

Fakultät für Elektrotechnik und Informationstechnik



Lehrstuhl für
Kommunikations-
technik

Bachelorarbeit B 14-2015

**Tasty Kanalmodell
für die drahtlose Kommunikation
zwischen Gebäuden und
Außeninstallationen**

von

Käpt'n Kevin Blaubär

Abgabedatum: 23. Januar 2018

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

Inhaltsverzeichnis

1 Einleitung	1
1.1 Section	1
2 DaViD	3
2.1 Einführung des DaViD	3
2.2 Systemmodell	4
2.2.1 Modulationsverfahren	5
2.2.2 DatenBlock	8
2.3 Anwendungsgebiete	8
3 Erste Methode	10
3.1 Allgemeine Struktur	10
3.2 Bildregistrierung	13
3.2.1 SURF	13
3.2.2 RANSAC	18
3.2.3 Bilder Umwandlung	22
3.3 Differenzbild Optimierung	33
3.4 Bildverarbeitung	36
3.5 QR Musters Detektion	37
4 Zweite Methode	41
4.1 Allgemeine Struktur	41
4.2 Binariesierung	41
4.3 Morphologie	43
4.4 Canny detection	48
4.5 Hough detection	48
5 Auswertung	49
5.1 Section	49
6 Zusammenfassung	50
6.1 Section	50

A Erster Anhang	51
A.1 Section	51
Abbildungsverzeichnis	53
Tabellenverzeichnis	55
Quellcodeverzeichnis	56
Literatur	57

1 Einleitung

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1.1 Section

Jetzt nur noch schreiben! :)

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacinia tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacinia congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacinia commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacinia. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacinia vel est. Curabitur consectetur.

2 DaVid

In diesem Kapitel werden das **Data transmission using Video devices (DaVid)** System beschrieben. Zuerst läuft die Vorstellung des DaVid Systems. Die Systemmodell und Arbeitsprinzip des Systems werden in anschließenden Abschnitt erläutert. Schließlich folgt die mögliche Anwendungsgebiete des Systems. [1]–[4]

2.1 Einführung des DaViD

DaVid ist ein neuartiges Verfahren zur optischen Freiraum-Datenübertragung zwischen einem Display als Sender und einer Kamera 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 die zusätzliche Datenmodulation für menschliche Betrachter nahezu unsichtbar ist, der Benutzer leitet ein hochauflösende Kamera oder ein Smartphone zur Bildschirm, um die Szene aufzunehmen. Durch der eingebaut Prozessor können die Signale decodiert werden.

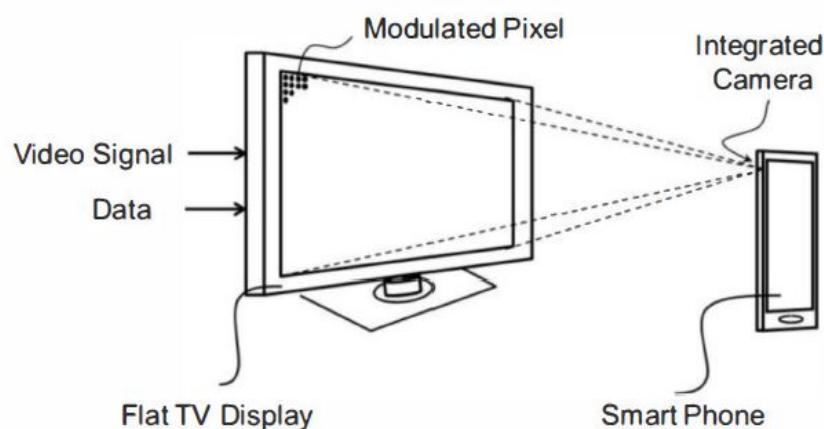


Abbildung 2.1: Eine beispielhafte Implementierung des DaVid-Systems.

2.2 Systemmodell

Bild in Display enthalten eine große Anzahl von Pixeln, die jeweils aus einer spezifischen Anordnung von Subpixelen für die RGB-Farbraum bestehen. Jeder einzelne Frame des Videos wird nämlich durch eine Matrix von Subpixelwerten dargestellt. DaVid System verwendet eine differentielle Modulationsmethod d.h. Teil der Videoinformationen muss wiederholt werden, indem Daten als ein symmetrischer Manchester-Code moduliert und zu den Videosignalkomponenten hinzugefügt werden. In Empfängerseite durch eine zeitliche Synchronisation können die zeitliche inter symbol interference (ISI) vermieden werden. Dann nach Verwendung einer örtlichen Synchronisation enthalten einen Differenzbild. Weil die Randbereich des Differenzbilds ungültig ist, verlässt sich die Modulationsgebiet durch die Verfahren in diese Arbeit entdecken. 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 von DaVid System.

2.2.1 Modulationsverfahren

Ein Modulationsverfahren, das die Videoqualität nicht offensichtlich reduziert garantiert, ist sehr wichtig für ein auf Videogerät basierendes Datenübertragungssystem. Die möglichen Modulationsverfahren in DaVid-System sind:

- Zeitliche differentielle Modulation der Luminanz
- Zeitliche differentielle Modulation der Chrominanz
- Örtliche differentielle Modulation der Luminanz
- Örtliche differentielle Modulation der Chrominanz

Zeitliche differentielle Modulationsverfahren lässt kontinuierliches Paar Frames den gleichen Luminanz- bzw. Chrominanz-Videoinhalt enthalten, d.h. durch Subtrahieren die mit daten addiert Kanal der Paar Frames die Differenzbild erhalten lassen können. Dagegen in örtlichen sind die benachbarte Pixel mit den gleichen Videoamplituden. Hier wird in dieser Arbeit nur zeitliche differentielle Modulation der Chrominanz verwendet. Abbildung 2.3 zeigt ein Blockdiagramm einer typischen Senderimplementierung durch zeitliche differentielle Modulation.

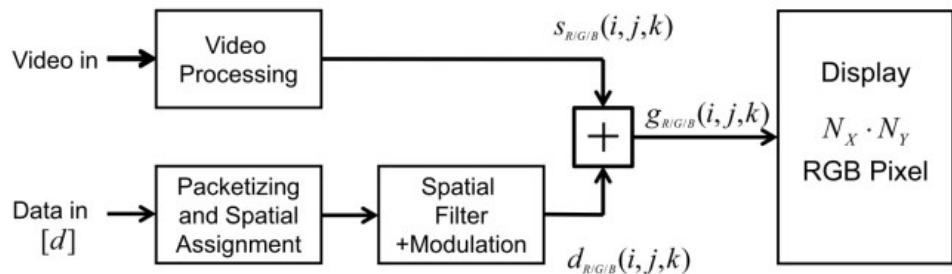


Abbildung 2.3: Blockschaltbild der Signalverarbeitung in zeitlicher differentieller Modulation.

Wir nehmen eine Diaplay an, die in horizontale Richtung N_x Pixel stehen, dagegen in vertikale Richtung N_y Pixel. Das Videoeingangssignal wird verarbeitet, um eine Anzeigeeingabe $s(i, j, k)$ zu liefern. Mit zeitliche differenziell Modulation ist die Videoinhalt des kommenden Frame dasselbe. Die Indizes i und j bedeuten die horizontale und vertikale Pixelposition auf dem Bildschirm, während k die Nummer des reproduzierten Bildes ist. Indiz m heißt den Zähler des Frames in einer Videosequenz. Der Amplitudengrenzenbereich des Videosignals sollte begrenzt sein, um die Addition kleiner Datenamplituden ohne Übersteuern zu ermöglichen.

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

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

Vor Datenübertragung muss der Datenstrom in Schichten der Länge L aufgeteilt werden. Indiz L bedeutet die Menge der Daten, die in einem Framepaar übertragen werden können. Ein direkter Ansatz ist eine direkte Zuordnung von Datenbits zu Pixeltripeln Zeile für Zeile.

$$\begin{aligned}
 d(l) &\rightarrow d(i, j, m) \quad d(l) \in \{-1, 1\} \\
 0 \leq l < L, L &= N_x \cdot N_y \\
 i = l \bmod N_x &\quad j = \lfloor l/N_x \rfloor
 \end{aligned} \tag{2.2}$$

Die Modulationsamplitude A ist ein wichtiger Parameter für Datenübertragung. Im Prinzip kann die Amplitude in verschiedenen Kanal unabhängig gewählt werden, um die Systemleistung zu optimieren. In diesen Arbeiten setzen die Amplitude gleichwertig.

$$A_R = A_G = A_B = A \tag{2.3}$$

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:

$$\begin{aligned}
 s_{R/G/B}(i, j, k) &= s_{R/G/B}(i, j, m) + A_{R/G/B} \cdot (2 \cdot d(i, j, m) - 1) \\
 s_{R/G/B}(i, j, k+1) &= s_{R/G/B}(i, j, m) - A_{R/G/B} \cdot (2 \cdot d(i, j, m) - 1)
 \end{aligned} \tag{2.4}$$

Ein Beispiel einer modulierten Bildfolge ist in Abbildung 2.4 gezeigt. Das Hinzufügen der modulierten Daten (hier mit $A = 4$) zu dem Videoeingang ergibt die Anzeigearamplituden in der rechten Spalte.



Abbildung 2.4: Ein Beispiel einer modulierten Bildfolge.

Im Vergleich zu Luminanzteil Y die Anzeigequalität in U und V Komponente ist signifikant besser, wenn Informationen in Chrominanz wiederholt und moduliert werden. Auf diese Weise wird die Gesamtleuchtdichte eines Pixel-Triple durch die Datenmodulation nicht beeinflusst. Die Umwandlungsmatrix, angegeben von test (ITU-R BT.709) für **high definition television** (HDTV) Display von Standard (R, G, B) zur Standard (Y, U, V) läuft:

$$T = \begin{pmatrix} 0,213 & 0,715 & 0,072 \\ -0,115 & -0,385 & 0,5 \\ 0,5 & -0,454 & -0,0458 \end{pmatrix} \quad (2.5)$$

Das Videosignal $s(i, j, k)$ muss vor dem Anwenden der Modulation in Y-, U- und V-Komponenten umgewandelt werden. Die nachfolgende inverse Konvertierung erklärt das Display-Eingangssignal in Abbildung 2.3:

$$\begin{aligned} \begin{pmatrix} g_R(i, j, k) \\ g_G(i, j, k) \\ g_B(i, j, k) \end{pmatrix} &= T^{-1} \cdot \left(T \cdot \begin{pmatrix} S_R(i, j, m) \\ S_G(i, j, m) \\ S_B(i, j, m) \end{pmatrix} + \begin{pmatrix} 0 \\ A_U \cdot d(i, j, m) \\ A_V \cdot d(i, j, m) \end{pmatrix} \right) \\ \begin{pmatrix} g_R(i, j, k+1) \\ g_G(i, j, k+1) \\ g_B(i, j, k+1) \end{pmatrix} &= T^{-1} \cdot \left(T \cdot \begin{pmatrix} S_R(i, j, m) \\ S_G(i, j, m) \\ S_B(i, j, m) \end{pmatrix} - \begin{pmatrix} 0 \\ A_U \cdot d(i, j, m) \\ A_V \cdot d(i, j, m) \end{pmatrix} \right) \end{aligned} \quad (2.6)$$

Diese Art der Modulation kann als eine Modulation des roten und des blauen Subpixels betrachtet werden, während das grüne Subpixel verwendet wird, um die Änderung der Luminanz des Pixel-Tripels zu kompensieren. Durch Definition korreliert A_U mit A_B und A_V mit A_R .

$$\begin{pmatrix} A_R \\ A_G \\ A_B \end{pmatrix} = T^{-1} \cdot \begin{pmatrix} A_Y \\ A_U \\ A_V \end{pmatrix} \quad (2.7)$$

2.2.2 DatenBlock

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

$$\begin{aligned} d(l) &\rightarrow d(x, y, k) \quad 0 \leq l < L \\ L &= \lfloor N_X/B_X \rfloor \cdot \lfloor N_Y/B_Y \rfloor \\ x &= (l \cdot B_X) \bmod N_X + r_X, \quad r_X = 0 \dots (B_X - 1) \\ y &= \lfloor l/\lfloor N_X/B_X \rfloor \rfloor \cdot B_Y + r_Y, \quad r_Y = 0 \dots (B_Y - 1) \end{aligned} \quad (2.8)$$

Wenn die Anzahl der Pixel pro Zeile kein Vielfaches von B_X ist oder wenn die Anzahl der Pixel kein Vielfaches von B_Y ist, muss die Anzahl der Pixel und Zeilen, die für die Modulation in Gleichung (2.5) verwendet werden, ersetzt werden durch:

$$\begin{aligned} N_X &= \lfloor N_X/B_X \rfloor \cdot B_X \\ N_Y &= \lfloor N_Y/B_Y \rfloor \cdot B_Y \end{aligned} \quad (2.9)$$

In dieser Arbeit werden das DatenBlock für quadratische Blöcke gesetzt.

$$B_X = B_Y = B. \quad (2.10)$$

2.3 Anwendungsgebiete

Die Datenübertragungsrate des DaVid-Systems wird voraussichtlich erreicht bis zu 100 Mbit/s. Es gehöre zu einer Sichtlinienübertragung für kurze Verbindungen. Geeignete Abdeckungsbereichen hängen von der Größe des Displays und der Kameraoptik ab. Obwohl im Vergleich zum letzten WLAN- Versionen Institute of Electrical and Electronics Engineers (IEEE) 802.11,

die Leistung scheint nicht so attraktiv. Der Vorteil liegt nicht nur in der wachsenden Leistungsfähigkeit von Video-Display und Kamera, aber auch die Option zur Wiederverwendung der bestehenden Hardware, die zum Zeigen des Videos installiert wurde. Ein empfohlene praktische Anwendungsbereich des DaVids ist öffentlicher Ort, z.B. U-Bahn-Station, großes Stadion und so weiter. Annehmen eine Situation, wenn die Leute auf ihre U-Bahn warten, sie können ihre eigene Software aktualisieren, indem Sie einfach auf die zeigende Werbung in dem Bildschirm leiten.

Berücksichtigen der Eigenschaften des DaVids, d.h. die Synchronisation von Videospielen und Datenübertragung. Viele Anwendungsszenarien können in Betracht gezogen werden und scheinen sehr attraktiv zu sein. Die drei Hauptszenarien sind:

- Indoor-individuelle Kommunikation: Kurzstreckenverbindungen basieren auf relativ kleinen (Tablet-Größe) Bildschirm, Anwendungen z.B. die Übertragung von Hintergrundinformationen an Besucher im Museum, Kiosk.
- Indoor-Multicast-Kommunikation: Streckabstand ist länger als ersten Fall auf relativ größer (40-100") Bildschirm, Anwendungen z.B. während Videoabspielen Besucher die Anwendungsdaten oder Mediendateien herunterladen können im Kiosk, Restaurant.
- Freien Kommunikation: Größter Bildschirm wie im Einkaufszentren oder Sport-Arenen, Anwendungen können denen des zweiten Szenarios ähneln.

Sobald die Dienste auf öffentlichen Bildschirmen implementiert werden, kann Leute mit Hilfe eines modernen Smartphones, die mit einer geeigneten Kamera eingebaut ist, nach der Installation einer neuen App innovative wahrnehmen.

3 Erste Methode

In diesem Kapitel wird die Realisierung des Erste Verfahrens eingegangen werden. Die Bildregistrierung wird zuerst im Detail gegeben. Anschließen läuft die Differenzbild Optimierung und die benötigte Bildverarbeitung. QR Muster Detektion werden schließlich erläutert. Zur Implementierung dieses Verfahrens wird Matlab unter der Lizenz TU Dortmund verwendet.

3.1 Allgemeine Struktur

Dieses Verfahren verwendet die Charakteristiken der Datenmodulation des DaViD Systems, d.h. an jedem Eck des Datenebene ein QR Muster hinzugefügt und dann mit den Daten zusammen hinter dem Bild moduliert. Am Empfängen nach einigen Operationen wird QR Muster mit die in diese Kapitel vorstellte Verfahren detektiert und schließlich bestimmt der Modulationsbereich. Diese Methode löst effektiv den unschönen Effekt, der das QR-Modul direkt zum Bild hinzugefügt wird, und kann das Problem günstiger und effektiver durch die QR-Mustereigenschaft lösen. Abbildung 3.1 zeigt die Strukturdiagramm dieses Verfahrens.

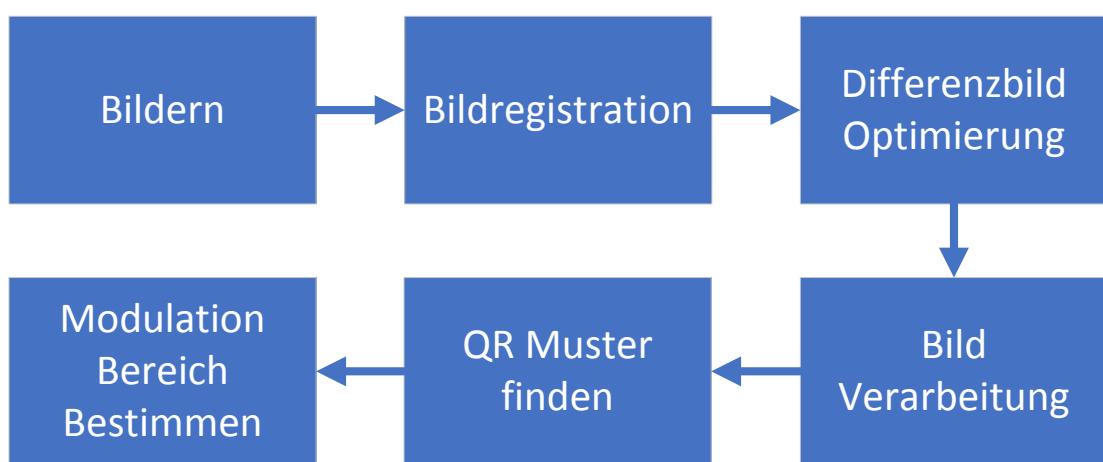


Abbildung 3.1: Strukturdiagramm.

Das Objekt, das mit dieser Methode bearbeitet wird, ist die aus einer spezifischen Handy App stammte Bildern. Diese App wurde speziell für das DaVid-System entwickelt und erstellt eine Reihe von Bildern bei jeder Aufnahme. In Allgemeiner wird die Kamera bei der Aufnahme in der Hand gehalten. Aufgrund von Handschütteln während des Aufnahme werden eine leichte Verschiebung zwischen den Bildern besteht. Um dieses Problem zu lösen, nehmen eine Bildregistrierung Operation, die zwei Bilder in dasselbe Koordinatensystem konvertiert. Zuerst durch Merkmalserkennung überprüfen die Merkmal der Bildern. Dann die entsprechenden Punkte zwischen den beiden Bilder werden durch Merkmalsextraktion und Merkmalanpassung erhalten. Hier braucht ein **R**ANdom **S**Amples **C**onsensus (RANSAC) Algorithmus, um die falschen Korrespondentpunkt zu beseitigen und die Genauigkeit zu verbessern. Aus dem Kameramodell wird dann ein mathematisches Transformations Modell zwischen den entsprechenden Punkten in den beiden Bildern erstellt. Hier bemerkt, der Prozess des Lösung dieser Transformationsmodell kann als ein nichtlineares Optimierungsproblem angesehen werden. Durch Lösung dieses Problems kann die Transformationsmatrix erhalten werden. In diesen Arbeit wählen das erste Bild immer als Referenzbild aus und nehmenn einen Bildregistrierung nacheinander mit anderen Bildern. Danach bekommen eine Reihe Bilder, die in dasselbe Koordinatensystem umgewandelt. Auswählen je zwei Bilder und subtrahieren, dann erhalten eine Reihe Differenzbildern. Um die folgend Detektion zu vereinfachen, machen eine Optimierungsoperation mit der Hilfe der geometrischen Eigenschaft des QR Musters und dann erstellen ein detektierendes Bild. Anschließen eine Reihe von Bildverarbeitung bzw. Binarisierung, Medianfilter, Morphologie Operation wird behandelt. Dadurch können die kleinen Punkte und Lücken, die durch Rauschen und Fehler verursacht werden, entfernt werden. Schließlich wird es durch die Charakteristik des QR Musters, das Breiteverhältnis $1 : 1 : 3 : 1 : 1$ beträgt, um QR Muster zu detektieren. Also bestimmen die Modulationsbereich. Die Flussdiagramm wird in Abbildung 3.2 gezeigt. Die Details jedes Teils werden in den folgenden Abschnitten beschrieben.

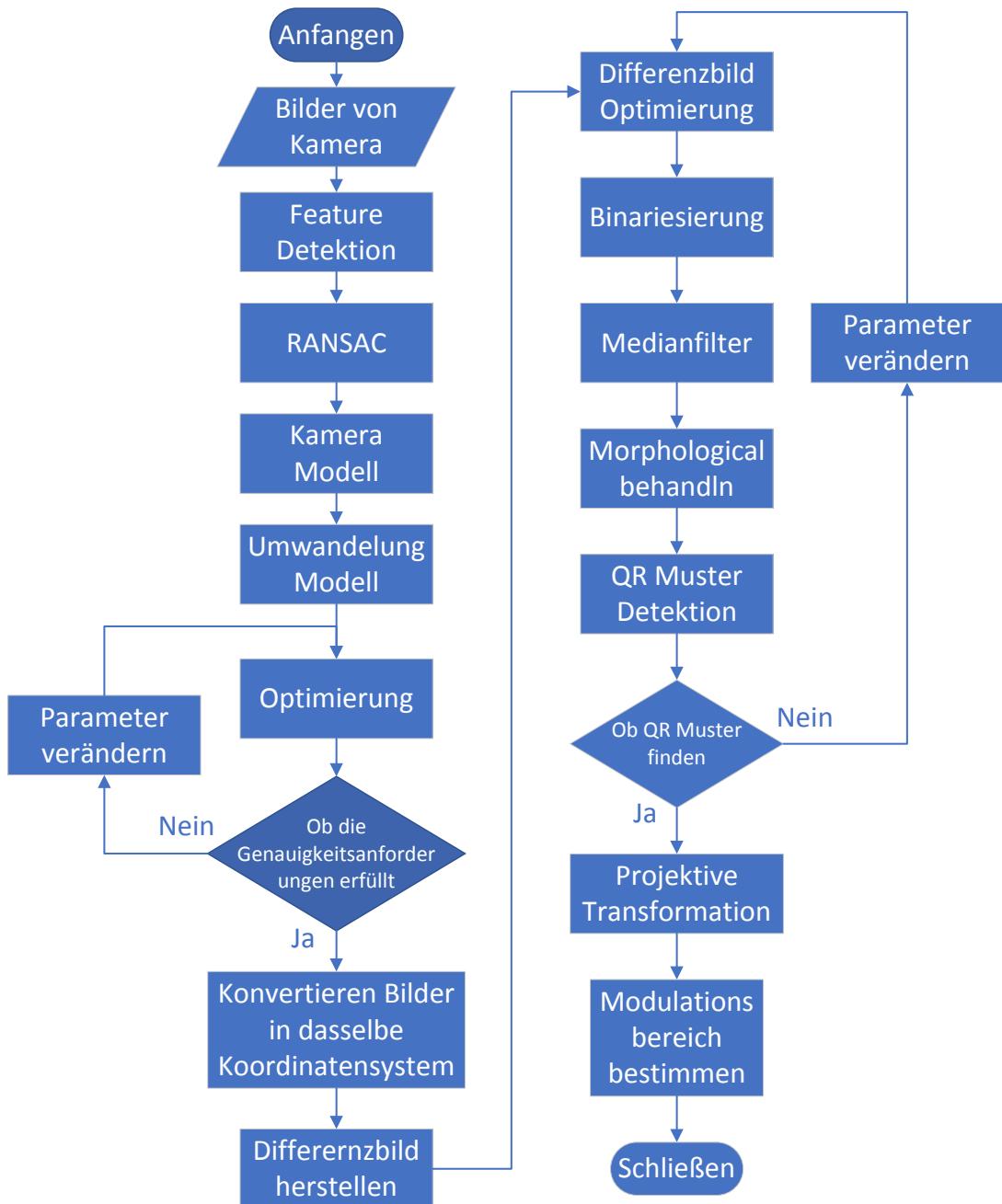


Abbildung 3.2: Flussdiagramm der Methode.

3.2 Bildregistrierung

Annehmen eine Szene, wenn eine Rhein Fotos mit einer Handheld-Kamera aufnehmen, kommt es aufgrund von Handbewegungen zu einer leichten Verschiebung zwischen den beiden benachbarten Fotos. Wenn Sie diese Fotos abziehen, um Differenzbild zu erhalten, wird der Ergebnisse sehr schlecht sein. Um diese Problem zu lösen, wird hier Bildregistrierung eingeführt. Ein Flussdiagramm der Bildregistrierung wird in Abbildung 4.2 gezeigt.

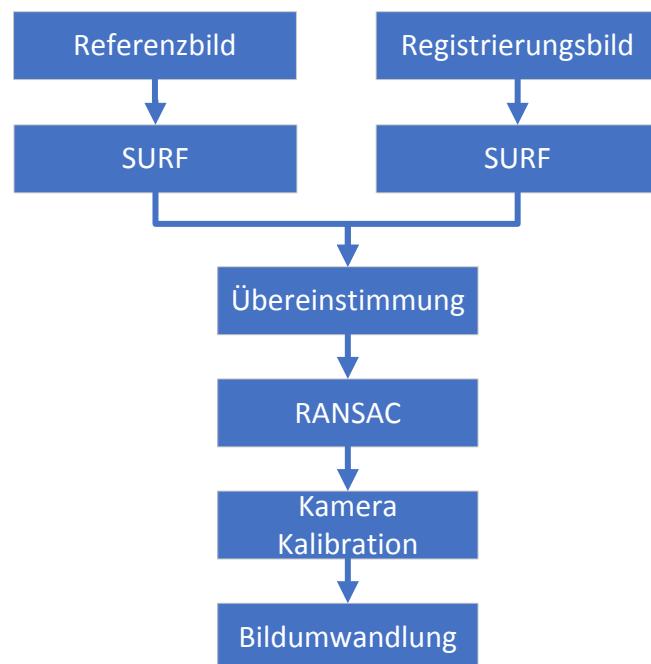


Abbildung 3.3: Flussdiagramm der Bildregistrierung.

3.2.1 SURF

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

Merkmalserkennungen sind wie z.b. Kante Detektion, Ecke Detektion, Tropfen Detektion und so weiter.

Hier in dieser Arbeit lass das SURF [5] nutzen. Das ist ein patentierter lokaler Merkmal-Detektor und Deskriptor und kann für Aufgaben wie Objekterkennung, Bildregistrierung, Klassifizierung oder 3D-Rekonstruktion verwendet werden. [5] Merkmalserkennungen ist eine verbesserte Version von SIFT Merkmalserkennungen, die Haar-Wavelet verwendet, um die Gradientenoperation in der SIFT-Methode anzunähern und gleichzeitig verwendet eine Integralgraph-Technik für schnelle Berechnungen. Die Faltung bezieht sich nur auf das vorherige Bild, mit Erhöhung der Größe des Bildkerns können das Heruntertaktung-Verfahren realisiert werden. Die Geschwindigkeit von SURF ist 3-7 mal die von SIFT mit der in den meisten Fällen entspricht Leistung von SIFT. Daher wurde es 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 die Funktion $f(z, y)$ angenommen, dass die Hesse-Matrix H setzt sich aus Funktionen und partiellen Ableitungen zusammen:

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

Diskriminante der H-Matrix läuft:

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

Der Wert der Diskriminante ist der Eigenwert der H-Matrix. Durch dessen positiven und negativen 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. Nutzen eine Zweite-Order Gaussian Function als Filter. Die zweiten Partielle Ableitungen können durch Faltung zwischen bestimmten Kernen berechnet werden. Dadurch können die Werte der drei Matrix-elemente der H-Matrix auch berechnet werden, nämlich die H-Matrix berechnet:

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

Hier $L_{xx}(\mathbf{x}, \sigma)$ bedeutet die Faltung der zweiter Gaussian Ableitung $G(\sigma)$ mit dem Bild I in Punkt $\mathbf{x}(x,y)$, ähnlich 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 zu feststellen. Zur einfacheren Anwendung schlägt Herbert Bay[5] vor, L mit einer Approximation ersetzen. Um den Fehler zwischen dem genauen Wert und der Approximation auszugleichen, kann die H-Matrix-Diskriminante wie folgt ausgedrückt werden:

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

- **Erstellen Maßstab Raum**

Der Maßstabsraum $L(\mathbf{x}, \sigma)$ des Bildes ist die Darstellung dieses Bildes bei unterschiedlichen Auflösungen(Skalierung). Im Bereich der Computer Vision wird der Maßstabsraum 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, und das Bild muss in der Größe angepasst werden. Daher hat diese Berechnungsmethode eine große Kosten in Berechnung. Im Vergleich dazu ist es in SURF durch die Erhöhung der Größe des Bildkerns. Diese ist ein Unterschied zwischen dem SIFT Algorithmus und dem SURF Algorithmus bei der Verwendung des Pyramidenprinzips. Der Algorithmus ermöglicht, dass mehrere Bilder des Maßstabsraums gleichzeitig verarbeitet werden, ohne dass das Bild unterabgetastet wird, wodurch die Leistung des Algorithmus verbessert wird. Das linke Bild in Abbildung 4.3 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 4.3 behält das ursprüngliche Bild unverändert und ändert nur die Filtergröße.

- **Präzise Lokalisierung von Feature-Punkten**

Vergleichen die Größe jedes Pixel, die von der hessischen Matrix verarbeitet wird, mit die 26 Punkten in seiner drei Dimensionen Raum, wie in Abbildung 4.4 zeignt. Wenn es das Maxi-

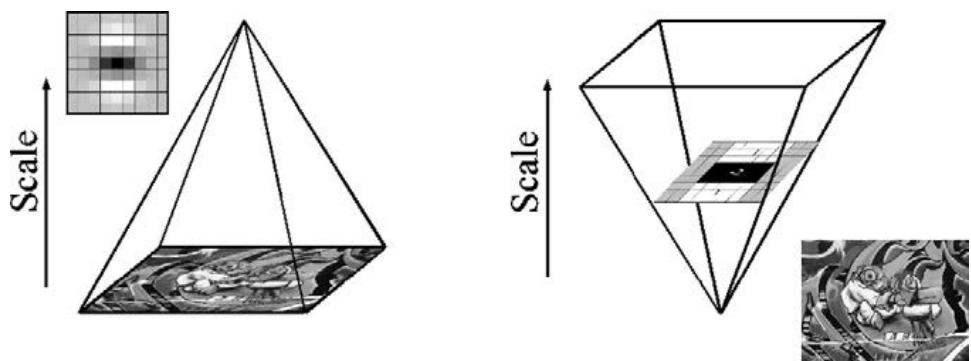


Abbildung 3.4: Scale space.

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

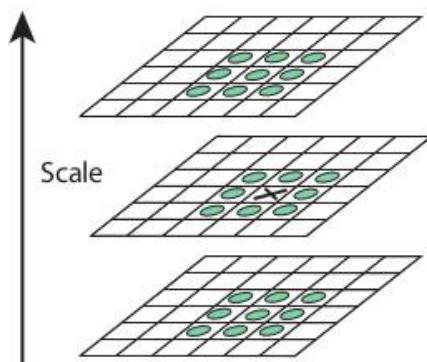


Abbildung 3.5: Extreme Wert Erkennung.

• Hauptrichtungsermittlung

SIFT wählt die Hauptrichtung des Merkmalspunkts unter Verwendung des Gradientenhistogramms im Merkmalspunktbereich aus. Die Richtung, in der der Bin-Wert des Histogramms der größte und oder 80% maximale Bin-Wert überschreitet, wird als Hauptrichtung des Merkmalspunkts genommen. Dagegen beim SURF wird das Gradientenhistogramm nicht statistiken, sondern das Harr-Wavelet-Eigenschaft im Merkmalspunktbereich wird statistisch analysiert. Das heißt, im Bereich der Merkmalspunkt (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($\pi/3$) werden gezählt. Die Größe des Haar Wavelets stellt als $4s$, so dass für jeden Sektor einen Wert bekommt. Dann wird 60-Grad-Sektor in einem bestimmten Intervall gedreht, schließlich lassen die Richtung des Sektors mit Maximalwert als Hauptrichtung des

Merkmalspunkts nehmen. Ein schematisches Diagramm des Prozesses ist wie folgt in Abbildung 4.5.

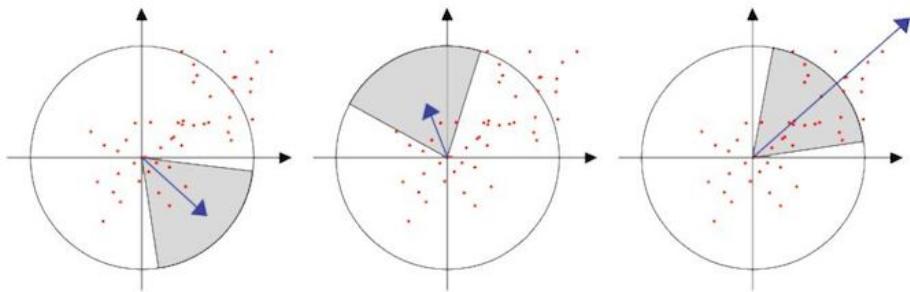


Abbildung 3.6: Dominante Orientierung Feststellen.

- **Merkmalspunkt Deskriptor Generierung**

SURF nehmen eine quadratische Rahmen um den Merkmalspunkt. Die Seite der Rahmen ist $20s$ (s ist die Skala, bei der der Merkmalspunkt erkannt wird). Die Richtung des Rahmens ist natürlich die Hauptrichtung, die in vorliegendem Schritt erfasst wird. Die Rahmen wird dann in 16 Unterbereiche unterteilt, von denen jeder die Haar-Wavelet-Merkmale der horizontalen und vertikalen Richtungen von 25 Pixeln berechnen. Hier die horizontalen und vertikalen Richtungen sind 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 4.6 zeigt dieses Prozesse.

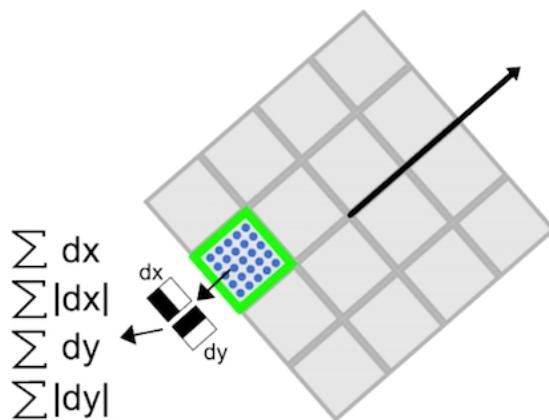


Abbildung 3.7: Merkmalspunkt Deskriptor.

Auf diese Weise hat jeder kleine Bereich 4 Werte, so dass jeder Merkmalspunkt ein $16 \times 4 = 64$ dimensionaler Vektor verfügt, der halb so klein wie Sift(128 Dimension) ist, deswegen den Anpassungsprozess beim Merkmalanpassungsprozess stark beschleunigt. Die folgende Abbil-

dung 4.7 zeigt den Merkmalspunkt, den wir durch den SURF-Algorithmus erhalten haben.

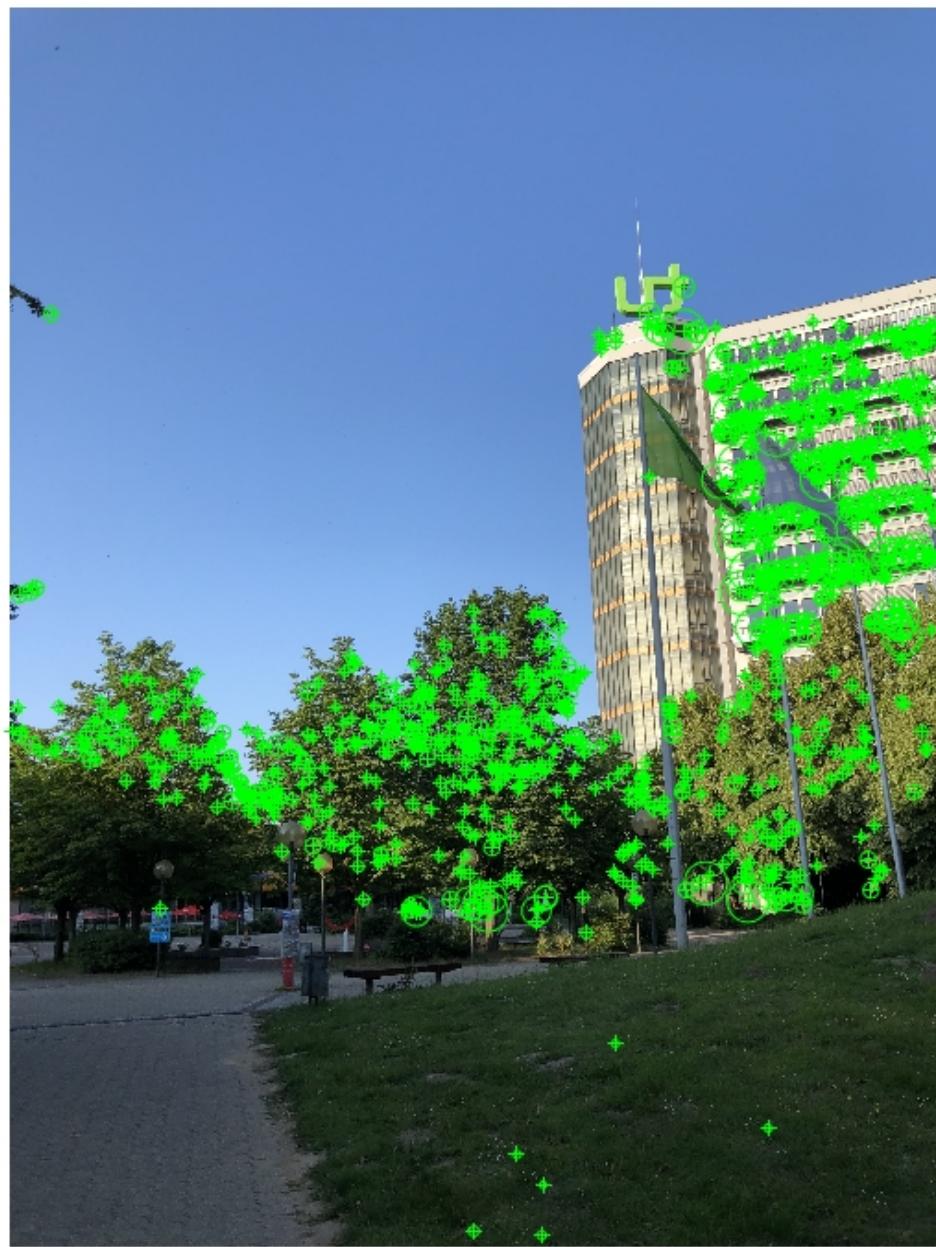


Abbildung 3.8: SURF Merkmal.

3.2.2 RANSAC

Nach SURF Merkmalserkennung wird die Merkmal von zwei benachbarten Bildern bekommen. Diese Merkmale werden dann mit extrahiert und abgeglichen, um entsprechende Punkte in den benachbarten zwei Bildern zu erhalten. Leider darin durch diese Operation gibt es

immer noch viele fehlerhafte zusammenpassendes Paar. Deswegen wird hier RANSAC eingeführt, um die falschen Punkt zu beseitigen.

RANSAC Algorithmus, der von Fischler und Bolles [6] vorgeschlagene im Jahr 1981, 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 kleinsten Quadranten, die von der Computer Vision Community aus der Statistik-Literatur übernommen wurden, wurde RANSAC aus der Computer-Vision-Community entwickelt.

Ein einfaches Beispiel ist in der Abbildung 4.7 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 verwenden ,um diese Linie zu finden, wie auf der linken Seite gezeigt, kann es leider nicht richtig finden, da die Methode der kleinsten Quadrate von alle Datenpunkte beeinflusst wird. Dagegen mit RANSAC kann das Modell nur von der inlierer Punkte berechnet werden und die Ergebnisse wie auf der rechten Seite zeigt.

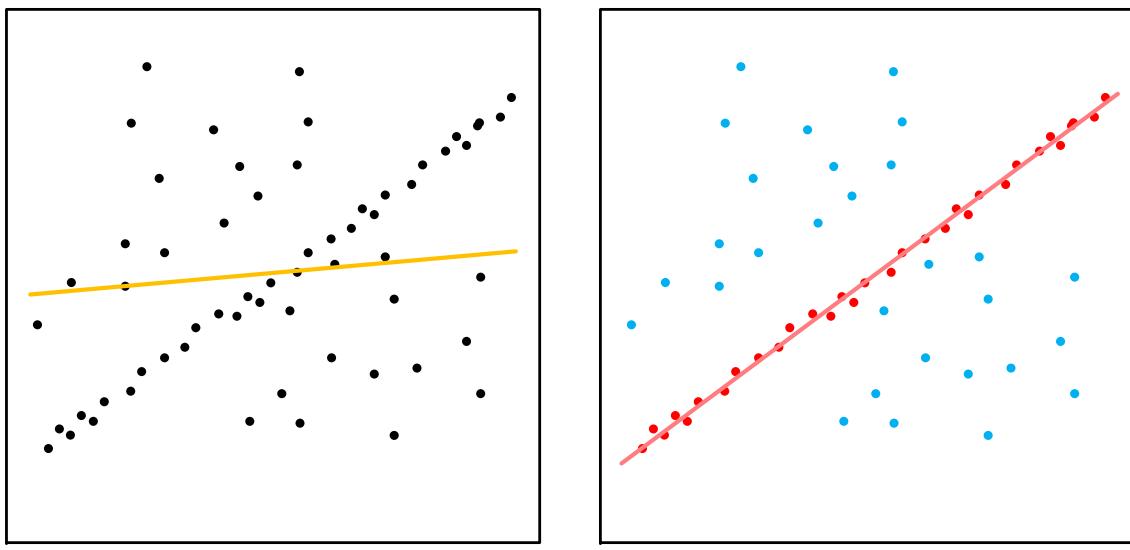


Abbildung 3.9: Linien Detektion.

RANSAC ist ein Wiederholungsprobennahme Verfahren, das durch die minimalen Anzahl von Beobachtungspunkten (Datenpunkten) die Kandidatenlösungen generiert. Diese Datenpunkten sind die erforderlich, um die zugrunde liegenden Modellparameter zu schätzen. Darauf haben Fischler und Bolles [6] hingewiesen, zur Erhalten einer anfängliche Lösung und Beschneidung der Ausreißern RANSAC Verfahren baraucht nicht so viele Daten, sondern

verwendet die kleinste mögliche Menge und fährt fort, diese Menge mit einer konsistenten Datenpunkten zu vergrößern.

Der grundlegende Algorithmus ist wie folgt zusammengefasst:

- Zufällig wählen die Mindestanzahl der Punkten aus, die erforderlich sind, zum Bestimmen der Modellparameter.
- Lösen die Parameter des Modells.
- Bestimmen wie viele Punkte aus der Menge aller Punkte mit einer vordefinierten Toleranz ϵ übereinstimmen
- Wenn der 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 wieder.
- Ansonsten wiederholen die Schritte 1 bis 4 (maximal N-mal).

N bedeutet die Anzahl der Iterationen. Es wird hoch genug gewählt, um die Wahrscheinlichkeit p (normalerweise auf 0,99 gesetzt) sicherzustellen, dass mindestens eine der Gruppen von Stichproben keinen Ausreißer enthält. Dann die Wahrscheinlichkeit, dass bei N Mal Iterationen mit erforderlich minimalen Anzahl Punkte (hier m annahmen) mindestens ein Ausreißer mit ausgewählt wird, läuft:

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

Hier u stellen die Wahrscheinlichkeit dar, dass jeder ausgewählte Datenpunkt ein Inlier ist. Dagegen $v = 1 - u$ heißt die Wahrscheinlichkeit, dass jeder ausgewählte Datenpunkt ein Ausreißer ist. Durch einige Gleichheitsumwandlung können die Anzahl der Iterationen ausgedrückt werden als:

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

Abbildung 4.8 zeigt die passende Punkt durch Merkmalübereinstimmung mit **Speeded Up Robust Features (SURF)** Detektion. Es ist ersichtlich, dass es viele fehlerhafte Kombinationen gibt. Dagegen durch die Anwendung von RANSAC kann dieses Problem effektiv lösen und die übereinstimmenden Punkte verfeinern, die Ergebnisse wie in Abbildung 4.9 zeigt.



Abbildung 3.10: OhneRANSAC.

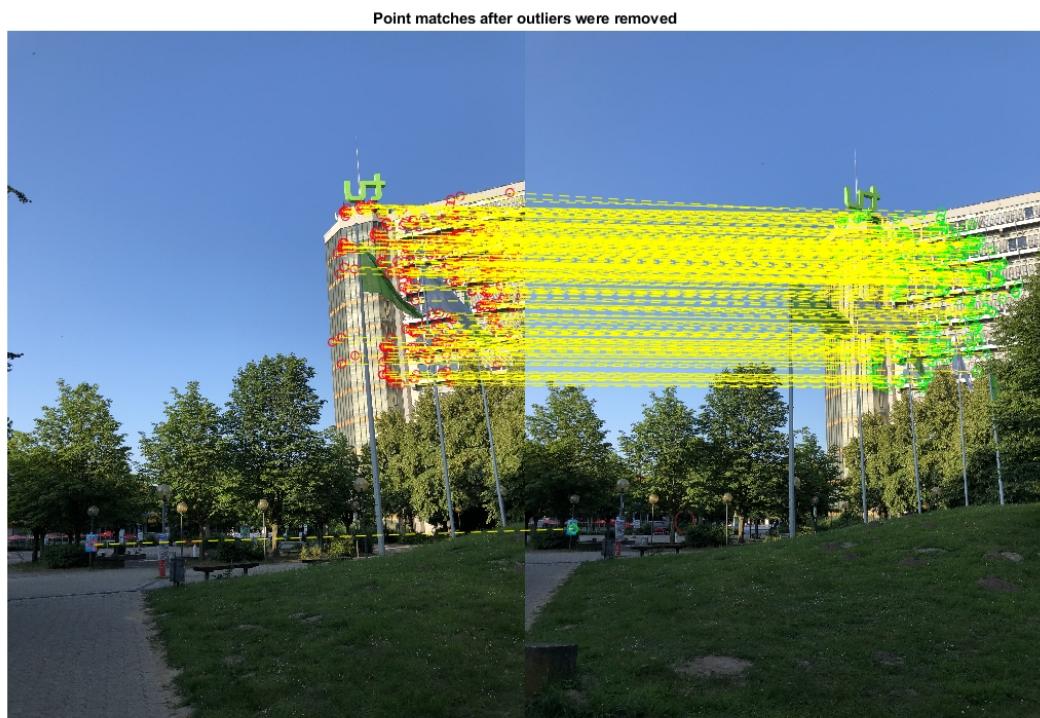


Abbildung 3.11: MitRANSAC.

3.2.3 Bilder Umwandlung

Wie in den vorliegend Abschnitten vorgestellt, mit Verwenden des SURF finden die übereinstimmende Punkte in aufeinanderfolgenden Bildern, anschließen durch RANSAC lassen die Ausreißer verwerfen. Das Ziel dieses Abschnitts besteht darin, das Bild in dasselbe Koordinatensystem zu konvertieren. Der erste Schritt ist, ein Kameramodell zu erstellen und dann darauf die Umwandlungsbeziehung zwischen den entsprechenden Punkten in den zwei Bilder erhalten, schließlich durch den Optimierungsalgorithmus, um die endgültige Transformationsmatrix zu bestimmen. Die verschiedenen Teile 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), f ist die Kamerabrennweite.

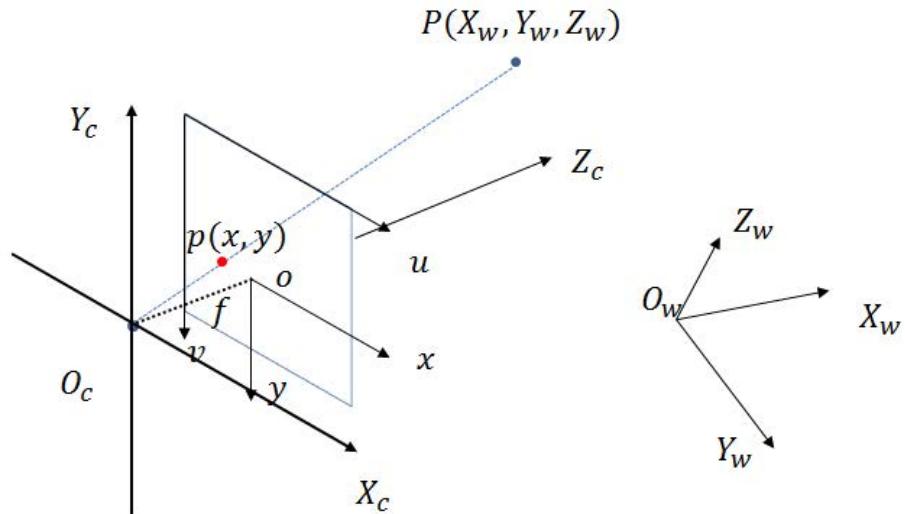


Abbildung 3.12: Modell einer Lochkamera.

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 und nur durch Rotation und Parallelverschiebung. Diese Transformation wird in Abbildung 3.13 gezeigt. R bedeutet Rotationsmatrix und T ist Translationsmatrix.

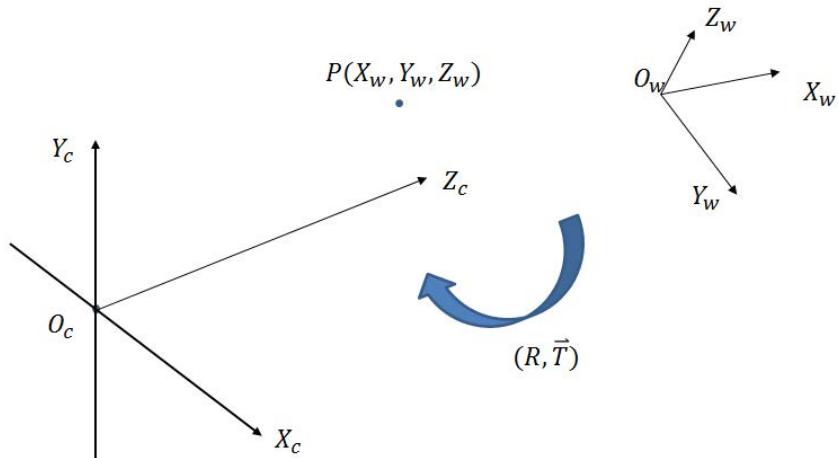


Abbildung 3.13: Transformation vom Weltkoordinatensystem zum Kamerakoordinatensystem.

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

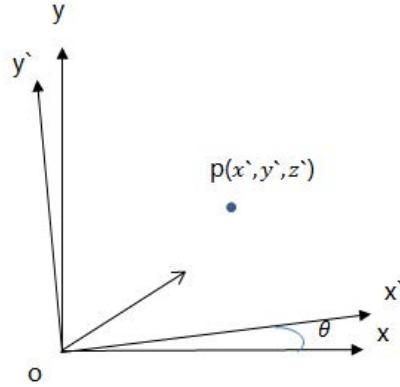


Abbildung 3.14: Rotation um Z-Achse.

Aus dem Bild können wir leicht bekommen:

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

In Matrixform wie folgend 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, um die x-Achse, y-Achse dreht sich um φ und ω Grad, bekommen:

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

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

Dann können die Rotationsmatrix erhalten werden:

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

Kombinieren das Obige, können die Koordinaten von Punkt P im Kamerakoordinatensystem erhalten.

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

Im homogenen Koordinatensystem darstellt:

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

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

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

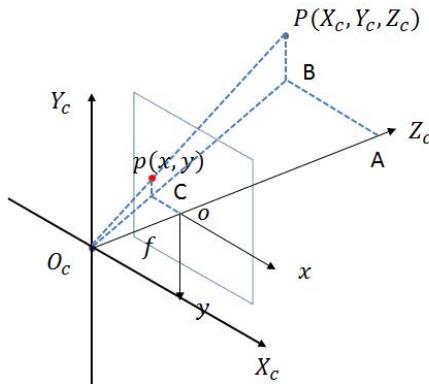


Abbildung 3.15: Transformation vom Kamerakoordinatensystem zum Bild Koordinatensystem.

Es gibt zwei Paare ähnlichen Dreiecken:

$$\begin{aligned} \triangle ABC &\sim \triangle OCO_c \\ \triangle PBC &\sim \triangle PBO_c \end{aligned} \quad (3.14)$$

Aus ähnlichen Dreiecksbeziehungen können diese Gleichung Verfügbar sein:

$$\frac{AB}{OC} = \frac{AO_c}{OO_c} = \frac{PB}{PC} = \frac{X_c}{x} = \frac{Z_c}{f} = \frac{Y_c}{y} \quad (3.15)$$

Durch die Gleichung Transformation können es erhalten:

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

Im homogenen Koordinatensystem darstellt:

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

Zu dieser Zeit ist die Einheit des Projektionspunkts p noch nicht Pixel, sondern mm 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, aber die jeweiligen Ursprünge und Maßeinheiten sind 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 mm, die zu der physikalischen Einheit gehört, und die Einheit des Pixelkoordinatensystems ist Pixel. Wir beschreiben gewöhnlich, dass ein Pixel welche Zeilen und Spalten ist. So ist der Übergang zwischen den beiden wie folgt in Abbildung 3.16.

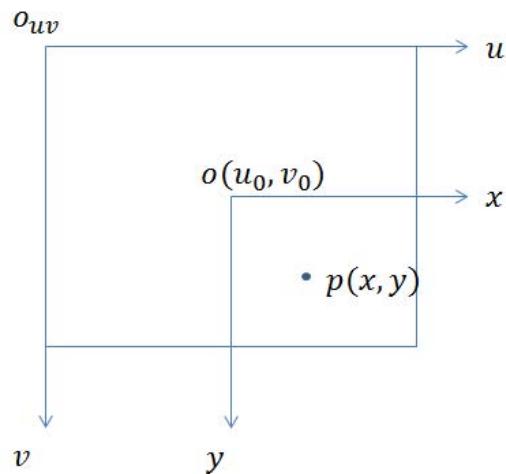


Abbildung 3.16: Konvertierung von Bildkoordinatensystem zu Pixelkoordinatensystem.

Hier dx , dy ist die Größe jedes Pixels 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 darstellt:

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

Es ist erwähnenswert, dass Bildkoordinatensystem ist eine zweidimensionale Ebene(Bildebene), in praktisch die Oberfläche des Kamera-CCD-Sensors ist. Jeder CCD-Sensor hat eine bestimmte Größe und eine bestimmte Auflösung. Diese beide bestimmt die Konvertierungsbeziehung. Geben eine simple Beispiel, eine Größe des CCD-Sensors ist $8\text{ mm} \times 6\text{ mm}$, die Auflösung dafür ist $640 \text{ pixels} \times 480 \text{ pixels}$, dann die Beziehung zwischen mm und Pixel läuft 80 pixel/mm . Lassen Sie die physikalische Größe jedes Pixels des CCD-Sensors $d_x \times d_y$ sein, entspricht läuft $d_x = d_y = \frac{1}{80} \text{ mm}$.

Dann durch die Umwandlung der obigen vier Koordinatensysteme kann ein Punkt vom Weltkoordinatensystem zum Pixelkoordinatensystem erhalten.

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

Die erste Matrix der rechten Gleichung ist die allgemein bekannt interne Referenz der Kamera. Dagegen ist die zweite Matrix die externe Referenz der Kamera. Beide Parameter der Kamera durch Zhang Zhengyou [7] Kalibrierung erhalten werden. Einige typisch Kamera Parameter von Manufaktur liegt in Tabellen 3.1.

Tabelle 3.1: Parameter des Kameras im Vergleich.

Parameter	Google Pixel	Google Pixel2	Iphone 10
Sensor Größe "	1/2.3	1/2.6	1/3
Bild Auflösung <i>pixels</i>	4048 × 3036	4032 × 3024	4032 × 3024
Pixel Größe μm	1.544	1.4	1.22
Brennweite <i>mm</i>	4.67	4.47	3.99
Formatfaktor 35mm	5.55	6.04	7.02

Aus der obigen Formel wenn die internen und externen Parameter der Kamera bekannt sind, ist nämlich die Projektionsmatrix bekannt, und zu diesem Zeitpunkt können die entsprechenden Bildkoordinaten für jeden beliebigen räumlichen Punkt erhalten werden. Dagegen wenn die Position $m(u, v)$ in Bildkoordinate bekannt ist, und auch die Parameter innerhalb und außerhalb der Kamera bereits bekannt sind, kann die entsprechenden Punkt in Wert Koordinate nicht eindeutig bestimmt werden. Die Grund dafür ist, die Z_c Information während des Projektionsprozesses eliminiert wird.

Umwandlung Modelle

Als nächstes machen die Umwandlungsbeziehung zwischen den entsprechenden Punkten in den beiden Bilder. Zuerst wird in diese Arbeit eine vereinfachte Situation betrachten, d.h. nur mit Rotations Einfluss. Abbildung 3.17 zeigt das 3D-Rotationsbewegungsmodell der Kamera. Darin 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 ist der interne Parameter der Kamera, wie zuvor definiert. λ ist der unbekannte Skalierungsfaktor, d.h. unter dem Kameramodell die Quelle der Bildpunktkoordinaten einem Strahl zugeordnet ist.

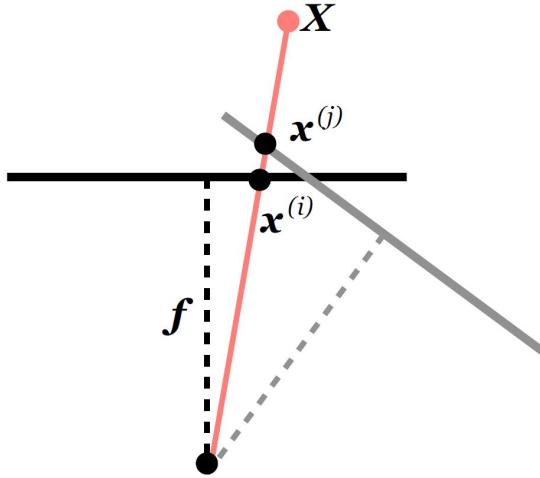


Abbildung 3.17: Rotationsbewegungsmodell.

Herleiten nun die Beziehung zwischen Bildpunkten in einem Framepaar für zwei verschiedene Kameraausrichtungen (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

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

Anordnen diese Gleichungen weiter und ersetzen 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 wird nur die Beziehung zwischen zwei Bildern desselben Videos betrachtet. Lockern diese Einschränkung, indem Frames von einer Kamera, die sich gemäß R dreht, zu einer anderen Kamera, die sich gemäß R' dreht, abbilden. Es gibt eine Hypothese, dass beide Kamerazentren sich im Ursprung befinden. Dann die Warping-Matrix, die Punkte von einer Kamera auf die andere abbildet, definiert werden können:

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

Es wird hier angenommen, dass das erste Bild als Referenz Bild nehmen und der Rotationswinkel 0 ist. Dann die Gleichung kann vereinfacht als:

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

Kombiniert mit Formel 4.23 können es ausgedrückt als:

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

Diese Formel 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. derzeit nicht nur mit Rotations Einfluss, sondern auch Translations Einfluss nehmen. Das Ableitungsverlauf ist im Allgemeinen gleich. Der Hauptunterschied ist Anzahl der Parametern, die von original nur 3 Rotationsparameter zur jetzt 6 Parameter einschließlich 3 Rotationsparameter und 3 Translationsparameter. Das neu Warping-Matrix darstellt wie folgen:

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

Transformation Optimierung

In den letzten beiden Abschnitten wurde die Transformationsmatrix zwischen den beiden Bilder aus dem Kameramodell abgeleitet. Hier in diesen Abschnitt werden die Parameter der Transformationsmatrix berechnet, also bestimmen das Transformationsmatrix. Hier sind die entsprechenden Punkte von zwei benachbarten Bildern x_i, x_j bekannt, und die Umwandlungsbeziehung dazwischen wird als die Formel 4.26 gezeigt. Angesichts dieser Bedingung kann die Berechnung des Transformationsmatrix als ein Optimierungsproblem formulieren, wobei den Fehler J bei der Mittelwertbildung im Quadrat aller Punktkorrespondenzen minimieren wollen:

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

Beachten, dass dies ein nichtlineares Optimierungsproblem ist. Einige nichtlinearer Optimierer könnte verwendet werden, um diese Zielfunktion zu minimieren. Jedoch ist es gefunden, dass Koordinatenabstieg durch direkte objektive Funktionsbewertung schnell konvergiert. Jedes Mal, wenn einen Schritt gemacht wird, bei dem die Zielfunktion J nicht abnimmt, kehren die Schrittrichtung um und verringern die Schrittweite des entsprechenden Parameter. Der Algorithmus endet, sobald die Schrittgröße für alle Parameter unter einen gewünschten Schwellenwert fällt (d.h. Wenn eine Zielgenauigkeit erreicht haben).

Der detaillierte Algorithmus ist wie folgt darstellt. Einige Definitionen werden hier einführen. P_0 heißt der Vektor, die Anfangswerte der Parameter speichert. Anschließen in den Vektor d_p speichert die Schrittgröße jedes Parameters. $temp_{dp}$ ist der Vektor, in den die gewünschten Schwellenwert der Parameter speichert. D bedeutet die Anzahl der Unbekannter Parameter und W das Transformationsmatrix.

1. Zuerst durch Anfangswert P_0 berechnen den anfänglich Fehler J_0 .
 2. Verändern einen Parameter in einer entsprechenden Schrittrichtung mit entsprechenden Schrittgröße.
 3. Berechnen den neuen Fehler J_{new} .
 4. Vergleichen die beiden Fehler, ob der neu Fehler J_{new} kleiner ist. Wenn Ja bestimmt, behalten diesen Wert und ändern des Objekt auf den nächsten Parameter. Kehren zum zweiten Schritt zurück. Dagegen kehren die Schrittrichtung um und verringern die Schrittgröße des Parameter auf ein Drittel des Originals. Ändern des Objekt auf den nächsten Parameter und kehren zum zweiten Schritt zurück.
 5. Wenn alles Schrittgröße in d_p die zuvor festgelegter Schwellenwert erreicht, beenden das Algorithmus und geben den minimalen Fehler J und entsprechenden Transformationsmatrix aus.

Das Flussdiagramm des Algorithmus ist wie in Abbildung 3.19 darstellt.

Dann durch Transformationsmatrix können die Koordinaten des zweiten Bildes in die Koordinaten des ersten Bildes umgewandelt werden. Abbildung 3.18 zeigt diese Verlauf. Dann durch Subtraktion können die Differenzbild erhalten.

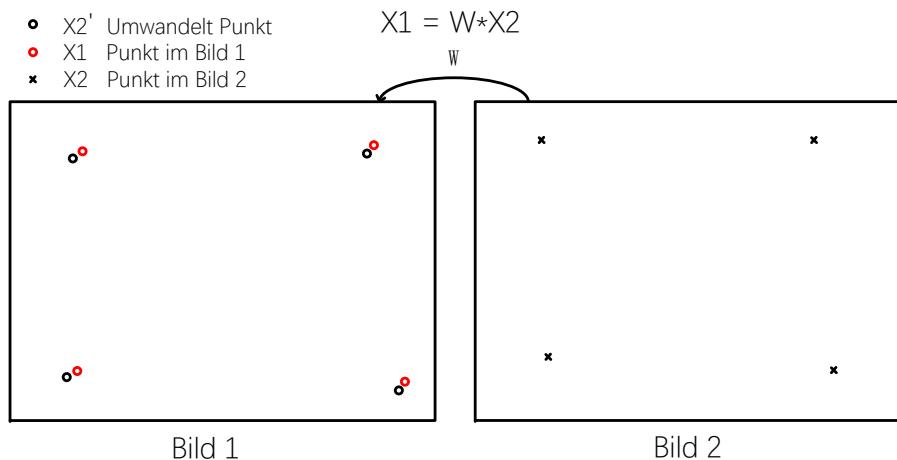


Abbildung 3.18: Transformation in desselben Koordinate.

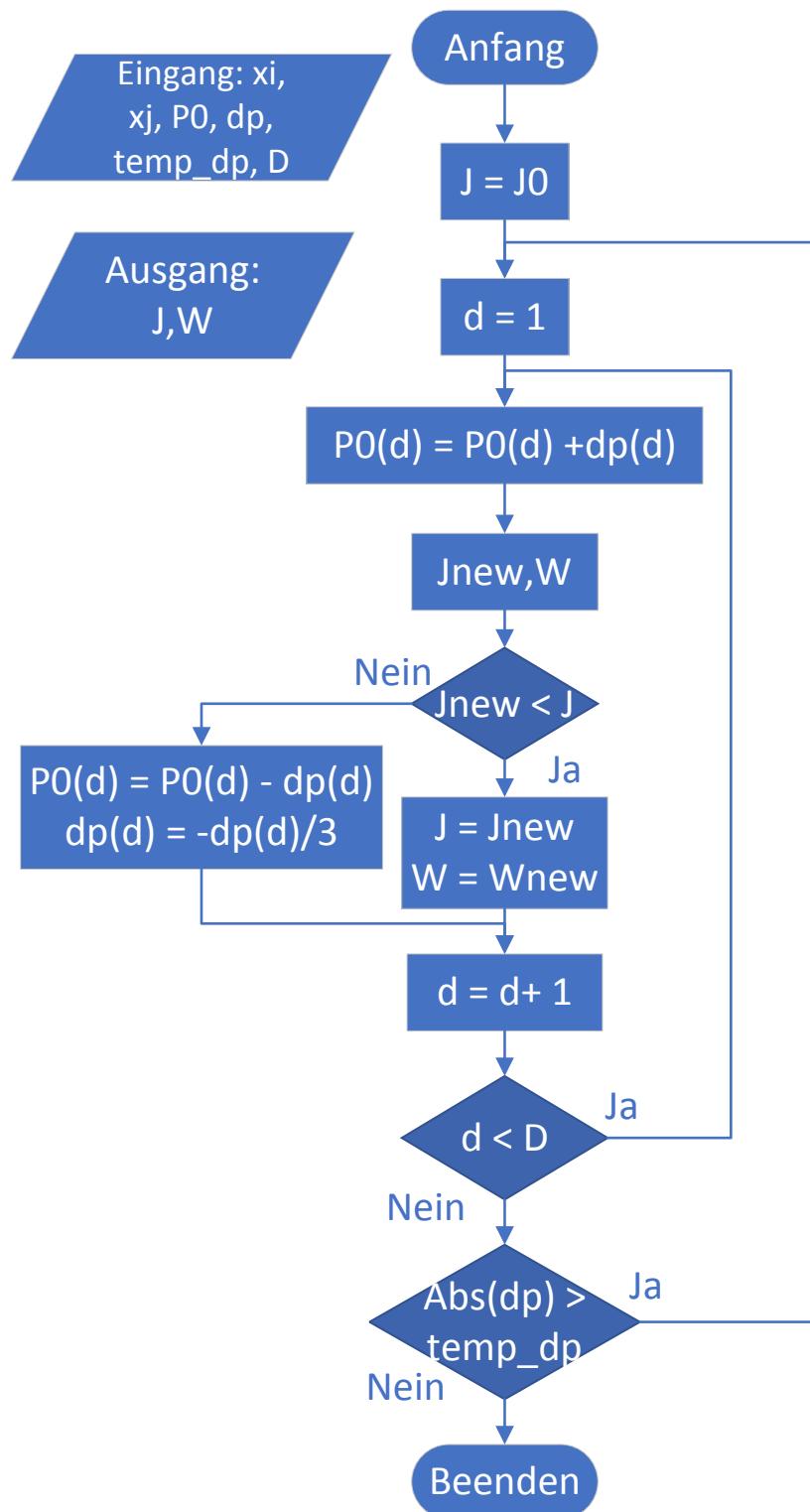


Abbildung 3.19: Flussdiagramm für Optimierung.

3.3 Differenzbild Optimierung

Durch Bildregistration erhalten eine Reihe Bilder von der Kamera, deren Koordinaten in das-selbe Koordinatensystem umgewandelt wurden. Auswählen je zwei Bilder und subtrahieren, um eine Reihe Differenzbildern zu enthalten. Die Ziel in diesem Abschnitt ist von diese Differenzbildern ein detektierendes Bilder herstellen, dadurch die QR Muster Detektion vereinfachen werden können. Es sollte hier beachtet werden, dass aufgrund der Zeitsynchronisation die QR Muster in Differenzbilder einige unerwartete Situationen auftreten sein können. Einige mögliche Formen der Differenzbilder wie folgend gezeigt. Nehmen an, dass es in der vertikalen Richtung ist.

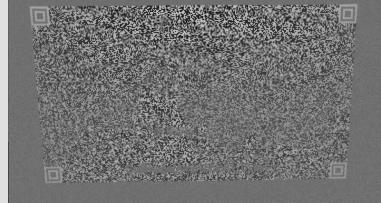
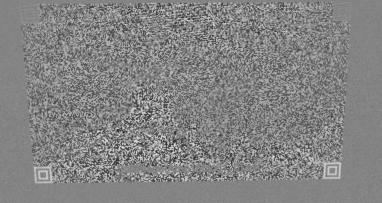
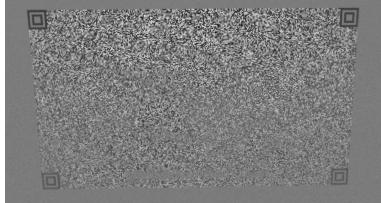
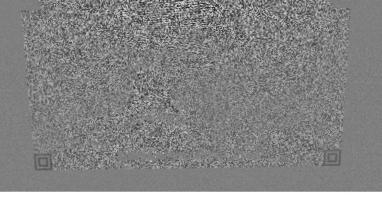
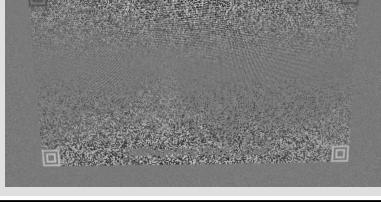
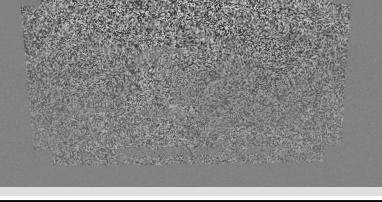
- total Schwarz-Weiß-Schwarz-Weiß-Schwarz Ordnung.
- halb Schwarz-Weiß-Schwarz-Weiß-Schwarz Ordnung, halb nicht gezeigt.
- total Weiß-Schwarz-Weiß-Schwarz-Weiß Ordnung.
- halb Weiß-Schwarz-Weiß-Schwarz-Weiß Ordnung, halb nicht gezeigt.
- halb Schwarz-Weiß-Schwarz-Weiß-Schwarz Ordnung, halb Weiß-Schwarz-Weiß-Schwarz-Weiß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

Tabelle 3.2 zeigt solche Situationen und aufgelistet, ob sie zu folgend Detektion angepasst sind.

Offensichtlich ist die direkte Verwendung dieser Differenzbilder ein kniffliges Problem für die nächste Detektion. Um dieses Problem zu lösen, wurde ein Algorithmus zur Optimierung der Differenzbilder entwickelt.

Der Struktur eines QR Musters zeigt in Formel 4.29. Die äußerste "1" Schicht ist hier eine Trennmuster und die Zentralbereich bedeutet das QR Muster. Aufgrund der Modulations-eigenschaften des DaVid Systems wird nur die Pixelwert des Punkts mit Element "1" nach Modulation offensichtlich verändert. Im Vergleich dazu werden die Pixelwert des Punkts mit Element "0" nur klein verändert. Deswegen durch eine Absolutwertoperation, werden die QR Muster als Schwarz-Weiß-Schwarz-Weiß-Schwarz Ordnung darstellen. Dies ist die erwartende Modellstruktur, die in nächster Schritt operieren werden. Der detailliertes Operieren wird in Abschnitt "QR Muster Detektion" gegeben.

Tabelle 3.2: Die mögliche Differenzbilder.

Ob angepasst zu folgende Detektion?	Differenzbild	Ob angepasst zu folgende 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 werden ein Begriff “Energie” vorstellen. Die auf numerischer Ebene bedeutet den Mittelwert des quadrierten Pixels des Bildes. Es ist bekannt sein, dass aufgrund der Subtraktion die Pixelwerte in das Gebiet, welches rund um die Modulationsbereich legt, gegeneinander aufgehoben werden. Deswegen in dieses Gebiet die Pixelwerte werden nur von Rausch beeinflusst und sehr klein sein werden. Im Vergleich dazu beträgt die Pixelwerte im Modulationsbereich ein relativ größer Wert $\pm 2A$, hier A beutet die Modulationsamplitude des

Systems. Überlegen die Einfluss von Zeitsynchronisation, indem die Pixelwerte in einem Teil der Modulationsbereich werden aufgrund Synchronisation auch ausgeglichen, nämlich die „Energie“ aufgehoben werden. Dann nehmen eine Regelmäßigkeit, die Größe der „Energie“ hängt hauptsächlich von die Pixelwerte im Modulationsbereich. Je größer die Pixelwerte sind, je höher der „Energie“ des Bildes ist, und je klarer das QR Muster werden. Gemäß dieser Regel berechnen die „Energie“ jedes Differenzbildes und anordnen in absteigender Reihenfolge. Um die folgende Detektion zu vereinfachen, fügen die ersten paar Bilder hinzu und erhalten einen zu detektierende Bilder. Die Praxis hat bewiesen, dass es ausreicht, die ersten drei Bilder zu nehmen.

Die Formel der „Energie“ wird in Gleichung 4.30 darstellt. Hier m, n repräsentiert die Anzahl der Zeilen und Spalten der Matrix. $diff(i, j)$ bedeutet die Pixelwert des Punkts in der m -ten Zeile die n -te Spalte der Matrix.

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

Als folgend legt die detailliert Schritt dieses Algorithmus-

1. Wähle zwei beliebige Bilder aus und subtrahiere, um Differenzbild zu erhalten.
2. Nehmen eine Absolutoperation für jedes Differenzbild und berechnen dessen Energie.
3. Sortiere in absteigender Reihenfolge und nehmen die erste 3 Differenzbildern.
4. Addieren diese 3 Differenzbildern, um eine detektierend Bild zu erhalten.

Das Flussdiagramm wird in Abbildung 3.20 gezeigt.

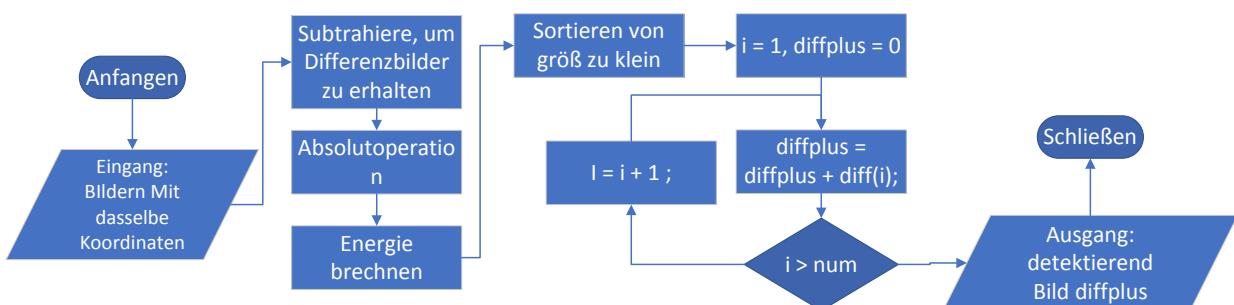


Abbildung 3.20: Differenzbild Flussigdiagramm.

Eine Beispiel für detektierendes Bild wird in Abbildung 3.21 gezeigt. Aus der Abbildung ist ersichtlich, dass der Modulationsbereich besser dargestellt ist und an den vier Eck des Bereichs ein eindeutiges QR Muster vorhanden ist. Als nächstes werden die Bildverarbeitung einführen.



Abbildung 3.21: Ein detektierendes Bild.

3.4 Bildverarbeitung

Durch die Differenzbild Optimierung, wird ein detektierendes Bild enthältet. Es ist noch ein Graustufenbild und muss noch einige Bildverarbeitung nehmen. Dadurch können kleine Punkte und Lücken, die durch Rauschen und Fehler verursacht werden, entfernt werden, um die nachfolgende Detektion zu erleichtern. Der detailliert Inhalt der Bildverarbeitung wurde in der anderen Methode eingeführt, hier ist nur eine kurze Beschreibung der verwendeten Funktionen in Matlab.

Bild Binarisierung

Für die Schwellenwertbildung und das Erstellen eines Binärbildes wurde eine Funktion namens "imbinarize" verwenden. Diese Funktion erhält das Bild und verwendet ein anpassungsfähige Schwellwert, um das Schwarz-Weiß-Bild zurückzugeben. Das ist genug für QR-Pattern Detktion, weil es nur dunkle und helle Module enthält, die binär 1 bzw. 0 sind.

Medianfilter

Der Grund für das Median-Filtern ist, dass manchmal beim Prozess Binarisierung aus einem Bild die Muster wie Salz- und Pfeffergeräusche erzeugt werden können. Um diesen Fehler

zu vermeiden, ist Median Filterung eine leistungsfähig Methode. Das Skript ist so einfach wie `medfilt2(img)`. Zur Verbesserung der Ergebnisse kann natürlich die verschiedenen Fenstergrößen für die Matlab-Funktion verwenden.

Morphologie

Mit öffnenden und schließenden Filtern können die Lücken zwischen Blöcken und die kleinen Punkte von Rausch stark reduziert werden, was das resultierende binär Bild zu einer guten Schätzung des QR-Pattern macht.

3.5 QR Musters Detektion

Nach Bildverarbeitung wird nun die QR Musters Detektion ausgeführt. Die Ziel einer QR Musters Detektion ist die Zentrum des Musters im Bild zu lokalisieren und dadurch die Bild zu rekonstruieren. Abbildung 3.22 zeigt eine geometrische Struktur des QR Musters. An jedem Ecke der Modularisationsbereiche gibt es eine solche QR Muster.

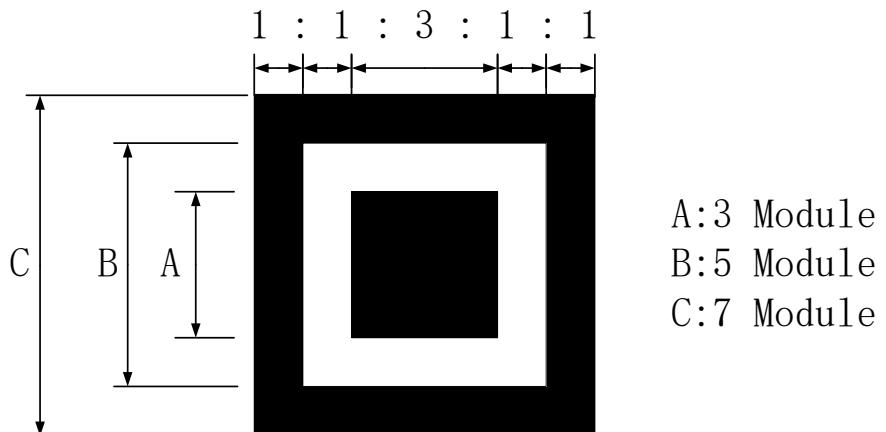


Abbildung 3.22: QR Pattern.

Aus der geometrischen Sicht kann jedes Muster als drei konzentrische Quadrate betrachtet werden und besteht aus einer schwarzen (dunklen) 7×7 Modulen, einer weißen (hellen) 5×5 Modulen und schließlich einer dunklen 3×3 Modulen. Von Kenntnissen der Geometrie können bekannt sein, in jeder Richtung das Breiteverhältnis der alternativen Schwarz- und Weißmodul in einem Muster eine Beziehung 1 : 1 : 3 : 1 : 1 beträgt, wie es in Abbildung 3.23 zeigt. Diese wichtige Eigenschaft hilft uns, die Lokalität der QR Muster zu finden. In der Praxis rund um die Muster gibt es noch ein Trennmuster, das ein Funktionsmuster mit aller weißen (hellen)

Module(Breit ein Modul). Es spielt eine Rolle als eine Grenze zwischen den Mustern und dem Datenbereich, um die Verwechslungen zwischen Muster und Daten zu vermeiden.

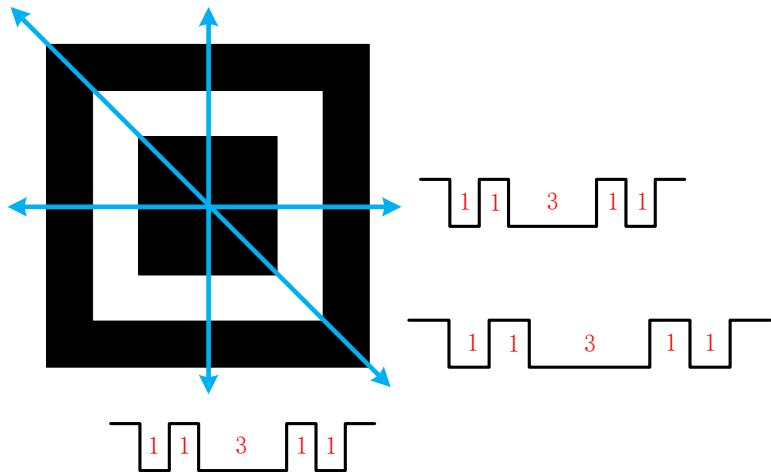


Abbildung 3.23: QR Pattern Ratio.

Als nächstes werden die detaillierten Schritte der Detektion eingeführt und es wird mit Funktion "detectFIPin Matlab implementiert.

Schritt 1:

Zuerst überlegen die große Berechnungsaufwand für die Analysen des ganzen Bilds, teilen einige kleine Bereiche auf, die QR Muster enthalten können.

Schritt 2:

Scannen jede Zeile dieses kleinen Bereiches und speichern die Länge der Schwarz und Weiß Module in eine fünf Element Vektor. Die Länge der Module heißt die Anzahl aufeinanderfolgende Pixel in einer Zeile mit die gleiche Farbe. Speicherreihenfolge in diesem Vektor ist laut 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, nehmen die fünf Elemente und ein Urteil machen, ob die Beziehung der Zählungen nahe genug an den $1 : 1 : 3 : 1 : 1$ Verhältnissen ist. Wenn die Bedingung erfüllt ist, gehen zum nächsten Schritt. Dagegen verschieben den Vektor um zwei nach links und werfen die ersten und zweiten Elemente des

Vektors weg. AnschliSeßen gehen zurück zu Schritt 2, um weiter Scannen und Zählen zu beginnen.

Schritt 4:

Verarbeiten die Elemente vom Vektor, um das ungefähre horizontale Zentrum zu erhalten. Mache eine Kreuzprüfung an diesem Punkt, welche besteht aus den Schritten 2 und 3, der Unterschied dazwischen wird der horizontale Scan durch einen vertikalen Scan ersetzt. Anschließen machen ein Urteil, ob eine vertikale Zentrum gefunden ist. Wenn Ja bestimmt, machen eine Kreuz-Kreuzprüfung mit horizontalem Scan, um die Ergebnis zu optimieren. Dies wird hauptsächlich benötigt, um die reale horizontale Mitte des Musters in extremer Schräglage Fällen zu lokalisieren. Speichern der potenzielle Zentrum, danach leeren die Elemente des Vektors und wieder zu zweiten Schritt, um einen neuen Scan machen. Ansonsten verschieben den Vektor um zwei nach links und werfen die ersten und zweiten Elemente des Vektors weg. Gehen zurück zu Schritt 2, um weiter Scannen und Zählen zu beginnen.

Schritt 5:

Verarbeiten die Ausgabe des vorherigen Schritts, falls es nicht nur eine potentiell Muster Zentrum gefunden, nutzen einen "SelectBestPattern", um die beste zu auswählen. Es sollte angemerkt werden, dass wenn die mögliche Muster nach dem Ende der Erkundung nicht gefunden wird, ein spezielle Signal zurückgegeben wird und die System zu Schritt Differenzbild Optimierung zurückkehren wird. Die Operation besteht darin, der ursprünglichen Bild, die aus drei Differenzbild besteht, ein mehre Differenzbild hinzuzufügen.

Schritt 6:

Durch die gefunden Muster Zentren, machen eine projective Transformation, un die Ecke des Bildes zu bestimmen.

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

Einige Beispiel für QR Muster Detektion zeigt in Abbildung:

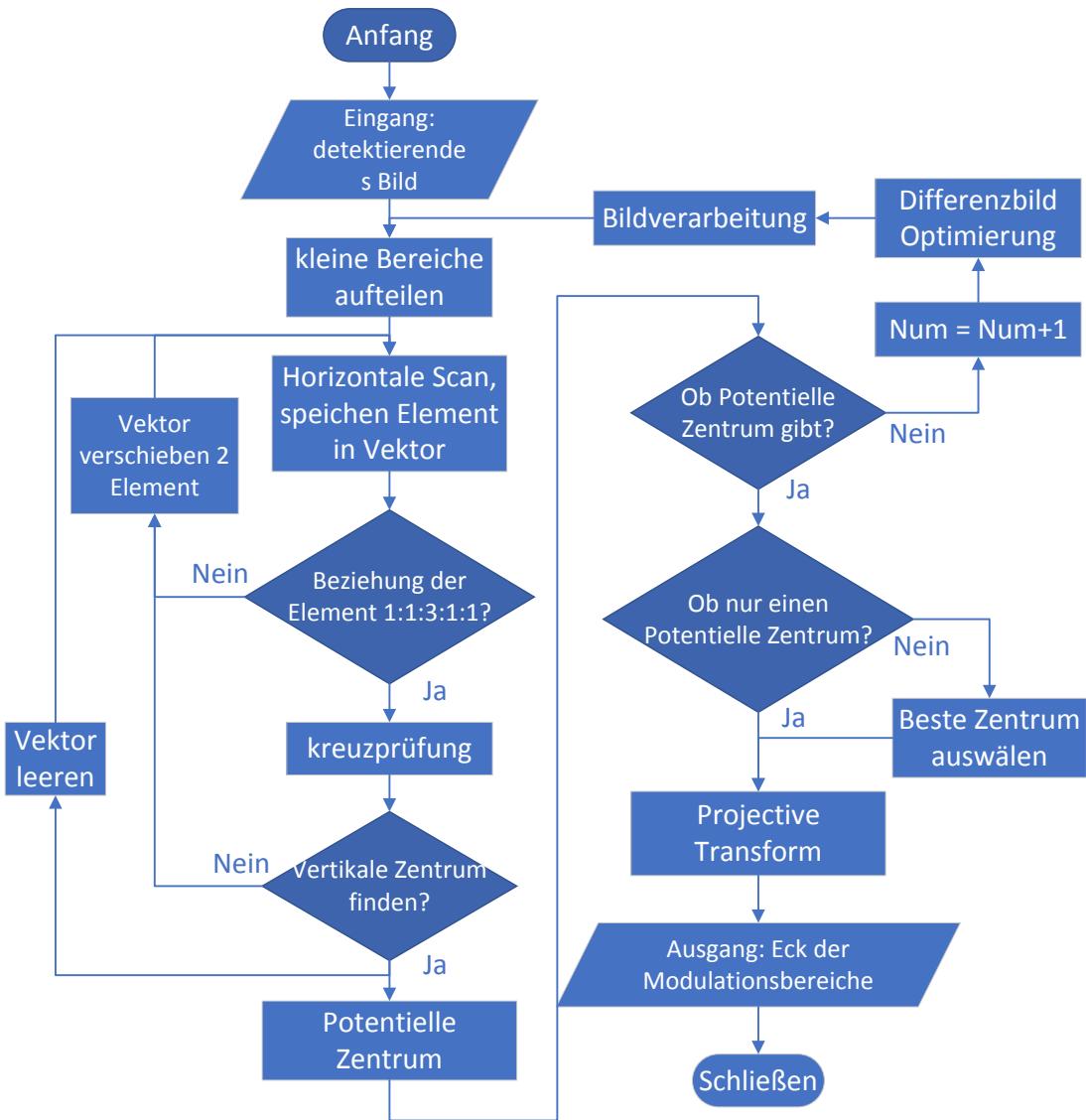


Abbildung 3.24: Flussdiagramm der QR Muster Detektion.

4 Zweite Methode

In diesem Kapitel werden die erste Erfahrung beschrieben. Zuerst läuft die Vorstellung des DaVid-Systems. Die Systemmodell und Arbeitsprinzip des Systems werden in anschließenden Abschnitt erläutert. Schließlich folgt die mögliche Anwendungsgebiete des Systems.

Nicht vergessen, dass Überschriften nicht aufeinander folgen dürfen...

4.1 Allgemeine Struktur

4.2 Binariesierung

Grundlegende globale Schwelle Methode

Wenn die Histogrammspitzen und -täler des Bildes offensichtlich sind und Doppelpeaks aufweisen, ist die Wirkung dieser Methode besser. Es basiert auf der visuellen Überprüfung des Histogramms und der Schwellenwert wird durch eine iterative Methode erhalten. Die grundlegende Algorithmus ist wie folgt:

1. Wählen eine Parameter t und einen anfänglichen Schwellenwert T_0 aus, wobei der Durchschnitt der maximalen Grauwerte I_{max} und minimalen Grauwerte I_{min} verwendet wird. $T_0 = (I_{max} + I_{min})/2$
2. Segmentieren das Bild mit dem Schwellenwert T_0 . Dann das Bild besteht aus zwei Teilen: G_1 besteht aus den Pixeln mit deren Grauwert größer als T_0 und dagegen G_2 deren Grauwert kleiner oder gleich als T_0 .
3. Berechnen den durchschnittlichen Grauwert aller Pixeln in u_1 und u_2 und den neuen Schwellenwert $T_1 = (u_1 + u_2)/2$.
4. Falls $|T_0 - T_1| < t$, dann nehmen T_1 als optimalen Schwellenwert. Andernfalls weisen T_1 zu T_0 und wiederholen die Schritte 2 ~ 4, bis der optimale Schwellenwert erhalten ist.

Grundlegende adaptive Schwellenwert Binarisierungsmethode

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

Ein Verfahren zur Verarbeitung besteht darin, das Bild weiter in Unterbilder zu unterteilen, um unterschiedliche 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, die Position im Bild, also solche Schwellen sind adaptiv.

Ein Verfahren zum Unterteilen von Unterbildern wird für das Bild übernommen. Hier werden drei Arten von Unterteilungsverfahren ausgewählt, und Unterbilder mit einer Größe von $32 \times 32, 4 \times 4, 16 \times 16$ Pixeln werden jeweils geteilt, und die durchschnittliche Graustufe der Unterbilder wird als Schwellenwert für die Binarisierung ausgewählt.

OTSU adaptive Schwellen Methode

OTSU[8], auch bekannt als die maximale Interklassenvarianz, wurde 1979 vom japanischen Gelehrten Otsu vorgeschlagen und ist eine adaptive Schwellenwertbestimmungsmethode, die auch als Otsu bekannt ist.

Die Grundidee des Otsu-Algorithmus ist: Verwenden das Histogramm des Bildes, gemäß der Varianz zwischen dem Vordergrund und dem Hintergrund, um den optimalen Schwellenwert dynamisch zu bestimmen. Setze die Anzahl der Pixeln in einem Bilden ist N , die Graustufe ist $L(0, 1, \dots, L - 1)$. Die Anzahl der Pixeln mit dem Grauwert i ist n_i , dann die Wahrscheinlichkeit von i läuft $P_i = \frac{n_i}{N}$. Für das Bild stellt das T die Segmentierungsschwellenwert zwischen Vordergrund und des Hintergrund dar. Vordergrund entspricht die Grauwerte von 0 zu $T - 1$, dagegen Hintergrund die Grauwerte von T zu $L - 1$.

Die Wahrscheinlichkeit der Vordergrundgebiet W_0 und der durchschnittliche Grauwert U_0 sind

$$w_0 = \sum_{i=0}^{T-1} p_i, \quad u_0 = \sum_{i=0}^{T-1} ip_i / w_0 \quad (4.1)$$

dagegen die Wahrscheinlichkeit der Hintergrundpunkte W_1 und der durchschnittliche Grauwert U_1 sind

$$w_1 = \sum_{i=T}^{L-1} p_i = 1 - w_0, \quad u_1 = \sum_{i=T}^{L-1} ip_i / w_1 \quad (4.2)$$

Dann das gesamte durchschnittliche Grauwert des Bildes ist

$$U = w_0 u_0 + w_1 u_1 \quad (4.3)$$

Die Infra-Klassen-Varianz ist definiert als

$$\sigma^2 = w_0(u_0 - U)^2 + w_1(u_1 - U)^2 = w_0 w_1 (u_0 - u_1)^2 \quad (4.4)$$

Nehmen Schwellenwert T von 0 zu $L - 1$. Wenn σ^2 Maximum ist, wählen der entsprechende T als der optimale Schwellenwert aus.

Die Ostu-Methode verwendet Grauwert Histogramm zur Bestimmen des Schwellenwerts. Es ist eine automatisch non-parametrische Schwellenauswahlmethode. Diese Methode ist einfach zu berechnen, und wird nicht von der Kontrast- und Helligkeitsänderung unter bestimmten Bedingungen beeinflusst und kann das Objekt zufriedenstellend vom Hintergrundbereich trennen.

4.3 Morphologie

Nach der Binarisierung kann das Bild zahlreiche Unvollkommenheiten erhalten. Insbesondere ist das Binärbild, die durch einfache Schwellenwert erteilt werden, durch Rauschen und Textur 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. Diese Eigenschaft macht es besonders geeignet für die Verarbeitung von Binärbildern. Natürlich ist Mengenoperation gleichmäßig gültig für Graustufenbilder. Hier in dieser Arbeit wird es nur in binäre Bilder benutzt. Es erstellt ein neues Binärbild, bei dem das Pixel einen nicht-Null Wert hat, nur wenn der Operation an dieser Position im Eingabebild erfolgreich ist.

Die Eingabedaten für die mathematischen morphologischen Operationen sind zwei Bilder: das zu bearbeitende Bild A und ein Strukturelement B. B ist normalerweise eine kleine Pixelmatrix mit jeweils einem Wert von Null oder Eins. Einige Eigenschaft des Strukturelements werden wie folgend gelegt.

- Die Matrixdimensionen geben die Größe des Strukturelements an.

- Das Muster, das aus Einsen und Nullen bestehen, gibt die Form des Strukturelements an.
- Ein Ursprung des Strukturelements ist üblicherweise eines dessen Pixel, es verlassen sich auch außerhalb des Strukturelements liegen.

Eine übliche Praxis besteht darin, dass ungerade Dimensionen der Pixelmatrix stellen und den Ursprung als das Zentrum der Matrix definieren. Einige grundlegende Strukturelement legt als folgen:

0	1	0
1	1	1
0	1	0

3*3 Kreuzform

1	1	1
1	1	1
1	1	1

3*3 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

5*5 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

5*5 Quadrat

Abbildung 4.1: Einige grundlegende Strukturelement.

Als folgend werden zuerst die zwei grundlegende morphologischen Operatoren Erosion und Dilatation beschrieben. Anschließend vorstellen die darauf basierend Operationen, wie als Öffnung und Schließung bekannt sind.

Dilatation

Die Dilatation Operation bewirkt, dass das Objekt nach größer wächst. Das wächst Ausmaß hängt von der Art und Form des Strukturelements ab. Die Formel eine Dilatation wird darstellt als:

$$A \oplus B = \{z \mid \widehat{(B)}_z \cap A \neq \emptyset\} \quad (4.5)$$

Hier $\widehat{(B)}_z$ heißt Strukturelement B um seinen Ursprung reflektiert und um z verschoben. Die entsprechende Dilatation für das Bild A mit B ist die Menge aller Verschiebungen z, die \widehat{B} und

A mindestens ein gemeinsames Element habe. Das Ergebnis der Dilatation läuft, die Pixeln um den Grenze des Objekts hinzufügen. Außerdem die Dilatation Operation wird im wesentlichen verwendet um die Löcher (fehlende Pixel) in einem kontinuierlichen Objekt zu füllen. Es beeinflusst die Intensität an diesem Bereich und kann als ein Unschärfe Effekt beobachtet werden, nämlich eine räumlichen Tiefpassfilter, die beim linearen Filtern des Bildes verwendet werden. Abbildung 4.2 zeigt eine typische Dilatation Operation.

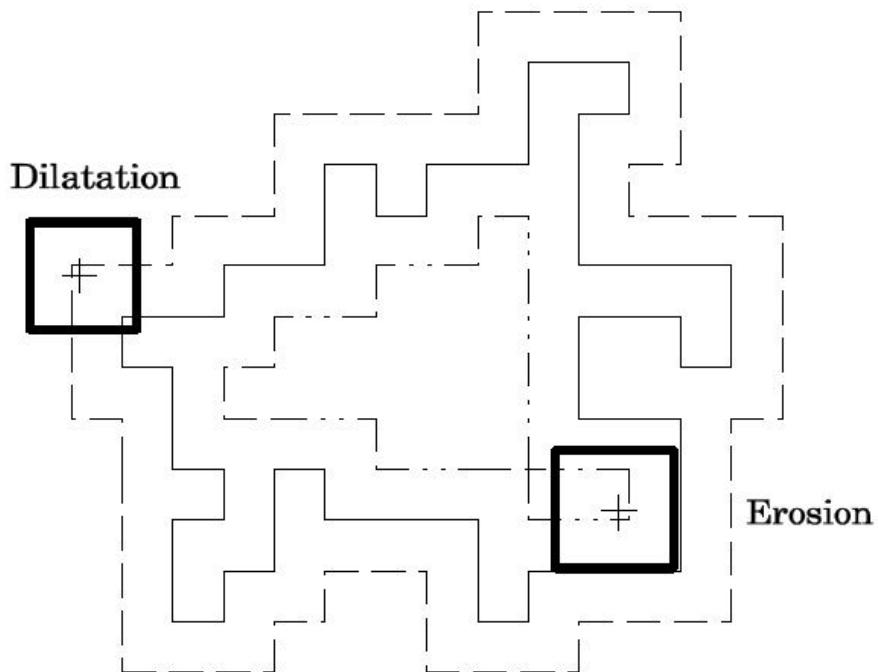


Abbildung 4.2: Dilatation und Erosion.

Erosion

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

$$A \ominus B = \{z \mid (B)_z \subseteq A\} \quad (4.6)$$

Hier die Erosion des Bilds A ist die Menge aller Punkte z mit derart, dass das Strukturelement B von verschieben wird, zu eine Teilmenge des Bildes A gehört. Diese Operation führt zu einem Verlust von Grenzpixeln des Objekts.

Die Erosion Operation entfernt solche Strukturen, die eine kleinere Größe als das strukturierende Element haben. Dann kann es verwendet wird, um die verrauschte "Verbindung zwischen zwei Objekten zu entfernen. Da die unerwünschten Pixel entfernen werden, ist der Effekt als ein Schärfen des Objekts entsprechen. Als folgend in Abbildung 4.2 zeigt eine typische Erosion Operation.

Öffnung

Die Öffnung Operation eines Bildes ist eine kombinatorische Operation von Erosion und Dilatation, d.h. zuerst nehmen eine Erosion Operation, danach eine Dilatation Operation. Praktisch werden Bild A durch beide Operationen mit dem gleichen Strukturelement B ausgeführt. Die Formel einer Öffnung Operation ist definiert als

$$A \circ B = (A \ominus B) \oplus B \quad (4.7)$$

Die Grenze des geöffneten Objekts sind die Punkte, dass Strukturelement B die äußersten Punkte der Grenze von Objekt erreicht, wenn B innerhalb dieser Grenze entlangfahren. Feine strukturierte Details, kleiner als das Strukturelement werden demnach beim Öffnung Operation entfernt, dünne Verbindungen zwischen größeren Teilen aufgelöst. Eine Öffnung Operation wird in Abbildung 4.3 gezeigt.

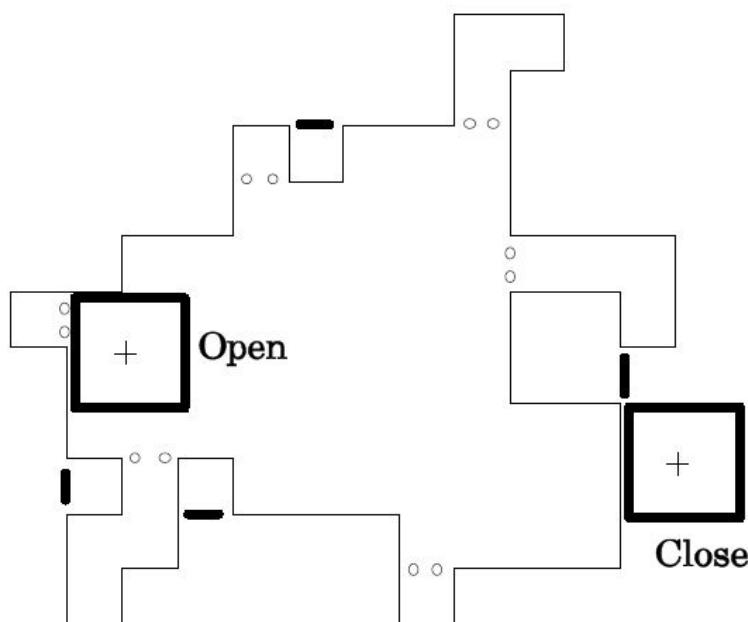


Abbildung 4.3: Einige grundlegende Strukturelement.

Schließung

Gleichfalls mit Öffnung Operation ist die Schließung Operation auch eine kombinatorische Operation von Erosion und Dilatation. Der Unterschied dazwischen legt der Reihenfolge der Operation, d.h. hier zuerst eine Dilatation Operation, danach eine Erosion Operation mit dem gleichen Strukturelement. Das Schließung eines Bildes A durch das Strukturelement B ist definiert als

$$A \bullet B = (A \oplus B) \ominus B \quad (4.8)$$

Die Grenze des geschlossenen Objekts sind die Punkte, dass Strukturelement B, die die äußersten Punkte der Grenze von Objekt erreichen, wenn B außerhalb dieser Grenze entlangfahren. Kleinere Risse, Lücken und feine Details werden dagegen aufgefüllt und mit den großen Teilen zusammengeschlossen. Abbildung 4.3 zeigt eine Vorgang der Schließung Operation.

Öffnung und Schließung Operation besitzen die folgenden 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.
- Bezeichnet ein Bild als B-Öffnet, wenn es bezüglich gleicher Öffnung Operation unverändert bleibt.
- Bezeichnet ein Bild als B-Schließt, wenn es bezüglich gleicher Schließung Operation unverändert bleibt.

4.4 Canny detection

4.5 Hough detection

5 Auswertung

Text hier zwischen. Referenz auf ein Bild mit cleverref (siehe ??). Und dann noch ein paar Zitierungen [9]–[11].

5.1 Section

Jetzt nur noch schreiben! :)

6 Zusammenfassung

Nicht vergessen, dass Überschriften nicht aufeinander folgen dürfen...

6.1 Section

Anhang A

Erster Anhang

A.1 Section

Jetzt nur noch schreiben! :)

Anhang A Erster Anhang

Abbildungsverzeichnis

2.1 Eine beispielhafte Implementierung des DaVid-Systems	3
2.2 Schematische Darstellung von DaVid System	4
2.3 Blockschaltbild der Signalverarbeitung in zeitlicher differentieller Modulation	5
2.4 Ein Beispiel einer modulierten Bildfolge	7
3.1 Strukturdiagramm	10
3.2 Flussdiagramm der Methode	12
3.3 Flussdiagramm der Bildregistrierung	13
3.4 Scale space	16
3.5 Extreme Wert Erkennung	16
3.6 Dominante Orientierung Feststellen	17
3.7 Merkmalspunkt Deskriptor	17
3.8 SURF Merkmal	18
3.9 Linien Detektion	19
3.10 OhneRANSAC	21
3.11 MitRANSAC	21
3.12 Modell einer Lochkamera	22
3.13 Transformation vom Weltkoordinatensystem zum Kamerakoordinatensystem	23
3.14 Rotation um Z-Achse	24
3.15 Transformation vom Kamerakoordinatensystem zum Bild Koordinatensystem	25
3.16 Konvertierung von Bildkoordinatensystem zu Pixelkoordinatensystem	26
3.17 Rotationsbewegungsmodell	29
3.18 Transformation in desselben Koordinate	31
3.19 Flussdiagramm für Optimierung	32
3.20 Differenzbild Flussigdiagramm	35
3.21 Ein detektierendes Bild	36
3.22 QR Pattern	37
3.23 QR Pattern Ratio	38
3.24 Flussdiagramm der QR Muster Detektion	40
4.1 Einige grundlegende Strukturelement	44
4.2 Dilatation und Erosion	45

Abbildungsverzeichnis

4.3 Einige grundlegende Strukturelement	47
---	----

Tabellenverzeichnis

3.1 Parameter des Kameras im Vergleich	28
3.2 Die mögliche Differenzbilder	34

Quellcodeverzeichnis

Literatur

- [1] 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.
- [2] R. Kays, C. Brauers und J. Klein, „DaViD : Data Transmission Using Video Devices-An Innovative System for Smart Media Applications“, 2016.
- [3] 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.
- [4] 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.
- [5] H. Bay, A. Ess, T. Tuytelaars und L. V. Gool, „SURF:Speeded Up Robust Features“, Bd. 110, 2008, S. 346–359.
- [6] 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>.
- [7] Z. Zhengyou, „A Flexible New Technique for Camera Calibration“, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jg. 22, S. 1330–1334, Dez. 2000.
- [8] 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.
- [9] R. Reinhold und R. Kays, „Improvement of IEEE 802.15.4a IR-UWB for time-critical industrial wireless sensor networks“, in *2013 IFIP Wireless Days (WD)*, Nov. 2013, S. 1–4.
- [10] T. K. Moon, *Error Correction Coding, Mathematical Methods and Algorithms*, 1. Aufl. Hoboken, NJ: Wiley-Interscience, Juni 2005, 800 S.
- [11] *Ieee standard for local and metropolitan area networks, part 15.4: Low-rate wireless personal area networks*, IEEE Std 802.15.4-2011, 2011.

Eidesstattliche Versicherung

Blaubär, Käpt'n Kevin
Name, Vorname

123456
Matr.-Nr.

Ich versichere hiermit an Eides statt, dass ich die vorliegende Bachelorarbeit mit dem Titel
Tasty Kanalmodell für die drahtlose Kommunikation zwischen Gebäuden und Außeninstallationen
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, 24. August 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, 24. August 2018
Ort, Datum

Unterschrift