-Sehens, nach dem die menschliche Tiefenwahrnehmung funktioniert. Unsere Augen betrachten ein Objekt von unterschiedlichen Positionen aus und können so dessen Entfernung bestimmen. Für technische Verfahren muss man hierbei zwischen passiven und aktiven Stereo-Verfahren unterscheiden, wie sie in 2.4 und 2.5 beschrieben werden.

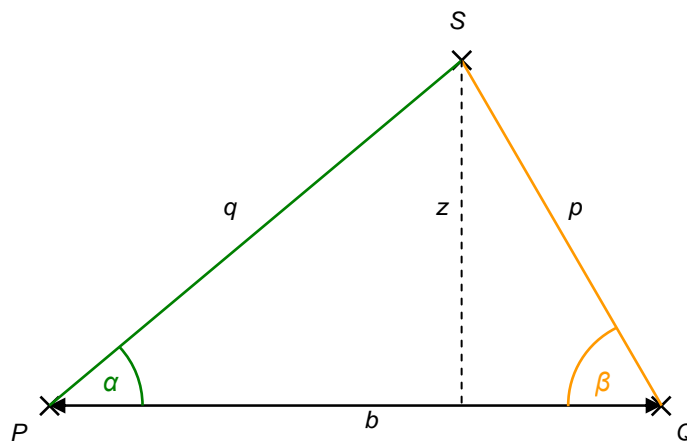


Abbildung 7 Grundprinzip der Triangulation. Der Abstand eines Punktes S von der Verbindung/Basis zweier Punkte P und Q lässt sich durch die Berechnung der Höhe eines Dreiecks ausdrücken.

Um verschiedene Verfahren der 3D-Objekterfassung überhaupt untereinander vergleichen zu können, müssen einige Begriffe geklärt werden, die die Beschreibung der Eigenschaften und deren Einordnung anhand methodischer Grundlagen erlauben ([Jähne 1991], [Jähne 2002], [Forster 2005]):

- Verschiedene Begriffe im Zusammenhang der 3D-Objekterfassung beschreiben oder meinen oft das Gleiche – das sind unter anderem *3D-Messung*, *Entfernungsmessung*, *Abstandsmessung*, *Objekterfassung*.
- *Messergebnis*: Hier wird grundsätzlich darin unterschieden, ob das System nur reine 3D-Oberflächendaten, 3D mit texturierten Oberflächen oder komplette Objektmodelle mit zugehörigem Volumen bestimmen kann.
- *Tiefenkarte*: Hiermit wird die Interpretation einer dreidimensionalen Szene als zweidimensionales Bild bezeichnet, die eine einfache Darstellung auf einem Bildschirm ermöglicht. Jedem Bildpunkt (u, v) wird dabei der entsprechende Tiefenwert z_C der Kamerakoordinaten (x_C, y_C, z_C) zugewiesen und in Farbwerten ausgedrückt. Objektdaten und Entfernungsdaten sind somit äquivalent.
- Die (laterale) *Auflösung* einer Tiefenkarte entspricht der Anzahl der Pixel, aus denen tatsächliche Messdaten gewonnen werden können. Idealerweise würde dies der Auflösung der verwendeten Kamera(s) entsprechen.
- Die (*geometrischen*) *Eigenschaften* eines 3D-Sensors bezeichnen dessen Maße, die erlaubte Entfernung eines Messobjekts, den Arbeitsraum, etc.
- Die *Genauigkeit* eines Messsystems ist die wohl wichtigste Vergleichsgröße – und man findet zugleich eine große Menge verschiedener Definitionen in der Literatur. In dieser Arbeit bezeichnet die Genauigkeit den Zusammenhang zwischen dem mittleren Fehler einer Tiefenmessung Δz und der zugehörigen Standardabweichung σ_z in einem definierten Sensorabstand z .

- Die *Robustheit* eines Messsystems erlaubt Aussagen über die möglichen Einsatzbereiche eines Sensors und bezieht sich auf mögliche Störgrößen, wie beispielsweise das Sonnenlicht.
- Die *Gefährlichkeit* für die Gesundheit schließlich – z.B. für das Augenlicht – ist ein weiterer Punkt, der über den Einsatz eines Messsystems entscheidet.

Die folgenden Kapitel beschreiben nun einige aktuelle Verfahren der 3D-Objekterfassung, angefangen mit allgemeinen Messmethoden bis hin zu spezifischen Stereoskopieverfahren.

2.3 Allgemeine Methoden der 3D-Objekterfassung

Es existieren verschiedene Ansätze zur dreidimensionalen Erfassung von Objekten. Neben den stereoskopischen Verfahren, die ab Kapitel 2.4 beschrieben werden, werden im folgenden allgemeine Verfahren zur 3D-Objekterfassung beschrieben. Diese stellen einen wichtigen Schritt zur 3D-Vermessung dar, da sie teilweise aus Beobachtungen der Natur entstanden sind und somit an das natürliche menschliche oder tierische Tiefensehen angelehnt sind. In der Literatur finden sich bereits einige Übersichten verschiedener Ansätze ([Salvi et al. 2004], [Bowyer et al. 2006], [Stoykova et al. 2007]), die sich jedoch jeweils nur mit einem thematischen Teilgebiet befassen. Dieses Kapitel hingegen soll einen allgemeinen Überblick über die Grundlagen der jeweiligen Verfahren und Ansätze liefern und dabei auch Schnittstellen zu einigen nicht-stereoskopischen Verfahren aufzeigen.

2.3.1 Ultraschall-Verfahren und Laufzeitmessung

Messsysteme, die mit Ultraschall arbeiten, basieren auf der Laufzeitmessung von Schallwellen ([Ballard and Brown 1982]). Einige Tiere, wie Fledermäuse oder Delfine, nutzen diese Technik, um sich zu orientieren oder ihre Beute zu finden. Die physikalischen Grundlagen liegen in der Ausbreitungsgeschwindigkeit c_S von Schallwellen und der Messung deren Reflexionen. Eine technische Anwendung stellen beispielsweise Rückfahrwarnsysteme dar, wie sie in vielen PKWs verbaut sind. Abbildung 8 zeigt den prinzipiellen Messaufbau einer solchen Anwendung.

Ein Sender sendet zu einer bestimmten Zeit t_0 einen Schallimpuls aus. Trifft dieser auf ein Objekt, so werden Schallwellen reflektiert, die ein Empfänger zur Zeit t_E detektieren kann. Durch den Laufzeitunterschied $\Delta t = (t_E - t_0)$ lässt sich daraus die Entfernung z des Objektes berechnen:

$$z = \frac{c_S(t_E - t_0)}{2} = \Delta t \frac{c_S}{2} \quad (27)$$

Ein Nachteil dieser Messmethode liegt zum einen in der vergleichsweise langen Messdauer bei größeren Entfernungen, zum anderen in der Genauigkeit und Auflösung der Messdaten, die im Allgemeinen nur eine Hinderniserkennung erlauben. Ein Vorteil dage-

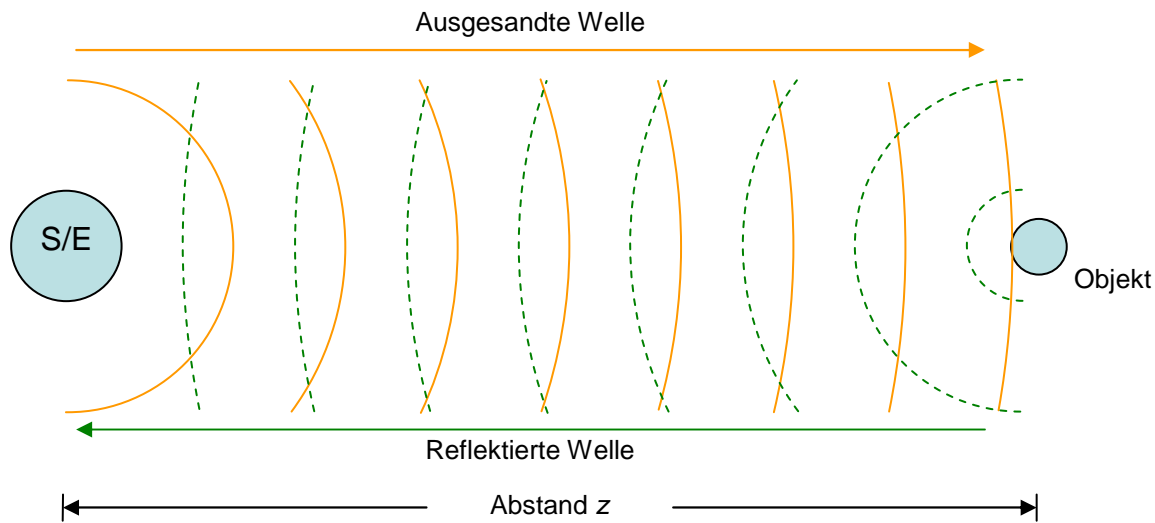


Abbildung 8 Funktionsprinzip einer Entfernungsmessung mittels Ultraschall. Ein Sender S strahlt Ultraschallwellen aus, die von einem Objekt reflektiert werden. Ein Empfänger E , zumeist am gleichen Ort wie der Sender, detektiert diese Wellen, wodurch der Abstand z über die Schallgeschwindigkeit bestimmt werden kann.

gen ist, dass die ausgesandten Schallwellen nicht vom menschlichen Gehör wahrgenommen werden können. Des Weiteren sind Messfehler unabhängig von der Objektentfernung, da nur die Zeitmessung einen direkten Einfluss auf das Ergebnis hat, wie folgende Gleichung zeigt:

$$\sigma(z) = \frac{c_s}{2} \sigma(\Delta t) \quad (28)$$

Für eine Entfernungs- oder Abstandsmessung sind Laufzeitmessungen also durchaus geeignet. Allerdings erlauben sie keine exakte Objekterkennung, da sie lediglich ein grobes Bild von Objekten mit äußerst geringer lateraler Auflösung wiedergeben können.

2.3.2 Time-Of-Flight

Das Prinzip der sogenannten Time-Of-Flight-Verfahren (u.a. [Jähne 2002]) basiert, wie die Ultraschall-Verfahren, auf der Messung von Laufzeitunterschieden. Allerdings handelt es sich um ein optisches Verfahren zur Detektion von Lichtwellen. Es liegt also nicht mehr die Schallgeschwindigkeit c_s als physikalische Konstante vor, sondern die Lichtgeschwindigkeit c . Der Messaufbau dagegen ist grundsätzlich identisch – ein Sender sendet zu einer bekannten Zeit t_0 einen Licht- oder Laserimpuls aus, dessen Reflexion an einem Objekt detektiert werden kann. Die Berechnung der Objektentfernung erfolgt daraus analog zu Gleichung (27), mit dem gleichen Vorteil der Unabhängigkeit des Messfehlers von der Objektentfernung (28). Schwieriger gestaltet sich hierbei die Laufzeitmessung, da die Lichtgeschwindigkeit deutlich kürzere Messdauern verlangt als die Schallgeschwindigkeit. Es kommen also in der Regel hochlichtempfindliche Empfänger zum Einsatz.

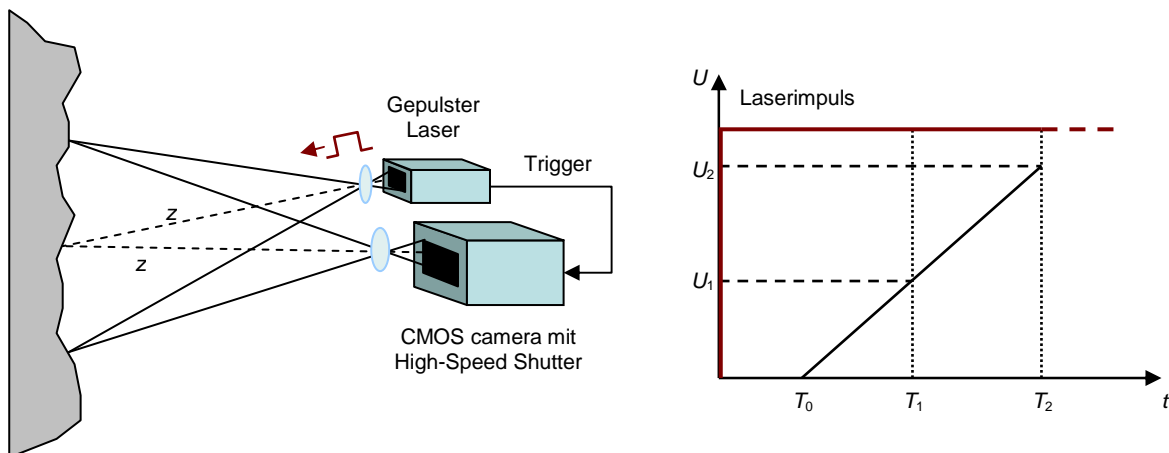


Abbildung 9 Time-Of-Flight-Prinzip am Beispiel der MDSI. Ein Objekt wird von einem gepulsten Laserlicht beleuchtet und von einer synchron gesteuerten Kamera aufgenommen, was zu einem Spannungsanstieg U in einem Sensorelement führt. Durch Integration zu verschiedenen Zeitpunkten T_1 und T_2 und über die Lichtgeschwindigkeit c lässt sich somit die Entfernung z eines Objektpunkts bestimmen.

Allerdings ist auch diese Methode aufgrund mangelnder Auflösung nicht für eine exakte Objekterkennung geeignet. Eine Erweiterung für diesen Anwendungsbereich stellt die Variante der „Multiple-Double Short Time Integration“ (MDSI) dar ([Mengel et al. 2001], [Elkhalili et al. 2004]), wie sie in Abbildung 9 dargestellt ist.

Die technischen Voraussetzungen für dieses Messverfahren sind eine Kamera mit einer sehr kurzen Belichtungszeit und linearem Sensor-Verhalten, sowie Laser- bzw. Fotodioden mit einer kurzen Anstiegszeit. Die Kamera beginnt, synchronisiert mit einem Laserimpuls, ihre Aufnahme. Trifft das Licht auf einen Punkt der Szene, der von der Kamera erfasst wird, so bewirkt dies einen Anstieg der Spannung U im entsprechenden Bildpunkt des Sensors, der wegen der Laufzeit entfernungsabhängig erst zum Zeitpunkt T_0 beginnt. Wird dieser Wert zu verschiedenen Zeitpunkten T_1 und T_2 ausgelesen, so ist eine Berechnung der Entfernung z über die Laufzeit des Lichtes gemäß folgender Formel möglich:

$$z = \frac{c(U_2 T_1 - U_1 T_2)}{2(U_2 - U_1)} \quad (29)$$

Ein klarer Vorteil im Vergleich zu anderen Laufzeitverfahren ist die Tatsache, dass die Messung für alle Pixel simultan durchgeführt wird, wodurch man mit einer Aufnahme mehrere Bildpunkte und somit auch mehrerer Punkte eines Objekts vermessen kann. Man erhält also eine tatsächliche Punktwolke, aus denen sich ein Modell der aufgenommenen Szene rekonstruieren lässt. Das Verfahren ist durch die doppelte Integration weitgehend unabhängig von Umgebungslicht und aufgrund des Messprinzips ist der Fehler unabhängig von der Objektentfernung. Allerdings ist die Auflösung eines Sensors kaum für eine echte Erkennung von Objekten geeignet, da aktuell verfügbare Sensoren lediglich eine grobe Rasterung von 8×64 Bildpunkten erlauben. Zudem liegt die Genauigkeit im Zentimeterbereich, was für tatsächliche Messaufgaben nicht ausreichend ist. Dennoch ver-

spricht die Weiterentwicklung der CMOS-Technik, vor allem durch die Möglichkeit der Integration zusätzlicher Sensoreigenschaften, eine Erhöhung der Auflösungen und Empfindlichkeit, so dass die genannten Nachteile bei zukünftigen Entwicklungen unter Umständen nicht mehr zum Tragen kommen.

2.3.3 Interferometrie

Die Grundlagen der interferometrischen Messtechnik wurden Ende des 19. Jahrhunderts mit der Erfindung des Interferometers durch Albert Michelson entwickelt [Bauer 2003]. Während die ursprüngliche Anordnung benutzt wurde, um den sogenannten „Lichtäther“ als Medium für die Ausbreitung des Lichts zu untersuchen, erlauben moderne Methoden der Bildverarbeitung den Einsatz für genaue Topographie-Vermessungen. Allen Verfahren ist jedoch gemein, dass sie Interferenzphänomene bei der Verwendung von kohärentem Licht beobachten. Man kann die Interferometrie auch als spezielle Form der Time-of-Flight-Verfahren mit periodisch modulierter Signalamplitude beschreiben [Jähne 2002], da die Entfernungsmessung ein Vielfaches der verwendeten Wellenlänge durch gezielte Phasenmessung ergibt. Eine Lichtquelle wird dabei durch einen Strahlteiler in zwei Teilwellen aufgespalten und anschließend nach Durchlaufen unterschiedlicher Wege wieder zur Interferenz gebracht.

Moderne Weißlicht-Interferometer bestehen aus einer Lichtquelle mit einer Kohärenzlänge im Bereich von μm , einem Strahlteiler, einem Referenzspiegel und einer Kamera, typischerweise eine CCD-Kamera mit angepasstem Objektivsystem. Sie können als Mikroskopsystem oder mit telezentrischen Objektiven aufgebaut werden, was einen Einfluss auf den Messbereich und die Genauigkeit hat. Je nach Aufbau sind mit einem Weißlicht-Interferometer laterale Auflösungen zwischen $1\mu m$ und $50\mu m$ möglich.

Um diesen Messbereich flexibel zu halten und nicht durch die verwendete Wellenlänge beschränken zu lassen, ist es möglich, durch Überlagerung (Modulo-Subtraktion) einzelner Wellenlängen λ_1 und λ_2 eine synthetische Wellenlänge Λ zu erzeugen [Meixner et al. 2004]:

$$\Lambda = \frac{\lambda_1 \lambda_2}{2 \cdot |\lambda_1 - \lambda_2| \cdot \cos \rho} \quad (30)$$

Der Winkel ρ bezeichnet dabei den Winkel zwischen der Messflächen-Normale und der Beleuchtungs-/Beobachtungsrichtung. Allerdings sind solche Verfahren eben wegen der Anwendung von zwei Wellenlängen relativ empfindlich gegenüber Objektbewegungen, da dabei durch die Dekorrelationen zwischen den Wellenlängen erhebliche Kontrastverluste auftreten können. Um dies zu vermeiden, werden mehrere Kameras eingesetzt, die simultan verschiedene Wellenlängen aufzeichnen können.

Interferometrische Verfahren benötigen zwar nur eine sehr kurze Aufnahmezeit, die Auswertung der Interferogramme ist jedoch relativ komplex. So liegen die Erfassungsraten bei

Berechnung mit aktuellen Prozessoren typischerweise bei 1-2fps (frames per second oder auch Datensätze pro Sekunde). Zudem können die Messaufbauten nur eine relativ kleine Fläche von wenigen Zentimetern abdecken, weshalb Interferometer in erster Linie für industrielle Anwendungen ruhender, kleiner Objekte und Oberflächen interessant sind.

Unter dem Begriff der Interferometrie lassen sich aber noch weitere Methoden der Tiefen- oder Entfernungsmessung zusammenfassen. Radio-Interferometer arbeiten nach dem von Michelson entwickelten Prinzip, allerdings mit deutlich höheren Wellenlängen als die des sichtbaren Lichts. Ebenfalls auf der Beobachtung und Auswertung von Interferenzmustern basieren Methoden der Moiré-Interferometrie. Hierbei werden künstlich projizierte Gitterstrukturen durch parallel ausgerichtete Gitter aufgenommen, so dass definierte Bedingungen für Interferenzen der Muster bestehen ([Pallek 2007], [Klette et. al. 1996]). Obwohl dieses Vorgehen eine hohe Auflösung und Genauigkeit ermöglicht, ist es anfällig für Oberflächentexturen und nur auf ausreichend glatten Oberflächen anwendbar, so dass interferometrische Verfahren dieser Art in erster Linie für technische Anwendungen in Frage kommen.

2.4 Passive Stereoskopie

Zu der Gruppe der passiven Stereo-Verfahren gehören all diejenigen Messmethoden, die auf einem bildbasierten Ansatz unter Verwendung von Kameras aufbauen und sich in ihrer Grundkonzeption in gewisser Weise am menschlichen Sehen und Erkennen orientieren. Der Begriff „passiv“ bezieht sich also auf die reine Auswertung von Sensor-Informationen, die ein Abbild der Umgebung ohne zusätzlich eingebrachte Informationen darstellen. Den Gegensatz dazu stellen aktive Komponenten dar, die in Kapitel 2.5 behandelt werden.

2.4.1 Depth from Focus

Die Tiefenberechnung auf Basis der Bildschärfe (Focus) ist genau betrachtet kein Stereo-Verfahren, da sie auf den Ergebnissen einer einzelnen Kamera basiert. Da die eigentliche Berechnung auf einer Triangulation beruht, wird sie dennoch der Stereoskopie zugeordnet.

Ein Objekt wird von einer Kamera, oder vielmehr einem Bildsensor, genau dann scharf aufgenommen, wenn es sich im Bereich der Tiefenschärfe von dessen Optik befindet (vgl. Abbildung 10). Genau diesen Effekt kann man sich zunutze machen, indem die geometrischen Grundlagen der Triangulation nicht auf zwei, sondern mehrere optische Pfade angewandt werden. Die Triangulationsbasis bildet dabei der Durchmesser der Optik, die verschiedenen Abbildungen des Unschärfekreises (vgl. 2.1.1.2) mit seinem jeweiligen Radius erlauben Rückschlüsse auf den Objektabstand. Es ist somit möglich, durch eine Folge von Aufnahmen mit definierten Tiefenschritten die Oberflächenstruktur eines Objekts zu bestimmen.

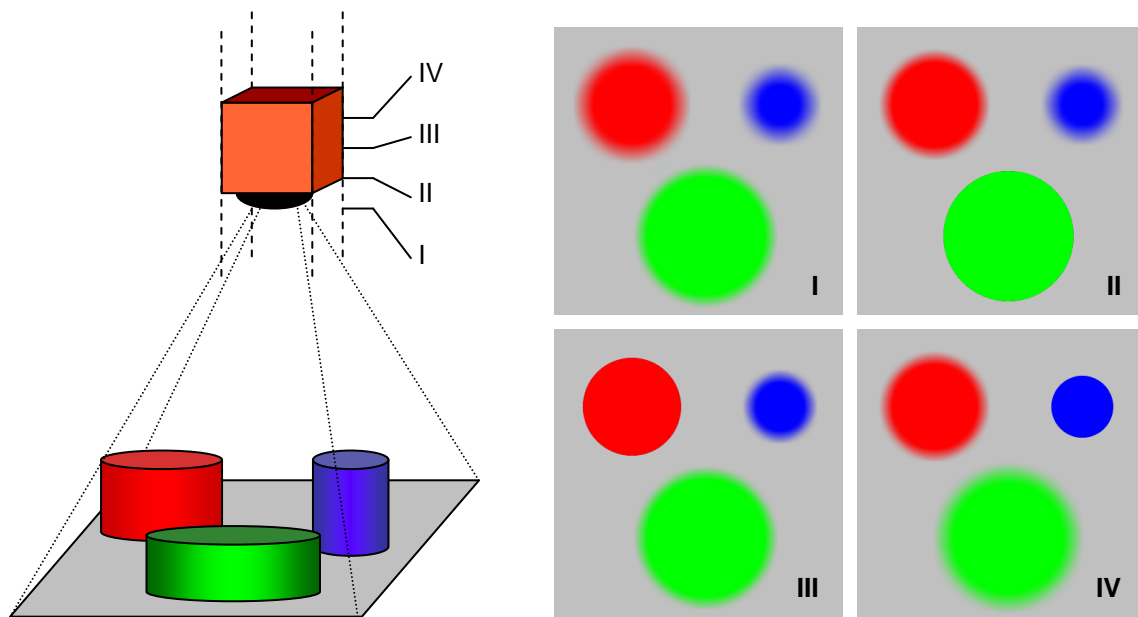


Abbildung 10 Depth-from-Focus. Eine Szene wird von verschiedenen Positionen aus aufgenommen (I-IV), wobei nur Objekte im Schärfebereich fokussiert abgebildet werden. Daraus lassen sich die Tiefeninformationen ermitteln.

[Steurer et al. 1986] schlagen dafür folgenden Ablauf vor:

Es wird eine Bildserie mit verschiedenen, konstanten Tiefenschritten der Optik aufgenommen. Ein Hochpassfilter dient dazu, in diesen Bildern Bereiche mit signifikanten Intensitätsänderungen zu maskieren. Vergleicht man die entsprechende Region in allen aufgenommenen Bildern, so ergibt der maximale Differenzwert eines Bildes den entsprechenden Tiefenwert für diese Region. Da sich dadurch keine flächendeckenden Werte bestimmen lassen, ist eine Interpolation unabdingbar, um die gesamte Oberfläche darstellen zu können.

Diese Methode kann eine Tiefenkarte in Kameraauflösung liefern, deren Genauigkeit im Bereich der verwendeten optischen Schritte (etwa $2\mu m$) liegt. Allerdings erfordert die Aufnahme Zeit, da mehrere Bilder aufgenommen, sowie die Tiefenschritte sehr genau eingestellt werden müssen. Außerdem ist eine Bestimmung scharf abgebildeter Bildbereiche nur für strukturierte Oberflächen möglich. Glatte Oberflächen beinhalten zu wenig Information, als dass eine Tiefenberechnung durch Depth from Focus möglich wäre.

2.4.2 Shape from Shading

Die Tiefenberechnung oder Entfernungsabschätzung aus nur einer einzelnen Ansicht ist eine Fähigkeit des Menschen, die auf Erfahrungen im Betrachten von Objekten beruht. Es ist uns durchaus möglich, beim Betrachten eines Fotos Annahmen über die Form, Entfernung und Beschaffenheit verschiedener Objekte zu treffen. Für eine Maschine ist dies nicht ohne weiteres möglich – die Umsetzung eines solchen Ansatzes wird Shape from

Shading (SfS) ([Klette et. al. 1996], [Zhang et al. 1999]) genannt, also die Berechnung einer Form aus deren Schattierung.

Für den Aufbau eines SfS-Systems ist nur eine einzelne Kamera nötig. Da allerdings eine zusätzliche konstante Objektbeleuchtung für die Zuverlässigkeit des Systems erforderlich ist und die Tiefenerfassung sich am menschlichen Sehen orientiert, ist diese Methode hier den Stereo-Verfahren zugeordnet.

Eine Tiefenberechnung auf Basis der Objektschattierung ist möglich, da Geometrie und Reflexionsverhalten einer Oberfläche das Licht beeinflussen, welches schließlich auf den Bildsensor trifft. Entscheidend hierfür ist der Farb- oder Grauwerteindruck im aufgenommenen Bild, die sogenannte Irradianz, die sich aus der eingestrahnten Lichtenergie auf ein bestimmtes Flächenelement berechnet. Sie kann von vier grundlegenden Größen beeinflusst werden – der Beleuchtung, der Oberflächenreflexion, der Objektgeometrie und von den Aufnahmeeigenschaften des Sensors selbst. Anschaulich lässt sich die Aufgabenstellung des SfS mit der dreidimensionalen Darstellung von Objekten auf zweidimensionalen Bildschirmen vergleichen, wobei in diesem Fall der umgekehrte Weg beschränkt wird. Jedoch gelten für die Messmethoden einige Einschränkungen. So muss zunächst die Position der Quelle der Beleuchtung bekannt sein. Bei zusätzlicher künstlicher Beleuchtung ist dies noch relativ einfach zu modellieren, es existieren verschiedene Umsetzungen für parallele Beleuchtungen oder Punktlichtquellen. Für diffuses oder natürlich einfallendes Licht sind exakte Modellierungen dagegen schwer durchführbar.

Da das aufgenommene Bild eine Interpretation des von der Oberfläche reflektierten Lichts darstellt, hat das Reflexionsverhalten des Objekts einen ebenso großen Einfluss wie die Beleuchtung. Für die Oberfläche wird im Allgemeinen die Lambertsche Reflexion angenommen, was eine homogene Struktur der gesamten Szene bedeutet. Dieses Zusammenspiel erlaubt die Modellierung einer sogenannten Reflektanzkarte, die der Oberflächenorientierung eindeutige Irradianzwerte zuordnet. Eine grobe Abschätzung von Flächenorientierungen ist damit bereits möglich.

Weitere Einschränkungen und Voraussetzungen an die Oberflächenbeschaffenheit machen Sensoren nach diesem Verfahren für eine allgemeine Anwendung auf unbekannte Objekte unbrauchbar, weshalb auf diese Probleme hier nicht näher eingegangen wird. Zudem bleibt die Szenenrekonstruktion mittels SfS sehr aufwendig. So wurde bereits von [Forsyth and Zisserman 1991] zusammengefasst, dass es durch SfS unmöglich ist, exakte, dichte und genaue Szenenrekonstruktionen zu erhalten. Da auch neuere Methoden stets mit starken Einschränkungen auskommen müssen oder A-Priori-Wissen verwenden, bleiben SfS-Methoden ein interessanter Aspekt der Tiefenberechnung, da sie sich mit fundamentalen Fragen der Auswirkung und Kompensation verschiedener Arten der Objektbeleuchtung beschäftigen. Dennoch werden sie für technische Anwendungen, vor allem aber als biometrische Schnittstelle, nicht als praktikabel angesehen.

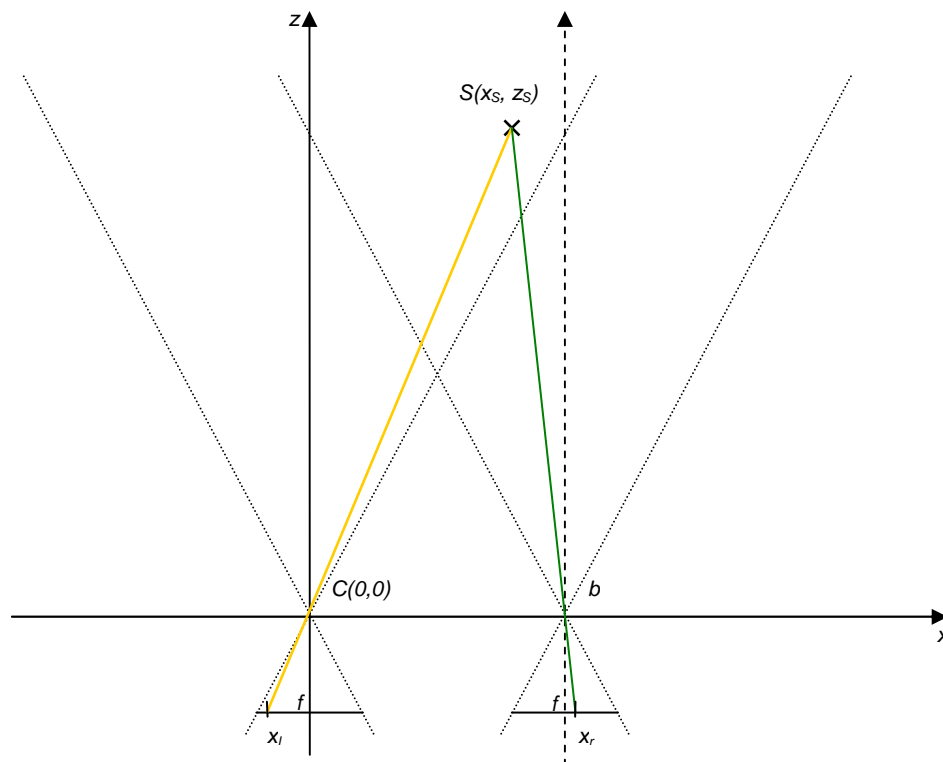


Abbildung 11 Passive Stereoskopie mittels zweier parallel angeordneter Bildsensoren. Über den Schnitt der Sichtgeraden beider Sensoren und die entsprechenden Bildkoordinaten x_l und x_r lässt sich die Entfernung z_s eines Objektpunkts S berechnen.

2.4.3 Passives Stereo

Die ursprüngliche passive Stereoskopie ist grundsätzlich eine direkte Übertragung des menschlichen Sehens auf ein Computer-Interface zur Bild- und Datenerfassung.

Der Mensch erkennt seine Umwelt aus dem Benutzen seiner beiden Augen. Sie ermöglichen es uns, Entfernungen abzuschätzen, Objektgrößen einzuordnen und uns in unserer dreidimensionalen Welt zu bewegen. Fällt eines unserer Augen aus, so sind wir in unserer Wahrnehmung stark eingeschränkt, da wir dann nur noch eine zweidimensionale Abbildung unserer Umgebung erhalten und uns zur Tiefenbestimmung auf unsere Erfahrung verlassen müssen (siehe Depth-from-Focus oder Shape-from-Shading).

Das dreidimensionale Erfassen von Objekten mittels passiver Stereoskopie lässt sich durch zwei Bildaufnahmegeräte, zum Beispiel einfache Kameras, in einer technischen Anwendung modellieren. Um so ein Stereo-System aufzubauen, werden zwei Kameras in einem definierten Abstand b , der Basis, nebeneinander angeordnet. Die Sichtachsen beider Kameras verlaufen dabei parallel, das Zentrum des Koordinatensystems liegt im optischen Zentrum einer der beiden Kameras. Abbildung 11 zeigt eine solche Anordnung.

Wird von beiden Kameras ein Punkt S erfasst, so lässt sich mittels einer Triangulation die Entfernung des Punktes von den Kameras berechnen. Dazu müssen die Koordinaten des Punktes in beiden Kamera-Systemen bekannt sein. Diese erlauben dann allgemein folgende Beschreibung:

$$\frac{x_l}{f} = \frac{x_S}{z_S} \quad \wedge \quad \frac{x_r}{f} = \frac{x_S - b}{z_S} \quad (31)$$

Mit den Umformungen

$$\frac{x_r}{f} = \frac{z_S x_l - b}{z_S} \quad (32)$$

$$x_r = \frac{z_S x_l - bf}{z_S} \quad (33)$$

$$z_S (x_l - x_r) = bf \quad (34)$$

lässt sich schließlich der Abstand z_S des Punktes S berechnen:

$$z_S = \frac{bf}{x_l - x_r} \quad (35)$$

Allerdings bedingt eine parallele Kameraanordnung, dass keiner der beiden verwendeten Bildsensoren auf seiner vollen Fläche ausgenutzt wird, da dies die Positionierung beider Sensoren auf den gleichen Koordinaten bedeuten würde. Um also die volle Auflösung beider Sensoren zu nutzen, ist es nötig, eine der beiden Kameras zu drehen. Der Winkel zwischen den Sichtachsen, der nun entsteht, wird als Triangulationswinkel θ bezeichnet. Abbildung 12 zeigt eine solche Anordnung zweier Kameras, von denen eine (rechts) „frei“ zur Referenzkamera (links) positioniert wurde.

Dieser Winkel zwischen den Kameras sorgt dafür, dass die sichtbaren Bereiche beider Sensoren besser in Deckung gebracht werden können, als dies bei einer parallelen Anordnung möglich wäre. Somit lässt sich eine größere Fläche abtasten, es können mehr Punkte vermessen werden. Zudem erlaubt eine nicht-parallele Anordnung, den Abstand der Kameras zu vergrößern. Dies ermöglicht die Bestimmung genauerer Tiefendaten, da sich der relative Fehler eines Tiefenpunkts durch eine Vergrößerung der Basis verringert (siehe auch 3.6.2). Die Tiefenberechnung ändert sich dabei entsprechend, da die verschiedene Ausrichtung der beiden Koordinatensysteme der Kameras eine Transformation der Koordinatensysteme auf Basis der Kameraorientierung und –position notwendig macht.

Hierzu wird eine imaginäre Kamera definiert, deren Zentrum dem der gedrehten Kamera $C' = (b, z_R)$ entspricht und die die Orientierung einer regulären, parallelen Anordnung be-

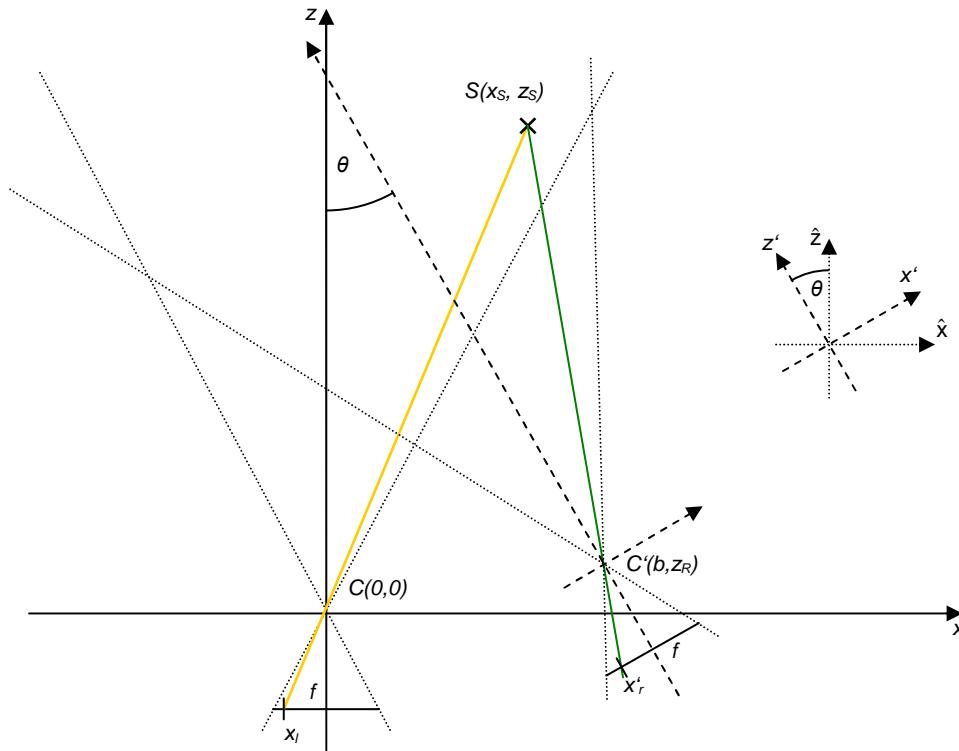


Abbildung 12 Passive Stereoskopie mittels zweier Bildsensoren unter einem Triangulationswinkel θ . Ähnlich der parallelen Anordnung lässt sich die Entfernung z_S eines Objektpunkts S über den Schnitt der Sichtgeraden beider Sensoren und die entsprechenden Bildpunkte x_l und x'_r berechnen.

sitzt. Dies geschieht entsprechend der folgenden Beziehung der Koordinatentransformation (siehe zur Verdeutlichung auch Abbildung 12):

$$\begin{aligned}\hat{x} &= x' \cos \theta - z' \sin \theta \\ \hat{z} &= x' \sin \theta + z' \cos \theta\end{aligned}\quad (36)$$

Für die Kamerakoordinate der rechten Seite, die zur Tiefenberechnung benötigt wird, gilt folglich:

$$\hat{x}_r = -f \frac{x'_r \cos \theta + f \sin \theta}{x'_r \sin \theta - f \cos \theta}\quad (37)$$

Die Gleichungen (31)-(35) ändern sich dementsprechend, so dass sich die Beziehungen innerhalb der beiden Koordinatensysteme und damit auch die Tiefenberechnung umformulieren lassen zu:

$$\frac{x_l}{f} = \frac{x_S}{z_S} \quad \wedge \quad \frac{\hat{x}_r}{f} = \frac{\hat{x}_S}{\hat{z}_S} = \frac{x_S - b}{z_S - z_R}\quad (38)$$

$$\hat{x}_r = \frac{\left(\frac{z_S x_l}{f} - b\right) f}{z_S - z_R} = \frac{z_S x_l - bf}{z_S - z_R} \quad (39)$$

$$z_S = \frac{bf - \hat{x}_r z_R}{x_l - \hat{x}_r} \quad (40)$$

Unabhängig von der Kameraanordnung ist es allerdings eine größere Schwierigkeit bei der Tiefenberechnung, die entsprechenden Punkte in beiden Sensoren korrekt zuzuordnen. Dieses sogenannte Korrespondenz-Problem fällt bei strukturierten Oberflächen relativ gering aus, da sich durch verschiedene Verfahren, wie beispielsweise ein Template-Matching, zuverlässig korrespondierende Bereiche in zwei Bildern finden lassen. Mehrdeutigkeiten oder Fehlzuordnungen ergeben sich dagegen besonders bei einfarbigen, homogenen Oberflächen oder Oberflächen mit periodischer Struktur. Stellt man sich hierbei zwei Aufnahmen einer weißen Wand mit verschiedenen Kameras vor, so ist eine Zuordnung entsprechender Punkte nahezu unmöglich. In solchen Fällen versagen passive Stereo-Verfahren oder benötigen das Hinzufügen zusätzlicher Oberflächenstrukturierungen. Üblicherweise behilft man sich in solchen Fällen durch das Aufbringen von Markern oder Strukturspray, wodurch der Vorteil eines berührungslosen Messverfahrens allerdings verloren geht. Um diesen Vorteil zu erhalten, wird deshalb oft eine andere Möglichkeit zusätzlicher Strukturierung gewählt, die mit der aktiven Stereoskopie im folgenden Kapitel vorgestellt wird.

2.5 Aktive Stereoskopie

Im Gegensatz zur passiven bezeichnet die aktive Stereoskopie diejenigen Verfahren, die zwar auf dem Stereo-Prinzip basieren, dieses aber durch eine aktive Komponente unterstützen. Diese aktive Komponente, in der Regel realisiert durch eine Beleuchtung mit einem bestimmten Muster, bringt dabei zusätzliche Information in eine zu vermessende Szene, die eine Tiefenberechnung erleichtert.

2.5.1 Grundprinzip

Im einfachsten Fall erleichtert diese zusätzliche Beleuchtung das Lösen des Korrespondenzproblems, da die Zuordnung entsprechender Bildbereiche zwischen zwei Kameras erleichtert wird (siehe Abbildung 13). Es kann also ohne mechanische Beeinträchtigung berührungslos eine Struktur auf eine neutrale Oberfläche aufgebracht werden, die eine klassische Stereoskopie auf beliebige Oberflächen vereinfacht (vgl. auch [Vuylsteke et al. 1990], [Knoll and Sasse 1993]). Hat man dabei noch genauere Informationen über die Anordnung des projizierten Musters, dessen geometrischer Struktur und die Art der Projektion, so ist es sogar möglich, eine Kamera komplett durch eine aktive Beleuchtung zu ersetzen. Die Tiefenberechnung nach dem Stereo-Prinzip bleibt dabei erhalten, da ein Projektor wie eine Kamera als zweidimensionale Abbildung einer Ebene modelliert werden kann. In der Projektion identifizierte Bildpunkte können demnach auf Sensorkoordinaten des Projektionsmusters zurückgeführt werden und erlauben eine Modellierung, die der einer Kamera entspricht.

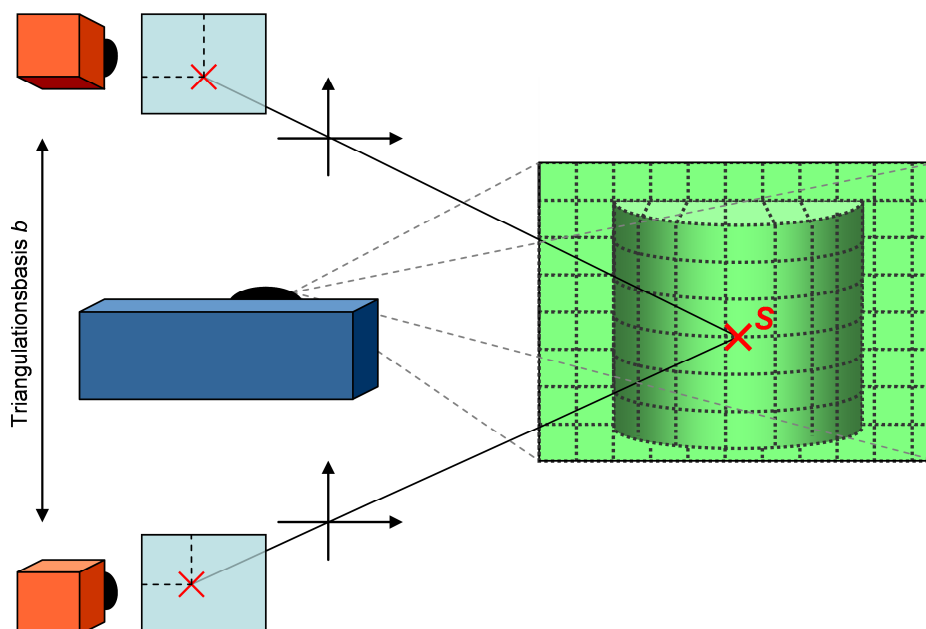


Abbildung 13 Aktive, strukturierte Beleuchtung. Durch Projektion eines strukturierten Musters wird das Lösen des Korrespondenzproblems eines Stereosystems erleichtert, vor allem bei unstrukturierten oder glatten Oberflächen.

2.5.2 Laservermessung und Lichtschnitt

Die einfachste Anordnung einer solchen aktiven Beleuchtung stellen dabei die Methoden der Laservermessung bzw. des Lichtschnitts dar. Aufgrund des einfachen Aufbaus und der Robustheit der Messmethode findet dieses Verfahren in vielen technischen Bereichen Anwendung. Abbildung 14 zeigt den Aufbau eines solchen Systems, das meist aus einem Linienlaser mit aufgefächertem Lichtstrahl und einer Kamera besteht.

Mit Kenntnis der geometrischen Anordnung von Kamera und Laser lassen sich die Grundlagen der Triangulation vom Beginn dieses Kapitels (siehe Abbildung 7) anwenden und somit die Entfernung von Oberflächenpunkten berechnen. Aber auch ohne genaue Kenntnis der relativen Anordnung von Kamera und Laser können Aussagen über Objektdeformationen oder Bewegungen getroffen werden. Dabei kommen die geometrischen Abbildungseigenschaften des Verfahrens zum Tragen. Trifft der Laserstrahl auf eine ebene Fläche, so erscheint er als Linie, eine gekrümmte Oberfläche bewirkt eine Verformung des Strahls. In der Kamera ist dann die zweidimensionale Repräsentation der Laserlinie mit ihrer vom beleuchteten Objekt hervorgerufenen Deformation deutlich sichtbar.

Um eine exakte Tiefenberechnung durchzuführen, muss bei diesem Verfahren zunächst die Position der Laserlinie genau in der Kamera bestimmt werden. Da die einzelne Linie jedoch als scharfe Hell-Dunkel-Grenze erkennbar ist, ist dies durch eine einfache Kantendetektion und -verfolgung durch eine Bildverarbeitung leicht zu lösen und erlaubt eine Positionsbestimmung bis auf Subpixelgenauigkeit. Allerdings erlaubt eine einzelne Linie auch nur die Betrachtung eines Ausschnitts eines Objekts oder einer Szene. Um ein komplettes kontinuierliches Modell zu erhalten, müssen mehrere Messungen kombiniert werden. Hierfür gibt es verschiedene Ansätze:

Das Messobjekt kann vor dem Sensor horizontal oder vertikal bewegt werden, der Sensor kann bewegt oder geschwenkt werden oder lediglich der Laser wird in fest definierten Schritten rotiert. All diese Umsetzungen erfordern jedoch eine hohe Genauigkeit bei evtl. eingesetzten Schrittmotoren, da fehlerhafte Positionierungen einen direkten Einfluss auf die Messgenauigkeit haben.

Ebenfalls denkbar ist eine Kombination mehrerer Kameras und Laser, was zugleich die simultane Aufnahme mehrerer Linien ermöglichen würde. Allerdings ist hierbei eine eindeutige Linienzuordnung in der jeweiligen Kamera unmöglich. Abhilfe schafft in einem solchen Fall eine abwechselnde Aufnahme oder die Verwendung verschiedener Wellenlängen bei den Lasern und Bandpassfiltern bei den Kameras. Somit entstehen unabhängige Laser-Kamera-Paare, die eine simultane Aufnahme erlauben.

Welche Kombination auch verwendet wird – ein Laser als Projektionseinheit erlaubt eine sehr genaue Oberflächen- und Tiefenvermessung, da der Strahl entlang der projizierten Laserlinie auf einen schmalen Bereich fokussiert ist. Nebeneffekte wie die Specklebildung, hervorgerufen durch das Auftreffen eines kohärenten Lichtbündels auf optisch raue

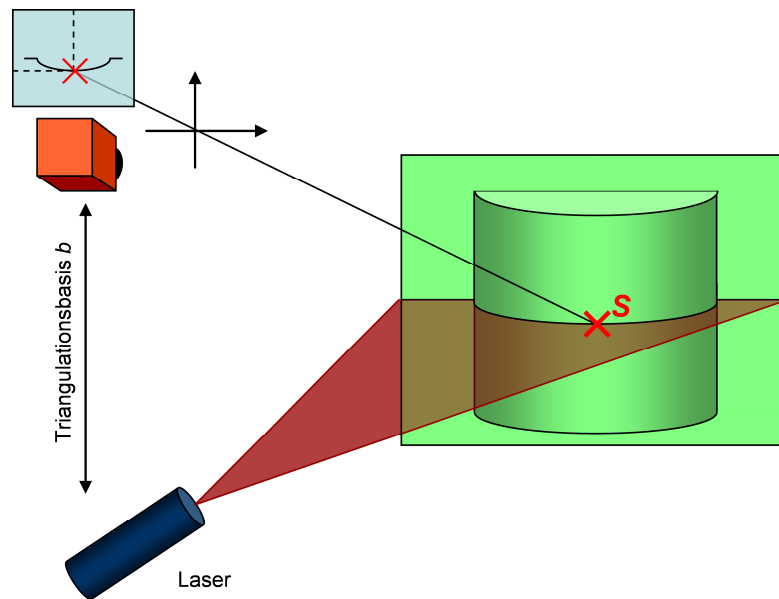


Abbildung 14 Lichtschnitt mittels Laser. Die projizierte Laserlinie kann im Kamerabild eindeutig identifiziert werden und ermöglicht somit den Schnitt der abgestrahlten Lichtebene mit der Sichtgeraden der Kamera zur Tiefenberechnung.

Oberflächen, können dabei durch die anschließende Bildverarbeitung einschließlich der Kantendetektion minimiert werden.

Eine Alternative zur Verwendung von Lasern zur Linienprojektion stellt die Verwendung eines Streifenprojektors dar. Damit können durch den Einsatz eines einfachen Video-Projektors oder eines Dias mehrere Lichtstreifen auf eine Szene projiziert werden (Abbildung 15). Grundsätzlich lässt sich dann für jede projizierte Linie eine Triangulation ausführen. Aufgrund der Eigenschaften der Projektion, die nicht fokussiertes Licht verwendet, nimmt jeder Lichtstreifen eine bestimmte Fläche ein. Dadurch wird keine exakte Linie beschrieben, weswegen es für die Messgenauigkeit von Vorteil ist, zur Messung lediglich die Kanten zwischen den Streifen zu betrachten.

Ein Kantenübergang zwischen einem hellen und einem dunklen Streifen entspricht nämlich einer genau definierten Linie und keiner projizierten Fläche. Gelingt es also, die Koordinaten einer solchen Linie zu bestimmen, so lässt sich diese wiederum als Grundlage einer Triangulation verwenden. Das Auffinden von Kanten ist in der digitalen Bildverarbeitung ein bekanntes Problem. Es existieren etliche Verfahren, deren Weg üblicherweise über die erste Ableitung der Helligkeitswerte der Bildpunkte unter Verwendung des Sobel-Operators mit folgenden Filtermatrizen führt:

$$\underline{S}_{ver} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{und} \quad \underline{S}_{hor} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (41)$$

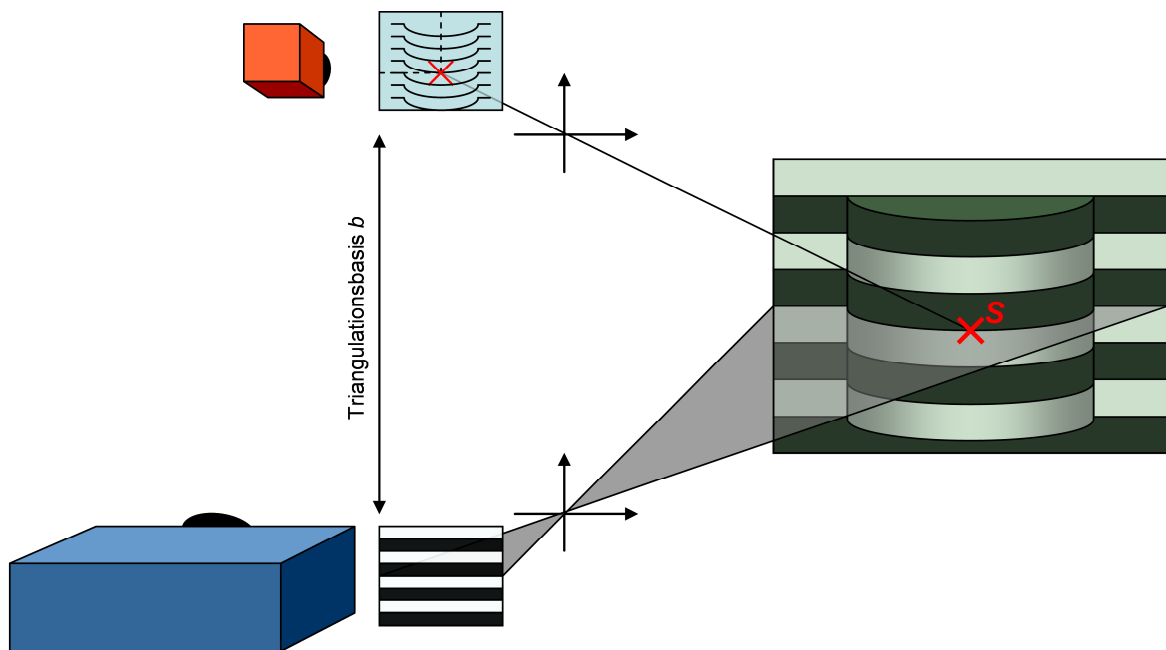


Abbildung 15 Projektion einfacher Lichtstreifen. Zur Tiefenberechnung können entweder die Streifen selbst, oder aber die Kanten dazwischen herangezogen werden – das Vorgehen ist dabei vergleichbar mit einem mehrfachen Lichtschnitt.

Durch Faltung eines Bildes mit einer dieser Matrizen wird ein Gradientenbild erzeugt, dessen Intensitätswerte in den Bereichen starker Helligkeitsänderung – also an Kantenpositionen – am höchsten sind. Wendet man diese Operatoren auf ein mit Streifen beleuchtetes Bild an, so lassen sich die Positionen der Kanten genau bestimmen. Das Ergebnis einer solchen Operation ist schematisch im Bild der Kamera in Abbildung 15 dargestellt. Der Verlauf der deformierten Kanten stellt die auf das Objekt auftreffenden Lichtebenen dar, deren Schnitt mit den Sichtgeraden der Kamera wiederum die Tiefenberechnung ermöglicht.

Die erreichbare Genauigkeit der Tiefenwerte bei diesen Verfahren ist dabei durchaus mit einer Laserlinien-Projektion vergleichbar. Allerdings stellt sich bei der Streifenbeleuchtung das gleiche Problem wie bei der Verwendung mehrerer Linienlaser – die Identifikation und Unterscheidung der einzelnen Kanten. Es müssen deshalb entweder stets alle Kanten im Bild sichtbar sein, um ein Abzählen zu ermöglichen, oder Methoden angewandt werden, die eine eindeutige Identifikation zulassen. Diese erlauben dann für jede Position eine exakte und genaue Tiefenberechnung. Um diese Einschränkungen zu vermeiden, lässt sich das projizierte Lichtmuster codieren, was im Folgenden erläutert wird.

2.5.3 Codiertes Licht

Messsysteme, die codiertes Licht verwenden, bezeichnen im Allgemeinen die Kombination ein oder mehrerer Kameras und einer Projektionseinheit, die zusätzliche Information in codierter Form in eine zu vermessende Szene einbringt und somit das Lösen des Korrespondenzproblems erleichtert.

2.5.3.1 Prinzip

Um Information in einem Lichtmuster eindeutig zu codieren, wurden in den vergangenen Jahrzehnten etliche Verfahren vorgestellt. Zurückführen lassen sich dabei die meisten Systeme auf den Grundsatz der aktiven Stereoskopie – eine Kamera wird durch eine Projektionseinheit ersetzt, deren geometrischer Aufbau (bzw. deren Sensorfläche) durch ein oder mehrere Muster in der Art codiert wird, dass eine eindeutige Zuordnung in der projizierten Darstellung möglich ist. Das Ziel ist also, Bildpunkte, Bildzeilen oder Bildspalten durch

Tabelle 1 Ausgewählte Verfahren zur Codierung von Lichtmustern (nach [Pagès et al. 2003], erweitert). Die Tabelle zeigt die wichtigsten bzw. bekanntesten Vertreter der 3D-Erfassung mittels codierten Lichts, sowie deren entsprechende Referenzen und Erscheinen in der Literatur. Ebenfalls aufgelistet finden sich dabei die Art der Codierung (periodisch oder absolut), die Eignung für bewegte oder statische Objekte sowie die verwendete Farbtiefe der Verfahren (binär – 2bit, Graustufen – 8bit, mehrfarbig – 24bit).

	Verfahren	Beschrieben unter anderem in	Code	Szene	Farbtiefe
Zeitliche Codierung	Binäre Codierung, Gray-Code	[Potsdamer and Altschuler 1982]	abs.	stat.	2bit
		[Inokuchi et al. 1984]	abs.	stat.	2bit
		[Minou et al. 1981]	abs.	stat.	2bit
		[Trobina 1995]	abs.	stat.	2bit
		[Valkenburg and McIvor 1998]	abs.	stat.	2bit
		[Rocchini et al. 2001]	abs.	stat.	24bit
n-näre Codierung	[Caspi et al. 1998] [Horn and Kiryati 1997]	abs.	stat.	24bit	
		abs.	stat.	8bit	
Phasenshift, hybride Methoden	[Bergmann 1995] [Gühring 2001] [Sato 1996] [Hall-Holt and Rusinkiewicz 2001] [Guan et al. 2003]	abs.	stat.	8bit	
		abs.	stat.	2bit	
		abs.	stat.	2-8bit	
		abs.	bew.	2bit	
		abs.	bew.	8bit	
Direkte / räumliche Codierung	De Bruijn Sequenzen	[Monks 1994]	abs.	bew.	24bit
		[Vuylsteke and Oosterlinck 1990]	abs.	bew.	2bit
		[Pajdla 1995]	period.	bew.	2bit
		[Salvi et al. 1998]	abs.	bew.	24bit
		[Zhang et al. 2002]	abs.	st./bew.	24bit
	M-Arrays	[Spoelder et al. 2000] [Morano et al. 1998]	abs.	bew.	2bit
			abs.	stat.	2-24bit
	Graustufen	[Carrihill and Hummel 1985] [Miyasaka et al. 2000]	abs.	stat.	8bit
abs.			bew.	8bit	
abs.			bew.	8bit	
Farb-Codierung	[Smutny and Pajdla 1996] [Wust and Capson 1991] [Forster 2005]	abs.	stat.	24bit	
		period.	bew.	24bit	
		abs./per.	bew.	24bit	

ein Muster oder eine Abfolge von Mustern derart darzustellen, dass diese in aufgenommenen Szenen eindeutig identifiziert und ihrer Ursprungsposition zugeordnet werden können. Das grundsätzliche Vorgehen ist hierfür bei allen Verfahren ähnlich, die mit codiertem Licht arbeiten. Es werden bestimmten Bereichen der Projektionsebene – dies kann ein Dia, ein Videoprojektor oder ein ähnliches Gerät sein – eindeutige Bezeichnungen zugewiesen, üblicherweise durch eine Nummerierung. Die entsprechenden Identifikationsnummern werden dann in Muster umgesetzt, wobei wahlweise durch Farben, Formen oder zeitliche Folgen eine eindeutige Identifikation sichergestellt wird. Kann dann bei einer Bildaufnahme ein Punkt des Projektionsmusters von der Kamera eindeutig identifiziert werden, ist das Korrespondenzproblem gelöst – eine Triangulation kann unter Durchführung eines Schnitts der Sichtgeraden der Kamera mit der „Lichtgeraden“ oder „Lichtebene“ des Projektors erfolgen.

Tabelle 1 zeigt eine Übersicht der gängigsten Verfahren, deren wichtigste Grundprinzipien im Folgenden beschrieben werden. Auf die in der Tabelle ebenfalls erwähnten Verfahren der De-Bruijn-Sequenzen, die aus der kombinatorischen Mathematik stammen, und M-Arrays, die zumeist mit zufällig (pseudo-random) angeordneten farbigen Punkten zur Codierung arbeiten, wird dabei jedoch nicht näher eingegangen, sondern auf die angegebene Literatur verwiesen. Im Vergleich mit den anderen gelisteten Verfahren liefern diese Methoden oftmals Ergebnisse, die für exakte Vermessungen nicht geeignet sind oder deren Verarbeitung vergleichsweise rechenintensiv ist, wie sich in Experimenten gezeigt hat (vgl. [Salvi et al. 2004]). Das Hauptaugenmerk liegt deshalb auf zeitcodierten Verfahren, Phase-Shifting und Kombinationen davon, sowie der direkten Codierung mit monochromer oder farbiger Beleuchtung.

2.5.3.2 Zeitcodierte Verfahren

Die zeitliche Codierung von Lichtmustern ist die älteste und wohl am weitesten verbreitete und erforschte Methode zur dreidimensionalen Vermessung mit codiertem, strukturiertem Licht. Die Codierung erfolgt hierbei durch eine Sequenz von Mustern, deren Kombination erst eine eindeutige Identifikation möglich macht. Diese zeitliche Codierung erlaubt aus zwei Gründen eine relativ hohe Genauigkeit und Robustheit der Codierung.

Zum einen ist der „Grundwortschatz“ (Primitive) des Codes zumeist relativ klein, da lediglich Binärwerte (schwarz-weiß) zum Einsatz kommen, die einfach zu unterscheiden sind. Zum anderen erlaubt die sequentielle Abfolge eine immer feiner werdende Positionsbestimmung des einzelnen Bildpunktes, was die letztendliche Genauigkeit der Bestimmung weiterhin erhöht. In den letzten Jahrzehnten sind einige Verfahren entwickelt worden, die diese Strategien zeitlicher und monochromer Projektionen verfolgen. Diese „klassische“ Form der Binärcodierung, die aufgrund ihrer eingeschränkten Farbtiefe etliche Vorteile bei der Bildverarbeitung bietet, wird nun anhand verschiedener Beispiele näher beschrieben.

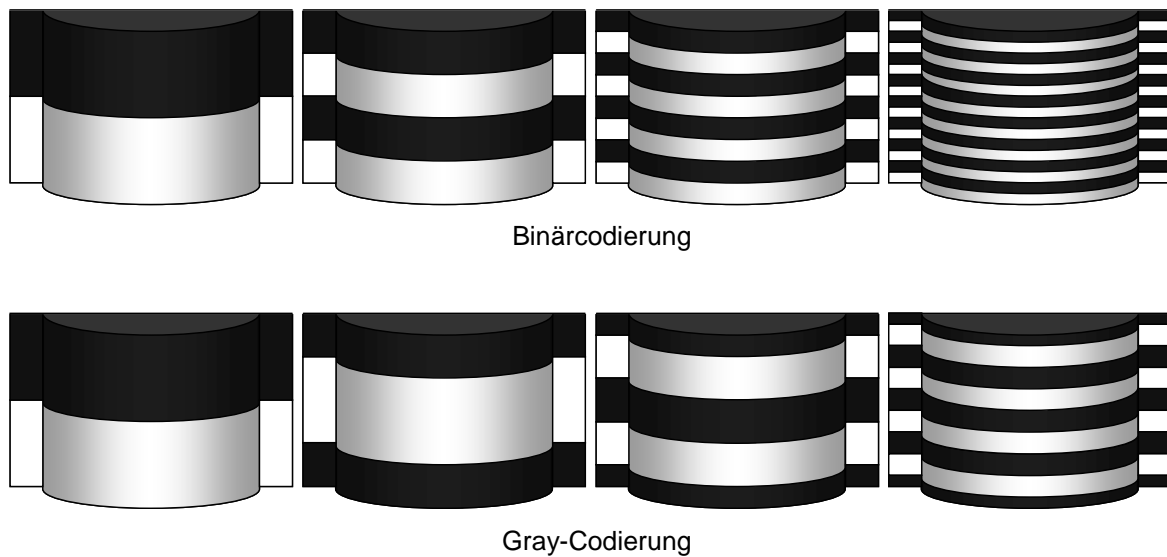


Abbildung 16 Zeitliche Codierung einer Szene. Ein Lichtstreifen kann entweder „echt“ binär codiert sein, so dass die Streifen strikt durchnummeriert sind, oder durch den codierungs-theoretisch geschickteren Gray-Code, bei dem die Hamming-Distanz zwischen benachbarten Streifen stets 1 beträgt (s.a. Tabelle 2).

Bereits Anfang der 80er Jahre stellten [Potsdamer and Altschuler 1982] ein Verfahren zur zeitlichen Mustercodierung vor, bei dem durch eine Sequenz von m Mustern 2^m Streifen eindeutig bestimmt werden. Jedes Codewort, bzw. jeder Streifen, kann somit durch m binäre Symbole dargestellt werden. In der Projektion wurden helle (weiße) Streifen der 1 zugeordnet, dunkle (schwarze) Streifen dementsprechend der 0 (vgl. Abbildung 16). Die maximale Anzahl der Streifen ist dabei prinzipiell nur durch die Auflösung des Projektors begrenzt, es könnten also Streifen einer Breite von einem Bildpunkt projiziert werden.

Allerdings gilt es zu beachten, dass diese Streifen für eine Kamera und die anschließende Bildverarbeitung auch detektierbar sein müssen. In der Praxis ist also die Anzahl und Breite der Codestreifen durch das Sampling-Theorem bestimmt, so dass die Auflösung der Kamera und der Musterprojektion aufeinander abgestimmt werden müssen. Sind die Streifen schließlich identifiziert und die Positionen decodiert, kann eine Triangulation zur Tiefenberechnung durchgeführt werden. Die Berechnung erfolgt auf Grundlage des inzwischen bekannten Schnitts einer Lichtebeine und einer Sichtgeraden. Wie unter Abschnitt 2.5.2 dargestellt, muss hierfür wieder entweder das Zentrum des feinsten Lichtstreifens bestimmt werden oder man verwendet die Kante zwischen zwei codierten Streifen.

Das Verfahren der Binärcodierung wurde unter anderem durch [Inokuchi et al. 1984] mit der Einführung des Gray-Codes zur Streifencodierung verbessert. Der Vorteil einer Gray-Codierung im Vergleich zu einer einfachen Binär-Codierung liegt darin, dass aufeinanderfolgende Codeworte stets eine Hamming-Distanz von 1 haben, wie Tabelle 2 zeigt, und somit robuster gegen auftretende Störungen sind. Der untere Teil von Abbildung 16 zeigt die entsprechende Codierung der Szene mit einem Gray-Code. Aus beiden Grafiken ist

Codewort-Nr	Binärcodierung	$h_{\text{Vorgänger}}$	Gray-Codierung	$h_{\text{Vorgänger}}$
00	0 0 0 0	4	0 0 0 0	1
01	0 0 0 1	1	0 0 0 1	1
02	0 0 1 0	2	0 0 1 1	1
03	0 0 1 1	1	0 0 1 0	1
04	0 1 0 0	3	0 1 1 0	1
05	0 1 0 1	1	0 1 1 1	1
06	0 1 1 0	2	0 1 0 1	1
07	0 1 1 1	1	0 1 0 0	1
08	1 0 0 0	4	1 1 0 0	1
09	1 0 0 1	1	1 1 0 1	1
10	1 0 1 0	2	1 1 1 1	1
11	1 0 1 1	1	1 1 1 0	1
12	1 1 0 0	3	1 0 1 0	1
13	1 1 0 1	1	1 0 1 1	1
14	1 1 1 0	2	1 0 0 1	1
15	1 1 1 1	1	1 0 0 0	1

Tabelle 2 Verschiedene Verfahren der zeitlichen Codierung. Dargestellt ist der Einfluss der verwendeten Codierungsmethode (Binärcodierung oder Gray-Code) auf die resultierenden Codeworte und deren Hamming-Distanzen untereinander. Beim Gray-Code liegt diese Distanz stets bei 1 (s.a. Abbildung 16).

ersichtlich, dass sich mit $m = 4$ Mustern $2^m = 16$ Streifen codieren lassen, wodurch sich die in der Tabelle aufgelisteten Codeworte ergeben.

Es wurden zudem weitere Varianten der Gray-Codierung oder Kombinationen des Gray- und Binärcodes entwickelt ([Minou et al. 1981], [Trobina 1995]). Die Grundlage für eine erfolgreiche Decodierung steht und fällt aber bei allen Verfahren mit der erfolgreichen Detektion und Interpretation der Helligkeit eines Pixels. Während bei homogenen, einfarbigen Oberflächen noch eine einfache Schwellwertdetektion denkbar ist, so ist dies für texturierte Oberflächen nicht immer möglich. Deshalb ist es hilfreich, die Schwellwerte für bestimmte Bildpunkte dynamisch zu bestimmen. Hierfür werden bei den beschriebenen Verfahren üblicherweise voll- und nicht-beleuchtete Referenz-Bilder mit aufgenommen. Alternativ können auch für jedes projizierte Muster invertierte Aufnahmen verwendet werden, wodurch speziell eine Kantendetektion zwischen zwei Streifen vereinfacht wird.

Vor allem in den letzten Jahren wurden die Verfahren der Binärcodierung dahingehend verbessert, die Exaktheit der Codierung und der Detektion auf Subpixel-Genauigkeit zu verbessern (u.a. [Rocchini et al. 2001]). Denn wie bereits erwähnt, kann die Codierung des Musters nur so fein erfolgen, wie es die Kamera zulässt. Bei der Kombination eines handelsüblichen Video-Projektors mit einer Auflösung von 1024x768 Bildpunkten und einer Kamera mit Standard-Video-Auflösung von 780x580 ist deshalb eine Codierung von mehr als 256 Streifen wenig sinnvoll, da sonst eine eindeutige Detektion kaum möglich ist. Daraus ergeben sich Streifen bestimmter Breite, deren Zwischenwerte für eine feinere Auffächerung der Lichtebenen entweder interpoliert oder durch weitere Methoden bestimmt

werden müssen. Eine Möglichkeit dafür liefert beispielsweise die im folgenden Punkt beschriebene Phasenverschiebung.

Ein grundsätzlicher Nachteil aller Verfahren der zeitlichen Codierung ist aber, dass sie in erster Linie für statische Szenen eingesetzt werden können. Denn bei der Aufnahme bewegter Objekte verschieben sich zwangsläufig die einzelnen Muster zueinander, was sich in Rekonstruktionsfehlern widerspiegeln kann.

2.5.3.3 Phasenverschiebung

Ein Muster, das auf dem Gray-Code oder einer vergleichbaren binären Codierung basiert, hat den großen Vorteil, dass eine Positionsbestimmung für jeden Bildpunkt einzeln und ohne Betrachtung der Nachbarschaft möglich ist. Dennoch limitiert genau diese diskrete Natur der Codierung auch die Gesamtauflösung eines solchen Musters, da stets Streifen bestimmter Breite detektiert werden. Ganz im Gegensatz dazu erlauben Verfahren wie die Phasenverschiebung (Phase-Shifting) aufgrund ihrer kontinuierlichen Natur auch eine kontinuierliche Positionsbestimmung. Man formuliert dabei keine Codierung des Musters im eigentlichen Sinne, sondern vollzieht eine Positionsbestimmung unter Verwendung phasenverschobener, periodischer Intensitätsverteilungen.

Abbildung 17 zeigt drei sinusförmige, um 120° phasenverschobene Muster, die eine exakte Positionsbestimmung innerhalb jeder Periode ermöglichen. Der Intensitätsverlauf der einzelnen Projektionen ist durch folgende Gleichungen bestimmt:

$$\begin{aligned} \text{I. } I_0 &= \hat{I} + A \cdot \sin(\varphi) \\ \text{II. } I_{120} &= \hat{I} + A \cdot \sin\left(\varphi + \frac{2}{3}\pi\right) \\ \text{III. } I_{240} &= \hat{I} + A \cdot \sin\left(\varphi + \frac{4}{3}\pi\right) \end{aligned} \quad (42)$$

Dabei bezeichnet φ die Position des projizierten Bildpunktes innerhalb einer Periode, \hat{I} den Offset, bzw. Mittelwert des Intensitätsintervalls der Projektion, und A die Amplitude, die nötig ist, um den Intensitätsumfang des verwendeten Projektors abzudecken. Beim Einsatz eines 8bit-Graustufen-Video-Projektors würde dies Werten von $\hat{I} = A = 127,5$ entsprechen. Nimmt man mit einer Kamera die entsprechenden Projektionen auf einer Objektoberfläche auf, so erhält man drei Intensitätswerte I_0 , I_{120} und I_{240} . Diese ermöglichen schließlich ein Lösen des Gleichungssystems aus (42) und erlauben die Bestimmung der „Position“ φ :

$$\varphi = \arctan\left(\frac{2I_0 - I_{120} - I_{240}}{\sqrt{3}(I_{120} - I_{240})}\right) \quad (43)$$

Eine solche Positionsbestimmung liefert sehr genaue Ergebnisse, da sie unabhängig von der Textur des Objekts erfolgt und nur die aufgenommenen Intensitätswerte als Variablen benötigt. Allerdings beruhen die Berechnungen auf kontinuierlichen Grauwertverteilungen gemäß der trigonometrischen Winkelfunktionen, die sich nur mit analogen Projektionen

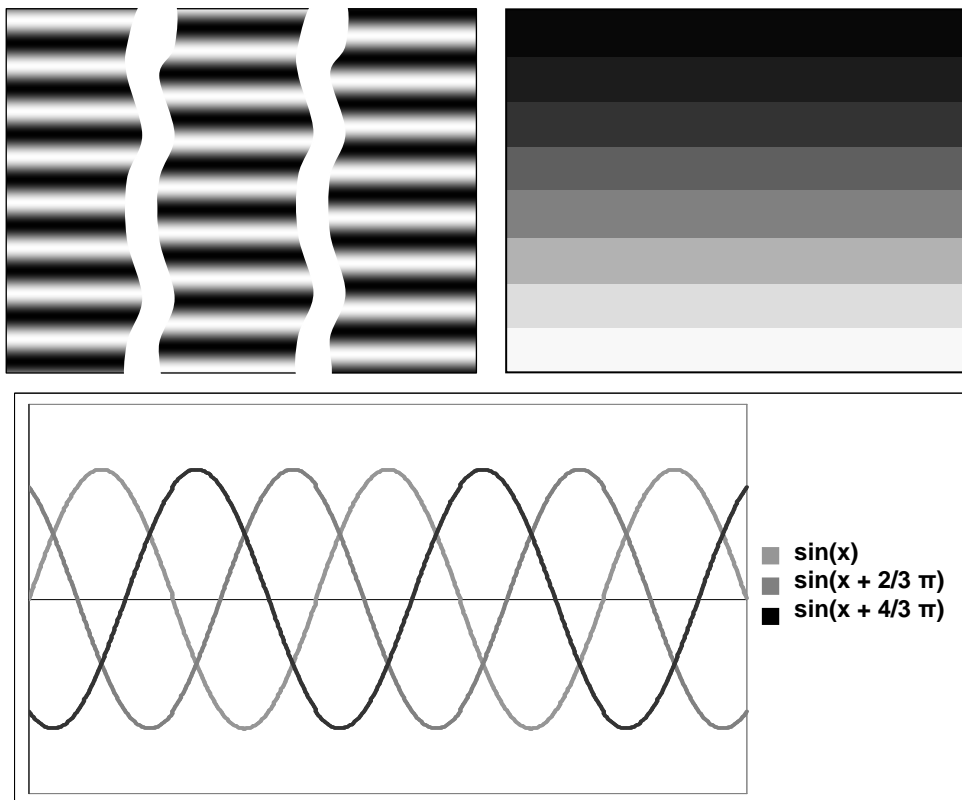


Abbildung 17 Phasenshift-Verfahren. Durch zueinander verschobene Sinuskurven (links oben und unten) können Positionen innerhalb eines bestimmten Bereichs, exemplarisch dargestellt durch verschiedene Graustufen rechts oben, eindeutig codiert und bestimmt werden.

wirklich erreichen lassen. Digitale Projektoren können stets nur diskrete Werte liefern, so dass mit kleinerer Periode auch die Anzahl der dargestellten Werte abnimmt und sich damit die Genauigkeit der Positionsbestimmung reduziert. Zudem ist es nur möglich, die Position innerhalb einer Periode festzustellen – nicht aber die Lage der Periode selbst. Ein sinnvoller Einsatz der Tiefenbestimmung mittels Phasenverschiebung ist also nur unter Verwendung weniger, oder gar nur einer Periode der Sinusmuster möglich.

Allerdings bieten Hybridansätze mit einer Kombination aus einer binären Codierung und einer anschließenden Phasenverschiebung interessante Möglichkeiten zur Effizienzsteigerung einer Messung (z.B. [Bergmann 1995]). Erlauben zeitcodierte Verfahren eine robuste Bestimmung von definierten Bereichen (Streifen) fester Breite in einer Szene, so lassen sich durch gezielte Projektion phasenverschobener Muster mit einer Periodenlänge der binären Streifenbreite die Positionen innerhalb eines Binärstreifens exakt rekonstruieren. Abbildung 17 stellt diese Kombination schematisch dar, wobei die sinusförmigen Muster jeweils einen der durch Graustufen gekennzeichneten Streifen abdecken. Somit ist also eine flächendeckende, kontinuierliche Muster-Rekonstruktion und damit verbunden eine theoretisch flächendeckende Tiefenberechnung möglich – mit dem Nachteil, dass ein sol-

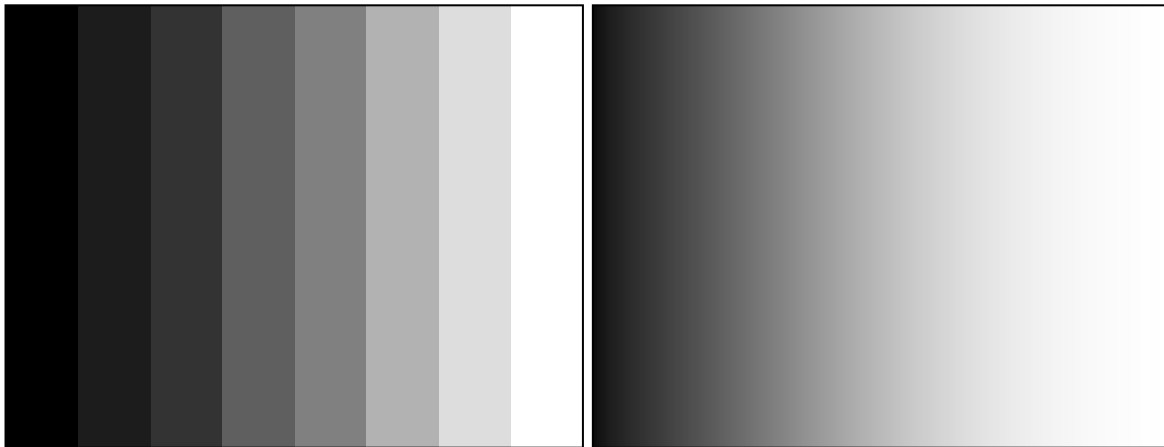


Abbildung 18 Direkte Codierung durch einen Graukeil. Je nach Farbtiefe kann eine Szene grob (z.B. 3bit, links) oder fein (z.B. 8bit, rechts) direkt codiert werden, wobei jeder Graustufe ein bestimmter Bereich bzw. Lichtstreifen zugeordnet wird.

ches Verfahren wiederum nur für statische Szenen geeignet ist, da viele Muster nacheinander projiziert und aufgenommen werden müssen.

2.5.3.4 Direkte Codierung

Der offenbar größte Nachteil der bisher behandelten Verfahren ist die vergleichsweise lange Dauer einer Objekterfassung, meist bedingt durch die notwendige Aufnahme mehrerer Bilder oder Muster. Es existieren aber, wie bereits in Tabelle 1 aufgelistet, verschiedene Möglichkeiten, ein Muster derart zu codieren, dass jeder Punkt eindeutig identifiziert werden kann. In diesem Fall spricht man von einer direkten Codierung.

Es ist naheliegend, dies durch das Erhöhen der Farbtiefe, verbunden mit dem Einsatz von Graustufen oder Farben, zu erreichen. Denn bereits dadurch wird der Informationsgehalt eines Musters vergrößert und erlaubt einen umfangreicheren Wortschatz für die Codierung. Allerdings gilt vor allem für den Einsatz verschiedener Farbkanäle zu beachten, dass projizierte Farben durch Oberflächentexturen verfälscht werden können. Zudem ist das Betrachten feiner aufgelöster Intensitätsunterschiede anfälliger für Störungen, als dies bei binären Werten der Fall ist.

Doch bereits mit dem Einsatz uncodierter Intensitätsunterschiede, ohne Farbinformationen, ist die direkte Codierung einer ganzen Szene vorstellbar. Bereits 1985 entwickelten [Carrhill and Hummel 1985] den sogenannten „Intensity Ratio Depth Sensor“, der eine Szene mit nur einem Graustufenmuster direkt codiert. Dieses Muster ist ein linear ansteigender Graukeil, wie in Abbildung 18 dargestellt.

Jedem Bildpunkt, bzw. jeder Bildspalte, wird darin ein bestimmter Intensitätswert zugewiesen, dessen relatives Verhältnis zu einer konstanten Beleuchtung eine eindeutige Identifikation ermöglichen soll. Da dies allerdings eine Referenzaufnahme – eben unter konstanter, gleichmäßiger Beleuchtung – erfordert, sind stets zwei Aufnahmen für eine Be-

rechnung erforderlich. Carrhill und Hummel bauten das System aus einem einfachen Projektor und einer monochromen Kamera mit 8 Bit Farbtiefe auf, erreichten aber nur eine relativ grobe Auflösung. Im Jahr 2000 bauten [Miyasaka et al. 2000] diesen Sensor mit aktueller Hardware neu auf und verwendeten dabei einen LCD-Projektor und eine 3-Chip CCD-Kamera. Dieser Aufbau wurde 2002 unter Verwendung eines DLP-Projektors von [Miyasaka and Araki 2002] aktualisiert. Unter Berücksichtigung der Oberflächentextur und mit einer getrennten Untersuchung der drei Farbkanäle konnte mit diesen Aufbauten eine deutlich höhere Genauigkeit erzielt werden. Da die Berechnung jedoch stets auf dem Verhältnis der Intensitäten eines „codiert“ beleuchteten und eines neutralen (gleichmäßig oder nicht beleuchteten) Bildpunkts basiert, ist sie relativ anfällig für etliche Störfaktoren, wie Fremdlicht, Signalrauschen oder ausgeprägte Oberflächentexturen. Deshalb lassen sich mit diesen Verfahren beliebige Szenen und Objekte nicht ohne Einschränkungen erfassen.

2.5.3.5 Farbcodierte Verfahren

Das Ziel der farbcodierten Verfahren ist grundsätzlich das Gleiche wie der im vorigen Kapitel dargestellten Verfahren mit monochromer Beleuchtung – das Codieren einer beliebigen Szene mit möglichst wenigen Mustern, so dass möglichst viele Bildpunkte eindeutig identifiziert werden können. Dabei verspricht der Einsatz von Farben einen erheblichen Vorteil, da damit prinzipiell 3 Intensitätskanäle simultan genutzt werden können. Mit dem Einsatz von farbigen Mustern ist auch das Verwenden von Farbkameras verbunden, deren grundsätzliche Funktionsweise in Abschnitt 2.1.2.3 vorgestellt wurde.

An der direkten Codierung durch Intensitätsverteilung orientieren sich Projektionen eines Regenbogenmusters, wie es unter anderem von [Smutny and Pajdla 1996] entwickelt wurde. Dabei werden Intensitätsänderungen auf alle Farbkanäle verteilt, was eine exaktere Positionsbestimmung ermöglichen soll. Mit diesen Verfahren sind zwar sehr dichte und hochauflösende Tiefenkarten möglich, allerdings ist die Genauigkeit aus ähnlichen Gründen wie beim ursprünglichen Graukeilverfahren relativ gering. Zudem können die Messungen durch Texturen der Objektoberfläche beeinflusst werden.

Auf Basis der Phasenverschiebung entwickelten [Wust and Capson 1991] ein Verfahren, welches mit drei um 90° verschobenen Sinusmustern arbeitet, die auf die drei Farbkanäle verteilt werden. Daraus resultiert das in Abbildung 19 dargestellte Muster. Durch diese Überlagerung ist eine Positionsbestimmung innerhalb einer Periode des Musters mit einer einzigen Aufnahme möglich, so dass grundsätzlich ein einzelnes Bild zur Tiefenberechnung ausreicht. Allerdings verlangt dieses Vorgehen eine relativ farbneutrale Oberfläche des aufzunehmenden Objekts, da die Teilmuster aufgrund ihrer Begrenzung auf einen schmalen Wellenlängenbereich durch Oberflächentexturen beeinflusst werden können.

Eine vergleichbare Methode wurde von [Zhang and Huang 2006] implementiert, wobei jedoch nur indirekt Farben oder Farbkanäle genutzt wurden. In diesem Fall wurde ein DLP-Projektor durch Entfernen des Farbrades derart modifiziert, dass jeder Farbkanal le-

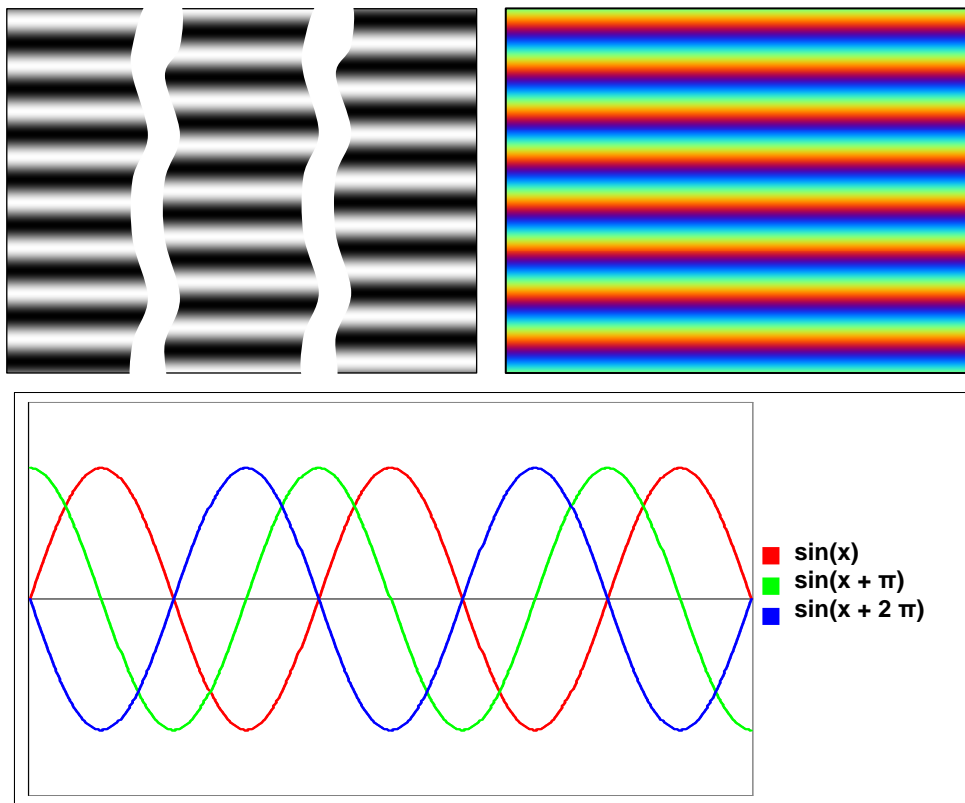


Abbildung 19 RGB-Phasenshift. Durch farbige Projektion und farbige Bildaufnahme kann ein dreifacher Phasenshift in einem einzelnen Bild aufgenommen werden, indem jeder Phase ein Farbkanal zugeordnet wird.

diglich zur reinen Graustufendarstellung fähig ist. Dieser Projektor wurde mit einer High-Speed-Graustufen-Kamera und einer Farb-CCD-Kamera gekoppelt, um einerseits sehr schnell die drei nötigen Phasenbilder aufnehmen zu können, wodurch das Erfassen bewegter Objekte ermöglicht wird, und andererseits eine zweidimensionale Texturaufnahme des Objekts zu erhalten. Obwohl der Hardwareaufwand für diesen Vorschlag komplexer ist als bei einfachen Kamera-Projektor-Kombination, erlaubt er die dreidimensionale Modellierung selbst bewegter Objekte inklusive ihrer Oberflächentextur.

Ein bestehendes Problem beider Verfahren bleibt aber die Unterscheidung der einzelnen Perioden untereinander. Denn je mehr Perioden eine Szene codieren, desto schwieriger ist die eindeutige Periodenzugehörigkeit eines Bildpunkts zu identifizieren. Mit wachsender Zahl der Perioden wird dann entweder zusätzliche Information zur Unterscheidung nötig, oder die Tiefeneindeutigkeit wird beschränkt, um Zweideutigkeiten zu vermeiden. Weiterhin gilt es auch hier die von der Phasenverschiebung bekannte Diskretisierung der Intensitätswerte zu berücksichtigen. Diese bringt den Nachteil einer nur hinreichend genauen Abbildung der kontinuierlichen trigonometrischen Funktionen mit sich.

Diese Probleme durch den Einsatz von Farben zur eindeutigen Codierung einer Szene zu lösen, ist aber dennoch möglich. Hierfür wurde von Forster und anderen (z.B. [Forster et

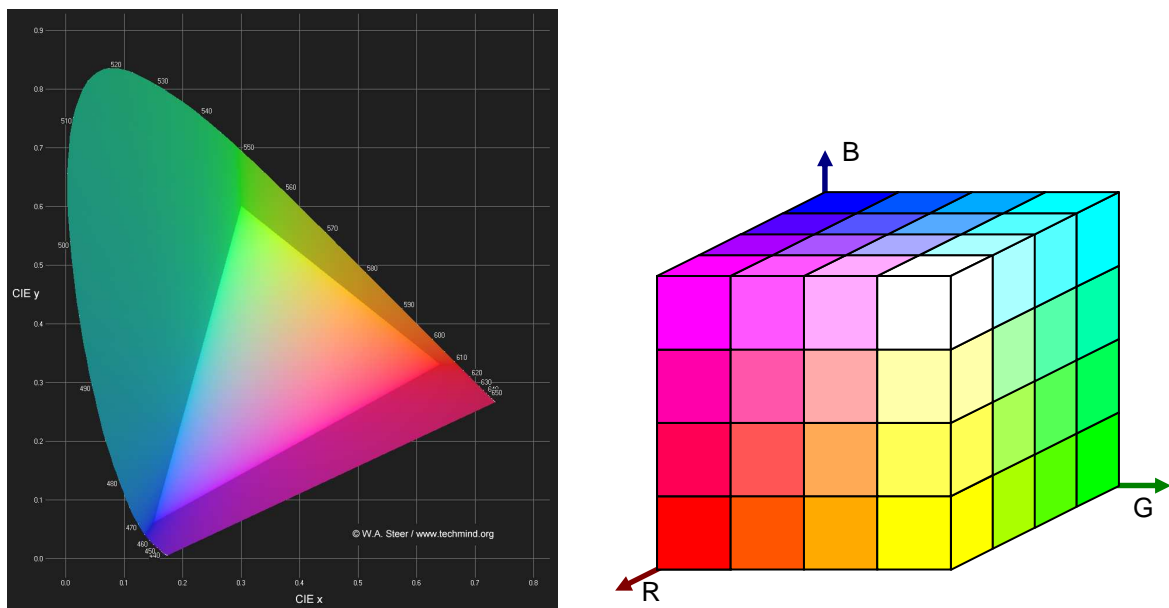


Abbildung 20 CIE-Farbtabelle und RGB-Farbwürfel. Das CIE-Normvalenzsystem (links, 1931 CIE Chromaticity diagram, [Steer 2008]) stellt eine Relation zwischen der menschlichen Farbwahrnehmung und den physikalischen Ursachen der Farbvalenz her. Der Farbwürfel (rechts) baut sich aus den diskreten Abstufungen der drei Grundfarben – Rot, Grün und Blau – und deren Mischfarben auf.

al. 2002], [Forster 2005]) das Verfahren der farbcodierten Triangulation (Color Coded Triangulation, CCT) entwickelt. Die Idee des Verfahrens ist es, eine komplette Szene mittels eines einzigen farbigen Musters eindeutig zu codieren, so dass keine weiteren Aufnahmen zur 3D-Rekonstruktion erforderlich sind. Als Farbraum werden die 8 Farben verwendet, die die Ecken des CIE-Farbwürfels bilden (vgl. Abbildung 20) – Rot, Grün, Blau, Gelb, Cyan, Magenta, Weiß und Schwarz.

Diese Farben werden als Streifen projiziert, die derart in Untergruppen angeordnet sind, dass die jeweilige Zusammenstellung im Muster einmalig ist, so wie es in Abbildung 21 symbolisch dargestellt ist. Die einzelnen Farben sind so gewählt, dass jeder Farbkanal für sich betrachtet durch Binärwerte dargestellt wird. Dementsprechend sind den einzelnen Farben Symbole in Form von Zahlen zugewiesen, die analog einer 3bit-Zuordnung die Werte zwischen 0 und 7 annehmen können, abhängig von der Präsenz eines jeden Kanals. Durch Zusammenfassen mehrerer Symbole entstehen dann Codeworte, die aufgrund der Vielzahl der möglichen Kombinationen einmalig im gesamten Muster sind. Der eigentliche Algorithmus der 3D-Berechnung ist dabei wie bei den meisten streifenprojizierenden Verfahren kantenbasiert, um die Lage der betrachteten Lichtebenen genauer bestimmen zu können. Eine Lichtebene wird also nicht direkt durch die Farben selbst beschrieben, sondern durch die Farbübergänge an der Kante zwischen den Streifen definiert. Ein Codewort, das beispielsweise aus 4 Farben aufgebaut ist, bestimmt 3 eindeutige Farbübergänge an den 3 eingeschlossenen Kanten. Das darauf folgende Wort ist aus den letzten 3 Farben des aktuellen Codewortes plus einer weiteren Farbe aufgebaut und liefert im Ganzen wie-

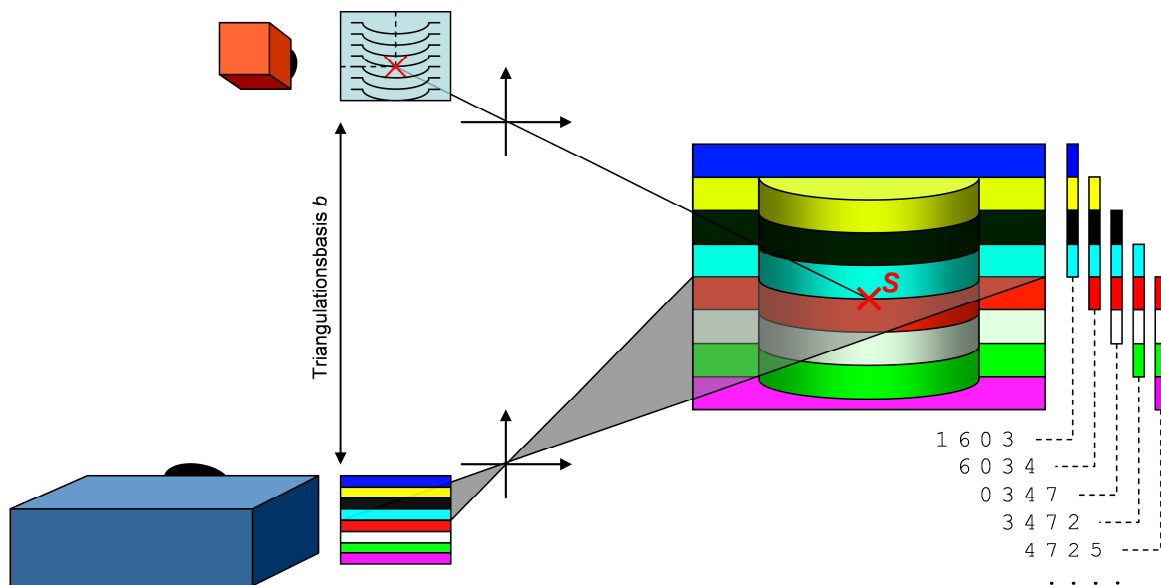


Abbildung 21 Grundaufbau der farbcodierten Triangulation. Dargestellt ist ein aktives Stereosystem, das eine Szene mit einem farblich und, über Nachbarschaftsbeziehung, räumlich codierten Muster beleuchtet und somit eine Tiefenberechnung ermöglicht.

der 3 Kantenübergänge, deren Zusammenstellung eindeutig ist. Nach diesem Verfahren wird ein Muster aufgebaut, das aus 106 Codewörtern mit jeweils 4 Symbolen (Farben) besteht. Ist eine Kante in dem auf ein Objekt projizierten Muster schließlich eindeutig identifiziert und lokalisiert, erfolgt die Tiefenberechnung nach dem bekannten Vorbild des Schnitts der Projektor-Lichtebene mit der Kamera-Sichtgeraden.

Durch die Notwendigkeit nur eines Projektionsmusters kann ein Messsystem nach diesem Verfahren aus jedem Farb-Video-Projektor, aber auch mit einem einfachen Dia-Projektor, aufgebaut werden. Dieser Ein-Bild-Ansatz erlaubt dabei relativ hohe Aufnahmezeiten, die grundsätzlich nur von der eingesetzten Hardware und folglich der Erfassungsgeschwindigkeit der Kamera und der Rechenleistung des Systems abhängen. Dadurch können selbst bewegte Objekte aufgenommen werden. Die Auflösung der Tiefendaten hängt dabei von der verwendeten Kombination aus Kamera und Projektor ab und entspricht horizontal der Anzahl der projizierten Streifen und vertikal der Kameraauflösung. Die Genauigkeit ist mit etwa $0,3\text{mm}$ für Objekte im Abstand von 700mm ebenfalls als hoch zu bezeichnen [Forster 2005]. Allerdings führt die Projektion von sichtbarem Licht zu einer sehr starken Blendwirkung, die eine länger andauernde Aufnahme eines Gesichts, beispielsweise für eine dreidimensionale Video-Sequenz, erschweren und die Akzeptanz eines Sensors herabsetzen können.

Farbcodierte Verfahren versprechen also eine gute Ausgangsbasis für ein echtzeitfähiges System zur Aufnahme beliebiger Objekte. Wenn die Nachteile der sichtbaren Beleuchtung gemindert werden können, könnten sie auch bedenkenlos für biometrische Anwendungen eingesetzt werden.

3 Sensoren zur dreidimensionalen Gesichtserfassung

Im vorangehenden Kapitel wurden verschiedene Verfahren zur dreidimensionalen Objekterfassung vorgestellt. Diese Verfahren ermöglichen es, in vielen Bereichen die Effektivität und Genauigkeit von Mess- und Steuerungsoperationen zu verbessern und zu erhöhen. Einen ebensolchen effizienzsteigernden Effekt können dreidimensionale Modelle menschlicher Gesichter für die Sicherheitstechnik (z.B. [Phillips et al. 2007], [Mpiperis et al. 2007], [Tsalakanidou and Malassiotis 2007]), aber auch für die Unterhaltungstechnik (z.B. virtuelle Avatare, [Jung and Kopp 2003], [Lee et al. 2005]) bewirken. Allerdings sind die bisher beschriebenen Verfahren aufgrund ihrer Genauigkeit, Geschwindigkeit oder Projektionsweise kaum für diese Zwecke geeignet. Das folgende Kapitel befasst sich deshalb mit der Entwicklung eines Sensorsystems zur dreidimensionalen, berührungslosen Erfassung des menschlichen Gesichts. Es baut dabei auf der farbcodierten Triangulation auf, wie sie in 2.5.3.5 vorgestellt wurde. Dieses Verfahren wurde aufgrund seiner Robustheit und Messgenauigkeit als Grundlage für weitere Entwicklungen gewählt.

3.1 Anpassung des Wellenlängenbereichs

Wie in Kapitel 2 dargestellt arbeiten aktive 3D-Scanner nach dem Stereo-Prinzip in der Regel mit einer Kombination von Kameras und projizierten Mustern. Üblicherweise werden dabei Projektoren verwendet, die im sichtbaren Lichtspektrum zwischen $380nm$ und $780nm$ (DIN 5031-7, [Reidenbach et al. 2005]) arbeiten – vor allem deshalb, da der Großteil der handelsüblichen Kameras für diesen Wellenlängenbereich ausgelegt ist. Für industrielle Anwendungen ist dieses Vorgehen auch durchaus sinnvoll, vereinfacht es zudem den Aufbau und die Einrichtung eines Sensors, da das projizierte Muster während der Konstruktion, Ausrichtung und Kalibrierung für das Auge beobachtbar ist. Bei der Erfassung von Gesichtern allerdings führt Licht im sichtbaren Wellenlängenbereich zu einer

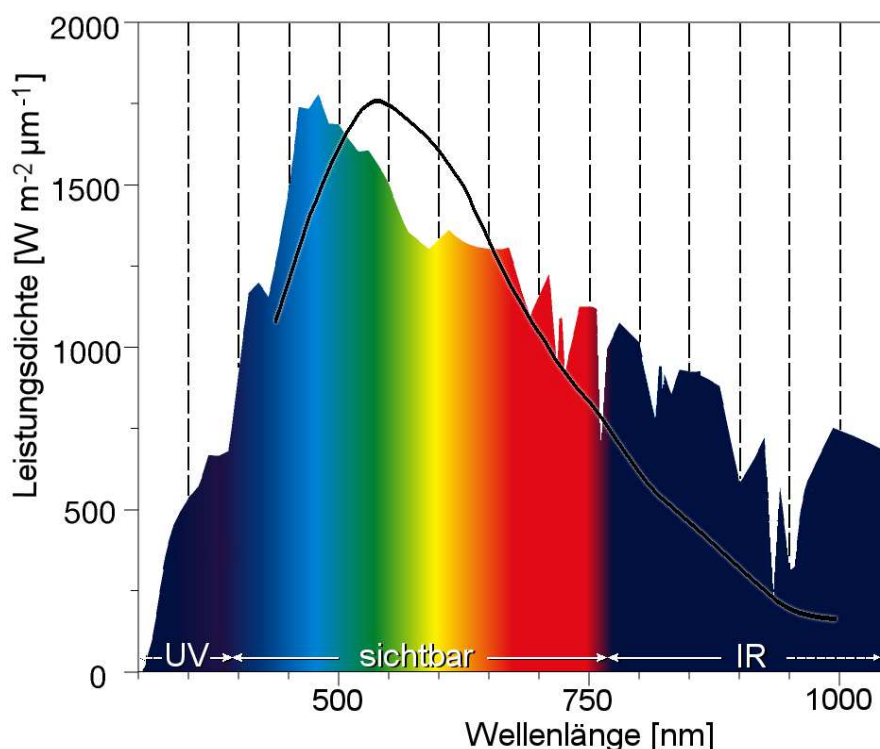


Abbildung 22 Sichtbares Lichtspektrum und Intensität des Sonnenlichts. Die Farben des sichtbaren Spektrums zwischen 380nm und 780nm sind entsprechend ihrer Wahrnehmung dargestellt, die typische Leistungsdichte des Sonnenlichts ist absolut angegeben (Quelle: [HMI 2008]). Die zusätzlich eingetragene Kurve entspricht der relativen spektralen Empfindlichkeit eines handelsüblichen CCD-Sensors (Sony ICX-415AL, [Sony 2008]).

starken Blendwirkung. Dies ist für einzelne Aufnahmen zwar hinzunehmen, für einen alltäglichen Gebrauch, wie beispielsweise bei einer Zugangs- oder Grenzkontrolle, dagegen nicht wünschenswert. Denn neben einer möglichen Schädigung der Netzhaut durch eine starke Beleuchtung bewirkt die Blendwirkung eine geringe Akzeptanz bei potentiellen Nutzern eines Sensors [BioSec 2005]. Um diesen Blendwirkungen entgegenzuwirken, bleibt lediglich die Möglichkeit, das sichtbare Lichtspektrum zu verlassen und entweder in das ultraviolette (UV) oder infrarote (IR) Spektrum zu wechseln (siehe Abbildung 22).

Betrachtet man das Spektrum des Sonnenlichts, so ist ein starker Abfall der Intensität im ultravioletten Anteil unterhalb von 400nm zu beobachten. Eine Verwendung einer Beleuchtung in diesem Wellenlängenbereich erscheint also durchaus sinnvoll, da dort mit geringen Störungen zu rechnen ist. Allerdings kann ultraviolette Strahlung zu einer noch stärkeren Schädigung der Netzhaut führen, als dies bei Licht im Sichtbaren der Fall ist ([Reidenbach et al. 2005], [HCN 1993]). Dies liegt am kurzwelligen Spektrum des UV-Lichts und der damit einhergehenden höheren Photonenenergie. Hierdurch können photochemische Reaktionen hervorgerufen werden (z. B. die sog. Photoretinitis), deren Auswirkungen zu irreversiblen Schäden des Sehapparates führen können. Je länger die Wellenlänge des Lichts dagegen ist, desto weniger Auswirkungen und Schäden sind zu befürchten.

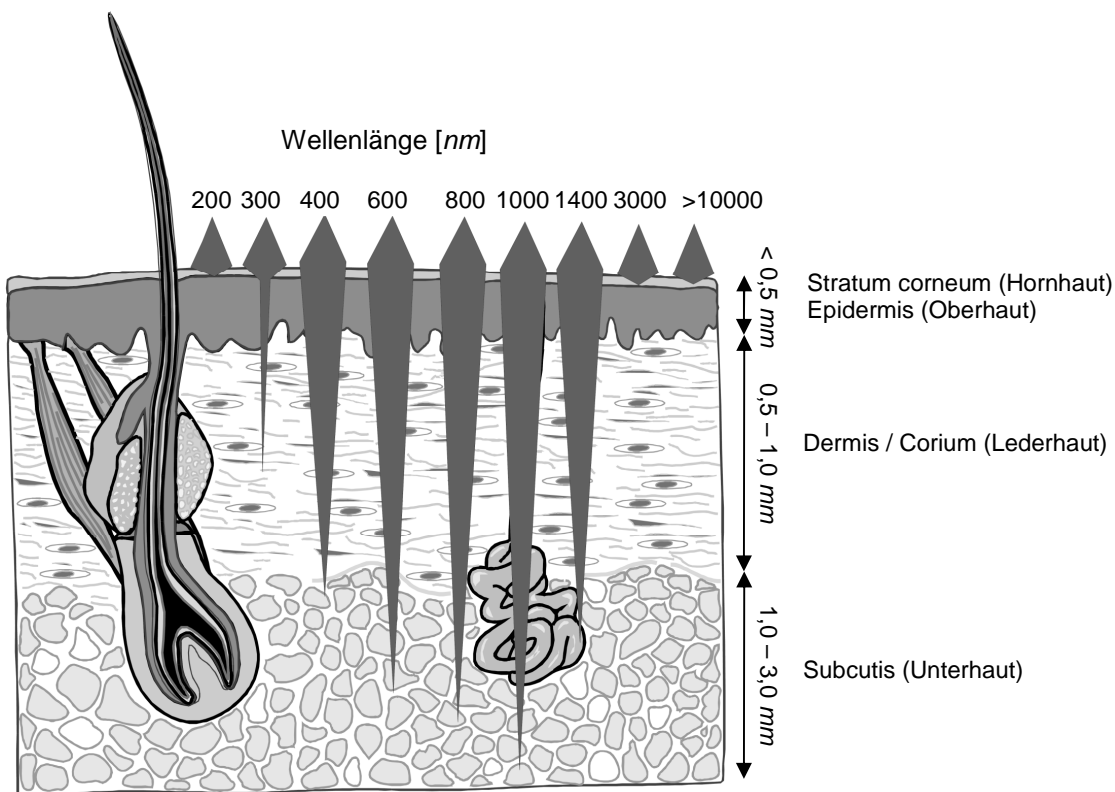


Abbildung 23 Eindringtiefe des Lichts in die menschliche Haut [Reidenbach et al. 2005]. Je nach Wellenlänge kann das Licht verschieden tief in die menschliche Haut eindringen.

ten. So ist beispielsweise die Photonenenergie infraroter Strahlung zu gering, um photochemische Reaktionen hervorzurufen. Aus diesen Gründen ist eine Lichtquelle im Infraroten für eine Projektionen auf das menschliche Gesicht in jedem Fall einer ultravioletten Variante vorzuziehen.

Die Empfindlichkeit der verwendeten Kameras ist ein weiterer entscheidender Punkt der Anpassung des Wellenlängenbereichs, da diese die projizierten Muster auch empfangen und auswerten können müssen. Es gibt zwar lichtempfindliche Sensoren für nahezu jeden Bereich des Lichtspektrums, die jedoch aufgrund ihrer exklusiven und geringen Fertigungszahl sehr teuer ausfallen können und für Alltagsanwendungen nicht in Frage kommen. Betrachtet man dagegen wie in Abbildung 22 die Kurve eines handelsüblichen Sensors, so stellt man zwar einen Abfall der Empfindlichkeit mit sinkender Wellenlänge fest, aber auch eine noch offenbar ausreichende Sensitivität im nahen Infrarotbereich (nahes Infrarot, near infrared, NIR). Es ist also naheliegend, einen Sensor mit „unsichtbarer“, nicht wahrnehmbarer Beleuchtung in diesem Bereich des Lichtspektrums zu entwerfen.

Das NIR-Licht ab einer Wellenlänge von 700nm aufwärts ist dabei vor allem auch deshalb interessant, da das Auge diese Strahlungen als kaum noch oder gar nicht mehr störend empfindet, aber die meisten handelsüblichen Kamerasensoren über diesen Bereich hinaus ausgelegt und deshalb noch ausreichend sensitiv für diese Art der Beleuchtung sind. Es ist

also nicht nötig, für einen Sensor, basierend auf infraroter Beleuchtung, spezielle Aufnahmesensoren zu verwenden.

Für die Anwendung zur Erfassung menschlicher Gesichter allerdings müssen einige Eigenschaften des Infrarotlichts berücksichtigt werden. Denn im Gegensatz zu Licht im sichtbaren Spektrum hat infrarotes Licht die Fähigkeit, bis zu einer gewissen Tiefe in die menschliche Haut einzudringen ([Reidenbach et al. 2005], [HCN 1993], [Meglinski and Matcher 2002]), wie es in Abbildung 23 dargestellt ist. Üblicherweise wird diese eindringende Strahlung als Wärme empfunden, wenn eine Körperregion beleuchtet wird. Abhängig von der jeweiligen Wellenlänge kann das Licht dabei mehrere Millimeter tief durch die Hautschichten vordringen. Dies hat zur Folge, dass auf die Haut projizierte Muster deformiert erscheinen können, da ursprünglich diskrete Kanten unscharf abgebildet werden können. Diese Eigenschaften sind bei der Wahl der Wellenlänge zu berücksichtigen, so dass die Eindringtiefe möglichst gering gehalten wird.

Dabei bietet die notwendige Festlegung auf eine oder verschiedene feste Wellenlängen auch zusätzliche Vorteile. Während man über das breite Spektrum des sichtbaren Lichts stets gegen ein entsprechend breites Spektrum an Störlicht „strahlen“ muss, so erlaubt die Begrenzung der Wellenlänge auf einen schmalen Bereich den Einsatz von Bandpassfiltern. Denn diese Begrenzung auf einen bestimmten Bereich des Spektrums schwächt den Einfluss des Störlichts ab, da dessen Energie durch die Filterung ebenfalls geringer ausfällt. Auch diese Eigenschaften sind beim Entwurf eines Sensors zu berücksichtigen und werden in den folgenden Kapiteln erläutert.

Tabelle 3 Kantenübergänge einer farbigen Codierung. Ein Kanal kann an einer Kante entweder hinzukommen (+), verschwinden (-) oder seinen aktuellen Zustand beibehalten (0). Mehrfaches Hinzukommen oder Verschwinden in Folge ist jedoch ebenso unmöglich wie das doppelte Verwenden der gleichen Farbe in benachbarten Streifen. Hervorgehoben sind die Übergänge mit mehr als einer Kanaländerung.

		Rot	Grün	Blau	Cyan	Magenta	Gelb	Schwarz	Weiß
		1 0 0	0 1 0	0 0 1	0 1 1	1 0 1	1 1 0	0 0 0	1 1 1
Rot	1 0 0	xxxxx	- + 0	- 0 +	- + +	0 0 +	0 + 0	- 0 0	0 + +
Grün	0 1 0	+ - 0	xxxxx	0 - +	0 0 +	+ - +	+ 0 0	0 - 0	+ 0 +
Blau	0 0 1	+ 0 -	0 + -	xxxxx	0 + 0	+ 0 0	+ + -	0 0 -	+ + 0
Cyan	0 1 1	+ - -	0 0 -	0 - 0	xxxxx	+ - 0	+ 0 -	0 - -	+ 0 0
Magenta	1 0 1	0 0 -	- + -	- 0 0	- + 0	xxxxx	0 - +	- 0 -	0 + 0
Gelb	1 1 0	0 - 0	- 0 0	- - +	- 0 +	0 - +	xxxxx	- - 0	0 0 +
Schwarz	0 0 0	+ 0 0	0 + 0	0 0 +	0 + +	+ 0 +	+ + 0	xxxxx	+ + +
Weiß	1 1 1	0 - -	- 0 -	- - 0	- 0 0	0 - 0	0 0 -	- - -	xxxxx

3.2 Reduktion der Farbkanäle

Der elementare Bestandteil der farbcodierten Triangulation ist die Unterscheidung verschiedener Farben, explizit der drei Farbkanäle des RGB-Farbraumes. Um im nahen Infrarotbereich ein farbcodiertes Muster verwenden zu können, müssen nun ebenfalls verschiedene Farben, genauer gesagt verschiedene Wellenlängen, unterschieden werden. Bei der bereits beschriebenen Methode der farbcodierten Triangulation im sichtbaren Licht bauen die drei Grundfarben Rot, Grün und Blau, sowie deren Kombinationen Gelb, Cyan, Magenta, Schwarz und Weiß das codierte Muster auf.

Es stehen also 8 Farben zur Verfügung, um ein Muster zu erzeugen. Abhängig von verschiedenen Faktoren, die die Robustheit, die Fehleranfälligkeit, die Rekonstruierbarkeit und weitere Punkte betreffen, ergibt sich daraus die Zahl der möglichen Kombinationen einer Codierung. Wenn bei einem streifenbasierten Verfahren die Kanten zwischen den Farben zur Codierung verwendet werden, so ist auch das Verhalten der Kanäle an eben diesen zu betrachten. Es können dabei für jeden Farbkanal sowohl ein Ansteigen als auch ein Abfallen des Kanals berücksichtigt werden, die sich abhängig vom Vorgänger und Nachfolger ergeben. Tabelle 3 verdeutlicht diese einzelnen Kombinationen beim Übergang der 8 Farben untereinander und listet das jeweilige Kantenverhalten der Kanäle auf.

Um aus diesen Farben und ihren Kombinationen ein Muster aufzubauen, müssen bestimmte Randbedingungen festgelegt werden, wie die Länge eines einzelnen Wortes oder die Anzahl der Kanaländerungen pro Kante. Im Einzelnen wird an einem Übergang dabei das Hinzukommen einer Farbe durch ein „+“ symbolisiert, das Reduzieren durch ein „-“ und ein sich nicht ändernder Kanal durch eine „0“. Es ergeben sich damit, in Abhängigkeit der Anzahl der Farbkanäle m , der Anzahl der Symbole pro Codewort s und der Mindestanzahl c der Kanaländerungen an einem Streifenübergang

$$N = 2^m \cdot \left[\sum_{i=c}^m \binom{m}{i} \right]^{s-1} \quad (44)$$

mögliche Kombinationen zum Aufbau einer Codierung. Tabelle 4 stellt den Umfang einiger Optionen für den Aufbau eines Codes mit $s = 4$ Symbolen pro Codewort dar. Dabei

Tabelle 4 Codewortumfang eines farbcodierten Streifenmusters mit 4 Symbolen pro Codewort. Die Anzahl der Codeworte hängt dabei von der Zahl der verwendeten Farbkanäle sowie den geforderten Kanaländerungen pro Kantenübergang ab.

Anzahl Farbkanäle	Anzahl mögl. Farben	Kanalübergänge pro Kante	Mögliche Codeworte	Gültige Codeworte
3	8	3	8	8
3	8	2	512	500
3	8	1	2744	2426
2	4	2	4	4
2	4	1	108	104

kann man deutlich erkennen, welchen Einfluss die Anzahl der verwendeten Farbkanäle hat. Außerdem muss die Zahl der rechnerisch möglichen Codeworte weiter reduziert werden, da eben durch die Betrachtung der Kantensymbole unterschiedliche Farbkombinationen die gleiche Kantencodierung hervorrufen können.

Betrachtet man nämlich beispielsweise für eine RGB-Codierung die Farbübergänge von Gelb (1 1 0) auf Schwarz (0 0 0) und von Weiß (1 1 1) auf Blau (0 0 1), so erhält man in beiden Fällen eine identische Kanteninterpretation mit einem Abfallen des Rot- und Grün-Kanals (- - 0), obwohl die reinen Farbinformationen offensichtlich verschieden sind. Auch bei der Reihung mehrerer Farben, wie beispielsweise für ein Codewort mit 4 Symbolen, können Streifenkombinationen trotz unterschiedlicher Farben die gleichen Kanteninformationen enthalten. Somit lassen sich auch die Abweichungen zwischen den möglichen und gültigen Codewort-Zahlen aus Tabelle 4 erklären.

Trotz solcher Reduktionen erlaubt der Einsatz von 3 Farbkanälen den Aufbau eines Codemusters, das zusätzliche Eigenschaften enthalten kann, die eine Rekonstruktion und Interpretation des Musters auf beleuchteten Objekten vereinfacht. Es können beispielsweise über die gesamte Codierung eine feste Hamming-Distanz eingehalten oder Methoden zur Erkennung und Korrektur von Fehlern eingebettet werden.

Für eine „unsichtbare“ Projektion ließen sich grundsätzlich die Farbkanäle des sichtbaren Lichts direkt in das infrarote Spektrum übertragen, indem man jede Farbe als bestimmte Wellenlänge interpretiert. Das würde allerdings voraussetzen, dass diese Wellenlängen von einer Kamera auch eindeutig identifiziert werden können – so wie es unter Abschnitt 2.1.2.3 für Farbkameras mit der Unterscheidung roter, grüner und blauer Komponenten beschrieben wurde. Theoretisch ließen sich im Infraroten die gleichen Verfahren verwenden, also ein Mosaik-Muster oder die Kombination mehrerer Sensoren. Da solche Kameras aber Spezialanfertigungen darstellen würden, wäre ihr Einsatz wenig praktikabel, vor allem wenn man die lange Geschichte und Entwicklungszeit bedenkt, die nötig waren, um heutige Digitalkameras zur Farberfassung herzustellen.

Daher ist es sinnvoll, die Zahl der verwendeten Wellenlängen und damit der Farbkanäle zu reduzieren, um somit das Problem der eindeutigen Unterscheidung zu vereinfachen. Für den Aufbau eines Code-Musters im Infraroten werden deshalb lediglich die Kombinationen zweier Wellenlängen λ_1 und λ_2 betrachtet. Dies führt zu einer starken Reduktion des Codewortschatzes, wie es in Tabelle 4 ersichtlich ist. Unter Berücksichtigung der doppeldeutigen Kanten ergibt sich ein Umfang von 104 Codeworten, deren Kanten lediglich eine Änderung in mindestens einem Kanal aufweisen. Die Symbole setzen sich dabei wie in Tabelle 5 dargestellt zusammen. Die Konsequenzen einer solchen Verkleinerung des Codewortschatzes werden im Folgenden betrachtet.

		λ_1	λ_2	--	$\lambda_1 \lambda_2$
		1 0	0 1	0 0	1 1
λ_1	1 0	XXXXXX	- +	- 0	0 +
λ_2	0 1	+ -	XXXXXX	0 -	+ 0
--	0 0	+ 0	0 +	XXXXXX	+ +
$\lambda_1 \lambda_2$	1 1	0 -	- 0	- -	XXXXXX

Tabelle 5 Kantenübergänge einer Codierung mit 2 Grundfarben und deren Mischungen. Die Übergänge unter den Farben der Codestreifen sind dabei analog zu Tabelle 3 dargestellt.

3.2.1 Auswirkungen einer Kanalreduktion auf die Codierung

Eine Reduktion der Farbkanäle beim Übergang vom sichtbaren ins infrarote Lichtspektrum hat also, wie eben beschrieben, einen direkten Einfluss auf die Codierung des Projektionsmusters und damit indirekt auch auf die Auflösung und Genauigkeit des Messverfahrens der farbcodierten Triangulation.

Wie bereits dargestellt hängt die Auflösung eines optischen Messverfahrens, das auf der farbcodierten Triangulation basiert, von der Dichte und Struktur des projizierten Musters ab. Bei den streifenbasierten Mustern ist dies die Anzahl der eindeutig codierten Lichtebenen. Ohne weitere Einschränkungen oder Codierungsvorschriften ließen sich also mit einem farbcodierten Streifenmuster, das aus 4 Symbolen mit jeweils 3 Farbkanälen besteht, 2426 Lichtebenen eindeutig voneinander unterscheiden (siehe Abschnitt 3.2, Tabelle 4). Dies überschreitet deutlich die Auflösung heutiger Projektoren, sowie der für Videoaufnahmen geeigneten Kamerasensoren. Zudem ist eine gewisse Streifenbreite des projizierten Musters auf dem aufgenommenen Objekt vonnöten, um Streifen und Kanten zu detektieren und der folgenden Bildverarbeitung vernünftige Ergebnisse zur Szenenrekonstruktion zur Verfügung zu stellen. In zahlreichen Tests hat sich eine Streifenbreite von 3-5 Pixeln bei der Aufnahme auf dem Kamerasensor als optimal herausgestellt, was etwa 200 projizierten Streifen auf Seiten des Projektors entspricht – bemessen an einer aufzunehmenden Fläche von $0,5m \times 0,5m$ in einem Abstand von etwa $0,7m$ und einer horizontalen Kameraauflösung von etwa 800 Bildpunkten. Mit der hohen Anzahl an Kombinationen bei 3 Farbkanälen lässt eine Codierung hierbei große Redundanzen zu. Eine Reduktion der Farbkanäle erlaubt dagegen lediglich den Aufbau eines deutlich geringeren Wortschatzes. Die einfachste Möglichkeit zur Lösung dieser Problematik wäre eine Erhöhung der Symbolanzahl pro Codewort. Bei einer Erhöhung von 4 auf 5 Symbole pro Wort würde sich somit für zwei Farbkanäle die Zahl der rechnerisch möglichen Codeworte auf 324 erhöhen, abzüglich doppeldeutiger Kanten. Allerdings ruft jede Erhöhung der Symbole einen erhöhten Decodierungsaufwand hervor und verlangt eine größere zusammenhängende und codierte Fläche zur Identifikation aller Symbole. Eine Erhöhung der Symbolzahl pro Codewort ist demnach wenig praktikabel.

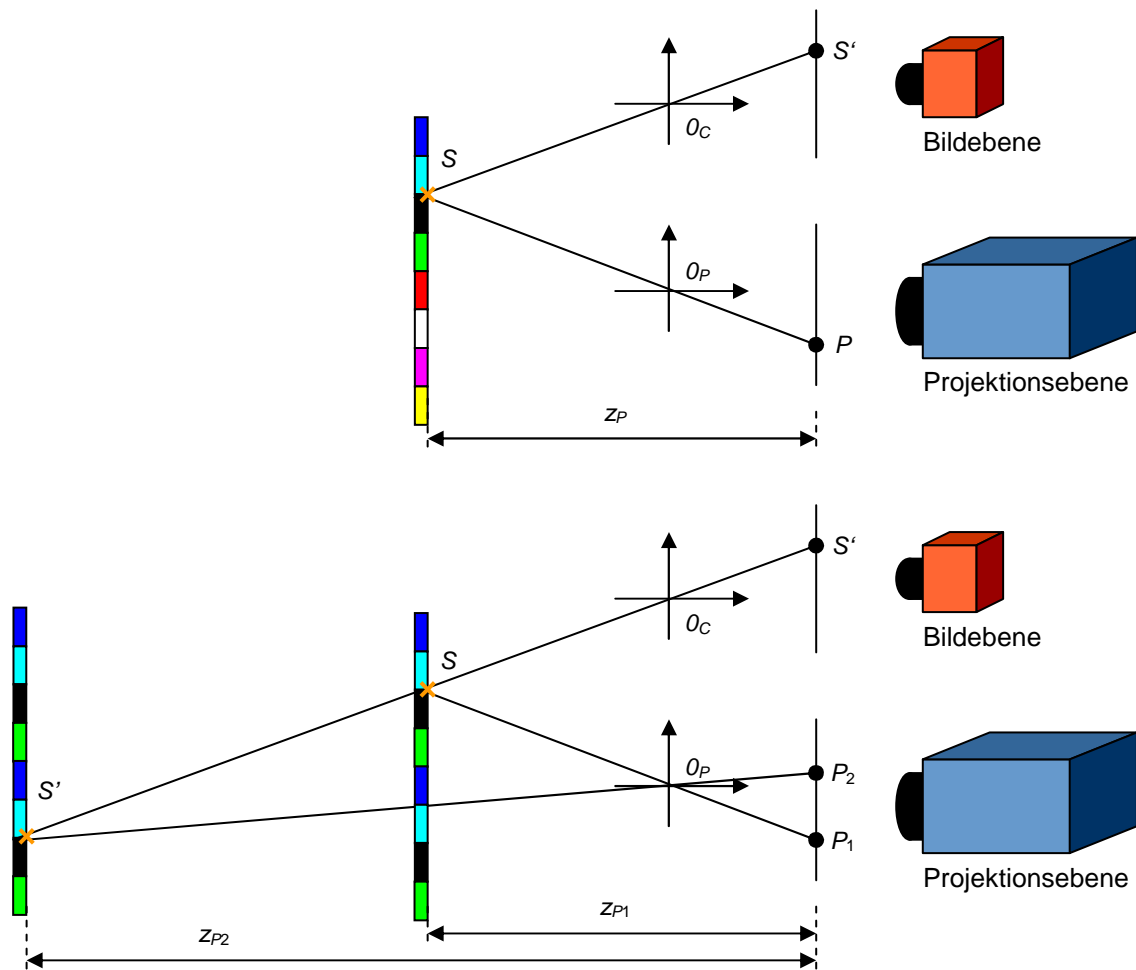


Abbildung 24 Tiefeneindeutigkeit eines periodischen Musters. Sind alle Elemente eines Codes eindeutig identifizierbar (oben), so lässt sich für jedes Codewort ein eindeutiger Tiefenwert berechnen. Wiederholen sich Codeworte und existieren somit Doppeldeutigkeiten, so herrscht keine Tiefeneindeutigkeit – diese muss beispielsweise über eine Arbeitsraumbegrenzung hergestellt werden.

Eine bessere Möglichkeit dagegen ist, das codierte Muster zu wiederholen. Dies erlaubt es, mit einer relativ geringen Anzahl von Codeworten ein beliebig großes Muster aufzubauen. Allerdings tritt dabei das Problem der Tiefeneindeutigkeit auf, das bereits von reinen Streifenprojektionen und Phase-Shifting-Verfahren (siehe Abschnitt 2.5.2 sowie 2.5.3) bekannt ist. Während bei einem komplett codierten Muster jede Lichtebene zu einem eindeutigen Tiefenwert führt (Abbildung 24 oben), kann bei einem periodischen Muster eine Lichtebene mehrfach interpretiert werden und dadurch zu unterschiedlichen Werten führen (Abbildung 24 unten). Diese Doppeldeutigkeit lässt sich jedoch durch eine Beschränkung des Arbeitsraumes weitgehend ausschließen. Denn lässt man nur Werte eines bestimmten Tiefenbereichs als Ergebnisse zu, so können Fehlinterpretationen von vornherein ausgeschlossen werden. Diese Arbeitsraumeinschränkung beeinflusst dabei den Messbereich eines Sensorsystems bei einmaliger Wiederholung eines Musters kaum, da die Opti-

ken der Projektoren und Kameras ohnehin nur einen bestimmten Tiefenschärfebereich aufweisen und somit der Arbeitsraum bereits dadurch festgelegt wird. Mit steigender Periodenanzahl sind jedoch immer weitere Einschränkungen des Arbeitsraumes hinzunehmen, so dass eine beliebige Wiederholung eines Codes zwangsläufig zu weiteren Identifikationsproblemen, wie bei uncodierten Verfahren, führen würde.

Merklich beeinflusst werden durch eine Kanalreduktion dagegen die Möglichkeiten, Zusatzbedingungen in die Codierung einzubauen. So ist es bei Verwendung aller zulässiger Codeworte und Berücksichtigung einfacher und mehrfacher Kantenübergänge nicht möglich, eine konstante Hamming-Distanz über die komplette Codierung beizubehalten. Ebenfalls erlaubt es die mangelnde Redundanz nicht, ungültige Codeworte herauszufinden – außer, es handelt sich um nicht definierte Farbkombinationen. Während also eine Korrektur fehlender oder nicht vollständig erkannter Symbole bei einem fehlerredundanten Code durch die Codierung selbst durchführbar ist, muss dies bei einer Kanalreduktion anderweitig vollzogen werden. Wie dies ohne weitere Beeinträchtigung der Messergebnisse möglich ist, wird unter 3.5 näher erläutert.

3.2.2 Robustheit gegenüber Störeinflüssen

Um einen Sensor zur Bild-, Daten- oder 3D-Aufnahme effektiv einsetzen zu können, sollte er eine gewisse Robustheit gegenüber möglichen Störeinflüssen aufweisen. Um den Einfluss von Störungen ausschließen oder bekämpfen zu können, müssen diese aber zunächst identifiziert werden.

Bei optischen Messverfahren, die auf der Projektion codierter Muster basieren, kann dabei folgenden drei Faktoren der größte Einfluss zugeschrieben werden:

- *Bewegungsartefakte*
Abbildungsfehler, hervorgerufen durch nicht-ruhende Objekte.
- *Oberflächenbeschaffenheit und Struktur*
Mögliches Auftreten von Abbildungsfehlern oder Interferenzen, wie beispielsweise beim Moiré-Effekt (vgl. auch Abbildung 26).
- *Fremd-, bzw. Umgebungslicht*
Unter anderem Tageslicht, Raumbeleuchtung oder Oberflächenreflexionen.

Bei der Aufnahme von menschlichen Gesichtern liegt der eigentliche Störeinfluss oftmals im aufzunehmenden Objekt selbst, bzw. in dessen Bewegung während der Aufnahmen. Bei ruhenden Gegenständen, wie in technischen Anwendungen, kann dieser Punkt dagegen außer Acht gelassen werden. Aber bei einer biometrischen 3D-Erfassung von Gesichtern oder sonstigen „natürlichen“ Objekten lassen sich Objektbewegungen kaum vermeiden. Die Frage ist allerdings, wie schwer diese Störungen ins Gewicht fallen und wie groß ihre Auswirkungen sind.

Es lassen sich dabei grundsätzlich zwei Folgeerscheinungen von Objektbewegungen unterscheiden. Zum einen die Bewegungsunschärfe innerhalb einer Aufnahme, zum anderen Positionsänderungen zwischen zwei Bildern. Diese sind vor allem bei Verfahren entscheidend, die ihre Daten aus mehreren Aufnahmen beziehen, so wie es z.B. bei binär codierten Mustern der Fall ist. Da hierfür mehrere Bilder aufgenommen werden müssen – für eine Codierung von 256 Streifen sind immerhin 8 Bilder nötig – macht sich eine Objektbewegung deutlich bemerkbar.

Die Auswirkung lässt sich anschaulich nachvollziehen, indem man, wie in Abbildung 25 dargestellt, den Strahlengang einer Bildaufnahme betrachtet. Eine Objektbewegung kann nämlich erst bemerkt werden, wenn sich die Abbildung auf der Sensorebene ändert. Das bedeutet, die Positionsänderung des Objekts muss auf der Sensorebene eine Verschiebung um mindestens einen Pixel bewirken. Gemäß der geometrischen Grundlagen und der Prinzipien der Kameraabbildungen gilt dabei folgende Beziehung:

$$\frac{d_y}{f} = \frac{\Delta y}{z} \Rightarrow \Delta y = z \frac{d_y}{f} \quad (45)$$

Geht man beispielsweise von einem Sensor mit einer quadratischen Pixelgröße von $8,3\mu\text{m}$ (z.B. Sony ICX 415AL) und einem Objektiv mit einer Brennweite von $8,5\text{mm}$ auf der Aufnahmeseite aus, mit dem man ein Objekt in 650mm Entfernung erfassen möchte, so ergibt sich gemäß Gleichung (45) für eine Verschiebung um einen Pixel eine Objektbewegung von $0,635\text{mm}$. Dies entspräche bei einer Aufnahmezeit von 25 Bildern pro Sekunde, also einer Belichtungszeit von 40ms , einer Geschwindigkeit des Objekts von $0,0159\text{m/s}$.

Natürlich ist es nicht zu garantieren, derart ruhig vor einer Kamera zur Bildaufnahme zu stehen. Eine gewisse Bewegungsunschärfe tritt also nahezu bei jeder Bildaufnahme eines Menschen auf, fällt allerdings nur bei schnelleren Bewegungen und damit größeren Verschiebungen ins Gewicht. Denn im Normalfall lassen sich geringe Unschärfen oder andere Ergebnisse einer Bewegungsunschärfe, wie „Verwischungen“, durch entsprechende Bildverarbeitungsschritte ausgleichen. Bei kantenbasierten Verfahren wird beispielsweise der kontinuierliche Übergang zweier Streifen betrachtet und die Kantenposition aus dem Extremwert ermittelt. Die leichte Unschärfe kommt dem Vorgehen dabei sogar eher zugute, da das Bild geglättet wird. Eine Reduktion der Farbkanäle hätte somit bezüglich eventueller Bewegungsartefakte keinen merklichen Einfluss auf die Bildqualität, da die entscheidenden Parameter, die die Bildaufnahme betreffen, unverändert bleiben.

Einleuchtend ist dagegen, dass Algorithmen, die auf der Aufnahme vieler Einzelbilder basieren, bei üblichen Belichtungszeiten nur unzureichend zur Aufnahme bewegter Objekte geeignet sind. Bewegt sich ein Objekt mit der oben berechneten Geschwindigkeit, so würde sich zwischen dem ersten und letzten Bild einer Serie eine Verschiebung von mehreren Pixeln einstellen, die dann sehr wohl das Messergebnis oder die Oberflächenrekonstruktion verfälschen könnten.

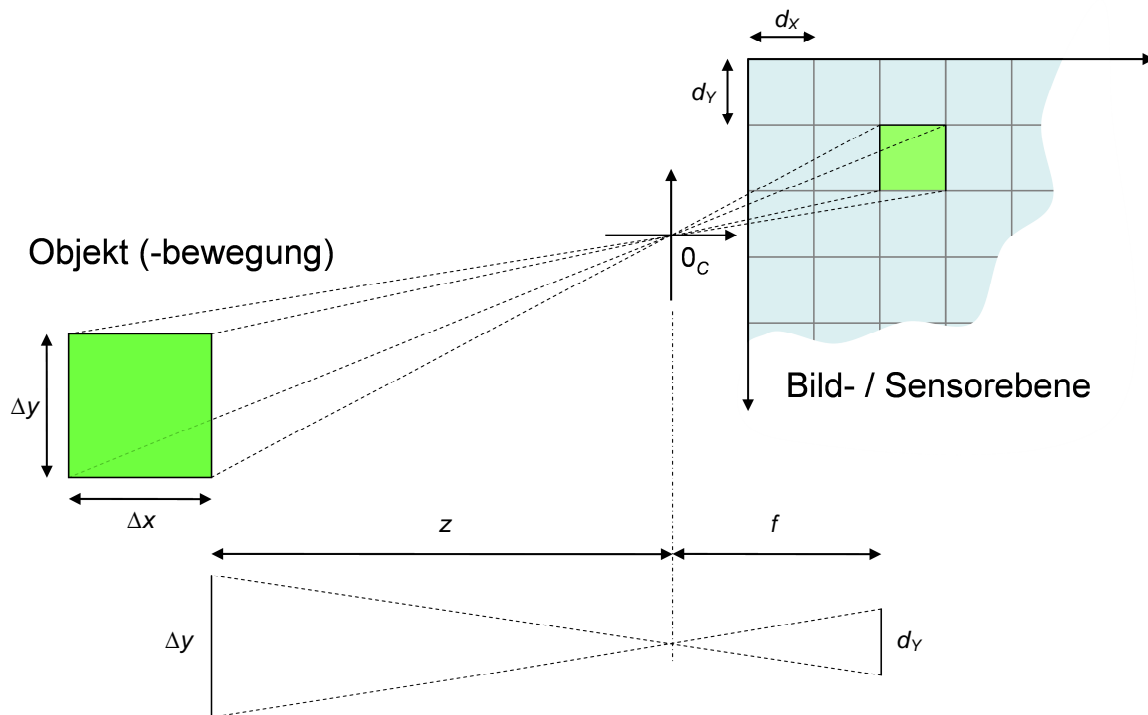


Abbildung 25 Folgen einer Objektbewegung für die Bildaufnahme. Der Einfluss einer Objektbewegung auf die Pixelabweichung im aufgenommenen Bild ist abhängig von der Objektentfernung z und der Objektivbrennweite f .

Ähnliche Auswirkungen auf die Bild- oder Kantenschärfe kann das Oberflächenmaterial des aufzunehmenden Objekts hervorrufen, wie es bereits in 3.1 mit dem Eindringen bestimmter Wellenlängen in die menschliche Haut erläutert wurde. Ebenso können die Struktur und Beschaffenheit einer Oberfläche eine Oberflächenrekonstruktion erschweren und zu Bildstörungen führen. Denn entscheidend für eine Szenenrekonstruktion ist, dass die Oberfläche mit dem auf sie projizierten Muster auf dem Sensor der Kamera so abgebildet wird, dass die Codierung erkannt werden kann. Weist ein Objekt dabei sehr feine Strukturen auf, die zudem noch in Tiefe und Abstand variieren, wie beispielsweise die menschliche Kopfbehaarung, so können auf einen einzelnen Bildpunkt keine eindeutigen Informationen abgebildet werden. Eine Rekonstruktion wäre für einen solchen Fall nur möglich, wenn die Auflösung der aufnehmenden Kamera derart hoch wäre, dass die entsprechenden Strukturen unterschieden werden können.

Ebenfalls durch Strukturen auf der Oberfläche hervorgerufen wird der Moiré-Effekt. Der Moiré-Effekt ist ein Interferenzmuster, das durch Überlagerung zweier Gitterstrukturen hervorgerufen wird, wenn diese gegeneinander verdreht sind oder verschiedene Gitterabstände aufweisen [Jähne 2002]. Ein periodisches und regelmäßiges projiziertes Muster könnte also auf einer ebenso regelmäßig strukturierten Oberfläche scheinbare Interferenzen hervorrufen. Da dies bei einem Farbcode aufgrund der verhältnismäßig groben Gitter-

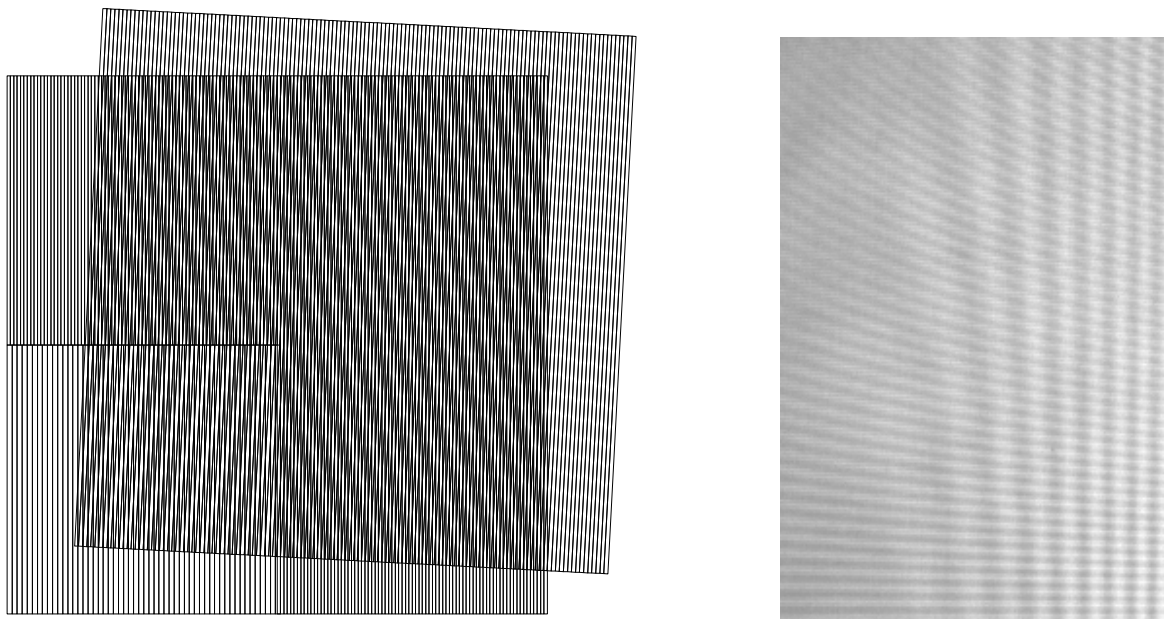


Abbildung 26 Moiré-Effekt. Die Folgen einer Abtastung verdrehter Gitter, sowie verschiedener Gitterabstände, sind in der linken Abbildung (nach [Jähne 1991]) deutlich sichtbar. Das gedrehte „abtastende“ Gitter hat den 1,2-fachen Abstand des Ursprunggitters, die linke untere Ecke den 1,5-fachen. Die rechte Abbildung zeigt die Auswirkungen in der Realität anhand der Aufnahme einer weiß projizierten Fläche bei einer Auflösung des Projektors von 800×600 Punkten und der Kamera von 782×582 Punkten.

auflösung jedoch eher kaum zutrifft, müssen die Strukturen der Projektionseinheit und des Kamerasensors betrachtet werden. Verwendet man einen Videoprojektor zur Darstellung des codierten Musters, so wird das projizierte Bild aus einzelnen Pixeln aufgebaut. Je nach der zugrunde liegenden Technologie des Projektors (siehe hierzu auch 3.6.3) ist dabei eine Pixelstruktur in Form eines Gitters innerhalb der Projektion erkennbar. Da diese Abbildung vom Sensor der Kamera abgetastet wird, der ebenfalls ein Gitter einzelner Bildpunkte darstellt, findet de facto eine Überlagerung zweier verschiedener Gitterstrukturen statt. Dies kann zu ausgeprägten Moiré-Effekten führen (siehe Abbildung 26).

Diesem Phänomen liegt das Abtasttheorem (Nyquist-Shannon-, bzw. Whittaker-Kotelnikow-Shannon-Theorem) zugrunde [Koch et al. 1998]. Dessen Kernaussage fordert dabei für die Abtastung analoger Signale eine Abtastfrequenz, die mehr als das Doppelte der Signalfrequenz beträgt:

$$f_{\text{Abtast}} > 2 \cdot f_{\text{Signal}} \quad (46)$$

Anders ausgedrückt bedeutet das, dass eine periodische Struktur aus den Werten einer Abtastung nur dann richtig rekonstruiert werden kann, wenn sie mit mindestens der doppelten Strukturfrequenz abgetastet wird [Jähne 1991]. Bei der Projektion zur farbcodierten Triangulation müsste also entweder die Pixelzahl der Kamera doppelt so hoch sein wie die des verwendeten Projektors oder das projizierte Muster dürfte keine erkennbare periodische Struktur aufweisen. Würde man also beispielsweise Dias oder ähnliche Filter zur

Mustererzeugung verwenden, so könnte eine Quelle für Bildstörungen beseitigt werden. Eine Erhöhung der Kameraauflösung dagegen ist für ein System, das echtzeitfähig arbeiten soll, weniger praktikabel, da mit steigender Pixelzahl auch die Bearbeitungszeit der Tiefenberechnung ansteigen würde. Die direkten Folgen des Moiré-Effekts auf die Oberflächenrekonstruktion werden im Rahmen einer Evaluierung unter 3.6 weiter ausgeführt.

Im Vergleich zu den bisher dargestellten Störeinflüssen stellen die Auswirkungen von Fremdlicht, das von der Kamera erfasst wird, und damit verbunden die optimale Ausleuchtung eines Objekts wohl die größte Herausforderung dar. Vor allem das Sonnenlicht kann man als den „größten Feind“ der Bildaufnahme und Szenenausleuchtung bezeichnen. Denn während man in geschlossenen Räumen mit künstlicher Beleuchtung meist konstante Beleuchtungssituationen vorfindet, so kann sich die Strahlungsintensität der Sonne, bedingt durch Bewölkung oder Ähnliches, innerhalb kürzester Zeit verändern.

Für ein aktives Stereosystem bedeutet dies, dass sich das projizierte Muster stets von der ohnehin vorhandenen Oberflächenreflexion und Textur eines Objekts absetzen bzw. unterscheiden lassen muss. Betrachtet man hierbei die ausgestrahlte Leistung der Sonne im sichtbaren Spektralbereich, wie sie beispielsweise in Abbildung 22 dargestellt wurde, so ist klar, dass eine zur Projektion verwendete Lichtquelle eine relativ hohe Leuchtkraft haben muss. Denn arbeitet man im sichtbaren Wellenlängenbereich und verwendet die drei Grundfarben sowie deren Mischungen, so erstreckt sich die Bandbreite zwischen 380nm und 780nm . Dementsprechend muss eine Lichtquelle über diesen gesamten Bereich in der Lage sein, die Lichtleistung der Sonne zu überstrahlen. Mit Hochleistungslampen in Videoprojektoren ist das, bis zu einer gewissen Objektentfernung, durchaus gut möglich. Die hohe Lichtleistung solcher Projektoren führt aber bei einer Gesichtserfassung zwangsläufig zu Blendwirkungen und kann, ab einer Leistung von $10\text{W}/\text{m}^2$, zudem zu Schädigungen der Netzhaut führen [Reidenbach et al. 2005]. Für eine Reduktion der Farbkanäle, bzw. das Design einer Projektionseinheit, ist es deshalb erstrebenswert, mit einer geringeren Lichtleistung ein gleichwertiges Ergebnis zu erhalten.

Dabei bringt die Verschiebung in den infraroten Spektralbereich nicht unbedingt Vorteile mit sich. Das dort abgestrahlte Fremdlicht ist zwar für den Menschen unsichtbar, aber nur ungleich leistungsschwächer als im sichtbaren Bereich. Das bedeutet demzufolge, dass dieses Licht von einer für diese Wellenlängen empfindlichen Kamera ebenfalls mit aufgenommen wird und sich ein projiziertes Muster deshalb weiterhin darüber absetzen können muss. Auf eine Kanalreduktion hat das freilich keinen Einfluss, wohl aber auf die verwendbaren Leuchtmittel. Da für den infraroten oder nahen infraroten Wellenlängenbereich keine Projektoren existieren, die mit Video-Projektoren vergleichbar sind, muss auf andere Leuchtmittel zurückgegriffen werden. Diese sind tendenziell leistungsschwächer, weshalb der Fremdlichteinfluss anderweitig reduziert werden muss.

Dabei kommt es der Anwendung zugute, dass es sich lediglich um zwei Farbkanäle mit zwei diskreten Wellenlängen handelt. Deshalb lässt sich der Spektralbereich, beispielsweise durch Bandpassfilter, derart begrenzen, dass nur noch Licht der erwünschten Wellen-

längen auf den Sensor der Kamera gelangt. Somit ist es möglich, die volle Leistung der Projektion zu nutzen und überflüssig empfangenes Fremdlicht zu filtern. Für künstliche Beleuchtung, die überwiegend im sichtbaren und kaum im infraroten Spektrum aktiv abstrahlt, ergeben sich dadurch noch größere Vorteile, da dort theoretisch alles Fremdlicht gefiltert werden könnte. Es sollte also grundsätzlich möglich sein, durch eine Verschiebung in den nahen Infrarotbereich und eine damit verbundene Kanalreduktion ein für Umgebungslicht robusteres System mit geringerer Lichtleistung aufzubauen, als dies im sichtbaren Spektrum der Fall ist.

Inwieweit eine praktische Ausführung der hier beschriebenen Maßnahmen tatsächlich Wirkung zeigt, wird ausführlich unter 3.6 im Rahmen weiterer Experimente, Laboraufbauten und Vergleichstest verschiedener Verfahren erläutert.

3.3 Aktive NIR-Stereoskopie unter Verwendung zweier Wellenlängen

Die im vorangehenden Kapitel beschriebenen Überlegungen dienen als Grundlage zum Aufbau eines Sensorsystems, das mittels aktiver Stereoskopie eine hochauflösende dreidimensionale Objekterfassung ermöglicht, ohne dabei eine Blendwirkung auf menschliche Benutzer auszuüben. Der Grundgedanke dabei ist die Erzeugung eines codierten Musters aus zwei Kanälen unter Anwendung verschiedener Wellenlängen [Modrow et al. 2007]. Um dafür eine einfache Abstraktion des nicht sichtbaren Lichts zu schaffen, wird im Folgenden weiterhin der Begriff des „Farbkanals“ zur Unterscheidung der verwendeten Wellenlängen verwendet. Schließlich bezieht sich die Benennung der Grundfarben prinzipiell auch auf drei diskrete Wellenlängen.

3.3.1 Erläuterung des Prinzips

Der Transfer des trichromatischen Farbraums mit drei der Projektion dienenden Farbkanälen in einen anderen Spektralbereich ließe sich, wie in den voranstehenden Kapiteln beschrieben, ohne Probleme vollziehen, indem jede der drei Grundfarben durch eine spezifische Wellenlänge im Infraroten interpretiert wird. Ein weiteres Problem bei einer Verschiebung des Spektralbereichs wurde bereits ebenfalls eingegrenzt – die Aufnahme der beleuchteten Szene mittels einer geeigneten Kamera.

Die Idee des im Folgenden beschriebenen Verfahrens ist der Aufbau eines infraroten, aktiven Stereoskopiesensors aus möglichst einfachen Standardkomponenten. Das Vorgehen gliedert sich folglich in mehrere Teile:

- Erzeugen eines farbcodierten Musters aus zwei Farbkanälen
- Aufbau einer Projektionseinheit zur infraroten Musterdarstellung
- Aufbau einer Aufnahmeeinheit mit spezifischer Wellenlängendetektion

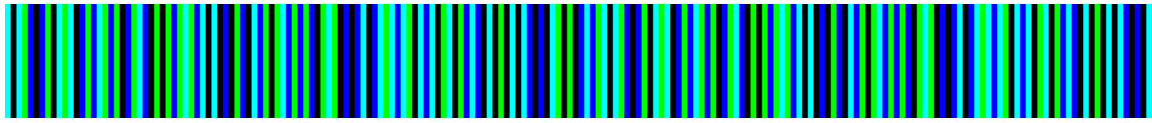


Abbildung 27 Ergebnis einer Codierung mit 2 Kanälen. Die Kanäle, bzw. Symbole, werden dargestellt durch die Grundfarben Grün und Blau, die Mischfarbe Cyan und die Farbe Schwarz.

3.3.2 Erzeugen eines farbcodierten Musters aus zwei Farbkanälen

In Abschnitt 3.2 wurden bereits der Übergang von 3 auf 2 Wellenlängen und die daraus resultierenden Folgen beschrieben. Eine Kanalreduktion ermöglicht demnach ohne größere Auswirkungen den Aufbau eines farbcodierten Musters, allerdings mit einem reduzierten Wortschatz.

Für den zum Aufbau des hier vorgestellten Sensorsystems verwendeten Code werden folgende Rahmenparameter festgelegt:

- Das Muster ist räumlich codiert und besteht aus parallel angeordneten Streifen konstanter Breite.
Jedes Symbol wird also durch einen Farbstreifen repräsentiert, zwei Farbstreifen bilden eine definierte Kante, eine Kombination mehrerer Streifen und damit auch Kanten bildet ein Codewort.
- Ein Codewort besteht aus 4 Symbolen.
Somit besteht ein ausgeglichenes Verhältnis zwischen Codewortumfang und Verarbeitungszeit zur 3D-Berechnung.
- Benachbarte Streifen dürfen nicht die identische Farbe besitzen.
Dies würde eine Kantendetektion unmöglich machen und die Forderung einer konstanten Streifenbreite verletzen.
- Jede Farbkombination ist einmalig.
Eine Codewortwiederholung innerhalb einer Periode des Musters würde zu unerwünschten Zweideutigkeiten führen.
- Jede Kantenkombination ist einmalig.
Dies ist die endgültige Bedingung zum Ausschluss von Mehrdeutigkeiten trotz verschiedener Farben (siehe Abschnitt 3.2.1).

Gemäß dieser Bedingungen ergibt sich ein Codewortschatz von 104 Codeworten zum Aufbau des farbcodierten Musters. Um daraus sinnvoll und effizient einen Gesamtcode aufzubauen, gehen wie bei der CCT die Codeworte ineinander über. Die Sequenz `ADCBDACB` bestünde also aus den Codeworten `ADCB`, `DCBD`, `CBDA`, `BDAC` und `DACB`. Bei einer binären Kanalinterpretation mit $\overline{\lambda_1 \lambda_2} = 00 \equiv 0$, $\overline{\lambda_1 \lambda_2} = 01 \equiv 1$,

$\lambda_1 \overline{\lambda_2} = 10 \equiv 2$ und $\lambda_1 \lambda_2 = 11 \equiv 3$ und den eben daraus entstehenden Dezimalzahlen $\{0; 1; 2; 3\}$ stellt folgende Zeichenkette eine mögliche Codierung dar:

```
3032101203230121321201231020212321303010210312023121 ...
... 2102320101301323132031302302131030203031010320201312
```

Das daraus resultierende Muster ist in Abbildung 27 mit $\lambda_1 \approx 436nm \equiv \text{blau}$ und $\lambda_2 \approx 546nm \equiv \text{grün}$ dargestellt. Um eine größere Fläche zu beleuchten, bzw. eine optimale Streifenbreite auf dem verwendeten Kamerasensor zu erzeugen, besteht das Muster dabei aus 200 Farbstreifen. Der Code wurde also einmal annähernd komplett wiederholt, was bei der Tiefenberechnung berücksichtigt werden muss, um Zweideutigkeiten der Codierung auszuschließen.

3.3.3 Aufbau einer Projektionseinheit zur infraroten Musterdarstellung

Beim Aufbau der Projektionseinheit gilt es zunächst herauszufinden, welche Art der Beleuchtung im nahen Infrarotbereich sinnvoll ist. Aufgrund ihrer Größe, Leistungsaufnahme und Lichtleistung kommen eigentlich nur Laser- oder Leuchtdioden in Frage – vor allem, da sich mit diesen bereits diskrete Wellenlängen projizieren lassen. Entladungs-, Glüh- oder Leuchtstofflampen, die ebenfalls einen Anteil ihres Lichts im Infraroten abgeben können, müssten hierfür mit speziellen Bandpassfiltern versehen werden. Zudem haben Dioden den Vorteil, dass sich ihre abgestrahlte Leistung durch die Stromzufuhr sehr akkurat kontrollieren und den geforderten Bedingungen anpassen lässt. Dabei führt der hohe Wirkungsgrad von Leuchtdioden dazu, dass kaum elektrische Energie in Wärme umgesetzt wird und verloren geht, sondern beinahe vollständig in Strahlungsenergie umgewandelt wird.

Vergleicht man die Projektion einer Laserdiode mit der einer Leuchtdiode, so lassen sich beim Laser aufgrund der Kohärenz der Lichtstrahlen so genannte Speckle ausmachen, die durch die Struktur einer optisch rauen Oberfläche verursacht werden. Dieses Phänomen macht man sich beispielsweise bei der Speckle-Interferometrie ([Koch et al. 1998], [Meixner et al. 2004]) zunutze, um eben solche Oberflächen räumlich zu vermessen. Für die Projektion von codierten Streifen sind diese Speckle dagegen von Nachteil, da sie für „Ausreißer“ zwischen zwei Lichtstreifen sorgen und somit die notwendige Kantendetektion verfälschen könnten. Deshalb werden für den hier beschriebenen Aufbau Leuchtdioden verwendet, bzw. Cluster von mehreren Leuchtdioden, deren Licht über eine Linse gebündelt wird. Die Wellenlängen der Cluster können dabei grundsätzlich beliebig gewählt werden. Jedoch empfiehlt es sich, aufgrund der Empfindlichkeit handelsüblicher Kamerasensoren, Wellenlängen im tatsächlich nahen Infrarotbereich ($< 900nm$) zu wählen. Die Wellenlängen sollten also möglichst nah am Sichtbaren liegen, dabei aber weit genug auseinander, um unterscheidbar zu sein. Es muss zudem berücksichtigt werden, dass durch den Einsatz vieler LEDs in einem Cluster abhängig vom Betrachtungswinkel die abgestrahlte Wellenlänge fertigungsbedingt um $\pm 20nm$ von der gewünschten Wellenlänge abweichen

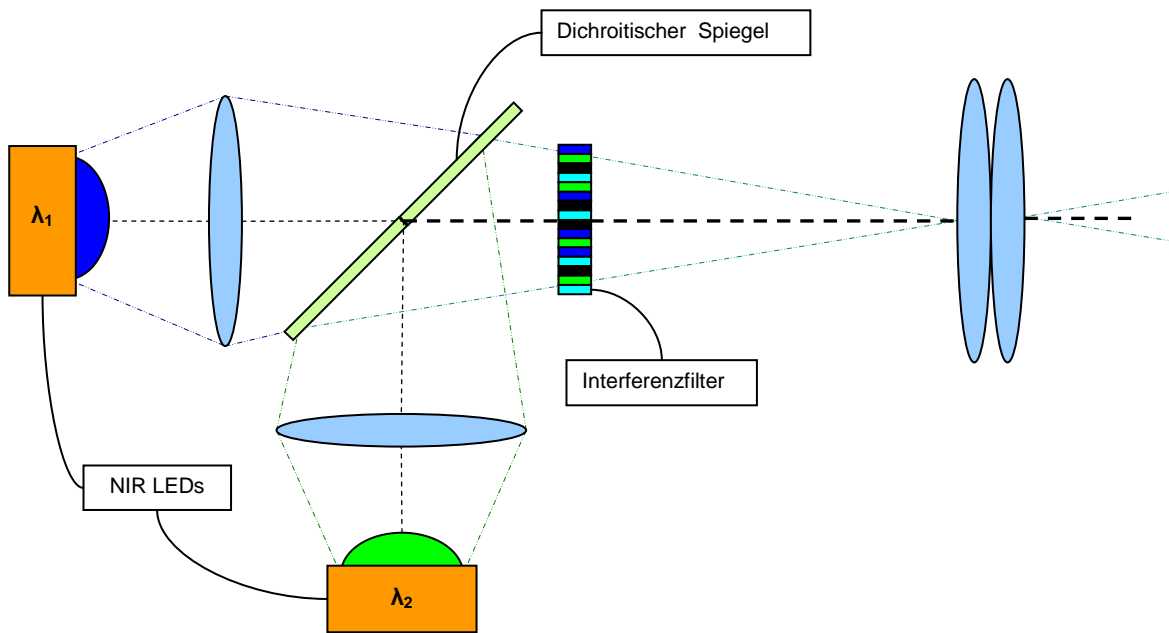


Abbildung 28 Projektionseinheit des Zwei-Wellenlängen-Ansatzes. Die Wellenlängen λ_1 und λ_2 werden über einen dichroitischen Spiegel kombiniert. Unter Verwendung eines Interferenzfilters wird schließlich das codierte Muster in die Szene projiziert.

kann. Berücksichtigt man all diese Parameter, so ist der Einsatz von LED-Clustern mit $\lambda_1 = 760nm$ und $\lambda_2 = 850nm$ möglich, welche im Prototyp getestet und verwendet wurden.

Das Licht dieser LEDs wird über einen dichroitischen Spiegel kombiniert, so dass eine der Wellenlängen nahezu vollständig reflektiert und die andere transmittiert wird. In Abbildung 28 wird dieses Prinzip verdeutlicht. Das Ergebnis nach diesem Spiegel ist ein Licht- oder vielmehr Wellenlängenbündel, das sowohl λ_1 als auch λ_2 enthält. Dieser dichromatische Lichtstrahl durchläuft dann eine Zwischenabbildung auf das codierte Streifenmuster, welches, ähnlich einem Diaprojektor, durchstrahlt und somit projiziert wird.

Für dieses „Dia“ wurde ein Interferenzfilter ausgewählt. Ein Interferenzfilter wird durch mehrere dünne Schichten mit bestimmten optischen Eigenschaften aufgebaut, die auf einer Trägerschicht – meist Glas oder ein Glassubstrat – aufgebracht werden. An den Grenzflächen der Schichten interferieren dabei die reflektierten und transmittierten Anteile eines auftreffenden Lichtstrahls, so dass es zur Auslöschung und Verstärkung bestimmter Wellenlängen am Ausgang des Filters kommen kann. Diese Eigenschaften auf bestimmte Wellenlängen sind abhängig von den aufgetragenen Schichtdicken und vom Einfallswinkel der auftreffenden Lichtstrahlen.

Nach diesem Verfahren werden Filterstreifen definiert, die entweder für keine, eine oder beide Wellenlängen durchlässig sind und dadurch das Projektionsmuster aufbauen. Ein großer Vorteil gegenüber beispielsweise organischen Dias liegt dabei in der Eigenschaft, dass Interferenzfilter in der Regel nur einen geringen Anteil der auftreffenden Strahlung absorbieren und sich deshalb kaum aufheizen, was deren Lebensdauer und geometrische

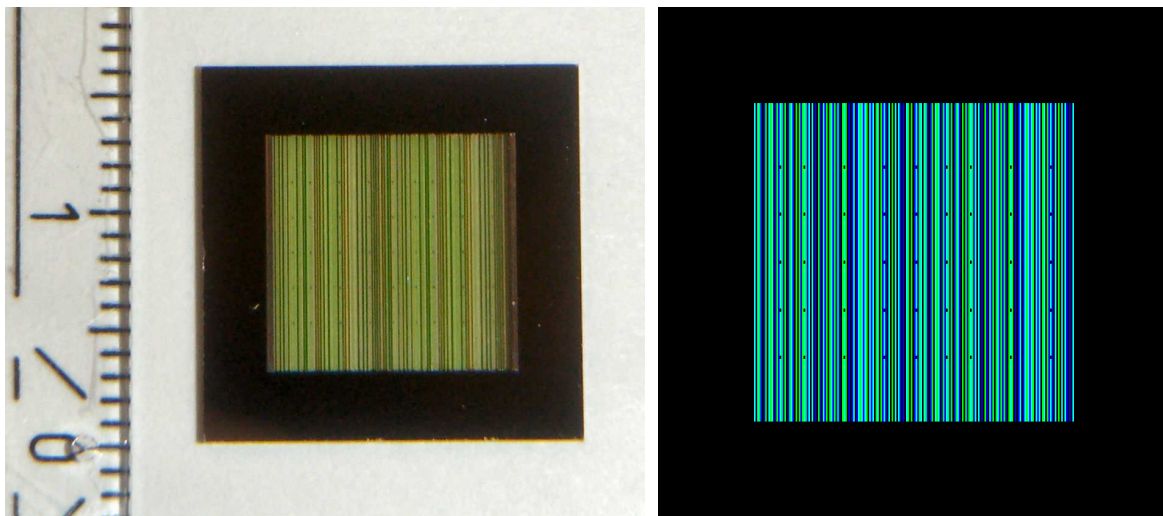


Abbildung 29 Interferenzfilter zur Musterprojektion. Die rechte Seite zeigt die schematische Darstellung mit Kanalinterpretation in Grün und Blau, die linke Seite zeigt vergrößert das real gefertigte Filter.

Stabilität deutlich erhöht. Aufgrund der Projektionseigenschaften, der Kameraparameter und der geplanten Tiefengenauigkeit und Tiefeneindeutigkeit wurde das Filter quadratisch auf einer Fläche von 10mm^2 mit einer Streifenbreite von $50\mu\text{m}$ angelegt und enthält damit das definierte Muster mit 200 Streifen. Abbildung 29 zeigt den schematischen Aufbau dieses Filters sowie das gefertigte Interferenzfilter. In beiden Abbildungen sind zusätzlich zu den Streifen Marken zur Projektorkalibrierung enthalten, auf die unter 3.5.3 näher eingegangen wird. Nachdem der Kombinationslichtstrahl das Filter passiert hat, durchläuft er eine klassische Projektionsoptik und wird somit über ein Objektiv in den Raum projiziert, bzw. auf das zu erfassende Objekt.

3.3.4 Kamera-Seite: Aufbau und Funktion

Bei der Aufnahme der codierten Szene gilt es, zwei Schwierigkeiten zu überwinden – zum einen, einen für die projizierten Wellenlängen empfindlichen Sensor zu verwenden, zum anderen, die einzelnen Wellenlängen zu unterscheiden.

Für die Aufnahme im Infraroten bieten sich grundsätzlich spezifische Infrarotkameras an. Diese sind allerdings, verglichen mit Standardkameras des sichtbaren Spektrums, deutlich teurer und vor allem für einen Einsatz im langwelligeren Infrarotspektrum entworfen. Da die Sensoren von „normalen“ Kameras einen breiteren Wellenlängenbereich als den Sichtbaren abdecken und durchaus im NIR-Bereich – also auch bei den verwendeten Wellenlängen – noch empfindlich sind, ist es möglich, handelsübliche Kameras für die Aufnahme im infraroten Spektrum zu verwenden. Dies wurde in Abbildung 22 mit der Empfindlichkeitskurve eines CCD-Sensors schon aufgezeigt.

Das größere Problem bei der Aufnahme liegt in der Identifikation der einzelnen Wellenlängen. Wie in Abschnitt 2.1.2.3 beschrieben kann dies durch den Einsatz mehrerer Senso-

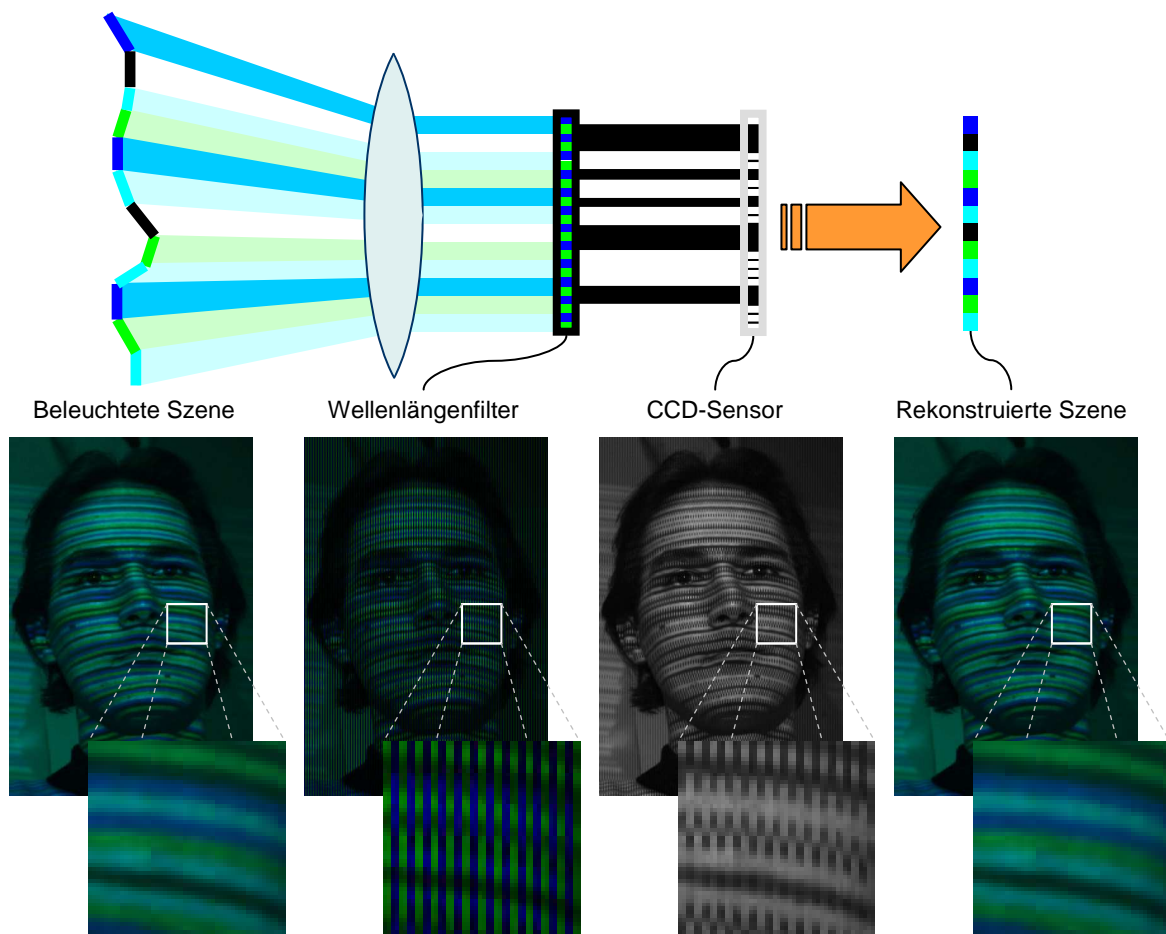


Abbildung 30 Prinzip der 1-Shot Infrarotaufnahme. Die mit dem codierten Muster beleuchtete Szene wird durch den Wellenlängenfilter in für jeweils eine Wellenlänge sensitive Bereiche aufgeteilt, die auf den CCD-Sensor abgebildet werden und in Graustufen erscheinen. Da die Struktur dieses Filters bekannt ist, kann das Ursprungsbild rekonstruiert werden.

ren (3-Chip-Design) geschehen, oder es wird durch die Anwendung eines Bayer-Filters gelöst. Ein ähnliches Vorgehen ist auch für das vorgestellte Verfahren im Infraroten denkbar. Allerdings ist es mit vertretbarem Aufwand nicht möglich, ein entsprechendes Filter pixelgenau und direkt auf die Sensorfläche zu bringen, so wie es bei der industriellen Fertigung von Farbsensoren praktiziert wird.

Dennoch wird ein vergleichbarer Ansatz verwendet, um zwischen den verwendeten Wellenlängen zu unterscheiden. Da das projizierte Muster aus einzelnen Streifen besteht, ist in der aufzunehmenden Szene eine sprunghafte Wellenlängenänderung in Richtung des Streifenverlaufs im Verhältnis zur Richtung des Musters nur in geringen Abstufungen zu erwarten. Betrachtet man also beispielsweise die Muster in Abbildung 29, so verlaufen die Streifen vertikal und das Muster horizontal. Eine wechselnde Schachbrettanordnung zur Betrachtung einzelner Bildpunkte ist für eine Wellenlängenbestimmung demnach nicht zwingend nötig, es genügt in diesem Fall eine zeilenweise oder, je nach Projektions-

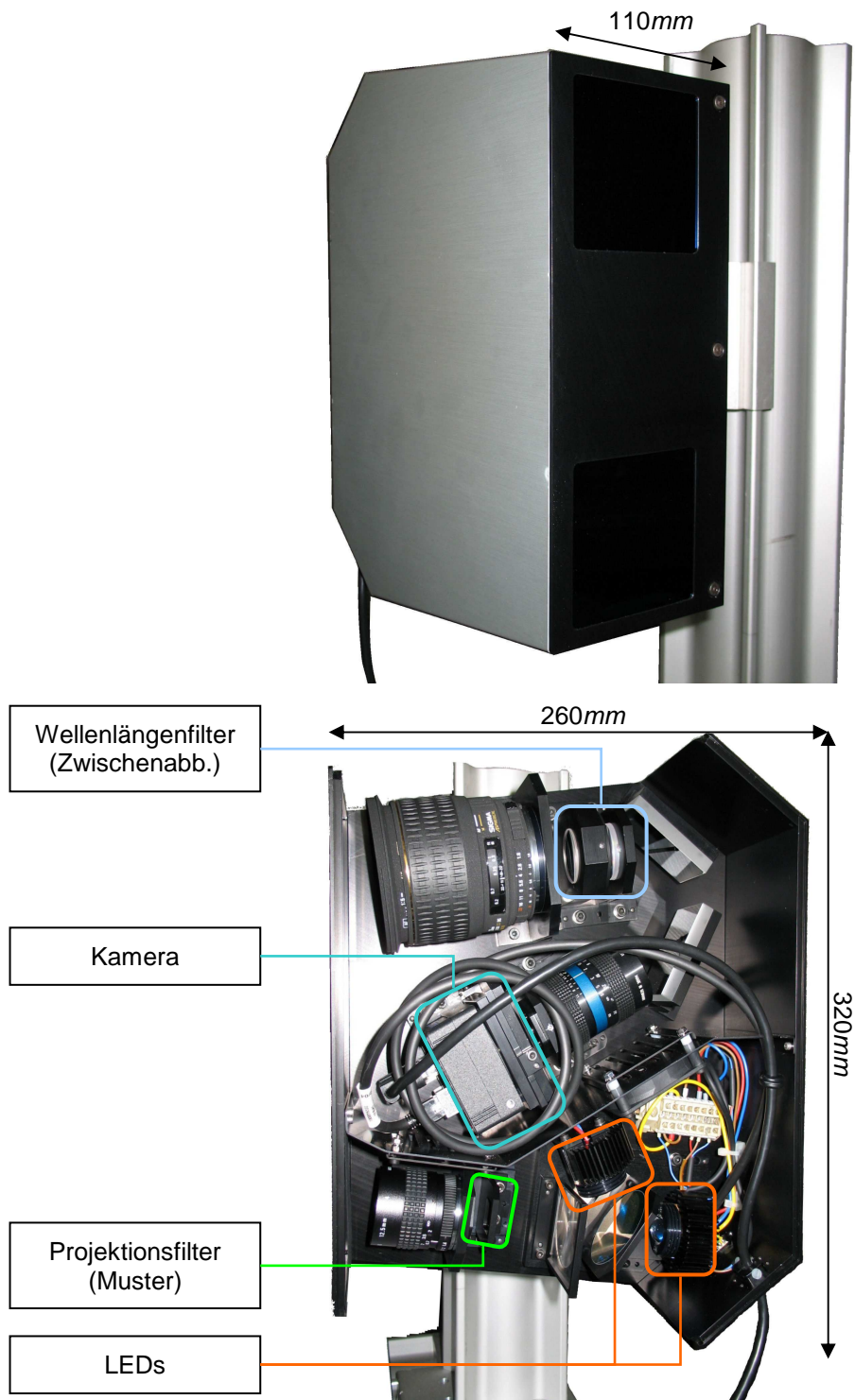


Abbildung 31 Sensor nach dem Zwei-Wellenlängen-Prinzip. Die obere Aufnahme zeigt die Außenansicht des aufgebauten Sensors, die untere Aufnahme erlaubt einen Blick auf die verbauten Komponenten. Das große Objektiv im oberen Teil bildet die Szene auf das Wellenlängenfilter ab, was eine Rekonstruktion im Kamerabild (Mitte) erlaubt. Der untere Teil beinhaltet die beiden LEDs sowie das Projektionsfilter.

richtung, auch spaltenweise Betrachtung des Bildes. Abbildung 30 verdeutlicht dieses Vorgehen. Die aufzunehmende Szene wird hierfür in eine Zwischenebene abgebildet und anschließend durch ein Wellenlängenfilter geführt, dessen Streifen abwechselnd für die einzelnen Wellenlängen durchlässig sind.

Es ist, wie das zur Projektion verwendete Filter, ebenfalls als Interferenzfilter aufgebaut, besteht also aus einzelnen auf ein Glassubstrat aufgetragenen Schichten. Die so gefilterte Zwischenabbildung wird schließlich über eine weitere Optik auf den Kamerasensor abgebildet.

Weil die Wellenlängenfilterung nicht direkt auf dem Sensor geschieht, lässt sich die Wellenlängenprojektion nicht pixelgenau (bzw. zeilen-/spaltengenau) durchführen. Es ist also eine Kalibrierung der Wellenlängenfilterung nötig, deren Ergebnis eine Look-Up-Table mit einer Pixel-Wellenlängen-Zuordnung ist. Mit dieser Information lassen sich letztendlich aus dem infrarot beleuchteten und in Graustufen aufgenommenen Bild eines Objekts die Kanalinformationen herausfinden und entsprechend interpolieren, so dass eine Rekonstruktion beider Wellenlängen möglich ist.

Dabei soll der Sensor Messergebnisse liefern, die dem einer vergleichbaren Kamera-Projektor-Kombination im sichtbaren Licht ähneln. Die effektive Auflösung der Kamera muss deshalb Bilder in der üblichen Video-Auflösung von 780×580 Pixeln liefern können. Durch die Rekonstruktion gehen allerdings de facto Bildinformationen verloren, weshalb für den NIR-Ansatz eine höher auflösende Kamera verwendet werden muss, um die Verluste der Interpolation zu kompensieren. Es kommt deshalb im Versuchsaufbau eine CCD-Kamera mit 1392×1040 Bildpunkten (Sony ICX 285AL) zum Einsatz.

Neben der höheren Auflösung bietet dieser Sensor eine relativ hohe Empfindlichkeit über das sichtbare Spektrum hinaus. Diese liegt bei den verwendeten Wellenlängen von $760nm$ und $850nm$ zwischen 30% und 50% [Sony 2008] der maximalen sichtbaren Empfindlichkeit, was nach einer Bandpassfilterung der nicht benötigten Wellenlängenbereiche für eine Bilderfassung ausreichend ist.

3.3.5 Aufbau und Test des Prototypen

Die bisher beschriebenen Komponenten zur Musterprojektion und Objektaufnahme können nun zu einem aktiven Stereosystem kombiniert werden – eine entsprechende Anordnung ist in Abbildung 31 zu sehen. Der dort abgebildete Sensor ist das Resultat der beschriebenen Entwicklung des aktiven NIR-Sensors unter Verwendung zweier Wellenlängen.

Bei der Kalibrierung und Objektvermessung zeigen sich dabei allerdings Schwierigkeiten. Während sich der Ansatz im Theoretischen relativ einfach beschreiben lässt und auch Simulationen im Sichtbaren mit einer Kombination aus Video-Projektor und Kamera sehr gute Ergebnisse liefern, die der Leistungsfähigkeit einer „regulären“ farbcodierten Trian-

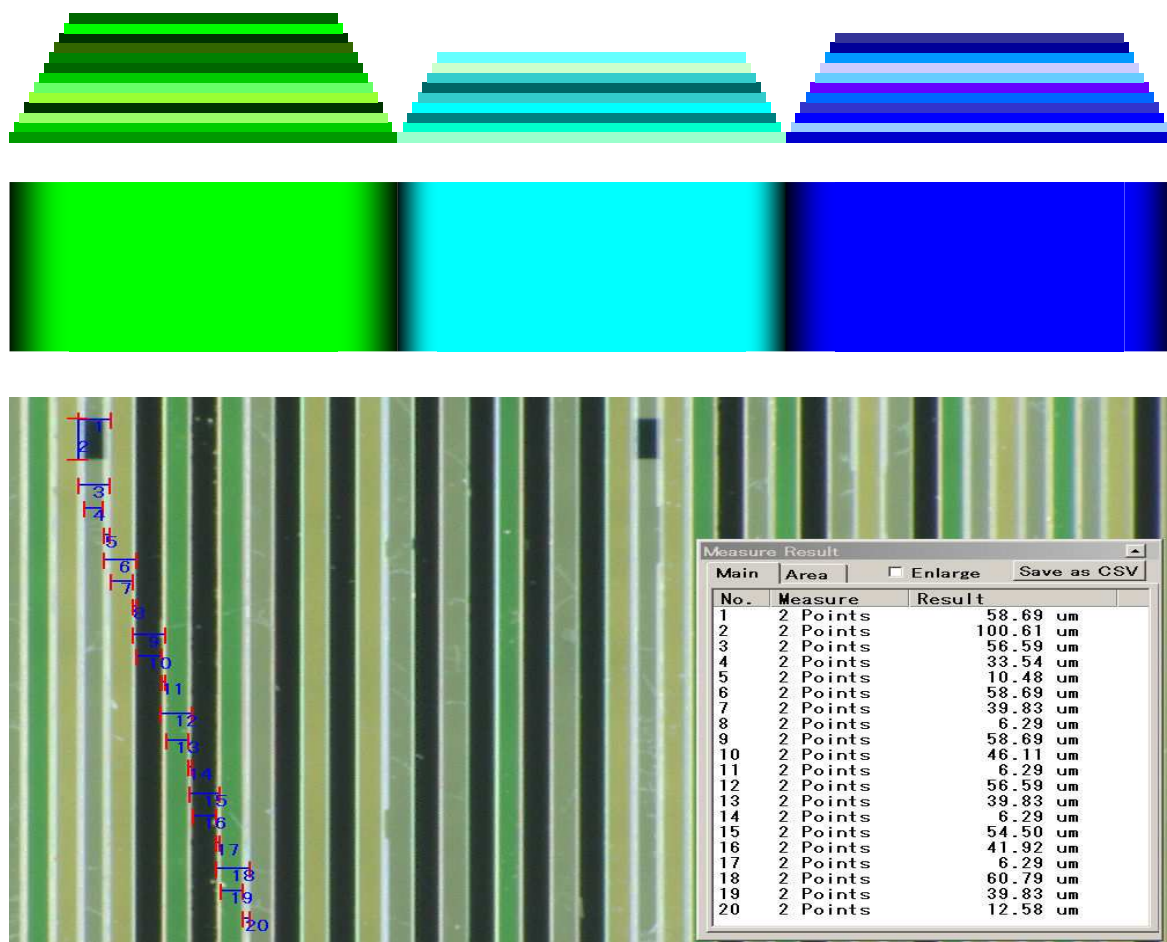


Abbildung 32 Aufbau eines Interferenzfilters. Der obere Teil zeigt schematisch den Aufbau der Filterstreifen in mehreren Schichten, der untere Teil die Aufnahme und Vermessung des realen Filters unter dem Mikroskop. Deutlich sichtbar sind unbeschichtete Bereiche verschiedener, unregelmäßiger Breite zwischen den einzelnen Streifen, sowie daraus resultierende, verschiedene Streifenbreiten des Codes.

gulation entsprechen, unterscheiden sich die Aufnahmen des tatsächlichen NIR-Aufbaus von diesen Ergebnissen. Als problematisch erweist sich dabei nicht das Ausrichten und Anordnen der Optiken und Filter, um die Projektion und die notwendige Zwischenabbildung zu bewerkstelligen. Vielmehr offenbart das Fertigungsverfahren der Interferenzfilter Schwierigkeiten. Wie weiter oben unter 3.3.3 beschrieben besteht ein Interferenzfilter aus mehreren dünnen Schichten mit verschiedenen optischen Eigenschaften, deren Kombination im Ergebnis eine gezielte Wellenlängenfilterung erlaubt. Werden diese Schichten aber als schmale Streifen aufgetragen, wie für die vorgeschlagene Mustererzeugung, so ergibt sich zwischen den einzelnen Filterstreifen eine undefinierte Treppenstruktur (siehe Abbildung 32). Dort, wo eigentlich eine exakte und gerade Kante verlaufen sollte, gehen die einzelnen Schichten stufenweise ineinander über.

Dieser Effekt kann, abhängig von der Oberflächenbeschaffenheit und Oberflächenstruktur des aufzunehmenden Objekts, „Geisterstreifen“ hervorrufen, zu Fehlinterpretationen des

projizierten Musters führen und somit eine fehlerhaften Decodierung bewirken. Bis zu einem gewissen Grad lässt sich dies korrigieren, vor allem, wenn man Kontextwissen über das aufzunehmende Objekt in der Tiefenberechnung berücksichtigt. Dennoch muss man zusammenfassen, dass das vorgestellte Verfahren der 3D-Objekterfassung mittels eines 1-Bild-Ansatzes unter Verwendung zweier Wellenlängen aufgrund des aktuellen Stands der Fertigungstechnik besser zur Aufnahme technischer und in ihrer Beschaffenheit homogener Oberflächen geeignet ist, als zur Erfassung menschlicher Gesichter oder beliebiger, unbekannter Objekte und Szenen. Auf Basis dieser Erkenntnisse wird im folgenden Kapitel die Entwicklung eines NIR-Sensors beschrieben, der die Aufnahme hochauflösender Tiefenkarten für allgemeine Anwendungen ermöglicht.

3.4 Aktive NIR-Stereoskopie unter Verwendung einer Wellenlänge

Der eben beschriebene Ansatz zur unsichtbaren dreidimensionalen Objekterfassung enthält zwei Punkte, die eine einfache Implementierung und eine exakte Szenenrekonstruktion erschweren. Diese sind auf der einen Seite die aufwändige Optik, die zur Erkennung und Interpolation der unterschiedlichen Farbkanäle notwendig ist, sowie der zugehörige Filter. Auf der anderen Seite erschwert das Herstellungsverfahren der Interferenzfilter, welches keinen exakten Kantenverlauf eines Streifens garantiert, eine genaue und sichere Musterrekonstruktion.

Um diese möglichen Fehlerquellen auszuschließen und die erwünschte Leistung eines im sichtbaren Licht arbeitenden Systems mit einer Kombination aus Video-Projektor und Kamera zu erzielen, wurde ein zweiter Ansatz entwickelt, der eine dreidimensionale Szenenrekonstruktion unter Verwendung von nur einer Wellenlänge ermöglicht [Modrow et al. 2008a].

3.4.1 Erläuterung des Prinzips

Der Grundgedanke des in 3.3.2 beschriebenen Prinzips der farbcodierten Triangulation mittels eines zweikanaligen Musters bleibt dabei für diesen Ansatz weitgehend erhalten. Allerdings werden die Kanäle nicht simultan unter Verwendung zweier Wellenlängen projiziert, sondern zeitlich versetzt hintereinander. Daraus ergeben sich einige Vorteile.

Zunächst einmal erlaubt es den Einsatz von lediglich einer Wellenlänge zur Projektion, welches den Aufwand der Kanalkombination und Wellenlängenfilterung reduziert. Es müssen lediglich einfache Graustufenbilder aufgenommen werden, was auf Seiten der Kamera den Umfang der Optik deutlich mindert und eine Bildrekonstruktion durch Interpolation unnötig macht. Dieses Verfahren baut somit zwar auf einer Farbcodierung eines Musters auf, die jedoch als zeitlich versetzte Projektion umgesetzt wird – weshalb man sozusagen von einer „timed-color-coded“ Triangulation sprechen könnte. Da die Aufnah-

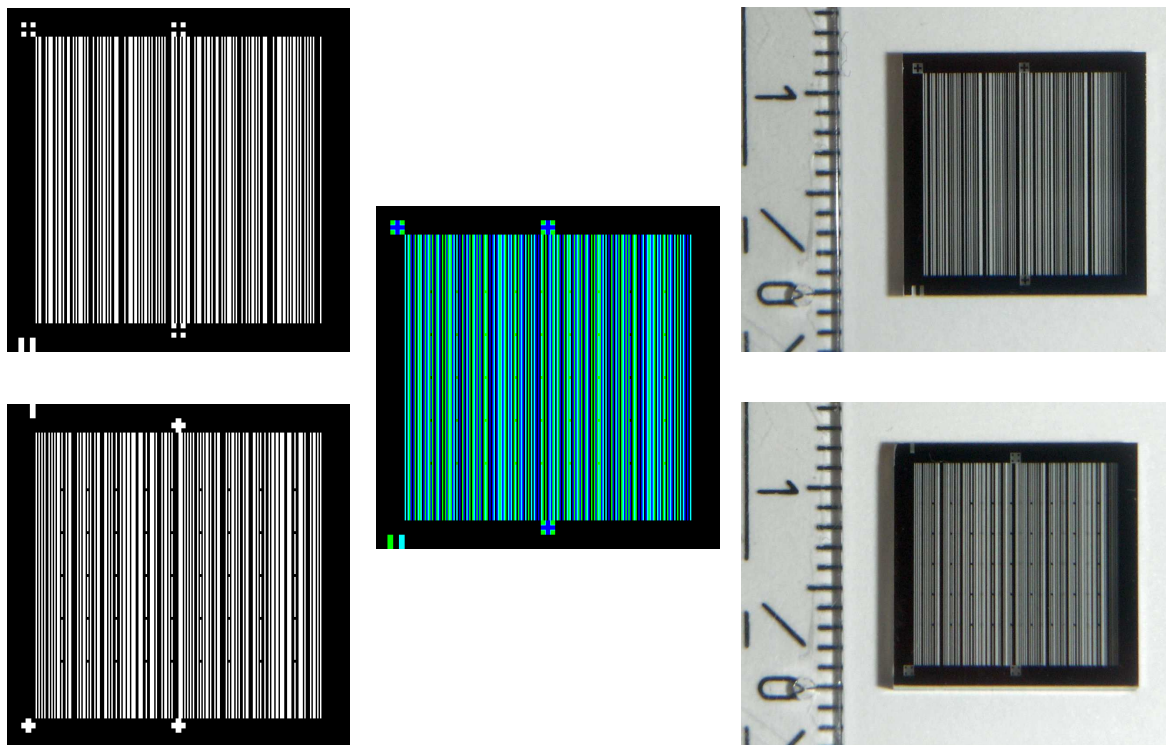


Abbildung 33 Chrommasken des „timed-color-coded“-Verfahrens. Auf der linken Seite sind die Schemadarstellungen der einzelnen Teilmuster zu sehen, deren Kombination das codierte Streifenmuster ergibt (Mitte). Rechts dargestellt sind die dementsprechend gefertigten Chrommasken.

me der benötigten Bilder sehr schnell hintereinander erfolgen kann, bleiben die Vorteile bisheriger Ein-Bild-Ansätze weitgehend erhalten, so dass keine Notwendigkeit einer statischen Szene besteht.

Das in 3.3.2 beschriebene Muster dient dabei wieder der Codierung einer Szene, allerdings wird es in seine beiden Basiskanäle aufgeteilt. Erst in Kombination lässt sich das Projektionsmuster decodieren und die Triangulation ausführen (siehe Abbildung 33). Um diese Musterdarstellung zu ermöglichen, muss wieder eine geeignete Möglichkeit der Projektion gefunden werden.

3.4.2 Projektions-Seite: Aufbau und Funktion

Die Projektionseinheit des 2-Bild-Ansatzes muss es ermöglichen, zwei Muster derart zu projizieren, dass sie auf dem Aufnahmeobjekt deckungsgleich erscheinen. Das bedeutet also, dass die Muster über dieselbe Optik und denselben Strahlengang des Projektors abgestrahlt werden sollten. Außerdem muss der Projektor eine schnelle sequentielle und mit der Kamera synchronisierte Projektion der einzelnen Muster zulassen. Diese Aufgabe wäre mit handelsüblichen Video-Projektoren zumindest teilweise lösbar – allerdings müssten diese für eine Anwendung im Infraroten umgebaut oder ein eigener Projektor entworfen werden. Unter Verwendung eines reflektiven dynamischen Panels wäre dies rein technisch

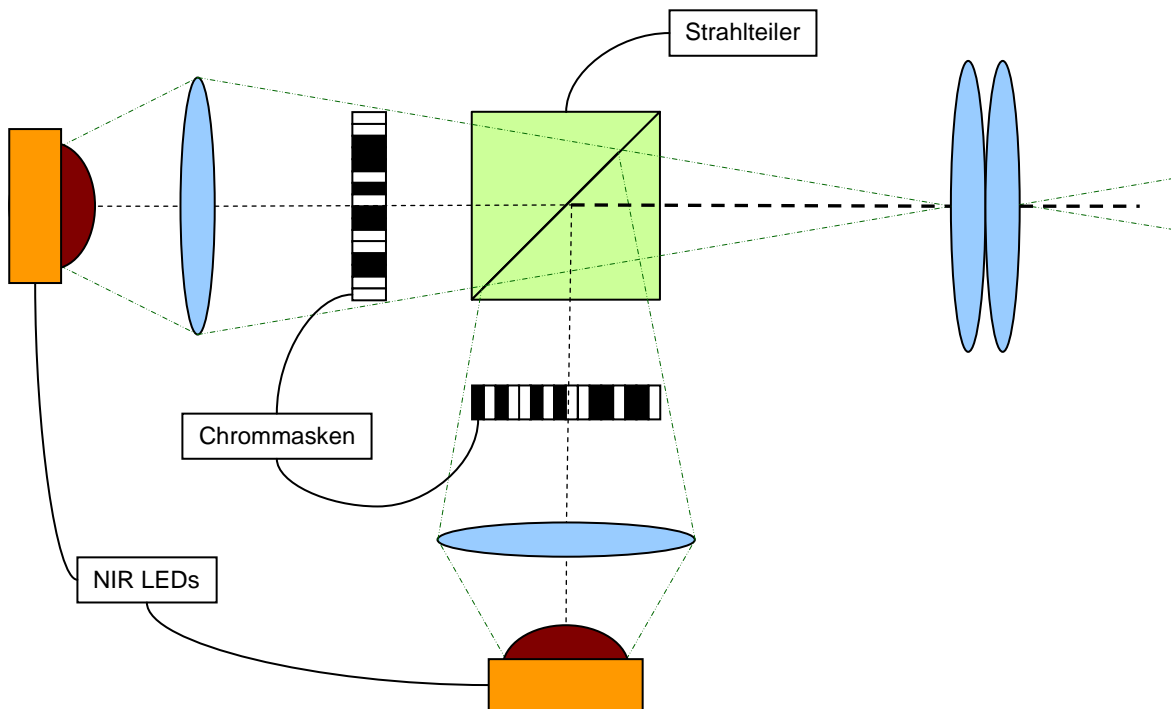


Abbildung 34 Projektionseinheit des Ein-Wellenlängen-Ansatzes. Durch LEDs werden die Chrommasken abgebildet, die die einzelnen Teile des Musters enthalten. Diese beiden optischen Pfade werden über einen Strahlteiler kombiniert und durch eine gemeinsame Optik projiziert.

durchaus realisierbar, würde aber neben einer aufwändigen Konstruktion auch einen erheblichen Kostenaufwand für eine einfache Projektionsaufgabe bedeuten. Deswegen wird, wie bereits in Kapitel 3.3, auf die Möglichkeit statischer Filter zurückgegriffen. Abbildung 34 zeigt schematisch den Aufbau der Projektionseinheit.

Die beiden einzelnen Projektionskanäle sind dabei in ihrem Grundaufbau identisch. Als Lichtquelle dient jeweils ein LED-Array, das über eine Linse das Filter mit dem Teil-Code durchstrahlt. Mittels eines Strahlteilers werden die einzelnen Kanäle dann zusammengeführt und über eine gemeinsame Optik projiziert. Der erste Unterschied zum 1-Bild-Ansatz liegt hierbei in der verwendeten Wellenlänge der LEDs, die nun einheitlich 735nm beträgt. Diese Wellenlänge erlaubt es, empfindlichere Bereiche der Kameras anzusprechen, da sie näher am sichtbaren Spektrum liegt, jedoch weit genug im Infraroten, um nicht vom menschlichen Auge als störend oder blendend empfunden zu werden. Den zweiten Unterschied liefert der Einsatz eines kubischen Strahlteilers anstelle des dichroitischen Spiegels. Ein speziell beschichteter Spiegel ist für dieses Verfahren nicht mehr vonnöten, da auch bei der Projektion nicht mehr zwischen Wellenlängen unterschieden werden muss. Der Zweck des Strahlteilers ist also nunmehr die gleichmäßige Kombination der beiden Projektionsmuster. Vorteilhaft erweist es sich hier, dass ein gleichmäßig durchlässiger (50/50) Strahlteiler zu den optischen Standardkomponenten zählt und keine Spezialanfertigung ist. Zudem verhindert die kompakte kubische Bauweise eventuelle Verzer-

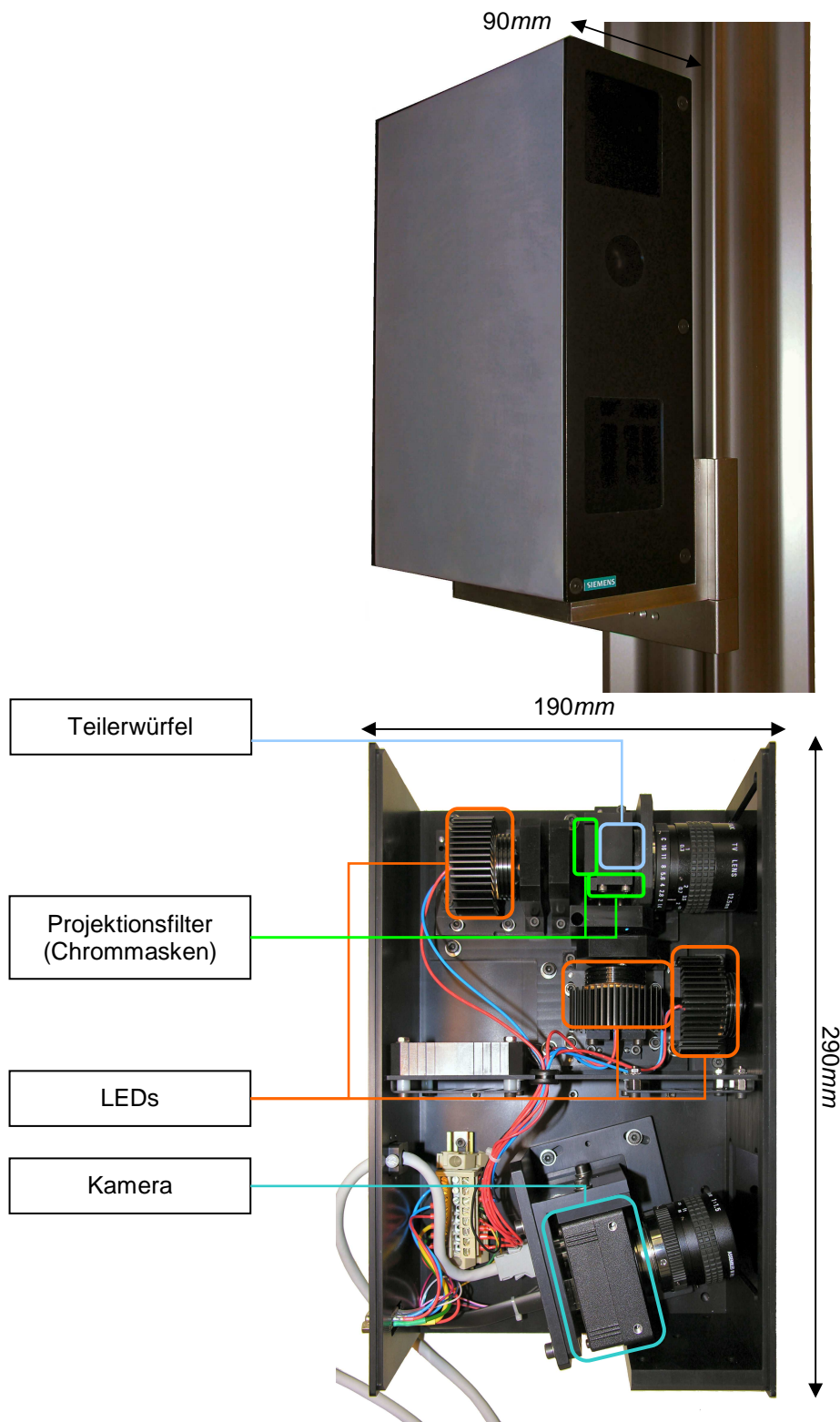


Abbildung 35 Sensor nach dem Ein-Wellenlängen-Ansatz. Das obere Bild zeigt die Außenansicht des Sensors, das untere den Innenaufbau. Im oberen Teil des Sensors befindet sich die Projektionseinheit mit den LEDs, in deren Projektionsrichtung sich die Chrommasken befinden. Vor dem Objektiv liegt der Teilerwürfel, der die beiden Pfade kombiniert.

rungen der Abbildung, wie sie nach der Montage eines dünnen Spiegels durch Materialspannungen auftreten können.

Ein weiterer wichtiger Punkt ist die Wahl der Filter zur Musterprojektion – vor allem, um bekannte fertigungstechnische Schwierigkeiten zu vermeiden. In diesem Fall der Projektion sind keine Interferenzfilter von Nöten, da nun spezifische Wellenlängen nicht mehr gemeinsam gefiltert werden müssen. Es genügt hier, lichtdurchlässige und undurchlässige Schichten bestimmter Breite auf einem Filter zu schaffen. Für diesen Zweck bietet sich der Einsatz einer Chrombeschichtung an, wie sie normalerweise zur Fertigung von Belichtungsmasken verwendet wird. Dabei ist es ausreichend, das Chrom als einzelne Schicht auf ein Glassubstrat aufzubringen, da eine solche Schicht vollständig lichtundurchlässig ist. Daraus ergeben sich, neben einem einfacheren Herstellungsverfahren, weitere Vorteile gegenüber einem Interferenzfilter. Denn durch das Wegfallen zusätzlich benötigter Schichten bleiben der beobachtete Treppeneffekt und die daraus resultierenden „Geisterstreifen“ aus. Das Resultat sind also Streifen exakter Breite mit einem geraden Kantenverlauf, wie sie von der anschließenden Bildverarbeitung benötigt werden.

Verglichen mit einem einzelnen Mehrkanalfilter oder einem dynamischen Mustergenerator bleibt ein kleiner Nachteil jedoch bestehen. Die deckungsgleiche Projektion in einem gemeinsamen Strahlengang über eine einzelne Optik wird durch eine Kombination der Kanäle im Strahlteiler erreicht, was ein exaktes Ausrichten der Filtermasken erfordert. Dies ist jedoch ein eher mechanisches Problem, das durch eine entsprechende Konstruktion der Filterhalterungen gelöst werden kann. Dadurch ist ein Ausrichten der Filter auf wenige μm genau möglich, wodurch eine konforme Projektion sichergestellt werden kann.

3.4.3 Aufbau und technische Daten

Mit der Projektionseinheit ist der wichtigste Teil des Sensors vorhanden, da für die Kamera zur Bildaufnahme keine spezielle Filterung benötigt wird. Es kann also eine normale CCD- oder CMOS-Kamera verwendet werden, die für eine Wellenlänge von 735nm ausreichend empfindlich ist. Die Verschiebung der benutzten Wellenlänge näher an das sichtbare Spektrum bringt dabei zusätzlich Vorteile zum 2-Wellenlängen-Verfahren, da man sich damit weiter in Richtung des Normalbereichs der Standard-Sensoren bewegt. Außerdem kann die Bildauflösung wieder auf 780×580 Pixel (die erwünschte Video-Auflösung) reduziert werden, da keine Wellenlängendetektion stattfindet und somit ein Bild bereits die kompletten Intensitätswerte eines Kanals enthält.

Für diesen Ansatz wurde zunächst eine Kamera mit einem CCD-Sensor (Sony ICX 415AL) gewählt, deren relative spektrale Empfindlichkeit bei 735nm etwa bei 50% liegt. Weitere Kameras wurden im Verlauf dieser Arbeit ebenfalls auf ihre Tauglichkeit im nahen Infrarotbereich getestet, worauf in 3.6.5 noch eingegangen wird. Die Kamera muss nun so positioniert werden, dass sie die codierte Szene unter einem geeigneten Triangula-

tionswinkel erfassen kann. Es ergibt sich daraus der in Abbildung 35 dargestellte Aufbau für den 2-Bild-Ansatz unter Verwendung einer Wellenlänge.

Die technischen Daten lassen sich wie folgt beschreiben:

- Arbeitsbereich (Entfernung zum Sensor): ca. 60cm
- Arbeitsraum: $50 \times 50 \times 50\text{cm}^3$
- Auflösung (Anzahl möglicher Messwerte): 200×782 (Kanten \times vert. Auflösung)
- Tiefengenauigkeit: $< 0,3\text{mm}$
- Messgeschwindigkeit: $\approx 10 - 20\text{fps}$

Diese Daten lassen die Schlussfolgerung zu, dass der beschriebene Ansatz bestens für Anwendungen zur dreidimensionalen Erfassung menschlicher Gesichter geeignet ist. Trotz der sequentiellen Bildaufnahme liegt die Erfassungsgeschwindigkeit beinahe im Bereich normaler Videoaufnahmen, so dass auch Bewegungen eines Objekts in einem gewissen Rahmen zulässig sind. Zudem erlaubt die filterlose Kamera eine zusätzliche Texturaufnahme bei uncodierter Flächenbeleuchtung, so dass neben den Tiefeninformationen auch die Oberflächenstruktur mit aufgenommen werden kann. Dies erlaubt das Erstellen eines vollständigen Modells, wie es beispielsweise für Sicherheitssysteme oder die Unterhaltungsindustrie (z.B. Avatare, virtuelle Umgebungen) von Nutzen ist. Abbildung 36 zeigt verschiedene Modelle menschlicher Gesichter nach diesem Aufnahmeprinzip. Die Personen wurden dabei frontal aufgenommen, die Darstellung zeigt das reine 3D-Modell sowie das Modell mit zusätzlicher Textur.

Weitere Daten zur Fehleranfälligkeit, Messungenauigkeit und Robustheit des aufgebauten Sensors werden in 3.6 behandelt und mit entsprechenden Verfahren im sichtbaren Licht verglichen.



Abbildung 36 Beispielaufnahmen menschlicher Gesichter. Die Darstellungen zeigen die Ergebnisse einer einzelnen Frontalaufnahme der Personen, aus denen ein komplettes Modell des Gesichts resultiert. Durch eine zusätzliche Graustufenaufnahme lassen sich die Texturen mit den 3D-Modellen verbinden.

3.5 Erweiterung und Ergänzung bisheriger Projektionsverfahren

Wie zu Beginn dieses Kapitels beschrieben, führt eine Reduktion der für ein Muster zur Verfügung stehenden projizierten Kanäle zwangsläufig zu einer Verringerung des Wortschatzes, der zur Codierung verwendet werden kann. Um dennoch eine Szene großflächig, in hoher Auflösung und vor allem eindeutig zu codieren, ist es deshalb nötig, die Rekonstruktionsalgorithmik für die speziellen Verfahren anzupassen. Hierfür bieten sich verschiedene Ansätze an, die nun beschrieben werden. Da zudem eine statische Musterprojektion keine separate Projektion eines Kalibrierungsmusters für die Projektionseinheit erlaubt, muss auch hierfür ein Verfahren gefunden werden, das im Folgenden ebenfalls erläutert wird.

3.5.1 A-priori-Wissen über das aufzunehmende Objekt

Generell sollte ein Sensorsystem zur Tiefen- und Entfernungsmessung in der Lage sein, beliebige Objekte und Formen verarbeiten zu können. In den meisten Anwendungsfällen lassen sich aber bereits im Vorfeld gewisse Aussagen über die Beschaffenheit und Art des aufzunehmenden Gegenstandes treffen. Lediglich bei einer „blinden“ Betrachtung völlig unbekannter Objekte kann man keine Gemeinsamkeiten, bzw. Kontinuitäten über die Aufnahmen formulieren – doch dies ist in den meisten Einsatzszenarien nicht der Fall. Deshalb ist es möglich, sich das Wissen über ein Aufnahme-Objekt a-priori zur Unterstützung und Erweiterung des Algorithmus der Datenrekonstruktion zunutze zu machen.

Wie in den vorigen Kapiteln beschrieben beruhen aktive Stereoverfahren üblicherweise auf einer Szenenrekonstruktion mittels einer codierten Beleuchtung. Seien dies nun binär oder farblich codierte Verfahren, im Endeffekt erfolgt in der Bildverarbeitung eine Decodierung des aufgenommenen Musters. Genau an diesem Punkt sind solche Systeme relativ anfällig, da die Bildaufnahme durch viele Faktoren beeinflusst werden kann.

So haben neben der reinen Bildschärfe, die durch Objektive und Linsen definiert wird, auch die Umgebungsbeleuchtung, die Oberflächenstruktur des Objekts, dessen Farbe und Textur Einfluss auf die Qualität des aufgenommenen Bildes. Interagieren diese objektspezifischen Eigenschaften mit dem Projektionsmuster, so kann dies Fehler in der Decodierung hervorrufen, die wiederum in fehlerhaften Entfernungswerten resultieren können. Um nun diese mögliche Fehlinterpretation zu erkennen, könnte man fehlerredundante Codes einsetzen, was aber aus den bereits genannten Gründen eines ohnehin schon reduzierten Wortschatzes nicht umsetzbar ist. Dennoch lassen sie sich ausgleichen, indem man das Vorwissen über das aufzunehmende Objekt nutzt. Betrachtet man beispielsweise als Messaufgabe eine industrielle Fertigung, so enthält diese im Allgemeinen definierte Objekte mit bekannten Ebenen, Kanten und Flächenverläufen. Aus diesen bekannten Informationen lassen sich dann bereits mit unvollständigen Datensätzen die gewünschten

Messwerte extrahieren. Andererseits lassen sich unter Zuhilfenahme von gemessenen Eckpunkten in Kombination mit vorhandenen Modellen aufgenommene Körper rekonstruieren.

Auch in nicht-industriellen Anwendungen eröffnet ein solches Vorgehen viele Möglichkeiten. Möchte man zum Beispiel Gesichter für virtuelle Avatare aufnehmen, so kann man fundamentale Erkenntnisse über die Physiologie der menschlichen Gesichter nutzen ([Lengagne et al. 2000], [D'Apuzzo 2002], [Enciso et al. 2003]). Diese erlauben etwa den Aufbau eines initialen Gesichtsmodells, das als Basis für weitere Aufnahmen verwendet werden kann. Sind nun neue Aufnahmen fehlerhaft, lassen sich diese Fehler anhand der Modellierung erkennen und korrigieren. Dadurch ist es möglich, auch bei augenscheinlich unregelmäßigen Objekten mit vorhandenen Eckdaten aus einer Messung komplette Modelle zu rekonstruieren.

3.5.2 Datenrekonstruktion und Fehlerkorrektur

Neben dem Ausnutzen des a-priori-Wissens über ein Aufnahmeobjekt lassen sich auch gezielt Fehler in der Detektion des Codemusters erkennen und somit die Datenrekonstruktion korrigieren. Obwohl das codierte Muster im vorliegenden Fall keine Redundanzen besitzt, die beispielsweise falsche Farbdetektionen oder nicht vorhandene Codeworte relativ einfach erkennen lassen, kann eine Fehlerkorrektur auf Basis der Nachbarschaftsbeziehungen, unter Annahme gewisser kontinuierlicher Oberflächenverläufe, vollzogen werden.

Die einfachste Lösung wäre in diesem Fall eine Interpolation von Daten durch das simple Auffüllen von Löchern in der aus der Messung resultierenden Tiefenkarte. Da solche Interpolationen aufgrund der Verarbeitungsgeschwindigkeit üblicherweise linear ausgeführt werden, stellen sie für ebene Oberflächen auch ein adäquates Mittel dar. Beliebige Objekte dagegen können durch dieses Vorgehen fehlerhaft abgebildet werden. Zudem muss das Auffüllen von tatsächlichen Löchern verhindert werden und durch die Formulierung von Randbedingungen Regeln geschaffen werden, die Interpolationen über die geometrischen Grenzen eines Objekts hinaus verhindern.

Sinnvoller ist es, bereits in einem früheren Schritt der Datenrekonstruktion Fehler zu detektieren und eventuell daraus entstehende Lücken zu vermeiden. Denn üblicherweise basieren Löcher oder Sprünge in der resultierenden Tiefenkarte nicht auf Berechnungsfehlern, sondern sind eine Folge fehlerhafter Interpretationen der Bildverarbeitung oder Decodierung. Um diese Fehler zu erkennen, wird deshalb bereits das Muster der detektierten und decodierten Kanten genau analysiert. Da die Reihenfolge der Kantennummern des Codemusters bekannt ist und monoton über die gesamte Szene ansteigt, können von vornherein Fehldetektionen anhand unerwarteter Kantenpositionen ausgeschlossen oder korrigiert werden.

Anhand des Verlaufs der Kanten lassen sich zudem bereits vor der Tiefenberechnung Lücken im Ergebnis vermeiden und außerdem eine homogene Oberflächenstruktur erreichen.

Durch Verfolgung einer Kante entlang ihrer gesamten Positionen innerhalb des Bildes können dabei zum einen fehlerhafte Code-Detektionen korrigiert werden. Zum anderen können die Positionen der einzelnen Punkte der Kante analysiert und gegebenenfalls geglättet werden, sollten sie unerwartet stark von ihren benachbarten Punkten abweichen. Natürlich ist bei allen beschriebenen Korrekturen stets die Beschaffenheit des aufzunehmenden Objekts zu berücksichtigen. Ansonsten können durch Einstellung zu „weicher“ Parameter und Toleranzen Feinheiten in der Tiefenberechnung verloren gehen.

3.5.3 Kalibrierung des Systems

Einer der wichtigsten Punkte überhaupt bei einem Messsystem ist dessen Kalibrierung, da auf dieser die gesamte Datenerfassung beruht und sich Fehler deshalb über alle Messungen fortpflanzen können. Die Kalibrierung einer Kamera wurde bereits in 2.1.3 beschrieben – und wird in dieser Form auch für die Infrarot-Sensoren verwendet.

Bei der Kalibrierung der Projektionseinheiten aktiver Stereosysteme ist die übliche Vorgehensweise, mit den auf die Kalibriertafeln aufgebrachten Marken vergleichbare Muster auf eben diese Tafeln zu projizieren. Daraus werden dann die intrinsischen Parameter des Projektors, sowie dessen Lage und Position im Raum bestimmt. Dies ist möglich, da nach einer Kamerakalibrierung die Position der Kalibriertafel exakt definiert ist und als „Ursprungsebene“ für eine Projektion angesehen werden kann. Auf Basis dieser Ebene werden dann die Koordinaten der projizierten Marken des Musters des Projektors bestimmt. Da deren geometrische Parameter bekannt sind, kann eine Kalibrierung unter Verwendung eines üblichen Kameramodells vollzogen werden – denn optisch unterscheidet sich der Vorgang einer Projektion nicht von dem einer Bildaufnahme. In beiden Fällen findet eine Abbildung von einer zweidimensionalen, durch Bildpunkte bestimmten Fläche über ein Linsensystem statt.

Für die Infrarotsensoren ergibt sich hierbei die Schwierigkeit, dass deren Projektoren mit statischen Mustern arbeiten und deshalb keine beliebigen Projektionsmuster darstellen können. Zunächst ein einzelnes Kalibriermuster zu montieren stellt keine Lösung dar, da erst nach der Kalibrierung das eigentliche Codemuster eingesetzt werden kann. Durch diese mechanische Beeinträchtigung, aber auch durch eventuelle Materialunterschiede oder Fertigungstoleranzen zwischen Code- und Kalibriermuster, könnten die Ergebnisse verfälscht werden. Aus diesen Gründen sind die Marken zur Kalibrierung bereits in die Codemuster integriert worden, wie es in Abbildung 29 und Abbildung 33 erkennbar ist. Diese befinden sich in jeweils einem Streifen des Musters und beeinträchtigen die eigentliche Tiefenberechnung dadurch nicht. Somit ist durch horizontale und vertikale Kantendetektion die Bestimmung der Koordinaten der Kalibriermarken möglich. Die Nummerierung und Unterscheidung der Marken kann durch Decodierung des Codemusters sichergestellt werden. Denn dadurch wird jeder Streifen des Projektionsmusters eindeutig bestimmt, womit auch dessen Lage und die der eingebetteten Kalibriermarken innerhalb des Musters bekannt sind und eine erste Unterscheidung erlauben. Die endgültige Position einer Marke

ergibt sich schließlich aus deren Lage innerhalb des Code-Streifens. Mit diesen Informationen schließlich kann die Kalibrierung des Projektors durchgeführt werden.

3.6 Evaluierung der aktiven NIR-Stereoskopie

Ein neu entwickelter Sensor ist nur dann als Fortschritt zu betrachten, wenn er sich in seiner Leistungsfähigkeit und Effizienz von bisherigen Verfahren oder Techniken absetzen kann. Was die beschriebene aktive Stereoskopie im Infraroten betrifft, so bietet sie aufgrund ihrer Technik vor allem einen Vorteil in Bezug auf die Benutzerfreundlichkeit und Akzeptanz für den täglichen Einsatz (siehe hierzu auch [BioSec 2005]). Wie sich dabei die entwickelten Verfahren auf die Messgenauigkeit, die Anfälligkeit für äußere Störfaktoren und weitere Einflüsse auswirken, wird nun in den folgenden Punkten dargestellt.

3.6.1 Messfehler eines Stereosystems

Die Genauigkeit eines Messsystems nimmt bei einer Evaluierung wohl die maßgebende Stellung ein. Im Folgenden werden deshalb zunächst anhand eines allgemeinen Ansatzes der Stereographie die Ursachen für Messfehler analysiert, sowie entsprechende Formeln zur Berechnung und Abschätzung der Messgenauigkeit hergeleitet.

Dabei wird ein parametrisiertes Modell eines Stereosystems gemäß Abbildung 37 in einem XYZ-System verwendet. Als Zentrum dieses Modells wird das optische Zentrum der (linken) Kamera festgelegt. Zudem gilt die Annahme, dass die y -Achsen beider Sensoren parallel verlaufen, was für ein derartiges System die Vorgabe der maximal möglichen Sensorflächen-Nutzung darstellt und in der Praxis auch mit hinreichender Präzision umgesetzt werden kann. Eine Abweichung von dieser Annahme ist also vernachlässigbar. Zur Modellierung der Abbildungseigenschaften wird das Lochkammermodell als hinreichend genaue Annäherung einer Linsenkamera verwendet, da entsprechende Abweichungen durch eine Kalibrierung oder weitere Berechnungsschritte angeglichen werden können. Schließlich wird für beide Sensoren – ob Kamera oder Projektor – die gleiche Brennweite f angenommen, was zwar in der Praxis nicht realistisch ist, aber durch eine entsprechende Skalierung ebenfalls angepasst werden kann.

Für die Berechnung der Messunsicherheit ausschlaggebend sind Angaben über den Lokalisationsfehler auf dem Kamerasensor und den Projektionsfehler des Projektors. Der Lokalisationsfehler findet seine Ursache dabei in der diskreten Pixelabbildung des abtastenden Sensors. Auch wenn Berechnungen auf Subpixel-Genauigkeit durchgeführt werden, um die diskrete, pixelgenaue Abbildung auf Seiten des Sensors auszugleichen, ist eine exakte Positionsbestimmung nur bis zu gewissen Abweichungen Δx_l , bzw. Δx_r , möglich (siehe Abbildung 37). Denn neben dem Sampling des Bildes, also der digitalen Bildabtastung, können auch durch Blurring (allgemein: Unschärfeerscheinungen) in höher auflösenden

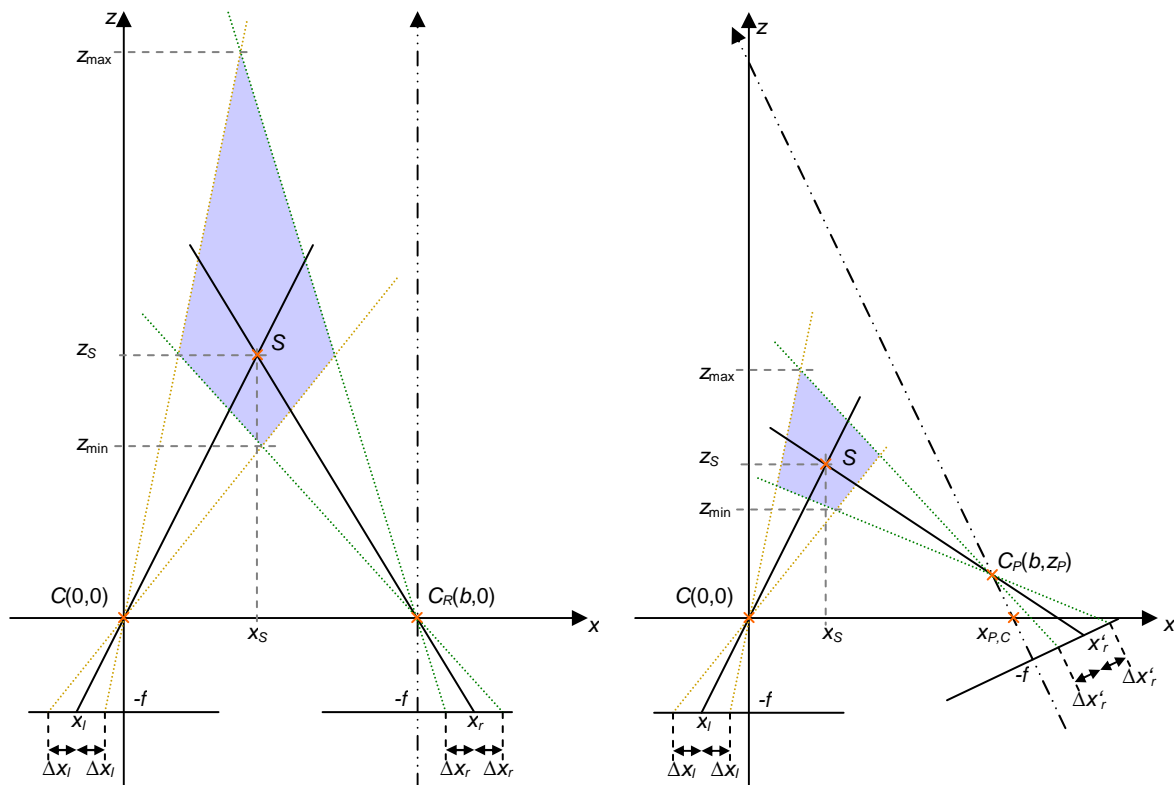


Abbildung 37 Abweichung der Tiefenberechnung durch Fehldetektion auf dem Bildsensor. Bei der linken, parallelen Anordnung fällt eine Pixelabweichung stärker ins Gewicht als bei der rechten, gedrehten Variante eines Stereosystems.

Kameras oder durch Linsenverzerrungen verursachte Fehler einen Einfluss auf die Pixelabbildung ausüben.

Dem Lokalisationsfehler entsprechend lassen sich die Ursachen des Projektionsfehlers eines aktiven Stereosystems beschreiben – bei einem passiven System ist dieser nicht vorhanden, dafür wäre in beiden abtastenden Systemen mit einem Lokalisationsfehler zu rechnen. Eine Ursache des Projektionsfehlers liegt in der Pixelstruktur des bilderzeugenden Sensors, da dessen einzelne Pixel nicht direkt aneinander angrenzend liegen. Dieser Fehler kann bei kontinuierlichen Mustern oder Masken nicht auftreten. Dennoch können durch die Linse Effekte wie Blurring oder chromatische Aberrationen verursacht werden, so dass im Allgemeinen für die Projektionsseite zunächst ebenfalls eine Abweichung von Δx_r berücksichtigt werden muss. Weitere Messungenauigkeiten, wie Modellierungsfehler, Kalibrierfehler oder Fehldetektionen, werden im Folgenden nicht berücksichtigt, da sie sich entweder durch entsprechende Verarbeitungsschritte (z.B. bei Fehldetektionen) ausschließen lassen oder ihr Einfluss im Vergleich zu den anderen Faktoren als gering einzuschätzen ist.

Somit kann, basierend auf der Tiefenberechnung eines regulären Stereosensors gemäß Gleichung (35), der Messfehler einer Entfernungsmessung als Tiefenfehler Δz wie folgt beschrieben werden:

$$\Delta z = \frac{bf}{x_l - x_r} - \frac{bf}{(x_l + \Delta x_l) - (x_r + \Delta x_r)} \quad (47)$$

Hierbei entsprechen Δx_l und Δx_r den eben und in Abbildung 37 beschriebenen Abweichungen. Mit $(\Delta x_l - \Delta x_r) = \Delta x$ und $(x_l - x_r) = d$ lässt sich Gleichung (47) umformen zu:

$$\Delta z = \frac{bf \cdot \Delta x}{d(d + \Delta x)} \quad (48)$$

Es lässt sich hierauf die Annäherung $d(d + \Delta x) \approx d^2$ zur Vereinfachung anwenden. Denn aufgrund der typischen Anwendung des Sensors im Nahbereich (z.B. zur Gesichtserkennung) ergibt sich ein Objektabstand von 500-1500mm. Bei einer üblichen hierfür angenommenen Anordnung der Hardwarekomponenten beträgt der Abstand der Sensorelemente zur Aufnahme und Projektion, die Basis b , etwa 250mm. Die Seitenlänge eines Pixels liegt für Kamera und Videoprojektor bei etwa 0,006-0,010mm, je nach Sensorauflösung. Aufgrund der räumlichen Anordnung der Sensoren und der angenommenen Objektentfernung ist d , das im Bereich der Sensorabmessungen liegt, stets deutlich größer als Δx , welches im Bereich der Pixelgröße liegt, so dass die beschriebene Annahme getroffen werden kann. Aus Gleichung (48) ergibt sich also unter Verwendung von (35) für den Fehler der Tiefenberechnung:

$$\Delta z = \frac{bf \cdot \Delta x}{d^2} = \frac{\Delta x \cdot z_s^2}{bf} \quad (49)$$

Dies gilt für eine reguläre, parallele Anordnung der beiden Sensoren. Da diese aber üblicherweise unter einem Winkel zueinander gedreht erscheinen (siehe Abbildung 37), muss dies in der Tiefenberechnung berücksichtigt werden, wie es durch Gleichung (40) ausgedrückt wird. Für den Tiefenfehler gilt in diesem Fall analog zu (47):

$$\Delta z = \frac{bf - \hat{x}_r z_p}{x_l - \hat{x}_r} - \frac{bf - (\hat{x}_r + \Delta x_r) z_p}{(x_l + \Delta x_l) - (\hat{x}_r + \Delta x_r)} \quad (50)$$

Gemäß den Überlegungen, die zur Herleitung von Gleichung (49) führten, ergibt sich damit für Δz :

$$\Delta z = \frac{\Delta x (bf - \hat{x}_r z_p) + \hat{d} \cdot \Delta x_r z_p}{\hat{d}^2} \quad (51)$$

Diese Gleichung erlaubt die Bestimmung des Tiefenfehlers unter Berücksichtigung aller vorhandenen formalen Parameter, erschwert aufgrund ihrer Komplexität bei der Berücksichtigung des Rotationswinkels und der entsprechenden Transformationen allerdings die Interpretation des Einflusses eben dieser Rotation.

Zur weiteren Umformung von Gleichung (51) wird nun der Faktor Δx_r im hinteren Teil des Terms betrachtet. Dieser entspricht im vorliegenden Fall der Abweichung des rechten Sensors, also des Projektionsmusters. Während er für einen Sensor zur Aufnahme als Lokalisationsfehler durchaus von Bedeutung ist, so fällt er für ein digitales Projektionsmuster oder eine Projektionsmaske kaum ins Gewicht. Beobachtungen bestätigen, dass durch die präzise Fertigung heutiger Projektoren die Pixelabweichung minimal für die meisten Messsysteme ist. Benutzt man keinen Projektor, sondern eine feste Maske, so lässt sich ein Fehler völlig ausschließen, da eine Maske einerseits keine Pixelstruktur besitzt und andererseits eventuelle Maskenfehler oder Ungenauigkeiten exakt im Vorfeld vermessen werden können.

Berücksichtigt man dies mit $\Delta x_r \approx 0$, so vereinfacht sich (51) zu:

$$\Delta z = \Delta x \frac{bf - \hat{x}_r z_p}{\hat{d}^2} = \frac{\Delta x \cdot z_s^2}{bf - \hat{x}_r z_p} \quad (52)$$

Weitere Umformungen stellen schließlich eine Beziehung des Tiefenfehlers Δz_{rot} einer rotierten Anordnung zu dem einer regulären (Δz_{reg}) gemäß Gleichung (49) her, so dass gilt:

$$\Delta z_{rot} = \frac{\Delta x \cdot z_s^2 \cdot \frac{1}{bf}}{1 - \frac{z_p}{b} \cdot \frac{\hat{x}_r}{f}} = \frac{\Delta z_{reg}}{1 + \frac{z_p}{b} \cdot \frac{\hat{x}_r}{-f}} = \frac{\Delta z_{reg}}{1 + \frac{z_p}{b} \cdot \frac{x' \cos \theta + f \sin \theta}{x' \sin \theta - f \cos \theta}} \quad (53)$$

Diese Formulierung erlaubt es schließlich, den direkten Einfluss des Triangulationswinkels θ auf den Tiefenfehler und damit auf die Messgenauigkeit auszudrücken. Deren Abhängigkeit von Position und Winkel erlaubt aber noch weiterführende Überlegungen.

Betrachtet man den Winkel θ , unter dem sich die Sichtachsen von Kamera und Projektor schneiden, so ergibt sich ein Schnittpunkt zwischen der z -Achse des Projektors und der x -Achse der Kamera, der sich zu $x_{P,C} = b + z_p \tan \theta$ berechnet (vgl. Abbildung 37). Aufgrund des vernachlässigbaren Projektionsfehlers und der optischen Abbildungseigenschaften der Projektionslinsen lässt sich damit das Projektionszentrum auch durch die Koordinaten $C_P = (b + z_p \tan \theta, 0)$ ausdrücken (siehe auch [Blais 2004]). Betrachtet man weiter die Tatsache, dass die Linsen zur Projektion im vorliegenden Fall keine Weitwinkeloptik darstellen, wie sie für eine großflächige Aufnahme im Nahbereich benötigt werden würden, so lassen sich die Winkel, unter denen die Lichtebenen des Musters erscheinen, allgemein durch den Winkel θ approximieren. Damit lässt sich der Fehler einer stereoskopischen Triangulation unter Anwendung von Gleichung (49) allgemein ausdrücken durch:

$$\Delta z = \frac{\Delta x \cdot z_s^2}{f \cdot (b + z_p \tan \theta)} = \frac{\Delta x \cdot z_s^2}{bf \cdot \left(1 + \frac{z_p}{b} \tan \theta\right)} = \frac{\Delta z_{reg}}{1 + \frac{z_p}{b} \tan \theta} \quad (54)$$

Die Messunsicherheit Δz wird dadurch unter Verwendung der geometrischen Parameter der Sensoranordnung und durch die Entfernung eines Objektpunktes z_S als einzigem szenenrelevantem Faktor vollständig beschrieben. Eine direkte Beziehung zwischen einer regulären und rotierten, verschobenen Anordnung ist damit ebenfalls jederzeit herstellbar.

3.6.2 Statistische Auswertung des Messfehlers

Mit den eben beschriebenen Formeln des Messfehlers eines allgemeinen stereoskopischen Sensors ist es nun möglich, Aussagen über die statistische Fehlerverteilung einer Messung zu treffen. Hierbei kommt zum Tragen, dass der Tiefenfehler Δz bei einer bekannten Objektentfernung z eine monotone Funktion über den Lokalisations- und Projektionsfehler, bzw. deren Differenz Δx , beschreibt.

Für den Lokalisationsfehler eines Bildsensors wird im Allgemeinen angenommen (z.B. [Blostein and Huang 1987], [Rodriguez and Aggarwal 1990], [Chang et al. 1994]), dass dieser, basierend auf der effektiven Größe eines Bildpunkts dx , im Bereich $[-dx/2; +dx/2]$ in etwa gleichmäßig verteilt ist. Die Wahrscheinlichkeitsdichtefunktion des Lokalisationsfehlers ergibt sich damit für dieses Intervall zu $f_{\Delta x}(x) = 1/dx$ und zu 0 für Werte außerhalb des Intervalls. Der Erwartungswert des Lokalisationsfehlers liegt bei einer Varianz des Fehlers von $dx^2/12$ folglich ebenfalls bei 0.

Da diese Analyse lediglich aktive Systeme betrachtet, die auf der Anwendung strukturierter Lichts basieren, kann, wie im vorigen Kapitel, der Projektionsfehler vernachlässigt werden. Mit $\Delta x = \Delta x_l$ liegt der Tiefenfehler Δz , abhängig von der Objektentfernung z , gemäß Gleichung (49) innerhalb folgender Grenzen:

$$\Delta z \in [\Delta z_{\min}; \Delta z_{\max}] \in \left[-\frac{dx \cdot z^2}{2 \cdot bf}; +\frac{dx \cdot z^2}{2 \cdot bf} \right] \quad (55)$$

Innerhalb dieser Grenzen ergibt sich dann die Wahrscheinlichkeitsdichtefunktion des Tiefenfehlers zu:

$$f_{\Delta z}(\Delta z) = f_{\Delta x}(\Delta z) \cdot \frac{bf}{z^2} = \frac{bf}{dx \cdot z^2} \quad (56)$$

Außerhalb des Intervalls $[\Delta z_{\min}; \Delta z_{\max}]$ liegt sie bei 0.

Der Erwartungswert des Tiefenfehlers liegt, wie der des Lokalisationsfehlers, bei 0, seine Varianz $Var(\Delta z)$ und Standardabweichung $\sigma(\Delta z)$ lassen sich beschreiben durch:

$$Var(\Delta z) = \frac{dx^2}{12} \cdot \left(\frac{z^2}{bf} \right)^2 \quad \text{und} \quad \sigma(\Delta z) = \frac{dx \cdot z^2}{\sqrt{12} \cdot bf} \quad (57)$$

Mit den bekannten Annäherungen zur Bestimmung des Tiefenfehlers einer allgemeinen Anordnung, die im vorigen Kapitel beschrieben wurden (vgl. Gleichung (54)), lässt sich nun auch direkt die Wahrscheinlichkeitsdichtefunktion desselbigen herleiten:

$$f_{\Delta z_{rot}}(\Delta z_{rot}) = \frac{f \cdot (b + z_p \tan \theta)}{dx \cdot z^2} \quad (58)$$

Die Berechnung von Varianz und Standardabweichung folgt dementsprechend in allgemeingültiger Formulierung:

$$\text{Var}(\Delta z) = \frac{dx^2}{12} \cdot \left(\frac{z^2}{f \cdot (b + z_p \tan \theta)} \right)^2 \quad \text{und} \quad \sigma(\Delta z) = \frac{dx \cdot z^2}{\sqrt{12} \cdot f \cdot (b + z_p \tan \theta)} \quad (59)$$

Diese Gleichungen werden nun auch für die Analyse der Sensoren und zur Beschreibung der theoretischen Messunsicherheiten verwendet. Zur tatsächlichen empirischen Analyse des Messfehlers und der Messabweichungen wird die Vermessung einer planen Fläche durchgeführt. Hierfür wird als Basis eine Tiefenkarte der Kalibrierplatte erstellt und analysiert, da diese aufgrund ihrer Materialeigenschaften (Glas mit einer Dicke von 1,5cm) ideale Voraussetzungen für eine flache Oberfläche bietet. Zudem erlauben die aufgebrachten Kalibriermarken eine Bestimmung der Lage der Platte im Raum, so dass die gemessenen Tiefenkoordinaten im Weltkoordinatensystem der Platte angegeben werden können. Das bedeutet, der Erwartungswert der gemessenen z -Koordinaten liegt bei 0, die abweichenden Werte streuen um diesen Wert. Die Platte wird dabei im Arbeitsraum des Sensors platziert, so dass ihre gesamte Oberfläche im Schärfebereich des Projektors und der Kamera liegt und innerhalb der Grenzwerte vermessen werden kann.

Zum Vergleich werden die gleichen Messungen mit einer Anordnung aus einem LCD-Videoprojektor im sichtbaren Licht und einer Kamera mit gleichen intrinsischen Parametern (Sony ICX 415AQ) ausgeführt. Die Ergebnisse dieser Messungen sind in Tabelle 6 aufgelistet. Dabei werden neben dem durchschnittlichen Tiefenwert (also dem Erwartungswert der Messung) sowohl die gemessene, als auch die gemäß Gleichung (59) vorhergesagte Standardabweichung der Messwerte angegeben. Aus der Auflistung ist ersichtlich, dass bei beiden Sensorsystemen die Messwerte unterhalb der theoretischen Grenzwerte liegen und dass das NIR-System gemäß seiner Messdaten ähnliche Leistungsdaten bezüglich der Genauigkeit besitzt wie ein System im sichtbaren Licht.

Tabelle 6 Statistischer Tiefenfehler aktiver Stereosysteme. Es wird der vorgestellte NIR-Sensor mit einem CCT-System unter Verwendung eines LCD-Projektors verglichen. Die Ergebnisse beruhen auf der Aufnahme einer glatten Kalibrierplatte, die als Referenzebene mit einem Erwartungswert von 0 dient. Der tatsächliche Erwartungswert, sowie die gemessene und statistische Standardabweichung (nach (59)) werden ebenfalls angegeben und verglichen.

Sensor	Abstand [mm]	$E(z)$ [mm]	$\sigma(\Delta z)$ (gemessen)	$\sigma(\Delta z)$ (berechnet)
NIR	625	0,08	0,14	0,204
NIR	665	0,07	0,14	0,231
LCD	745	0,08	0,16	0,219
LCD	795	-0,02	0,17	0,250

3.6.3 Sampling und Bildrekonstruktion

Bei den Überlegungen zur Störanfälligkeit unter 3.2.2 wurden bereits die möglichen Einflüsse von Sampling-Artefakten angedeutet. Neben diesen Phänomenen können ähnliche Effekte durch die Farbbilderfassung hervorgerufen werden, wie sie bei einem farbcodierten System im sichtbaren Licht zur Aufnahme der codierten Szene vonnöten ist (vgl. [Modrow et al. 2008b]). Dieser Abschnitt befasst sich mit diesen Einflüssen auf die farbcodierte Triangulation und vergleicht hierbei verschiedene Kamera-Projektor-Kombinationen mit dem vorgestellten NIR-Sensor.

3.6.3.1 Testumgebung

Für den Aufbau und die Untersuchung der sichtbaren, farbcodierten Triangulation werden Videoprojektoren mit den drei meist verbreiteten Techniken verwendet:

- LCD (Liquid Crystal Display):
Panasonic PT-LB50NTE
- LCoS (Liquid Crystal on Silicone):
Canon XEED-SX50
- DLP (Digital Light Processing):
Proxima Ultralight X350 & Toshiba ff1

Ausführlichere Informationen zu den einzelnen Projektionstechniken finden sich unter 5.1 mit der Beschreibung der verschiedenen Sensor-Prinzipien.

Als Kamera kommt der bereits bekannte CCD-Sensor Sony ICX 415 in seiner Farbversion mit einem Bayer-Filter zum Einsatz (ICX 415AQ), der eine polychromatische (mehrfarbige) Projektion des Musters mit allen Farbkanälen in einer Aufnahme verarbeiten kann. Zur Simulation einer 3-Chip-Kamera wird die Schwarz-Weiß-Version des gleichen Sensors (ICX 415AL) verwendet, indem drei Aufnahmen der verschiedenen Farbkanäle kombiniert werden (zur Farberzeugung siehe auch 2.1.2.3). Im Folgenden wird diese Anordnung deshalb als 3-Kanal-Kamera bezeichnet. Auch wenn diese simulierte Kamera lediglich statische Objekte aufnehmen kann, ermöglicht sie es, den Einfluss bzw. das Fehlen einer Bildrekonstruktion direkt zu beobachten. Um die notwendige Statik der Szene zu gewährleisten, wird als Objekt eine glatte, weiße Ebene verwendet, so dass die rekonstruierten Tiefendaten ebenfalls eine Ebene darstellen.

Betrachtet man nun für diese Kombinationen die Aufnahmen der farbcodierten Triangulation genauer (vgl. auch Abbildung 39), so fallen in den rekonstruierten Oberflächen zwei Arten von Tiefen-Artefakten auf – nämlich einerseits Abweichungen, die den projizierten Streifen zu folgen und dadurch code-abhängig scheinen, sowie andererseits unregelmäßige Abweichungen, die die Oberfläche rau erscheinen lassen.

3.6.3.2 Code-abhängige Rekonstruktionsfehler

Zur Analyse der erstgenannten Rekonstruktionsfehler betrachten wir die Auswertung der Kanten, an die die Fehler gekoppelt zu sein scheinen. Dabei stellt man fest, dass die detektierten Positionen der Kanten der einzelnen Farbkanäle voneinander abweichen und nicht den gleichen Positionen zugeordnet werden, was die erwähnte Abweichung bei der folgenden Tiefenberechnung erklärt. Abbildung 38 verdeutlicht diese Kantenverschiebung der Kanäle zueinander an verschiedenen Positionen des aufgenommenen Bildes bzw. des projizierten Musters. Bei einer von links nach rechts laufenden Codierung werden Positionen am linken und rechten Rand sowie in der Mitte betrachtet.

Dabei fällt auf, dass sich die Verschiebungen der Farbkanäle bei polychromatischer Projektion – Diagramme (1) und (2) – relativ zueinander ändern, indem die Kanäle Rot und Blau über eine deckungsgleiche Anordnung in der Bildmitte ihre Positionen quasi vertauschen. Projiziert man dagegen die einzelnen Kanäle des Musters mit einer monochromen Beleuchtung sequentiell – es werden also alle Kanäle nacheinander beispielsweise durch eine Grün-Schwarz-Projektion dargestellt – so treten diese Verschiebungen nicht auf (Abbildung 38 (3)). Berechnet man die durchschnittliche Abweichung der Kanäle zueinander, so ergeben sich je nach Lage innerhalb des Bildes Differenzen von bis zu einem Pixel, wodurch sich die deutlichen Tiefenfehler erklären lassen. Ein Fehler auf Seiten der Projektion kann dabei ausgeschlossen werden, da der beobachtete Fehler bei allen Projektoren in ähnlicher Ausprägung auftritt. Sogar bei den DLP-Projektoren, die aufgrund ihres Aufbaus nur ein Panel besitzen und die drei Grundfarben nacheinander darstellen, zeigen sich Kanalabweichungen. Die Ursache hierfür liegt vielmehr in den chromatischen Aberrationen der Linsen des Projektors und der Kamera. Diese sind in einem gewissen Maße jedoch korrigierbar, wie Abbildung 39 (3) zeigt. In diesem Beispiel wurde eine lineare Korrektur des chromatischen Fehlers angewendet. Dafür werden die relativen Kanten- und Kanalverschiebungen zueinander analysiert, die Positionsänderungen gemittelt und die Ergebnisse dementsprechend farbspezifisch angepasst, wodurch sich das Gesamtergebnis bereits deutlich verbessern lässt. Alternative Möglichkeiten zur Korrektur chromatischer Aberrationen werden unter anderem auch von [Mallon and Whelan 2007] vorgeschlagen.

Die beste Darstellung lässt sich jedoch mit einer sequentiellen monochromen Projektion und monochromer Aufnahme erzielen, wie Abbildung 39 (4) und (5) zeigen – so wie es auch beim NIR-Sensor praktiziert wird.

3.6.3.3 Unregelmäßige Tiefenabweichungen

Neben den Tiefenabweichungen entlang der Codestreifen fällt bei der polychromatischen Variante der farbcodierten Triangulation zudem die relativ unebene, raue Oberfläche auf (siehe Abbildung 39 (1)). Eine Beobachtung, die auf eine mögliche Ursache hierfür hinweist, wurde bereits in Abschnitt 3.2.2 gemacht – die Einflüsse des Moiré-Effekts.

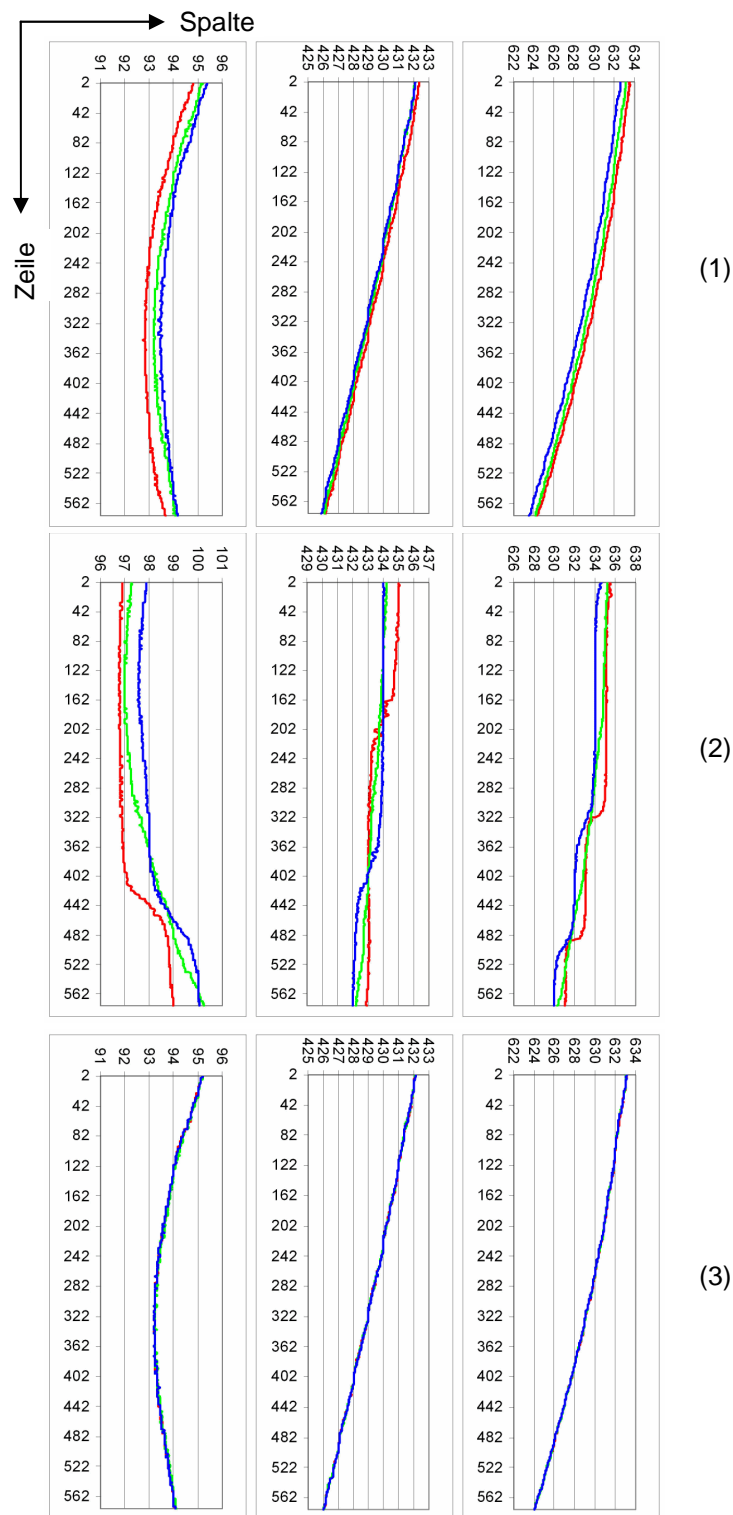


Abbildung 38 Kantenpositionen der drei Farbkanäle des farbcodierten Musters. Die polychromatische Projektion zeigt bei Aufnahme einer 3-Kanal-Kamera (1) eine Abweichung der Kanäle, abhängig von der Bildposition, bei Aufnahme über ein Bayer-Filter (2) zudem eine deutliche Treppenstruktur. Bei einer sequentiellen monochromen Projektion und Aufnahme mittels Schwarz-Weiß-Kamera (3) sind erwartungsgemäß keinerlei Abweichungen zu erkennen.

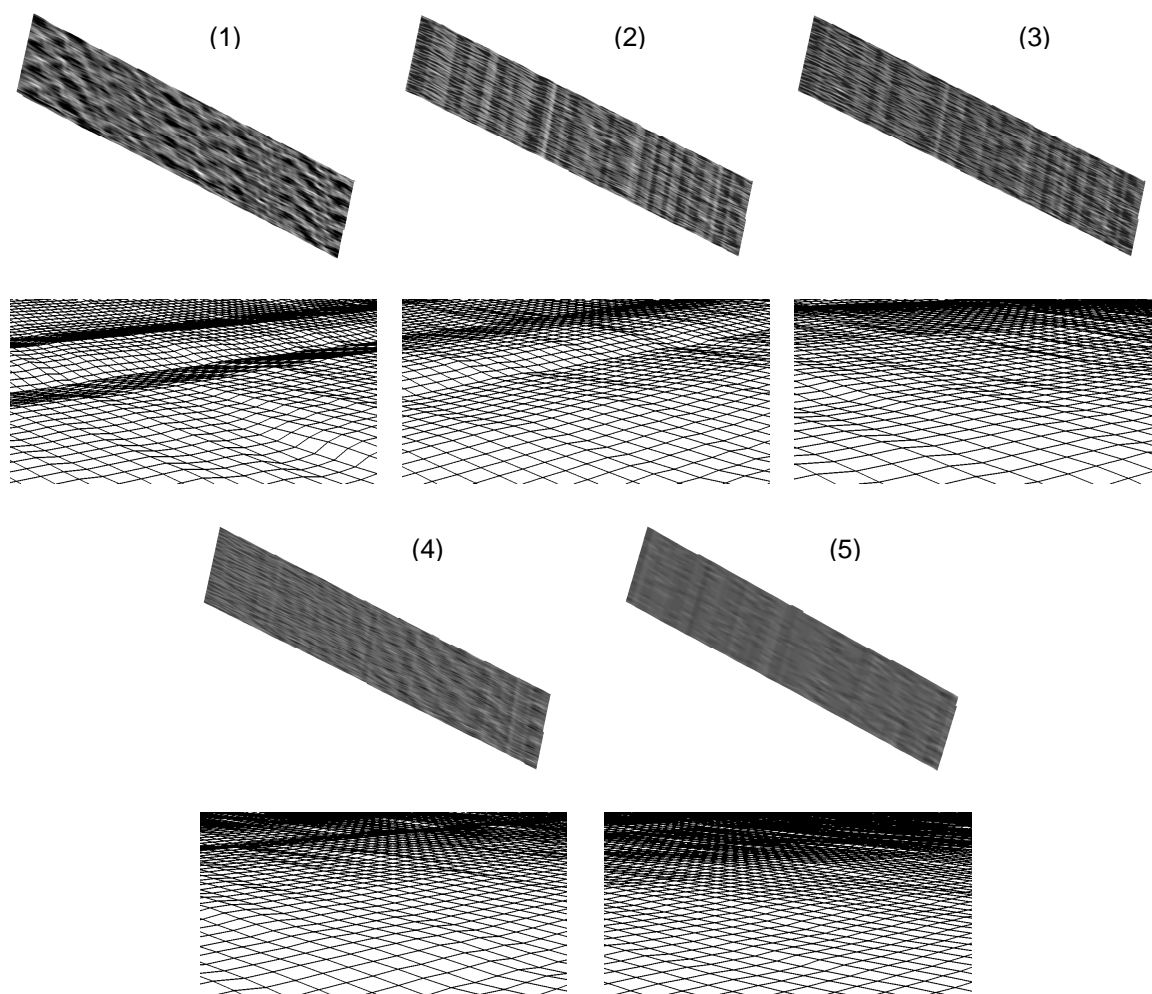


Abbildung 39 Tiefenkarten und Oberflächen einer Ebene, aufgenommen mit verschiedenen Verfahren. Digitale Polychromatische Projektion mit Bayer-Kamera (1), digitale polychromatische Projektion mit 3-Kanal-Kamera (2), digitale polychromatische Projektion mit 3-Kanal-Kamera und chromatischer Korrektur (3), digitale sequentielle monochrome Projektion mit Schwarz-Weiß-Kamera (4) und NIR-Sensor mit analoger monochromer Projektion mit Schwarz-Weiß-Kamera (5).

Das durch Überlagerung der Pixelstrukturen hervorgerufene Interferenzmuster besitzt eine Periode p_i , die sich in Abhängigkeit der Gitterabstände von Projektor (p_p) und Kamera (p_c) bei paralleler Anordnung folgendermaßen ausdrücken lässt:

$$p_i = \frac{p_p \cdot p_c}{|p_p - p_c|} \quad (60)$$

Da davon auszugehen ist, dass man keine ideale, parallele Anordnung verwenden kann, muss die Rotation der Gitterrichtungen, ausgedrückt durch den Zwischenwinkel α , berücksichtigt werden, wodurch sich (60) zu folgender Formulierung ändert:

$$p_i = \frac{p_p \cdot p_c}{\sqrt{p_p^2 + p_c^2 - 2 \cdot p_p \cdot p_c \cdot \cos \alpha}} \quad (61)$$

Grundsätzlich ließe sich damit bei bekannter Anordnung der Sensoren der Einfluss des Moiré-Effekts für weitere Berechnungen berücksichtigen.

Versuche zeigen jedoch, dass dessen Auswirkung generell relativ gering ausfällt. Vor allem aber ist der Effekt nur innerhalb eines schmalen Ausschnitts des Schärfebereichs zu beobachten. Er tritt also nur in Bereichen auf, in denen sowohl die Projektion des Musters, als auch die Schärfentiefe der Kamera gleichermaßen exakt fokussiert sind. Dies lässt sich aber beispielsweise durch leichtes Defokussieren eines der Objektivs ausschließen. Da der beobachtete Effekt auf die Tiefendaten dabei aber weiterhin auftritt, müssen die Ursachen hierfür einen anderen Grund haben. Betrachtet man Abbildung 38 (2), so fällt eine gewisse Treppenstruktur im Rot- und Blaukanal mit plötzlichen Sprüngen über mehrere Pixel auf, deren Positionen sich mit den Tiefenabweichungen decken. Es ist zu beobachten (siehe Abbildung 38 (1) und (2)), dass dieses Phänomen nur bei der Bayer-Kamera, nicht jedoch bei der 3-Kanal-Kamera auftritt. Die Ursache liegt also offensichtlich in der Art der Farb-rekonstruktion, also dem so genannten Demosaicking des Bayer-Filters.

Wie in Abschnitt 2.1.2.3 beschrieben, lag der Entwicklung und Anwendung des Bayer-Filters oder vergleichbarer Farbfilter zugrunde, dass sich die Farbinformationen innerhalb einer räumlich hinlänglich kleinen Bildregion kaum ändern. Die Annahme dieses kontinuierlichen Farbverlaufs bildet dabei die Grundlage für die Interpolation aller drei Kanalinformationen aus benachbarten Bildpunkten, was für das Auge ein reales Abbild der Umgebung ergibt. Allerdings ist genau diese Annahme kontinuierlicher Farbübergänge eine Eigenschaft, die für die farbcodierte Triangulation nicht zutrifft. Denn da die Algorithmen der Tiefenberechnung auf der Lage einer Kante zwischen Farbstreifen basieren, betrachtet man einen diskreten, abrupten Farbübergang. Weil man dessen Position möglichst exakt bestimmen möchte, ist deshalb eine möglichst auffällige Farbänderung explizit erwünscht. Der direkte Einfluss des Demosaickings ist dabei in Abbildung 40 zu sehen. Während die eigentlichen Farbbilder (links) sich nur in „ausgefranzten“ Streifen unterscheiden, die auch nur in der vergrößerten Ansicht wirklich sichtbar sind, zeigt die Aufspaltung der Farbkanäle (Mitte) ein anderes Bild. Hier erkennt man bereits einen deutlichen Unterschied zwischen den Farbkanälen, wobei sich besonders der rote und blaue Kanal mit ihrem Streifenverlauf vom grünen unterscheiden. Betrachtet man den Verlauf der Kanten dieser Kanäle (rechts), so erkennt man in diesen beiden Kanälen eine Treppenstruktur, die sich bereits im mittleren Diagramm (2) aus Abbildung 38 gezeigt hat. Der Fehler, der sich durch diese Bildrekonstruktion ergibt, lässt sich anschaulich durch die mittlere quadratische Abweichung (RMSD, root mean squared deviation) gemäß folgender Formel beschreiben:

$$RMSD(P_{ref}, P_{rec}) = \sqrt{\frac{1}{N} \sum_{i \in N} (P_{ref}(i) - P_{rec}(i))^2} \quad (62)$$

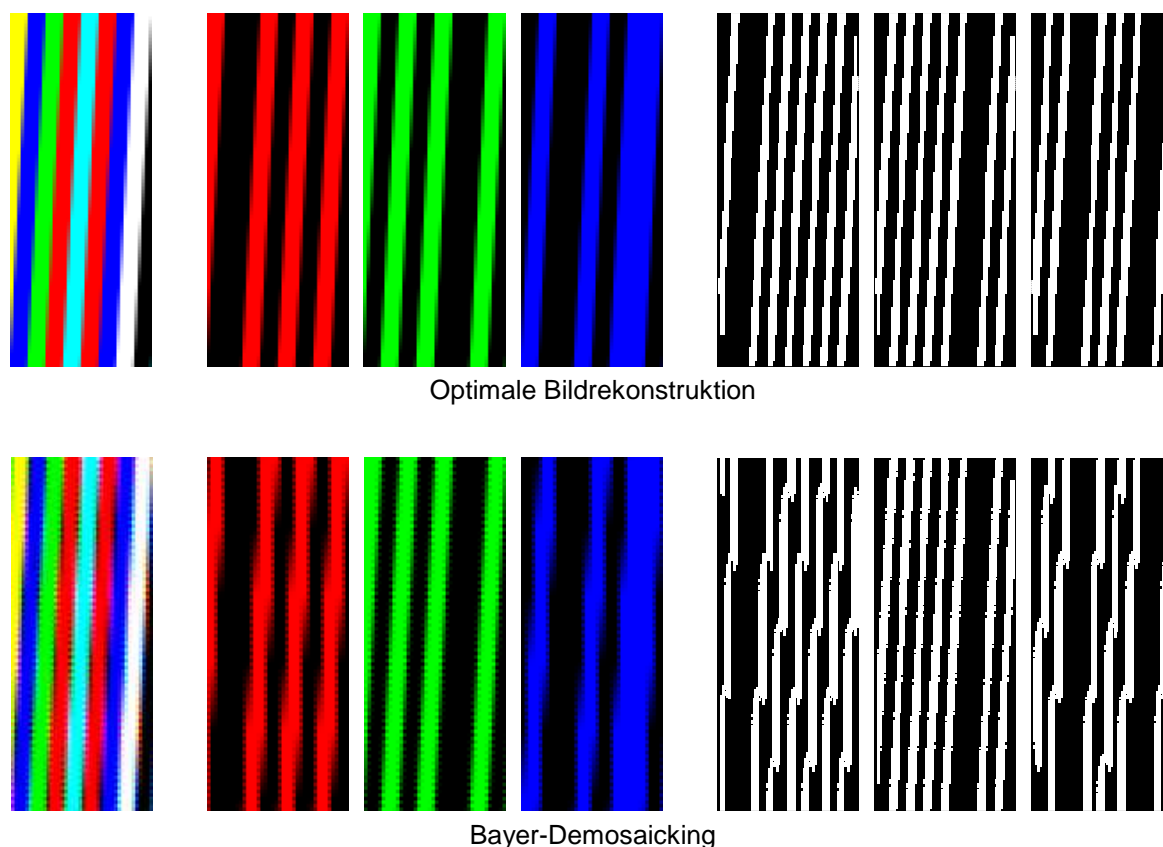


Abbildung 40 Einfluss des Demosaickings auf den Kantenverlauf einer Codierung. Die obere Zeile stellt die optimale Aufnahme eines Streifenmusters dar, die daraus folgende Interpretation der drei Farbkanäle, sowie die Positionen der sich daraus ergebenden Kanten. Die untere Zeile liefert die entsprechenden Ergebnisse einer Aufnahme mittels eines Bayer-Filters mit einer Bildrekonstruktion durch Demosaicking. Deutlich erkennbar ist die Treppenstruktur der roten und blauen Kanten.

Berechnet wird die Abweichung auf Basis eines über das Filter aufgenommenen (recorded) Bildes und des idealen Referenz-Bildes (reference), der Wert P_i steht dabei, je nach Betrachtung des Fehlers, entweder für den Intensitätswert eines Kanals oder für die Position einer Kante im jeweiligen Bild.

Für die einzelnen Farbkanäle ergibt sich hierbei bei einer Farbtiefe von 8bit pro Kanal eine Abweichung der Intensitätswerte von etwa 40 Stufen für den Rot- und Blaukanal und von 20 Stufen für den Grünkanal. Dieser Wert hängt von der Verkippung des Musters, bzw. der Streifen, relativ zu den Pixelspalten ab, wie es Tabelle 7 auflistet. Dass die Werte des Grünkanals etwa der Hälfte des Rot- und Blaukanals entsprechen, ist dabei zu erwarten, da das Bayer-Pattern zu 50% aus grünen und nur zu jeweils 25% aus roten und blauen Bereichen besteht. Aus den Intensitätsunterschieden resultiert schließlich der Fehler in der Bestimmung der Position einer Kante, der sich in der rauen Oberflächenstruktur äußert. Diese Positionsabweichung liegt, ebenfalls abhängig von der Verkippung des Codemusters, etwa im Bereich von 0,3 Pixel, wodurch die sichtbaren Höhenunterschiede in der Tiefenkarte hervorgerufen werden.

Verkipfung	<i>RMSD(rot)</i>	<i>RMSD(grün)</i>	<i>RMSD(blau)</i>	<i>RMSD(Pos)</i>
0°	51,90219	25,34369	50,47343	0,33196
1°	38,55651	17,91452	37,72368	0,27670
2°	38,32293	17,91222	37,52121	0,27927
3°	38,06589	17,94588	37,29069	0,27811
4°	37,83792	17,99994	37,08006	0,27778
5°	37,64285	18,08970	36,88473	0,27674

Tabelle 7 Mittlere quadratische Abweichung der Bildrekonstruktion eines Bayer-Filters. Die Werte entsprechen den Abweichungen der rekonstruierten Intensitätswerte bei 8bit Farbauflösung von den Originalen, aufgespalten nach den drei Farbkanälen bei verschiedener Verkipfung gegenüber den Bildspalten. Die daraus resultierende Positionsabweichung einer Kante in absoluten Pixeln ist ebenfalls angegeben.

Es existieren zwar Ansätze für feinere Demosaicking-Algorithmen, Bildverbesserungen nach der Bildrekonstruktion oder alternative Filter-Strukturen (siehe z.B. [Lukac et al. 2005], [Gorokhovskiy et al. 2006]), dennoch bleibt all diesen Ansätzen ein Problem gemein – sie müssen nicht-vorhandene Daten rekonstruieren, woraus sich zwangsläufig Fehler ergeben. Da diese besonders an diskreten Übergängen sichtbar werden, bleibt der Einfluss auf die kantenbasierte Triangulationsberechnung stets bestehen. Wird dagegen das Codemuster mit einer 3-Kanal-Kamera aufgenommen, so kann das Bild mit seinem vollen Informationsgehalt bearbeitet werden, was zu einer wesentlich besseren Darstellung führt (siehe Abbildung 39 (2)). Eine vergleichbare Verbesserung lässt sich auch durch monochrome Musterprojektion erreichen, beispielsweise durch Aufspalten des Farbmusters in seine drei Kanäle, wodurch zusätzlich die Einflüsse durch chromatische Aberrationen entfallen (Abbildung 39 (4)).

All diese positiven Eigenschaften sind dabei im NIR-Ansatz vereint:

Es wird monochrom, über eine einzelne Wellenlänge projiziert. Dadurch bedingte chromatische Aberrationen treten also in allen Kanälen oder Teilbildern gleichmäßig auf und verfälschen die Position zusammengefasster Kanten nicht. Die einzelnen Kanäle des Musters werden getrennt aufgenommen, so dass jeder Kanal seine volle Information behält. Es findet also keine Bildrekonstruktion statt. Zuletzt erfolgt die Projektion noch analog, wodurch eventuelle Einflüsse durch Samplingartefakte ausgeschlossen werden können. Dies führt schließlich zu einer deutlich verbesserten dreidimensionalen Aufnahme, als es mit digitalen Projektionsverfahren im sichtbaren Licht erreicht werden kann (Abbildung 39).

3.6.4 Robustheit gegenüber Umgebungslicht

Für die Alltagstauglichkeit eines Sensors von großer Bedeutung ist dessen Unabhängigkeit gegenüber umweltbedingten Störeinflüssen. Bei einem Stereoskopiesystem mit aktiver Beleuchtung betrifft dies vor allem das natürliche oder künstliche Umgebungslicht, welches ein Objekt oder eine Szene zusätzlich beleuchtet.

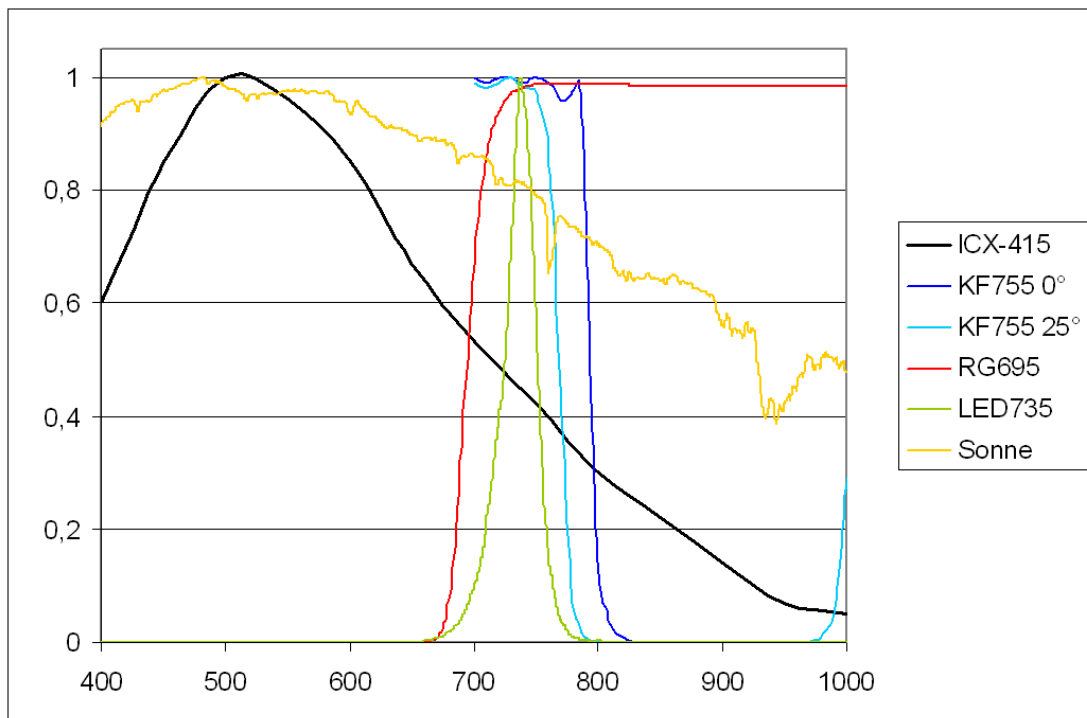


Abbildung 41 Aufbau eines NIR-Bandpassfilters. Das Diagramm zeigt die, jeweils auf ihr Maximum normierten, Kurven der relativen Empfindlichkeit des verwendeten CCD-Sensors, sowie die Spektren des Sonnenlichts und der verwendeten LED. Das Bandpassfilter ist eine Kombination aus einem Rotglas RG695 und einem Kantenfilter bei 755nm, weshalb es eine gewisse Winkelabhängigkeit besitzt

Von diesem Umgebungslicht muss sich eine Musterprojektion derart absetzen können, dass die abgebildeten Informationen im aufgenommenen Bild erkannt und decodiert werden können. Das lässt sich freilich mit einer ausreichend starken Beleuchtungsintensität sicherstellen, wie sie beispielsweise Videoprojektoren mit Hochdrucklampen (oder auch UHP-/UHE-/VIP-Lampen) mit einem Lichtstrom von über 1000 ANSI-Lumen problemlos erreichen können. Dass dies jedoch für das Erfassen von Gesichtern aufgrund von Blendwirkungen und möglicher Schädigung des Auges vollkommen ungeeignet ist, wurde bereits ausführlich beschrieben.

Deshalb ist ein Kompromiss aus einer geringen Lichtleistung und einer gewissen Robustheit gegenüber Umgebungslicht nötig, um einen Sensor alltagstauglich nennen zu können. Beim beschriebenen NIR-Sensor wird dies zum einen durch die kaum wahrnehmbare Wellenlänge erreicht, die dabei von, verglichen mit normalen Projektorlampen, leuchtschwachen LEDs abgestrahlt wird. Zum anderen erlaubt es die Beschränkung auf eine spezifische Wellenlänge, das übrige Umgebungslicht durch einen Bandpass zu filtern. Somit lässt sich erreichen, dass lediglich der Wellenlängenbereich, der Informationen beinhaltet, vom Bildsensor erfasst werden kann. Im Gegensatz dazu muss ein System, das im sichtbaren Licht mit verschiedenen Farben arbeitet, für das gesamte sichtbare Lichtspektrum empfindlich sein. Da hierfür die Codierung breitbandig angelegt ist, können auch

	$E_{\text{Fremdlicht}}$	$K(\text{NIR}),$ ungefiltert	$K(\text{NIR}),$ gefiltert	$K(\text{LED}),$ sichtbar
Kein Umgebungslicht	0 lx	0,889	0,819	0,769
Leuchtstofflampe	150 lx	0,564	0,790	0,618
Halogenlampe	400 lx	0,136	0,401	0,183
Tageslicht	1200 lx	0,044	0,124	0,040

Tabelle 8 Robustheit gegenüber Fremdlicht. Die Tabelle listet die gemessenen Kontrastwerte zwischen einem hellen und einem dunklen Code-Streifen auf. Die Szene wird dabei direkt von verschiedenen Fremdlichtquellen bestrahlt. Verglichen wird der NIR-Sensor mit und ohne Bandpassfilter mit einem vergleichbaren System im sichtbaren Lichtspektrum.

Störeinflüsse breitbandig einwirken. Abbildung 41 zeigt die Funktionsweise eines solchen NIR-Bandpassfilters auf. Das Diagramm vereint dabei den Intensitätsverlauf des Sonnenlichts und der verwendeten LED, sowie die Empfindlichkeit eines typischen CCD-Sensors. Die einzelnen Kurven des Diagramms sind jeweils normiert und relativ zu ihrem Maximum gemessen und berechnet. Der Bandpass setzt sich aus zwei Einzelfiltern, einem für das sichtbare Licht und einem für das angrenzende Infrarote, zusammen und schließt damit das Spektrum der Lichtquelle direkt ein. Weil das Sonnenlicht im Infraroten aber ebenfalls noch einen erheblichen Anteil besitzt, lässt es sich nicht völlig ausschließen, da dessen absolute Leistung bei direkter Bestrahlung durch den Einsatz von LED-Technik mit vertretbarem technischen Aufwand derzeit nicht übertroffen werden kann.

Als Maß für die Robustheit gegenüber der Umgebungsbeleuchtung wird der Kontrast K zwischen einem hellen (max) und einem dunklen (min) Streifen des Codes durch Messung der jeweiligen Lichtleistung Φ verwendet, der sich folgendermaßen berechnen lässt:

$$K = \frac{\Phi_{\max} - \Phi_{\min}}{\Phi_{\max} + \Phi_{\min}} \quad (63)$$

Der Kontrastwert ist deshalb ein gutes Maß für die Robustheit, da die eigentliche Decodierung des NIR-Sensors auf der Detektion der Kanten zwischen hellen und dunklen Streifen der beiden Kanäle beruht und eben dadurch direkt kontrastabhängig ist.

Als Vergleichssystem wurde eine Kombination einer baugleichen Kamera und eines LED-Projektors (Toshiba ff1) mit vergleichbarer Lichtleistung verwendet. Dabei wurde der gleiche Algorithmus angewendet und die gleichen Muster projiziert, so dass die gleichen Rahmenbedingungen für das System im sichtbaren Licht und den NIR-Sensor vorherrschen. Gemessen wurde die abgestrahlte Lichtleistung in direkter Projektionsrichtung. Dabei wurde die Messung zunächst im Dunkeln durchgeführt und anschließend durch verschiedene, alltägliche Beleuchtung beeinflusst. Tabelle 8 zeigt die Ergebnisse dieser Messungen, sowie die jeweilige Beleuchtungsstärke des Fremdlichts, wobei auch der Einfluss des Bandpassfilters anhand „ungefilterter“ Messungen verdeutlicht wird.

Aus dieser Auflistung ist ersichtlich, dass die Beschränkung auf eine spezifische Wellenlänge, also die monochrome Projektion, erhebliche Vorteile bringt, die sich durch die Anwendung eines Bandpasses nutzen lassen. Somit lässt sich der Kontrastwert im Vergleich zu einem System im sichtbaren Licht um das bis zu dreifache erhöhen, was den Sensor gegenüber Umgebungslicht deutlich robuster werden lässt. Der dargestellte Sensor ist also durchaus für den täglichen Gebrauch in verschiedenen Umgebungen geeignet, vor allem wenn man berücksichtigt, dass sich die abgestrahlte Lichtleistung beispielsweise durch Laser-LEDs weiter erhöhen und das Spektrum, bzw. die Streuung um die Wellenlänge, zusätzlich einengen lassen, wodurch selbst ein Einsatz bei Tageslicht vorstellbar wird.

3.6.5 Untersuchung verschiedener Kameratypen

Im Rahmen der Untersuchungen zur Robustheit gegenüber Fremd- und Störlicht wurden ebenfalls verschiedene Kameras auf ihre Tauglichkeit zur Infrarotaufnahme verglichen. Folgende Typen, bzw. Sensoren, standen dabei zur Auswahl:

- Basler A312f – Sony ICX 415AL (1/2" progressive CCD-Sensor)
- AVT Guppy F044B NIR – Sony ICX 429ALL (1/2" interlaced CCD-Sensor)
- IDS uEye UI-1220-M – Micron MT9V032 (1/3" Global Shutter CMOS-Sensor)

Bei der Auswahl einer Kamera für den NIR-Ansatz ist eines der ersten entscheidenden Kriterien die spektrale Empfindlichkeit im nahen Infraroten, besonders bei der zur Projektion verwendeten Wellenlänge. Die Datenblätter der Sensorhersteller geben hierfür Werte zwischen 30% und 60% an, relativ gemessen zur maximalen Empfindlichkeit, die bei allen Typen bei einer Wellenlänge von etwa 550nm liegt. Die Empfindlichkeit ist besonders deshalb entscheidend, da von ihr die Belichtungszeit einer Aufnahme abhängt und mit dieser der Einfluss von Bewegungsartefakten oder Bildrauschen. Besonders bei einem Verfahren, welches mehr als eine Aufnahme zur Datenauswertung benötigt, sind dies kritische Punkte. Zudem bedeutet eine kurze Belichtungszeit auch eine kürzere Zeit, die für die Musterprojektion benötigt wird. Das eröffnet die Möglichkeit, die zur Beleuchtung ver-

Tabelle 9 Vergleich verschiedener Kamerasensoren auf ihre Infrarot-Eignung. Betrachtet werden der statistische Fehler $\sigma(\Delta z)$ für eine passende Sensor-Anordnung, das Rauschverhalten wird über die Standardabweichung der Intensität $\sigma(I)$ bei 8bit Farbtiefe um einen mittleren Grauwert von 128 innerhalb einer Bilderserie untersucht. Die relativ empfangene Lichtintensität wird für verschiedene Belichtungszeiten bei fester Blendeneinstellung beobachtet. Betrachtet werden stets Vollbilder, auch beim interlaced-Sensor.

Sensor	Typ	Auflösung	dx [μm]	$\sigma(\Delta z)$ [mm]	$\sigma(I)$ (Rauschen)	Intensität [%] bei Belichtungszeit [ms]			
						10	20	30	40
ICX 415AL	1/2"-CCD	782 x 582	8,3	0,231	1,274	0,164	0,326	0,485	0,641
ICX 429ALL	1/2"-CCD	752 x 582	8,6	0,239	0,977	0,297	0,569	0,827	0,951
MT9V032	1/3"-CMOS	782 x 480	6,0	0,237	1,290	0,229	0,463	0,699	0,897

wendeten LEDs im getakteten Betrieb unter höherer Spannung zu verwenden und dadurch die Lichtausbeute zu erhöhen. Im Dauerbetrieb würde eine derartige Spannungserhöhung diese dagegen zerstören.

Es zeigt sich, dass die verschiedenen Sensoren deutliche Unterschiede in ihrer Empfindlichkeit aufweisen. So kann mit dem speziell für den nahen Infrarotbereich ausgelegten Sensor (ICX 429ALL) die beste Empfindlichkeit und damit eine kürzere Belichtungszeit erreicht werden als mit den anderen Typen. Allerdings bedingt die interlaced-Technik bei diesem Typ die Aufnahme von Halbbildern, wodurch Bewegungsartefakte durch die versetzte Abtastung entstehen können.

Ein weiterer Punkt, der zwar indirekt die Intensität, aber vielmehr die statistische Messunsicherheit berührt, ist die Pixelgröße auf dem Sensor und die damit verbundene Sensorfläche und Brennweiteinstellung des Objektivs. Die beiden CCD-Sensoren weisen hier nahezu identische Werte auf, der CMOS-Sensor besitzt eine kleinere Pixelgröße. Da deshalb die Sensorfläche ebenfalls kleiner ausfällt, ist eine andere Brennweite für eine vollständige Szenenabbildung vonnöten. Dies führt dazu, dass der nach Gleichung (59) berechenbare Fehler für alle Sensortypen nahezu gleich ausfällt.

Das Rauschverhalten der einzelnen Sensoren wurde ebenfalls untersucht. Hierfür wurde unter konstanter Beleuchtung eine Serie von Bildern so aufgenommen, dass sich ein Mittelwert von 50% der maximalen Intensität einstellt – also ein Graustufenwert von 128 bei einer 8bit Darstellung. Aus diesen Bildern wurde für jedes Pixel der Mittelwert, sowie die Standardabweichung von diesem bestimmt, aus denen schließlich der angegebene Wert gemittelt wurde. Wird die Helligkeit nicht nachträglich verstärkt, zum Beispiel durch Einstellen des Gain-Wertes, so liegt diese Abweichung im Bereich von einer Graustufe. Erst höhere Gain-Werte erhöhen auch den Rauschpegel der Kameras merklich. In Tabelle 9 sind alle eben beschriebenen Ergebnisse zusammengefasst. Es zeigt sich dabei, dass auch „normale“ Kameratypen ohne spezielle NIR-Eignung durchaus für den Einsatz zur Infrarotaufnahme geeignet sind.

Eine Besonderheit der nach dem interlaced-Prinzip arbeitenden Kamera sollte in diesem Zusammenhang noch erwähnt werden. Im Grunde ist das Aufnehmen und Zusammensetzen von Halbbildern für die Aufnahme natürlicher und bewegter Objekte eher ein Nachteil, da sich Informationen in benachbarten Zeilen voneinander unterscheiden können. Durch gezielte Ansteuerung der Projektion ist es beim „timed-color-coded“-Verfahren dagegen denkbar, pro Halbbild der Aufnahme eines der beiden Muster darzustellen. Eine Berechnung der Tiefenkarte ist somit bereits nach einer einzelnen Vollbildaufnahme möglich, allerdings bei reduzierter Auflösung. Je nach Anwendung, die diese Daten verwertet, ist eine solche Objektrekonstruktion aber uneingeschränkt ausreichend. Denn die Anzahl der codierten Lichtebenen wird durch ein solches Vorgehen nicht reduziert, die erreichbare Tiefengenauigkeit bleibt also bestehen, die Anzahl der Messwerte wird halbiert. Beispielsweise zur Positionsbestimmung oder Detektion einer Objektrotation ist eine solche schnelle Tiefenberechnung vorstellbar.

4 Anwendungsgebiete der aktiven (infraroten) Stereoskopie

Wie in der Einleitung zu dieser Arbeit erwähnt, sind die Einsatzbereiche von Systemen zur dreidimensionalen Objekterfassung sehr vielseitig. Dieses Kapitel befasst sich mit drei Anwendungsbereichen, die vom Einsatz dreidimensionaler Objektdaten deutlich profitieren und bei denen im Verlauf dieser Arbeit beschriebene und entwickelte Verfahren zum Einsatz gebracht werden konnten. Die Bereiche umfassen dabei die biometrische Gesichtserkennung, den Einsatz zur Analyse animalischer Anatomie in der Biologie, sowie die Anwendung in multimedialen Umgebungen.

4.1 Biometrie und 3D-Gesichtserkennung

Da die Entwicklung eines infraroten Stereoskopiesensors von der Idee einer dreidimensionalen Gesichtserkennung motiviert wurde, nimmt diese als mögliche Applikation einen sehr hohen Stellenwert ein. Denn die Verwendung zusätzlicher dreidimensionaler Informationen bei einer automatischen Identifikation verspricht eine höhere Zuverlässigkeit und Sicherheit bei der Verifizierung von Gesichtsdaten als bei reinen zweidimensionalen Ansätzen ([Phillips et al. 2007]). Um sich diesen Informationsgewinn einer weiteren Dimension zunutze zu machen, kann für die eigentliche Gesichtserkennung nach verschiedenen Strategien vorgegangen werden.

Relativ naheliegend ist es, eine direkte Verifizierung auf Basis der dreidimensionalen Daten zu vollziehen. Hierbei werden aus einer oder mehreren Referenzaufnahmen eines Gesichts bestimmte Eigenschaften extrahiert, die dann bei einer Erkennung mit dem aktuellen Objekt verglichen werden. Das grundsätzliche Vorgehen ist dabei vergleichbar mit bekannten, zweidimensionalen Verfahren. Es werden allgemeine Bereiche oder charakteristische Punkte (sog. Feature-Points) definiert, deren Lage automatisch detektierbar sein muss. Zwischen diesen Punkten können dann Bedingungen definiert werden, die sich im Dreidimensionalen durch Abstände, Pfade oder Kurvenverläufe beschreiben lassen. Aller-

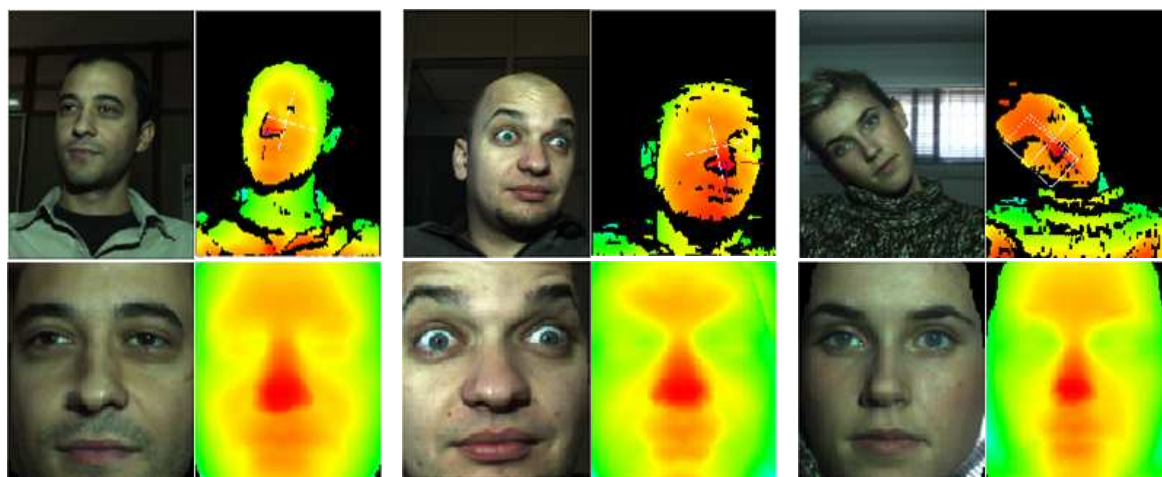


Abbildung 42 Lage- und Beleuchtungskorrektur. Verschiedene Positionen des Gesichts können durch Erfassen der 3D-Daten in eine Frontalansicht gebracht werden, wodurch die Texturdaten (links) und Tiefenkarten (rechts, Farbinterpretation der Tiefe) stets in ähnlicher Form zur Erkennung verwendet werden können (Bilder aus [Tsalakanidou 2006]).

dings können genau diese Merkmalsextraktionen für allgemeingültige Verfahren relativ rechenintensiv ausfallen, so dass eine echtzeitfähige Anwendung schwierig wird.

Eine Abhilfe schaffen in diesem Fall hybride Ansätze (siehe u.a. [Tsalakanidou et al. 2005], [Malassiotis and Strintzis 2005], [BioSec 2005], [Mpiperis et al. 2007]), die zwei- und dreidimensionale Daten kombinieren. Ein beispielhaftes Vorgehen dafür wird im Folgenden beschrieben.

Den hybriden Ansätzen liegt die Tatsache zugrunde, dass zweidimensionale Systeme zur Gesichtserkennung bereits sehr gute und zuverlässige Algorithmen bieten, die einen hohen Sicherheitsstandard gewährleisten. Die größten Schwierigkeiten der zweidimensionalen Ansätze liegen dabei noch in der Positionierung des Gesichts vor der Kamera bzw. der Drehung des Kopfes relativ zur Sichtebene. Aus einer variierenden Szenenbeleuchtung oder anderen sich ändernden äußeren Parametern können außerdem Kontrastunterschiede im Texturbild entstehen. Richtet man zusätzlich das Augenmerk auf die Fälschungssicherheit zweidimensionaler Verfahren, so können die aufzeichnenden Sensoren mancher Systeme durch einfache Fotos von Gesichtern getäuscht werden. All dies sind Faktoren, die unter Verwendung einer dreidimensionalen Gesichtsmodellierung berücksichtigt und damit korrigiert oder vermieden werden können.

Wird ein Datensatz zur Gesichtserkennung aufgenommen, so bestimmt man zunächst dessen Lage relativ zum Sensor. Dies geschieht allein auf Basis der 3D-Daten, so dass die initiale Lokalisation bereits unabhängig von Texturen oder Umgebungsbeleuchtungen durchgeführt wird. Hierfür wird die Nase detektiert, deren Spitze bis zu einer gewissen Verdrehung stets den nächsten Punkt zum Sensor darstellt. Ist dieser Punkt gefunden, so kann der Verlauf des Nasenrückens verfolgt werden, der bereits eine Achse zur Orientierung des Gesichts festlegt. Eine weitere Achse lässt sich über die Verbindung zwischen

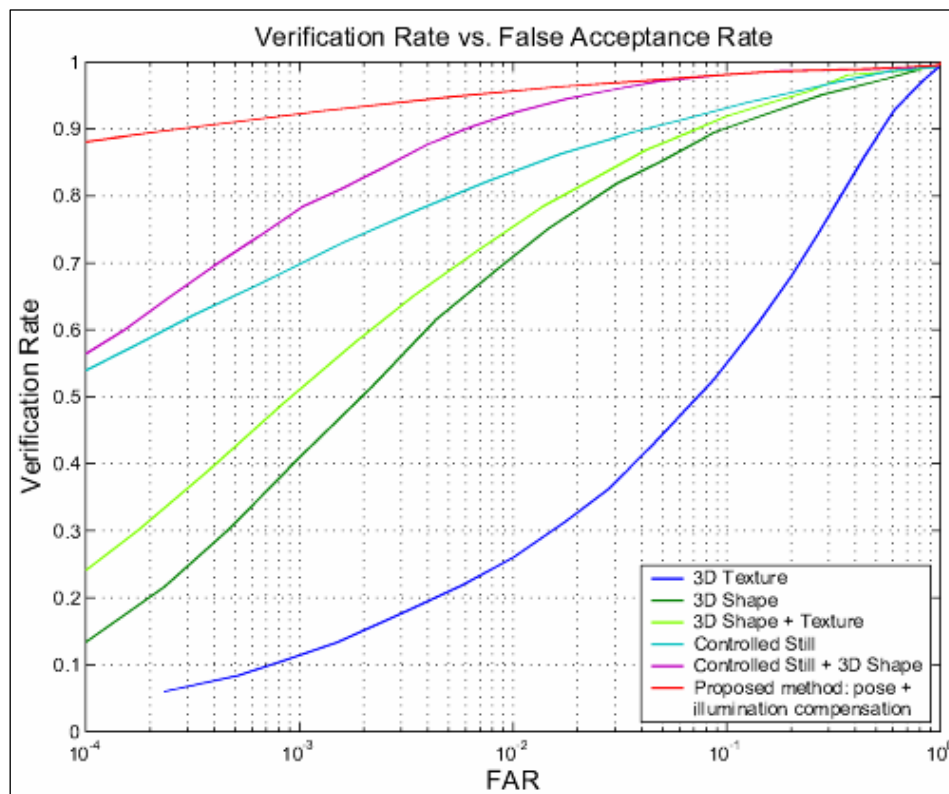


Abbildung 43 Vergleich verschiedener Ansätze zur Gesichtserkennung. Der rote Graph stellt den beschriebenen Ansatz im Vergleich zu den im FRGV [Phillips et al. 2007] vorgestellten Verfahren dar.

den Augenmittelpunkten definieren, wodurch die Gesichtsrotation mit hoher Zuverlässigkeit bestimmt werden kann. Werden diese Schritte bereits bei der Referenzaufnahme einer Person vollzogen, können vorhandene Unsymmetrien oder „Fehlstellungen“ im Gesicht berücksichtigt werden. Abbildung 42 zeigt Beispiele solcher Positions- und Rotationskorrekturen. Blickt nun also bei einer Verifizierung eine Person nicht optimal in die Kamera, so kann das 3D-Modell des Gesichts nachträglich problemlos in eben diese optimale Position rotiert werden. Dadurch erhält man dann ein ideal ausgerichtetes zweidimensionales Bild des Gesichts, was zur texturbasierten Gesichtserkennung verwendet werden kann und deren Robustheit bereits deutlich erhöht.

Durch die Analyse der Tiefenkarte des Gesichts lässt sich anschließend der Einfluss der Objektbeleuchtung kompensieren, so dass für alle Bilder unabhängig von der Blickrichtung und Orientierung stets gleiche Kontrastverhältnisse angewendet werden können. Hierfür wird angenommen, dass jede aufgenommene Person durch Fremdlicht primär aus einer bestimmten Richtung \underline{l} beleuchtet wird. Mehrere alternative Lichtquellen würden dagegen eine diffuse Beleuchtung ergeben, die weniger störende Schattierungen hervorrufen würde. Über die Schattierung des zweidimensionalen Texturbildes des Gesichts und die Daten der dreidimensionalen Tiefenkarte lässt sich die Richtung \underline{l} der Fremdlichtquelle hinreichend genau bestimmen. Zusammen mit der Richtung \underline{l}_0 der Projektionsbeleuchtung

Method	Frontal	Pose	Expression	Glasses	Illumination	Avg.
Texture	2.13 (99.68)	5.95 (97.93)	8.02 (97.91)	4.48 (98.00)	5.42 (97.43)	4.16 (98.29)
Depthmap	2.35 (99.76)	4.84 (98.62)	6.20 (99.00)	4.07 (98.25)	5.73 (97.43)	4.11 (98.76)
Hybrid	1.42 (99.92)	3.21 (99.54)	5.70 (99.27)	2.74 (98.50)	4.43 (98.15)	2.62 (99.28)

Tabelle 10 Ergebnisse einer hybriden Gesichtserkennung unter Laborbedingungen. Die Daten werden dabei nach Frontalansicht, gedrehter oder verkippter Ansicht, Gesichtsausdrücken, dem Tragen einer Brille und dem Einfluss von Fremdlicht unterschieden. Der obere Wert stellt die equal error rate (EER) dar, der Wert in Klammern die entsprechende Erkennungsrate. (Daten aus [Tsalakanidou 2006])

und einer gerenderten Tiefenkarte R mit konstantem Albedo (Rückstrahlvermögen) lassen sich schließlich die beleuchtungs-kompensierten Intensitätswerte \tilde{I}_C aus den nicht-kompensierten Farbbildwerten I_C und Tiefendaten I_D an jeder Bildposition \underline{u} gemäß folgender Formel bestimmen [Tsalakanidou and Malassiotis 2007]:

$$\tilde{I}_C(\underline{u}) = I_C(\underline{u}) \frac{R(I_D, \underline{l}, \underline{u})}{R(I_D, \underline{l}_0, \underline{u})} \quad (64)$$

Im Rahmen eines EU-Projekts [BioSec 2005] konnte ein beleuchtungskompensierendes, hybrides Verfahren bei einer Pilotinstallation zur Zugangskontrolle an einem Flughafen getestet werden. Hierfür wurde eine Datenbank mit 36 Personen erstellt. Die Datensätze zum Training des Erkennungsalgorithmus umfassten dabei verschiedene Ansichten, Positionen, Beleuchtungsszenarien und Gesichtsausdrücke, so dass sich insgesamt etwa 3000 Trainingsdatensätze ergaben. Eine allgemeingültige, größere Datenbank, wie sie beispielsweise mit der FERET-Datenbank für zweidimensionale Verfahren existiert, konnte nicht verwendet werden, da eine solche zum Evaluierungszeitpunkt nicht zur Verfügung stand und für den Zweck eines Feldversuchs zudem wenig geeignet wäre. Als 3D-Sensor kam eine farbcodierte Triangulation im sichtbaren Licht zum Einsatz, da die hier vorgestellte NIR-Variante noch nicht einsatzbereit war. Mit dem beschriebenen Verfahren konnte eine Erkennungsrate von 95% bei einer false acceptance rate (FAR) von 0,5% erreicht werden. Unter Laborbedingungen konnten noch bessere Werte erzielt werden, die in Tabelle 10 aufgelistet sind. Dabei wird deutlich, dass selbst auf hohem zweidimensionalen Erkennungsniveau eine Fusion von 2D- und 3D-Daten eine deutliche Verbesserung bewirken kann. Ein graphischer Vergleich der Methode mit anderen Verfahren ist in Abbildung 43 dargestellt, die den eben beschriebenen Ergebnissen die Resultate des „Face Recognition Grand Vendor Test“ (FRGV) [Phillips et al. 2007] gegenüberstellt. Dabei konnten ebenfalls durch die hybride Methode deutlich bessere Ergebnisse erzielt werden.

4.2 Biologische Analysen

Ein im weitesten Sinne der Biometrie ähnliches Teilgebiet ist die Analyse humaner oder animalischer Anatomie, auch wenn dies nicht als Primärziel dreidimensional abbildender Sensoren gesehen werden kann. Dabei können dreidimensionale Daten wichtige Aufschlüsse über den ontogenetischen Entwicklungsprozess verschiedener Gattungen, deren Abstammung oder deren Evolutionsprozess liefern.

Im Rahmen dieser Arbeit wurden zwei Projekte unterstützt, die aufgrund ihrer ungewöhnlichen Thematik im Umfeld technischer Sensoren durchaus Erwähnung finden sollten. Dabei handelt es sich zum einen um die Vermessung und Untersuchung des Rachenraums von Eidechsen [Ph-T35 2008]. Diese besitzen aufgrund ihres geringen Ohrenabstands einen speziellen Hörmechanismus – das so genannte Druck-Gradienten-Gehör ("pressure-gradient receiver"). Durch eine Kopplung der beiden Trommelfelle über kopfinterne Höhlen unterscheidet sich dieses Hörsystem grundsätzlich von dem der meisten Säugetiere. Zur weiteren Untersuchung ist dabei jedoch Kenntnis über die Geometrie des Rachenraums der Tiere notwendig, die sich bei lebenden – aber auch bei bereits verstorbenen – Tieren nahezu unmöglich durch direkte Messung gewinnen lässt. Nur mit einer dreidimensionalen Modellierung der Rachenhöhle lassen sich entsprechende Daten gewinnen, die für tiefgreifende Analysen geeignet sind. Hierfür wurden Kunststoffabgüsse des Rachenraums durch ein aktives Stereoskopieverfahren auf Basis der farbcodierten Triangulation dreidimensional vermessen. Die einzelnen Tiefenkarten wurden schließlich zu vollständigen, 360° umfassenden Rundummodellen zusammengefügt. Endgültige Ergebnisse dieser Analysen standen zum Ende dieser Arbeit nicht zur Verfügung, lassen sich aber in der aktuellen Forschung [Ph-T35 2008] weiter verfolgen.

Das andere Projekt befasst sich mit der Untersuchung der Entwicklung von Buntböcken (*Damaliscus pygargus pygargus*) [Burghardt and Wasilewski 2007]. Diese vom Aussterben bedrohte Gattung ist aus biologischer Sicht zur Entwicklungsanalyse besonders interessant, da sie durch einen genetischen Flaschenhals mit einer Population von lediglich 70 Tieren ging, auf denen die heutige Population von etwa 2300 Tieren beruht. Der Genpool dieser Population ist damit, verglichen mit anderen Gattungen, sehr eingeschränkt. Um die Zucht zu kontrollieren, Inzucht zu vermeiden und dadurch das Genmaterial zu stärken, ist es nötig, Tiere individuell unterscheiden zu können. Um aber das Verhalten wild lebender Tiere nicht zu beeinträchtigen, ist hierfür ein möglichst einfaches Verfahren nötig, das berührungslos und ohne Markierung der Tiere auskommt. Da die Forschung gezeigt hat, dass sich die Tiere in der Struktur ihrer Hörner so deutlich unterscheiden, wie wir Menschen es beispielsweise an unseren Fingerabdrücken tun, empfiehlt sich eine Vermessung der Hörner zur Identifikation. Um hierfür erste Daten zu sammeln, wurden die Hörner und Schädel verstorbener Tiere durch aktive Stereoskopie gescannt und zu dreidimensionalen Modellen zusammengesetzt, die das Erstellen eines allgemeinen Modells zur Identifikation eines Tieres erlauben. Abbildung 44 zeigt Beispiele dieser Aufnahmen, an denen die

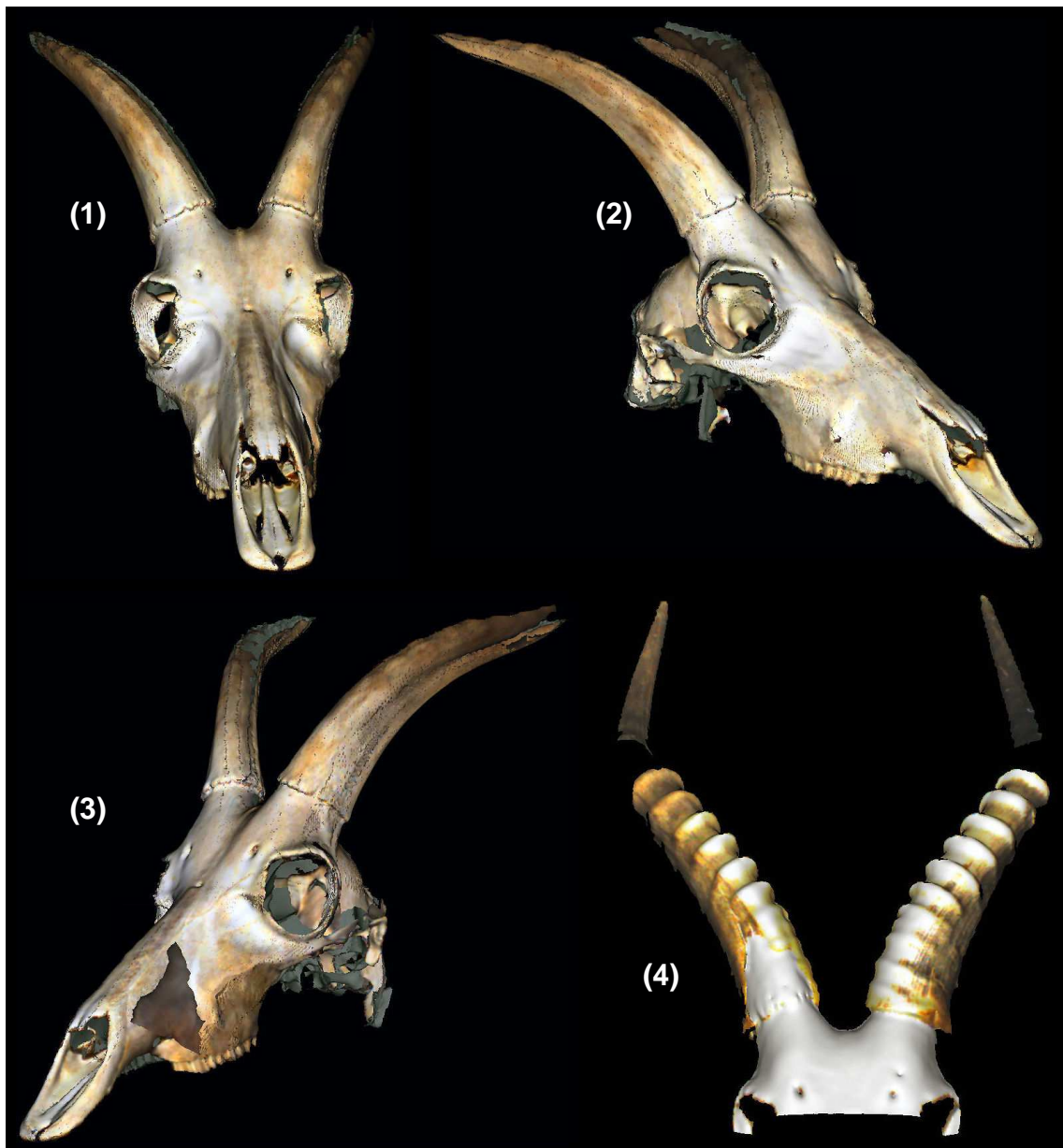


Abbildung 44 Dreidimensionale Modelle von Buntbock-Schädeln. Die Bilder zeigen das Rundummodell eines Schädels (1)-(3), sowie die Frontalaufnahme der Hörner eines anderen Schädels (4). Anhand der Struktur der Hörner (4) können Tiere individuell identifiziert werden.

Unterschiede der Hornstrukturen bereits deutlich sichtbar werden. Die Auswertung und Modellierung dieser Daten dauert weiter an, weitere Ergebnisse dieser Arbeiten finden sich auch in [Burghardt and Wasilewski 2007].

4.3 Multimediale Anwendungen

Zu den multimedialen Anwendungen eines Sensors wird hier allgemein der Einsatz als Mensch-Maschine-Schnittstelle außerhalb biometrischer Sicherheitsanwendungen gezählt. Dies umfasst ein weites Themenspektrum, das vom Einsatz zur Modellierung menschlicher Gesichtsstrukturen, beispielsweise für die Schaffung virtueller Avatare, über die Analyse menschlicher Emotionen aus dreidimensionalen Daten bis hin zu neuen Informationskanälen für netzwerkgebundene Kommunikation führt.

Mit diesem gesamten Themenspektrum befasst sich unter anderem das Forschungsprojekt „Psychologically Augmented Social Interaction Over Networks“ [PASION 2008], dessen Zielsetzung es ist, die Effizienz der Kommunikation innerhalb verschiedener Gruppen zu steigern und die Arbeit durch Bereitstellen „sozialer“ Netzwerke zu erleichtern. Hierfür werden verschiedene Ansätze untersucht, die neben sogenanntem „Social Gaming“ auf die Kommunikation unter Bereitstellung vielfältiger Informationen hinauszielen. Dabei sind Szenarien angedacht, die traditionelle audio-visuelle Kommunikation mit neuen Medien verschmelzen, so dass man beispielsweise Daten über die aktuelle Stimmung auf Basis von Emotionserkennung übermitteln kann. Auch die vollständige Repräsentation durch ein virtuelles Ebenbild ist eine angestrebte Modalität. Die Aufgabe der 3D-Sensorik ist es in diesem Fall, als Interface zur Emotionserkennung zu dienen, aber auch Modelle für virtuelle Avatare bereitzustellen.

Ganz allgemein bewirkt der Einsatz virtueller, menschenähnlicher Avatare auch eine Effizienz-Steigerung von Hilfs- und Informationsanwendungen, da sie ein natürliches Kommunikationsverhalten des Nutzers provozieren (z.B. [Jung and Kopp 2003]). Dabei hängt die Akzeptanz allerdings in großem Maße vom natürlichen Eindruck des virtuellen Charakters ab. Eine genaue Modellierung menschlicher Gesichtszüge und Mimik durch akkurate Sensorerfassung ermöglicht diese Illusion und erleichtert es, Avatare zu kreieren.

Auch für die immer weiter zunehmende Personalisierung der eigenen Erscheinung im Internet, beispielsweise in virtuellen Umgebungen wie „Second Life“ (© Linden Research Inc., <http://secondlife.com>), erlauben dreidimensionale Modellierungen der eigenen Person eine Annäherung der virtuellen an die reale Erscheinung. Dies ist auch denkbar für Spiele oder ähnliche Applikationen, in denen man in die Rolle anderer Personen schlüpfen möchte, die dabei dem eigenen Charakter oder Auftreten ähneln soll.

Das Einsatzspektrum dreidimensionaler Sensoren umfasst also viele Bereiche. Dabei eröffnen vor allem die Einfachheit der Datenerfassung farbcodierter Verfahren und die berührungs- und störungsfreie Methode der NIR-Verfahren zahlreiche Möglichkeiten, Sicherheit zu erhöhen, Messaufgaben zu erweitern oder ganz allgemein die Nutzerfreundlichkeit und den Komfort etlicher Anwendungen zu verbessern.

5 Erweiterte Ansätze zur aktiven 3D-Objekterfassung

Das Hauptziel dieser Arbeit, das auf einer Verbesserung und Vereinfachung der dreidimensionalen Sensorik für biometrische Anwendungen lag, konnte durch die in Kapitel 3 dargestellten Verfahren erreicht werden. Dabei wurden im Verlauf der Entwicklung auch weitere Möglichkeiten und Erweiterungen zur aktiven Stereoskopie im sichtbaren und infraroten Wellenlängenbereich untersucht. Die Überlegungen zu diesen Erweiterungen und die entwickelten Ansätze werden in diesem Kapitel vorgestellt.

5.1 Dynamische NIR-Stereoskopie

Aus den verbliebenen technischen Schwierigkeiten der vorgestellten NIR-Verfahren mit statischer Musterprojektion geboren ist die Idee eines dynamischen Verfahrens der infraroten dreidimensionalen Objekterfassung. Dabei rechtfertigt die zu erwartende höhere Flexibilität der Projektion den eventuell notwendigen, höheren Hardwareaufwand.

Betrachtet man bestehende Verfahren, die mit verschiedenen statischen Mustern arbeiten, so fallen in erster Linie drei nachteilige Punkte auf – eine Reduktion der effektiven Lichtleistung durch den Einsatz von dichroitischen Spiegeln oder Strahlteilern, der notwendige Einbau von Marken zur Sensorkalibrierung in das Projektionsmuster und die exakte Ausrichtung der einzelnen Muster zueinander. All dies sind Punkte, die beim Einsatz eines handelsüblichen Videoprojektors keine große Rolle spielen, da dieser die Projektion beliebiger Muster und Marken erlaubt und eine korrekte Musterausrichtung, auch bei mehrsensorigen Projektoren, durch den Fertigungsprozess sichergestellt ist. Auf Basis der Videoprojektionstechnik wird nun der Entwurf einer dynamischen Projektionseinheit für die NIR-Stereoskopie vorgestellt. Als Grundlage dienen hierbei bekannte und verbreitete Projektionstechniken, also LCD/TFT, DLP/DMD oder LCoS-Pannels. Da sich die Techniken grundlegend unterscheiden, werden die Arbeitsprinzipien und ihre Tauglichkeit zur Infra-

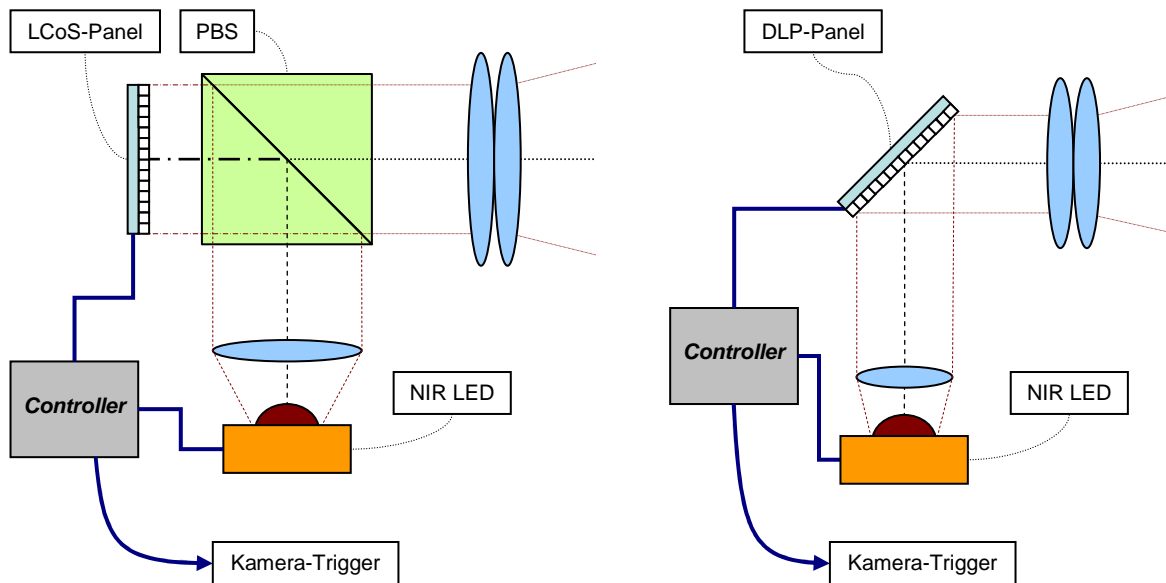


Abbildung 45 Projektionseinheiten mit LCoS- und DLP-Technik. Die linke Seite zeigt den Aufbau einer LCoS-Projektion, bei der das Licht durch einen „polarizing beam splitter“ (PBS) auf den LCoS-Sensor gelenkt wird, wo das Bild erzeugt und zur Projektion reflektiert wird. Die rechte Seite zeigt eine DLP-Projektion, bei der das Licht direkt auf den DLP-Sensor gelenkt wird, der Intensitäten pro Pixel zeitlich regelt und das Licht direkt auf das Objektiv reflektiert.

rotprojektion im Folgenden kurz erläutert. Da eine Farberzeugung für die geplante Anwendung eher von nebensächlichem Interesse ist, wird hierauf nicht näher eingegangen.

Die LCD-Technik (auch als TFT-Display, Thin Film Transistor, bekannt) ist die älteste zur digitalen Projektion eingesetzte Technik. Sie funktioniert ähnlich einem Diaprojektor, wobei das durchstrahlte Medium aus gitterförmig angeordneten Flüssigkristallzellen besteht. Dies erlaubt ein punktgenaues Einstellen der Intensität des durchstrahlenden Lichts durch Beeinflussung seiner Polarisationsrichtung. Durch einen bestimmten Schaltzustand können also beliebige Graustufenbilder dargestellt werden. Allerdings können die eingesetzten Flüssigkristalle nach einer gewissen Betriebsdauer „ausbleichen“, was besonders durch den UV- und Infrarotanteil des durchstrahlenden Lichts hervorgerufen wird. Da dies jedoch vor allem bei hohen Lichtleistungen und langer Betriebsdauer passiert, ist dieser Effekt nicht zwingend als Einschränkung zu betrachten. Größere Auswirkungen dagegen hat die Art und Weise der elektronischen Verschaltung auf der Sensorfläche. Da nur die Flüssigkristallschicht durchstrahlt wird, die Elektronik dagegen aber lichtundurchlässig ist, erscheint das Panel als gekachelte Struktur aus durchlässigen Pixeln und undurchlässigen Zwischenräumen. Damit ergibt sich lediglich ein Füllfaktor des Sensors von bis zu 70%, welcher sich durch eine ausgeprägte Gitterstruktur in der Projektion äußert, die wiederum die Anfälligkeit gegenüber Sampling-Artefakten erhöhen kann (siehe auch 3.2.2). Weitere Informationen und Verweise zur LCD-Projektionstechnik finden sich unter anderem in [Klußmann and Wiegelmann 2005].

Mit einem gänzlich anderen Konzept arbeiten DLP-Projektoren, deren Bildwandler ein so genanntes DMD, Digital Micromirror Device, darstellt. Dies ist ein integrierter Schaltkreis, auf dem jeder einzelne Bildpunkt durch einen kleinen Spiegel realisiert wird, der durch elektrische Impulse zwischen zwei festen Endzuständen hin und her gekippt werden kann. Diese beiden Zustände sind binär zu betrachten – ein Pixel ist also entweder „an“ oder „aus“, das einfallende Licht wird gänzlich in Projektionsrichtung reflektiert oder von dieser abgelenkt. Eine Graustufendarstellung ist nur durch den schnellen Zustandswechsel der Spiegel möglich, der bei heutigen Panels (Stand 2007/2008) eine Schaltfrequenz von etwa 16kHz erlaubt. Die gezielte zeitliche Ansteuerung der einzelnen Punkte ermöglicht dabei eine direkte, linear quantisierte Intensitätssteuerung in der Projektion bei einem hohen Kontrastverhältnis. Die Technik erlaubt zudem einen relativ hohen Füllfaktor von bis zu 90%, da die Ansterelektronik unter den Spiegeln liegt. Lediglich die Zwischenräume, die zur Verkipfung nötig sind, bewirken eine leichte Gitterstruktur, die aber deutlich geringer ausfällt als bei LCD-Projektionen. Da die Projektion durch Reflektion erfolgt, ist es möglich, nahezu beliebige Wellenlängen zu verwenden, Alterungseffekte der Sensoroberfläche oder betriebsdauerabhängige Lichtabschwächungen sind bisher nicht bekannt. Deshalb ist diese Technik auch besonders gut für eine NIR-Projektion geeignet. Die DLP-Technik wurde von Texas Instruments entwickelt, wo sich auch weiterführende Literatur zu dieser Projektionstechnik findet ([TI 2005]).

Eine Art Kombination der eben beschriebenen Techniken stellen LCoS-Projektoren dar. Auf einer reflektierenden Oberfläche ist dabei eine Flüssigkristallschicht aufgebracht, die für jeden Bildpunkt eine Intensitätsregelung durch Polarisationsänderung ermöglicht. Die Darstellung bzw. Erzeugung von Graustufenbildern geschieht also analog zur LCD-Technik mit dem großen Unterschied, dass sämtliche Ansteuerungselektronik unterhalb der einzelnen Flüssigkristall-Pixel angebracht ist. Möglich ist dies, da die Schicht aufgrund der reflektiven Arbeitsweise nicht durchstrahlt werden muss. Dadurch ist ein sehr hoher Füllfaktor der Sensoroberfläche möglich (mehr als 90%), Gitterstrukturen sind im erzeugten Muster kaum zu erkennen. Aufgrund dieser Eigenschaften sind LCoS-Panels ebenfalls zum Aufbau einer neuen Projektionseinheit geeignet – allerdings nur im direkt an das Sichtbare anschließenden, nahen Infrarotbereich, da die Spezifikationen der Flüssigkristalle und der Oberflächenreflektanz gemäß der technischen Beschreibungen nur für eben dieses Spektrum ausgelegt sind. Weitere Spezifikationen und Informationen zu dieser Projektionstechnik finden sich unter anderem in [Holo 2008].

Mit den beschriebenen Techniken der LCoS- und DLP-Projektion ist es also möglich, eine Projektionseinheit aufzubauen, die für einen aktiven Stereoskopie-Sensor geeignet ist. Abbildung 45 zeigt die dabei nötigen Anordnungen für beide Methoden. Nach diesen Schemata konnten unter Einsatz herstellerepezifischer Platinen Prototypen für Projektionseinheiten aufgebaut werden. Dabei bestätigen die resultierenden Bilder die Angaben bezüglich der Geschwindigkeit, des Füllfaktors und der Lichtausbeute. Die Herstellung und Konstruktion eines 3D-Sensors unter Verwendung einer dynamischen Projektion ist

damit zwar aufwändiger als beim Einsatz statischer Muster, dafür erhält man aber einen hohen Grad an Flexibilität bei der Musterdarstellung, der den Einsatz vieler verschiedener Ansätze des strukturierten Lichts erlaubt.

5.2 Hybride intensitätsbasierte Laufzeitmessung

Mit aktuellen und weiterentwickelten Projektionstechniken ist es auch möglich, Verfahren der aktiven Stereoskopie, die in früheren Entwürfen an technischen Einschränkungen der Projektionstechnik litten, sowohl für sichtbare, als auch für infrarote Beleuchtung nutzbar zu machen. Auch die weiterentwickelte Kameratechnik mit schnelleren und empfindlicheren Sensoren bringt dabei weitere Vorteile. Dieser Abschnitt beschäftigt sich mit einer Erweiterung der unter Abschnitt 2.5.3.4 beschriebenen direkten Szenencodierung durch Intensitätsverläufe. Es wird ein Verfahren beschrieben, das eine hochauflösende Szenencodierung zur Berechnung von Tiefendaten erlaubt.

Ausgangspunkt ist dabei das zu projizierende Lichtmuster, welches als „Graukeil“ bezeichnet wird, da es einen linearen Intensitätsabfall von Weiß nach Schwarz darstellt. Die besondere Art der Codierung bringt dabei den großen Vorteil mit sich, dass jede Position unabhängig von benachbarten Bildpunkten eindeutig definiert ist. Bei ersten Vorstellungen solcher Verfahren wurden hierfür statische Dias oder LCD-Projektoren verwendet, bei denen in beiden Fällen jedoch wegen der Art der Mustererzeugung von keiner Linearität der Grauwertverteilungen ausgegangen werden kann. Diese Einschränkung lässt sich beispielsweise durch den Einsatz eines DLP-Panels zur Musterprojektion beheben. Denn durch diese Technik kann für jeden Bildpunkt die Projektionszeit exakt eingestellt werden, wodurch ein linearer Anstieg der Helligkeit über die gesamte Bildauflösung sichergestellt werden kann. Im Folgenden wird dieser Graukeil nun als eine Art Lichtvorhang interpretiert, der innerhalb seiner Projektionsdauer T „aufgezogen“ wird. Ein heller Punkt des Projektionsmusters wird hierbei zeitlich gesehen vor einem dunkleren Punkt projiziert und für eine längere Zeit beleuchtet. Besteht ein Graukeil also aus N Graustufen, so beginnt jede einzelne dieser N Beleuchtungen zum Zeitpunkt $t_i \in [t_0; t_N]$ (vgl. Abbildung 46).

Jede Projektionsspalte – oder Zeile, je nach Ausrichtung des Musters – ist demnach eindeutig über ihre Intensität codiert und kann für eine Triangulationsberechnung herangezogen werden. Hieraus resultiert auch die grundsätzlich mögliche hohe Auflösung der Tiefenkarte. Üblicherweise werden nach der Projektion des Graukeils und eventuell weiterer Referenzmuster die Lichtebenen analysiert und in eine Tiefenberechnung überführt. Die technische Realisierbarkeit einer solchen Decodierung zur Tiefenberechnung wurde bereits mehrfach nachgewiesen (u.a. [Carrhill and Hummel 1985], [Miyasaka et al. 2000] oder [Miyasaka and Araki 2002]). Wird nach einer Decodierung des Projektionsmusters die Tiefenberechnung jedoch durch Triangulation durchgeführt, so fallen für jeden Bildpunkt mehrere Berechnungsschritte an. Da dies relativ zeitintensiv ist, sind intensitätsbasierte Codierungen bisher kaum für echtzeitfähige Anwendungen geeignet. Aus diesem

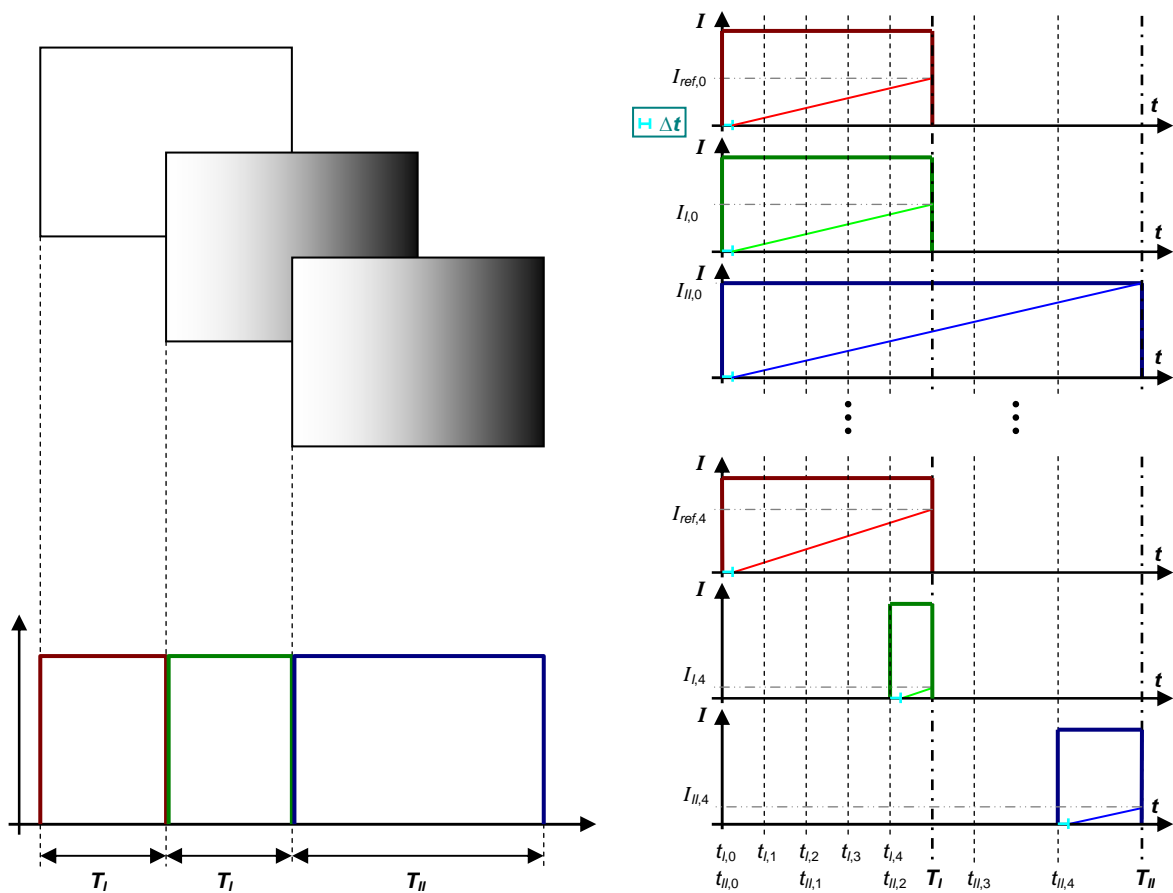


Abbildung 46 Direkte Szenencodierung durch Projektion einer Intensitätsverteilung. Nach der Projektion einer Referenzbeleuchtung der Dauer T_I wird ein Graukeil für verschiedene Zeiten T_I und T_{II} projiziert (links), so dass sich für jeden Bildpunkt und jede Musterposition andere Intensitätswerte einstellen. Die rechten Diagramme zeigen dies exemplarisch für einen Graukeil aus 5 Intensitätsstufen ($[0 \dots 4]$). Die verschiedenen Steigungen der Sensorantworten repräsentieren unterschiedliche Oberflächentexturen. Die oberen Diagramme stellen Pixel im linken, helleren Graukeilbereich dar, äquivalent der Stufe 0 und den Projektionszeitpunkten $t_{I,0}$ und $t_{II,0}$. Die unteren entsprechen rechten, dunkleren Bereichen zu Zeitpunkten $t_{I,4}$ bzw. $t_{II,4}$. Dies erlaubt eine Tiefenberechnung in jedem Punkt der beleuchteten Szene.

Grund wird im Folgenden ein Verfahren vorgestellt, das eine alternative, schnellere Entfernungsberechnung auf Basis einer Intensitätscodierung ermöglicht. Es basiert zunächst auf dem klassischen Vorgehen der Positionsdetektion innerhalb des Projektionsmusters, welche vorab beschrieben wird.

Zur Minderung des Rauschens und zur Erhöhung der Robustheit der Intensitätsmessung für eine genauere Positionsdetektion werden in der Regel mehr als eine einzelne Musteraufnahme benötigt. Oft wird dafür eine zusätzliche unbeleuchtete Aufnahme oder ein gegenläufiger Graukeil verwendet. Das hier angewendete Verfahren arbeitet mit drei Aufnahmen, wie sie in Abbildung 46 dargestellt sind. Für die Projektion und die damit verbundene Aufnahme werden zwei Belichtungszeiten T_I und T_{II} in der Art gewählt, dass T_I idealerweise die Hälfte von T_{II} beträgt. Hinreichend ist aber die Bedingung, dass T_I kleiner

als T_{II} ist. Die Szene wird zunächst für die Dauer T_I mit voller Beleuchtung, also „Weiß“, projiziert, im Anschluss erfolgt eine Graukeilaufnahme mit T_I und schließlich ein Graukeil der Dauer T_{II} . Die rechte Seite von Abbildung 46 zeigt die resultierenden Sensorantworten für Positionen am linken und rechten Ende des Musters. Die jeweiligen Zeitfenster der Projektionsmuster sind zur besseren Unterscheidung in der Grafik durch Farben gekennzeichnet. Der Zeitversatz Δt zwischen Musterprojektion und Pixelantwort entspricht der Laufzeit zwischen dem Sensor und dem Objekt, die abhängig von der Belichtungszeit, der Objektentfernung und der Kameraempfindlichkeit vorerst vernachlässigt werden kann. Die verschiedenen Steigungen der Intensitätsverläufe stehen symbolisch für verschiedene Texturen der Oberfläche. Durch die Referenzaufnahme entfällt nämlich im Gegensatz zu ursprünglichen Verfahren die Annahme einer homogenen Struktur.

Die gewählte Projektionsfolge der Muster unterstützt bereits die später beschriebene neue Berechnungsmethode. Sie ist dabei inspiriert von der in Abschnitt 2.3.2 beschriebenen MDSI-Technik, jedoch mit dem Ziel einer flächendeckenden, hochauflösenden und durch Kombination verschiedener Techniken robusten und genauen Tiefenberechnung. Das erste, weiß beleuchtete Muster dient dazu, die Sensorantwort jedes einzelnen Bildpunkts der Kamera zu bestimmen und eine Referenz für die Intensitätsmessung zu erhalten, so dass der Einfluss von wechselnden Texturen minimiert werden kann. Der lineare Anstieg m_i der Intensität I eines Bildpunkts i berechnet sich dabei unter Verwendung der aufgenommenen Referenzintensität $I_{ref,i}$ zu:

$$m_i = \frac{I_{ref,i}}{T_I} \quad (65)$$

Mit dem Wissen dieses Wertes erlauben es die Aufnahmen der beiden Graukeilbilder, den jeweiligen Projektionszeitpunkt $t_{I,i}$ und $t_{II,i}$ und damit verbunden die Position p_i auf dem Projektionssensor zu berechnen. Da der Intensitätswert eines Kamerapixels $I_{I,i}$ bzw. $I_{II,i}$ mit der Belichtungszeit T_I oder T_{II} der Beziehung

$$I_{I-i} = m_i \cdot \left(T_I - \frac{p_i T_I}{N} \right) \quad \text{und} \quad I_{II-i} = m_i \cdot \left(T_{II} - \frac{p_i T_{II}}{N} \right) \quad (66)$$

gehört (N steht hierbei für die Anzahl der codierten Zeilen/Spalten), kann die Position, bzw. die entsprechende Zeilen- oder Spaltennummer, durch folgende Berechnung direkt bestimmt werden:

$$p_i = N \cdot \left(1 - \frac{I_{I,i}}{m_i T_I} \right) \quad \text{und} \quad p_i = N \cdot \left(1 - \frac{I_{II,i}}{m_i T_{II}} \right) \quad (67)$$

Alternativ – oder als zusätzliche Verifizierung – erlauben die verschiedenen Aufnahmezeiten der Intensitätsverteilungen eine weitere Berechnung der Position durch gleichzeitige Betrachtung der beiden resultierenden Intensitäten $I_{I,i}$ und $I_{II,i}$, so dass sich die Position p_i auch folgendermaßen berechnen lässt:

$$p_i = \frac{N}{m_i} \cdot \left(m_i - \frac{I_{II,i} - I_{I,i}}{T_{II} - T_I} \right) \quad (68)$$

Wurden nun durch eine der Gleichungen (67) oder (68) auf der beleuchteten Oberfläche die Positionen im Projektionsmuster bestimmt, ist es möglich, für jeden Bildpunkt eine Triangulationsberechnung durch den Schnitt einer Lichtebene und einer Sichtgeraden durchzuführen, wie es für Lichtschnittverfahren üblich ist. Wie bereits beschrieben erfordern die hierfür durchzuführenden Schritte – Aufbau einer Ebene, Koordinatenberechnung und Schnittpunktbestimmung – mehrere Rechenoperationen. Die hier neu vorgestellte Methode der Tiefenberechnung stellt dagegen die vorerst theoretische Betrachtung dar, einen eigentlich Intensitäten messenden Sensor, die Kamera, zur Messung von Laufzeiten zu verwenden. Damit ist eine schnellere Tiefenberechnung möglich, da sich die Zahl der notwendigen Berechnungen deutlich reduziert.

Orientiert man sich an den laufzeitbasierten Verfahren aus Abschnitt 2.3, so beginnt die eigentliche Bildaufnahme, also die Sensorantwort auf die Projektion, erst nach der bisher vernachlässigten Zeit Δt , die dem Zurücklegen der doppelten Objektentfernung bei Lichtgeschwindigkeit entspricht. Abbildung 46 enthält diese Komponente bereits in ihrer schematischen Darstellung. Die Gleichungen (65)-(68) haben diese Komponente bisher nicht berücksichtigt, was einen leichten Fehler der Streifennummerierung im Nachkommabereich verursacht. Da aber diese Nummern stets ganzzahlig sein müssen und die Berechnung mehrfach verifiziert und abgeglichen werden kann, ist eine solche Vernachlässigung zulässig. Berücksichtigt man nun die Laufzeit exakt, so ändert sich Gleichung (66) zu:

$$\tilde{I}_{I,i} = m_i \cdot \left(T_I - \left(\frac{p_i T_I}{N} + \Delta t_i \right) \right) \quad \text{und} \quad \tilde{I}_{II,i} = m_i \cdot \left(T_{II} - \left(\frac{p_i T_{II}}{N} + \Delta t_i \right) \right) \quad (69)$$

Mit den Projektionszeitpunkten $t_{I,i}$ und $t_{II,i}$ anstelle der Positionsdaten p_i ändert sich diese Formulierung zu:

$$\tilde{I}_{I,i} = m_i \cdot (T_I - (t_{I,i} + \Delta t_i)) \quad \text{und} \quad \tilde{I}_{II,i} = m_i \cdot (T_{II} - (t_{II,i} + \Delta t_i)) \quad (70)$$

Damit lässt sich die Laufzeit Δt_i eines individuellen Bildpunkts i anhand der gemessenen Intensitäten $I_{I,i} \equiv \tilde{I}_{I,i}$ und $I_{II,i} \equiv \tilde{I}_{II,i}$ bestimmen:

$$\Delta t_i = \frac{I_{II,i}(T_I - t_{I,i}) - I_{I,i}(T_{II} - t_{II,i})}{(I_{II,i} - I_{I,i})} \quad (71)$$

Die Entfernung d_i eines einzelnen projizierten Pixels berechnet sich dann, wie bereits durch Abschnitt 2.3.1 und Gleichung (27) bekannt, zu:

$$d_i = \frac{c}{2} \Delta t_i \quad (72)$$

Die Halbierung beruht auf dem Zurücklegen des doppelten Weges durch das Licht.

Eine Tiefenberechnung ist damit flächendeckend möglich, jedoch hängt die Genauigkeit von den Belichtungszeiten und der Farbtiefe des Kamerasensors ab. Denn je kürzer die Belichtungszeit eines Bildes, desto größer sind die Auswirkungen der Laufzeitunterschiede und desto zuverlässiger die Berechnung. Gleiches gilt für die Farbtiefe, da ein größerer Intensitätsumfang wiederum eine genauere Abgrenzung der einzelnen Werte voneinander erlaubt.

Diese beiden Punkte werden durch die technischen Parameter der eingesetzten Komponenten begrenzt. Die minimale Belichtungszeit T_{\min} wird weniger von der Kamera, sondern vielmehr vom Projektionssensor und dessen Projektionsgeschwindigkeit bestimmt. Möchte man einen DLP-Chip zur Projektion verwenden, so müssen während einer Belichtung N Schaltvorgänge ausgeführt werden, um ein Muster mit N Lichtebenen zu codieren. Für einen Sensor mit einer Auflösung von 1024×768 Bildpunkten entspräche dies bei einer Belichtungszeit von 40ms einer Schaltfrequenz von $25,6\text{kHz}$, was sich bereits dem Maximum der kommenden Sensorgeneration (ca. 32kHz , Stand 2007/2008) nähert. Für eine reine laufzeitbasierte Messung ließe sich in diesem Fall die Musterauflösung reduzieren, da diese keinen direkten Einfluss auf die Auflösung der Tiefendaten hat, die für jeden Pixel allein auf Intensitätsbasis vollzogen wird. Für eine Triangulation hingegen wird das Lichtmuster in Ebenen codiert, was bei einer Reduktion der Musterauflösung auch zu einer reduzierten Tiefenauflösung führt. Die Genauigkeit der Tiefendaten hängt dagegen bei einer intensitätsbasierten Messung von der Kamera und deren maximaler Farbtiefe ab.

Derzeitige Massenprodukte erlauben die Aufnahme mit einer Farbtiefe von bis zu 12bit, was in einem Bild 4096 Graustufen entspricht. Da diese nur diskret ausgegeben werden können, ist dieser Wertebereich für eine Laufzeitmessung unzulänglich. Es müssten also spezialisierte Kameras verwendet werden, die eine höhere Farbtiefe liefern oder das direkte Auslesen der Sensorantwort erlauben, beispielsweise in analoger Form als anliegende Spannung an einem Bildpunkt. Auch empfindlichere Sensoren, die kürzere Belichtungszeiten erlauben, erhöhen die Genauigkeit der Berechnung.

Dieses neue Berechnungsverfahren zeigt also, dass in der Theorie intensitätsbasierte Verfahren durchaus echtzeitfähig für eine hochauflösende stereoskopische Objekterfassung eingesetzt werden können. Die technische Umsetzung hängt dabei von den verfügbaren Hardwarekomponenten ab, so dass derzeit lediglich eine „einfache“ Intensitätsmessung und Positionsbestimmung gemäß den Gleichungen (67) und (68) und eine Tiefenberechnung durch Triangulation möglich ist. Aber die kommenden Generationen von Projektions- und Aufnahmesensoren könnten durchaus Alternativen für eine schnellere und ebenso robuste Tiefenberechnung auf Intensitäts- und Laufzeitbasis liefern.

5.3 Radiale Lichtcodierung

Eine weitere Einschränkung eines aktiven Stereo-Systems aus einer Kamera und einer Projektionseinheit ist, dass Objekte nicht flächendeckend aufgenommen werden können. Durch den Triangulationswinkel und die Abbildungseigenschaften der Objektive können Verdeckungen auftreten, die aufgrund ihrer geometrischen Eigenschaften ganze Bereiche von einer Tiefenberechnung ausschließen. Dieses Kapitel stellt Ansätze vor, die sich der Lösung dieser Problematik annehmen.

5.3.1 Grundsätzliche Überlegungen

Ein aktives Stereosystem besteht in der Regel aus der Kombination einer Kamera und einer Beleuchtungseinheit in Form eines digitalen Videoprojektors oder eines statischen Musters. Betrachtet man sowohl das abgestrahlte Muster, aber auch analog den Erfassungsbereich der Kamera, als geometrische Körper, so entsprechen diese der Form von Pyramiden mit rechteckiger Grundfläche, deren Spitze im optischen Zentrum der Abbildung liegt. Nimmt man Abstand von der Abbildung einer rechteckigen Sensorfläche und betrachtet allein die Abbildungseigenschaften der Objektive, so wird jeweils innerhalb der Grenzen eines Kegels projiziert und aufgenommen.

Da zur dreidimensionalen Erfassung nach den Grundlagen der Triangulation ein Objekt stets von beiden optischen Systemen erfassbar sein muss – also entweder Kamera und Kamera oder Projektor und Kamera – verhindern Verdeckungen oder Abschattungen eine vollständige Aufnahme. Abbildung 47 verdeutlicht dies anhand der Aufnahme einer Kugel, durch deren Oberflächenbeschaffenheit Oberflächennormalen verschiedener Orientierungen ohne Verdeckungen simuliert werden können. Die effektiv abtastbare Fläche ist dabei nicht die gesamte Oberfläche einer Halbkugel, sondern hängt von der relativen Position von Projektor und Kamera zueinander und von deren Abstand zum Objekt ab. Da sowohl das Projektionsfeld als auch das Aufnahme-feld tangentialen Kegeln auf der Kugel entsprechen, berechnet sich die für 3D-Erfassungen nutzbare Fläche durch den Schnitt zweier Kegel mit einer Kugel.

Die resultierende Fläche eines solchen Schnitts ist ein sogenanntes Kugelzweieck, dessen Inhalt sich durch folgende Formel berechnen lässt:

$$A = 2 \cdot \frac{\varepsilon}{180^\circ} \cdot \pi \cdot r^2 \quad (73)$$

Diese Gleichung interpretiert das Kugelzweieck als zwei kongruente sphärische Dreiecke. Deren sphärischer Exzess ε , der zur weiteren Berechnung benötigt wird, lässt sich durch den Halbumfang s ausdrücken:

$$s = \frac{a + b + c}{2} \quad (74)$$

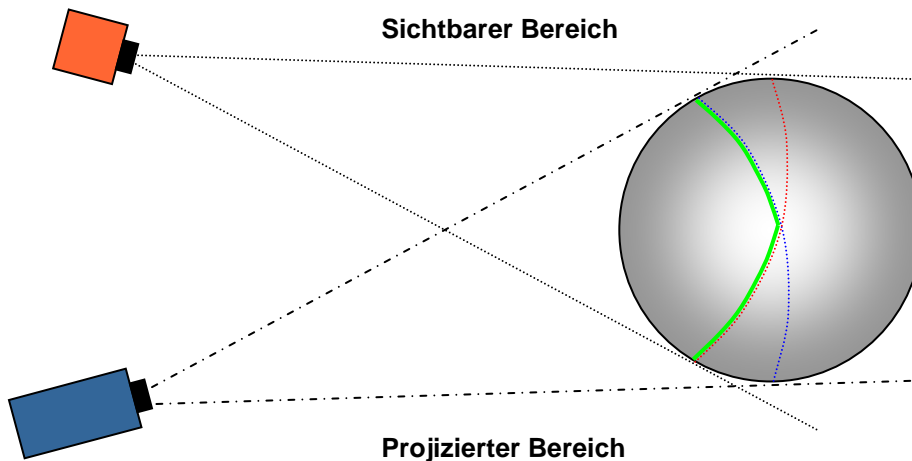


Abbildung 47 Erfassungsbereich eines aktiven Stereosystems. Der Schnitt des Projektionsbereichs (blau) mit dem sichtbaren Bereich (rot) ergibt den effektiven Erfassungsbereich (grün), der mit einer solchen Anordnung aufgenommen werden kann.

Hierbei stehen a , b und c für die Seiten des Kugeldreiecks. Der Exzess ε wird daraus über die l'Huiliersche Gleichung bestimmt:

$$\varepsilon = 4 \cdot \arctan \left(\sqrt{\tan\left(\frac{s}{2}\right) \cdot \tan\left(\frac{s-a}{2}\right) \cdot \tan\left(\frac{s-b}{2}\right) \cdot \tan\left(\frac{s-c}{2}\right)} \right) \quad (75)$$

Das Ergebnis einer solchen Schnittflächenberechnung ist in Abbildung 48 dargestellt. Der Flächeninhalt bezieht sich in diesem Beispiel auf die Berechnung anhand einer „Einheitskugel“ mit einem Radius von $1m$. Generell ist die Schnittfläche abhängig vom Triangulationswinkel zwischen Projektor und Kamera, sowie vom Abstand des Objekts. Der Einfluss dieser Variablen ist aus der Grafik ersichtlich, die effektiv nutzbare Fläche wächst mit steigendem Abstand und fallendem Winkel, so dass bei deckungsgleicher Position von Kamera und Projektor in unendlicher Entfernung das Optimum einer halben Kugeloberfläche erreicht wäre.

Wünschenswert wäre freilich eine Methode, die die Rundumaufnahme der kompletten Oberfläche ohne Einschränkungen unter realen Bedingungen ermöglicht. Dies ist allerdings durch Objektverdeckungen, räumliche Einschränkungen oder die Tatsache einer einzelnen Beleuchtung, die nicht hinter oder durch Materialien scheinen kann, unmöglich. Möchte man aber die flächendeckende Aufnahme einer Halbkugel erreichen, oder allgemein einer 180° umfassenden Sphäre, um beim gewählten Beispiel zu bleiben, ist eine Kombination mehrerer Kameras möglich. Da eine Kamera alleine mehr als einem Drittel der Oberfläche einer Halbkugel abdecken kann, genügen drei Kameras, um deren gesamte Oberfläche zu betrachten. Damit könnte also ein projiziertes Muster vollständig aufgenommen und betrachtet werden. Die Kombination mehrerer Kameras oder Ansichten für eine möglichst umfassende Objekt- oder Szenenerfassung ist dabei nicht neu (u.a. [Murray

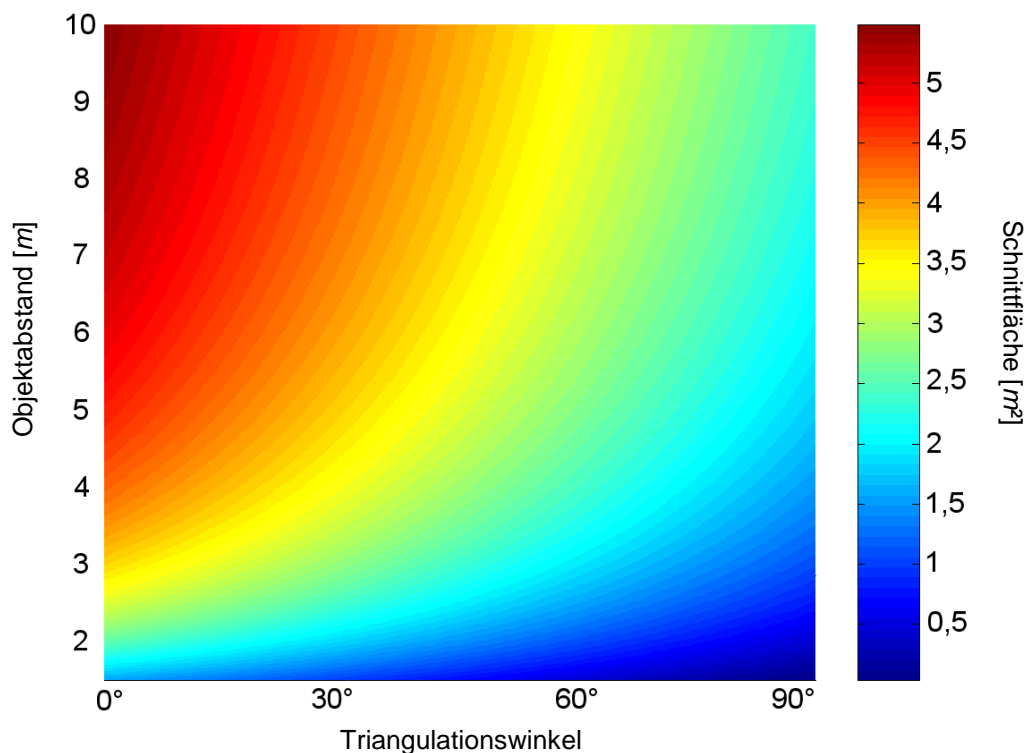


Abbildung 48 Mit einer regulären Kamera-Projektor-Kombination erfassbare Fläche auf einer Kugel. Berechnet wird die Schnittfläche einer kegelförmigen Projektion und Aufnahme auf einer Einheitskugel (Radius $1m$) in Abhängigkeit des Triangulationswinkels und des Objektstandes gemäß Gleichung (73).

1995], [Grest and Koch 2004] oder [Seitz et al. 2006]). Da jedoch die üblichen Projektionsmuster aktiver Stereoskopiesysteme zeilen- oder spaltenorientiert angeordnet sind, würden sich für jede verwendete Kamera entweder der anzuwendende Algorithmus oder die erreichbare Genauigkeit und der zu erwartende Tiefenfehler ändern. Man benötigt also einen Algorithmus mit einem Projektionsmuster, welches unabhängig von der Lage zum Projektor eine allgemeine Positionsbestimmung und Tiefenberechnung ermöglicht. Aufgrund der geometrischen Anordnung und der beschriebenen Projektionseigenschaften bedeutet dies ein rotationssymmetrisches Muster, das seine Projektionsfläche eindeutig codieren kann. Ein Verfahren auf dieser Grundlage wird im folgenden Abschnitt vorgestellt.

5.3.2 Mehr-Kamera-System mit kegelförmiger Beleuchtung

Die einfachste Möglichkeit, einen größeren effektiven Erfassungsbereich zu erreichen, ist also die Kombination einer rotationssymmetrischen Projektion mit mehreren Kameras. Zum Aufbau des Musters wird als Basis eine Farbcodierung verwendet, die aus den 8 Farben des Farbwürfels und Codeworten einer bestimmten Länge besteht – so wie es generell für kanalcodierte Verfahren bereits beschrieben wurde. Allerdings werden die Farben nicht in parallelen Streifen angeordnet, sondern in Kreisen, für deren Zentrum das optische Zentrum des Projektors gewählt wird, bzw. dessen Repräsentation auf der Bildebene. Der

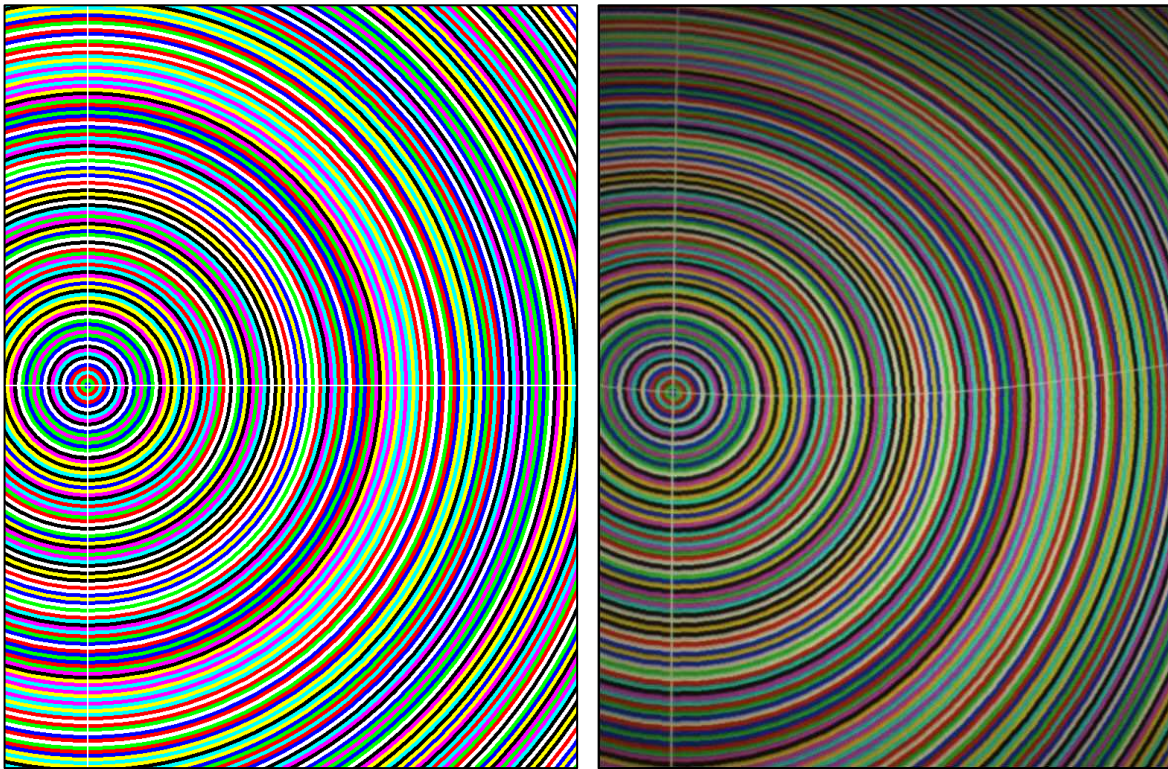


Abbildung 49 Rotationssymmetrisches Projektionsmuster. Die linke Seite zeigt ein farblich codiertes, rotationssymmetrisches Muster, dessen Projektion auf einer gewölbten Oberfläche rechts zu sehen ist. Das Zentrum der Kreise liegt im optischen Zentrum des Projektors, die weißen Linien dienen der Identifikation des Zentrums im aufgenommenen Bild.

Radius jedes Kreises wächst mit steigender Position um einen konstanten Wert, so dass bisherige Ringe stets von einem neuen Ring vollständig umschlossen werden. Das Ergebnis eines solchen Musters ist in Abbildung 49 dargestellt.

Die Tiefenberechnung beruht dabei aus Gründen der Genauigkeit, Detektierbarkeit und Eindeutigkeit wieder auf der Position der Kanten, also den Übergängen von einem Ring zum anderen. Die Decodierung erfolgt ebenfalls anhand der Kantenübergänge und der Farbänderungen der einzelnen Kanäle. Jedoch kann die Kantendetektion im vorliegenden Fall nur bedingt über die Gradienten des Bildes mit dem Sobel-Filter erfolgen, da dieser primär auf horizontale oder vertikale Kantenverläufe anwendbar ist. Durch eine gezielte Kombination beider Richtungen wäre eine radiale Kantenextraktion zwar möglich, besser geeignet ist aber beispielsweise der Laplace-Operator, der in einer isotropen Approximation als folgende Matrix dargestellt werden kann:

$$\underline{L} = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (76)$$

Der bekannte Marr-Hildreth-Operator beispielsweise basiert ebenfalls auf dem Laplace-Filter und ist auch als „Laplace of Gaussian“ bekannt [Jähne 2002].

Der Laplace-Operator bildet die zweite Ableitung eines Bildes und kann zur Detektion von Kanten verwendet werden, da diese in der zweiten Ableitung als Nulldurchgänge erscheinen. Der große Vorteil ist dabei die Richtungsunabhängigkeit des Filters, der zugleich horizontale und vertikale Kanten hervorheben kann. Der Filter besitzt also die Eigenschaft eines Hochpasses, der strukturierte Bereiche hervorhebt und weniger strukturierte Bereiche unterdrückt.

Sind die Kantenpositionen und deren Codierung bekannt, kann mit der Tiefenberechnung begonnen werden. Diese basiert nun nicht mehr auf dem Schnitt einer Ebene mit der Sichtgeraden der Kamera, sondern auf dem Schnitt einer Sichtgeraden mit einem Lichtkegel. Denn jeder Kreis des Musters wird über das Projektionszentrum wieder auf einen Kreis abgebildet. Das Projektionszentrum ist also für jeden projizierten Kreis die Spitze eines Doppelkegels zwischen Szene und Musterebene. Der Mantel eines solchen Kegels, der zugleich die Kante und den Übergang zum nächsten Ring bildet, lässt sich wie folgt beschreiben:

$$(\underline{x} - \underline{c}) \cdot \underline{w} = |\underline{x} - \underline{c}| \cdot |\underline{w}| \cdot \cos \omega_i \quad (77)$$

Dabei bezeichnet \underline{x} die Punkte auf dem Mantel, \underline{c} die Kegelspitze, also das optische Zentrum der Projektion, und \underline{w} die Richtung, bzw. die Mittelachse des Kegels, die sich aus der Geraden durch das optische Zentrum und dessen Abbildung auf der Bildebene ergibt. Der Winkel ω_i ist der Öffnungswinkel des Kegels i und liegt zwischen Mittelachse und Mantellinie, lässt sich also auch folgendermaßen ausdrücken:

$$\cos \omega_i = \frac{\underline{m}_i \cdot \underline{w}}{|\underline{m}_i| \cdot |\underline{w}|} \quad (78)$$

Dabei repräsentiert \underline{m}_i einen beliebigen Punkt auf dem Mantel des Kegels. Mit dieser Interpretation lässt sich (77) ohne trigonometrische Funktion vereinfachen zu:

$$(\underline{x} - \underline{c}) \cdot \underline{w} = \frac{|\underline{x} - \underline{c}|}{|\underline{m}_i|} \cdot (\underline{m}_i \cdot \underline{w}) \quad (79)$$

Jeder Punkt dieser Mantellinie, der von der Kamera erfasst wird, bildet schließlich eine Sichtgerade zur Kamera durch deren optisches Zentrum \underline{k} und gehorcht dabei folgender Beschreibung:

$$\underline{x}(t) = \underline{k} + t \cdot \underline{v} \quad \Leftrightarrow \quad \underline{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} k_x + t \cdot v_x \\ k_y + t \cdot v_y \\ k_z + t \cdot v_z \end{bmatrix} \quad (80)$$

Der Vektor \underline{v} beschreibt hierbei die Richtung der Geraden vom optischen Kamerazentrum zum betrachteten Bildpunkt.

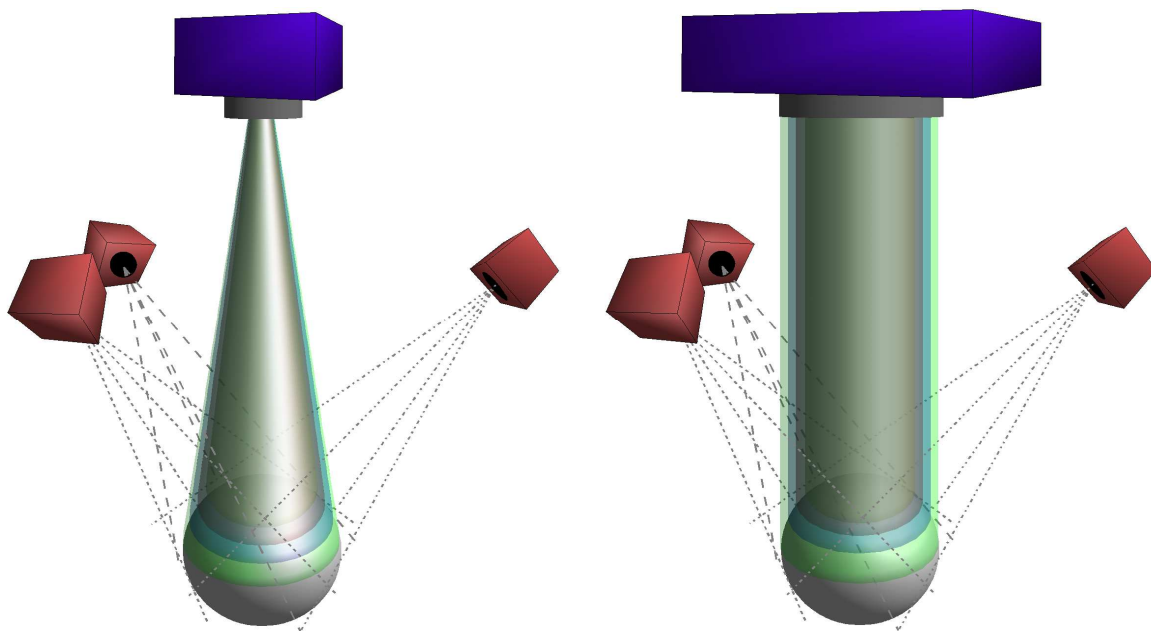


Abbildung 50 Mehr-Kamera-Systeme zur flächendeckenden Objekterfassung. Die linke Seite zeigt den Aufbau mit einer regulären Projektion, deren Abbildung kegelförmig verläuft. Die rechte Seite dagegen verwendet ein telezentrisches Objektiv zur Beleuchtung, so dass eine vollständige Halbkugel durch zylindrische Projektion abgedeckt werden kann.

Mit den Gleichungen (79) und (80) schließlich ist die Tiefenberechnung als Schnitt von Kegel und Gerade möglich, indem man die Koordinaten von \underline{x} in der Kegelformel durch $\underline{x}(t)$ ersetzt und das resultierende Gleichungssystem nach der verbliebenen Variablen t auflöst. Da die Kegelformel einen Doppelkegel beschreibt, ergibt sich für t eine quadratische Gleichung, deren Lösungen auf Plausibilität überprüft werden müssen.

Auch wenn die Berechnung aufwändiger ausfällt als beispielsweise für streifenförmige Codierungen, so erlaubt das beschriebene Verfahren eine beliebige Anordnung der Kameras relativ zum Projektor, so dass eine flächendeckende Aufnahme möglich wird. Abbildung 50 (links) zeigt den Aufbau eines solchen Systems aus 3 Kameras mit einer zentralen Musterprojektion. Die Musterabdeckung jedoch bleibt aufgrund ihrer Zentralprojektion eingeschränkt, so dass keine sphärische 180°-Abdeckung erzielt werden kann. Dieses Problem lässt sich aber durch eine andere Form der Projektionsoptik lösen, wie sie im nächsten Abschnitt beschrieben wird.

5.3.3 Mehr-Kamera-System mit telezentrischer Beleuchtung

Obwohl mit dem im vorigen Punkt beschriebenen Verfahren eine umfassende Aufnahme mit einem positionsunabhängigen Algorithmus beschrieben wurde, erlaubt die Art der Projektion dennoch keine vollständige dreidimensionale Erfassung einer 180°-Sphäre. Das Problem hierbei ist die Beleuchtung, die aufgrund ihrer Projektionseigenschaften lediglich die Schnittfläche eines Kegels mit einer Kugel abdecken kann.

Wird dagegen eine andere Art der Projektion gewählt, so kann man diese Einschränkung ausgleichen. Telezentrische Objektive etwa bilden Szenen innerhalb eines parallelen Strahlengangs ab. Betrachtet man eine Projektion durch ein solches Objektiv, so findet diese innerhalb der Grenzen eines Zylinders statt. Die Beleuchtung einer Kugel, wie im vorangehenden Kapitel beschrieben, würde also eine Hälfte vollständig abdecken können. Somit liefert die Kombination aus einer telezentrischen Projektion und mehreren Kameras zur Aufnahme die notwendigen Voraussetzungen zur Erfassung innerhalb einer 180° umschließenden Sphäre. Abbildung 50 (rechts) zeigt eine entsprechende Anordnung.

Die Tiefenberechnung erfolgt dabei analog zu dem in 5.3.2 beschriebenen Vorgehen, wobei die Projektion nun nicht mehr kegelförmig, sondern zylindrisch modelliert wird. Das rotationssymmetrische Muster wird also derart abgebildet, dass jeder Kreis einen Zylinder gemäß folgender Formel beschreibt:

$$r_i^2 = (\underline{x} - \underline{c})^T \cdot (\underline{E} - \underline{w}\underline{w}^T) \cdot (\underline{x} - \underline{c}) \quad (81)$$

Dabei bezeichnet r_i den Radius des aktuellen Zylinders und \underline{c} einen beliebigen Punkt auf dessen Mittelachse, wobei hierfür idealerweise die Abbildung des optischen Zentrums gewählt werden sollte. \underline{E} bezeichnet die Einheitsmatrix und \underline{w} die Richtung der Mittelachse des Zylinders, also die Projektionsrichtung. Tiefenwerte erhält man analog dem bisherigen Vorgehen durch den Schnitt der Sichtgeraden der Kamera mit der parametrisierten Beschreibung der Zylinder. Am grundsätzlichen Vorgehen zur Detektion und Identifikation der Kanten des im vorigen Abschnitt beschriebenen Kegelansatzes ändert sich dabei nichts, da beide Körper – Zylinder und Kegel – rotationssymmetrische Objekte sind.

Die technischen Umsetzungen telezentrischer Objektive limitieren dabei jedoch die Anwendbarkeit des Verfahrens stark. Da eine Parallelprojektion stattfindet, bedeutet dies, dass die Linsen des Objektivs auf jeden Fall mindestens den gleichen Durchmesser besitzen müssen wie das aufzunehmende Objekt selbst. Die Aufnahme größerer Objekte, wie etwa des menschlichen Gesichts, scheidet daher für dieses Verfahren aus. Zudem sind telezentrische Optiken in ihrer Tiefenschärfe ziemlich begrenzt, so dass das beschriebene zylindrische Verfahren eher für die dreidimensionale Vermessung kleinerer Objekte und Oberflächen geeignet ist.

Für beide Verfahren, ob kegelförmig oder zylindrisch, lässt sich generell aussagen, dass die Messgenauigkeit stark von der Art der Mustererzeugung abhängt. Verwendet man hierfür einen digitalen Bildsensor, wie etwa einen Videoprojektor, so bedeutet das, dass die Kreise zur Projektion durch diskrete Pixel und damit diskrete Positionen auf dem Sensor repräsentiert werden. Da ein Kreis aber per definitionem einen konstanten Radius besitzt, führt dieses Vorgehen zwangsläufig zu Fehlern in der Tiefenberechnung. Es sollten also beispielsweise hochauflösende Filtermasken verwendet werden, die eine Annäherung an eine kontinuierliche Kreismodellierung garantieren können. Auch digitale Sensoren mit Auflösungen im Megapixelbereich kommen hierfür in Frage. Da aber die derzeit verfügbaren Komponenten diese Voraussetzungen nicht erfüllen können, lassen sich mit den

dargestellten rotationssymmetrischen Verfahren nur schlechtere Ergebnisse erzielen als mit den bisher beschriebenen „regulären“ aktiven Stereoskopiesystemen. So müssen durch Berücksichtigung der generischen Fehler im Rahmen einer Kalibrierung oder durch Anpassung der Optiken für kontinuierlichere Kantenübergänge die Messergebnisse korrigiert werden, um aussagekräftige Tiefenwerte zu erhalten.

Dennoch verspricht die Anwendung einer radialsymmetrischen Codierung mit der Kombination mehrerer Kameras eine einfach aufzubauende und kostengünstige Methode einer umfassenden, dreidimensionalen Objekterfassung.

6 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurde die generische 3D-Objekterfassung zur Aufnahme und Modellierung von Gegenständen und Körpern behandelt. Im Allgemeinen lag der Schwerpunkt hierbei auf aktiven Stereoskopieverfahren, im Speziellen auf echtzeitfähigen Messsystemen für technische und biometrische Anwendungen, die vom Einsatz exakter Modellierungen menschlicher Gesichter profitieren.

Es wurden hierfür zunächst die Grundlagen der Entfernung- und Tiefenmessung betrachtet und ein Überblick über den derzeitigen Stand der Technik in einigen Bereichen der 3D-Erfassung hergestellt. Beschrieben wurden laufzeitbasierte Messverfahren, reine optische Methoden und Verfahren auf Basis der Triangulation. Hierbei wurde zunächst das Vorgehen der passiven Stereoskopie beschrieben, auf dem aktive Verfahren aufbauen. Im Bereich dieser aktiven Methoden wurden bekannte und bewährte Verfahren der strukturierten Beleuchtung beschrieben, auf denen neue und aktuelle Methoden der Musterkodierung basieren, wie etwa durch zeitliche oder farbige Repräsentationen. Die Vor- und Nachteile dieser Verfahren wurden ebenso beschrieben wie ihre Eignung für den Einsatz als biometrische Schnittstellen.

Auf der Grundlage dieser Überlegungen schließlich konnten Verfahren entwickelt werden, die zum Aufbau von speziellen Sensorsystemen zur aktiven stereoskopischen Erfassung menschlicher Gesichter geführt haben. Diese neuen Ansätze basieren auf der dreidimensionalen Erfassung unter Anwendung von strukturiertem Licht, welches zur Vermeidung von Blendwirkungen in das infrarote Lichtspektrum verlagert wurde. Die Auswirkungen dieser Spektrenänderung wurden erforscht, der Einfluss auf die menschlichen Organe wie Haut und Augen dargestellt, sowie die Folgen auftretender Interaktionen auf die Datengenauigkeit betrachtet. Im Rahmen einer statistischen Untersuchung und einer empirischen Evaluation wurden die anhand der neuen Verfahren aufgebauten Prototypen vorhandenen und bewährten Messtechniken der aktiven Stereoskopie gegenübergestellt. Es konnte gezeigt werden, dass alle Erwartungen, die an Messmethoden im sichtbaren Licht gestellt werden, in Hinblick auf die Messgenauigkeit, die Aufnahmegeschwindigkeit, die Tiefenauflösung und die Robustheit gegenüber Störeinflüssen durch die neuen Methoden entwe-

der erfüllt oder sogar übertroffen werden. Dies beweist die Eignung der entwickelten Verfahren als biometrische Schnittstelle zur Erfassung dreidimensionaler Daten.

Diese Eignung konnte zudem durch Beispiele einzelner Projekte aufgezeigt werden, in denen die beschriebenen Methoden zum Einsatz kamen. Diese Projekte umfassten dabei sowohl technische Anwendungen als auch den Einsatz in biometrischen und biologischen Verfahren.

Neben diesen speziell für die dreidimensionale Gesichtserfassung ausgelegten Methoden wurden darüber hinaus Erweiterungen zu aktiven Stereoskopieverfahren entwickelt und vorgestellt. Diese Erweiterungen hatten eine Erhöhung der Flexibilität, eine Vergrößerung des Messumfangs oder eine Verbesserung der Auflösung und Genauigkeit zum Ziel. Hierbei wurden Verfahren beschrieben, die etwa eine vollständige Szenencodierung mit einer schnellen Art der Tiefenberechnung über eine laufzeitbasierte Algorithmik vereinen. Eine Steigerung des effektiven Messfelds eines Stereosystems konnte durch die Kombination mehrerer Aufnahmesensoren und einer rotationssymmetrischen Musterkodierung erreicht werden.

Während all dieser Untersuchungen und Entwicklungen im Rahmen dieser Arbeit wurde deutlich, dass sich durch die Entwicklung neuer Komponenten, durch den Einsatz neuer Techniken oder durch die Verbesserung von Fertigungsprozessen auch ältere oder bekannte Methoden stets verbessern lassen. Dies wird vor allem an Verfahren offensichtlich, deren Realisierung in früheren Jahren aus Mangel an geeigneter Technik scheiterte und die nun durch den Einsatz etwa aktueller digitaler Projektionssysteme eine Art Renaissance erleben. So werden heutzutage in vielen technischen Anwendungen Methoden zur dreidimensionalen Datenerfassung eingesetzt, die bereits in den 80er Jahren entwickelt wurden.

Ähnliche Limitierungen der verfügbaren technischen Komponenten mussten auch teilweise bei den in dieser Arbeit vorgestellten Verfahren beobachtet werden. Es ist aber zu erwarten, dass in allen notwendigen Bereichen, angefangen von der Beleuchtungstechnik bis hin zu Projektionssensoren, neue Techniken dazu führen werden, dass auch diese Methoden in den andauernden Prozess der Forschung, Entwicklung und Verbesserung der Messtechnik einfließen. Somit liefert diese Arbeit wichtige Inhalte, Informationen und Ansätze zur aktuellen und zukünftigen Entwicklung von Sensoren, die als biometrische Schnittstelle und zur generischen dreidimensionalen Objekterfassung geeignet sind.

Übersicht häufig verwendeter Abkürzungen

2D, 3D	zweidimensional, dreidimensional
CCD	Charge-Coupled Device, ein digitaler Kamerasensor
CCT	Color Coded Triangulation, farbcodierte Triangulation
CMOS	Complementary Metal Oxide Semiconductor (hier unter anderem als digitaler Kamerasensor verwendet)
DLP	Digital Light Processing, eine digitale Projektionstechnik
DMD	Digital Micromirror Device, der Sensor eines DLP-Systems
EER	Equal Error Rate, Maß zur Fehlerberechnung
FAR	False Acceptance Rate, Sicherheitsmerkmal eines Erkennungssystems
fps	frames per second, Bilder/Datensätze/Aufnahmen pro Sekunde
FRGV	Face Recognition Grand Vendor Test
IR	Infrarot
LCD	Liquid Crystal Display, hier als digitale Projektionstechnik
LCoS	Liquid Crystal on Silicon, eine digitale Projektionstechnik
LED	Leuchtdiode
NIR	Near Infrared, naher Infrarotbereich
PBS	Polarizing Beam Splitter, ein spezieller Teilerwürfel in Projektoren
Pixel	Picture Element, Bildpunkt
RMSD	Root Mean Squared Deviation, mittlere quadratische Abweichung
RMSE	Root Mean Squared Error, mittlerer quadratischer Fehler
TFT	Thin Film Transistor, eingesetzt bei Matrix-LCD-Anordnungen
UV	Ultraviolett

Übersicht häufig verwendeter Symbole

(falls im zugehörigen Text nicht anderweitig bezeichnet)

b	Basislänge einer Triangulationsanordnung
c	Lichtgeschwindigkeit
c_S	Schallgeschwindigkeit
d_x, d_y	Maße eines Sensorelements in x-/y-Richtung
Δi	Kennzeichnung der Abweichung, des Unterschieds oder des Fehlers einer Variablen i
f	Brennweite einer Kamera oder eines Projektors
f_{Abtast}, f_{Signal}	Abtast- und Signalfrequenz
I	auf einem Sensor auftreffende/gemessene Lichtintensität
κ_i	i -ter Verzerrungskoeffizient einer Linsenverzerrung
λ	Wellenlänge von abgestrahltem Licht
Λ	synthetische erzeugte Wellenlänge
\underline{M}	Bezeichnung einer Matrix M
$m_{i,j}$	Element der i -ten Zeile und j -ten Spalte der Matrix \underline{M}
N_x, N_y	Anzahl der Elemente/Pixel eines Sensors in x-/y-Richtung
O	Ursprung eines Koordinatensystems
P	Bezeichnung eines Punktes P
p_x, p_y, p_z	x-, y-, z-Koordinate (Komponente) eines Punktes P (Vektors \underline{p})
\underline{P}	Projektionsmatrix P
\underline{R}	Rotationsmatrix R
$r_{i,j}$	Element der Rotationsmatrix
$\sigma(k)$	Standardabweichung, auch als Maß für eine Messungenauigkeit des Wertes k
$T, t (T_i, t_i)$	Bezeichnung einer Zeit oder eines Zeitraums (mit Indizierung i)
\underline{t}	Translationsvektor t
t_x, t_y, t_z	Komponenten des Translationsvektors
θ	Triangulationswinkel
U	elektrische Spannung
u, v	Koordinaten eines Bildpunkts in Pixel- bzw. Bildkoordinaten
\underline{v}	Bezeichnung eines Vektors v
x, y, z	Koordinatenbezeichnung

x_C, y_C, z_C	Koordinaten im Kamerakoordinatensystem
x_W, y_W, z_W	Koordinaten im Weltkoordinatensystem
x_i, y_i	Koordinaten auf der Sensor-/Bildebene einer Kamera
$x_{P^*}, y_{P^*}, z_{P^*}$	Koordinaten eines Punktes P im Koordinatensystem *
x_P, y_P, z_P	allgemein: Koordinaten eines Punktes P im aktuellen Koordinatensystem

Literaturverzeichnis

[Ballard and Brown 1982]

D. H. Ballard and C. M. Brown, "Computer Vision", Prentice-Hall, Englewood Cliffs, New Jersey, 1982

[Bauer 2003]

Norbert Bauer (Redaktion), "Leitfaden zu Grundlagen und Anwendungen der optischen 3D-Messtechnik", Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V., 2003

[Bayer 1976]

B. E. Bayer, "Color Imaging Array", United States Patent 3,971,065, 1976

[Bergmann 1995]

D. Bergmann, "New approach for automatic surface reconstruction with coded light", In T.F. Schenk, (Hrsg.) Proc. SPIE Vol. 2572 - Remote Sensing and Reconstruction for Three-Dimensional Objects and Scenes, pp. 2-9, 1995.

[BioSec 2005]

Research-Project "Biometrics and security" (BioSec), IST-2002-001766-FP6, 01.12.2003-30.11.2005,
http://cordis.europa.eu/fetch?CALLER=FP6_PROJ&ACTION=D&DOC=1&CAT=PROJ&QUERY=1203366579567&RCN=74601

[Blais 2004]

François Blais, "Review of 20 years of range sensor development", Journal of Electronic Imaging, Volume 13, Issue 1, pp. 231-243, 2004

[Blostein and Huang 1987]

S. D. Blostein and T. S. Huang, "Error Analysis in Stereo Determination of 3-D Point Positions", Transactions on Pattern Analysis and Machine Intelligence (IEEE PAMI), Vol. 9, Iss. 6, pp. 752-765, 1987

[Bowyer et al. 2006]

Kevin W. Bowyer, Kyong Chang, Patrick Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition", Computer Vision and Image Understanding, Vol. 101, Iss. 1, pp. 1-15, 2006

[Burghardt and Wasilewski 2007]

Tilo Burghardt and Anja Wasilewski, "Conservational Prospects of Computer-aided Visual Bontebok Identification Using Biometrical Features", 6th International Zoo and Wildlife Research Conference on Behaviour, Physiology and Genetics, pp. 248-250, October 2007

[Carrihill and Hummel 1985]

B. Carrihill and R. Hummel, "Experiments with the intensity ratio depth sensor", Computer Vision, Graphics and Image Processing, Vol. 32, pp. 337-358, Academic Press, 1985

[Caspi et al. 1998]

D. Caspi, N. Kiryati and J. Shamir, "Range Imaging with Adaptive Color Structured Light", Transactions on Pattern Analysis and Machine Intelligence (IEEE PAMI), Vol. 20, Iss. 5, pp. 470-480, 1998

[Chang et al. 1994]

C. Chang, S. Chatterjee and P. R. Kube, "A Quantization Error Analysis for Convergent Stereo", Proceedings of the IEEE International Conference in Image Processing (ICIP-94), Vol. 2, pp. 735-739, 1994

[Chen et al. 2007]

H.J. Chen, J. Zhang, D. J. Lv and J. Fang, "3-D shape measurement by composite pattern projection and hybrid processing", Optics Express, Vol. 15, Iss. 19, pp. 12318-12330, 2007

[D'Apuzzo 2002]

N. D'Apuzzo, "Modeling human faces with multi-image photogrammetry", Proceedings of the SPIE, Vol. 4661, pp.191-197, 2002

[DIN 1992]

H. G. Freeman, „Wörterbuch technischer Begriffe mit 6500 Definitionen nach DIN“, 4. Auflage, DIN Deutsches Institut für Normung e.V., Berlin 1992

[Elkhalili et al. 2004]

O. Elkhalili, O. M. Schrey, P. Mengel, M. Petermann, W. Brockherde and B. J. Hosticka, „A 4 x 64 Pixel CMOS Image Sensor for 3-D Measurement Applications“, IEEE Journal of Solid-State Circuits, Vol. 39, No. 7, July 2004

[Enciso et al. 2003]

Reyes Enciso, Alex M. Shaw, Ulrich Neumann and James Mah, "Three-dimensional head anthropometric analysis", Medical Imaging 2003, Proceedings of SPIE, Vol. 5029, pp. 590-597, May 2003

[Faugeras 1993]

O. Faugeras, "Three-Dimensional Computer Vision: A Geometric Viewpoint", MIT Press, Cambridge, MA, 1993

[Forster 2005]

F. Forster, "Real-Time Range Imaging for Human-Machine Interfaces", Dissertation, TU München, Lehrstuhl für Mensch-Maschine-Kommunikation, 2005

[Forster et al. 2002]

F. Forster, M. Lang and B. Radig, "Real-Time Range Imaging for Dynamic Scenes Using Colour-Edge Based Structured Light", Proceedings of the International Conference on Pattern Recognition (ICPR2002), Vol. 3, pp. 645-648, 2002

[Forsyth and Zisserman 1991]

David Forsyth and Andrew Zisserman, "Reflections on Shading", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, Iss. 7, pp. 671-679, 1991

[Gorokhovskiy et al. 2006]

K. Gorokhovskiy, J. A. Flint and S. Datta, "Alternative color filter array layouts for digital photography", Research in Microelectronics and Electronics 2006, Ph. D., Issue 35631, pp. 265-268, 2006

[Grest and Koch 2004]

D. Grest and R. Koch, "Realtime multi-camera person tracking for immersive environments", IEEE 6th Workshop on Multimedia Signal Processing, pp. 387-390, 2004

[Guan et al. 2003]

C. Guan, L. Hassebrook, and D. Lau, "Composite structured light pattern for three-dimensional video", Optics Express, Vol. 11, Issue 5, pp. 406-417, 2003

[Gühring 2001]

J. Gühring, "Dense 3-d surface acquisition by structured light using off-the-shelf components", Videometrics and Optical Methods for 3D Shape Measurement, Proceedings of the SPIE, Vol. 4309, 220-231, 2001

[Haasbroek 1968]

N. D. Haasbroek, „Gemma Frisius, Tycho Brahe and Snellius and their Triangulations“, Publication of the Netherlands Geodetic Commission, Delft, 1968

[Hall-Holt and Rusinkiewicz 2001]

O. Hall-Holt and S. Rusinkiewicz, "Stripe boundary codes for real-time structured-light range scanning of moving objects", Eighth International Conference on Computer Vision (ICCV2001), pp. 359-366, 2001

[Hartley and Zisserman 2000]

R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, Cambridge, UK, 2000

[Hauske 1994]

Gert Hauske, „Systemtheorie der visuellen Wahrnehmung“, Teubner, Stuttgart, 1994

[HCN 1993]

Health Council of the Netherlands, "Optical radiation Health based exposure limits for electromagnetic radiation in the wavelength range from 100 nanometre to 1 millimetre", Gezondheidsraad, LM The Hague, 1993

[HMI 2008]

Hahn-Meitner-Institut,
http://www.hmi.de/pr/bildarchiv/solarenergie/grafiken/giffs_jpeg/sonnenspektrum.jpg,
Download am 03.01.2008

- [Holo 2008]
Holoeye Photonics, "LCOS Microdisplay Technology", Download am 20.03.2008
- [Horn and Kiryati 1997]
E. Horn and N. Kiryati, "Toward Optimal Structured Light Patterns", Proceedings of the International Conference on Recent Advances in 3-D Digital Imaging and Modeling, pp. 28-35, 1997
- [Inokuchi et al. 1984]
S. Inokuchi, K. Sato and F. Matsuda, "Range Imaging System for 3-D Object Recognition", IEEE 7th International Conference on Pattern Recognition (ICPR84), pp. 806-808, Montreal, 1984
- [Jähne 1991]
Bernd Jähne, „Digitale Bildverarbeitung“ (2. Auflage), Springer-Verlag, Berlin Heidelberg, 1991
- [Jähne 2002]
Bernd Jähne, „Digital Image Processing“ (5th Edition), Springer, Berlin Heidelberg, 2002
- [Jung and Kopp 2003]
B. Jung and S. Kopp, "FlurMax: An interactive virtual agent for entertaining visitors in a hallway", Intelligent Agents: 4th International Workshop (IVA2003), pp. 23–26, 2003
- [Klette et. al. 1996]
R. Klette, A. Koschan und K. Schlüs, „Computer Vision“, Vieweg Verlagsgesellschaft, Braunschweig/Wiesbaden, 1996
- [Klußmann and Wiegelmann 2005]
N. Klußmann and Jörg Wiegelmann, „Lexikon Elektronik“, Hüthig GmbH & Co. KG, Heidelberg, 2005
- [Knoll and Sasse 1993]
A. Knoll and Ralph Sasse, "An Active Stereometric Triangulation Technique Using a Continuous Colour Pattern", Graphics & Robotics, pp. 191-206, 1993
- [Koch et al. 1998]
A. W. Koch, M. W. Ruprecht, O. Toedter und G. Häusler, „Optische Messtechnik an technischen Oberflächen“, expert verlag, Renningen-Malmsheim, 1998
- [Lee et al. 2005]
Sang-Yup Lee, Ig-Jae Kim, Sang C Ahn, Myo-Taeg Lim, Hyung-Gon Kim, "Intelligent 3D Video Avatar for Immersive Telecommunication", AI 2005, Lecture Notes in Computer Science (LNCS 3809), pp. 726-735, Springer, 2005
- [Lengagne et al. 2000]
R. Lengagne, P. V. Fua, and O. Monga, "3D stereo reconstruction of human faces driven by differential constraints", Image and Vision Computing Journal (IVC), Vol. 18, Nr. 4, pp. 337-343, March 2000

[Lukac et al. 2005]

R. Lukac, B. Smolka, K. Martin, K. N. Plataniotis and A. N. Venetsanopoulos, "Vector Filtering for Color Imaging", IEEE Signal Processing Magazine, Vol. 22, Nr. 1, pp. 74-86, 2005

[Malassiotis and Srinatzis 2005]

S. Malassiotis and M. G. Srinatzis, "Robust Real-time 3D Head Pose Estimation from Range Data", Pattern Recognition, Vol. 38, Nr. 8, pp. 1153-1165, August 2005

[Mallon and Whelan 2007]

J. Mallon and P.F. Whelan, "Calibration and removal of lateral chromatic aberration in images", Pattern Recognition Letters, Vol. 28, No. 1, pp. 125-135, 2007

[Meglinski and Matcher 2002]

Igor V. Meglinski and Stephen J. Matcher, "Quantitative assessment of skin layers absorption and skin reflectance spectra simulation in the visible and near-infrared spectral regions", Physiol. Meas. 23, pp. 741-753, 2002

[Meixner et al. 2004]

Andreas Meixner, Andreas Purde, Hans Schweizer, Philipp Kirilenko, Markus Riemen-schneider, Thomas Zeh und Alexander W. Koch, "Methoden der interferometrischen Formerfassung nicht-ruhender technischer Oberflächen", Technisches Messen 71, pp. 211-217, Oldenburg Verlag, 2004

[Mengel et al. 2001]

Peter Mengel, Günter Doemens, Ludwig Listl, "Fast Range Imaging by CMOS Sensor Array Through Multiple Double Short Time Integration (MDSI)", Proceedings of the IEEE International Conference on Image Processing (ICIP), Vol. 2, pp. 169-172, 2001

[Merrill 1999]

Richard Billings Merrill, "Color separation in an active pixel cell array using a triple well structure", United States Patent 5,965,875, 1999

[Minou et al. 1981]

Minou M., Kanade T. and Sakai T. A Method of Time-Coded Parallel Planes of Light for Depth Measurement. Transactions of the IECE of Japan, Vol. E64(8): 521-528. 1981

[Miyasaka and Araki 2002]

Takeo Miyasaka and Kazuo Araki, "Development of Real-Time 3D Measurement System Using Intensity Ratio Method", Proc. ISPRS Commission III, Vol. 34, Part 3B, Photogrammetric Computer Vision (PCV 02), pp. 181-185, Graz, 2002

[Miyasaka et al. 2000]

T. Miyasaka, K. Kuroda, M. Hirose, K. Araki, "High speed 3-D measurement system using incoherent light source for human performance analysis", Proceedings of the 19th Congress of The International Society for Photogrammetry and Remote Sensing (ISPRS2000), pp. 65-69, Amsterdam, 2000

[Modrow et al. 2007]

D. Modrow, C. Laloni, G. Doemens and G. Rigoll, "3D Face Scanning Systems Based on Invisible Infrared Coded Light", *Advances in Visual Computing, LNCS 4841*, pp. 521–530, Springer, 2007

[Modrow et al. 2008a]

D. Modrow, C. Laloni, G. Doemens and G. Rigoll, "A novel sensor system for 3D face scanning based on infrared coded light", *Proceedings of the SPIE, Vol. 6805*, pp. 68050M-1-10, 2008

[Modrow et al. 2008b]

D. Modrow, C. Laloni and G. Rigoll, "The Influence of Color on Color Coded Triangulation", *Proceedings of the Tenth IASTED International Conference: Computer Graphics and Imaging*, pp. 243-248, Innsbruck, Austria, 2008

[Monks 1994]

P. Monks, "Measuring the Shape of Time-Varying Objects", PhD Thesis, University of Southampton, Faculty of Engineering and Applied Science, 1994

[Morano et al. 1998]

R. A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, J. Nissanov, "Structured light using pseudorandom codes", *Pattern Analysis and Machine Intelligence, Vol. 20, Iss. 3*, pp. 322–327, 1998

[Mpiperis et al. 2007]

I. Mpiperis, S. Malassiotis and M.G. Strintzis, "3D Face Recognition with the Geodesic Polar Representation", *IEEE Transactions on Information Forensics and Security, Vol. 2, Iss. 3, Part 2*, pp.537-547, September 2007

[Murray 2006]

D. W. Murray, "Recovering Range using Virtual Multi-camera Stereo", *Computer Vision and Image Understanding, Vol. 61 (2)*, pp. 285-291, 1995

[Pagès et al. 2003]

J. Pagès, J. Salvi, R. Garcia and C. Matabosch, "Overview of coded light projection techniques for automatic 3D profiling", *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '03), Vol. 1*, pp. 133-138, 2003

[Pajdla 1995]

T. Pajdla, "BCRF - Binary-Coded Illumination Range Finder Reimplementation", Technical report KUL/ESAT/MI2/9502, Katholieke Universiteit Leuven, ESAT, 1995

[Pallek 2007]

Dieter Pallek, "Moiré – Technik", Deutsches Zentrum für Luft- und Raumfahrt (DLR), http://www.dlr.de/as/desktopdefault.aspx/tabid-183/251_read-2738/
Download am 05.11.2007

[Panasonic 2007]

Panasonic, 3CCD-Technik, ECAM Camcorder, <http://www.panasonic.de/technologie/Camcorder/3CCDCamcorder.aspx>, letzte Einsicht am 17.08.2007

[PASION 2008]

Research-Project “Psychologically Augmented Social Interaction Over Networks” (PASION), IST-2006-027654-FP6, 01.01.2006-31.12.2009,
http://cordis.europa.eu/fetch?CALLER=FP6_PROJ&ACTION=D&DOC=1&CAT=PROJ&QUERY=1204448731318&RCN=80641

[Phillips et al. 2007]

P. J. Phillips, W. T. Scruggs, A. J. O’Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott und M. Sharpe, “FRVT 2006 and ICE 2006 large-scale results”, NISTIR 7408, 2007

[Ph-T35 2008]

Lehrstuhl T35 für Theoretische Biophysik an der TUM, „Auditorische Signalverarbeitung mit gekoppelten Tympani“, http://www.t35.ph.tum.de/research/lizard_hearing.html, letzte Einsicht am 18.02.2008

[Potsdamer and Altschuler 1982]

J. L. Potsdamer, M. D. Altschuler, Surface measurement by space-encoded projected beam systems, *Computer Graphics and Image Processing* 18 (1) (1982) 1–17

[Reidenbach et al. 2005]

Hans-Dieter Reidenbach (Redaktion), „Leitfaden ‚Nichtionisierende Strahlung‘ – Sichtbare und infrarote Strahlung“, Fachverband für Strahlenschutz e.V., Fachhochschule Köln, 2005

[Rocchini et al. 2001]

C. Rocchini, P. Cignoni, C. Montani, P. Pinci and R. Scopigno, “A low cost 3D scanner based on structured light”, *EUROGRAPHICS 2001 Proceedings*, Vol. 20(3), pp. 299-308, 2001

[Rodriguez and Aggarwal 1990]

J. J. Rodriguez and J.K. Aggarwal, “Stochastic Analysis of Stereo Quantization Error”, *Transactions on Pattern Analysis and Machine Intelligence (IEEE PAMI)*, Vol. 12, Iss. 5, pp. 467-470, 1990

[Salvi et al. 1998]

J. Salvi, J. Batlle, E. Mouaddib, “A robust-coded pattern projection for dynamic 3d scene measurement, *International Journal of Pattern Recognition Letters* (19), pp. 1055–1065, 1998

[Salvi et al. 2004]

Joaquim Salvi, Jordi Pagès and Joan Batlle, “Pattern codification strategies in structured light systems”, *Pattern Recognition* 37(4), pp. 827-849, 2004

[Sato 1996]

K. Sato, “Range imaging based on moving pattern light and spatio-temporal matched filter”, *IEEE International Conference on Image Processing (ICIP 1996)*, Vol. 1, pp. 33–36, 1996

[Schröder 1998]

Gottfried Schröder, „Technische Optik: Grundlagen und Anwendungen“, 8., überarb. Auflage, Vogel, Würzburg 1998

[Seitz et al. 2006]

S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms", IEEE Computer Vision and Pattern Recognition (CVPR'06), Volume 1, pp. 519-528, 2006

[Slama et al. 1980]

C. Slama, C. Theurer and S. Hendriksen (Editors), "Manual of Photogrammetry (4th Edition)", American Society of Photogrammetry, Falls Church, Va., 1980

[Smutny and Pajdla 1996]

V. Smutný and T. Pajdla, "Rainbow Range Finder and its Implementation at the CVL", Research Report No. K335-1996, Czech Technical University, Faculty of Electrical Engineering, March 1996

[Sony 2008]

Sony Corporation, CCD Image Sensor Datasheet Catalogue,
<http://www.sony.co.jp/~semicon/english/90203.html>,
letzte Einsicht am 03.01.2008

[Spoelder et al. 2000]

H. J. W. Spoelder, F. M. Vos, Emil M. Petriu and F. C. A. Groen, "Some Aspects of Pseudo Random Binary Array-Based Surface Characterization", IEEE Transaction on Instrumentation and Measurement, Vol. 49, Nr. 6, pp. 1331-1336, 2000

[Steer 2008]

A. Steer, 1931 CIE Chromaticity diagram, www.techmind.org
Download am 23.01.2008

[Steurer et al. 1986]

Johannes Steurer, H. Giebel und W. Altner, „Ein lichtmikroskopisches Verfahren zur zweieinhalbdimensionalen Auswertung von Oberflächen“, Mustererkennung 1986: 8. DAGM-Symposium, pp. 66-70, Paderborn, Germany, 1986

[Stoykova et al. 2007]

Elena Stoykova, Aydin Alatan, Philip Benzie, Nikos Grammalidis, Sotiris Malassiotis, Joern Ostermann, Sergej Piekh, Ventseslav Sainov, Christian Theobalt, Thangavel Thevar, Xenophon Zabulis, "3D Time-varying Scene Capture Technologies - A Survey", IEEE Transactions on Circuits and Systems for Video Technology, Special issue on 3DTV, Vol. 17, Iss. 11, pp. 1568-1586, 2007

[TI 2005]

Texas Instruments, "Single-Panel DLP Projection System Optics", Application Report, Discovery DLPA002, 2005

[Trobina 1995]

Marjan Trobina, "Error model of a coded-light range sensor", Technical Report BIWI-TR-164, Communication Technology Lab, Image Science Group, ETH, Zürich, 1995

[Tsai 1987]

Roger Y. Tsai, „A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses“, IEEE Journal of Robotics and Automation, Vol. RA-3, No.4, August 1987

[Tsalakanidou 2006]

Filareti Tsalakanidou, “Face Recognition from Three-dimensional Colour Recordings”, PhD Thesis, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Thessaloniki, Greece, April 2006.

[Tsalakanidou and Malassiotis 2007]

F. Tsalakanidou and S. Malassiotis, “Application and Evaluation of a 2D+3D Face Authentication System”, Proceedings of the IEEE 3DTV Conference 2007 (3DTV-CON 2007), pp. 1-4, 2007

[Tsalakanidou et al. 2005]

F. Tsalakanidou, S. Malassiotis and M. G. Strintzis, “Face Localization and Authentication Using Color and Depth Images”, IEEE Transactions on Image Processing, Vol. 14, Nr. 2, pp. 152-168, February 2005

[Turner et al. 2005]

Richard M. Turner, Richard F. Lyon, Rudolph J. Guttosch, Richard B. Merrill, “Vertical color filter detector group and array”, United States Patent 6,864,557, 2005

[Valkenburg and McIvor 1998]

R. J. Valkenburg and A. M. McIvor, “Accurate 3d measurement using a structured light system”, Image and Vision Computing, Vol. 16, Nr. 2, pp. 99–110, 1998

[Vuylsteke and Oosterlinck 1990]

P. Vuylsteke, A. Oosterlinck, “Range image acquisition with a single binary-encoded light pattern”, Transactions on Pattern Analysis and Machine Intelligence (IEEE PAMI), Vol. 12, No. 2, pp. 148–163, 1990

[Vuylsteke et al. 1990]

P. Vuylsteke, C. B. Price, A. Oosterlinck, “Image sensors for real-time 3D acquisition: part 1”, in “NATO ASI Series F: Traditional and non-traditional robotic sensors”, pp. 187-210, Springer-Verlag, New York, 1990

[Wust and Capson 1991]

C. Wust and D.W. Capson, “Surface Profile Measurement Using Color Fringe Projection”, Machine Vision and Applications, Vol. 4, Nr. 3, pp- 193-203, 1991

[Yadid-Pecht and Etienne-Cummings 2004]

O. Yadid-Pecht and R. Etienne-Cummings (Eds.), “CMOS Imagers: From Phototransduction to Image Processing”, Kluwer Academic Press, Boston, 2004

[Zhang 1998]

Zhengyou Zhang, “A Flexible New Technique for Camera Calibration”, Technical Report MSR-TR-98-71, Microsoft Research, Microsoft Corporation, Redmond, WA 98052, 1998

[Zhang and Huang 2006]

Song Zhang and Peisen S. Huang, "High-resolution, real-time three-dimensional shape measurement", *Optical Engineering*, Vol. 45, No. 12, December 2006

[Zhang et al. 1999]

Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer and Mubarak Shah, "Shape from Shading: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, Iss. 8, pp. 690-706, 1999

[Zhang et al. 2002]

L. Zhang, B. Curless, S. M. Seitz, "Rapid shape acquisition using color structured light and multi-pass dynamic programming", *Int. Symposium on 3D Data Processing Visualization and Transmission*, pp. 24-36, 2002