



Masterarbeit M 06-2017

**Untersuchung von Verfahren
zur Ausschnittsdetektion
für die Screen-Camera
Light Communication**

von

Xiaoyu Chen

Abgabedatum: 10. September 2018

Prof. Dr.-Ing Rüdiger Kays • Lehrstuhl für Kommunikationstechnik • TU Dortmund

Inhaltsverzeichnis

1 Einleitung	1
1.1 Motivation	1
1.2 Aufbau der Arbeit	2
2 Optische Freiraum-Datenübertragung mittels Videodisplays und Kameras	3
2.1 Einführung des Data transmission using Video devices (DaVid)s	3
2.2 Systemmodell	4
2.2.1 Modulationsverfahren	5
2.2.2 DatenBlock	7
2.3 Anwendungsgebiete	7
3 Untersuchung vom ersten Verfahren mittels Mustererkennung	9
3.1 Allgemeine Struktur	9
3.2 Bildregistrierung	12
3.2.1 Speeded Up Robust Features (SURF)	12
3.2.2 RANdom SAmple Consensus (RANSAC)	17
3.2.3 Bildumwandlung	21
3.3 Differenzbild-Optimierung	32
3.4 Bildverarbeitung	35
3.5 QR-Muster-Detektion	36
4 Untersuchung vom zweiten Verfahren mittels Formerkennung	41
4.1 Allgemeine Struktur	41
4.2 Bildverarbeitung	44
4.2.1 Binarisierung	44
4.2.2 Morphologie Operation	49
4.2.3 Kantendetektion	53
4.2.4 Cross Dilatation	57
4.3 Linienerkennung	58
4.3.1 Hough Transformation	58
4.3.2 Radon Transformation	60

5 Implementierung	63
5.1 Experimentelle Umgebung	63
5.2 Implementierung vom ersten Verfahren	63
5.3 Implementierung vom zweiten Verfahren	66
5.4 Implementierung auf einer Smartphone-GPU	68
6 Evaluierung	70
6.1 Evaluierung vom ersten Verfahren	70
6.2 Evaluierung vom zweiten Verfahren	75
7 Zusammenfassung	78
Abkürzungsverzeichnis	80
Symbolverzeichnis	81
Abbildungsverzeichnis	82
Tabellenverzeichnis	85
Literatur	86

1 Einleitung

In den letzten Jahren hat **Visible Light Communication (VLC)** sowohl in der Wissenschaft als auch in der Industrie große Aufmerksamkeit erregt. Aufgrund seiner Vorteile hinsichtlich Bandbreitenverfügbarkeit, Sicherheit und Datensicherheit wird VLC in vielen Anwendungen als eine bessere Alternative zur Radiofrequenz und Infrarot betrachtet, wie z.B. drahtloses Netzwerk mit Beleuchtungssystem[1], Innenraumpositionierung [2], optische Verbindungen zu elektronischen Chips [3], Körpersensornetzwerken [4], usw. Viele dieser VLC-Techniken wurden entwickelt, um bereits vorhandene Lichterzeugungsgeräte zu nutzen. Ein innovativer Ansatz ist die Verwendung von Display-Kamera-Paaren zur Datenübertragung. Aufgrund der steigenden Performance der beteiligten Schlüsselkomponenten erscheinen Datenraten von bis zu 100 Mbit/s realisierbar, während gleichzeitig eine Videopräsentation für menschliche Zuschauer auf dem gleichen Bildschirm zur Verfügung gestellt werden kann. Mit einem solchen System können viele innovative Medienanwendungen realisiert werden.

1.1 Motivation

In einer aktuellen Forschung untersucht der Lehrstuhl für Kommunikationstechnik an der TU Dortmund ein neuartiges Verfahren der Visible Light Communication: Das DaVid-System, welches eine optische Freiraum-Übertragung unter Verwendung verfügbarer Displays und Kameras durchführt. Ein Highlight dieses Systems ist, dass sie die ursprüngliche Funktionalität des Displays nicht beeinträchtigt; Während der Datenübertragung kann das Display immer noch statische Bilder oder Videos anzeigen, ohne dass menschliche Betrachter die versteckten Datensignale wahrnehmen. Die Übertragungsdaten werden in einem Videosignal in Form geringer Amplitudenänderungen differentiell überlagert. Am Empfänger lässt sich der Display durch eine Kamera oder Smartphone optische aufnehmen. Anschließend können die überlagerten Daten empfangen und decodiert werden. Hierzu ist es notwendig, den Modulationsbereich, der durch die optische Projektion verzerrt wird, mit einigen Operationen wiederherzustellen.

In dieser Arbeit werden zwei Verfahren zur Ausschnittsdetektion für Differenzbilder untersucht und implementiert. Die erste Methode verwendet die Charakteristiken der Datenmodulation

des DaVid Systems, d.h. auf dem Differenzbild wird das QR Muster detektiert, um den Modulationsbereich zu bestimmen. Das zweite Verfahren basiert auf der Geometrie des Displays. Mit Verwendung der Radon Transformation kann der rechteckige Modulationsbereich bestimmt werden. Darüber hinaus wird in dieser Arbeit ein Modul zu Bilderkennung entwickelt, welches die Videostabilität bei Aufnahme mit dem Smartphone aus der Hand verbessern kann. Außerdem durch Einführung des Begriffs der „*Energie*“ wurde eine Optimierungsmethode für Differenzbilder entwickelt. Im Anschluss wurde die Performance der beiden Verfahren evaluiert. Das zweite Verfahren wurde auf einer Smartphone-GPU implementiert.

1.2 Aufbau der Arbeit

Der Rest des Dokuments ist wie folgt organisiert. Das nachfolgende Kapitel gibt eine Beschreibung über das DaVid System. Es stellt die Struktur und Arbeitsweise des DaVid Systems vor und listet verschiedene Anwendungsbereiche auf, die möglicherweise vom DaVid-Konzept profitieren könnten. Der dritte und vierte Absatz enthält Informationen zu beiden Verfahren. Die Struktur und Zusammensetzung der Methoden werden vorgestellt und das Arbeitsprinzip jedes Teils wird detailliert beschreibt. Das nächste, also fünfte Kapitel, zeigt die Implementierung jeder Methode. Darin können die Wirkung jedes Teils in der Methode gesehen werden. Kapitel 6 enthält eine Evaluierung der beiden Verfahren. Die Performance des Verfahrens werden analysiert. Der letzte Abschnitt schließt die Ergebnisse ab und gibt einen Ausblick auf die zukünftige Arbeit.

2 Optische Freiraum-Datenübertragung mittels Videodisplays und Kameras

Diesem Kapitel gibt einen Beschreibung über DaVid System. Zuerst wird eine Einführung vorgestellt. Das Systemmodell und Arbeitsprinzip des Systems werden im anschließenden Abschnitt erläutert. Schließlich folgen mögliche Anwendungsgebiete des Systems. [5]–[8]

2.1 Einführung des DaVids

DaVid ist ein neuartiges Verfahren zur optischen Freiraum-Datenübertragung zwischen einem Display als Sender und einem digitalen Kamera (Smartphone) als Empfänger. Ein grundlegendes Übertragungskonzept von DaVid wird in Abbildung 2.1 gezeigt. Ein flaches Display, wie ein OLED- oder LCD-Bildschirm, zeigt ein Live-Video. Gleichzeitig werden die Daten hinter dem Bild auf die Pixel moduliert, während der Pixel auf dem Bild für menschliche Betrachter nahezu unsichtbar ist. Der Benutzer richtet eine hochauflösende Kamera beziehungsweise Smartphonekamera auf den Bildschirm um die Szene aufzunehmen. Durch den Smartphone Prozessor können die Signale in Daten decodiert werden.

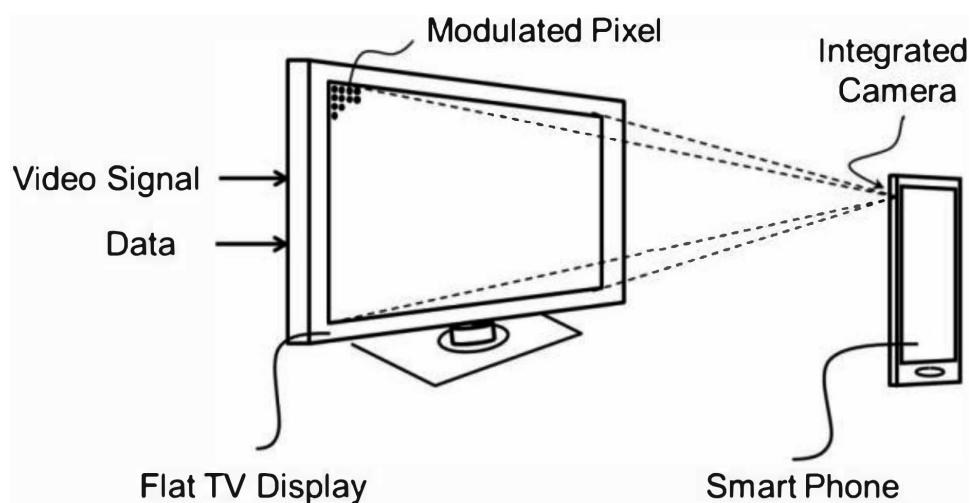


Abbildung 2.1: Eine beispielhafte Implementierung des DaVid-Systems.

2.2 Systemmodell

Das Display enthält eine große Anzahl von Pixel, die jeweils aus einer spezifischen Anordnung von Subpixel zur Darstellung des RGB-Farbraums bestehen. Jeder einzelne Frame des Videos wird durch eine Matrix von Subpixelwerten dargestellt. Das DaVid-System verwendet eine differentielle Modulationsmethode d.h. ein Teil der Videoinformationen wird wiederholt, sodass Daten als ein symmetrischer Manchester-Code moduliert und zu den Videosignalkomponenten hinzugefügt werden. Durch eine zeitliche Synchronisation der Empfängerseite kann die inter symbol interference (ISI) vermieden werden. Nach einer örtlichen Synchronisation mit dem modulierten Display, enthält der Empfänger ein Differenzbild. Da der Randbereich des Differenzbilds ungültig ist, wird das Gebiet durch das Verfahren dieser Arbeit beschränkt. Danach werden die überlagerten Datensequenz durch eine Reihe von Behandlungen, vom Videoinhalt getrennt. Abbildung 2.2 zeigt die schematische Darstellung des DaVid-Systems.

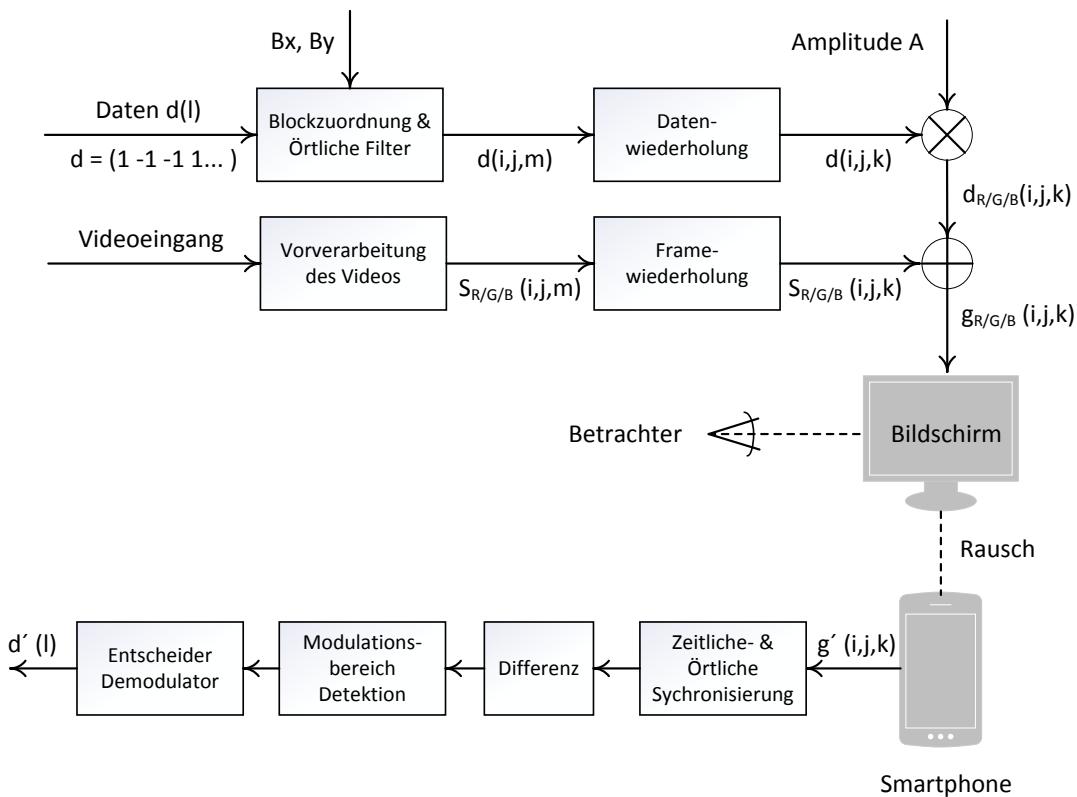


Abbildung 2.2: Schematische Darstellung vom DaVid-System.

2.2.1 Modulationsverfahren

Ein Modulationsverfahren, welches die Videoqualität nicht beträchtlich reduziert, ist sehr kritisch für ein auf Videogeräte basierendes Datenübertragungssystem. Die möglichen Modulationsverfahren im DaVid-System sind:

- Zeitlich-differentielle Modulation der Luminanz
- Zeitlich-differentielle Modulation der Chrominanz
- Örtlich-differentielle Modulation der Luminanz
- Örtlich-differentielle Modulation der Chrominanz

Zeitlich-differentielle Modulationsverfahren lassen eine Reihe von Frames paarweise den gleichen Luminanz- bzw. Chrominanz-Videoinhalt anzeigen. Dagegen haben die benachbarten Pixel bei örtlich-differentielle Modulation die gleichen Videoamplituden. In dieser Arbeit wird nur zeitlich-differentielle Modulation der Chrominanz verwendet, d.h. durch Subtrahieren dieser Frames entsteht ein Datensequenz Differenzbild, welches die beiden Verfahren in dieser Arbeit darauf basiert. Abbildung 2.3 zeigt ein Blockschaltbild einer typischen Senderimplementierung durch zeitlich-differentielle Modulation.

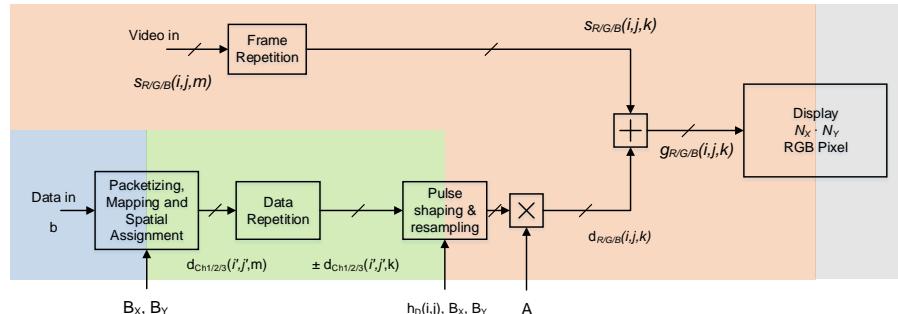


Abbildung 2.3: Blockschaltbild der Signalverarbeitung in zeitlich-differentieller Modulation.

Eine Display enthält in horizontaler Richtung N_x Pixel und in vertikaler Richtung N_y Pixel. Ein Videoeingangssignal wird verarbeitet, um eine Anzeigeeingabe $s(i, j, k)$ auszusenden. Mit zeitlich-differentieller Modulation gleicht der Inhalt dieses Frames des darauf folgenden. Die Indizes i und j definieren die horizontale und vertikale Pixelposition auf dem Display, während k die Nummer des reproduzierten Frames ist. Indiz m symbolisiert den Zähler der Frames in einer Videosequenz. Der Amplitudenbereich des Videosignals sollte begrenzt sein um die Addition kleiner Datenamplituden A ohne Übersteuern zu ermöglichen.

$$s_{R/G/B}(i, j, k+1) = s(i, j, k) \quad (2.1)$$

for $0 \leq i < N_x, 0 \leq j < N_y, k = 2m, m \in \mathbb{Z}$

Die Modulationsamplitude A ist ein wichtiger Parameter für die Datenübertragung. Im Prinzip kann die Amplitude in verschiedenen Kanälen unabhängig gewählt werden um die Systemleistung zu optimieren. In dieser Arbeit werden alle Amplituden auf den gleichen Wert gesetzt.

$$A_R = A_G = A_B = A \quad (2.2)$$

Das differentielle Modulationsverfahren ordnet jede Sequenz von $\{-A, A\}$ zu $d = 0$ bzw. $\{A, -A\}$ zu $d = 1$ zu. Modulierte Datensymbole und verarbeitete Videoamplituden werden addiert, um die Anzeigeeingabe $g(i, j, k)$ zu liefern:

$$\begin{aligned} g_{R/G/B}(i, j, k) &= s_{R/G/B}(i, j, m) + A_{R/G/B} \cdot (2 \cdot d(i, j, m) - 1) \\ g_{R/G/B}(i, j, k+1) &= s_{R/G/B}(i, j, m) - A_{R/G/B} \cdot (2 \cdot d(i, j, m) - 1) \end{aligned} \quad (2.3)$$

2.2.2 DatenBlock

Ein einfaches und unkompliziertes Verfahren um die Anforderungen an die Kameraauflösung zu lockern, ist die Zuordnung jedes Datenbits zu einem Block von $B_X \times B_Y$ Pixel.

$$\begin{aligned} d(l) \rightarrow d(x, y, k) \quad 0 \leq l < L \\ L &= \lfloor N_X/B_X \rfloor \cdot \lfloor N_Y/B_Y \rfloor \\ x &= (l \cdot B_X) \bmod N_X + r_X, \quad r_X = 0 \dots (B_X - 1) \\ y &= \lfloor l / \lfloor N_X/B_X \rfloor \rfloor \cdot B_Y + r_Y, \quad r_Y = 0 \dots (B_Y - 1) \end{aligned} \quad (2.4)$$

In dieser Arbeit ist der Datenblock quadratisch.

$$B_X = B_Y = B. \quad (2.5)$$

2.3 Anwendungsbereiche

Die Datenübertragungsrate des DaVid-Systems beträgt voraussichtlich bis zu 100 Mbit/s. Es gehört zu einer Sichtlinienübertragung für kurze Strecken. Geeignete Abdeckungsbereiche hängen von der Größe des Displays und der Kameraoptik ab. Im Vergleich zu den letzten WLAN-Versionen des Institute of Electrical and Electronics Engineers (IEEE) 802.11, erscheint die Leistung zunächst nicht attraktiv. Dieses spezielle Verfahren der Datenübertragung macht es jedoch zu einem Kandidaten für viele praktische Anwendungen. Außer den Hauptvorteilen von VLC liegt der Vorteil auch in der Option zur Wiederverwendung der bestehenden Hardware, die zum Anzeigen des Videos eingebaut wurde. Ein geeigneter praktischer Anwendungsbereich des DaVid-Systems ist ein öffentlicher Ort, wie eine U-Bahn-Station, ein großes Stadion und Ähnliches. Angenommen in einer Situation, wo Passanten auf eine U-Bahn warten und ihre Software mit Hilfe von Digitalen Werbetafeln aktualisieren können, indem sie ihre Kamera auf die Werbung richten.

Durch Berücksichtigung der Eigenschaften des DaVids, z.B. Synchronisation von Anwendungen und Datenübertragung, können viele attraktive Anwendungszenarien in Betracht gezogen werden. Drei Hauptzonen sind:

- Indoor-individuelle Kommunikation: Kurzstreckenverbindungen basierend auf relativ kleinen (Tablet-Größe) Bildschirmen, Anwendungen wie z.B. die Übertragung von Hintergrundinformationen an Besucher im Museum oder Kiosk.
- Indoor-Multicast-Kommunikation: Streckenabstand ist länger als im ersten Fall auf relativ großem (40-100") Bildschirm, Anwendungen wie z.B. Herunterladen von Anwendungsdateien oder Mediendateien im Kiosk oder Restaurant.
- Freie Kommunikation: Riesige Bildschirme wie in Einkaufszentren oder Sport-Arenen. Anwendungen können denen des zweiten Szenarios ähneln.

Sobald die Dienste auf öffentlichen Bildschirmen implementiert werden, können Benutzer mit Hilfe eines modernen Smartphones mit einer geeigneten Kamera nach der Installation einer neuen App Innovation wahrnehmen.

3 Untersuchung vom ersten Verfahren mittels Mustererkennung

In diesem Kapitel wird auf die Realisierung des ersten Verfahrens eingegangen. Zuerst wird die Bildregistrierung im Detail besprochen. Anschließend läuft die Differenzbild Optimierung und die benötigte Bildverarbeitung. Schließlich wird die QR Mustererkennung erläutert. Zur Implementierung dieses Verfahrens wird Matlab unter der Lizenz der TU Dortmund verwendet.

3.1 Allgemeine Struktur

Dieses Verfahren verwendet die Charakteristiken der Datenmodulation des DaVid Systems, d.h. an jeder Ecke der Datenebene wird ein QR Muster hinzugefügt und dann zusammen mit den Daten hinter dem Bild moduliert. Am Empfänger wird nach einigen Operationen das QR Muster mit den in diesem Kapitel vorgestellten Verfahren detektiert und schließlich der Modulationsbereich bestimmt. Diese Methode löst den unansehnlichen Effekt, dass das QR Muster direkt zum Bild hinzugefügt wird. Außerdem wird die Detektion mit der QR-Mustereigenschaft günstig und effektiv implementiert. Abbildung 3.1 zeigt das Strukturiagramm dieses Verfahrens.

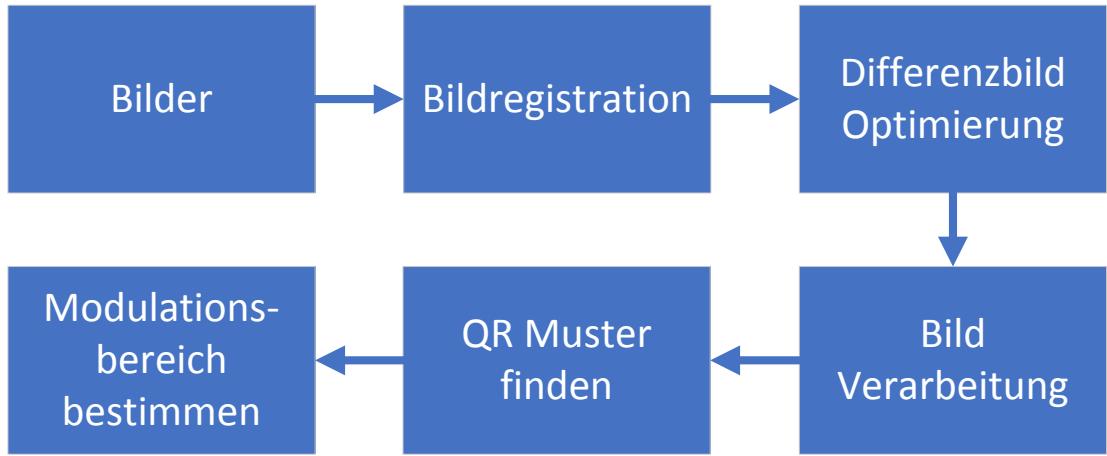


Abbildung 3.1: Strukturdiagramm.

Das Objekt, das mit dieser Methode bearbeitet wird, ist eine aus einer spezifischen Smartphone App "VLCReceiver" gesammelte Bilderreihe. Diese App wurde speziell für das DaVid-System entwickelt und erstellt eine Reihe von Bildern bei jeder Aufnahme. Im Allgemeinen wird die Kamera bei der Aufnahme in der Hand gehalten. Aufgrund von Handschütteln während der Aufnahme entsteht eine leichte Verschiebung zwischen den Bildern. Um dieses Problem zu beheben, wird eine Bildregistrierungsoperation verwendet, die zwei Bilder einer Aufnahme in dasselbe Koordinatensystem konvertiert. Zuerst werden durch SURF Merkmalserkennung Merkmale der Bilder verglichen. Durch Merkmalsextraktion und Merkmalsanpassung werden die Korrespondenzpunkte zwischen den verglichenen Bildern erhalten. Der RANSAC Algorithmus benötigt, um die irrelevanten Korrespondenzen zu beseitigen und die Genauigkeit zu verbessern. Aus dem Kameramodell wird anschließend ein mathematisches Transformationsmodell zwischen den entsprechenden Punkten in den beiden Bildern erstellt. Es ist zu beachten, dass der Prozess zur Lösung des Transformationsmodells als ein nichtlineares Optimierungsproblem angesehen werden kann. Durch die Lösung dieses Problems kann die Transformationsmatrix erhalten werden. In dieser Arbeit wird immer das erste Bild als Referenzbild gewählt um eine Bildregistrierung mit folgenden Bildern vorzunehmen. Die von der Bildregistrierung erhaltenen Bilder werden in dasselbe Koordinatensystem umgewandelt. Es werden je zwei Bilder subtrahiert, wodurch eine Reihe Differenzbilder entstehen. Um die folgende Detektion zu vereinfachen, wird eine Optimierungsoperation mit der Hilfe der geometrischen Eigenschaft des QR Musters unternommen woraufhin ein zu detektierendes Bild erstellt wird. Anschließend werden eine Reihe von Bildverarbeitungen bzw. Binarisierung, Medianfilter, Morphologie Operation vorgenommen. Dadurch können vereinzelt Punkte und Lücken, die durch Rauschen und Fehler verursacht wurden, entfernt werden. Schließlich wird es durch die

Charakteristik des QR Musters, dessen Breiteverhältnis $1 : 1 : 3 : 1 : 1$ beträgt um QR Muster zu detektieren. Dies bestimmt auch den Modulationsbereich. Das Flussdiagramm wird in Abbildung 3.2 gezeigt. Die Details jedes Teils werden in den folgenden Abschnitten beschrieben.

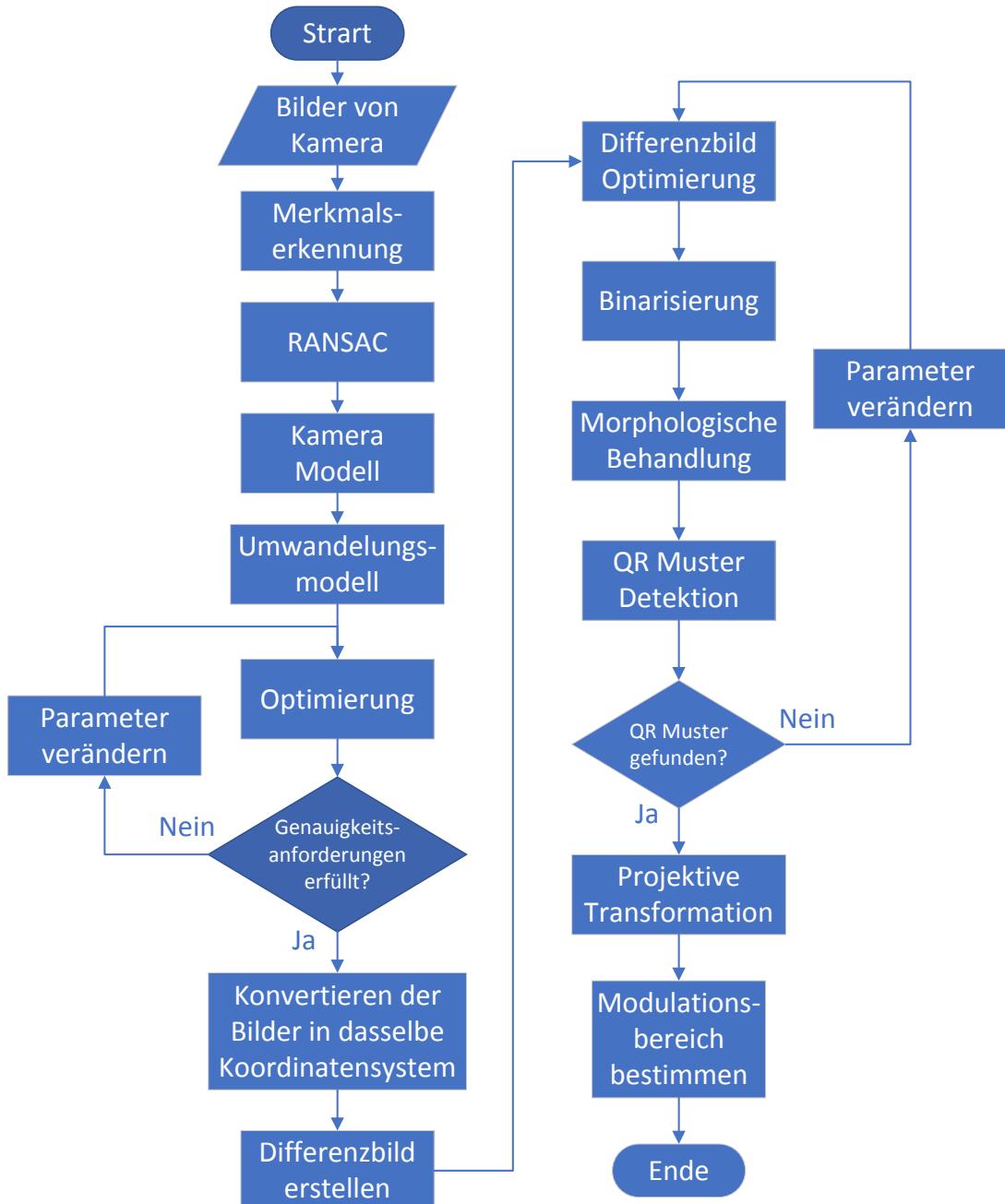


Abbildung 3.2: Flussdiagramm der ersten Methode.

3.2 Bildregistrierung

Es wird angenommen, eine Reihe Bilder wird mit einer Handheld-Kamera aufgenommen. Es kommt aufgrund von Handbewegungen zu einer leichten Verschiebung zwischen den beiden benachbarten Bildern. Wenn diese Bilder subtrahiert werden, um die Differenzbilder zu erhalten, wird das Ergebnis sehr schlecht sein. Um dieses Problem zu lösen, wird hier Bildregistrierung eingeführt. Ein Flussdiagramm der Bildregistrierung wird in Abbildung 4.2 gezeigt.

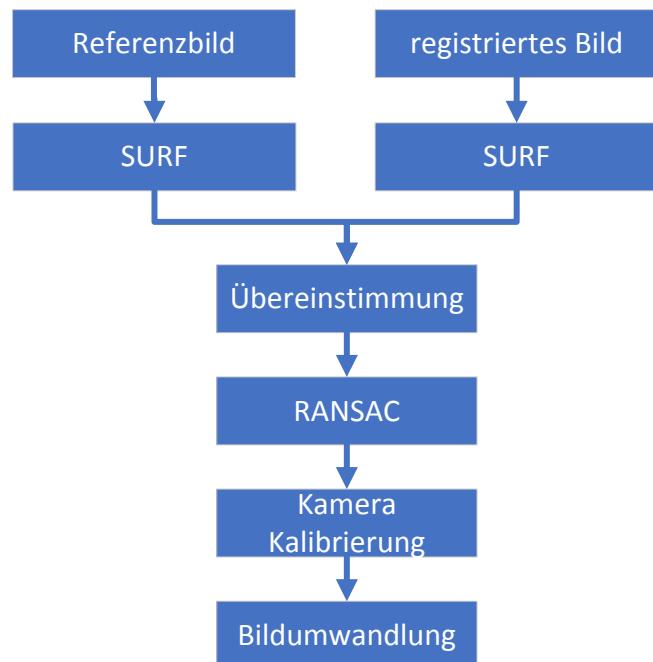


Abbildung 3.3: Flussdiagramm der Bildregistrierung.

3.2.1 SURF

In jedem Bild gibt es eindeutige Pixelwertpunkte, d.h. Merkmalspunkte des Bildes. Um die Bilder in dasselbe Koordinatensystem zu transformieren, müssen diese Merkmalspunkte erkannt und analysiert werden. Deswegen ist es in diesem Verfahren besonders wichtig, die Merkmalspunkte eines Bildes zu definieren und zu finden. Um dieses Problem zu lösen, wird das Konzept der Merkmalserkennung eingeführt. Diese wird oft in Computer Vision und Bildverarbeitungsbereichen verwendet. Die Merkmalserkennung beinhaltet Verfahren zum Berechnen von Abstraktionen von Bildinformationen und zum Treffen lokaler Entscheidungen an jedem Bildpunkt, ob es an einem Punkt ein Bildmerkmal eines bestimmten Typs gibt oder nicht.

Einige typische Merkmalserkennungen sind z.B. Kantenerkennung, Eckenerkennung, Tropfenerkennung und so weiter.

In dieser Arbeit wird das SURF [9] genutzt. Dieser ist ein patentierter lokaler Merkmal-Detektor und Deskriptor und kann für Aufgaben wie Objekterkennung, Bildregistrierung, Klassifizierung oder 3D-Rekonstruktion angewandt werden. Diese Merkmalserkennung ist eine verbesserte Version von **Scale Invariant Feature Transform** (SIFT) Merkmalserkennung, die Haar-Wavelet verwendet um die Gradientenoperation in der SIFT-Methode zu ersetzen, und gleichzeitig eine Integralgraph-Technik für schnelle Berechnungen verwendet. Die Faltung bezieht sich nur auf das vorherige Bild. Durch die Vergrößerung des Bildkerns kann das Heruntertaktungsverfahren realisiert werden. Die Geschwindigkeit von SURF ist 3-7 mal schneller als die von SIFT welche in den meisten Fällen der Leistung von SIFT entspricht. Daher wird SURF in vielen Anwendungen eingesetzt, insbesondere in Anwendungen in denen die Laufzeitanforderungen hoch sind.

Der Verlauf einer SURF Merkmalserkennung ist wie folgend:

- **Aufbau einer hessischen Matrix.**

Die Hesse-Matrix stellt den Kern des SURF Algorithmus dar. Zur Vereinfachung der Operation wird für die Funktion $f(x, y)$ angenommen, dass sich die Hesse-Matrix H aus partiellen Ableitungen und Funktionen zusammensetzt.

$$H(f(x, y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \cdot \partial y} \\ \frac{\partial^2 f}{\partial x \cdot \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (3.1)$$

Determinante der H-Matrix ist wie folgt:

$$\det(H) = \frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \cdot \partial y} \right)^2 \quad (3.2)$$

Der Wert der Determinante ist der Eigenwert der H-Matrix. Durch dessen positiven und negativen Wert wird bestimmt, ob der Punkt ein Extrempunkt ist oder nicht. Im SURF Algorithmus wird das Bildpixel $I(x, y)$ anstelle des Funktionswertes $f(x, y)$ verwendet. Es wird eine Gauß-Funktion zweiter Ordnung als Filter verwendet. Die zweiten partiellen Ableitungen können durch Faltung zwischen bestimmten Kernen berechnet werden. Dadurch können die Werte der drei Matrixelemente der H-Matrix ebenfalls berechnet werden.

$$\begin{aligned}\mathcal{H}(\mathbf{x}, \sigma) &= \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix} \\ L(\mathbf{x}, \sigma) &= G(\sigma) * I(\mathbf{x}) \\ G(\sigma) &= \frac{\partial^2 g(\sigma)}{\partial x^2}\end{aligned}\tag{3.3}$$

Hier bedeutet $L_{xx}(\mathbf{x}, \sigma)$ die Faltung der zweiten Gaußschen Ableitung $G(\sigma)$ mit dem Bild I im Punkt $\mathbf{x}(x,y)$, das selbe für $L_{xy}(\mathbf{x}, \sigma)$ und $L_{yy}(\mathbf{x}, \sigma)$. Auf diese Weise kann der Wert der Determinante für jedes Pixel in dem Bild berechnet werden und dieser Wert kann verwendet werden um den Merkmalspunkt festzustellen. Zur einfacheren Anwendung schlägt Herbert Bay[9] vor, L mit einer Approximation zu ersetzen. Um den Fehler zwischen dem genauen Wert und der Approximation auszugleichen, kann die H-Matrix-Determinante wie folgt ausgedrückt werden:

$$\det(\mathcal{H}_{\text{Approx}}) = D_{xx}D_{yy} - (0.9D_{xy})^2\tag{3.4}$$

• Erstellen des Maßstabraums

Der Maßstabraum $L(\mathbf{x}, \sigma)$ des Bildes ist die Darstellung dieses Bildes bei unterschiedlichen Auflösungen(Skalierungen). Im Bereich der Computer Vision wird der Maßstabraum symbolisch als Bildpyramide ausgedrückt, wobei die Eingangsbildfunktion wiederholt mit dem Kern der Gaußschen Funktion gefaltet und wiederholt unterabgetastet wird. Diese Methode wird hauptsächlich für die Implementierung des SIFT Algorithmus verwendet. Jede Bildschicht hängt jedoch von der vorherigen Bildschicht ab, weshalb das Bild in der Größe angepasst werden muss. Daher büßt diese Berechnungsmethode viel Leistung ein. Das Gegenstück in SURF ist die Erhöhung der Größe des Bildkerns. Dies ist ein Unterschied zwischen dem SIFT Algorithmus und dem SURF Algorithmus bei der Verwendung des Pyramidenprinzips. Der Algorithmus ermöglicht, dass mehrere Bilder des Maßstabraums gleichzeitig verarbeitet werden, ohne dass das Bild unterabgetastet wird, wodurch die Leistung des Algorithmus verbessert wird. Das linke Bild in Abbildung 3.4 ist eine Pyramidenstruktur, die auf herkömmliche Weise erstellt wird. Die Größe des Bildes wird geändert, und die Operation wird die Unterebene unter Verwendung der Gaußschen Funktion wiederholt glätten. Der SURF Algorithmus auf der rechten Seite in Abbildung 3.4 erhält das ursprüngliche Bild und ändert nur die Filtergröße.

• Präzise Lokalisierung von Feature-Punkten

Die von der Hesse-Matrix verarbeitete Größe jedes Pixels wird mit den 26 Punkten der drei Dimensionen im Raum, wie in Abbildung 3.5 gezeigt, verglichen. Wenn einer davon das Maxi-

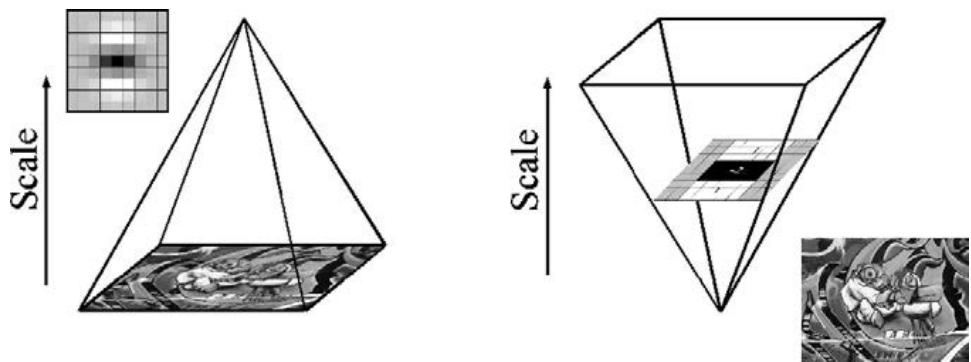


Abbildung 3.4: Scale space.

mum oder Minimum dieser 26 Punkte ist, wird dieser als vorläufiger Merkmalspunkt beibehalten. Das dreidimensionale lineare Interpolationsverfahren wird angewandt, um die Merkmalspunkte des Subpixel-Niveaus zu erhalten. Die Punkte, deren Werte kleiner als ein bestimmter Schwellenwert sind, werden ebenfalls entfernt.

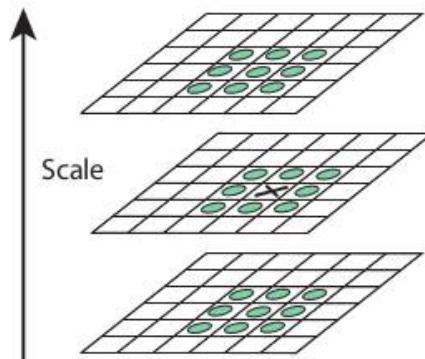


Abbildung 3.5: Extremwert Erkennung.

• Hauptrichtungsermittlung

SIFT wählt die Hauptrichtung des Merkmalspunkts unter Verwendung des Gradientenhistogramms im Merkmalspunktbereich aus. Die Richtung, darin der Bin-Wert des Histogramms der größte ist oder 80% des maximalen Bin-Werts überschreitet, wird als Hauptrichtung des Merkmalspunkts genommen. Dagegen beim SURF, wird nicht das Gradientenhistogramm, sondern die Haar-Wavelet-Eigenschaft im Merkmalspunktbereich statistisch analysiert. Das heißt, im Bereich der Merkmalspunkte wird (zum Beispiel innerhalb eines Kreises mit einem Radius von $6s$, wobei s der Maßstab ist auf dem der Punkt liegt), die Summe der Horizontal-Haar-Wavelet-Merkmale und der Vertikal-Haar-Wavelet-Merkmale aller Punkte im 60-Grad-Sektor($\pi/3$) gezählt. Die Größe des Haar-Wavelets beträgt $4s$, sodass jeder Sektor einen Wert bekommt. Anschließend wird der 60-Grad-Sektor in einem bestimmten Intervall gedreht. Schließlich lässt die Richtung des Sektors anhand des Maximalwerts als Hauptrichtung vom

Merkmalspunkt nehmen. Ein schematisches Diagramm des Prozesses ist wie folgt in Abbildung 3.6.

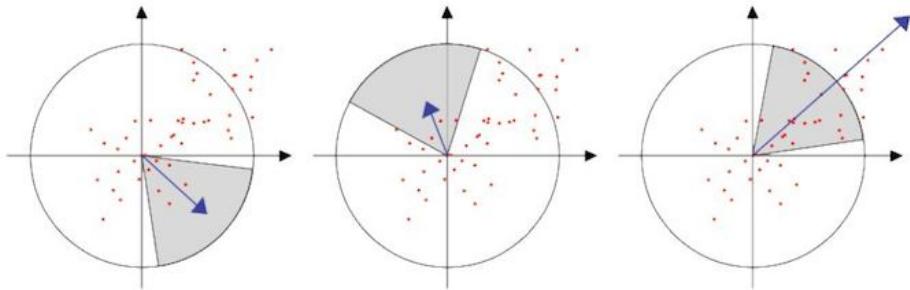


Abbildung 3.6: Dominante Orientierung feststellen.

- **Merkmalspunkt Deskriptor Generierung**

SURF generiert einen quadratischen Rahmen um den Merkmalspunkt. Die Größe der Seite des Rahmens ist $20s$ (s ist die Skala, dabei ein Merkmalspunkt erkannt wird). Die Richtung des Rahmens ist die Hauptrichtung, die im vorherigen Schritt erfasst wurde. Der Rahmen wird dann in 16 Unterbereiche unterteilt, von denen jeder die Haar-Wavelet-Merkmale der horizontalen und vertikalen Richtungen von 25 Pixeln berechnet. Hier sind die horizontalen und vertikalen Richtungen relativ zur Hauptrichtung. Das Haar-Wavelet-Merkmal ist die Summe der horizontalen Richtungswerte, die Summe der absoluten Werte in der horizontalen Richtung, die Summe der vertikalen Richtungen und die Summe der absoluten Werte in der vertikalen Richtung. Das schematische Diagramm in Abbildung 3.7 zeigt diesen Prozess.

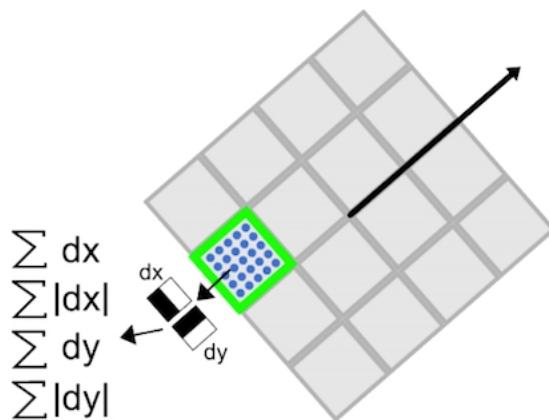


Abbildung 3.7: Merkmalspunkt Deskriptor.

Auf diese Weise hat jeder kleine Bereich 4 Werte, so dass jeder Merkmalspunkt über einen $16 \times 4 = 64$ dimensionalen Vektor verfügt, der nur halb so groß wie SIFT(128 Dimension) ist, und deshalb den Anpassungsprozess beim Merkmalanpassungsprozess stark beschleunigt. Die

folgende Abbildung 3.8 zeigt die Merkmalspunkte, die durch den SURF-Algorithmus erhalten werden.

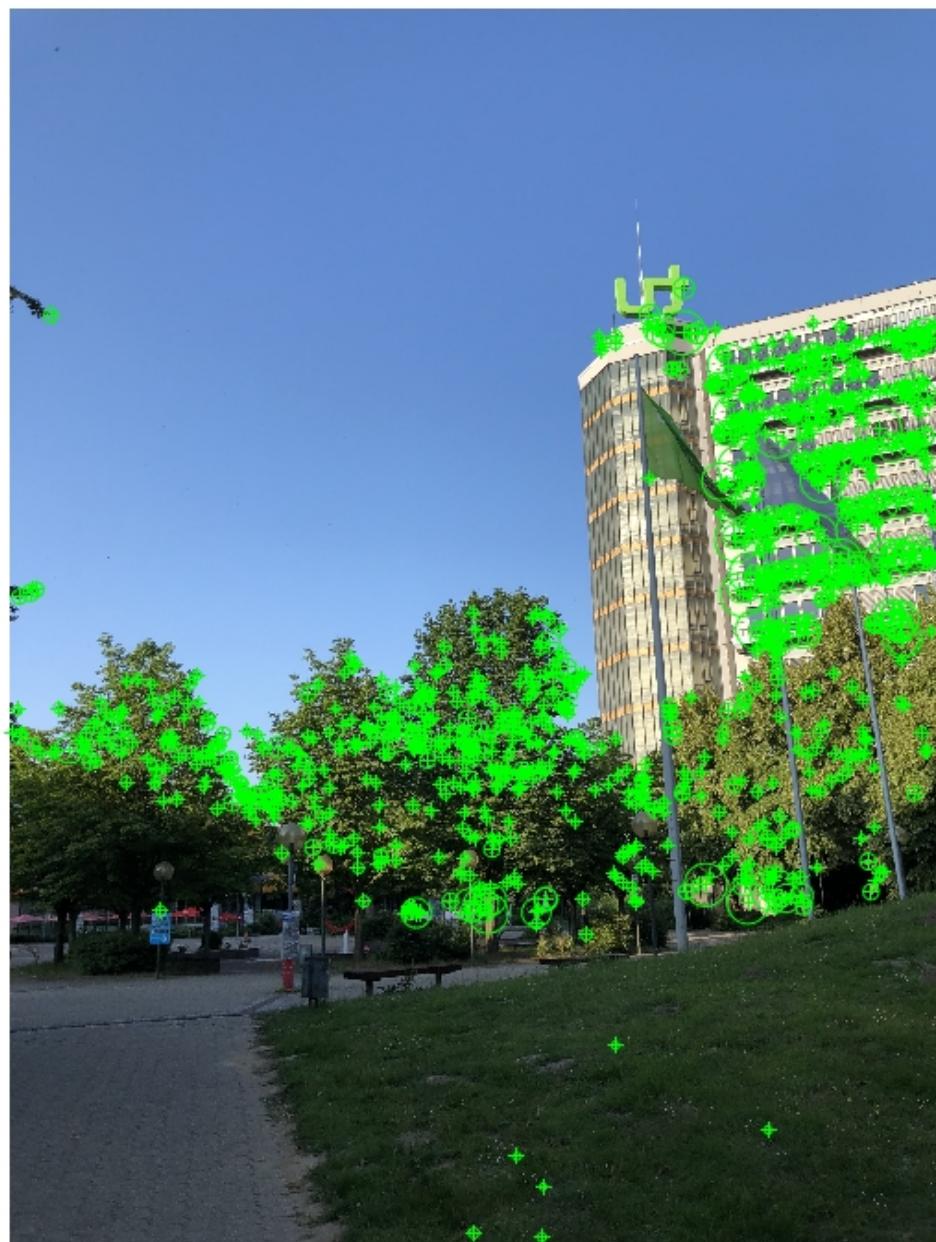


Abbildung 3.8: SURF Merkmale.

3.2.2 RANSAC

Nach der SURF Merkmalserkennung werden die Merkmale von zwei benachbarten Bildern generiert. Diese Merkmale werden dann extrahiert und abgeglichen um entsprechende Punk-

te in den benachbarten zwei Bildern zu erhalten. Leider verbleiben durch diese Operation immernoch viele fehlerhafte zusammenpassende Paare. Deswegen wird hier RANSAC verwendet, um die falschen Punkte zu beseitigen.

Der im Jahre 1981 vorgestellte RANSAC Algorithmus von Fischler und Bolles [10], ist ein allgemeiner Parameterschätzungsansatz um den großen Anteil von Ausreißern in den Eingabedaten zu bewältigen. Im Gegensatz zu vielen der üblichen robusten Schätzverfahren wie M-Schätzer und kleinste Quadrate, die von der Computer Vision Community aus der Statistik-Literatur übernommen wurden, wurde RANSAC von der Computer-Vision-Community entwickelt.

Ein einfaches Beispiel ist in der Abbildung 3.9 dargestellt. Das Ziel besteht darin, die am besten geeignete Linie unter einer Menge von Datenpunkten zu finden. Wenn es die einfache Methode der kleinsten Quadrate verwendet um diese Linie zu finden, wie auf der linken Seite gezeigt, kann es leider nicht richtig funktionieren, da die Methode der kleinsten Quadrate von allen Datenpunkten beeinflusst wird. Dagegen kann mit RANSAC das Modell nur mit Potenziell korrekten Punkten berechnet werden, wie die Ergebnisse auf der rechten Seite zeigen.

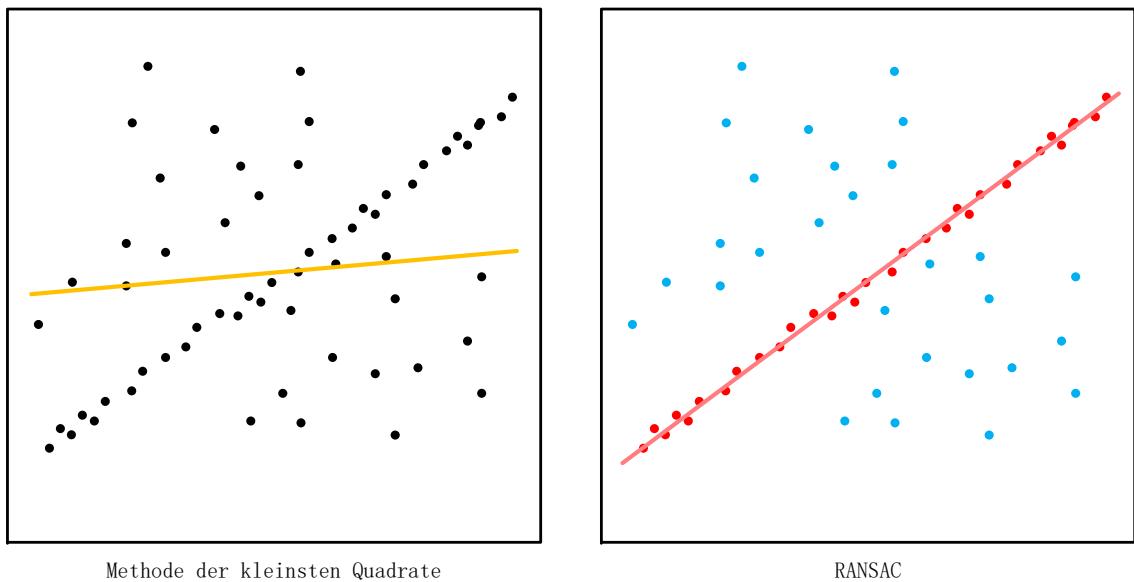


Abbildung 3.9: Linien Detektion.

RANSAC ist ein Wiederholungsprobennahme-Verfahren, welches durch die minimale Anzahl von Beobachtungspunkten (Datenpunkten) die Kandidatenlösung generiert. Diese Datenpunkte sind erforderlich, um die zugrundeliegenden Modellparameter zu schätzen worauf Fischler und Bolles [10] hingewiesen haben. Die Erhaltung einer anfänglichen Lösung und

Ausschließung der Ausreißer mit dem RANSAC Verfahren benötigt nicht so viele Daten, sondern verwendet die kleinste mögliche Menge und fährt fort, diese Menge mit konsistenten Datenpunkten zu vergrößern.

Der grundlegende Algorithmus ist wie folgt zusammengesetzt:

- Zufällige Auswahl der Mindestanzahl erforderlicher Punkte zum Bestimmen der Modellparameter.
- Lösen der Parameter des Modells.
- Bestimmung der Punkte aus der Menge aller Punkte, welche mit einer vordefinierten Toleranz ϵ übereinstimmen
- Wenn ein Bruchteil der Anzahl von Inlieren über die Gesamtzahl der Punkte in dem Satz einen vordefinierten Schwellenwert τ überschreitet, schätzen die Modellparameter mit allen identifizierten Inlieren und terminieren.
- Ansonsten Wiederholung der Schritte 1 bis 4 (maximal N-mal).

N ist die Anzahl der Iterationen. Diese wird hoch genug gewählt, um die Wahrscheinlichkeit p (normalerweise 0.99) sicherzustellen, dass mindestens eine der Gruppen von Stichproben keinen Ausreißer enthält. Die Wahrscheinlichkeit, dass bei N mal iterieren mit erforderlich minimaler Anzahl von Punkten (hier m annahmen) mindestens ein Ausreißer mit ausgewählt wird, berechnet sich folgendermaßen:

$$1 - p = (1 - u^m)^N \quad (3.5)$$

u stellt die Wahrscheinlichkeit dar, dass jeder ausgewählte Datenpunkt ein Inlier ist. Dagegen ist $v = 1 - u$ die Wahrscheinlichkeit, dass jeder ausgewählte Datenpunkt ein Ausreißer ist. Durch einige Gleichheitsumwandlungen kann die Anzahl der Iterationen ausgedrückt werden als:

$$N = \frac{\log(1 - p)}{\log(1 - (1 - v)^m)} \quad (3.6)$$

Abbildung 3.10 zeigt die passenden Punkte durch Merkmalübereinstimmung mit der SURF Detektion. Es ist ersichtlich, dass es viele fehlerhafte Kombinationen gibt. Durch die Anwendung von RANSAC kann dieses Problem effektiv gelöst und die übereinstimmenden Punkte verfeinert werden, wie das Ergebnis in Abbildung 3.11 zeigt.



Abbildung 3.10: Ohne RANSAC.

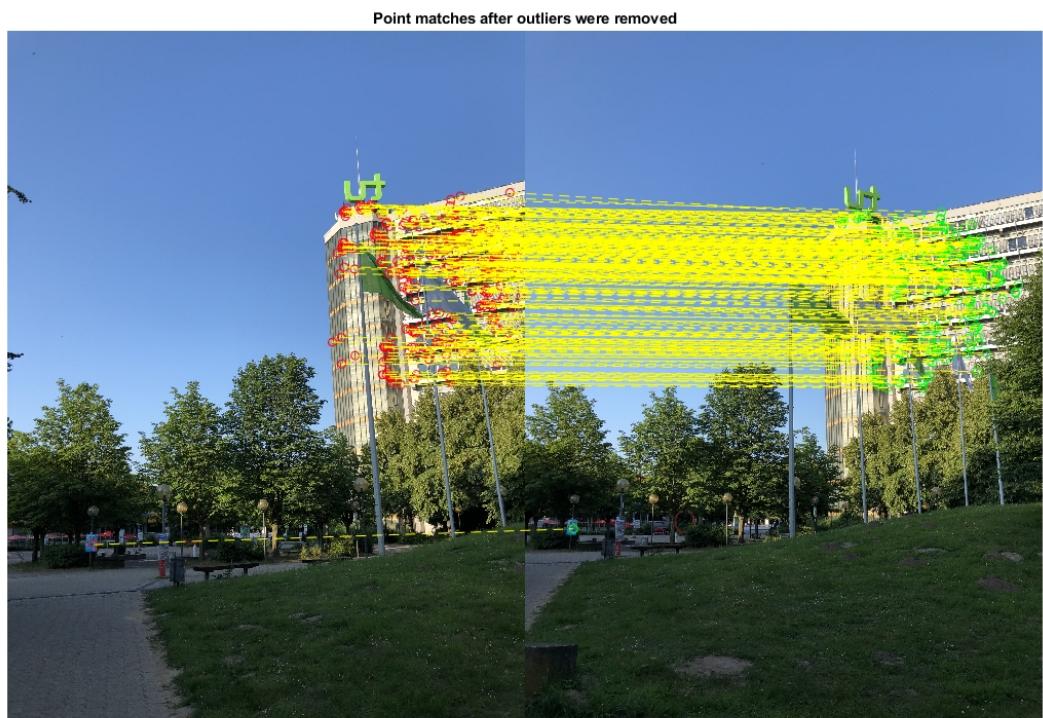


Abbildung 3.11: Mit RANSAC.

3.2.3 Bildumwandlung

Wie in dem vorherigen Abschnitt vorgestellt, kann durch Verwenden des SURF übereinstimmende Punkte in aufeinanderfolgenden Bildern gefunden werden, anschließend lassen sich durch RANSAC die Ausreißer ausschließen. Das Ziel dieses Abschnitts besteht darin, das Bild in dasselbe Koordinatensystem zu konvertieren. Der erste Schritt ist ein Kameramodell zu erstellen und anschließend die Umwandlungsbeziehung zwischen den entsprechenden Punkten in den zwei Bildern zu erhalten, um schließlich durch den Optimierungsalgorithmus die endgültige Transformationsmatrix zu erhalten. Die verschiedenen Schritte werden im Folgenden detailliert beschrieben.

Kamera Modell

Das Modell der Lochkamera ist in Abbildung 3.12 dargestellt. In dem Modell ist O_C das optische Zentrum(Fokus) und f ist die Kamerabrennweite.

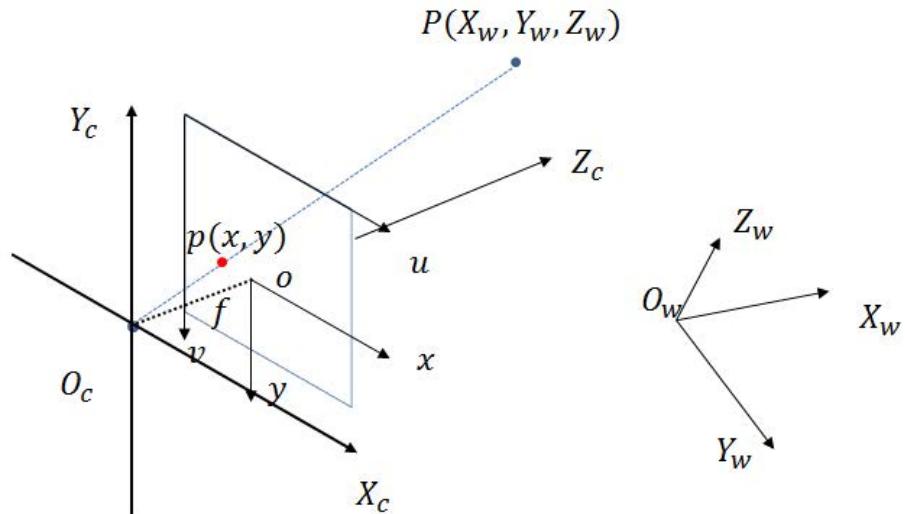


Abbildung 3.12: Modell einer Lochkamera.

Die vier Koordinatensysteme im Modell sind wie folgt definiert:

- 3D Weltkoordinatensystem $P(X_W, Y_W, Z_W)$
Punktkoordinaten werden durch homogene Koordinaten dargestellt: $\tilde{X}_W \sim (X_W, Y_W, Z_W, 1)^T$
- 3D Kamerakoordinatensystem $C(X_C, Y_C, Z_C)$
Punktkoordinaten werden durch homogene Koordinaten dargestellt: $\tilde{X}_C \sim (X_C, Y_C, Z_C, 1)^T$

- 2D Bildabbildung Koordinatensystem $p(x, y)$
Punktkoordinaten werden durch homogene Koordinaten dargestellt: $\tilde{x} \sim (x, y, 1)^T$
- 2D Bildpixel Koordinatensystem $I(u, v)$
Punktkoordinaten werden durch homogene Koordinaten dargestellt: $\tilde{u} \sim (u, v, 1)^T$

Unter diesem Modell wird ein 3D-Punkt im Weltkoordinatensystem durch drei Koordinaten den 2D-Bildpixelkoordinaten zugeordnet.

(1). 3D-Weltkoordinatensystem zum 3D-Kamera-Koordinatensystem.

Die Transformation vom Weltkoordinatensystem zum Kamerakoordinatensystem ist eine Starrekörpertransformation, d.h. das Objekt verformt sich nicht, sondern es rotiert nur und wird parallel verschoben. Diese Transformation wird in Abbildung 3.13 gezeigt. R ist die Rotationsmatrix und T die Translationsmatrix.

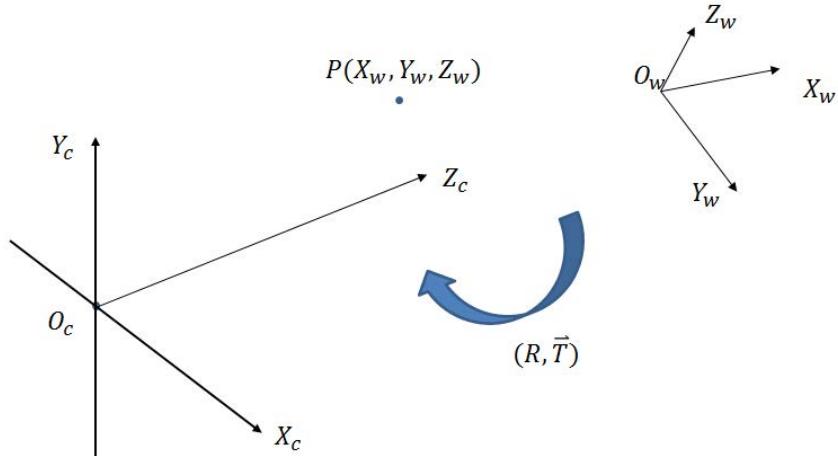


Abbildung 3.13: Transformation vom Weltkoordinatensystem zum Kamerakoordinatensystem.

Um die entsprechende Rotationsmatrix zu erhalten, wird die Koordinatenachsen um verschiedene Winkel gedreht. Ein simples Beispiel wird in Abbildung 3.14 gezeigt.

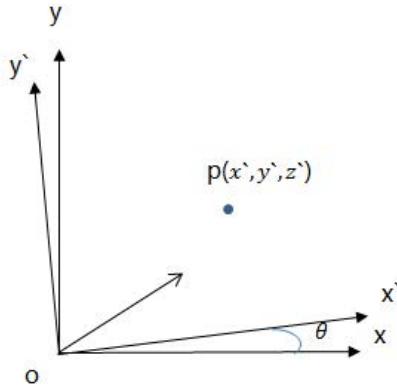


Abbildung 3.14: Rotation um Z-Achse.

Aus dem Bild kann leicht entnehmen werden:

$$\begin{cases} x = x' \cos \theta - y' \sin \theta \\ y = x' \sin \theta + y' \cos \theta \\ z = z' \end{cases} \quad (3.7)$$

In Matrixform wie folgt ausgedrückt:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_1 \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad (3.8)$$

In ähnlicher Weise dreht sich die x-Achse, y-Achse um φ und ω Grad wie folgt:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_2 \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad (3.9)$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos \omega & 0 & \sin \omega \\ 0 & 1 & 0 \\ -\sin \omega & 0 & \cos \omega \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_3 \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad (3.10)$$

Schließlich kann die Rotationsmatrix erhalten werden:

$$R = R_1 \cdot R_2 \cdot R_3 \quad (3.11)$$

Durch das Kombinieren der obigen Ergebnisse, lassen sich die Koordinaten von Punkt P im Kamerakoordinatensystem bestimmen:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = R \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T \quad (3.12)$$

Im homogenen Koordinatensystem dargestellt:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ \vec{0} & 1 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3.13)$$

(2). 3D-Kamera-Koordinatensystem zum 2D-Bildabbildung Koordinatensystem.

Die Transformation vom Kamerakoordinatensystem zum Bildkoordinatensystem gehört zur perspektivischen Projektionsbeziehung von 3D zu 2D, wie in Abbildung 3.15 gezeigt.

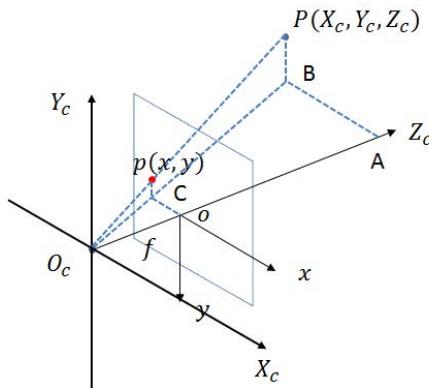


Abbildung 3.15: Transformation vom Kamerakoordinatensystem zum Bildkoordinatensystem.

Es gibt zwei Paare mit ähnlichen Dreiecken:

$$\begin{aligned} \triangle ABO_C &\sim \triangle oCO_c \\ \triangle PBO_C &\sim \triangle pCO_c \end{aligned} \quad (3.14)$$

Aus ähnlichen Dreiecksbeziehungen kann diese Gleichung erstellt werden:

$$\frac{AB}{oC} = \frac{AO_C}{oO_C} = \frac{PB}{pC} = \frac{X_C}{x} = \frac{Z_C}{f} = \frac{Y_C}{y} \quad (3.15)$$

Durch die Transformationsgleichung folgt:

$$x = f \cdot \frac{X_C}{Z_C}, y = f \cdot \frac{Y_C}{Z_C} \quad (3.16)$$

Im homogenen Koordinatensystem dargestellt:

$$Z_C \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \quad (3.17)$$

Zu dieser Zeit ist die Einheit des Projektionspunkts p noch nicht Pixel, sondern Millimeter und muss weiter in das Pixelkoordinatensystem umgewandelt werden.

(3). 2D-Bildabbildung Koordinatensystem zum 2D-Bildpixel Koordinatensystem.

Das Pixelkoordinatensystem und das Bildkoordinatensystem befinden sich alle auf der Abbildungsebene, jedoch sind die jeweiligen Ursprünge und Maßeinheiten unterschiedlich. Der Ursprung des Bildkoordinatensystems ist der Schnittpunkt der optischen Achse der Kamera und der Abbildungsebene, üblicherweise der Mittelpunkt der Abbildungsebene oder der Hauptpunkt. Die Einheit des Bildkoordinatensystems ist Millimeter welche eine physikalische Einheit ist. Die Einheit des Pixelkoordinatensystems ist Pixel. Wie gewohnt wird beschrieben in welcher Zeile und Spalte ein Pixel ist. So ist der Übergang zwischen den beiden Koordinatensystem wie in Abbildung 3.16 dargestellt.

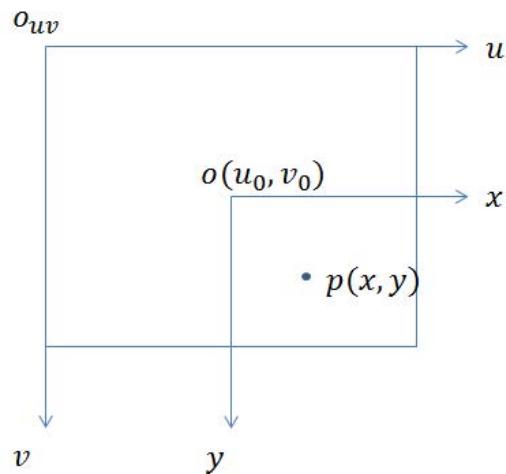


Abbildung 3.16: Konvertierung vom Bildkoordinatensystem zum Pixelkoordinatensystem.

d_x und d_y ist die Größe jedes Pixel in den X- und Y-Achsenrichtungen. Jedes Pixel des Bildes hat die folgende Beziehung zwischen den zwei Koordinatensystemen.

$$\begin{cases} u = \frac{x}{d_x} + u_0 \\ v = \frac{y}{d_y} + v_0 \end{cases} \quad (3.18)$$

Im homogenen Koordinatensystem dargestellt:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3.19)$$

Es ist erwähnenswert, dass das Bildkoordinatensystem eine zweidimensionale Ebene(Bildebene), praktisch die Oberfläche des Kamera-CCD-Sensors ist. Jeder CCD-Sensor hat eine bestimmte Größe und Auflösung. Diese beiden Faktoren bestimmen die Konvertierungsbeziehung. Gegeben sei ein simples Beispiel: Eine Größe des CCD-Sensors ist $8\text{ mm} \times 6\text{ mm}$, die Auflösung dafür ist $640\text{ pixels} \times 480\text{ pixels}$. Die Beziehung zwischen mm und Pixel ist 80 pixel/mm . Ist die physikalische Größe jedes Pixels des CCD-Sensors $d_x \times d_y$, entspricht diese $d_x = d_y = \frac{1}{80}\text{ mm}$.

Durch die Umwandlung der obigen vier Koordinatensysteme kann ein Punkt vom Weltkoordinatensystem zum Pixelkoordinatensystem extrahiert werden.

$$\begin{aligned} Z_C \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} R & t \\ \vec{0} & 1 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} R & t \\ \vec{0} & 1 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \end{aligned} \quad (3.20)$$

Die erste Matrix der rechten Gleichung ist die allgemein bekannte interne Referenz der Kamera. Dagegen ist die zweite Matrix die externe Referenz der Kamera. Beide Parameter der Kamera können durch Zhang Zhengyou [11] Kalibrierung erhalten werden. Einige typische Kamera Parameter vom Werk aus sind der Tabellen 3.1 zu entnehmen.

Wenn die internen und externen Parameter der Kamera bekannt sind, ist es möglich daraus die Projektionsmatrix zu erstellen um daraus die Bildkoordinaten jeden beliebigen räumlichen

Tabelle 3.1: Parameter der Kameras im Vergleich.

Parameter	Google Pixel	Google Pixel2	Iphone X
Sensor Größe "	1/2.3	1/2.6	1/3
Bild Auflösung pixels	4048×3036	4032×3024	4032×3024
Pixel Größe μm	1.544	1.4	1.22
Brennweite mm	4.67	4.47	3.99
Formatfaktor 35mm	5.55	6.04	7.02

Punkts zu erhalten. Wenn die Position $m(u, v)$ einer Bildkoordinate und die Parameter innerhalb und außerhalb der Kamera bekannt sind, kann der entsprechende Punkt in Weltkoordinate nur ungefähr bestimmt werden. Der Grund dafür ist, dass die Z_c Information während des Projektionsprozesses eliminiert wird.

Umwandlungsmodelle

Als nächstes wird die Umwandlungsbeziehung zwischen den entsprechenden Punkten in den beiden Bildern vorgenommen. Zuerst wird in dieser Arbeit eine vereinfachte Situation betrachtet, d.h. nur mit Rotationseinfluss. Abbildung 3.17 zeigt das 3D-Rotationsbewegungsmodell der Kamera. Die Position des optischen Zentrums ändert sich, während der Kamerabewegung im Drehbewegungsmodell der Kamera, nicht. Unter diesem Modell ist die Abbildungsbeziehung zwischen dem Punkt X im Weltkoordinatensystem und der Bildkoordinate x im homogenen Koordinatensystem dargestellt:

$$x = KRX, X = \lambda K^{-1}x \quad (3.21)$$

Wobei K der interne Parameter der Kamera ist, wie zuvor definiert. λ ist der unbekannte Skalierungsfaktor, d.h. dass unter dem Kameramodell die Quelle jeder Bildpunktkoordinate einem Strahl zugeordnet ist.

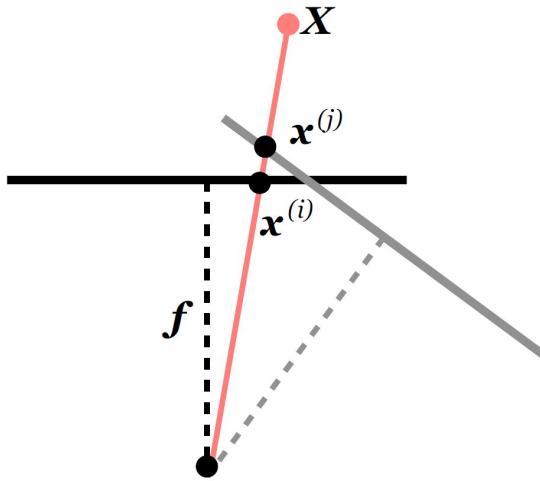


Abbildung 3.17: Rotationsbewegungsmodell.

Nun wird die Beziehung zwischen Bildpunkten in einem Framepaar zwei verschiedener Kameraausrichtungen hergeleitet (siehe Abbildung 3.17). Für einen Weltkoordinatepunkt X sind die projizierten Punkte x_i und x_j in der Bildebene von zwei Bildern i und j gegeben durch

$$x_i = KR_i X, x_j = KR_j X \quad (3.22)$$

Nach weiterem anordnen und ersetzen von X , wird eine Beziehung aller Punkte im Bildrahmen i auf alle Punkte im Rahmen j erhalten:

$$x_j = KR_j R_i^T K^{-1} x_i \quad (3.23)$$

Bisher wurde nur die Beziehung zwischen zwei Bildern der selben Aufnahme betrachtet. Diese Einschränkung wird gelockert, indem Frames von einer Kamera, die sich gemäß R dreht, zu einer anderen Kamera, die sich gemäß R' dreht, abgebildet werden. Es gibt eine Hypothese, dass beide Kamerazentren sich im Ursprung befinden. Mit der Warping-Matrix können die abgebildeten Punkte der Kameras wie folgt definiert werden:

$$W = KR' R^T K^{-1} \quad (3.24)$$

Hier wird das erste Bild als Referenzbild mit einem Rotationswinkel von 0 Grad gewählt, wodurch die Gleichung vereinfacht werden kann:

$$W = KRK^{-1} \quad (3.25)$$

Kombiniert mit Formel 3.23 kann es ausgedrückt werden als:

$$x_j = Wx_i \quad (3.26)$$

Diese Gleichung zeigt, dass jeder Punkt in dem Bild i nach der Transformationsmatrix W in einen entsprechenden Punkt in dem Bild j umgewandelt werden kann. Weiter in einem allgemeineren Fall, d.h. von nun an nicht nur mit Rotationseinfluss, sondern auch Translationseinfluss. Der Ableitungsverlauf ist im Allgemeinen gleich. Der Hauptunterschied ist die Anzahl der Parameter, die ursprünglich nur 3 Rotationsparameter aber jetzt 6 Parameter einschließlich 3 Rotationsparameter und 3 Translationsparameter beinhalten. Die neue Warping-Matrix sieht folgendermaßen aus:

$$W = \begin{bmatrix} f & 0 & \frac{w}{2} & 0 \\ 0 & f & \frac{h}{2} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} R_{11} & R_{21} & R_{31} & 0 \\ R_{12} & R_{22} & R_{32} & 0 \\ R_{13} & R_{23} & R_{33} & 0 \\ t1 & t2 & t3 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{f} & 0 & -\frac{w}{2f} & 0 \\ 0 & \frac{1}{f} & -\frac{h}{2f} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.27)$$

Transformationsoptimierung

In den letzten beiden Abschnitten wurde die Transformationsmatrix zwischen den beiden Bildern aus dem Kameramodell abgeleitet. Hier in diesem Abschnitt werden die Parameter der Transformationsmatrix berechnet, also wird nun die Transformationsmatrix bestimmt. Hier sind die entsprechenden Punkte von zwei benachbarten Bildern x_i, x_j bekannt und die Umwandlungsbeziehung dazwischen wird als die Formel 3.26 dargestellt. Angesichts dieser Bedingung kann die Berechnung der Transformationsmatrix als ein Optimierungsproblem betrachtet werden, wobei der Fehler J bei der Summenwertbildung im Quadrat aller Punktkorrespondenzen minimiert werden soll:

$$J = \sum_{(i,j)} \|x_j - Wx_i\|^2 \quad (3.28)$$

Zu Beachten ist, dass dies ein nichtlineares Optimierungsproblem ist. Einige nichtlineare Optimierer können ebenfalls verwendet werden um diese Zielfunktion zu minimieren. In der Kapitel Evaluierung wurde bewiesen, dass der Koordinatenabstieg durch direkte objektive Funktionsbewertung schnell konvergiert. Jedes Mal, wenn ein Schritt gemacht wird bei dem die Zielfunktion J nicht abnimmt, wird die Schrittrichtung umgekehrt und verringert die Schrittweite des entsprechenden Parameters. Der Algorithmus endet, sobald die Schrittgröße für alle Parameter unter einen gewünschten Schwellenwert fällt (d.h. Wenn die Zielgenauigkeit erreicht ist).

Der detaillierte Algorithmus ist wie folgt dargestellt. Einige Definitionen werden hier eingeführt. P_0 ist der Vektor, der die Anfangswerte der Parameter speichert. Anschließend speichert er die Schrittgröße jedes Parameters in den Vektor d_p . temp_{dp} ist der Vektor, in dem die gewünschten Schwellenwerte der Parameter gespeichert werden. D ist die Anzahl Unbekannter Parameter und W die Transformationsmatrix.

1. Zuerst durch Anfangswert P_0 Berechnung des anfänglichen Fehlers J_0 .
2. Veränderung eines Parameters in eine entsprechende Schrittrichtung mit entsprechender Schrittgröße.
3. Berechnung der neuen Fehler J_{new} .
4. Vergleichen der beiden Fehler, ob der neue Fehler J_{new} kleiner ist. Wenn Ja wird dieser neue Wert beibehalten. Wenn Nein wird die Schrittrichtung umgekehrt und die Schrittgröße auf ein Drittel des vorherigen Werts reduziert. Dann ändern das Objekt auf den nächsten Parameter und kehren zum zweiten Schritt zurück.
5. Wenn alle Schrittgrößen in d_p den zuvor festgelegten Schwellenwert erreichen, wird der Algorithmus, mit Ausgabe des minimalen Fehlers J und der entsprechenden Transformationsmatrix, terminiert.

Das Flussdiagramm des Algorithmus ist wie in Abbildung 3.19 dargestellt.

Durch die Transformationsmatrix können die Koordinaten des zweiten Bildes in die Koordinaten des ersten Bildes übertragen werden. Abbildung 3.18 zeigt diesen Verlauf. Schließlich kann durch Subtraktion das Differenzbild erzeugt werden.

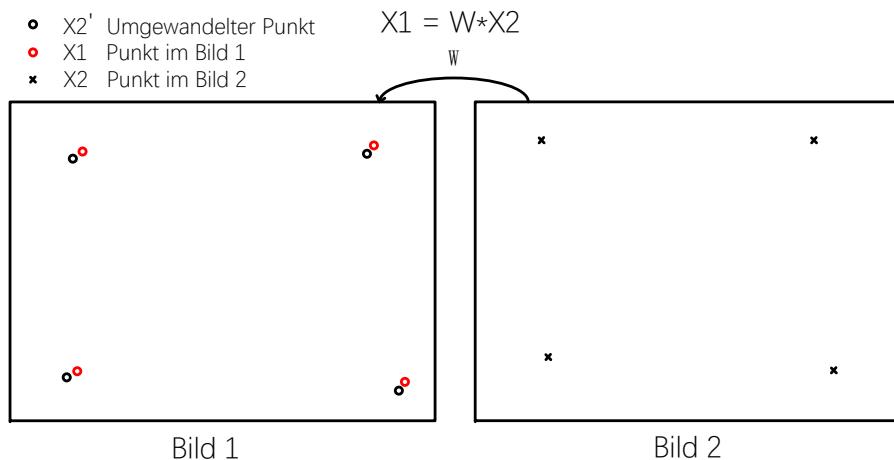


Abbildung 3.18: Transformation in das selbe Koordinatensystem.

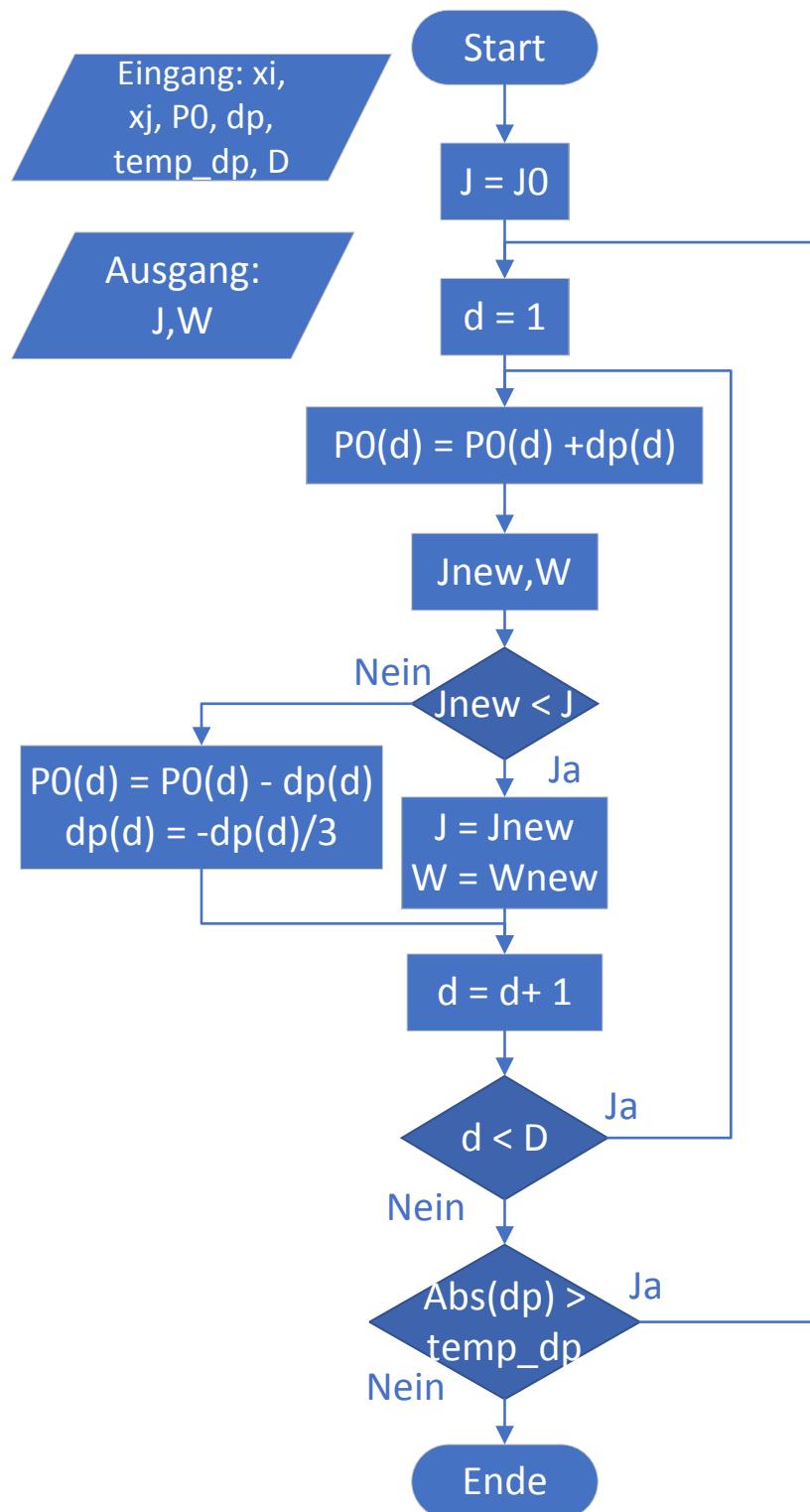


Abbildung 3.19: Flussdiagramm für Optimierung.

3.3 Differenzbild-Optimierung

Die Bildregistration liefert eine Reihe Bilder von der Kamera, deren Koordinaten in dasselbe Koordinatensystem umgewandelt wurden. Es werden je zwei Bilder ausgewählt und subtrahiert, um eine Reihe Differenzbilder zu erhalten. Das Ziel in diesem Abschnitt ist, von diesen Differenzbildern ein zu detektierendes Bild herzustellen, das die QR Muster Detektion vereinfacht. Es sollte hier beachtet werden, dass aufgrund der Zeitsynchronisation die QR Muster in den Differenzbildern einige unerwartete Effekte aufweisen könnten. Eine mögliche Formel der Differenzbilder wie folgt mit der Annahme dass es in vertikaler Richtung ist.

- total Schwarz-Weiß-Schwarz-Weiß-Schwarz Ordnung.
- halb Schwarz-Weiß-Schwarz-Weiß-Schwarz Ordnung, halb nicht gezeigt.
- total Weiß-Schwarz-Weiß-Schwarz-Weiß Ordnung.
- halb Weiß-Schwarz-Weiß-Schwarz-Weiß Ordnung, halb nicht gezeigt.
- halb Schwarz-Weiß-Schwarz-Weiß-Schwarz Ordnung, halb Weiß-Schwarz-Weiß-Schwarz-Weiß Ordnung.
- total nicht gezeigt.

Tabelle 3.2 zeigt solche Situationen und listet auf, ob sie zur folgenden Detektion angepasst sind. Die direkte Verwendung dieser Differenzbilder ist für die nächste Detektion ein kniffliges Problem. Um dieses Problem zu lösen, wurde ein Algorithmus zur Optimierung der Differenzbilder entwickelt.

Die Struktur eines QR Musters ist in Formel 3.29 aufgeführt. Die äußerste “1” Schicht ist hier ein Trennmuster und der Zentralbereich repräsentiert das QR Muster. Aufgrund der Modulationseigenschaften des DaVid Systems werden nur die Pixelwerte der Punkte mit dem Element “1” nach der Modulation verändert. Im Vergleich dazu werden die Pixelwerte der Punkte mit dem Element “0” nur gering verändert. Deswegen werden durch eine Absolutwertoperation die QR Muster als Schwarz-Weiß-Schwarz-Weiß-Schwarz Ordnung dargestellt. Dies ist die erwartete Modellstruktur die im nächsten Schritt operiert wird. Ein detailliertes Operieren wird im Abschnitt “QR Muster Detektion” besprochen.

Tabelle 3.2: mögliche Differenzbilder.

Ob angepasst zu folgender Detektion?	Differenzbild	Ob angepasst zu folgender Detektion?	Differenzbild
Ja		Nein	
Nein		Nein	
Nein		Nein	

$$QR_{base} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (3.29)$$

Als nächstes wird der Begriff “Energie” vorgestellt. Auf numerischer Ebene beschreibt dieser den Mittelwert des quadrierten Pixelwertes des Bildes. Es ist bekannt, dass die Pixelwerte in dem Gebiet rund um den Modulationsbereich sich durch Subtraktion der beiden Frames gegeneinander aufheben. In diesem Gebiet werden die Pixelwerte nur vom Rauschen beeinflusst und haben deshalb einen sehr kleinen Wert. Im Vergleich dazu betragen die Pixelwerte im Modulationsbereich einen relativ großen Wert $\pm 2A$, hier ist A die Modulationsamplitude des

Systems. Der Einfluss von Zeitsynchronisation gleicht die Pixelwerte in einigen Teilen des Modulationsbereichs aufgrund von Synchronisation aus, und zwar durch aufheben der „Energie“. Die Größe der „Energie“ hängt hauptsächlich von den Pixelwerten im Modulationsbereich ab. Je größer die Pixelwerte sind, desto höher die „Energie“ des Bildes, und desto klarer kann das QR Muster werden. Gemäß dieser Regel wird die „Energie“ jedes Differenzbildes berechnet und in absteigender Reihenfolge angeordnet. Um die folgende Detektion zu vereinfachen, werden die ersten paar Bilder hinzugefügt. Dann wird ein zu detektierende Bilder erhalten.

Die Formel der „Energie“ wird in Gleichung 3.30 dargestellt. Hier repräsentiert m, n die Anzahl der Zeilen und Spalten der Matrix. $diff(i, j)$ ist der Pixelwert des Punkts in der m -ten Zeile und n -ten Spalte der Matrix.

$$Energie = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n diff^2(i, j) \quad (3.30)$$

Es folgen die detaillierten Schritte dieses Algorithmus:

1. Wahl zwei beliebiger Bilder und Subtraktion dieser Bilder um das Differenzbild zu erhalten.
2. Berechnung der Energie durch eine Absolutoperation.
3. Sortierung in absteigender Reihenfolge und Auswahl einiger Bilder.
4. Addition der gewählten Differenzbilder um ein zu detektierendes Bild zu erhalten.

Dazu, das passende Flussdiagramm in Abbildung 3.20.

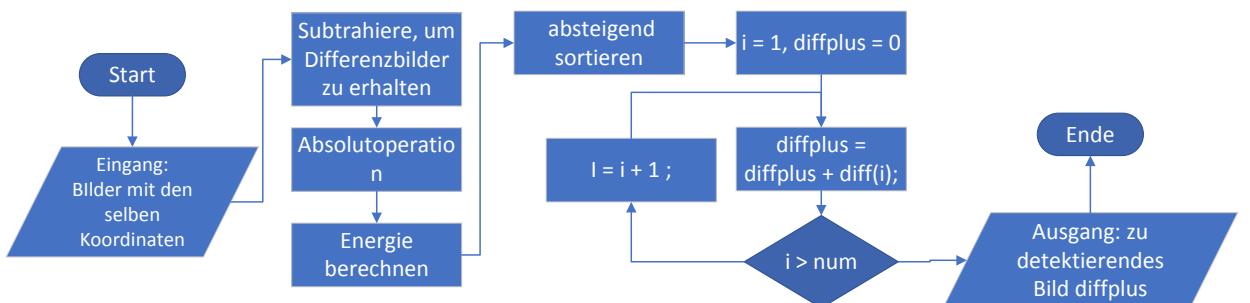


Abbildung 3.20: Differenzbild Flussdiagramm.

Ein Beispiel für ein zu detektierendes Bild wird in Abbildung 3.21 vorgeführt. Aus der Abbildung ist ersichtlich, dass der Modulationsbereich besser dargestellt ist und an den vier Ecken des

Bereichs ein eindeutiges QR Muster vorhanden ist. Als nächstes wird die Bildverarbeitung vorgestellt.

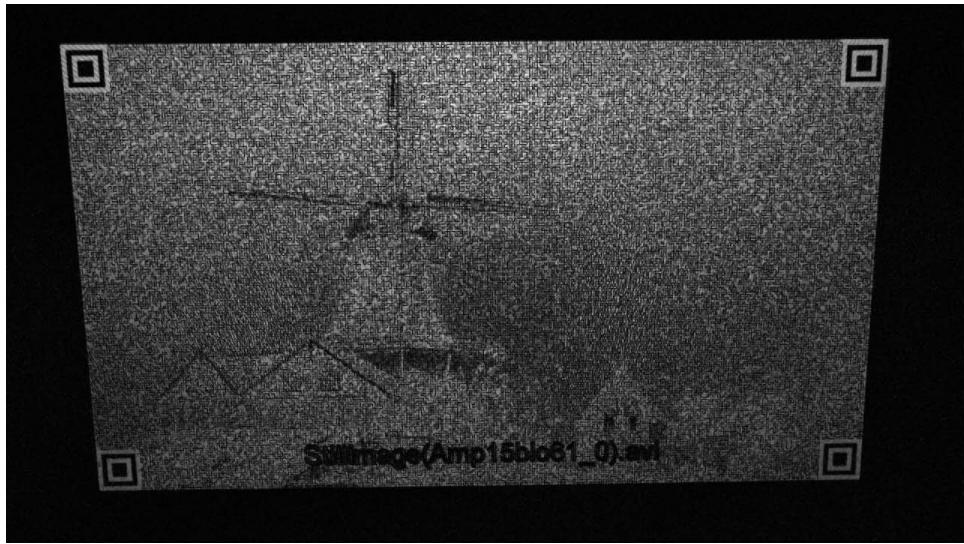


Abbildung 3.21: Ein zu detektierendes Bild.

3.4 Bildverarbeitung

Durch die Differenzbild-Optimierung, wird ein zu detektierendes Bild erhalten. Abbildung 3.21 ist noch ein Graustufenbild und muss noch einiger Bildverarbeitung unterzogen werden. Dadurch können kleine Punkte und Lücken, die durch Rauschen und Fehler verursacht wurden, entfernt werden um die nachfolgende Detektion zu erleichtern. Der detaillierte Inhalt der Bildverarbeitung wurde in der anderen Methode aufgeführt. Es folgt eine kurze Beschreibung.

Bild Binarisierung

Für die Schwellenwertbildung und das Erstellen eines Binärbildes wurde eine Funktion namens “*imbinarize*” in Matlab verwendet. Dieser Funktion wird das Bild übergeben, damit es einen anpassungsfähigen Schwellwert generiert um das Schwarz-Weiß-Bild zu erstellen. Dies ist ausreichend für die QR Muster Detektion, weil es nur schwarze und weiße Module enthält, die binär 1 bzw. 0 sind.

Morphologie

Mit öffnenden und schließenden Filtern können die Lücken zwischen Blöcken und die Punkte vom Rauschen stark reduziert werden, was das binär Bild zu einer guten Interpretation des QR Musters macht.

3.5 QR-Muster-Detektion

Nach der Bildverarbeitung wird nun die QR Muster Detektion ausgeführt. Das Ziel einer QR Muster Detektion ist, das Zentrum des Musters im Bild zu lokalisieren um dadurch das Bild zu rekonstruieren. Abbildung 3.22 zeigt die geometrische Struktur eines QR Musters. An jeder Ecke des Modularisationsbereichs gibt es ein solches QR Muster.

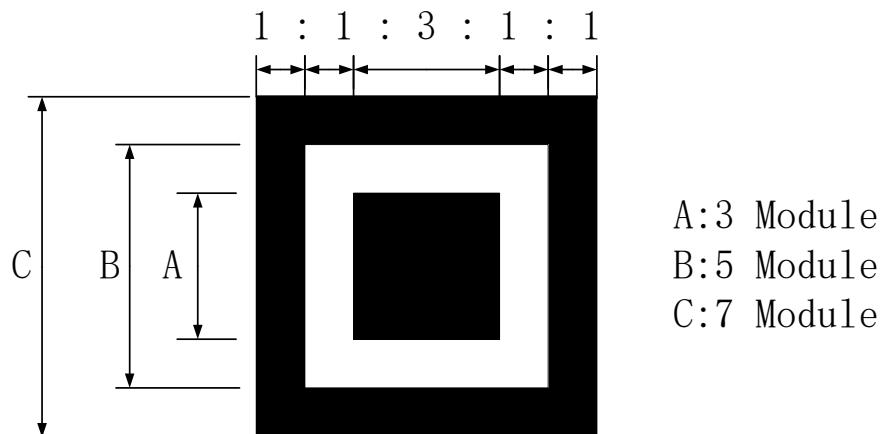


Abbildung 3.22: QR Muster.

Aus der geometrischen Sicht kann jedes Muster als drei konzentrische Quadrate betrachtet werden und besteht aus einem schwarzen (dunklen) 7×7 Modul, einem weißen (hellen) 5×5 Modul und schließlich einem schwarzen (dunklen) 3×3 Modul. Von Kenntnissen der Geometrie ist bekannt, dass in jeder Richtung das Breiteverhältnis der alternativen Schwarz- und Weißmodul in einem Muster die Beziehung $1 : 1 : 3 : 1 : 1$ existiert, wie es in Abbildung 3.23 vorgeführt ist. Diese wichtige Eigenschaft hilft uns, die Positionen der QR Muster zu finden. In der Praxis gibt es rund um die Muster noch ein Trennmuster, das ein Funktionsmuster aller weißen (hellen) Module (ein Modul breit) ist. Es dient als eine Grenze zwischen den Mustern und dem Datenbereich, um die Verwechslung zwischen Muster und Daten zu vermeiden.

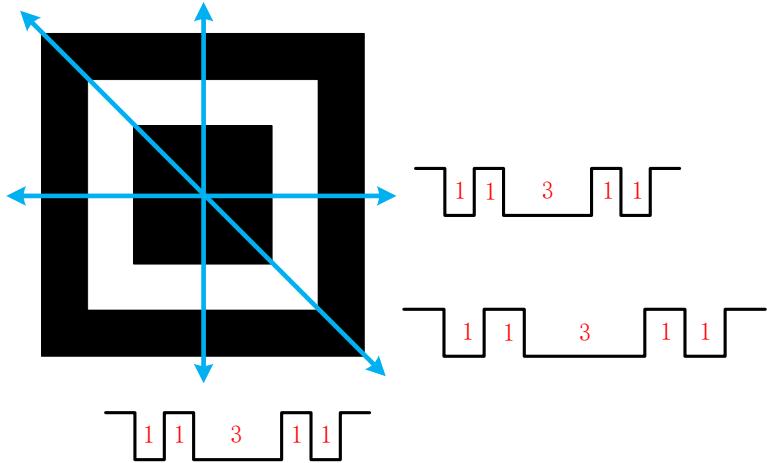


Abbildung 3.23: QR Muster Ratio.

Als nächstes werden die detaillierten Schritte der Detektion aufgeführt welche mit der Funktion "detectFIP" in Matlab implementiert wurde.

Schritt 1:

Zuerst wird der Berechnungsaufwand der Analyse des ganzen Bildes geschätzt, und anschließend wird der Modulationsbereich in kleine Bereiche unterteilt, die die QR Muster beinhalten.

Schritt 2:

Scannen jeder Zeile dieses kleinen Bereiches und Speicherung der Länge der Schwarz und Weiß Module in einen fünf Elementen Vektor. Die Länge der Module bedeutet die Anzahl aufeinanderfolgender Pixel in einer Zeile mit der gleichen Farbe. Die Speicherreihenfolge in diesem Vektor ist Schwarz-Weiß-Schwarz-Weiß-Schwarz. Es sollte hier angemerkt werden, dass das erste Element des Vektors die Anzahl der schwarzen Module enthält.

Schritt 3:

Immer wenn das fünfte Element des Vektors gezählt wird, werden die Werte ausgewertet, ob das Muster nahe genug an den $1 : 1 : 3 : 1 : 1$ Verhältnissen daran ist. Wenn die Bedingung erfüllt ist, gehen zum Schritt 4. Ansonsten wird der Vektor um zwei Elemente nach links verschoben und die ersten zwei Elemente des Vektors werden entfernt. Kehren zum Schritt 2 zurück.

Schritt 4:

Die Elemente vom Vektor werden verarbeitet, um das ungefähre horizontale Zentrum zu erhalten. Eine Kreuzprüfung wird an diesem Punkt vorgenommen, welche aus den Schritten 2 und 3 besteht, der Unterschied besteht darin, dass der horizontale Scan durch einen vertikalen Scan ersetzt wird. Anschließend wird überprüft, ob ein vertikales Zentrum gefunden wurde. Wenn Ja wird eine Kreuz-Kreuzprüfung mit dem horizontalen Scan vorgenommen, um die Ergebnisse zu optimieren. Dies wird hauptsächlich benötigt, um die reale horizontale Mitte des Musters in extremen Schräglage Fällen zu lokalisieren. Nach Speicherung des potenziellen Zentrums, werden die Elemente des Vektors geleert um wieder im zweiten Schritt einen Scan zu machen. Ansonsten wird der Vektor um zwei nach links verschoben und die ersten beiden Elemente des Vektors weggeworfen. In Schritt 2, wird wiederum erneut begonnen zu scannen und zu zählen.

Schritt 5:

Die Ausgabe des vorherigen Schritts wird verarbeitet. Falls mehrere potentielle Muster gefunden wurden, wird „*SelectBestPattern*“ verwendet, um das Beste Auszuwählen. Es sollte angemerkt werden, dass wenn die möglichen Muster nach dem Ende der Erkundung nicht gefunden wurden, ein spezielles Signal zurückgegeben wird und das System zur Schritt Differenzbild Optimierung zurückkehren wird. Die Operation besteht darin dem ursprünglichen Bild, welches aus einige Differenzbildern besteht, ein mehreres Differenzbild hinzuzufügen.

Schritt 6:

Durch die gefundenen Muster Zentren, wird eine projektive Transformation vorgenommen, um die Ecken des Bildes zu bestimmen.

Die folgende Abbildung 3.24 zeigt ein Flussdiagramm einer Detektion für QR Muster.

Ein Beispiel für QR Muster Detektion wird in Abbildung 3.25 vorgeführt:

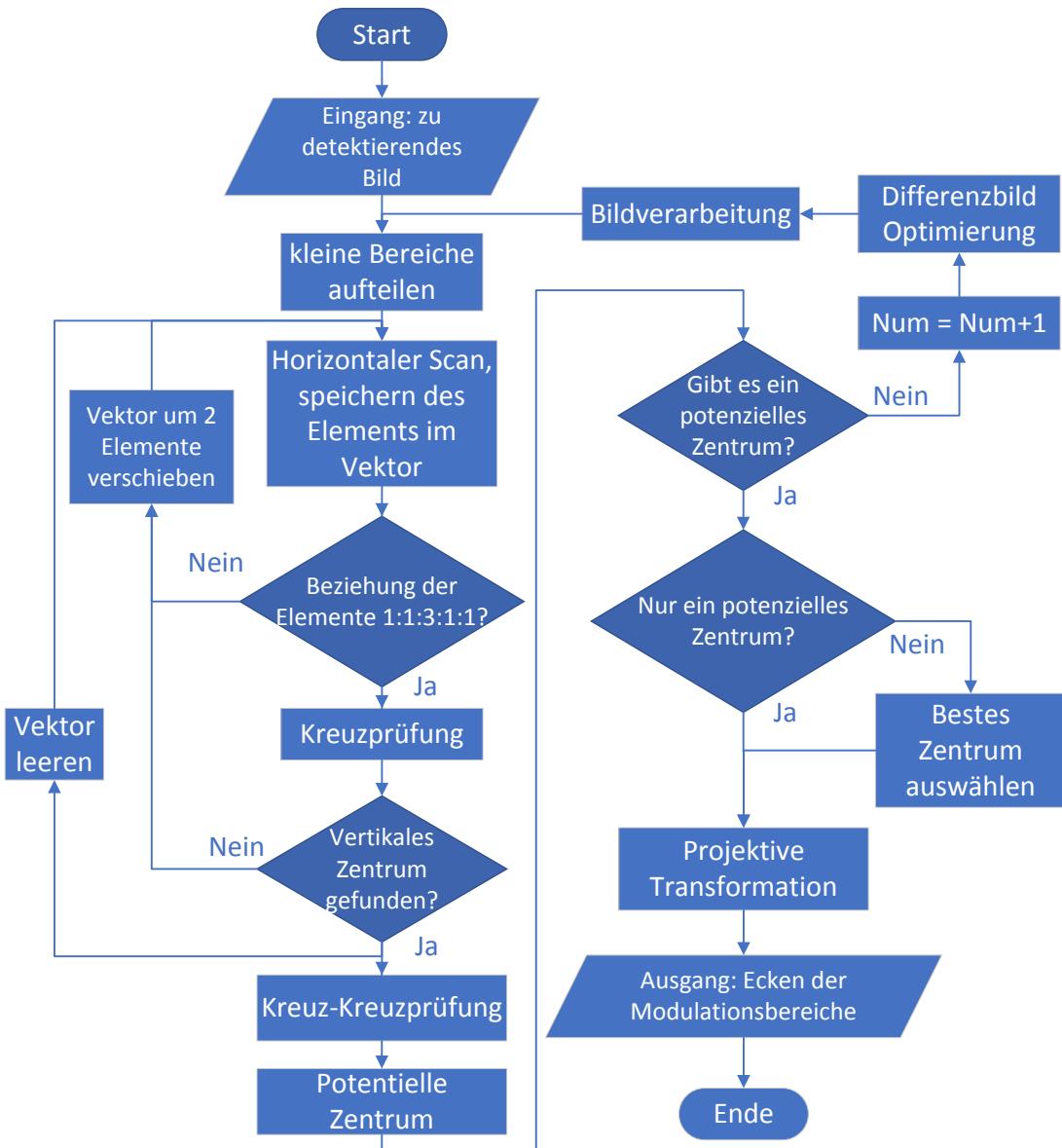


Abbildung 3.24: Flussdiagramm der QR Muster Detektion.

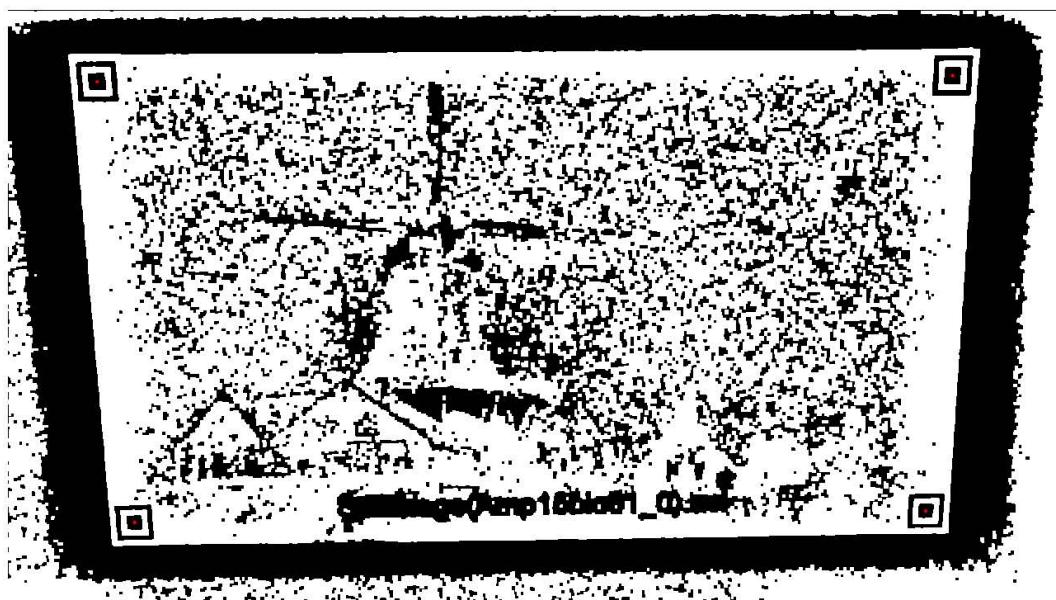


Abbildung 3.25: Beispiel einer QR Muster Detektion.

4 Untersuchung vom zweiten Verfahren mittels Formerkennung

In diesem Kapitel wird die Realisierung des zweiten Verfahrens eingegangen. Zuerst wird die allgemeine Struktur dieser Methode beschrieben. Dies liefert ein grobes Verständnis der Struktur und des Arbeitsprinzips der gesamten Methode. Anschließend läuft eine Erklärung der Bildverarbeitungen im Detail. Diese besteht aus folgenden Teilen: Binarisierung, Morphologie Operation und Canny Kanten Detektion. In diesem Abschnitt wird erläutert, wie diese Operationen funktionieren und was der Sinn dieser Behandlungen ist. Schließlich wird die Linienextrahierungsmethode (Hough- und Radon-Transformation) detailliert vorgestellt.

4.1 Allgemeine Struktur

Anders als bei der vorherigen Methode, indem durch die Charakteristiken des QR Musters die Modulationsbereiche detektiert wurden, wird hier eine neue Strategie mit Hilfe geometrischer Eigenschaften des Modulationsbereichs verwendet. Es ist bekannt, dass im DaVid System durch Aufnahme des Inhalts auf dem Bildschirm die versteckten Daten erkannt werden können. Dann entspricht der Modulationsbereich tatsächlich dem Bildschirmbereich, der im Allgemeinen über die Form eines Rechtecks verfügt. Deswegen kann hier das Problem eines Detektierens des Modulationsbereichs in ein Problem des Detektierens eines Rechtecks umgewandelt werden. Außerdem, lassen die geometrischen Eingenschaften des Rechtecks, welches aus 2 paaren paralleler Linien besteht, das Problem zu einem Linienextrahierungsproblem werden. Das heißt, diese Methode basiert auf dem extrahieren der Linien aus dem Bild, um den endgültigen Modulationsbereich zu erkennen und zu interpretieren. Abbildung 4.1 zeigt das Strukturdiagramm dieses Verfahrens.



Abbildung 4.1: Idee für Detektion.

Das Objekt, welches mit dieser Methode bearbeitet wird, ist eine Reihe von Bildern, die von einer Handkamera mit Stativ aufgenommen wurde. Durch das Stativ können diese Bilder in dasselbe Koordinatensystem eingespeist werden. Mit gegenseitiger Subtraktion wird eine Reihe Differenzbilder erhalten. Durch das Wissen des vorherigen Kapitels, wird die „*Energie*“ hauptsächlich aus dem Modulationsbereich gewonnen. Durch Addition der Differenzbilder mit der meisten „*Energie*“ kann ein zu detektierendes Bild erhalten werden um die folgende Operation durchzuführen. Diese Teilbereiche sind schon im vorherigen Kapitel beschrieben und werden hier nicht ins Detail erklärt. Um den Modulationsbereich vom Bild hervorzuheben, wird das zu detektierende Bild binarisiert, dadurch wird das Bild in zwei Teile geteilt, ein für das Objekt (hier der Modulationsbereich) und ein für den Hintergrund (hier der Bereich um den Modulationsbereich). Danach wird zur Beseitigung der zahlreichen Unvollkommenheiten (hier die kleinen Punkte und Lücken), die von Rauschen und Fehlern verursacht werden, wird hier eine Morphologie Operation benötigt, welche auf der Form und Struktur des Objekts basiert. Als Nächstes soll der Modulationsbereich erkannt werden. Um diese Operation zu implementieren, muss die rechteckige Grenze zwischen dem Modulationsbereich und dem Hintergrundbereich gefunden werden. Mit Hilfe einer Kantenextraktionsmethode, also dem Canny-Algorithmus, werden im Bild nur die Kanten übernommen die der Grenze des Modulationsbereichs entsprechen. Die Grenze des Modulationsbereichs hat eine rechteckige Form, bestehend aus vier Linien. Das Hough- und Radon-Transformationsverfahren wird verwendet, um alle zwei geraden Linien in der horizontalen Richtung bzw. der vertikalen Richtung zu erfassen. Schließlich wird das endgültige Rechteck, also der Modulationsbereich, durch die vier geraden Linien bestimmt. Das Flussdiagramm wird in Abbildung 4.2 gezeigt. Die Details jedes Teils werden in den folgenden Abschnitten beschrieben.

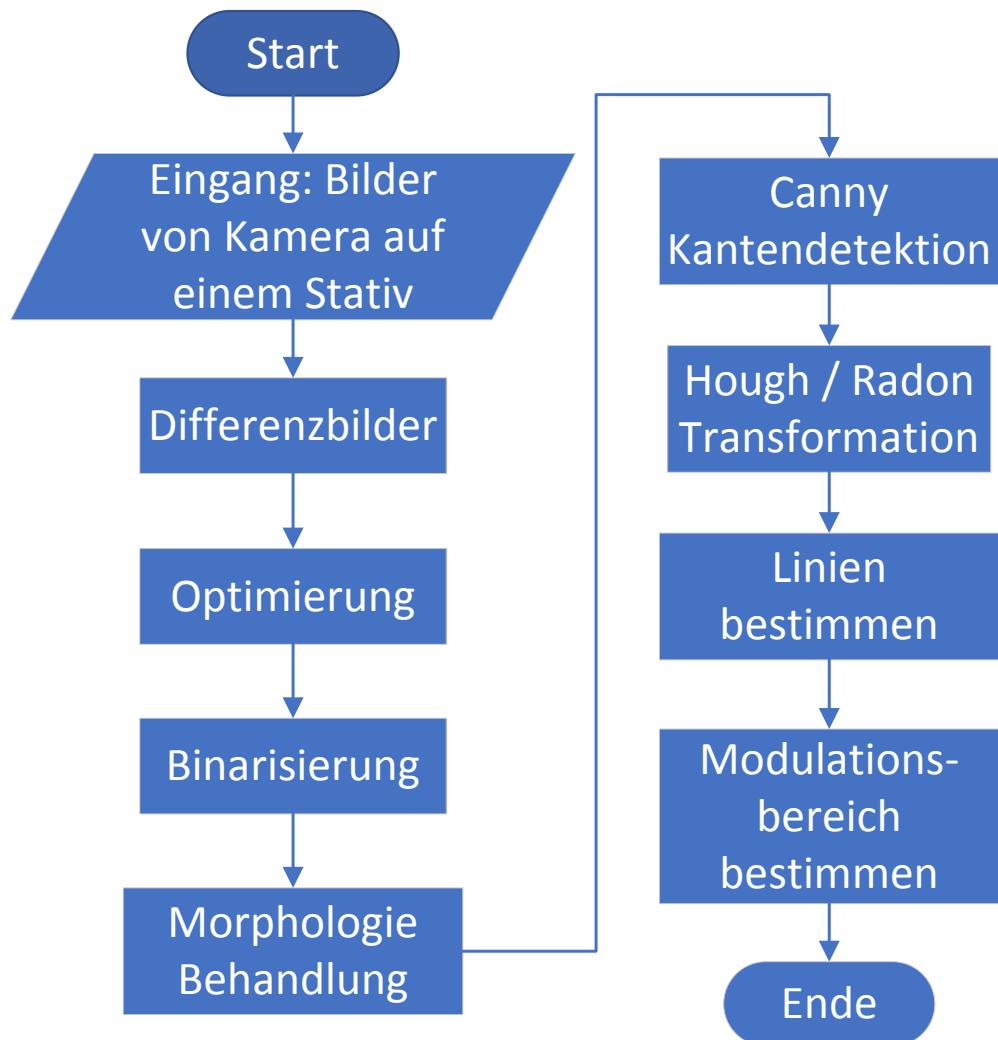


Abbildung 4.2: Flussdiagramm der zweiten Methode.

4.2 Bildverarbeitung

4.2.1 Binarisierung

In diesem Abschnitt wird der Binarisierungsvorgang der Bildverarbeitung beschrieben. Um den Modulationsbereich zu detektieren, muss dieser zuerst vom Bild hervorgehoben werden. Anschließend kann mit einer Binarisierungsoperation dieser Vorgang implementiert werden, indem das Bild in zwei Teile geteilt wird: ein Teil mit dem Pixelwert 1 für den Modulationsbereich und der andere Teil mit dem Pixelwert 0 für den umgebenden Hintergrund. Schließlich kann der Modulationsbereich mit einigen weiteren Operationen wie der Morphologie Operation und Kanten Detektion behandelt werden.

In der digitalen Bildverarbeitung nehmen binäre Bilder eine ausschlaggebende Position ein. Viele Anwendungen der digitalen Bildverarbeitung können als binäre Probleme betrachtet werden. Um binäre Bilder zu verarbeiten und zu analysieren, wird zunächst das Graustufenbild digitalisiert, um ein binäres Bild zu erhalten. Wenn das Bild weiterverarbeitet wird, bezieht sich die Sammeleigenschaft des Bildes nur auf die Position des Pixels, dessen Pixelwert 0 oder 1 ist, sodass der mehrstufige Wert des Pixels nicht länger vertreten ist um die Verarbeitung zu vereinfachen. Darüber hinaus ist die Menge an Datenverarbeitung und Komprimierung gering.

Um ein ideales Binärbild zu erhalten, werden im Allgemeinen geschlossene, zusammenhängende Grenzen verwendet, um Bereiche zu definieren die sich nicht überlappen. Hier ist ein einfaches Beispiel zur Veranschaulichung des Binarisierungsvorgangs gegeben. In einem bimodalen Histogramm eines Bildes, wie in Abbildung 4.3, gibt es zwei spitze Wellenberge, einen für das Objekt und einen für den Hintergrund. Die Binarisierungsschwelle T wird genau an die tiefste Stelle zwischen den beiden Wellenbergen gelegt.

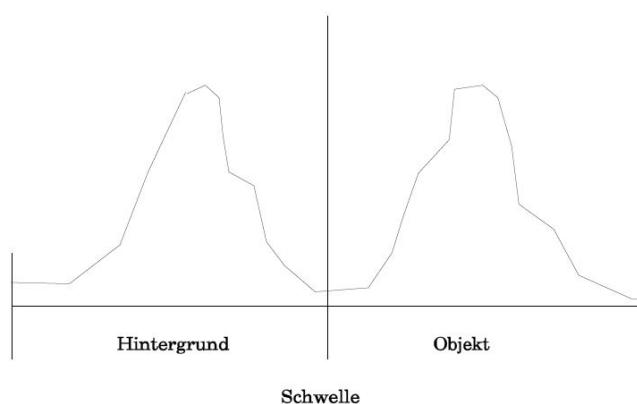


Abbildung 4.3: Histogramm eines einfachen Beispiels.

Alle Pixel, deren Graustufe größer oder gleich dem Schwellenwert T sind, werden zu einem bestimmten Objekt zugeordnet und deren Pixelwert auf 1 gesetzt. Andernfalls werden diese Pixel aus dem Objektbereich ausgeschlossen und deren Pixelwert auf 0 gesetzt, was den Hintergrund bzw. den Ausnahmeobjektbereich anzeigt. Die Formel dazu lautet:

$$y(x) = \begin{cases} 0, & \text{für } x < T \\ 1, & \text{sonst} \end{cases} \quad (4.1)$$

In der Praxis ist das Histogramm aufgrund der Helligkeit und dem Inhalt des Bildes nicht der obige bimodale Fall. Dies erfordert auch, dass die geeignete Binarisierungsmethode den tatsächlichen Bedürfnissen entsprechend im Binarisierungsprozess ausgewählt wird. Als folgende werden einige in dieser Arbeit versuchte Binarisierungsmethoden vorgestellt.

Grundlegende globale Schwelle Methode

Die grundlegende globale Schwellenmethode ist eine Erweiterung der festen Schwellenmethode, die vorher beschrieben wurde. Wenn die Histogrammspitzen und -täler des Bildes offensichtlich sind und Doppelpeaks aufweisen, ist diese Methode effektiver. Basierend auf der visuellen Überprüfung des Histogramms und der Erhaltung des Schwellenwerts durch eine iterative Methode. Die grundlegende Algorithmus ist wie folgt:

1. Wahl eines Parameters t und einen anfänglichen Schwellenwert T_0 , wobei der Durchschnitt der maximalen Grauwerte I_{max} und minimalen Grauwerte I_{min} verwendet wird.

$$T_0 = (I_{max} + I_{min})/2 \quad (4.2)$$

2. Segmentieren des Bildes mit dem Schwellenwert T_0 . Dadurch besteht das Bild aus zwei Teilen: G_1 besteht aus den Pixeln mit dem Grauwert größer als T_0 und G_2 aus denen, mit dem Grauwert kleiner oder gleich T_0 .

3. Berechnen des durchschnittlichen Grauwerts aller Pixel in u_1 und u_2 und den neuen Schwellenwert T_1 .

$$T_1 = (u_1 + u_2)/2 \quad (4.3)$$

4. Falls $|T_0 - T_1| < t$, dann ist T_1 der optimale Schwellenwert. Andernfalls wird T_1 zu T_0 zugewiesen und die Schritte 2–4 werden wiederholt, bis der optimale Schwellenwert gefunden ist.

Abbildung 4.6 zeigt ein Beispiel zur Binarisierung mit einer globalen Schwellenwertmethode.



Abbildung 4.4: Binarisierung mit globalen Schwellenwertmethode.

Grundlegende lokale Schwellenwertmethode

Ein Bildgebungs faktor ungleichmäßiger Helligkeit bewirkt, dass ein Histogramm, welches ansonsten für eine effiziente Segmentierung geeignet wäre, ein Histogramm wird, das nicht effektiv mit einem einzigen globalen Schwellenwert segmentiert werden kann.

Das Verfahren zur Verarbeitung besteht darin, das Bild weiter in Unterbilder zu unterteilen, um verschiedene Unterbilder mit unterschiedlichen Schwellenwerten zu segmentieren. Diese Methode wird als grundlegende adaptive Schwellenwert-Binarisierungsmethode bezeichnet. Das Hauptproblem bei diesem Ansatz besteht darin, das Bild zu unterteilen und den Schwellenwert für das resultierende Teilbild abzuschätzen. Da die Schwellen für jedes Pixel von dem Pixel in der Untergruppe abhängt ist die Position im Bild, also solche Schwellen, adaptiv.

Das Verfahren zum Unterteilen von Unterbildern wird auf das Bild angewendet. Hier werden drei Arten von Unterteilungsverfahren ausgewählt, Unterbilder einer jeweils Größe von 32×32 , 4×4 , 16×16 Pixeln werden unterteilt und die durchschnittliche Graustufe der Unterbilder wird als ein Schwellenwert für die Binarisierung ausgewählt. Das Ergebnis wird in Abbildung 4.5 gezeigt.

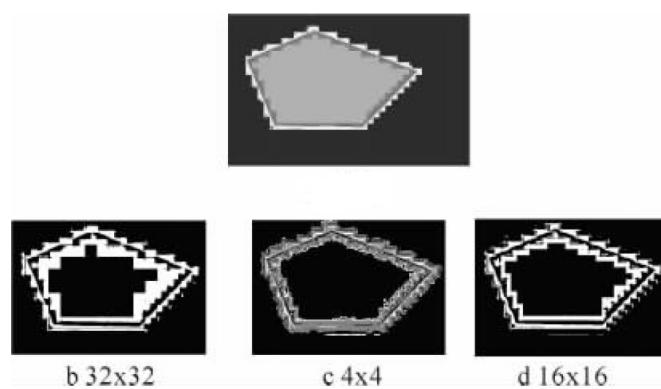


Abbildung 4.5: Binarisierung mit lokaler Schwellenwertmethode.

Otsu adaptive Schwellenmethode

Otsu[12], auch bekannt als die maximale Interklassenvarianz, wurde 1979 vom japanischen Gelehrten Otsu entworfen und ist eine adaptive Schwellenwertbestimmungsmethode, die auch als Otsu bekannt ist.

Die Grundidee des Otsu-Algorithmus ist, das Histogramm des Bildes gemäß der Varianz zwischen dem Vordergrund und dem Hintergrund zu verwenden um den optimalen Schwellenwert dynamisch zu bestimmen. Die Anzahl der Pixel in einem Bild ist N und die Graustufen $L(0, 1, \dots, L - 1)$. Die Anzahl der Pixel mit dem Grauwert i ist n_i , dadurch ist die Wahrscheinlichkeit von i : $p_i = \frac{n_i}{N}$. T stellt den Segmentierungsschwellenwert zwischen Vordergrund und Hintergrund dar für das Bild. Der Vordergrund entspricht den Grauwerten von 0 zu $T - 1$, dagegen der Hintergrund den Grauwerten von T zu $L - 1$.

Die Wahrscheinlichkeit des Vordergrundgebiets W_0 und des durchschnittlichen Grauwerts U_0 sind

$$w_0 = \sum_{i=0}^{T-1} p_i, \quad u_0 = \sum_{i=0}^{T-1} ip_i / w_0 \quad (4.4)$$

Die Wahrscheinlichkeit der Hintergrundpunkte W_1 und des durchschnittlichen Grauwerts U_1 sind

$$w_1 = \sum_{i=T}^{L-1} p_i = 1 - w_0, \quad u_1 = \sum_{i=T}^{L-1} ip_i / w_1 \quad (4.5)$$

Dann ist der gesamte durchschnittliche Grauwert des Bildes

$$u = w_0 u_0 + w_1 u_1 \quad (4.6)$$

Die Infra-Klassen-Varianz ist definiert als

$$\sigma^2 = w_0(u_0 - u)^2 + w_1(u_1 - u)^2 = w_0 w_1 (u_0 - u_1)^2 \quad (4.7)$$

Wenn sich der Schwellenwert T zwischen 0 und $L - 1$ befindet und σ^2 das Maximum ist, ist das entsprechende T der optimale Schwellenwert wie das Beispiel in Abbildung 4.6 zeigt.



Abbildung 4.6: Binarisierung mit Ostu.

Die Ostu-Methode verwendet Grauerthistogramme zur Bestimmung des Schwellenwerts. Es ist eine automatische nicht-parametrische Schwellenauswahlmethode. Diese Methode ist einfach durchzuführen, wird nicht von der Kontrast- und Helligkeitsänderung unter bestimmten Bedingungen beeinflusst und kann das Objekt zufriedenstellend vom Hintergrundbereich trennen.

Abbildung 4.7 zeigt ein Binärbild nach einer Binariesieungsoperation mit Ostu.

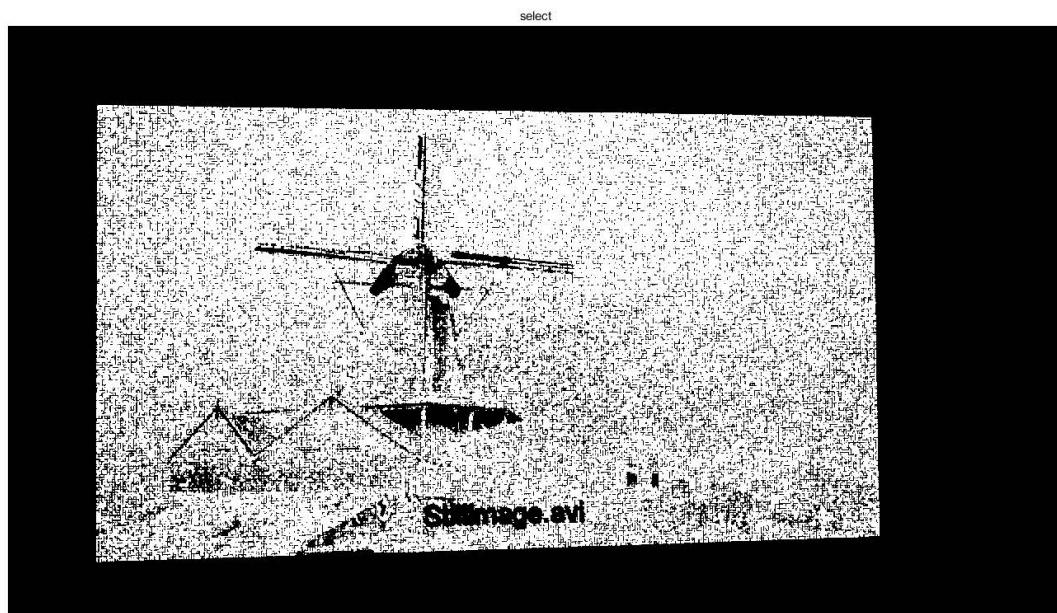


Abbildung 4.7: Binarisiertes zu detektierendes Bild.

4.2.2 Morphologie Operation

Nach der Binarisierung kann das Bild über zahlreiche Unvollkommenheiten verfügen. Insbesondere ist das Binärbild, welches durch einfache Schwellenwerte generiert wurde, durch Rauschen und Fehler verzerrt. Die morphologische Bildverarbeitung verfolgt das Ziel, diese Unvollkommenheiten zu beseitigen indem Form und Struktur des Objekts berücksichtigt werden.

Morphologische Operationen arbeiten auf der Grundlage der Mengenoperation und hängen von der relativen Ordnung des Pixelwerts ab. Durch diese Eigenschaft eignet sich diese Methode besonders für die Verarbeitung von Binärbildern. Die Mengenoperation ist gültig für alle Graustufenbilder. Hier in dieser Arbeit werden nur binarisierte Bilder benutzt. Die Mengenoperation erstellt ein neues Binärbild, bei dem das Pixel nur dann einen nicht-Null Wert erhält, wenn die Operation an dieser Position im Eingabebild erfolgreich war.

Die Eingabedaten für die mathematischen Morphologie Operationen sind zwei Bilder: das zu bearbeitende Bild A und ein Strukturelement B. B ist normalerweise eine kleine Pixelmatrix aus Nullen oder Einsen. Einige Eigenschaften des Strukturelements werden wie folgt gewählt.

- Die Matrixdimension ist gleichzeitig die Größe des Strukturelements.
- Das Muster, das aus Einsen und Nullen besteht, gibt die Form des Strukturelements an.
- Ein Ursprung des Strukturelements ist üblicherweise ein Pixel innerhalb des Strukturelements, welche aber auch außerhalb des Strukturelements liegen kann.

Eine übliche Anwendung besteht darin, dass der Ursprung ungerader Dimensionen der Pixelmatrixstellen als das Zentrum der Matrix definiert werden. Einige grundlegende Strukturelemente stehen wie folgt aus:

0	1	0
1	1	1
0	1	0

3x3 Kreuzform

1	1	1
1	1	1
1	1	1

3x3 Quadrat

0	0	1	0	0
0	1	1	1	0
1	1	1	1	1
0	1	1	1	0
0	0	1	0	0

5x5 Diamant

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

5x5 Quadrat

Abbildung 4.8: Einige grundlegende Strukturelemente.

Zuerst werden die zwei grundlegenden Morphologie Operationen, Erosion und Dilatation, beschrieben. Anschließend werden die darauf basierenden Operationen, die als Öffnung und Schließung bekannt sind, vorgestellt.

Dilatation

Die Dilatation Operation bewirkt, dass das Objekt wächst. Das Wachstum hängt von der Art und Form des Strukturelements ab. Die Formel einer Dilatation wird dargestellt als:

$$A \oplus B = \{z \mid (\widehat{B})_z \cap A \neq \emptyset\} \quad (4.8)$$

Hier ist $(\widehat{B})_z$ das Strukturelement B, welcher seinen Ursprung reflektiert und um z verschoben ist. Die entsprechende Dilatation für das Bild A mit B ist die Menge aller Verschiebungen z, wo \widehat{B} und A mindestens ein gemeinsames Element haben. Die Dilatation ist dazu da, um Pixel um die Grenze des Objekts hinzuzufügen. Außerdem wird die Dilatation Operation im wesentlichen verwendet, um die Löcher (fehlende Pixel) in einem soliden Objekt zu füllen. Diese beeinflusst die Intensität an einem Bereich und kann als ein Unschärfe Effekt bezeichnet werden, nämlich ein räumlicher Tiefpassfilter der beim linearen Filtern des Bildes verwendet wird. Abbildung 4.9 zeigt eine typische Dilatation Operation.

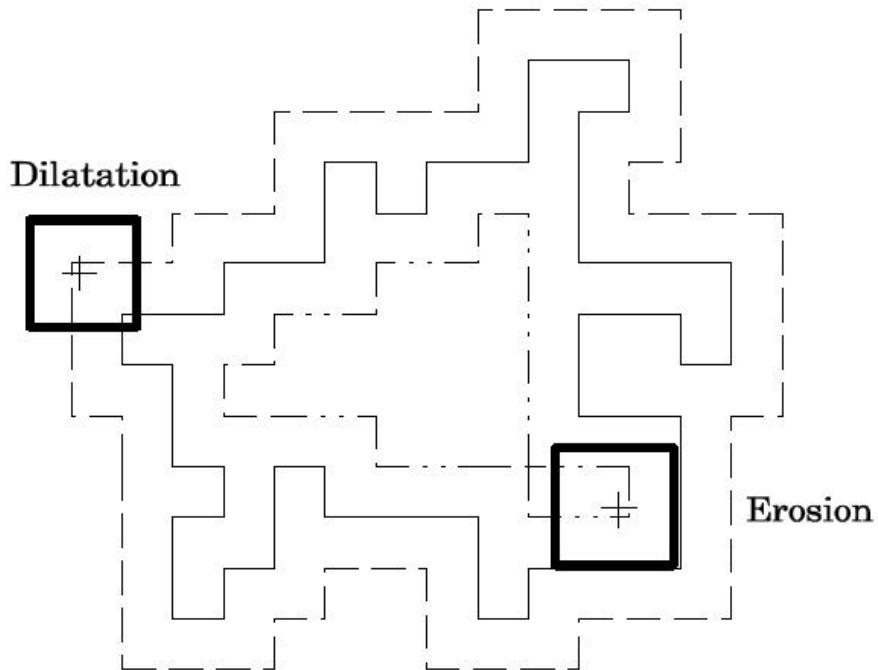


Abbildung 4.9: Dilatation und Erosion.

Erosion

Der Operationseffekt einer Erosion ist genau das Gegenteil von dem einer Dilatation. Die Erosion Operation bewirkt, dass das Objekt schrumpft. Die Erosion eines Bildes A ist durch das Strukturelement B definiert als

$$A \ominus B = \{z \mid (B)_z \subseteq A\} \quad (4.9)$$

Hier ist die Erosion des Bildes A die Menge aller Verschiebungen z, wo das Strukturelement B durch Verschiebungen zur Teilmenge des Bildes A gehört. Diese Operation führt zu einem Verlust von Grenzpixel des Objekts.

Die Erosion Operation entfernt Strukturen, die eine kleinere Größe als das strukturierende Element haben. Zudem kann es verwendet werden, um die verrauschte „Verbindung“ zwischen zwei Objekten zu entfernen. Da unerwünschte Pixel entfernen werden, entspricht der Effekt einem Schärfen des Objekts. Abbildung 4.9 zeigt eine typische Erosion Operation.

Öffnung

Die Öffnung Operation eines Bildes ist eine Kombination aus Erosion und Dilatation, d.h. auf eine Erosion Operation, folgt eine Dilatation Operation. In Anwendungsfall wird Bild A durch beide Operationen mit dem gleichen Strukturelement B ausgeführt. Die Formel der Öffnung Operation ist definiert als

$$A \circ B = (A \ominus B) \oplus B \quad (4.10)$$

Die Grenzen des geöffneten Objekts sind die Punkte, wo das Strukturelement B die äußeren Punkte der Grenze vom Objekt erreicht, während B entlang des Inneren der Kontur. Fein strukturierte Details, kleiner als das Strukturelement, werden demnach bei der Öffnung Operation entfernt und dünne Verbindungen zwischen größeren Teilen aufgelöst. Eine Öffnung Operation wird in Abbildung 4.10 gezeigt.

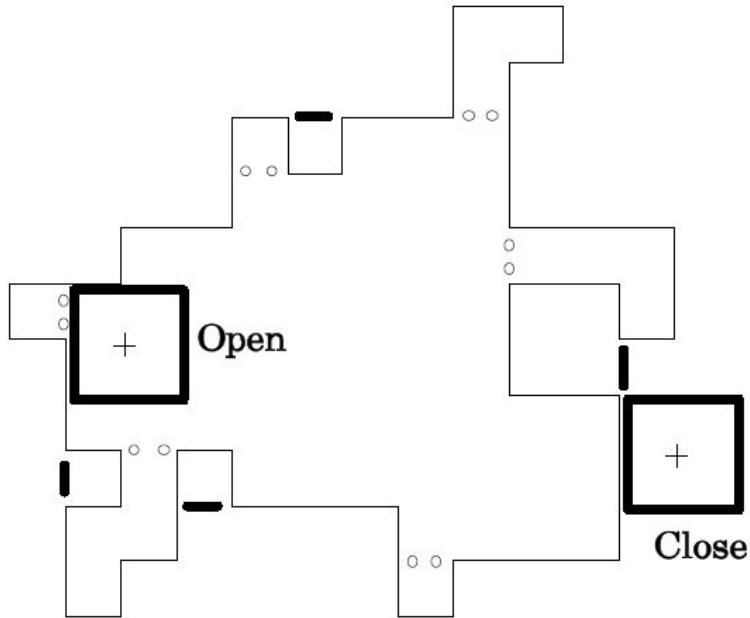


Abbildung 4.10: Öffnung und Schließung.

Schließung

Genau wie die Öffnung Operation ist die Schließung Operation auch eine Kombination aus Erosion und Dilatation. Der Unterschied liegt in der Reihenfolge der Operationen, d.h. zuerst wird eine Dilatation Operation, danach eine Erosion Operation mit dem gleichen Strukturelement ausgeführt. Die Schließung eines Bildes A durch das Strukturelement B ist definiert als

$$A \bullet B = (A \oplus B) \ominus B \quad (4.11)$$

Die Grenze des geschlossenen Objekts sind die Punkte, wo das Strukturelement B, die äußersten Punkte der Grenze von Objekt erreicht, während B entlang des Außen der Kontur. Kleinere Risse, Lücken und feine Details werden dagegen aufgefüllt und mit den großen Teilen zusammengeschlossen. Abbildung 4.10 zeigt den Vorgang einer Schließung Operation.

Öffnung und Schließung Operation besitzen folgende Eigenschaften:

- Öffnung und Schließung sind idempotent.
- Die Öffnung Operation ist anti-extensiv.
- Die Schließung Operation ist extensiv.

- Öffnung und Schließung sind dual bezüglich der Komplementierung.
- Ein Bild wird als B-Geöffnet bezeichnet, wenn es bei der gleichen Öffnung Operation unverändert bleibt.
- Ein Bild wird als B-Geschlossen bezeichnet, wenn es bei der gleichen Schließung Operation unverändert bleibt.

Das zu detektierende Bild aus Abbildung 4.7 nach einer morphologischen Operation wird in Abbildung 4.11 gezeigt.



Abbildung 4.11: Bild nach einer morphologischen Operation.

4.2.3 Kantendetektion

Nach der morphologischen Operation wird ein Binärbild erhalten, welches grob in zwei Teile geteilt werden kann. Ein Modulationsbereich mit dem Pixelwert 1 im mittleren Bereich und einen umgebenden Hintergrundbereich mit dem Pixelwert 0. Um die nächste Linie Detektion zu implementieren, wird eine Kanten Detektion, weil Kanten oft mit den Grenzen von Objekten in einer Szene verknüpft werden, durchgeführt. In anderen Worten, ist es die Grenze zwischen Modulationsbereich und Hintergrundbereich. [13]

Canny Algorithmus

Es gibt immer einige gute Kantenextraktionsmethoden, wie Sobel, Canny, Laplacian, Prewitt, Roberts uvm. In dieser Arbeit wird die leistungsfähigste Kantenerkennungsmethode bzw. der Canny Algorithmus benutzt. Die Canny-Methode unterscheidet sich von den anderen Kantenerkennungsmethoden darin, dass sie zwei verschiedene Schwellenwerte verwendet (um starke und schwache Kanten zu erkennen) und die schwachen Kanten in der Ausgabe nur dann einschließt, wenn sie mit starken Kanten verbunden sind. Diese Methode kann daher im Vergleich zur anderen Methode weniger Wahrscheinlichkeit verfügen, durch Rauschen getäuscht zu werden, und mehr Wahrscheinlichkeit verfügen echte schwache Kanten zu erkennen.

Das Ziel vom Canny Algorithmus ist es, einen optimalen Kantenextraktionsalgorithmus zu finden. Drei Kriterien für die optimale Kantendetektion werden vorgeschlagen:

- Gute Erkennung: Der Algorithmus kann so viele tatsächliche Kanten wie möglich im Bild erkennen.
- Gute Positionierung: Identifizieren der Kanten so nah wie möglich an den tatsächlichen Kanten im Bild.
- Minimale Antwort: Kanten in einem Bild können nur einmal identifiziert werden, und mögliches Bildrauschen sollte nicht als Kanten erkannt werden.

Die Implementierung des Canny Algorithmus funktioniert so:

1. **Gaußsche Unschärfe.** Das Hauptziel ist, Rauschen zu entfernen. Da Rauschen Hochfrequenzsignale ähnelt, wird es leicht als eine falsche Kante erkannt. Die Anwendung Gaußscher Unschärfe ist dazu da, um Rauschen zu entfernen und die Erkennung falscher Kanten zu reduzieren. Es sollte beachtet werden, dass ein zu großer Radius einige schwache Kanten nicht erkennen.
2. **Berechnen der Größe und Richtung des Gradienten.** Die Kanten des Bildes können in verschiedene Richtungen zeigen. Daher verwendet der klassische Canny Algorithmus vier Gradientenoperatoren, um die Gradienten in horizontaler, vertikaler und diagonaler Richtung zu berechnen. Jedoch werden im allgemeinen diese vier Gradientenoperatoren nicht verwendet, sondern mit Kantenunterschiedsoperationen (wie Rober, Prewitt, Sobel) die Differenz G_x und G_y in horizontaler und vertikaler Richtung berechnet. Die Berechnung der Gradientengröße und Richtung funktioniert wie folgt:

$$G = \sqrt{G_x^2 + G_y^2} \quad (4.12)$$

$$\theta = \arctan2(G_y, G_x)$$

Der Gradientenwinkel θ liegt im Bereich der Radianen $-\pi$ bis π , anschließend approximiert dieser in vier Richtungen, die horizontale, vertikale und zwei diagonale Richtungen ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) repräsentieren. Dieser kann durch $\pm i\pi/8$ ($i = 1, 3, 5, 7$) geteilt werden, sodass der in jeweils einem Bereich fallende Gradientenwinkel einen spezifischen Wert ergibt, der eine von vier Richtungen repräsentiert. Es folgt ein Beispiel einer Sobel Operator.

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * A \quad \text{and} \quad G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A \quad (4.13)$$

3. Nicht-maximale Unterdrückung. Nicht-maximale Unterdrückung ist eine Kantenverfeinerungsmethode. Die Gradientenkanten, die normalerweise abgeleitet werden, sind nicht einen, sondern mehrere Pixel breit, während Kriterium 3 erfordert, dass die Kante nur eine genaue Punktbreite hat. Nicht-maximale Unterdrückung kann dazu beitragen, den lokalen maximalen Gradienten beizubehalten, während alle anderen Gradientenwerte unterdrückt werden. Dies bedeutet, dass nur die schärfste Position in der Gradientenänderung beibehalten wird. Der Algorithmus funktioniert wie folgt:

- Vergleichen der Gradientenstärke des aktuellen Punkts mit der Gradientenstärke der positiven und negativen Gradientenrichtungspunkte.
- Wenn die Gradientenstärke des aktuellen Punkts im Vergleich zur Gradientenstärke anderer Punkte in der gleichen Richtung am größten ist, wird der Wert auf 1 beibehalten. Ansonsten wird dieser auf 0 gesetzt. Zum Beispiel, wenn die Gradientenrichtung des aktuellen Punktes in die Richtung von 90° direkt darüber zeigt, muss es mit den Pixeln in der vertikalen Richtung direkt darüber und darunter verglichen werden.

Es ist erwähnenswert, dass die positiven und negativen Richtungen keine unterschiedlichen Bedeutungen haben. Zum Beispiel ist die südöstliche Richtung sowie die nordwestliche Richtung als eine Richtung der Diagonalen zu betrachten. Vorher approximieren die Gradientenrichtung horizontal, vertikal und zwei Diagonalen in vier Richtungen, so dass jedes Pixel in einer dieser vier Richtungen entsprechend mit seiner eigenen Gradientenrichtung verglichen wird, um zu bestimmen ob es beibehalten wird.

4. Doppelte Schwelle. Ein allgemeiner Kantenerkennungsalgorithmus welcher einen Schwellenwert verwendet, um kleine Gradientenwerte zu entfernen die durch Rauschen oder Farbänderungen verursacht werden, während große Gradientenwerte beibehalten werden. Der Canny Algorithmus wendet einen doppelten Schwellenwert, d.h. einen hohen

und einen niedrigen Schwellenwert an, um Kantenpixel zu unterscheiden. Wenn der Kantenpixelgradientwert größer als der hohe Schwellenwert ist, wird er als ein starker Kantenpunkt betrachtet. Wenn der Kantengradientenwert kleiner als der hohe Schwellenwert und größer als der niedrige Schwellenwert ist, wird er als ein schwacher Kantenpunkt markiert. Punkte, die unterhalb der niedrigen Schwelle liegen werden unterdrückt.

5. **Hysteres Grenzenverfolgung.** Bisher könnten die starken Kantenpunkte als echte Kanten angesehen werden. Dagegen könnten schwache Kantenpunkte ebenfalls echte Kanten sein, jedoch könnten diese ebenfalls eine Ursache von Rauschen oder Farbänderungen sein. Um genaue Ergebnisse zu erhalten, sollten Kantenpunkte zweiter Art entfernt werden. Im Allgemeinen werden schwache Kantenpunkte und starke Kantenpunkte die durch reale Kanten verursacht werden, als verbunden betrachtet, während schwache durch Rauschen verursachte Kantenpunkte, welche keine Verbindung besitzen gefiltert werden. Der sogenannte Hysteres Grenzenverfolgung Algorithmus untersucht die angrenzenden Pixel eines schwachen Kantenpunktes, solange ein starker Randpunkt vorhanden ist, wird dieser schwache Kantenpunkt als echter Kantenpunkt betrachtet und dessen Wert bleibt 1 .

Zur Implementierung dieses Schritts werden alle verbundenen schwachen Kanten untersucht. Wenn ein Punkt einer verbundenen schwachen Kante mit einem starken Kantenpunkt verbunden ist, wird die schwache Kante beibehalten, ansonsten wird diese schwache Kante unterdrückt. Bei der Suche kann der Algorithmus "*Breitezuerst*" oder "*Tiefezuerst*" verwendet werden. Nach der gesamten Suche des Bildes, werden die Nicht-Kantenpunkte entfernt, d.h. der Wert dieser auf 0 gesetzt.

Abbildung 4.12 zeigt ein Ergebnis mit der Canny Kantenextraktion.

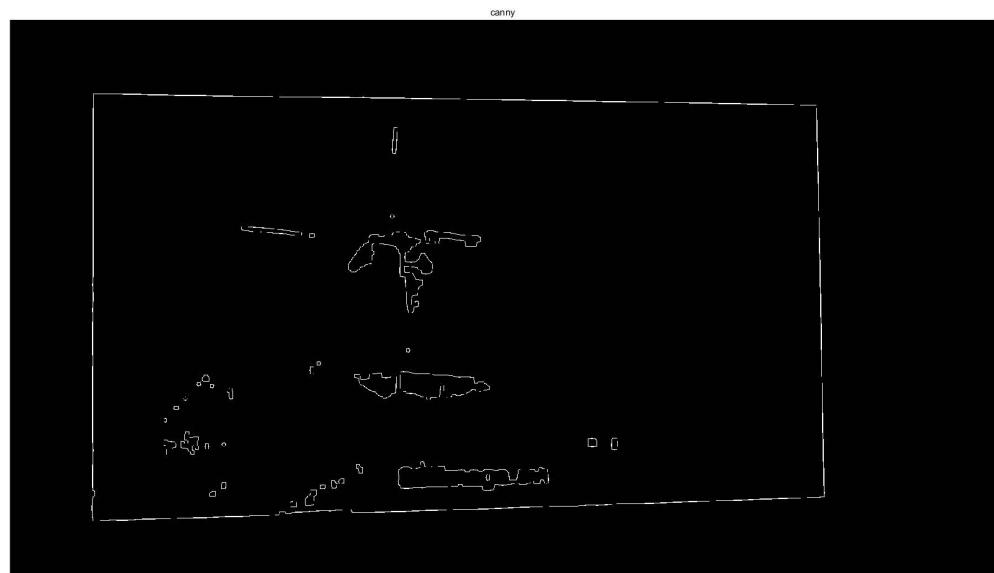


Abbildung 4.12: Binärbild nach der Canny detektion.

4.2.4 Cross Dilatation

Diese Operation erkennt Linien die Aufgrund von Kamera-Verzerrung nicht gerade aufgenommen wurden, wie Abbildung 1 zeigt. Dies kann die Erkennung des Rechtecks in praktischen Anwendungen beeinträchtigen. Um dieses Problem zu lösen und die Leistung der Methode zu erhöhen, wird die Cross Dilatation verwendet, welche jedes Pixel (x, y) zu einem Kreuzmuster erweitert. Alle Pixel, einschließlich der Ausgedehnten Pixeln, werden in der nächsten Operation behandelt.

Es ist angenommen, dass Einflussbereich jedes Pixels mit einer Region dargestellt wird, die um das Pixel herum zentriert ist. Wenn jedes Pixel auf eine Linie die Region erreicht, die durch das nächste Pixel der Linie beeinflusst wird, kann die gesamte Linie extrahiert werden. Um den Einflussbereich zu bestimmen, wird jedes Pixel (x, y) zu einem Kreuzmuster erweitert, indem benachbarte Pixel in den vier Richtungen des Pixels den Wert 1 bekommen, wie in Abbildung 1 zu sehen ist.

4.3 Linienerkennung

4.3.1 Hough Transformation

Hier werden die Grenzen des modulierten Rechtecks als Linien betrachtet. Um das Problem für eine Rechteck Detektion lösbar zu machen, müssen zunächst die Linien extrahiert werden. Um dies zu tun, müssen die Kanten die auf einer Linie im Bild liegen, erkannt werden. Die Hough Transformation ist eine beliebte Methode zum Extrahieren von Linien aus einem Bild. Diese ist in der Lage die benötigten Informationen zu liefern, indem die Spitzen an den Punkten der geraden Linien denen des binären Bildes entsprechen. Der Aufwand einer Hough Transformation hängt von der Größe des Bildes und der Anzahl der analysierten Winkel ab. Da die Winkelauflösung für die weitere Verarbeitung wichtig ist, ist eine maximale Kameraneigung vorgeschrieben, welcher in dieser Arbeit $\pm 10^\circ$ beträgt.

Um die Funktionsweise des HT Algorithmus zu beschreiben, müssen einige Definition eingefügt werden. In der Hough Transformation, kann jede Linie in der xy Ebene parametrisch beschreiben werden als:

$$\rho = x \cos \theta + y \sin \theta \quad (4.14)$$

Hier bedeutet ρ die Entfernung vom Ursprung der Koordinate zur Linie, θ der Winkel zwischen ρ und der positiven Richtung der x-Achse und (x, y) sind die Punkte auf der geraden Linie, wie Abbildung 4.13 zeigt.

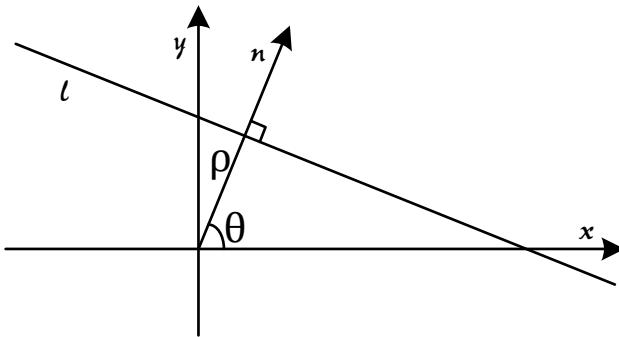


Abbildung 4.13: ρ, θ Parameterraum.

Die Gleichung 4.14 gilt für jedes Pixel auf dem Bild, d.h. für jeden Punkt auf dem Bild kann eine entsprechende trigonometrische Kurve nach der Hough-Transformation im Parameter-

raum (θ, ρ) gefunden werden. Es werden zufällig zwei Punkte auf dem Bild ausgesucht, die zwei trigonometrischen Kurven entsprechen, welche unvermeidlich einen Schnittpunkt (θ_0, ρ_0) erzeugen werden. Dieser Schnittpunkt wird dann in die Gleichung der Linie eingefügt, um eine gerade Linie zu bestimmen (und zwar die durch die zwei Punkte auf dem Bild festgelegte Linie).

Das heißt, die entsprechenden trigonometrischen Kurven, die von Punkten auf derselben Linie auf dem Bild erzeugt werden, schneiden sich an einem Punkt (θ_0, ρ_0) im Parameterraum. Je mehr Punkte auf der Linie sind, desto mehr Kurven schneiden sich an diesem Punkt. Wie der Algorithmus einer Hough Transformation folgendermaßen darstellt:

1. Erstellung eines Parameterraums mit einer geeigneten Quantisierungsstufe für die Entfernung ρ und den Winkel θ .
2. Erstellung eines Akkumulator Arrays $A(\rho, \theta)$ für alle (ρ, θ) .

$$A(\rho, \theta) = 0 \quad (4.15)$$

3. Für jeden Nicht-Hintergrundpunkt (x, y) wird im Bild, ρ mit jedem θ berechnet, ob diese Gleichung erfüllt wird: $\rho = x \cos \theta + y \sin \theta$. Wird diese Erfüllt, wird das Akkumulator-Array um 1 erhöht:

$$A(\rho, \theta) = A(\rho, \theta) + 1 \quad (4.16)$$

4. Suche des Spitzenwerts im Arrays A, welche die Linie im Parameterraum angibt.

Eine Beispiel Implementierung der Hough Transformation ist in Abbildung 4.14 zu sehen.

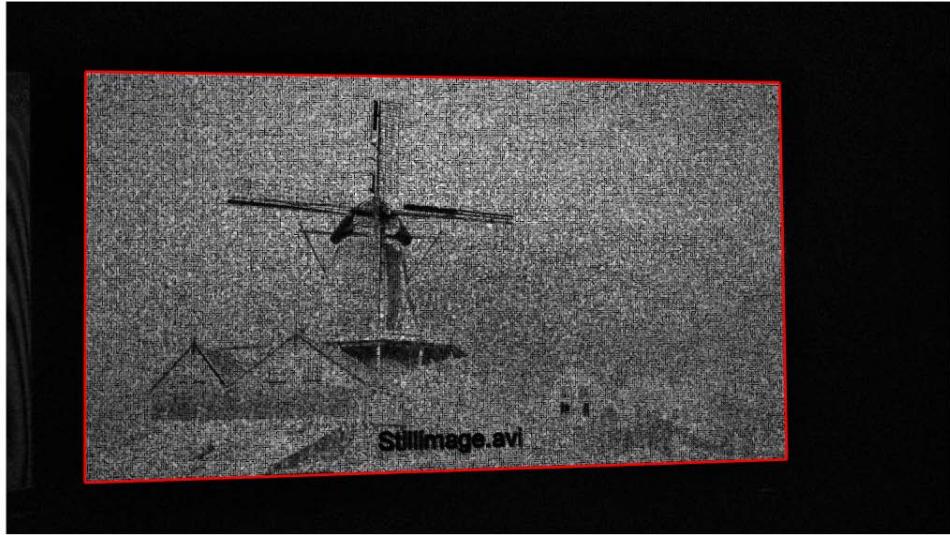


Abbildung 4.14: Houghdetektion.

4.3.2 Radon Transformation

In diesem Abschnitt wird Radon Transformation vorgestellt. Wie im vorherigen Abschnitt beschrieben, transformiert die Hough Transformation eine gegebene Kurve im Bildraum als ein Punkt im Parameterraum entsprechend dem Parameterausdruck der Kurve, wodurch mit dem Suchen nach Peaks in dem Parameterraum diese Kurven in dem Bildraum bestimmt werden können. Radon Transformation projiziert den Bildraum in den $\rho\theta$ -Raum in Form eines Linienintegrals (entspricht dem Parameterraum der Linie). Die Hough Transformation kann als eine diskrete Form der Radon Transformation betrachtet werden. Aufgrund der soliden mathematischen Basis ist die Radon Transformation im Gegensatz zu HT raffinierter und genauer.

Die Idee der Radon Transformation besteht darin, die ursprüngliche Funktion räumlich zu transformieren, das heißt, die ursprünglichen Punkte aus der XY-Ebene werden auf die AB-Ebene abgebildet. Dadurch sind alle Punkte einer geraden Linie in der XY-Ebene an demselben Punkt in der AB-Ebene. Durch Zeichnen der akkumulierten Dicke der Punkte auf der AB-Ebene, kann die Existenz der Linien auf der XY-Ebene erkannt werden.

Wie bei der Hough Transformation werden die Parameter ρ und θ zuerst verwendet, um die geraden Linien auf der XY-Ebene darzustellen. Diese Parametrisierte Linie wird in Abbildung 4.15 gezeigt und die Formel dazu lautet:

$$L(\rho, \theta) = \{(x, y) : x \cos \theta + y \sin \theta = \rho\} \quad (4.17)$$

Angenommen, es gibt eine Funktion $f(x, y)$, wie in Abbildung 4.16 gezeigt, dann ist das Integral der Funktion über die gerade Linie L :

$$\int_L f(x, y) ds \quad (4.18)$$

Wobei ds das Differential der Linie ist.

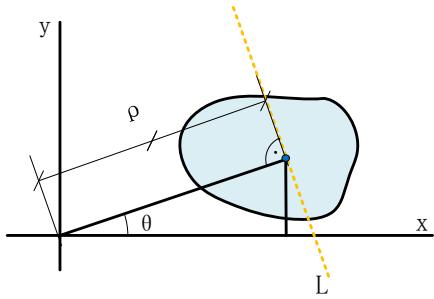


Abbildung 4.15: Parametrisierte Linie.

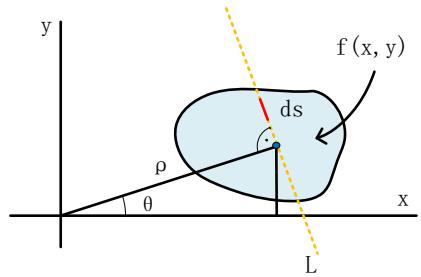


Abbildung 4.16: Integral der Funktion $f(x, y)$.

Die obigen Integrale für x, y sind leicht zu lösen. Eine der Lösungstechniken ist die Verwendung der Delta-Funktion. Das obige Integral kann geschrieben werden als:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - \rho) dx dy \quad (4.19)$$

Deshalb kann durch die Bestimmung des Satzes von (ρ, θ) ein Integral Wert entlang $L(\rho, \theta)$ erhalten werden. Somit wird die Radon Transformation zum Linienintegral der Funktion $f(x, y)$ umgewandelt, wie in Abbildung 4.17 gezeigt. Die Höhe g repräsentiert das Linienintegral von $f(x, y)$ auf der Linie L .

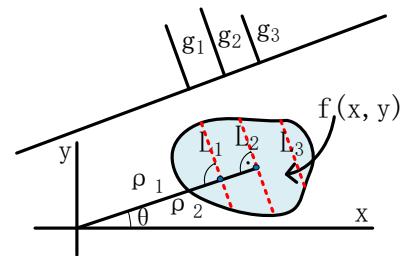
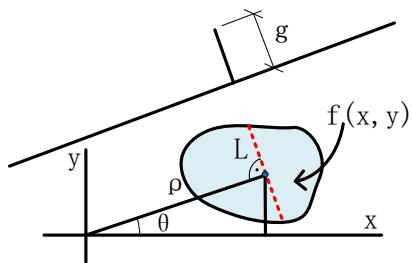


Abbildung 4.17: Radon Transformation (θ, ρ) .

Abbildung 4.18: Radon Transformation (θ) .

Wenn viele Linien parallel zu L sind, haben sie das gleiche θ , die radialen Koordinaten ρ sind dagegen verschieden. Für jede solcher parallelen Linien eins Linienintegrals von $f(x, y)$ werden viele Projektionslinien erzeugt, wie in Abbildung 4.18 zu sehen ist. Das heißt, die Radon Transformation eines Bildes in einem bestimmten Winkel erzeugt N Linienintegralwerte (Radon Transformation) und jeder Zeilenintegralwert entspricht einer Radialkoordinate. Die Radon Transformationswerte bei verschiedenen Winkeln werden kombiniert, um ein Radon Diagramm zu bilden.

5 Implementierung

In diesem Kapitel soll das zuvor vorgestellte Verfahren implementiert werden. Zuerst wird die experimentelle Umgebung auf dem Matlab vorgestellt. Anschließend die Implementierung der beiden vorherigen Methoden, welche die verarbeiteten Daten jeden Schrittes und dessen Demonstration der Endergebnisse enthalten. Schließlich läuft die Implementierung des Detektionsverfahrens auf einer Smartphone-Graphics Processing Unit (GPU).

5.1 Experimentelle Umgebung

Hier wird die Experimentelle Umgebung auf dem Matlab erläutert. Das Google Pixel wird hier verwendet, um die Bilder aufzunehmen. Dessen Parameter sind in der Tabelle 3.1 verfügbar. Die Anwendungssoftware “*VLCReceiver*” wird mit einem automatischen Einstellungsmodus benutzt. Entsprechend dem möglichen Einsatzbereich des David-Systems wurden im Experiment Bildschirme verschiedener Größen und Umgebungen ausgewählt. Die in dem Experiment verwendeten Bildschirme sind in der Tabelle 5.1 zu sehen. Die aufgenommenen Bilder werden auf der PC-Seite mit Matlab 2017b unter der Lizenz der TU Dortmund verarbeitet.

Tabelle 5.1: Verwendete Bildschirme.

Hersteller	Model	Beleuchtung	Auflösung	Größe	Frequenz
Lenovo Notebook	T440P	LED	1920 × 1280	14"	60
Samsung	SMB2440MH	LED	1920 × 1280	24"	60
Ausu	VG248	LCD	1920 × 1280	24"	60
LG	55EF950V	OLED	2160 × 3840	55"	120

5.2 Implementierung vom ersten Verfahren

Im Implementierungsprozess der Methode 1, stammt das Bild aus einer Reihe von acht Bildern, die mit einer Smartphonekamera aus die Hand aufgenommen wurden. Entsprechend den Eigenschaften der Bildregistrierung in dieser Arbeit sind die von der handgehaltenen Kamera aufgenommenen Objekte alles statische Videos. Dies bedeutet, dass sich der Inhalt des

Videos nicht ändert. Im tatsächlichen Betrieb entspricht es z.B. einer statischen elektronischen Werbetafel. Die Modulationsamplitude wird auf 6 gesetzt, der Daten Datenblock mit $4 \times 4\text{Pixel}$ und die QR Muster Größe als $72 \times 72\text{Pixel}$ initialisiert.

Aufgrund von Handschütteln während der Aufnahme entsteht eine leichte Verschiebung zwischen den Bildern. Um diesen Auswirkung zu verdeutlichen, wird das erste Bild und das achte Bild in Abbildung 6.1 verglichen. Abbildung 5.2 zeigt das Bild nach der Bildregistrierung. Wie zu sehen ist, ist der Abstand komprimiert worden.

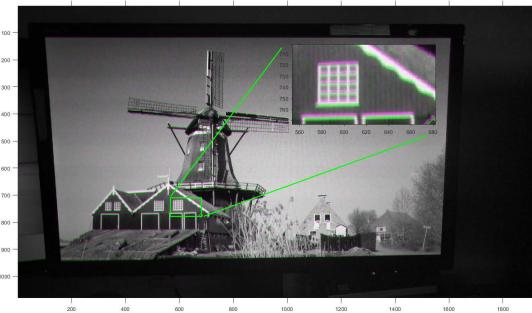


Abbildung 5.1: Bilder vor Bildregistrierung.

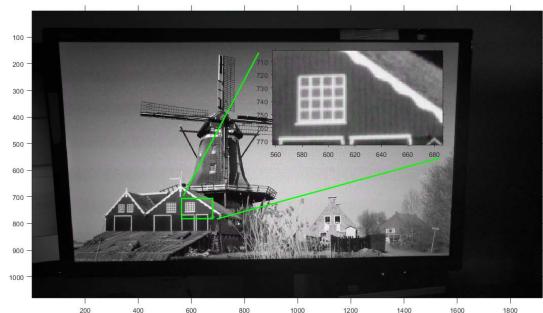


Abbildung 5.2: Bilder nach Bildregistrierung.

Aufgrund der Bildwiederholfrequenz werden die QR Muster in den Differenzbildern, wie in Tabelle 3.2 zu sehen ist, einige unerwartete Effekte aufweisen. Um diese zu neutralisieren, wird eine Absolutwertoperation durchgeführt. Danach werden hier 3 Differenzbilder welche die größte "Energie" beinhalten, überlagert um ein neues zu detektierende Bild entstehen. Das Ergebnis ist in Abbildung 5.3 zu sehen.

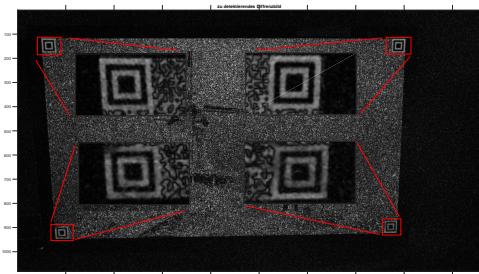


Abbildung 5.3: Ein zu detektierendes Bild.

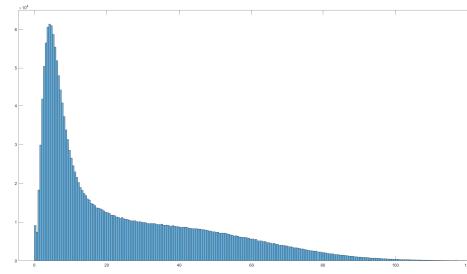


Abbildung 5.4: entsprechendes Histogramm.

Anschließend wird das Bild zur Bildverarbeitung durchgeführt. Abbildung 5.5 und Abbildung 5.6 zeigen die Ergebnisse durch die Binarisierung bzw. durch die Morphologie Operation. Nach der Binarisierungsvorgang mit Ostu wird der Modulationsbereich vom Bild grob extrahiert. Die nächste Morphologie Operation kann die Lücken und die kleinen Punkte vom Rauschen stark reduziert. Hier wird ein 5×5 quadratisches Strukturelement verwendet. Mit der

QR Muster Detektion kann das Zentrum des Musters bestimmen werden, wie in Abbildung 5.7 zu sehen ist. Dadurch lässt sich der Modulationsbereich bestimmen. In Abbildung 5.8 wird der Modulationsbereich durch ein rotes Rechteck gekennzeichnet. Schließlich kann mit einer projektiven Transformation das Endergebnis, welche in Abbildung 6.6 zu sehen ist, erhalten werden.



Abbildung 5.5: Binär.

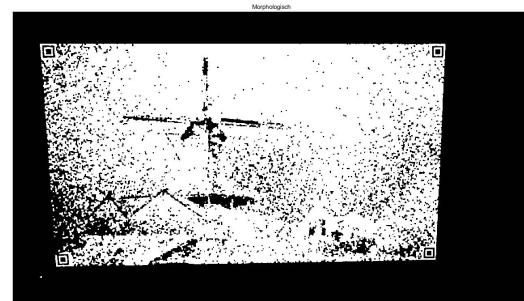


Abbildung 5.6: Morphologisch.



Abbildung 5.7: QR Muster.



Abbildung 5.8: Modulationsbereich.

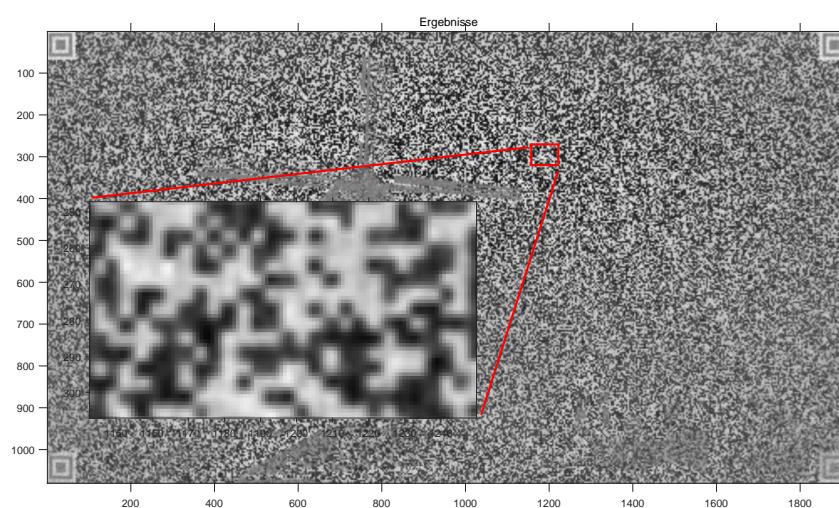


Abbildung 5.9: Modulationsbereich im U-Kanal.

5.3 Implementierung vom zweiten Verfahren

Hier wird die Implementierung der zweiten Methode vorgestellt. Das Bild stammt aus einer Reihe von acht Bildern, die durch ein Google Pixel mit Stativ aufgenommen wurden. In dieser Situation wird keine Bildregistrierung verwendet. Dies bedeutet, dass sich der Inhalt des Videos verändern lässt. Die Modulationsamplitude werden hier auf 6 und der Daten Datenblock auf $4 \times 4\text{Pixel}$ eingestellt.

Durch das Wissen des 4. Kapitels, werden von diesen Bildern eine Reihe Differenzbilder erstellt. Die 3 Differenzbilder mit der größten „Energie“ werden addiert, um ein zu detektierendes Bild zu erhalten, wie in Abbildung 5.10 zu sehen ist. Anschließend wird unter Verwendung der Binarisierungsmethode, Ostu Schwellenmethode, ein Binärbild generiert, indem der Modulationsbereich des Bildes extrahiert wird. Um kleine Punkte und Lücken zu beseitigen, lässt sich eine Morphologie Operation durchführen. Das hier benutzte Strukturelement ist ein 5×5 Quadrat. Die Wirkungen beider Operationen werden in Abbildung 5.11 und Abbildung 5.12 gezeigt. Um der Einfluss der Kamera-Verzerrung zu neutralisieren, wird das Binärbild mit einer Cross Dilatation verarbeitet, also $n_{around} = 1$. Struktur der Cross Dilatation wird in Abbildung 5.13 gezeigt und Abbildung 5.14 zeigt diese Operation. In der Radon Transformation lässt sich das Bildzentrum auf den Ursprung setzen. Dann durch Vergleich der Integralwerte kann die längste Linie, welche die oben, unten, links und rechts liegt, gefunden werden. Um die Berechnungskosten zu reduzieren, werden nur die Linien innerhalb von $\pm 10^\circ$ erkannt.

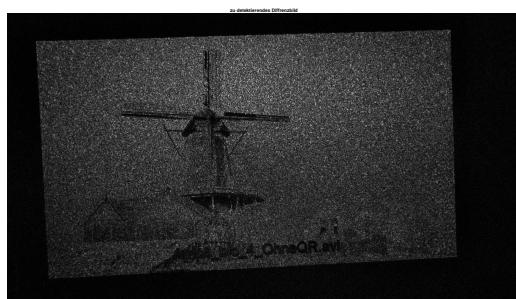


Abbildung 5.10: Ein zu detektierendes Bild.

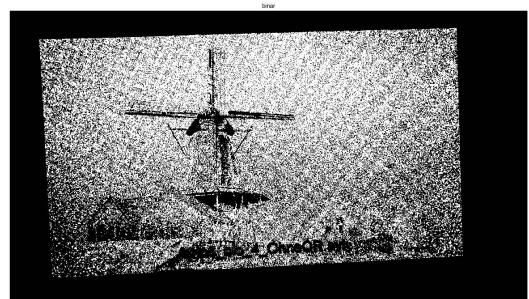


Abbildung 5.11: Binarisierung.

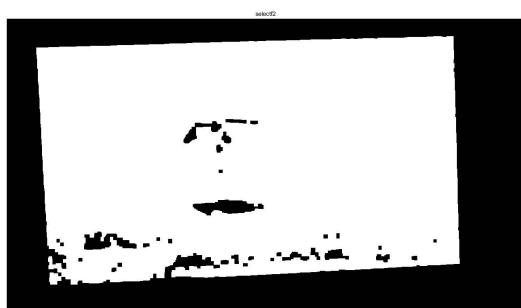


Abbildung 5.12: Morphologisch.

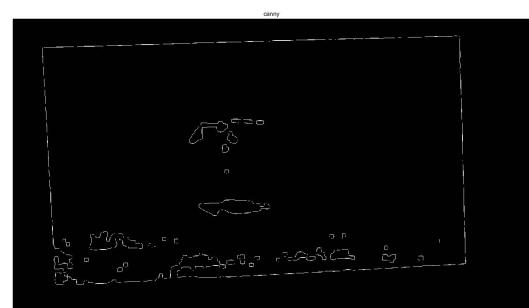


Abbildung 5.13: Canny Detektion.

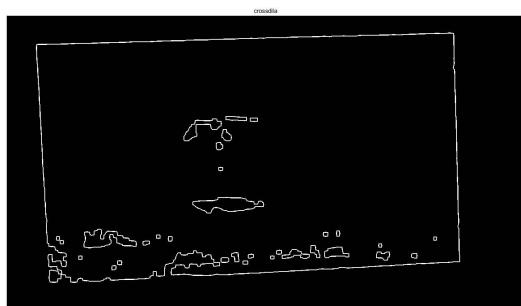


Abbildung 5.14: Cross Dilatation.



Abbildung 5.15: Radon Detektion.

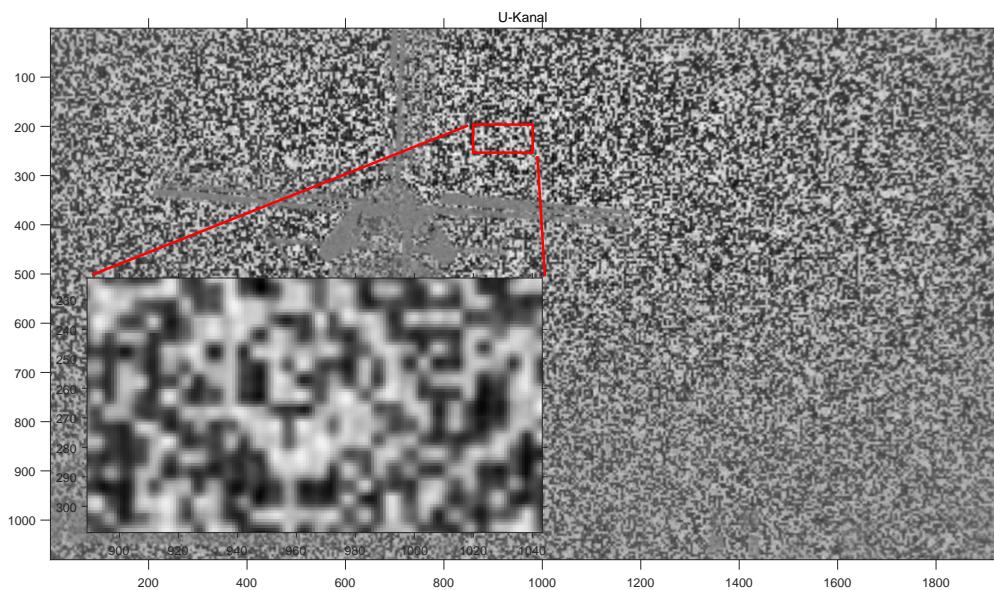


Abbildung 5.16: Modulationsbereich im U-Kanal.

5.4 Implementierung auf einer Smartphone-GPU

In diesem Abschnitt wird die Implementierung auf einer Smartphone-GPU erläutert. Hier wird Smartphone Xiaomi Mi3 benutzt. Dessen Parameter sind in Tabelle 5.2 zu sehen. Durch RenderScript als Framework für High Performance Computing auf der Android-Plattform, kann die vorherige Methode auf einer Smartphone-GPU parallel implementiert werden. Die verwendete Software auf der PC-Oberfläche ist Android Studio 3.1.4.

Tabelle 5.2: Grundlegende Parameter für Mi3.

Smartphone	GPU	CPU	RAM
Xiaomi Mi3	Qualcomm Snapdragon 800 MSM8274AB	Krait 400, 2300MHZ	2 GB, 800 MHZ

RenderScript

RenderScript stellt eine native High-Performance-Computing-API bereit, die in C (C99-Standard) geschrieben ist. Mit Renderscript können Anwendungssoftwares automatisch verschiedene Operationen parallel über alle verfügbaren Prozessorkerne ausführen. Sie bietet auch Unterstützung für verschiedene Verarbeitungsmedien wie CPU, GPU oder DSP. Renderscript ist nützlich für die Grafikverarbeitung, mathematische Modellierung oder jede andere Anwendung die viele mathematische Berechnungen erfordert. Darüber hinaus kann auf all diese Funktionen zugegriffen werden, um verschiedene Architekturen oder eine unterschiedliche Anzahl von Prozessorkernen zu unterstützen ohne Code zu schreiben. Es ist auch nicht notwendig, die Anwendungssoftware für verschiedene Prozessortypen zu kompilieren, da der RenderSkript-Code kompiliert wird, während er auf dem Gerät ausgeführt wird.

Die Vorteile für RenderScript sind folgende:

1. Kompatibilität: Renderscript ist für die Ausführung auf vielen Geräten in verschiedenen Prozessor- (CPU-, GPU- und DSP-Instanzen) Architekturen ausgelegt. Alle unterstützten Architekturen sind nicht für jedes bestimmte Gerät spezifisch, da sein Code zur Laufzeit auf dem Gerät kompiliert und zwischengespeichert wird.
2. Effizient: Renderscript stellt parallel eine Hochleistungs-Computing-API zur Verfügung, die den Kernel über das gesamte Gerät verteilt.
3. Einfach zu verwenden: Renderscript vereinfacht die Entwicklung, mit Möglichkeiten wie z.B. das Abbrechen von JNI-Code.

Das Ergebnis der Implementierung auf Xiaomi Mi-3 wird in Abbildung 5.17 gezeigt.

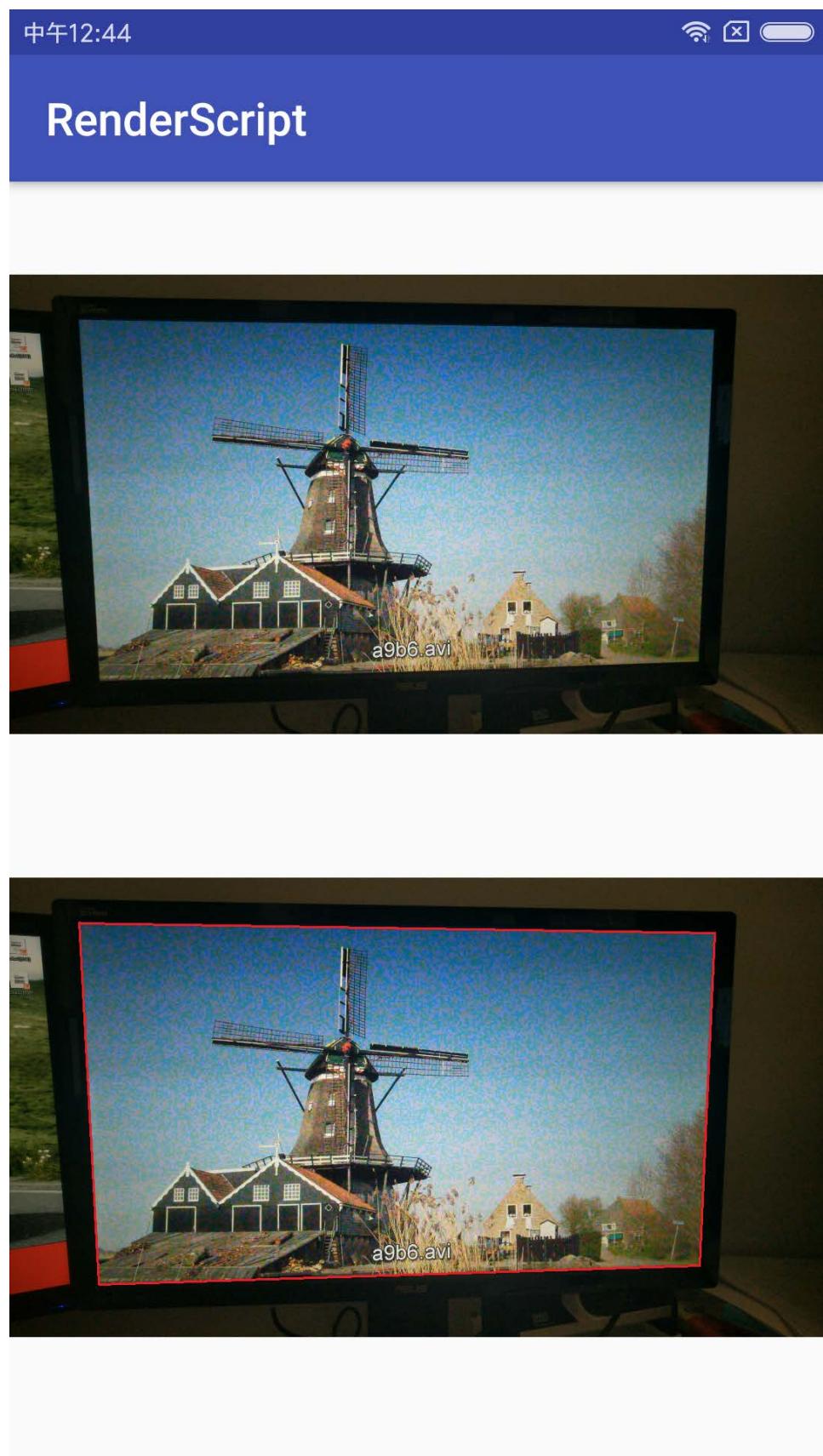


Abbildung 5.17: Ergebnis auf dem Xiaomi Mi3.

6 Evaluierung

In diesem Kapitel soll das zuvor implementierte Verfahren evaluiert werden. Es lassen sich eine Reihe von Testanalysen durchführen, um die Methode auf die Genauigkeit und die Leistungsfähigkeit zu testen.

6.1 Evaluierung vom ersten Verfahren

Hier wird das 1. Verfahren mit zusammen mit dessen Genauigkeit und Leistungsfähigkeit evaluiert. Wie in Kapitel 3 vorgestellt, ist diese Methode hauptlich für den Fall der Aufnahme über das Smartphone aus der Hand vorgesehen. Anschließend wird eine Bildregistrierungs Operation benötigt, um die Bilder in dasselbe Korordinatensystem zu transformieren. Die Funktion dieses Verfahrens wirkt sich direkt auf die Entstehung der Differenzbilder aus und hat einen wesentlichen Einfluss auf die spätere Erkennung. Deswegen wird zuerst eine Analyse durchgeführt.

Um die Leistung intuitiv auszudrücken, lassen sich einige Startparameter setzen. Die Modulationsamplitude beträgt 6, der Datenblock ist $4 \times 4\text{Pixel}$ und die QR Muster Größe ist $72 \times 72\text{Pixel}$. In der Reihe von Bildern wird das erste Bild und das letzte Bild verglichen, wie in Abbildung 6.1 zu sehen ist.

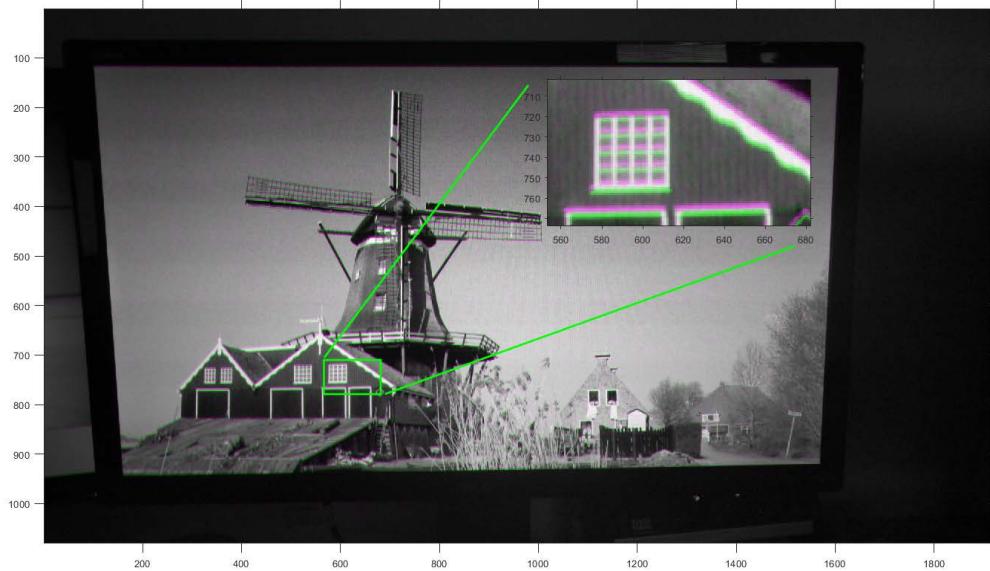


Abbildung 6.1: Verzerrung durch Handschütteln.

Aus der Abbildung ist deutlich zu erkennen, dass das Bild durch Handschütteln eine deutliche Verschiebung hat. Die durchschnittliche Entfernung der entsprechenden Pixel zwischen den zwei Bildern hier beträgt 2.3213 Pixel. Durch die Bildregistrierung, die in dieser Arbeit vorgestellt wurde, kann dieses Problem effektiv behoben werden. Abbildung 6.2 zeigt die Konvergenzkurve des Fehlers J bei der Summenwertbildung im Quadrat aller Korrespondenzpunkte.

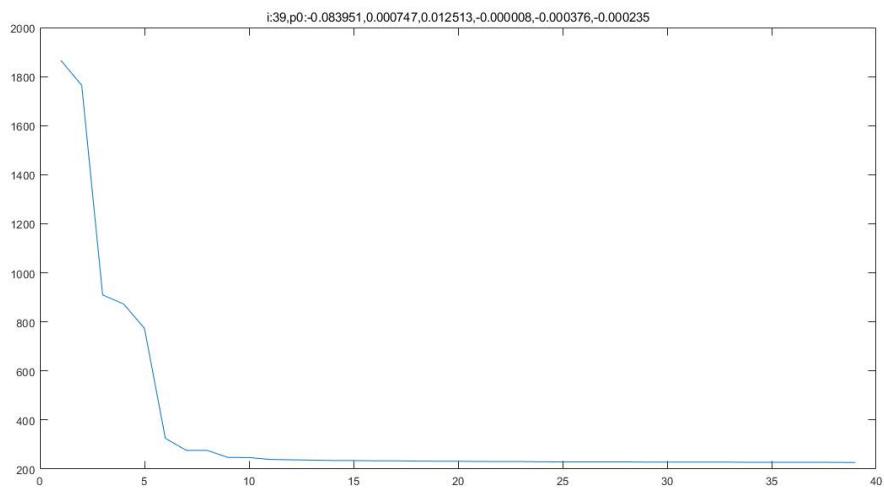


Abbildung 6.2: Konvergenzkurve von J .

Die vertikale Koordinate repräsentiert den Fehlerwert J und die horizontale Koordinate repräsentiert die Anzahl der Iterationen des Algorithmus. Wie in der Abbildung zu sehen ist, wird J nach etwa 10 Iterationen konvergiert. Die detailliert gesammelten Daten sind in Tabelle 6.1 abgebildet.

Tabelle 6.1: Daten des Fehlers.

Verglichene Bilder	Anzahl der Punkte	Vor Bildregistrierung		Nach Bildregistrierung	
		Gesamtfehler	Mittelwert	Gesamtfehler	Mittelwert
1-2	849	350.5674	0.4129	224.9265	0.2649
1-3	723	826.8834	1.1437	236.0301	0.3265
1-4	619	662.3621	1.0701	208.4760	0.3368
1-5	500	723.2307	1.4465	207.1174	0.4142
1-6	835	1607.4002	1.9250	231.7138	0.2775
1-7	802	1059.4579	1.3210	260.3104	0.3246
1-8	803	1863.9972	2.3213	246.7970	0.3073

Abbildung 6.3 zeigt den Wirkung, der in Abbildung 6.1 durch die Bildregistrierung behoben wurde. Wie zu sehen ist, ist der Fehler kompensiert. Abbildung 6.3 zeigt die Abweichung in horizontaler und vertikaler Richtung zwischen den entsprechenden Punkten nach der Operation. Die Abweichung in beiden Richtungen beträgt ca. 95% also unter 0.5 Pixel. Die durchschnittlichen Abweichungen in horizontaler und vertikaler Richtung betragen 0.1746 bzw. 0.1956 Pixel. Also beträgt die neue durchschnittliche Entfernung 0.3073 Pixel.

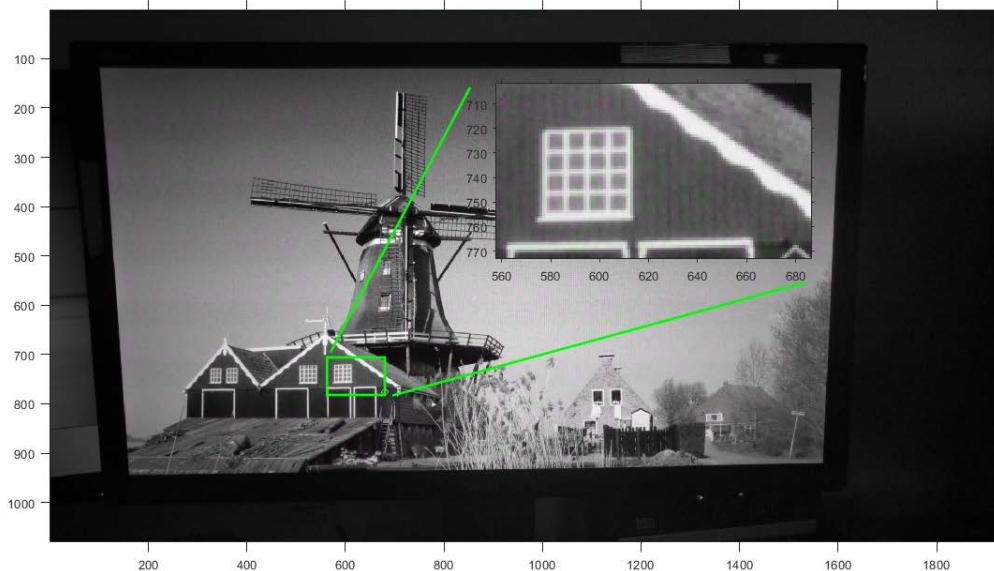


Abbildung 6.3: Bild nach Registrierung.

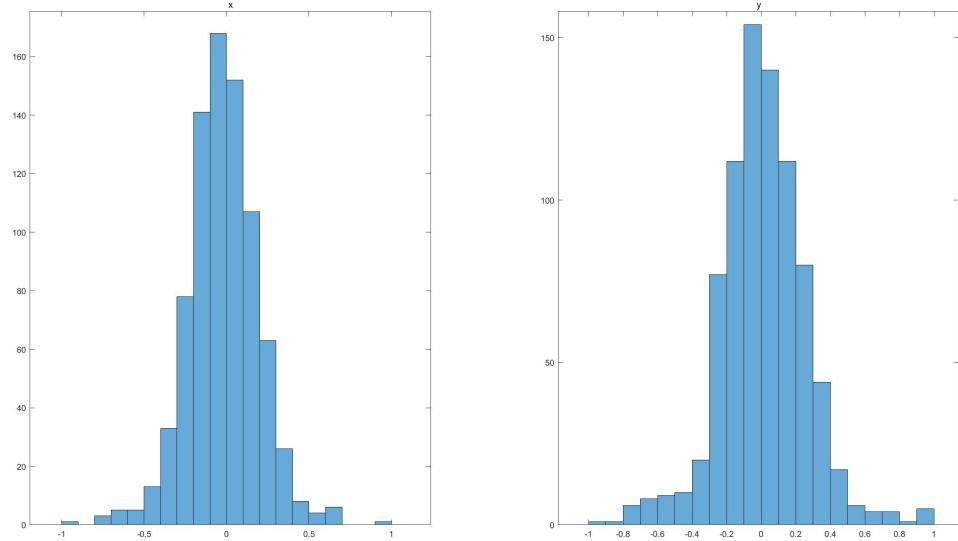


Abbildung 6.4: Histogramm für die Abweichungen in beiden Richtungen.

Als nächstes folgt eine Gesamtbewertung von Methode 1. Hier wird ein spezielles Bild für das Testen verwendet, d.h. das Gitter wird zum Originalbild hinzugefügt wessen Größe im Voraus bestimmt wurde. Hier wird ein Gitter von 200×200 ausgewählt. Abbildung 6.3 zeigt dieses spezielle Bild.



Abbildung 6.5: Testbild mit Gitter.

Der zu detektierende Modulationsbereich durch die 1. Methode wird in Abbildung 6.3 gezeigt.

Aus der Abbildung wurde die Abweichung jedes Gitterpunkts markiert. Der Untergraph in der oberen rechten Ecke der Abbildung ist eine Teilvergrößerung des roten rechteckigen Bereichs. Die durchgezogene Linie repräsentiert die Position des Gitters im Idealfall, und die gestrichelte Linie repräsentiert die tatsächliche Position des Gitters, nachdem das 1. Verfahren durchgeführt wurde. Der rote Pfeil zeigt die Abweichung des Punktes an. Die daraus berechnete durchschnittliche Abweichung beträgt 1.7680 Pixel.

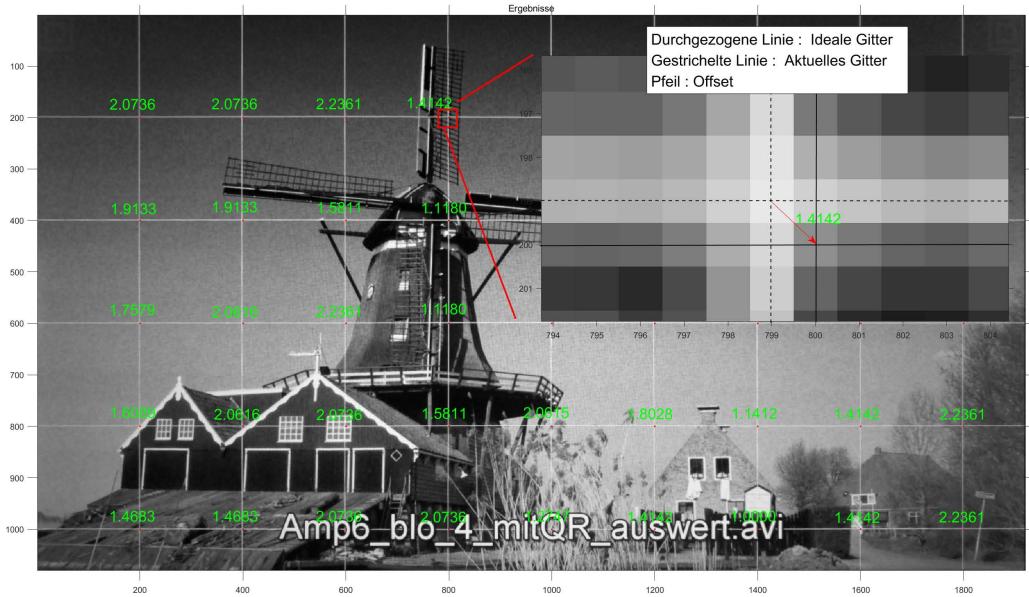


Abbildung 6.6: Ergebnis vom ersten Verfahren.

Die Laufzeitleistung bei Verwendung vom ersten Verfahren ist in Tabelle 6.2 aufgeführt, und die entsprechenden Detektionsergebnisse sind in Abbildung 6.3 dargestellt. Der Parameter A9B6 in der Tabelle repräsentiert eine Modulationsamplitude von 9 und die Größe des Datenblocks von 6×6 Pixel. Alle Experimente wurden als ein Reihe von acht Bildern eingegeben. Aus der Tabelle beträgt die durchschnittliche Zeit der QR-Detektion etwa 0.07 Sekunden.

Tabelle 6.2: Laufzeitleistung vom ersten Verfahren.

No.	Parameter	Display	Bildregistrierung[s]	Differenzbild Opt.[s]	Bildverarbeitung[s]	QR-Detektion[s]
1	A9B6	ASUS	56.5812	2.0524	0.0502	0.0709
2	A9B6	OLED	56.9922	2.1650	0.0489	0.0720
3	A6B4	ASUS	54.9934	2.1214	0.0512	0.0713
4	A6B4	OLED	55.9458	2.1593	0.0533	0.0789
5	A4B4	ASUS	57.8797	1.9985	0.0566	0.0717
6	A4B4	ASUS	56.9837	2.0786	0.0527	0.0752

Als Nächst werden einige Ergebnisse der Verwendungen von den 2. Methode angegeben.



Abbildung 6.7: Entsprechende Ergebnisse vom den ersten Verfahren.

6.2 Evaluierung vom zweiten Verfahren

In diesem Abschnitt wird das 2. Verfahren evaluiert. Wie in Kapitel 4 vorgestellt, ist diese Methode hauptlich für den Fall bei der Aufnahme der Bilder mit einem Smartphone auf einem Stativ. Infolgedessen wird hier die Bildregistrierungs Operation nicht mehr gebraucht. Die Gitterbilder aus dem vorherigen Abschnitt werden hier ebenfalls verwendet. Der Parameter ist eingestellt auf: Modulationsamplitude als 6, der Datenblock als $4 \times 4\text{Pixel}$ und die QR Muster

Größe als $72 \times 72\text{Pixel}$. Die Ergebnisse nach Methode 2 sind in der Abbildung 6.8 dargestellt. Die berechnete durchschnittliche Abweichung beträgt 1.3468 Pixel.

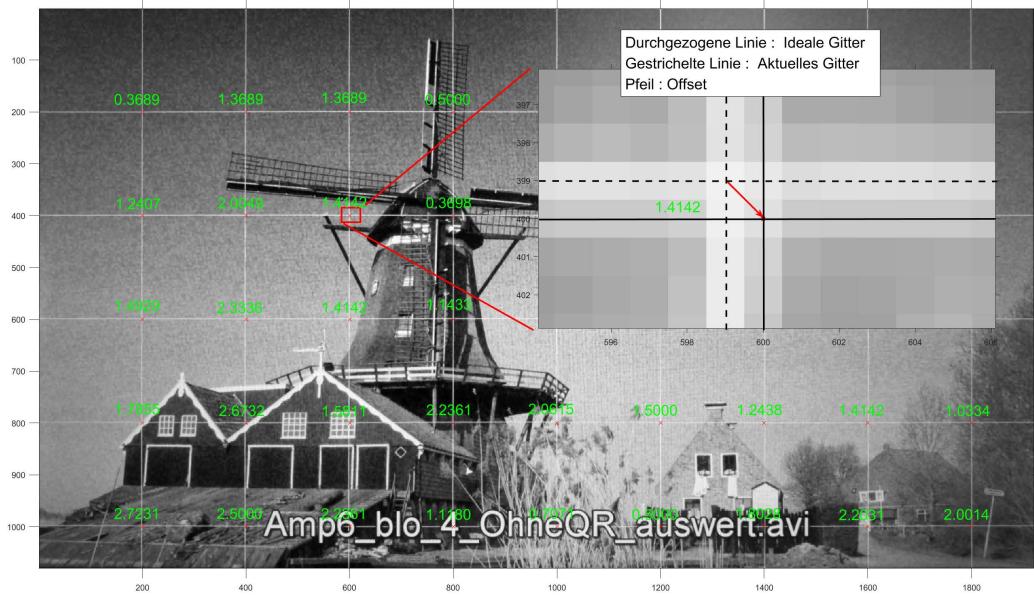


Abbildung 6.8: Ergebnis vom zweiten Verfahren.

Tabelle 6.2 zeigt die Zeitperformance bei Verwendung des zweiten Verfahrens. Die entsprechenden Detektionsergebnisse sind in Abbildung 6.3 zu sehen. Wie bei dem ersten Verfahren, repräsentiert A9 die Modulationsamplitude von 9 und B6 die Größe des Datenblocks von $6 \times 6\text{Pixel}$. Alle Experimente wurden als ein Reihe von acht Bildern eingegeben. Aus der Tabelle ist ersichtlich, dass die durchschnittliche Zeit der Radon Transformation etwa 0.3 Sekunden beträgt.

Tabelle 6.3: Laufzeitleistung vom zweiten Verfahren.

No.	Parameter	Display	Differenzbild Opt.[s]	Bildverarbeitung[s]	Radon[s]
1	A9B6	ASUS	2.1544	0.1275	0.2792
2	A9B6	OLED	2.3691	0.1283	0.2795
3	A9B6	Notebook	2.1059	0.1340	0.3009
4	A6B4	ASUS	2.0939	0.1309	0.3086
5	A6B4	OLED	2.0882	0.1293	0.2725
6	A6B4	OLED	2.0573	0.1321	0.3382
7	A4B4	ASUS	2.4397	0.1284	0.3726
8	A4B4	ASUS	2.4893	0.1296	0.3529



(a) No.1



(b) No.2



(c) No.3



(d) No.4



(e) No.5



(f) No.6



(g) No.7



(h) No.8

Abbildung 6.9: Entsprechende Ergebnisse vom den zweiten Verfahren.

7 Zusammenfassung

In dieser Arbeit werden zwei Verfahren zur Ausschnittsdetektion für die Screen-Camera Visible Light Communication untersucht und implementiert. Methode 1 ist vom differentiellen Modulationsverfahren des David-Systems inspiriert. d.h. Daten werden mit geringer Amplitude im Videosignal differentiell überlagert. In der Methode wird ein QR Muster an jeder Ecke der Datenebene hinzugefügt und dann mit den Daten zusammen hinter dem Bild moduliert. Diese Methode kann nicht nur den unschönen Effekt lösen, welcher durch direktes Hinzufügen des QR Musters zu dem Videobild verursacht wird, sondern kann auch den QR Muster durch seine spezifische geometrische Eigenschaft, d.h. in jeder Richtung beträgt das Breiteverhältnis $1 : 1 : 3 : 1 : 1$, effektiv erkennen. Dadurch wird schließlich der Modulationsbereich bestimmt. Das 2. Verfahren wird in den physikalischen Eigenschaften des Modulationsbereichs ausgeführt. Im DaVid System wirkt der Bildschirm als Übertragungsende, und der Bildschirmbereich ist der Modulationsbereich, der im Allgemeinen in Form eines Rechtecks ist. Dadurch kann dieses Problem in eine rechteckige Erkennung umgewandelt werden. Durch die Radon Transformation wird die längste gerade Linie in dem Bild detektiert, um ein Rechteck bzw. den Modulationsbereich, zu bestimmen.

zwei Aufnahmebedingungen wurden in dieser Arbeit berücksichtigt, eine ist die Aufnahme mit dem Smartphone aus der Hand und die andere auf einem Stativ. Handheld-Kamera-Shooting ist eine der häufigsten und bequemsten Aufnahmemethoden im Alltag, und hat eine unerlässliche Position in der praktischen Anwendung. Zu diesem Zweck wurde ein Bildregistrierungsmodul speziell für Handschütteln entwickelt, welche Merkmals Detektion, RANSAC, Kamera Modell und Nichtlineare Optimierung beinhaltet. Beim Testen beträgt die relative Verschiebung zwischen den entsprechenden Punkten des korrigierten Bildes c.a. 0.3 Pixel. Dies führt zu einem relativ guten Differenzbild. Dieser spielte eine wichtige Rolle weil beide Verfahren in dieser Arbeit auf der Detektion von Differenzbildern basieren. Dafür wird der Begriff "*Energie*" eingeführt, welcher die Klarheit des Modulationsbereichs repräsentiert. Gemäß der Energiesortierung werden die klarsten Bilder ausgewählt und hinzugefügt, um das zu detektierende Bild zu erhalten. In Methode 1 werden die beiden obigen Schritte implementiert, und durch Auswertung die durchschnittliche Abweichung ca. 1 – 2 Pixel bemessen. Es sollte hier angemerkt werden, dass aufgrund der Eigenschaften der Bildregistrierung das Video, das in Methode 1

aufgenommen wurde, ein statisches Video ist, d.h. der Inhalt des Videos hat sich nicht geändert. Der Grund dafür ist, dass der Punkt der Merkmals Detektion hauptsächlich auf dem Inhalt des Videos basiert. Wenn sich der Videoinhalt stark ändert, ist es schwierig, die Genauigkeit der Bildwiederherstellung zu garantieren. Einige bei der Arbeit versuchten Methoden war die Abschirmung des zentralen Bereichs während der Merkmals Detektion, d.h. nur die Merkmale in der Umgebung des Bildes wurden erkannt, was nicht zufriedenstellend ist. Methode 2 implementiert die Erkennung des Modulationsbereichs unter Verwendung eines Stativs. Nach der Erkennung beträgt die durchschnittliche Abweichung ebenfalls ca. 1 – 2 Pixel. In beiden Methoden kann der Modulationsbereich mit Parametern, wie die Modulationsapmlitude als 4, der Datenblock als $4 \times 4\text{Pixel}$, erfolgreich bestimmt werden.

Für dieses Thema gibt es einige mögliche weitere Forschung in der Zukunft. Zuerst kann wie oben erwähnt, ein Verfahren gefunden werden, welches beim Aufnehmen des dynamischen Videos mit dem Smartphone aus der Hand, das Handschütteln-Problem effektiv löst. Der Fokus liegt darauf, die Wirkung vom Video-Inhalt zu neutralisieren und das Bild nur auf die Merkmale des Hintergrunds zu filtern. Zweitens, die in dieser Arbeit verwendete Differenzbild Optimierungsmethode, kann aufgrund der Berechnung der „*Energie*“ für jedes Differenzbild zu einer großen Rechenlast führen. Ein effektives Verfahren wäre versuchen, ein zu detektierendes Bild, das auf dem bekannten Differenzbild basiert, effizient zu generieren. Schließlich wurde in Experiment festgestellt, dass unter denselben Parametern die Verwendung des gleichen Smartphones mit dem gleichen Video-Inhalt auf verschiedenen Bildschirmen, zu unterschiedlichen Ergebnissen führt. Dadurch können die Auswirkungen der Bildschirmfaktoren, wie z.B. Beleuchtungsart, Auflösung, Bildwiederholfrequenz, weiter erforscht werden.

Abkürzungsverzeichnis

abc	test
DaVid	Data transmission using Video devices
EEPROM	electrically erasable programmable read-only memory
GPU	Graphics Processing Unit
HDTV	high definition television
IEEE	Institute of Electrical and Electronics Engineers
ISI	inter symbol interference
LED	Light-Emitting Diode
NCC	Normalized Cross Correlation
RANSAC	RANDom SAmple Consensus
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Features
VLC	Visible Light Communication

Symbolverzeichnis

Å non-SI unit of length

A non-SI unit of length

Ω unit of electrical resistance

π ratio of circumference of circle to its diameter

Abbildungsverzeichnis

2.1	Eine beispielhafte Implementierung des DaVid-Systems	3
2.2	Schematische Darstellung vom DaVid-System	5
2.3	Blockschaltbild der Signalverarbeitung in zeitlich-differentieller Modulation	6
3.1	Strukturdiagramm	10
3.2	Flussdiagramm der ersten Methode	11
3.3	Flussdiagramm der Bildregistrierung	12
3.4	Scale space	15
3.5	Extremwert Erkennung	15
3.6	Dominante Orientierung feststellen	16
3.7	Merkmalspunkt Deskriptor	16
3.8	SURF Merkmale	17
3.9	Linien Detektion	18
3.10	Ohne RANSAC	20
3.11	Mit RANSAC	20
3.12	Modell einer Lochkamera	21
3.13	Transformation vom Weltkoordinatensystem zum Kamerakoordinatensystem . .	22
3.14	Rotation um Z-Achse	23
3.15	Transformation vom Kamerakoordinatensystem zum Bildkoordinatensystem .	24
3.16	Konvertierung vom Bildkoordinatensystem zum Pixelkoordinatensystem . .	25
3.17	Rotationsbewegungsmodell	28
3.18	Transformation in das selbe Koordinatensystem	30
3.19	Flussdiagramm für Optimierung	31
3.20	Differenzbild Flussdiagramm	34
3.21	Ein zu detektierendes Bild	35
3.22	QR Muster	36
3.23	QR Muster Ratio	37
3.24	Flussdiagramm der QR Muster Detektion	39
3.25	Beispiel einer QR Muster Detektion	40
4.1	Idee für Detektion	42
4.2	Flussdiagramm der zweiten Methode	43

4.3 Histogramm eines einfachen Beispiels	44
4.4 Binarisierung mit globalen Schwellenwertmethode	46
4.5 Binarisierung mit lokaler Schwellenwertmethode	46
4.6 Binarisierung mit Ostu	48
4.7 Binarisiertes zu detektierendes Bild	48
4.8 Einige grundlegende Strukturelemente	49
4.9 Dilatation und Erosion	50
4.10 Öffnung und Schließung	52
4.11 Bild nach einer morphologischen Operation	53
4.12 Binärbild nach der Canny Detektion	57
4.13 ρ, θ Parameterraum	58
4.14 Houghdetektion	60
4.15 Parametrisierte Linie	61
4.16 Integral der Funktion $f(x, y)$	61
4.17 Radon Transformation (θ, ρ)	62
4.18 Radon Transformation (θ)	62
 5.1 Bilder vor Bildregistrierung	64
5.2 Bilder nach Bildregistrierung	64
5.3 Ein zu detektierendes Bild	64
5.4 entsprechendes Histogramm	64
5.5 Binär	65
5.6 Morphologisch	65
5.7 QR Muster	65
5.8 Modulationsbereich	65
5.9 Modulationsbereich im U-Kanal	65
5.10 Ein zu detektierendes Bild	66
5.11 Binarisierung	66
5.12 Morphologisch	67
5.13 Canny Detektion	67
5.14 Cross Dilatation	67
5.15 Radon Detektion	67
5.16 Modulationsbereich im U-Kanal	67
5.17 Ergebnis auf dem Xiaomi Mi3	69
 6.1 Verzerrung durch Handschütteln	71
6.2 Konvergenzkurve von J	71
6.3 Bild nach Registrierung	72
6.4 Histogramm für die Abweichungen in beiden Richtungen	73

Abbildungsverzeichnis

6.5	Testbild mit Gitter	73
6.6	Ergebnis vom ersten Verfahren	74
6.7	Entsprechende Ergebnisse vom den ersten Verfahren	75
6.8	Ergebnis vom zweiten Verfahren	76
6.9	Entsprechende Ergebnisse vom den zweiten Verfahren	77

Tabellenverzeichnis

3.1	Parameter der Kameras im Vergleich	27
3.2	mögliche Differenzbilder	33
5.1	Verwendete Bildschirme	63
5.2	Grundlegende Parameter für Mi3	68
6.1	Daten des Fehlers	72
6.2	Laufzeitleistung vom ersten Verfahren	74
6.3	Laufzeitleistung vom zweiten Verfahren	76

Literatur

- [1] T. Komine und M. Nakagawa, „Integrated system of white LED visible-light communication and power-line communication“, *IEEE Transactions on Consumer Electronics*, Jg. 49, Nr. 1, S. 71–79, Feb. 2003.
- [2] X. Liu, H. Makino und Y. Maeda, „Basic study on indoor location estimation using Visible Light Communication platform“, in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2008, S. 2377–2380.
- [3] D. V. Plant und A. G. Kirk, „Optical interconnects at the chip and board level: challenges and solutions“, *Proceedings of the IEEE*, Jg. 88, Nr. 6, S. 806–818, Juni 2000.
- [4] X. Huang, B. Chen und A. Roscoe, „Multi-channel key distribution protocols using visible light communications in body sensor networks“, *Computer Science Student Conference 2012*, S. 15, 2012.
- [5] R. Kays, C. Brauers und J. Klein, „Modulation concepts for high-rate display-camera data transmission“, in *2017 IEEE International Conference on Communications (ICC)*, Mai 2017, S. 1–6.
- [6] R. Kays, C. Brauers und J. Klein, „DaViD : Data Transmission Using Video Devices-An Innovative System for Smart Media Applications“, 2016.
- [7] R. Kays, „Modulation concepts for visible light communication using video displays“, in *2015 IEEE 5th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, Sep. 2015, S. 388–392.
- [8] R. Kays, „Visible light communication using TV displays and video cameras“, in *2015 IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2015, S. 554–555.
- [9] H. Bay, A. Ess, T. Tuytelaars und L. V. Gool, „SURF:Speeded Up Robust Features“, Bd. 110, 2008, S. 346–359.
- [10] M. A. Fischler und R. C. Bolles, „Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography“, *Commun. ACM*, Jg. 24, Nr. 6, S. 381–395, Juni 1981. Adresse: <http://doi.acm.org/10.1145/358669.358692>.

- [11] Z. Zhengyou, „A Flexible New Technique for Camera Calibration“, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jg. 22, S. 1330–1334, Dez. 2000.
- [12] N. Otsu, „A Threshold Selection Method from Gray-Level Histograms“, *IEEE Transactions on Systems, Man, and Cybernetics*, Jg. 9, Nr. 1, S. 62–66, Jan. 1979.
- [13] J. Canny, „A Computational Approach to Edge Detection“, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jg. PAMI-8, Nr. 6, S. 679–698, Nov. 1986.

Eidesstattliche Versicherung

Chen, Xiaoyu

Name, Vorname

181225

Matr.-Nr.

Ich versichere hiermit an Eides statt, dass ich die vorliegende Masterarbeit mit dem Titel

Untersuchung von Verfahren zur Ausschnittsdetektion für die Screen-Camera Light Communication

selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäß Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Dortmund, 9. September 2018

Ort, Datum

Unterschrift

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -)

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Dortmund, 9. September 2018

Ort, Datum

Unterschrift