

Image Stitching

Das SIFT-Verfahren zur Merkmalerkennung

CS1025 Hauptseminar ROAD - Robots and Autonomous Driving

Benedikt Jensen

benedikt.jensen@gni.thm.de

Dozenten:

Jakob Czekansky, M.Sc.

Moritz Schauer, M.Sc.

Institut für Technik und Informatik
Technische Hochschule Mittelhessen, Gießen
6.7.2021

Inhaltsverzeichnis

1	Einleitung	1
1.1	Was ist Image Stitching?	1
1.2	Anwendungsbereiche	1
1.3	Herausforderungen	2
2	Technische Umsetzung	3
2.1	Hauptschritte des Image Stitching	3
2.1.1	Kalibrierung	3
2.1.2	Registrierung	3
2.1.3	Mischung	3
2.2	Image Stitching Techniken	3
2.2.1	Direkte Technik	4
2.2.2	Merkmalsbasierte Technik	4
2.3	Ablauf	4
2.3.1	Bilderfassung	4
2.3.2	Merkmalerkennung und -paarung	4
2.3.3	Homographie	5
2.3.4	Globale Ausrichtung	6
2.3.5	Projizierung	7
2.3.6	Mischung	9
2.3.7	Nahtschneiden	9
2.3.8	Weichzeichnen	9
3	Das SIFT-Verfahren	11
3.1	Schritte	11
3.1.1	Erkennung von Skalenraumextremen	11
3.1.2	Merkmalslokalisierung	11
3.1.3	Orientierungszuweisung	11
3.1.4	Merkmalsdeskriptor	12
4	Anwendung von SIFT zur Objekterkennung	12
4.1	Merkmalspaarung	12
4.2	Effizienzerhöhung	13
5	Vergleich von SIFT und SURF	13
6	Fazit	14
6.1	Zusammenfassung	14
6.2	Ausblick	14
	Abbildungsverzeichnis	I
	Tabellenverzeichnis	II
	Literatur	III

1 Einleitung

Image Stitching kombiniert die Informationen aus mehreren Fotos, welche sich in Teilbereichen überlappen, und fügt diese zu einem zusammenhängenden Bild zusammen. In den vergangenen Jahren wurden viele Algorithmen entwickelt, welche dieses Problem behandeln. Es bleibt allerdings nach wie vor eine Herausforderung dieses Verfahren zu perfektionieren.(vgl. [Kale and Singh \[2015\]](#): 284)

1.1 Was ist Image Stitching?

Image Stitching ist ein Unterbereich der Computervision. 'Image' ist Englisch für Bild und 'Stitching' bedeutet 'Nähen' oder 'Heften. Mehreren Einzelbilder, welche von einem ähnlichen Blickwinkel geschossen wurden, werden aufeinander gelegt und 'zusammengeñäht'. Es entsteht ein hochauflösendes Panoramabild. Das Wort 'Panorama' stammt aus dem Griechischen. 'Pan' bedeutet alles und 'horama' bedeutet Sicht. Ein Panoramabild ist also eine Gesamtansicht.(vgl. [Kale and Singh \[2015\]](#): 284)

Normalerweise ist es mit einer Kamera nur möglich solche Fotos zu schießen, welche sich auf ihren maximalen Sichtwinkel beschränken. Doch anhand der Technik des Image Stitching ist es möglich diese Schranke zu überwinden. Vom Standpunkt der Kamera aus wird eine Reihe von Fotos aufgenommen, jedes um einen gewissen Winkel versetzt zum vorhergehenden. Die so aufgenommenen Fotos werden mithilfe von Image-Stitching-Algorithmen verarbeitet und kombiniert.

1.2 Anwendungsbereiche

Image Stitching wird im Bereich des autonomen Fahrens in Industrie und im Verkehr angewandt. Hier werden Aufnahmen oft von mehreren Kameras gleichzeitig gemacht. Zusammenführen der verschiedenen Kamerawinkel verbessert gegebenenfalls das Ergebnis von Computervision-Algorithmen. Nun können Aufnahme nicht mehr nur noch getrennt betrachtet werden, sondern es entsteht eine ganzheitliche Beobachtung.

Ein weiterer Anwendungsbereich ist die Überwachung von Anbaufeldern. Herkömmliche Methoden der Überwachung der Anbauflächen, ist arbeitsintensiv und zeitaufwändig. Stattdessen können besagte Flächen von unbemannten Luftfahrzeugen(UAV) beobachtet werden. Die so gewonnenen Aufnahmen werden mithilfe von Image Stitching zusammengefügt und ausgewertet.(vgl. [Feng et al. \[2020\]](#): 1)

Dokument-Mosaikierung ist ein Vorgang, bei dem aus einzelnen Aufnahmen eines Dokumentes ein hochauflösendes Gesamtbild geschaffen wird. Ähnliches ist für Mikroskopaufnahmen möglich. In der Bildstabilisierung bei Digitalkameras wird die Distanz zwischen verschiedenen Momentaufnahmen berechnet, dann wird aufgrund dieser Berechnung das Wackeln ausgeglichen um ein unverwackeltes Video zu erhalten. Weitere Anwendungsbereiche sind die Fotomosaikierung für Karten und Satellitenbilder. Auch das bildgebende Verfahren in der Medizin nutzt Image Stitching. Bei der Komprimierung von Videodateien wird die Videomosaikierung angewandt. Desweiteren ist es möglich 360° Panoramas zu erstellen, welche z.B. in der Virtual Reality verwendet werden. Dazu werden mehrere Aufnahmen verknüpft, da eine Kamera immer nur einen begrenzten Aufnahmewinkel hat.

1.3 Herausforderungen

Einerseits gibt es die Herausforderung der Invarianz. Aufnahmen die zusammengefügt werden, haben unter Umständen unterschiedliche Werte von Rotation, Skalierung, Beleuchtung und Unschärfe. Unabhängig davon, sollten zusammenhängende Bilder verknüpft werden können.

Ein weiterer Punkt ist die Performanz. Für viele Anwendungsbereiche des Image Stitchings ist es notwendig, dass Ergebnisse augenblicklich, also in Echtzeit, vorliegen. Das lässt sich leicht verstehen, wenn das Image Stitching z.B. zur Hinderniserkennung im autonomen Fahren eingesetzt wird. Sind die Ergebnisse nicht in Echtzeit bereit, kann es sein, dass es bereits zum Zusammenprall gekommen ist.

Noch eine Herausforderung ist es, Teilaufnahmen im überlappenden Bereich wirklich zur Deckung zu bringen, denn haben Teilbilder keinen identischen Aufnahmewinkel, wodurch das Erkennen von Wiederkehrenden Objekten nicht trivial ist. Dadurch entsteht auch das Problem, dass Bilder sogar bei korrektem Zusammenlegen nicht perfekt zur Deckung kommen.

2 Technische Umsetzung

Um aus einer Reihe von Einzelbildern ein zusammenhängendes Bild zu generieren, müssen erst Operationen auf den Einzelbildern ausgeführt werden, um die notwendigen Anforderungen zu erfüllen. Dann muss eine Beziehung zwischen den Einzelaufnahmen definiert werden und letztendlich muss es ein Zusammenführen der Daten geben. Der Vorgang des Image Stitchings wird im Folgenden in Teilschritte aufgeteilt.

2.1 Hauptschritte des Image Stitching

2.1.1 Kalibrierung

Die Kalibrierung(engl.: image calibration) dient der Korrektur von Abweichungen der verwendeten Kombination aus Linsen im Vergleich zur idealen Linse. Ursache für solche Abweichungen sind z.B. Imperfektionen in einer der Linsen, ungenaue Ausrichtung der Linsen oder auch Lichtreflexe. Diese zeigen sich in der Aufnahme z.B. in Form von radialer, tonnen- oder kissenförmiger Verzerrung. Werden diese nicht korrigiert, führt das zu Ungenauigkeiten im resultierenden Panorama.(vgl. [Kale and Singh \[2015\]](#): 284)

2.1.2 Registrierung

In der Registrierung(engl.: image registration) werden die Einzelbilder übereinander gelegt und abgeglichen. Ein mit einer Kamera aufgenommenes Foto wird für gewöhnlich auf einer Ebene abgebildet. Die Lage dieser Ebenen zueinander wird kalkuliert und gegebenenfalls notwendige Transformationen werden durchgeführt. Ein wichtiger Vorbereitungsschritt dafür ist die Merkmalerkennung. Entscheidend hierbei sind Merkmale, welche sich in mehreren Aufnahmen gleichzeitig finden, also in den Bereichen einer Überlappung.(vgl. [Kale and Singh \[2015\]](#): 284)

2.1.3 Mischung

Die Mischung(engl.: image blending) ist das nahtlose Zusammenfügen der Einzelbilder zu einem Gesamtbild. Idealerweise sollten nach dem Stitching keine sichtbaren Nahtstellen mehr vorhanden sein. Fallspezifisch wird das Gesamtbild entweder auf eine flache, zylindrische oder sphärische Oberfläche projiziert. Aufgrund von variierendem Lichteinfall, Verzerrung und weiteren Unstimmigkeiten entstehen nach dem Überlagern der Bilder meist Nahtstellen. Ziel der Mischung ist es diese Nahtstellen unkenntlich zu machen. Zur Umsetzung der Mischung gibt es verschiedene Ansätze, welche abhängig von der Natur der Fotoaufnahmen anzuwenden sind.(vgl. [Kale and Singh \[2015\]](#): 284)

2.2 Image Stitching Techniken

Die bekannten Techniken des Image Stitching lassen sich größtenteils in zwei Kategorien einteilen: Die direkte Technik und die kennzeichenbasierte Technik. Bei der direkten Technik ist das Ziel die Unterschiedlichkeit der korrelierenden Pixel zu minimieren. Die kennzeichenbasierte Technik hingegen zielt darauf ab, wenige herausstechende Merkmale zu alignieren.(vgl. [Kale and Singh \[2015\]](#): 285)

2.2.1 Direkte Technik

Die direkte Technik vergleicht die Intensität sich überlappender Pixel. Es wird die Summe des Betrags all dieser Abstände gebildet. Diese Summe soll minimiert werden, denn eine kleine Summe ist gleichbedeutend mit geringer Unterschiedlichkeit.(vgl. [Kale and Singh \[2015\]](#): 285) Ein großer Nachteil dieser Technik ist, dass der Algorithmus komplexe Berechnungen durchführt und somit eine lange Laufzeit hat. Jeder Pixel wird mit jedem anderen verglichen. Desweiteren ist er sehr unflexibel, da er stark durch unterschiedliche Sichtbarkeit einzelner Objekte beeinflusst wird. Ein weiterer Nachteil dieser Technik ist, dass sie einiges an Initialisierung benötigt. Menschliche Interaktion ist von Nöten, um z.B. die Orientierung der zu kombinierenden Fotos zueinander zu bestimmen.(vgl. [Maponga et al. \[2017\]](#): 1)

2.2.2 Merkmalsbasierte Technik

In der merkmalsbasierten Technik extrahiert ein Merkmalsdetektor Metadaten aus dem Bild. Dazu gehört unter anderem die geometrische Lage der Pixel im 3D-Raum zueinander. Eine weitverbreitete Technik zur Merkmalerkennung ist die skaleninvariante Merkmalsstransformation(engl.: scale-invariant feature transform, Abk.: SIFT). Weitere bekannte Techniken sind Harris, SIFT, SURF, FAST, PCA-SIFT und ORB.(vgl. [Kale and Singh \[2015\]](#): 285) Etwas später werden wir die SIFT-Technik genauer betrachten.

Erkannte Schlüsselpunkte werden zu Objekten zusammengefügt. Es wird nach korrelierenden Merkmalen in einem Bildpaar gesucht, in zwei Einzelbildern welche sich überlappen. Wird das gleiche Objekt in einer anderen Teilaufnahme erkannt, kann anhand von Informationen zur Orientierung der Aufnahme bestimmt werden, ob es sich tatsächlich um das selbe Objekt handelt. SIFT liefert Ergebnisse mit großer Genauigkeit, jedoch ist ein großer Nachteil, dass der Algorithmus sehr rechenaufwendig ist und sich somit nicht für Echtzeitanwendungen eignet.(vgl. [Maponga et al. \[2017\]](#): 1)

Eine Alternative bietet die beschleunigte robuste Merkmalerkennung(engl. Speeded up robust features, Abk.: SURF). Diese Technik bietet nicht die gleiche Genauigkeit wie SIFT, wird aber ohnehin als robust angesehen und liefert Ergebnisse deutlich schneller.

2.3 Ablauf

2.3.1 Bilderfassung

Als Voraussetzung für das Image Stitching müssen sich dafür eignende Bilder aufgenommen werden. Für gewöhnlich wird hierzu eine Digital- oder Smartphonekamera benutzt.

2.3.2 Merkmalerkennung und -paarung

Als zweiter Schritt werden Merkmale erkannt. Dies wird häufig als der Hauptschritt des Image-Stitchings angesehen. Anstatt das gesamte Bild zu betrachten, werden markante Merkmale herausgepickt. Diese Herangehensweise spart Rechenzeit und ist zudem sehr flexibel. In vielen Anwendungsfällen des Image Stitchings sind Echtzeitberechnungen notwendig. Beispiele dafür sind SLAM(Simultane Positionsbestimmung und Kartierung, engl.: Simultaneous localization and mapping), 3D Rekonstruktion, sowie Video Stabilisierung. Um qualitative Ergebnisse zu erlangen ist es essentiell korrelierende Merkmale eindeutig zu bestimmen. Ein Merkmal des einen Bildes und ein Merkmal in einem

zweiten Bild, welche einander entsprechen, werden zu einem Merkmalspaar zusammengeführt. Eine gute Grundlage dazu bieten Eckpunkte von Objekten. Es werden einige Voraussetzungen an einen Merkmalsdetektor gestellt. Er muss Merkmale unabhängig von Translation, Rotation, Skalierung, Bildrauschen und Unschärfe erkennen. Ansonsten wäre es unmöglich Bilder zu vergleichen welche unterschiedliche Grade an Belichtung haben, aus verschiedenen Winkeln geschossen wurden oder auf andere Weise zwar die gleichen Objekte abbilden, allerdings nicht identisch sind. (vgl. [Kale and Singh \[2015\]](#): 286)

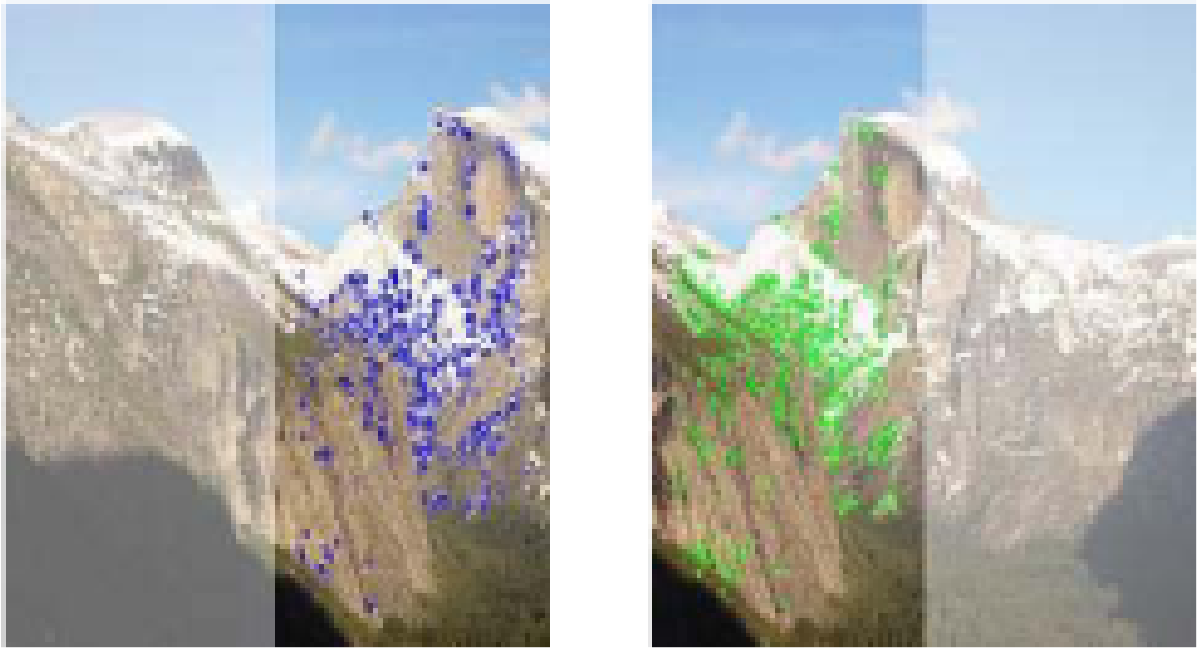


Abbildung 1: Merkmalsmatching

2.3.3 Homographie

Homographie ist die Transformation, anhand welcher sich eine Ebene im Raum auf eine zweite Ebene projizieren lässt. Im Prozess des Image Stitching ist Homographie von Interesse, weil ein mit einer Kamera aufgenommenes Foto sich normalerweise auf einer Ebene im Raum abbildet. Während der Aufnahme eines Panoramabildes ändert sich der Winkel der Kamera und somit auch die Lage der Eben, auf welche sich die Aufnahme abbildet. Folgende Abbildung veranschaulicht diesen Zusammenhang.

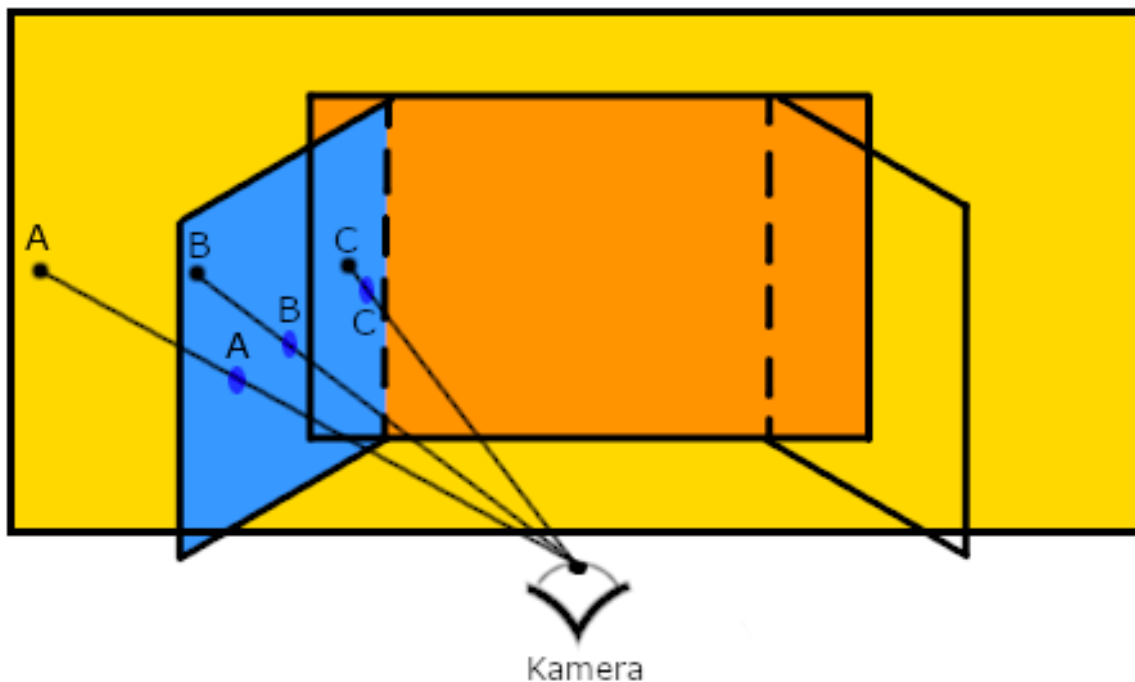


Abbildung 2: Homographie

Ein Foto wird als Basis festgelegt. Die Ebene auf der sich dieses Foto befindet, wird als Zielebene bestimmt. Das Basisfoto wird von der orangenen Fläche repräsentiert. Die Zielebene ist gelb. Die Blaue Fläche ist ein Foto welches mit leicht verändertem Winkel aufgenommen wurde. Die lilanen Punkte werden auf der Zielebene von gleichnamigen schwarzen Punkten repräsentiert. Trifft der projizierte Punkt innerhalb des orangenen Bereichs auf, korrelieren die beiden Punkte auf den verschiedenen Fotos zueinander. Kenntnis um diesen Zusammenhang ermöglicht die Berechnung der Homographie. Eine bekannte Technik. Vier korrelierende Merkmalspunkte aus beiden Fotos werden in Beziehung gestellt und die Lage der Ebenen zueinander wird geschätzt. Da Objekte auf einem Foto nicht tatsächlich auf einer Ebene liegen, reicht eine einzige Berechnung nicht, um ein zuverlässiges Ergebnis zu erhalten. Deshalb wird die Technik RANSAC (Übereinstimmung mit einer zufälligen Stichprobe, engl.: Random Sample Consensus) angewandt. Als Teil dieser Technik wird ein Resampling, also eine Stichprobenwiederholung durchgeführt, womit trotz Ausreißern in den Daten ein robustes Ergebnis erlangt werden kann. (vgl. [Kale and Singh \[2015\]](#): 286)

2.3.4 Globale Ausrichtung

Die ausschlaggebende Technik bezüglich der globalen Ausrichtung ist die Bündelblockausgleichung. Sie ist eine Technik aus dem Bereich der Photogrammetrie, also der Schätzung von Lage und Form eines Objektes mittels einer Fotografien. Mit ihr wird eine 3D-Rekonstruktion der aufgenommenen Szene erstellt. Dies ist notwendig um unerwünschte Effekte wie Parallax und Unschärfe zu vermeiden. Parallax ist der Effekte, der den Schein verursacht, dass naheliegende Objekte sich schneller bewegen als Objekte in der Ferne. (vgl. [Kale and](#)

Singh [2015]: 286)

2.3.5 Projizierung

Nachdem wir jetzt alle vorgegebenen Bilder registriert haben, ist der nächste Schritt deren Kombination zu einem zusammenhängenden Gesamtbild. Das beinhaltet die Entscheidung, auf welche Oberfläche das Bild projiziert wird. Zur Auswahl stehen flach, zylindrisch oder sphärisch. Bei wenigen Vorgabebildern ist ein üblicher Ansatz eines der Bilder als Referenz zu verwenden. Die Ebene in welcher sich dieses Bild befindet, ist dann gleichzeitig die Zielebene, auf welche wir unser Bild projizieren. Bei der Projektion auf eine zylindrische Oberfläche hilft es, sich in das Innere eines Zylinders zu versetzen. Die Innenseite des Zylinders ist die Fläche, auf welche wir projizieren. Es ist anzumerken, dass bei dieser Vorgehensweise ein Sichtwinkel von bis zu 360° in horizontaler Richtung möglich ist. Um den Sichtwinkel weiter zu maximieren, was bei Virtual Reality Anwendungen nützlich ist, kann alternativ auf die Innenseite einer Kugel projiziert werden. Nun sind alle Sichtrichtungen abgedeckt - es kann sogar nach oben und unten geblickt werden. Die Projektion der Ausgangsbilder auf die Zieloberfläche wird durch Mapping der einzelnen Bildpunkte auf Punkte der Zieloberfläche erreicht. Das kann in Vorwärtsrichtung geschehen. Jedem Punkt des Ausgangsbildes werden Koordinaten auf der Zieloberfläche zugewiesen. Jedoch entstehen dabei unter Umständen Lücken, wenn die Punktdichte der Zieloberfläche die der Ausgangsfläche überschreitet. Deswegen wird stattdessen das inverse Mapping angewandt. Jedem Punkt auf der Zieloberfläche werden Koordinaten auf dem Ausgangsbild bzw. den Ausgangsbildern zugewiesen. Da ein Punkt auch zwischen den vorhandenen Pixelkoordinaten liegen kann, wird zwischen anliegenden Pixeln interpoliert. Interpolation ist die Berechnung von unbekannten exakten Werten anhand von bekannten anliegenden Werten. Vereinfacht gesagt würde sich also für die Position zwischen einem weißen und einem schwarzen Punkt die Farbe grau ergeben. (vgl. Kale and Singh [2015]: 286-287)

Vor- und Nachteile der verschiedenen Zielebenen sind folgende. Bei der Projizierung auf eine zylindrische Oberfläche ist es möglich, eine All-Around-Ansicht zu generieren. Der Nebeneffekt ist allerdings, dass Linien die vorher gerade waren sich teilweise krümmen. Bei der Kugelprojizierung ist es sogar möglich die Sicht nach oben abzubilden.

Die Projizierung auf eine Fläche hingegen bietet einen ganz anderen Vorteil. Dieser erschließt sich, wenn man auf den Einsatz von Image Stitching im autonomen Fahren eingeht. Beim autonomen Fahren spielt die Objekterkennung eine große Rolle. Zur effektiven Objekterkennung ist es jedoch häufig notwendig, dass die Aufnahme auf der die Operationen durchgeführt werden, bestimmte Anforderungen erfüllt. Eine dieser Anforderungen ist, dass die Aufnahme nicht verzerrt ist. Somit würde für den Einsatz im autonomen Fahren hauptsächlich die Projizierung auf eine flache Ebene zum Einsatz kommen.

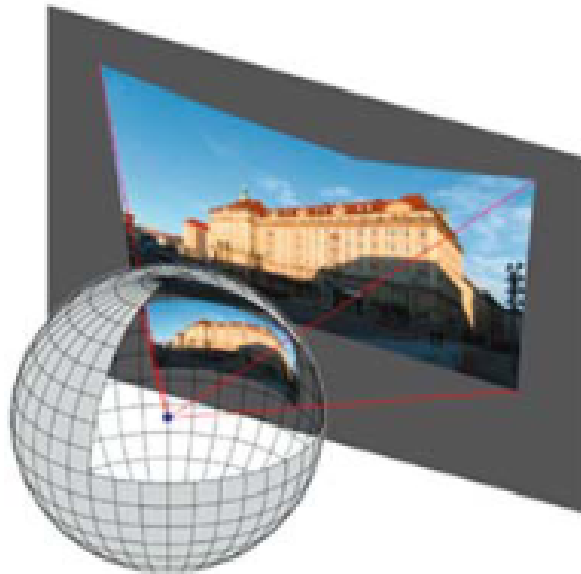


Abbildung 3: Kugelprojektion

Diese Abbildung zeigt die Projizierung einer Aufnahme auf eine kugelförmige Oberfläche. Diese Art der Zieloberfläche ermöglicht es eine All-Around-Ansicht zu generieren.

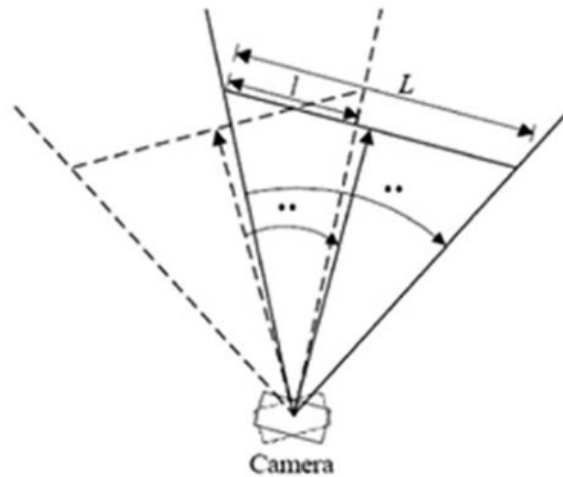


Abbildung 4: Kamerawinkel

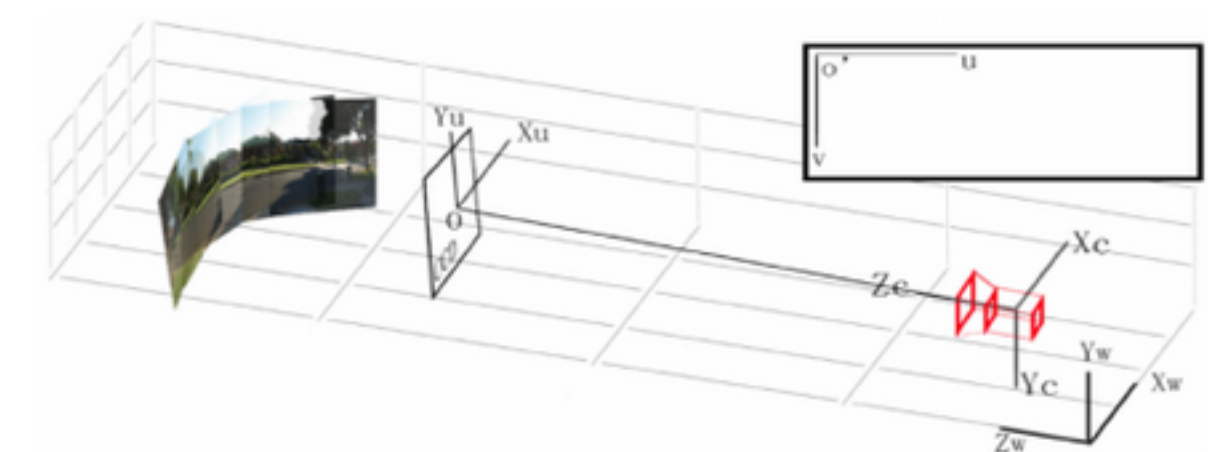


Abbildung 5: Illustrierung

2.3.6 Mischung

Nach der Projizierung ist der nächste Schritt die Mischung. Es ist fast unvermeidbar, dass zwischen den einzelnen Aufnahmen Unterschiede in Belichtung, Schärfe und ähnlichem entstehen. Das Resultat sind sichtbare Nähte im Endergebnis. Diese Nähte und andere Fehler wie variierende Helligkeit oder Farbtintensität unkenntlich zu machen ist Ziel der Mischung.

2.3.7 Nahtschneiden

Ein Lösungsansatz ist die Methode des Min-Cut/Max-Flow. Sie ist eine Methode zur Findung idealer Nähte. Dies sind Nähte, welche Diskontinuität im zusammengehefteten Gesamtbild minimal halten. Eine detaillierte Erklärung zur Min-Cut/Max-Flow Methode findet sich im Forschungsartikel von Qu et al. (Qu et al. [2018]).

2.3.8 Weichzeichnen

An den Kanten die durch das Nahtschneiden entstanden sind, ist ein plötzlicher Übergang erkennbar. Dies lässt sich selbst durch geschicktes Wählen von Nahtstellen selten verhindern. Deshalb wird zusätzlich dazu die ein Weichzeichner angewandt. Hier gibt es verschiedene Varianten mit Vor- und Nachteilen. Eine Methode ist, einen Weichzeichner in vertikale Richtung der Kante anzuwenden.



Abbildung 6: Ungünstige Naht

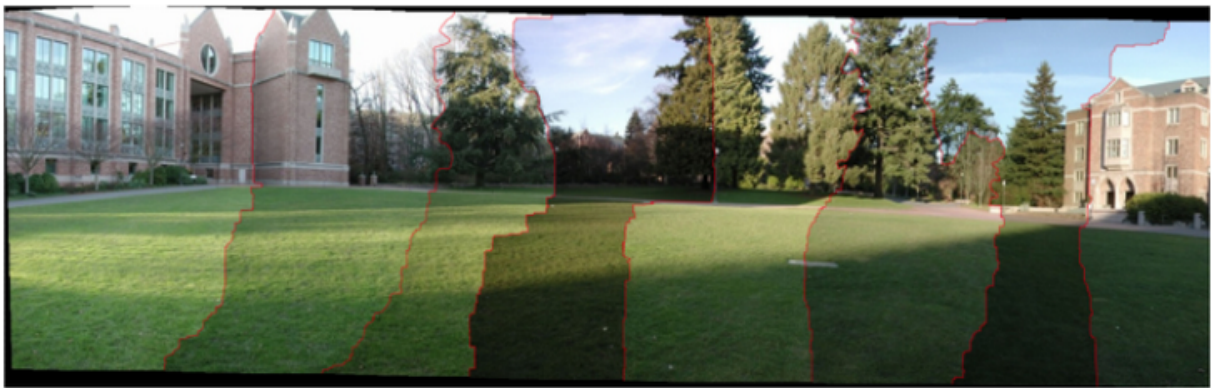


Abbildung 7: Beste Nahtstellen

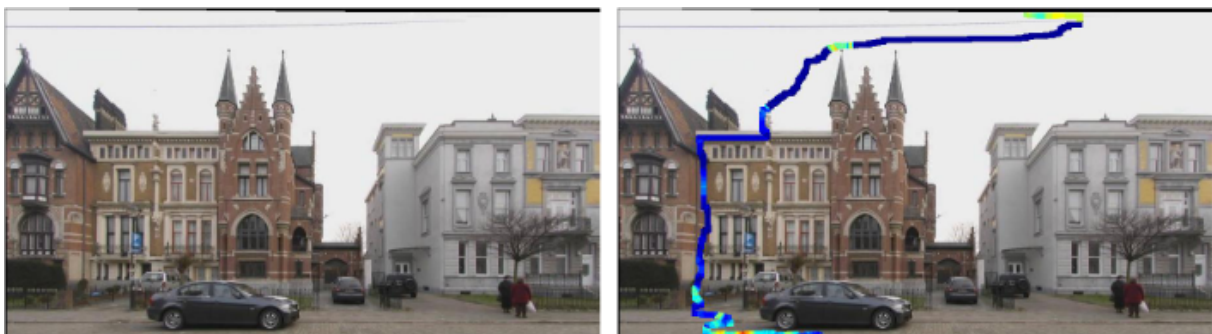


Abbildung 8: Günstige Naht

Folgende Abbildung zeigt die Ausnutzung des Nahtschneidens, um mehrere Teilaufnahmen eines Familienfotos zu einem ansehnlicheren Gesamtfoto zu kombinieren.



Abbildung 9: Fotokombinierung

3 Das SIFT-Verfahren

3.1 Schritte

3.1.1 Erkennung von Skalenraumextremen

Der erste Schritt im SIFT-Verfahren ist die Erkennung von Skalenraumextrema.(vgl. [Lowe \[2004\]](#): 1) In diesem Schritt wird zunächst ein Skalenraum geschaffen. Dazu wird die Ausgangsaufnahme auf verschiedene Größen skaliert. Das dient der Skaleninvarianz. Als nächstes wird für jede dieser Skaliergrößen das Bild über mehrere Stufen hinweg weichgezeichnet. Dieser Vorgang verhindert, dass rauschen im Bild und andere vernachlässigbare Merkmale die Merkmalerkennung beeinflussen. Nun wird die Technik der Gaußner Differenz angewandt. Hierzu wird ein Bild immer mit den danebenliegenden Bildern, also den Bildern der höheren und niedrigeren Weichzeichnestufe, voneinander abgezogen. Man erhält eine Differenz in der sich Merkmale verstärkt ausgeprägt zeigen.

3.1.2 Merkmalslokalisierung

Zu jedem Erkannten Merkmal werden Ort und Skalierung bestimmt. Wenig aussagekräftige Merkmale werden aussortiert.(vgl. [Lowe \[2004\]](#): 1) Dazu werden erst lokale Minima und Maxima gesucht. Dieser Vorgang geschieht durch den Vergleich eines jeden Pixels mit den benachbarten Pixeln. Benachbarte Pixel sind hierbei auch diejenigen Pixel, welcher auf den anliegenden Weichzeichnestufen liegen. Insgesamt ergeben sich 26 Bezugspixel. Ist der momentare Pixel entweder ein Minimum oder ein Maximum, dann wird er als Merkmal festgehalten.

3.1.3 Orientierungszuweisung

Eine oder mehrere Orientierung werden jedem Merkmal zugewiesen. Alle folgenden Operationen werden auf Grundlagen von Bilddaten durchgeführt, die zuvor relativ zu deren Orientierung, Skalierung und Ort transformiert wurden. Dieses Vorgehen gewährleistet, dass das Modell invariant zu diesen Transformationen ist. (vgl. [Lowe \[2004\]](#): 1) Zu jedem Merkmal werden Gradienten berechnet. Dazu wird die Differenz in der Intensität der an den zu untersuchenden Pixel anliegenden Pixel ausgerechnet. Über diese Gradienten ist es möglich auch auf die Orientierung, also die Richtung, des Merkmals zu schließen. Das geschieht über folgende Formel:

$$\text{Magnitude} = \sqrt{[(G_x)^2 + (G_y)^2]} = 16.64$$

$$\Phi = \text{atan}(G_y / G_x) = \text{atan}(1.55) = 57.17$$

Abbildung 10: Gradient und Orientierung

3.1.4 Merkmalsdeskriptor

Die lokalen Bildgradienten um jeden Merkmalspunkt werden abgemessen. Sie auf eine Weise transformiert, die es möglich macht ein wesentliches Maß an Verzerrung von Form und Änderung in der Belichtung zu tolerieren.(vgl. Lowe [2004]: 1)

4 Anwendung von SIFT zur Objekterkennung

Wie zuvor beschrieben können mit dem SIFT-Verfahren Merkmale aus einem Bild extrahiert werden. Nun kann eine Objekterkennung erfolgen. Dazu werden erst Merkmale gepaart. Viele dieser Paarungen werden vorraussichtlich inkorrekt sein, da es uneindeutige Merkmale. Deshalb wird nach Gruppen von mindestens 3 Merkmalen gesucht, welche auf ein Objekt bezogen stimmige Werte liefern. Diese Gruppen werden dann weiter geprüft und dementsprechend entweder als gültig angenommen oder verworfen.(vgl. Lowe [2004]: 19)

4.1 Merkmalspaarung

Merkmale werden anhand des zuvor kalkulierten Merkmalsdeskriptor gepaart. Naheliegenderste Nachbarn, also Nachbarn bei welchen die euklidische Distanz zwischen ihren Deskriptoren minimal ist, werden zusammengeführt.(vgl. Lowe [2004]: 19)

Zu vielen Merkmalen werden sich keine passenden Gegenstücke finden. Dies lässt sich durch Merkmale begründen, welche aus dem Hintergrund entstehen und wenig Aussagekraft über die im Fokus stehenden Objekte haben. Weiterhin gibt es Merkmale, die nur einmalig erkannt werden, da sie tatsächlich auf nur einer einzigen Aufnahme auftauchen. Merkmale beider Sorten sollten nicht beachtet werden. Um das zu erreichen wird eine Schwelle eingesetzt. Dabei handelt es sich allerdings nicht um eine globale Schwelle. Stattdessen wird die Differenz der Distanz zum naheliegendsten Nachbarn, und der Distanz zum zweitnaheliegendsten Nachbarn herangezogen. Überschreitet diese Differenz eine festgelegte Schwelle so wird das Merkmal verworfen. Um mit einem Merkmal arbeiten zu können, muss es sich eindeutig einem anderen zuordnen lassen. Gibt es mehrere Merkmale, deren Deskriptoren eine ähnliche euklidische Distanz zum Bezugsmerkmal haben,

könnte es sich um eine Falschpaarung handeln. Mit Bestimmtheit liefert diese Paarung keine sichere Information. (vgl. Lowe [2004]: 20)

4.2 Effizienzerhöhung

Es sind keine Algorithmen bekannt, mit Ausnahme der erschöpfenden Suche, die den naheliegenden Nachbarn exakt bestimmen. Stattdessen wird der approximierende Best-Bin-First-Algorithmus(BBF) eingesetzt. BBF erkennt den naheliegenden Nachbarn mit hoher Wahrscheinlichkeit. Der BBF funktioniert bei der Merkmalspaarung deswegen besonders gut, weil ein großer Unterschied in der Distanz zum naheliegendsten und zum zweitnaheliegendsten Nachbarn vorausgesetzt wird. (vgl. Lowe [2004]: 20)

5 Vergleich von SIFT und SURF

SIFT ist invariant zu Skalierung, Unschärfe, Rotation, Belichtung und Orientierung. Der Nachteil von SIFT allerdings ist, dass er für viele Anwendungsfälle nicht performant genug ist. Besonders im Bereich des autonomen Fahrens trifft dies zu, denn hier werden Echtzeitergebnisse gebraucht.

Herbert Bay et al. haben als Alternative zu SIFT den Algorithmus SURF(speeded-up robust features) vorgeschlagen. SURF soll demnach schneller als SIFT sein. SURF ist SIFT ähnlich. Jedoch nutzt er nicht zur Lokalisierung von Extrema nicht die Gaußner Differenz sondern die zweite Ableitung eines näherungsweisen Mittelwertfilters welche sich effizient implementieren lässt.(vgl. ? : 4)

Somit ist SURF meist geeigneter für Echtzeitanwendungen wie zum Beispiel das autonome Fahren.

6 Fazit

Das Image Stitching ist ein Bereich der weit verbreitet und bekannt im Bereich der Panoramabildaufnahme ist. Jedoch hat es noch viel tieferreichende Anwendungsfälle, welche oft im Hintergrund als Teil einer größeren Gesamtaufgabe ablaufen. Image Stitching ist somit eine Technik die sehr vielseitig anwendbar ist.

6.1 Zusammenfassung

Image Stitching hat zahlreiche Anwendungsbereiche und ermöglicht den Ausbau vieler Anwendungsbereiche in der Computervision. Beschränkungen bedingt durch einen begrenzten Kamerawinkel können überwunden werden. Das Imagestitching hat bereits ein Reifelevel erreicht, auf welchem Bilder zusammengeführt werden, bei denen kaum kenntlich ist, dass es sich anfangs um Teilbilder handelte.

Die Methode des Image Stitchings beinhaltet die Kalibrierung, die Merkmalerkennung und -zuordnung, die Berechnung der Homographie zwischen Aufnahmen, die globale Ausrichtung, die Projizierung und die Mischung.

6.2 Ausblick

In Zukunft könnten effizientere und sicherere Algorithmen zum Image Stitching noch mehr Möglichkeiten im Bereich des autonomen Fahrens, der Bildmosaikierung sowie Augmented und Virtual Reality bieten. Die Merkmalerkennung und -zuordnung ermöglicht es maschinell Objekte wiederzuerkennen oder kann auch in der Gesichtserkennung angewandt werden.

Abbildungsverzeichnis

1	Merkmalsmatching(vgl. Qu et al. [2020])	5
2	Homographie	6
3	Kugelprojektion(vgl. Mehta and Bhirud [2011])	8
4	Kamerawinkel(vgl. Mehta and Bhirud [2011])	8
5	Illustrierung(vgl. Qu et al. [2018])	9
6	Ungünstige Naht(vgl. Liao et al. [2019])	10
7	Beste Nahtstellen(vgl. Qu et al. [2018])	10
8	Günstige Naht(vgl. Liao et al. [2019])	10
9	Fotokombinierung(vgl. Szeliski [2011])	11
10	Gradient und Orientierung(vgl. SINGH)	12

Tabellenverzeichnis

Literatur

- Aijing Feng, Jianfeng Zhou, Earl Vories, and Kenneth A. Sudduth. Evaluation of cotton emergence using uav-based narrow-band spectral imagery with customized image alignment and stitching algorithms. *Remote. Sens.*, 12(11):1764, 2020. doi: 10.3390/rs12111764.
- Pranoti Kale and KR Singh. A technical analysis of image stitching algorithm. *International Journal of Computer Science and Information Technologies*, 6(1):284–288, 2015. URL <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.666.3271&rep=rep1&type=pdf>.
- Tianli Liao, Jing Chen, and Yifang Xu. Quality evaluation-based iterative seam estimation for image stitching. *Signal Image Video Process.*, 13(6):1199–1206, 2019. doi: 10.1007/s11760-019-01466-9.
- David G. Lowe. Distinctive imagefeaturesfromscale-invariantkeypoints. 2004. URL <https://people.eecs.berkeley.edu/~malik/cs294/lowe-ijcv04.pdf>.
- Ryan Maponga, Sung-Hyok Ro, and Se-Hun Kim. Image stitching techniques. 2017. URL https://sites.tufts.edu/eeseniordesignhandbook/files/2017/05/Yellow_Maponga_F1.pdf.
- Jalpa D. Mehta and S. G. Bhirud. Image stitching techniques. In S. J. Pise, editor, *Thinkquest~2010*, pages 74–80, New Delhi, 2011. Springer India. ISBN 978-81-8489-989-4.
- Zhong Qu, Tengfeng Wang, Shiquan An, and Ling Liu. Image seamless stitching and straightening based on the image block. *IET Image Process.*, 12(8):1361–1369, 2018. doi: 10.1049/iet-ipr.2017.1064.
- Zhong Qu, Jun Li, Kang-Hua Bao, and Zhi-Chao Si. An unordered image stitching method based on binary tree and estimated overlapping area. *IEEE Trans. Image Process.*, 29: 6734–6744, 2020. doi: 10.1109/TIP.2020.2993134.
- AISHWARYA SINGH. A detailed guide to the powerful sift technique for image matching (with python code). URL <https://www.analyticsvidhya.com/blog/2019/10/detailed-guide-powerful-sift-technique-image-matching-python/>.
- Richard Szeliski. *Computer Vision - Algorithms and Applications*. Texts in Computer Science. Springer, 2011. ISBN 978-1-84882-934-3. doi: 10.1007/978-1-84882-935-0.