

MASTER THESIS
Matthis Keppner

Vergleich von Clustering-Verfahren und Da- tenvorverarbeitungsvarianten bei Anwendung auf multispektralen Bilddaten

FAKULTÄT TECHNIK UND INFORMATIK
Department Informatik

Faculty of Engineering and Computer Science
Department Computer Science

Matthis Keppner

Vergleich von Clustering-Verfahren und Datenvorverarbeitungsvarianten bei Anwendung auf multispektralen Bilddaten

Masterarbeit eingereicht im Rahmen der Masterprüfung
im Studiengang *Master of Science Informatik*
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Tim Tiedemann
Zweitgutachter: Prof. Dr. Peer Stelldinger

Eingereicht am: 28. August 2023

Matthis Keppner

Thema der Arbeit

Vergleich von Clustering-Verfahren und Datenvorverarbeitungsvarianten bei Anwendung auf multispektralen Bilddaten

Stichworte

Clustering, Bildverarbeitung, Bildregistrierung, multispektral,multimodal, SIFT, Bildzerteilung

Kurzzusammenfassung

Diese Thesis beschäftigt sich mit dem Vergleich von Clustering-Verfahren und Datenvorverarbeitungsvarianten bei Anwendung auf multispektralen Bilddaten. Multispektrale Bilddaten werden in verschiedenen Bereichen wie Recycling und Landwirtschaft eingesetzt, um Objekte oder Materialien zu unterscheiden. Die Arbeit stellt einen neuen zweistufigen Ansatz für die Registrierung der einzelnen Kanäle vor, der verschiedene Transformationsmodelle und Filter berücksichtigt. Außerdem werden verschiedene Vorverarbeitungsschritte wie Skalierung, Normalisierung und Hauptkomponentenanalyse durchgeführt, um die Daten für die Clustering-Analyse zu optimieren. Die Clustering-Verfahren K-Means, HDBSCAN und Mean-Shift werden auf zwei Datensätzen angewendet, die aus Aufnahmen von Sperrmüll und Erdbeeren bestehen. Die Ergebnisse werden anhand von optischen und numerischen Metriken evaluiert und verglichen.

Matthis Keppner

Title of Thesis

Comparison of clustering methods and data preprocessing variants when applied to multispectral image data

Keywords

Clustering, Image processing, Image registration, multispectral,multimodal, SIFT, Image slicing

Abstract

This master thesis deals with the comparison of clustering methods and data preprocessing variants when applied to multispectral image data. Multispectral image data are used in various fields such as recycling and agriculture to distinguish objects or materials. The thesis presents a new two-stage approach for the registration of the individual channels, which considers different transformation models and filters. In addition, various preprocessing steps such as scaling, normalization and principal component analysis are performed to optimize the data for the clustering analysis. The clustering methods K-Means, HDBSCAN and Mean-Shift are applied to two datasets consisting of images of bulky waste and strawberries. The results are evaluated and compared using optical and numerical metrics.

Inhaltsverzeichnis

Abbildungsverzeichnis	vii
Tabellenverzeichnis	viii
1 Einleitung	1
1.1 Problem	1
1.2 Zielsetzung	2
1.3 Aufbau	3
2 Theoretischer Hintergrund	4
2.1 Bild Registrierung	4
2.1.1 Merkmalserkennung	4
2.1.2 Merkmalsabgleich	6
2.1.3 Entfernen von Ausreißern	7
2.1.4 Berechnung einer Transformationsfunktion	7
2.1.5 Bildtransformation	8
2.1.6 Evaluation	8
2.1.7 Registrierung von multispektralen Bildern	9
2.2 Clustering-Verfahren	11
2.2.1 Vorverarbeitungsschritte für Clustering-Verfahren	13
2.2.2 Evaluation	14
2.2.3 MS Clustering	15
3 Methodik	17
3.1 Datenaufnahme	17
3.2 Registrierung	18
3.3 Clustering-Verfahren	20

4 Ergebnisse	24
4.1 Registrierung	24
4.1.1 Sperrmülldatensatz	25
4.1.2 Erdbeerdatensatz	28
4.1.3 Zusammenfassung	30
4.2 Clustering-Verfahren	30
4.2.1 Sperrmülldatensatz	31
4.2.2 Erdbeerdatensatz	33
5 Diskussion	39
5.1 Multispektrale Registrierung	39
5.2 Clustering-Verfahren	40
6 Fazit und Ausblick	43
Literaturverzeichnis	45
Selbstständigkeitserklärung	48

Abbildungsverzeichnis

2.1	Gefundene Keypoints mit Gradienten-Richtungen	5
2.2	Vergleich von Anzahl vor und nach Ratio-Test	7
2.3	Vergleich der Falschfarbenbilder der Registrierungsmethoden	10
2.4	Starke Verzerrung nach Transformation	11
3.1	Einzelbilder des multispektralen Datensatzes	19
3.2	Teilbilder nach Zerlegung eines RGB-Bilds	22
4.1	Falschfarbenbilder der Registrierungsmethoden von einer Szene aus dem Sperrmülldatensatz	28
4.2	Falschfarbenbilder der Registrierungsmethoden von einer Szene aus dem Erdbeerdatensatz	28
4.3	HDBSCAN Cluster-Ergebnisse für eine Szene aus dem Sperrmülldatensatz	34
4.4	K-Means Cluster-Ergebnisse für eine Szene aus dem Sperrmülldatensatz .	34
4.5	Mean-Shift Cluster-Ergebnisse für eine Szene aus dem Sperrmülldatensatz	35
4.6	HDBSCAN Cluster-Ergebnisse für eine Szene aus dem Erdbeerdatensatz .	37
4.7	K-Means Cluster-Ergebnisse für eine Szene aus dem Erdbeerdatensatz . .	38
4.8	Mean-Shift Cluster-Ergebnisse für eine Szene aus dem Erdbeerdatensatz .	38

Tabellenverzeichnis

3.1	Übersicht über verwendete Sensorik und Spektralfilterkombinationen . . .	18
3.2	Funktion des Zerlegungsalgorithmus: Ausgangsmatrix (links) mit einer Auflösung von $(b * h)$, aufgeteilt in n^2 Teilmatritzen (rechts) mit einer Auflösung von $(\frac{b}{n} * \frac{h}{n})$ und $n =$ Skalierungsfaktor.	21
4.1	Durchschnittliche Anzahl an Matches vor und nach Ratio-Test für die Mülldaten	26
4.2	Durchschnittlicher RMSE Bild- und Matchesweise für den Mülldatensatz .	27
4.3	Durchschnittliche Anzahl an Matches vor und nach Ratio-Test für die Erdbeerdaten	29
4.4	Durchschnittlicher RMSE Bild- und Matchesweise für Erdbeerdatensatz .	29
4.5	Calinski-Harabasz Index Sperrmülldatensatz	32
4.6	Davies-Bouldin Index, Sperrmülldatensatz	33
4.7	Calinski-Harabasz Index, Erdbeerdatensatz	36
4.8	Davies-Bouldin Index, Erdbeerdatensatz	36

1 Einleitung

1.1 Problem

Aktuell ist das Spektrum der verfügbaren Sensoren und Anwendungen für multispektrale Bildgebung sehr groß und wird zum Beispiel in Bereichen wie Kulturerbe, Landwirtschaft, medizinischer Bildgebung und Recycling eingesetzt. Hierdurch können mehr Informationen gewonnen werden, da sich das aufgenommene Spektrum um die für den Menschen nicht sichtbaren Bereiche des Lichtspektrums erweitert. [11] Aufgrund der großen Mengen an multispektralen Daten, die aus Satellitenbildern, medizinischen Geräten, geografischen Informationssystemen (GIS), Bilddatenbankexplorationen usw. gewonnen werden können, ist es für die Nutzer teuer und schwierig, multispektrale Daten im Detail zu untersuchen. Clustering kann dazu beitragen, den Prozess der Analyse und das Verständnis multispektraler Daten zu verbessern. Es wird verwendet, um interessante Merkmale und Muster zu identifizieren und zu extrahieren. [9] Verschiedene Sensoren können verwendet werden, um multispektrale Bilddaten aufzunehmen. Dieser Ansatz kann als multimodal bezeichnet werden. Hierbei werden Bilder der gleichen Szene mit unterschiedlichen Sensoren aufgenommen, um eine höhere spektrale Auflösung dieser Szene zu erhalten. [27] Obwohl ein solches System von multispektralen Sensoren zwar mehr Informationen bieten kann, bringen diese auch einige Herausforderung mit sich. Zum einen sind die einzelnen Bilder der verschiedenen Sensoren gegebenenfalls nicht genau überlagernd, sodass hier eine Bildregistrierung nötig ist. Diese führt zu verschiedenen Herausforderungen, da sich multispektrale Bilddaten stark in ihren Intensitätswerten unterscheiden, und somit eine Registrierung erschwert. Zum anderen kann die hohe Dimensionalität von multispektralen Bilddaten zu hohen Komplexitäten bei Berechnungen führen. Auch ist die Anzahl an einzelnen Datenpunkten mit einer hohen Auflösung der Kameras sehr hoch, sodass auch der Berechnungsaufwand und Speicherbedarf steigt. Bereits im Vorfeld dieser Thesis wurde sich mit den verschiedenen Problem beschäftigt, sodass nun an den Problemen weitergearbeitet werden kann, um bessere Lösungen zu entwickeln.

1.2 Zielsetzung

Insgesamt soll in dieser Thesis von den Rohdaten verschiedener Kameras ausgehend, verschiedene Clustering-Verfahren verglichen werden, um eine erste Einschätzung von zwei Datensätzen zu erhalten und um zu entscheiden, ob ein solches Verfahren sinnvoll ist. Desweiteren werden verschiedene Vorverarbeitungsmethoden untersucht, die das Ergebnis beeinflussen können. Dabei wird ein neuer Ansatz für die Registrierung der einzelnen Kanäle entwickelt. Dieser berücksichtigt verschiedene Informationen über die Datenaufnahme und die verwendete Sensorik, um die Herausforderungen bei der Bildregistrierung von multispektralen Daten für den verwendeten multimodalen Aufbau zu lösen. Dieser Ansatz ist zweistufig aufgebaut. In der ersten Stufe werden die Bilder der einzelnen Kameras mit Hilfe einer partiellen affinen Transformation registriert. Hierdurch wird die Transformation auf Translation, Skalierung und Rotation beschränkt. Zudem wird ein Differenzfilter eingesetzt, der nur Matches von gefundenen Merkmalspunkten zulässt, die eine bestimmte Differenz in X- und Y- Richtung nicht überschreiten. Dadurch wird die Transformation weiter eingeschränkt, sodass nur eine minimale Rotation und Translation möglich ist. Diese Einschränkungen werden vorgenommen, da bei den Bildern pro Kamera lediglich kleine Verschiebungen und Rotationen zu erwarten sind. Eine Skalierung ist durch die Verschiebung des Fokuspunkts bei verschiedenen Wellenlängen möglich. Dieser Effekt nennt sich chromatische Aberration. [2] In der zweiten Stufe werden die Bilder der unterschiedlichen Kameras registriert. Hier wird eine projektive Transformation berechnet, welche anschließend auf alle Bilder der jeweiligen Kamera angewendet wird. Für den Vergleich der Clustering-Analyse werden zunächst verschiedene Vorverarbeitungsschritte durchgeführt. Es werden verschiedene Datensätze aufgebaut, um die Wirkung anschließend vergleichen zu können. Eine Vorverarbeitungsmethode bildet dabei eine Standardisierung, die für jeden Kanal berechnet wird. Zum anderen wird eine Hauptkomponentenanalyse (PCA) durchgeführt, und verschieden viele Axen für die Weiterverarbeitung genutzt. So kann überprüft werden, ob eine Dimensionsreduktion möglich ist. Da die multispektralen Daten eine große Menge an Datenpunkten erzeugen, werden geeignete Methoden verwendet, um diese möglichst effizient zu clustern. Hierfür wird eine Methode vorgestellt, um das Bild sinnvoll zu zerlegen. Abschließend soll untersucht werden, welche Clustering-Verfahren sich für multispektrale Bilddatenverarbeitung in den Bereichen Recycling und Landwirtschaft eignen.

1.3 Aufbau

Die Arbeit ist wie folgt aufgebaut. Kapitel 2 gibt einen Überblick über den theoretischen Hintergrund. Hier werden vor allem die Themen Bildregistrierung und Clustering von multispektralen Bilddaten beleuchtet und verschiedene Evaluationsmetriken vorgestellt. Im Kapitel 3 werden die Methoden der einzelnen Vorgehensweisen beschrieben. Die Ergebnisse werden anschließend in Kapitel 4 ausgewertet und in Kapitel 5 diskutiert. Im letzten Kapitel wird ein Fazit gezogen und ein Ausblick gegeben.

2 Theoretischer Hintergrund

Es folgt eine kurze Einführung in die Bildregistrierung und in die Clustering-Analyse. Hierfür werden verschiedene Vorverarbeitungsmethoden erläutert. Zudem werden für beide Themen Evaluationsmethoden vorgestellt.

2.1 Bild Registrierung

Bei der Bildregistrierung werden zwei oder mehr Bilder der selben Szene aufeinander ausgerichtet. Hierfür wird ein Bild als Zielbild ausgewählt und die anderen Bilder mit Hilfe von Transformationen so verändert, dass diese übereinander liegen. Tareen et al. [23] beschreiben Bildregistrierung in fünf Schritten:

1. Merkmalerkennung und -Extraktion
2. Merkmalsabgleich
3. Entfernen von Ausreißern
4. Berechnung einer Transformationsfunktion
5. Bildtransformation

2.1.1 Merkmalerkennung

Im ersten Schritt der Bildregistrierung werden Detektionsalgorithmen eingesetzt, um Merkmalspunkte (feature points) für das Ziel- und Quellbild zu finden. Bei Merkmalspunkten kann es sich um Ecken, Kanten, Kreuzungen, Linien oder Flächen handeln. [23] Jeder Merkmalspunkt besteht aus einem Merkmalsvektor, der Informationen über die Größe, das Muster und die dominante Richtung der Gradienten eines jeden Merkmalspunktes kodiert. [26] In Abbildung 2.1 wird ein Bild mit gefundenen Merkmalspunkten

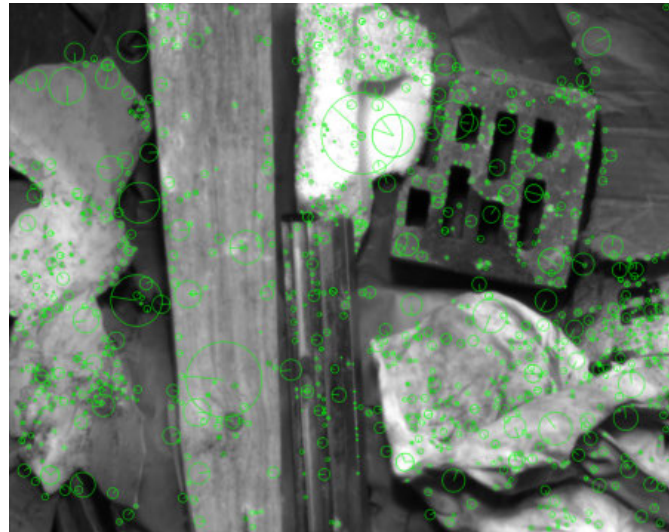


Abbildung 2.1: Gefundene Keypoints mit Gradienten-Richtungen

(grün) gezeigt. Dabei wird die Gradienten-Ausrichtung durch den Strich im Kreis dargestellt. Die Größe der Kreise ist abhängig von dem Skalenraum, auf dem der Merkmalspunkt gefunden wurde. Im folgenden werden weit verbreitete Algorithmen kurz vorgestellt.

Das 1999 von Lowe eingeführte SIFT-Verfahren (scale-invariant feature transform) ist eine merkmalsbasierte Methode, die in der Lage ist, Merkmale zu identifizieren, die gegenüber Rotation, Translation und Skalierung unverändert bleiben. [16] SIFT kann in folgende vier Berechnungsschritte aufgeteilt werden [17]:

- (1) Bestimmung von Skalenraum-Extrema
- (2) Lokalisierung von Merkmalspunkten
- (3) Orientierungszuweisung
- (4) Merkmalspunkteextraktion

Der SIFT-Detektor basiert auf dem Difference-of-Gaussians (DoG) Operator, der eine Annäherung an den Laplacian-of-Gaussian (LoG) ist. Merkmalspunkte werden durch die Suche nach lokalen Maxima unter Verwendung von DoG in verschiedenen Skalen der betreffenden Bilder ermittelt. [23] Anschließend werden die gefundenen Extrema durch eine Stabilitätsanalyse weiter untersucht. Werden die Bedingungen erfüllt, wird der Punkt zu einem Merkmalspunkt und die genaue Position sowie Skalierung berechnet. Jedem

Merkmalspunkte werden basierend auf lokalen Bildgradientenrichtungen ein oder mehrere Hauptorientierungen zugewiesen. Die lokalen Bildgradienten werden in der Umgebung jedes Merkmalspunkts auf der ausgewählten Skalierung gemessen. Diese werden in eine Darstellung umgewandelt, die starke, lokale Formverzerrung und Änderungen der Beleuchtung zulässt. [17]

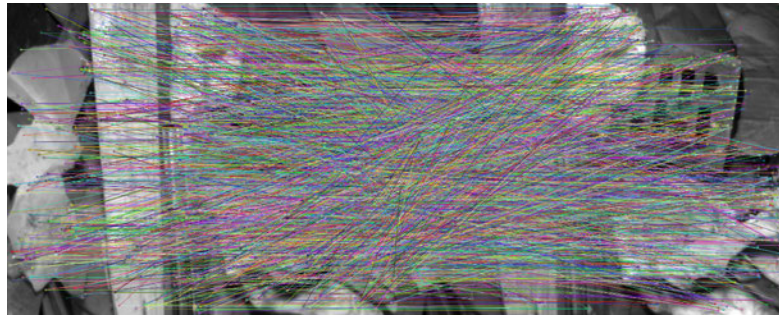
SIFT hat den SURF-Extraktionsalgorithmus teilweise inspiriert. SURF verwendet eine ganzzahlige Näherung des Determinanten des Hessischen Blob-Detektors, um Merkmalspunkte aus einem Bild zu finden. Der Determinant des Hessischen Blob-Detektors kann mit drei ganzzahligen Operationen unter Verwendung eines im Voraus berechneten Integralbildes berechnet werden. Dieser Merkmalspunkt-Extraktionsalgorithmus basiert auf der Summe der Haar-Wavelet-Komponenten um den Kontrollpunkt. SURF konzentriert sich auf skalierbare und rotationsinvariante Detektoren und Deskriptoren. Um die Skalenninvarianz der Merkmalspunkte zu gewährleisten, wandelt SURF das Bild in Koordinaten um, die eine Multiresolutionspyramidenteknik verwenden. Das Ergebnis ist ein Bild mit reduzierter Bandbreite. Dies schafft einen speziellen Unschärfefeekt auf dem Bild und erzeugt einen Skalenraum. Der Skalenraum macht die Merkmalspunkte skalierbar. [1]

2.1.2 Merkmalsabgleich

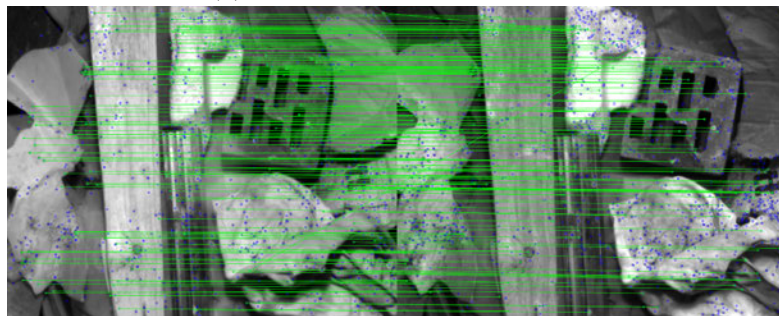
Beim Merkmalsabgleich werden die Merkmalspunkte der beiden einzelnen Bilder zusammengeführt. Dafür werden die Merkmalspunkte des einen Bildes mit denen des anderen Bildes verglichen. Bei einem Brute-Force-Matcher werden alle euklidischen Distanzen zwischen den Merkmalspunkten berechnet und jeweils die Kleinste ausgewählt. Ein Nachteil an Brute-Force-Matchern ist die hohe Komplexität und der hohe Rechenaufwand. [26] Andere Matching-Algorithmen sind zum Beispiel Nearest-Neighbor oder ein Nearest-Neighbor-Ratio. [23] Um gute Matches zu finden, hat Lowe in seinem Artikel [16] einen Ratio-Test eingeführt. Hier wird jeweils die Distanz des Erstbesten mit der des Zweitbesten Ergebnisses verglichen. Unterschreitet dieses Verhältnis einen Grenzwert (0.7) wird das Match verworfen. Die Berechnung ist in Formel

$$m.\text{dist} < 0.7 * n.\text{dist} \text{ mit } m = \text{bestMatch} \text{ und } n = 2\text{ndbestMatch} \quad (2.1)$$

In Abbildung 2.2 werden beispielhaft gefundene Matches vor und nach dem Ratio-Test von Lowe gezeigt.



(a) Matches vor dem Ratio-Test



(b) Matches nach dem Ratio-Test

Abbildung 2.2: Vergleich von Anzahl vor und nach Ratio-Test

2.1.3 Entfernen von Ausreißern

Im vorherigen Schritt können nicht alle falschen Matches (Ausreißer) aussortiert werden, weshalb ein Extraschritt erforderlich ist. Random Sample Consensus (RANSAC), M-estimator Sample Consensus (MSAC) und Progressive Sample Consensus (PROSAC) sind einige der robusten, probabilistischen Methoden, die zur Entfernung von Ausreißern aus abgeglichenen Merkmalen verwendet werden. In der Praxis kann dieser Schritt mit dem nachfolgenden Schritt kombiniert werden. [23]

2.1.4 Berechnung einer Transformationsfunktion

Im Schritt der Berechnung der Transformationsfunktion wird eine Transformation gefunden, welche eines der Bilder so transformiert, dass die Merkmalspunkte der beiden Bilder übereinander liegen. Dazu kann zum Beispiel eine Homographie-Matrix H (3×3 Matrix)

berechnet werden, die folgende Formel erfüllt:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = H \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}$$

Es werden mindestens vier Matches benötigt. Jedoch sind mehr Matches zu bevorzugen, da zumeist ein besseres Ergebnis erzielt werden kann. [26]

2.1.5 Bildtransformation

Im letzten Schritt wird eines der Bilder mit Rücksichtnahme auf das Zielbild und mit der in Schritt vier berechneten Transformation transformiert.

2.1.6 Evaluation

Verschiedene Verfahren können für die Evaluation einer Registrierung genutzt werden. Wang et al. verwenden die Wurzel des mittleren quadratischen Fehlers (root mean square error (RMSE)) von den Merkmalspunkten der Bilder und einem experten-generierten ground truth als Hauptkriterium der Registrierung. Wenn der RMSE unter 10 Pixeln liegt, sei die Registrierung erfolgreich. [25]

In [27] werden drei mögliche Fehlerkategorien zur Evaluation der Registriergenauigkeit beschrieben: Der ersten Kategorie werden Lokalisierungsfehler der Merkmalspunkte zugeordnet. Ein größerer Fehler spricht für eine schlechtere Lokalisierung der Merkmalspunkte. Dieser Fehler ist nicht vom Bild aus messbar, da sich dieser nur intern im Detektionsalgorithmus finden lässt. Als zweite Kategorie werden Matching-Fehler beschrieben. Hierbei handelt es sich um fehlerhafte Zuordnungen der einzelnen Merkmalspunkte zwischen den Bildern. Merkmalspunkte die tatsächliche Matches sind, können verifiziert werden, indem sie von unterschiedlichen Matching-Algorithmen zugeordnet werden. Falsche Matches können eine Transformation stark verzerren und unbrauchbar machen. Als letzter Fehler wird die Differenz zwischen dem Registrierungsmodell und der tatsächlichen geometrischen Verzerrung zwischen den Bildern beschrieben. Dieser Fehler wird Ausrichtungsfehler genannt. Dieser Fehler kann aus zwei Gründen auftreten. Zum einen kann das gewählte Transformationsmodell nicht zur eigentlichen Verzerrung passen. Zum anderen können durch eine schlechte oder fehlerhafte Lokalisierung der Merkmalspunkte, die Transformation nicht korrekt berechnet werden. Ein Ausrichtungsfehler ist

im Allgemein immer vorhanden. [27]

In der Literatur wird der Erhöhung der Anzahl der zu extrahierenden Merkmalspunkten große Bedeutung beigemessen. Es hat sich gezeigt, dass eine größere Anzahl von Kontrollpunkten zu mehr guten Korrespondenzen und damit zu genaueren Registrierungen führt. [26]

2.1.7 Registrierung von multispektralen Bildern

Durch die steigende Anzahl und Entwicklung von kompakten, multispektralen Kameras, wie beispielsweise von MicaSense¹ ist auch die Nachfrage nach geeigneten Bildregistrierungen für multispektrale Bilder gestiegen. Zudem können verschiedene multispektrale Kamerasysteme, bestehend aus verschiedenen Sensoren mit verschiedenen optischen Eigenschaften, wie Brennweite oder Sensorgröße, verwendet werden. So werden Szenen von verschiedenen Blickpunkten aufgenommen, was zusätzliche Verzerrungen der verschiedenen Bilder bewirkt. Hierdurch und durch die nichtlinearen Intensitätsunterschiede von multispektralen Bildern bleibt ein erfolgreiches Registrieren eine Herausforderung. [10] Jhan et al. zeigen einen Ansatz, von einem normalisierten SURF (N-SURF), womit die Anzahl der Merkmalspunkte zwischen den unterschiedlichen, multispektralen Bildpaaren gesteigert werden kann. So ist eine einstufige Registrierung möglich. Auch wurden verschiedene Transformationsmodelle verglichen. Die besten Ergebnisse wurden mit einer, um Linsen-Verzerrungsparameter erweiterte projektiven Transformation, erzielt. [10] Shaharom et. al. kommen in ihrem Übersichtspaper zum Schluss, dass SIFT und SURF für das Matching von multispektralen Bildern oft genutzt wird, jedoch zu schlechten Leistungen und Ergebnissen führen kann. Hauptsächlich treten viele fehlerhafte Matches auf, welche jedoch durch geeignete Matching-Algorithmen verringert werden können. Auch eine Veränderung und Kombination mit anderen Algorithmen verbessert die Performanz von SIFT und SURF. [21] In [26] wird ein datenbasiertes Framework gezeigt. Hierbei wird der Matching-Algorithmus nach der durchschnittlichen Anzahl der Merkmalspunkte ausgewählt. Es wurden neben SIFT und SURF auch ORB und BRISK verwendet. SIFT fand gegenüber den anderen Algorithmen die meisten Merkmalspunkte und wurde somit überwiegend verwendet. ORB fand sehr wenig Merkmalspunkte und zum Teil gar keine auf dem blauen Kanal der RGB-Bilder. In [2] wird eine multispektrale Filterrada-Kamera vorgestellt. Dazu werden verschiedene spektrale Bandpassfilter sequentiell vor

¹<https://support.micasense.com/hc/en-us>

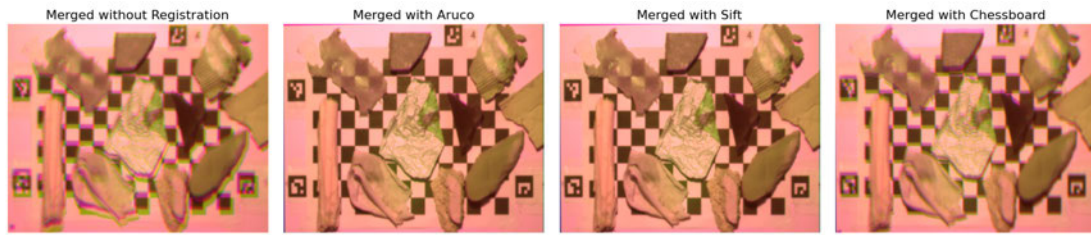


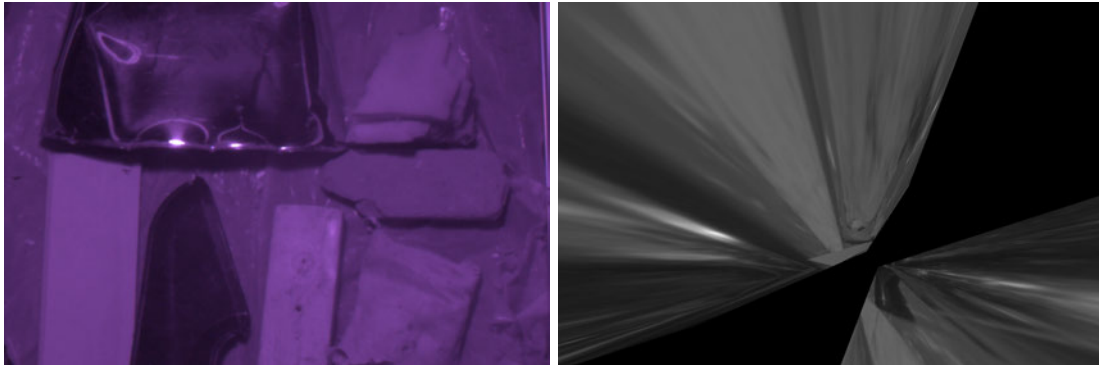
Abbildung 2.3: Vergleich der Falschfarbenbilder der Registrierungsmethoden

den Sensor geschoben, um eine multispektrale Aufnahme zu erhalten. Sie zeigen, dass affine Transformationen gute Modelle für chromatische Aberrationen darstellen können. Es werden zwei Algorithmen vorgestellt. Der erste basiert auf kleinen Regionen des Bildes. Hier werden Verschiebungsvektoren berechnet und für die Annäherung der Transformationsparameter benutzt. Im globalen Ansatz werden die gesamten Bilder genutzt. Genutzt werden iterativ, vorläufige geometrische Transformation mit verschiedenen Parametern und eine Ähnlichkeit optimiert. Beide Ansätze sind in der Lage, die Bilder zu registrieren. [2]

Im Vorfeld dieser Arbeit wurde bereits an dem Problem der Registrierung von multispektralen Bildern gearbeitet. [15] Das zugrundeliegende System besteht aus drei verschiedenen Kameras, welche mit spektralen Bandpassfiltern sequentiell multispektrale Bilddaten aufnehmen. Dazu wurden verschiedene Ansätze ausprobiert. Zum einen wurde eine Registrierung aufgrund von einem, vor der Serie aufgenommenen, Schachbrettmusters durchgeführt. Durch das bekannte Muster können Punkte subpixelgenau bestimmt werden. Ein Problem mit diesem Ansatz ist, dass Verschiebungen innerhalb der Serie, durch zum Beispiel Filterwechsel oder chromatische Aberration, nicht für die Berechnung der Transformationen berücksichtigt werden. Bei einem anderen Ansatz wurden Aruco-Marker im Blickfeld der Kameras platziert. Die Positionen dieser Marker kann pixelgenau in allen benutzten Spektren erkannt werden, sodass auch hier ein sehr kleiner Lokalisierungsfehler erzielt werden kann. Aus diesen Punkten wurden Transformationen auf ein gemeinsames Zielbild berechnet. Während dieser Ansatz gute Registrierungsergebnisse erbringen kann, ist er nur für Aufnahmen in Labor-Umgebung sinnvoll einzusetzen. Marker auf einem Feld oder in einer Recyclinghalle anzubringen, ist auf Grund der örtlichen Gegebenheiten nicht exakt möglich. Zudem wird der Arbeitsbereich durch die Marker eingeschränkt. Als letzten Ansatz wurde SIFT verwendet, um Merkmalspunkte aus den einzelnen Bildern zu extrahieren. Diese wurden anschließend benutzt, um eine perspektivische Transformation für jeden Kanal auf einen Zielkanal zu berechnen. Dieser Ansatz kann als zielbasiert beschrieben werden. In Abbildung 2.3 finden sich Falschfarben-Bilder

der verschiedenen, genutzten Methoden. Hierbei wurden alle Einzelbilder auf ein Bild mit drei Farbkanälen zusammengeführt. Auffällig ist die oben angesprochene verbliebene Verzerrung bei dem Bild "Merged with Chessboard"(rechts). Ein Problem bei dem SIFT-Ansatz ist, dass einige Bildpaare sehr wenig Matches der Merkmalspunkte besitzen, was zu schlechten und stark verzerrten Ergebnissen führt. In Abbildung 2.4 ist eine solche Verzerrung zu sehen.

Auf der Grundlage dieser Arbeiten wird in dieser Thesis ein neuer Ansatz vorgestellt.



(a) Bild vor der Transformation

(b) Bild nach der Transformation

Abbildung 2.4: Starke Verzerrung nach Transformation

Dieser ist zweistufig aufgebaut, sodass verschiedene Transformationsmodelle für unterschiedliche erwartete Verzerrungen genutzt werden können. Zudem können so Bildpaare mit einer größeren Anzahl an Matches für eine Transformation genutzt werden.

2.2 Clustering-Verfahren

Clustering ist eine nützliche Strategie, um Gruppen und interessante Verteilungen und Muster in den Daten zu entdecken. Das Clustering-Problem besteht darin, einen gegebenen Datensatz in Gruppen aufzuteilen, sodass die Datenpunkte in einem Cluster einander ähnlicher sind, als Punkte in verschiedenen Clustern. [9] Halkidi et al. beschreiben folgende Arten von Clustering: Das partitionierende Clustering versucht den Datensatz direkt in eine Menge von disjunkten Clustern zu zerlegen. Sie versuchen eine Anzahl von Partitionen zu bestimmen, die eine bestimmte Kriterienfunktion optimieren. Die Kriterienfunktion kann die lokale oder globale Struktur der Daten betonen. Ihre Optimierung ist ein iteratives Verfahren. [9]

Die hierarchische Cluster-Analyse erfolgt sukzessive, indem kleinere Cluster zu größeren

zusammengefasst oder größere Cluster aufgeteilt werden. Das Ergebnis des Algorithmus ist ein Baum von Clustern (Dendrogramm), der zeigt, wie die Cluster miteinander verwandt sind. Durch das Abschneiden des Dendrogramms auf einer gewünschten Ebene, wird eine Cluster-Bildung der Datenobjekte in disjunkte Gruppen erreicht.

Bei dichte-basierten Clustering-Verfahren werden benachbarte Objekte eines Datensatzes auf der Grundlage von Dichtekriterien in Cluster gruppiert.

Rasterbasierte Clustering-Verfahren von Algorithmen, werden hauptsächlich für die räumliche Datenanalyse genutzt. Hierbei wird der Raum in eine endliche Anzahl von Zellen quantisiert um anschließend alle Operationen im quantisierten Raum durchzuführen. [9] K-Means ist ein Beispiel für ein partitionierendes Clustering-Verfahren. Die Methode wurde von MacQueen 1967 in [18] veröffentlicht. Der Algorithmus beginnt mit der Initialisierung von k Clusterzentren. Dann weist er jedes Objekt des Datensatzes dem Cluster zu, dessen Zentrum am nächsten ist, und berechnet die Zentren neu. Der Prozess wird fortgesetzt, bis sich die Zentren der Cluster nicht mehr ändern oder eine bestimmte Anzahl an Iterationen erreicht ist. [9]

Der Meanshift-Algorithmus ist ein weiteres Verfahren zur Cluster-Analyse, welches jeden Datenpunkt einem Modus zuordnet. Hierbei handelt es sich um ein Maxima einer Dichtefunktion auf einer Teilmenge der Daten. Die Dichtefunktionen werden nach dem Konzept des Kernel Density Estimation (KDE) berechnet. Ein Gauß-Kernel wird häufig eingesetzt. Für jede Teilmenge wird der Kernel in Richtung des berechneten Mittelwerts verschoben, bis diese konvergiert. Die Anzahl der Cluster wird nicht im Voraus festgelegt, sondern ergibt sich aus dem Algorithmus. [5]

Ein dichte-basiertes Verfahren bildet DBSCAN [8]. Jeder Punkt in einem Cluster muss in seiner Nachbarschaft mit einem bestimmten Radius ε eine Mindestanzahl von Punkten enthalten, um nicht als Ausreißer zu gelten. Die Cluster ergeben sich aus den Punkten, welche am dichtesten zusammen sind. DBSCAN kann mit Rauschen (Ausreißern) umgehen und Cluster beliebiger Form finden. [9]

Hierarchisches DBSCAN (HDBSCAN) ist eine Erweiterung von Campello et al.. Dieses Verfahren führt DBSCAN mit verschiedenen Werten für ε durch, sodass Cluster mit unterschiedlicher Dichte gefunden werden. [19] Um möglichst viele Informationen aus den Daten zu erhalten kann eine Vorverarbeitung nützlich sein. Zudem kann eine Vorverarbeitung helfen, eine große Menge an Daten zu verarbeiten, da diese im Clustering sonst sehr große Speicher- und Berechnungsanforderungen haben können. Im folgenden werden einige Methoden zur Vorverarbeitung vorgestellt.

2.2.1 Vorverarbeitungsschritte für Clustering-Verfahren

Ein Problem mit Datensätzen, die aus Merkmalen verschiedener Sensoren aufgenommen wurden, kann sein, dass diese unterschiedliche Einheiten oder Wertebereiche nutzen. Hierdurch kann es zu einer Gewichtung einzelner Merkmale kommen. Um dieses Verhalten zu minimieren, können die Daten normalisiert und standardisiert werden. [24] Für die Standardisierung werden die Werte merkmalsweise verändert, indem der Mittelwert entfernt und auf die Einheitsvarianz skaliert wird. So können unterschiedlich verteilte Merkmale besser verglichen werden. Ziel der Normalisierung ist es, die Werte der einzelnen Merkmale in einen gemeinsamen Wertebereich zu übertragen, zumeist zwischen 0 und 1. [11] zeigt, dass eine Normalisierung genutzt werden kann, um Intensitätsschwankungen durch geometrische Eigenschaften innerhalb der Objekte zu minimieren. Trebuna et al. kommen zu dem Schluss, dass diese Vorverarbeitungsmethoden sinnvoll für eine Cluster-Analyse sein können. Jedoch können diese auch zu einer Verschlechterung der Cluster-Ergebnisse führen. [24]

Verschiedene Quellen [22, 11, 13] zeigen die Problematik einer hohen Dimensionalität in den Daten auf. Durch mehr Dimensionalität steigen der Speicherbedarf und Berechnungsanforderungen für Ähnlichkeitsberechnungen. [11] Ein weiteres Problem mit hoher Dimensionalität ist, dass eine feste Anzahl von Datenpunkten mit zunehmender Dimensionalität immer "spärlicher" wird. Hierbei wird auch vom "Fluch der Dimensionalität" gesprochen. Für Cluster-Analysen ist dieser Effekt an der Ähnlichkeitsberechnung, welche zumeist über die Distanz berechnet wird, zu erkennen. Der Unterschied zwischen der kleinsten und größten Distanz innerhalb der Daten wird mit zunehmender Dimensionalität kleiner oder geht sogar gegen null. [22]

Für dieses Problem kann als Vorverarbeitungsmethode eine Dimensionsreduktion vorgenommen werden. Eine einfache Methode wäre, lediglich verschiedene Merkmale auszuwählen, die für die Clustering-Analyse genutzt werden. Jedoch fehlt für die Auswahl zumeist Vorwissen, sodass relevante Merkmale ausgeschlossen werden könnten. [13] beschreibt verschiedene Dimensionsreduktionsmethoden. Durch diese Methoden werden die Daten in einer verringerten Dimensionalität dargestellt. Ein Beispiel hierfür ist die PCA. Die Hauptidee hinter der Hauptkomponentenanalyse besteht darin, eine Rotation im Hyperraum \mathbb{R}^{n_D} zu finden, die die maximale Streuung der Datenproben in den Hauptachsen erfasst. Basierend auf dieser Rotation wird die Daten auf einen niedriger dimensionierten Raum \mathbb{R}^{n_P} projiziert, der als Hauptunterraum bekannt ist, indem nur die ersten n_P Vektor-Koeffizienten erhalten bleiben [11]. Diese Methode ist weit verbreitet und wird auch für die mutlispektrale Analyse eingesetzt [11]. Journaux et al. kommen

in ihrer Untersuchung [13] zu dem Schluss, dass die PCA, obwohl sie nur lineare Zusammenhänge berücksichtigt, gute Ergebnisse liefert.

Bei großen Datensätzen kann es sinnvoll sein, diese zu reduzieren, da beispielsweise die Speicherkapazität oder die Rechenleistung nicht ausreicht. [4] Um die Daten zu reduzieren, kann beispielsweise eine kleinere (repräsentative) Probe ausgewählt werden. Jedoch ist die Bestimmung dieser, schwierig und ein großer Zusammenhang kann verloren gehen. [6] beschreiben eine Möglichkeit große Bilddaten mithilfe von einer pyramidsichen Skalierung zu Clustern. Hierfür wird das Ursprungsbild durch die Berechnung des pyramidsichen Raums skaliert. Für jedes Level der Pyramide wird zunächst ein Gauß-Filter angewendet und anschließend die Auflösung geviertelt. Die Cluster-Analysen werden zunächst auf einem hohen Level der Pyramide, also einem runterskalierten Bild durchgeführt. Anschließend werden repräsentative Pixel jedes gefundenen Cluster ausgewählt und auf das nächst kleinere Level der Pyramide übertragen. Diese Schritte werden solange durchgeführt, bis die ursprüngliche Auflösung des Bildes wieder hergestellt ist. Auch in [4] wird ein großer Datensatz in Kleinere aufgeteilt. Hierfür wird ein "mod 2Skalierer verwendet, der eine Datensatz X in $X_{mod2=1} = x_1, x_3, \dots, x_{2m}$ und $X_{mod2=0} = x_2, x_4, \dots, x_{2m}$ mit $m = n/2$ aufteilt. Auch die partitionierten Datensätze können weiter aufgeteilt werden, um diese weiter in ihrer Größe zu halbieren. Anschließend werden die Clustering-Verfahren auf den Partitionierungen des Datensatzes durchgeführt, um die Cluster-Zentren für den ursprünglichen Datensatz zu verwenden. Cluster, die in verschiedenen Partitionierungen dicht bei einander liegen, werden zusammengeführt. Hierdurch entsteht ein Clustering für den großen Datensatz. [4]

Jordan et al. präsentieren in [12] ein Verfahren mit Superpixeln. Anstatt jedes Pixel eines multispektralen Bildes zu clustern, werden diese auf Superpixeln berechnet. Superpixel beschreiben dabei homogene Bereiche im Bild, die eine starke Ähnlichkeit haben. Dabei sind die Pixel in den Superpixeln verbunden. Hierdurch wird die Gitterstruktur des Bildes ersetzt. [12]

2.2.2 Evaluation

Verschiedene Verfahren wurden entwickelt, um Cluster-Analysen zu evaluieren. Da für die in dieser Thesis benutzen Daten keine zusätzlichen Informationen, wie einem ground truth, vorhanden sind, werden diese Metriken nicht betrachtet. Die Evaluation erfolgt somit auf dem berechneten Clustering-Model selbst. Hier können verschiedene Eigenschaften, wie die Ähnlichkeit innerhalb und zwischen Clustern genutzt werden. Der

Silhouetten-Koeffizient von [20] ist eine solche Metrik. Dieser misst die Ähnlichkeit zwischen den Clustern und auch die Ähnlichkeit der Objekte innerhalb der einzelnen Cluster. Gebildet wird er aus dem Durchschnitt der Silhouetten-Koeffizienten der einzelnen Punkte (s), die wie folgt berechnet werden:

$$s = \frac{b - a}{\max(a, b)} \quad (2.2)$$

Hierbei ist a die durchschnittliche Intra-Cluster-Distanz und b die durchschnittliche Distanz zum nächstgelegenen Cluster. Der Silhouetten-Koeffizient reicht von -1 bis +1. Hohe Werte bedeuten, eine große Ähnlichkeit innerhalb der Cluster und eine große Abweichung zwischen den Clustern, also eine großen Trennung. Ein Nachteil für die Berechnung ist, dass alle Paarweisen-Distanzen berechnet werden müssen, was bei großen Datensätzen zu einem hohen Aufwand wird. [20]

Eine weitere Metrik für die Evaluation von Clustering-Verfahren, ist der Calinski-Harabasz Index. Der Index misst, wie gut die Daten in Cluster eingeteilt sind, indem er die Summe der Abweichungen innerhalb und zwischen den Clustern vergleicht, wobei die Abweichung als Summe der Quadrate der Abstände berechnet wird. Dichte und gut separierte Cluster ergeben einen höheren Index. [3]

[7] beschreiben den Davies-Bouldin Index. Dieser Index stellt die durchschnittliche Ähnlichkeit zwischen den Clustern dar, wobei die Ähnlichkeit als ein Maßstab definiert wird, der den Abstand zwischen den Clustern im Verhältnis zu ihrer Größe bewertet. Der niedrigste erreichbare Wert ist Null. Werte, die sich näher an Null befinden, deuten auf eine effektivere Partitionierung hin. [7]

2.2.3 MS Clustering

Verschiedene Clustering-Verfahren werden für die Untersuchung von multispektralen Daten genutzt. [6] benutzen für ihren Ansatz auf den multispektralen Daten K-Means als Verfahren. Hiermit werden gute Ergebnisse erzielt. [12] nutzt eine Variation auf den Mean-Shift Algorithmus. Dieser wird auf zuvor berechneten Superpixeln angewendet. Zudem ist Mean-Shift für eine Bildsegmentierung weit verbreitet. [11]

Im Vorfeld dieser Thesis wurden bereits erste Versuche mit Clustering-Verfahren auf multispektralen Bilddaten durchgeführt. [14] Hier wurden verschiedene Verfahren auf einer kleinen Datenprobe von circa 70.000 Pixelwerten eines 15-Dimensionalen multispektralen Bildes durchgeführt. Für die Daten waren zum Teil die ground truth Label bekannt,

sodass eine erste Einschätzung verschiedener Clustering-Verfahren möglich ist. Auch wurde der Einfluss einer Standardisierung und einer Dimensionsreduktion durch eine PCA untersucht. Die höchste Genauigkeit kann bei standardisierten Daten mit dem K-Means festgestellt werden. Mean-Shift und HDBSCAN erzielen eine geringe Genauigkeit. Eine Untersuchung der Clustering-Verfahren unabhängig von den ground truth Daten wurde nicht durchgeführt, sodass keine Aussagen über Ähnlichkeiten zwischen und innerhalb der Cluster getroffen werden können. Zudem wurde nur eine kleine Probe an Daten genutzt.

Darauf aufbauend werden für diese Thesis die Untersuchungen fortgesetzt. Die verschiedenen Vorverarbeitungsmethoden, Standardisierung, Skalierung und Dimensionsreduktion durch eine PCA werden zusammen mit den Clustering-Verfahren K-Means, Mean-Shift und HDBSCAN für die Anwendung auf multispektralen Bilddaten verglichen.

3 Methodik

In diesem Kapitel wird zunächst kurz erläutert, wie die Datensätze aufgenommen wurden. Anschließend wird das erarbeitete Verfahren für die multispektrale Bildregistrierung und die vorgehensweise für die Clustering-Analyse vorgestellt.

3.1 Datenaufnahme

Die in dieser Arbeit verwendeten Datensätze wurden mit einem multimodalen Sensoransatz aufgenommen. Es werden drei verschiedene Kameras verwendet, welche in unterschiedlichen spektralen Bereichen Bilddaten aufnehmen können. Um eine feinere spektrale Aufteilung zu erhalten, werden zusätzlich verschiedene spektrale Bandpassfilter verwendet, welche nur einen kleinen Teil des Lichtspektrums durchlassen. Diese werden vor die Objektive der Kamera positioniert. Die Filter werden entweder manuell oder mithilfe eines Filterrads gewechselt. 3.1 zeigt die verwendete Hardware und die Filter-Kamera-Kombinationen. Bei dem Filter mit der ID 2 handelt es sich um einen Dual-Bandpassfilter. Dieser lässt zwei kleine Bereiche des Lichtspektrums durch. Mit diesem Setup werden sequenziell zwölf verschiedene Bilder in unterschiedlichen Spektren aufgenommen. Durch die Aufteilung des RGB-Bildes (ID 3) werden daraus 15 Kanäle eines multispektralen Bilds. Die Kameras sind so ausgerichtet, dass sich die Blickfelder möglichst stark überschneiden, sodass möglichst viele Objekte pro Szene aufgenommen werden können. Zudem sind die Sensor-Ebenen möglichst parallel ausgerichtet, um unnötige Verzerrungen zu vermeiden. Für die Aufnahmen unter Laborbedingungen werden die Kameras senkrecht zur Arbeitsfläche positioniert. Eine Beleuchtung dieser findet über einen Halogenstrahler statt, da dieser Licht im Spektralbereich von UV bis SWIR ausstrahlt. Da sich die Blickfelder und die Auflösung der Kameras unterscheiden, müssen die einzelnen Bilder im Vorfeld zugeschnitten werden. Um bei diesem Crop-Vorgang keine Verzerrungen zu

erzeugen, werden die Bildausschnitte mit gleichem Seitenverhältnis zugeschnitten. Anschließend werden diese Bildausschnitte auf das gleiche Format von 1280 * 1080 Pixel gebracht. Für die Szenarien werden im ersten Datensatz verschiedene Müllobjekte genutzt.

Kamera	Spektralbereich[nm]	ID
PCO.ultraviolett	190 - 1100	0
Objektiv: 25mm f/2.5-16	290 - 365	1
FoV: 14.8° x 11.1°	375 - 425	2
Pixel Format: 1392 x 1040	745 - 970	
Basler Ace 2	400 - 1000	3
Objektiv: 12.5mm f/1.4-16	730 - 755	4
FoV: 30.4° x 19.0°	830 - 865	5
Pixel Format: 1920 x 1200	845 - 930	6
	928 - 955	7
Allied Vision Goldeye G-130	400 - 1700	8
Objektiv: 16mm f/1.4-16	930 - 1030	9
FoV: 22.9°x18.3°	1290 - 1310	10
Pixel Format: 1280 x 1024	1440 - 1460	11
	1485 - 1645	12

Tabelle 3.1: Übersicht über verwendete Sensorik und Spektralfilterkombinationen

Im zweiten Datensatz werden verschieden reife Erdbeeren aufgenommen. Abbildung 3.1 zeigt beispielhaft eine multispektrale Szene mit allen Einzelkanälen.

3.2 Registrierung

Da es sich bei dem hier verwendeten multispektralen Kamerasystem um einen multimodalen Aufbau handelt, werden im vorgeschlagenen Ansatz zur Registrierung verschiedene zusätzliche Informationen mit berücksichtigt.

Um möglichst viele Merkmalspunkte zu finden, werden die einzelnen Bilder, wie bereits bei der vorherigen Arbeit, normiert. [15] Als weiterer Schritt der Vorverarbeitung werden einige Bilder morphologisch transformiert, um Unreinheiten und eine grobe Körnung auf dem Bild zu verringern. Um diese Verunreinigungen zu minimieren wurde eine morphologische Öffnung benutzt, welche eine Erosion, gefolgt von einer Dilation, durchführt.

Die vorgeschlagene Methode für die Registrierung ist in zwei Stufen aufgebaut. In der ersten Stufe werden die Bilder jeder einzelnen Kamera registriert. Hierfür wird ein Bild

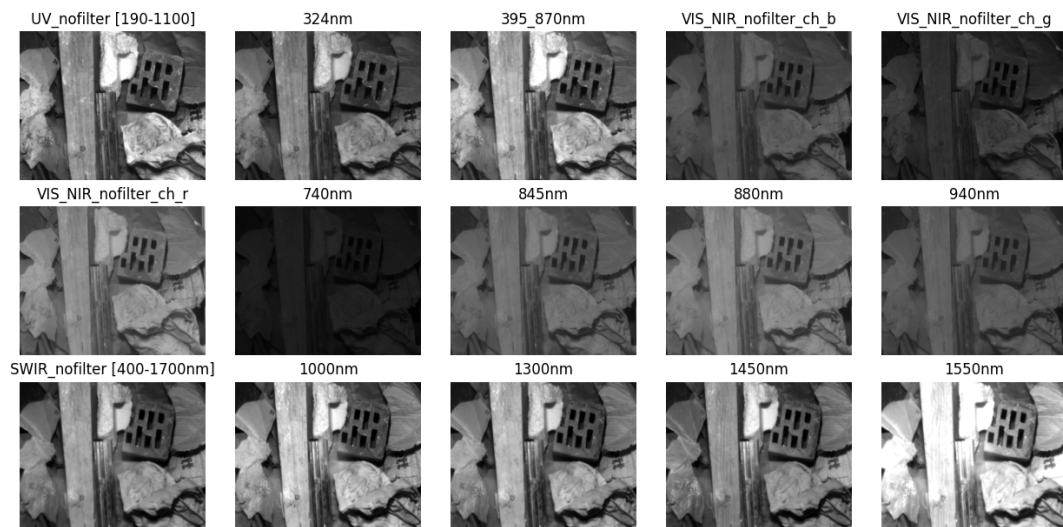


Abbildung 3.1: Einzelbilder des multispektralen Datensatzes

pro Kamera als Zielbild ausgewählt. Anschließend werden SIFT Merkmalspunkte für jedes Bild gesucht. Nach einem Matching mithilfe von einem Nearest-Neighbor wird zunächst der von Lowe et. al. vorgestellte Verhältnistest durchgeführt, um falsche Matches auszusortieren. Da bei den Bildern der gleichen Kamera davon ausgegangen werden kann, dass sich die Bilder nur minimal durch chromatische Aberration oder Verschiebungen bei Filterwechseln unterscheiden, wird ein zusätzlicher Distanzfilter für die Matches eingeführt. Dieser vergleicht die Positionen der gematchten Merkmalspunkte zweier Bilder und entfernt die Matches, bei der die Differenz zwischen den Koordinaten größer als ein Schwellwert ist. Hierdurch können weitere Matches ausgeschlossen werden, die zu einer zu großen Rotation, Translation und Skalierung führen können. Aus den verbliebenen Matches wird eine partielle affine Transformation als Grundlage verwendet. Diese lässt nur die bereits oben genannten geometrischen Transformationen zu. Dadurch bleiben parallele Linien parallel. Eine partielle affine Transformationsmatrix für zweidimensionale Punkte lässt sich wie folgt darstellen.

$$\begin{bmatrix} \cos(\theta) * s & -\sin(\theta) * s & t_x \\ \sin(\theta) * s & \cos(\theta) * s & t_y \end{bmatrix}$$

Hierbei ist θ der Rotationswinkel, s der Skalierungsfaktor und t_x und t_y die Translationen in x - und y -Richtung.

Diese berechnete Transformationsmatrix wird anschließend genutzt, um die Bilder der Kameras zu transformieren. So sind die Bilder pro Kamera registriert, sodass in der zweiten

Stufe jeweils nur eine Transformation für die Bilder eines Kamerapaars gefunden werden muss.

Für diesen Schritt wird zunächst eine Zielkamera bestimmt. Aus den Bildern dieser Kamera wird ein Zielbild ausgewählt, sodass die anderen Kameras auf das selbe Zielbild registriert werden. Das Zielbild wird so gewählt, dass möglichst viele nutzbare Matches zwischen Bildern beider anderer Kameras gefunden werden können. Die Quellbilder der anderen Kameras werden nach der Anzahl der Matches zum Zielbild ausgewählt. Mit den Matches wird pro Kamerapaar eine Transformationen berechnet. Da die Bilder mit unterschiedlichen Sensoren und Optiken aufgenommen werden, wird hierbei ein perspektivisches Transformationsmodell verwendet. Die Methode wird in Python 3.11 mit der Bibliothek Open Computer Vision Library (OpenCV)¹ implementiert. OpenCV ist eine open-source Bildverarbeitungsbibliothek, die in verschiedenen Programmiersprachen verfügbar ist. Sie bietet verschiedene Implementierung für Merkmalsextraktion und Matching. Für diese Thesis wird die Methode SIFT verwendet, welche über das *OpenCV Contrib package* verfügbar ist. Als Matcher wird die ebenfalls in OpenCV eingebundene *Fast Library for Approximate Nearest Neighbors (FLANN)* genutzt. Für die Transformationsberechnung in der ersten Stufe wird die Methode *estimateAffinePartial2D*² benutzt, während für die zweite Stufe die Methode *findHomography*² genutzt wird. Die Berechnungen werden RANSAC-basierend durchgeführt, wodurch eine robuste Annäherung der Transformationen gefunden wird.

3.3 Clustering-Verfahren

Ein großes Problem bei der Cluster-Analyse von multispektralen Bilddaten ist die Größe der Datensätze. Wenn die einzelnen Pixel als Datenpunkte genutzt werden, ergeben sich pro spektralen Kanal $Bildbreite * Bildhöhe = Anzahl_{Pixel}$. Bei Bildern mit einer Auflösung von $1280 * 1024$ führt das zu 1310720 Pixeln pro Bildkanal. Ein weiteres Problem von multispektralen Bilddaten kann die hohe Dimensionalität sein, welche bereits in Kapitel 2 beschrieben wurde. Da die in dieser Thesis verwendeten multispektralen Bilddaten mit verschiedenen Sensoren aufgenommen wurden, können die Intensitätswerte stark variieren. Dadurch können einige Kanäle mehr Gewichtung bei der Cluster-Analyse erhalten. Um dieses Verhalten zu beheben können Verfahren der Standardisierung und Normierung angewendet werden. Hierdurch werden die Daten-Kanalweise zunächst normiert,

¹<https://opencv.org/>

²https://docs.opencv.org/3.4/d9/d0c/group__calib3d.html

P1,1	P1,2	P1,3	P1,4	P1,5	P1,6	P1,7	P1,8	P1,1	P1,5	P1,2	P1,6	P1,3	P1,7	P1,4	P1,8
P2,1	P2,2	P2,3	P2,4	P2,5	P2,6	P2,7	P2,8	P5,1	P5,5	P5,2	P5,6	P5,3	P5,7	P5,4	P5,8
P3,1	P3,2	P3,3	P3,4	P3,5	P3,6	P3,7	P3,8	P2,1	P2,5	P2,2	P2,6	P2,3	P2,7	P2,4	P2,8
P4,1	P4,2	P4,3	P4,4	P4,5	P4,6	P4,7	P4,8	P6,1	P6,5	P6,2	P6,6	P6,3	P6,7	P6,4	P6,8
P5,1	P5,2	P5,3	P5,4	P5,5	P5,6	P5,7	P5,8	P3,1	P3,5	P3,2	P3,6	P3,3	P3,7	P3,4	P3,8
P6,1	P6,2	P6,3	P6,4	P6,5	P6,6	P6,7	P6,8	P7,1	P7,5	P7,2	P7,6	P7,3	P7,7	P7,4	P7,8
P7,1	P7,2	P7,3	P7,4	P7,5	P7,6	P7,7	P7,8	P4,1	P4,5	P4,2	P4,6	P4,3	P4,7	P4,4	P4,8
P8,1	P8,2	P8,3	P8,4	P8,5	P8,6	P8,7	P8,8	P8,1	P8,5	P8,2	P8,6	P8,3	P8,7	P8,4	P8,8

Tabelle 3.2: Funktion des Zerlegungsalgorithmus: Ausgangsmatrix (links) mit einer Auflösung von $(b * h)$, aufgeteilt in n^2 Teilmatritzen (rechts) mit einer Auflösung von $(\frac{b}{n} * \frac{h}{n})$ und $n =$ Skalierungsfaktor.

und anschließend in einen Wertebereich um 0 gebracht. Damit liegen alle Kanäle im gleichen Wertebereich. Wie [24] zeigt, können diese Verfahren für einige Datensätze zu einer Verschlechterung der Cluster-Analyse führen.

Um diese Probleme zu untersuchen, werden die genutzten Datensätze vor der Cluster-Analyse mit verschiedenen Vorverarbeitungsmethoden in unterschiedliche Varianten unterteilt. Als erste Variante werden die Rohdaten verwendet, um den Einfluss der Vorverarbeitungsmethoden vergleichen zu können. Die zweite Variante verändert die Rohdaten durch eine kanalweise Standardisierung. Um die Dimensionalität zu reduzieren werden die Rohdaten für die Varianten drei und vier zunächst durch eine Hauptkomponentenanalyse umgewandelt. Anschließend wird das erklärte Varianzverhältnis (explained variance ratio) berechnet. Dieses gibt eine Einschätzung darüber, wie viel Prozent der Gesamtvarianz durch jede der Hauptkomponenten erklärt wird. Sie ist eine entscheidende Hilfe, um die Anzahl der Hauptkomponenten zu bestimmen, die verwendet werden sollen. Dafür wird die kumulative Summe genutzt, bis eine gewünschte Gesamtvarianz erreicht ist. Für die Variante drei werden so viele Hauptkomponenten genutzt, dass die Gesamtvarianz über 99% beträgt, während Variante vier 95% der Gesamtvarianz abbilden.

Um die großen Datenmengen verarbeiten zu können, wird in dieser Thesis ein Zerlegungsalgorithmus vorgestellt. Eine Zerlegung wird beispielhaft in Tabelle 3.2 dargestellt. Ein Ausgangsbild, dargestellt als Matrix wird so aufgeteilt, dass die Pixel ($p_{zeile,spalte}$) jeder n -ten Zeile und Spalte ein neues Teilbild erzeugen. Dabei ist n ein wählbarer Skalierungsfaktor. Somit besteht ein Teilbild aus den Pixeln der Zeilen und Spalten $x, x + n, x + 2n, \dots$. Die Auflösung der Teilbilder ist somit um den Faktor n^2 kleiner. Die einzelnen Pixelwerte werden nicht verändert, was zum Teil bei anderen Bildskalierungen durch Interpolation verschiedener Pixelwerte geschehen kann. Die Szene in den Teilbildern unterscheidet sich nur gering von der vom Ursprungsbild. Es bleiben die

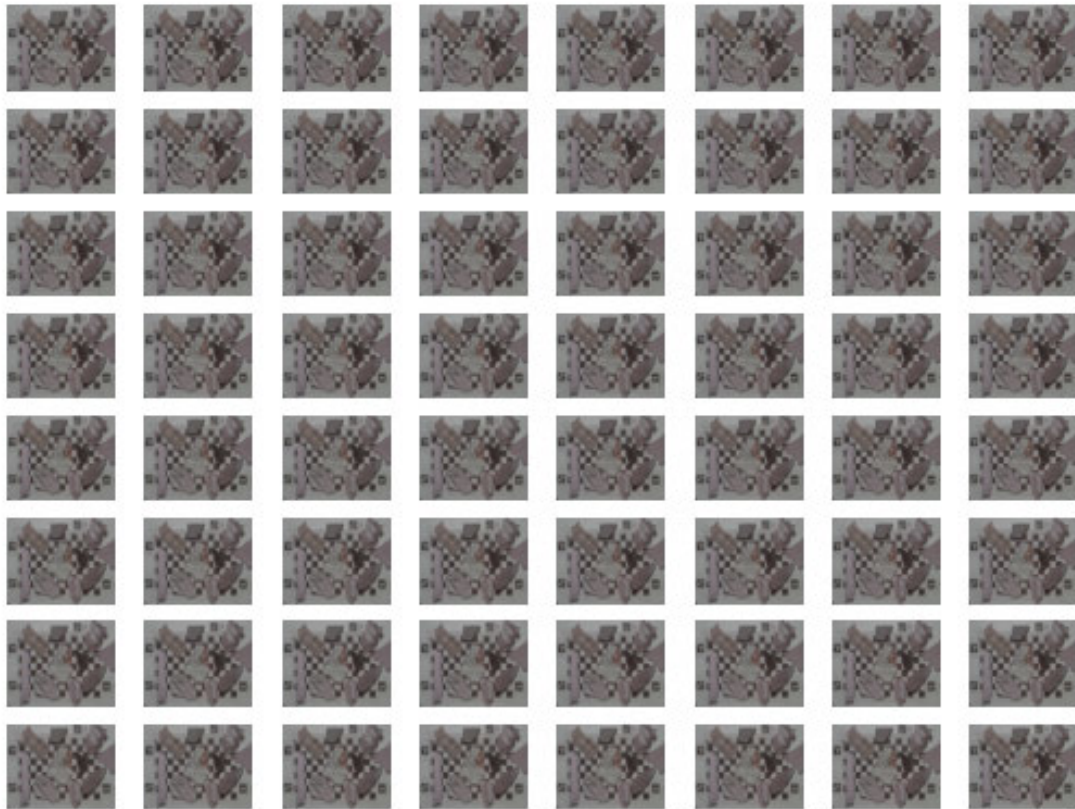


Abbildung 3.2: Teilbilder nach Zerlegung eines RGB-Bilds

Strukturen erhalten. Abbildung 3.2 zeigt die Zerlegung eines Bildes, mit einem Faktor 8. Die Cluster-Analyse wird anschließend sequentiell auf den Teilbildern durchgeführt. Hierdurch erhält man für jedes Teilbild ein Cluster-Ergebnis. Durch eine einfache Umkehrung des Zerlegungsalgorithmuses kann jedoch kein Gesamtergebnis erzeugt werden, da die unterschiedlichen Teil-Clusterings gegebenenfalls verschiedene Label für ähnliche Cluster vergeben haben. Clustering-Verfahren geben einfache Label, die keiner Klassifizierung entsprechen. Deshalb müssen die Label der Teil-Cluster zugeordnet werden. Hierfür werden die Cluster-Zentren betrachtet. Jedes Label L_x aus einem Teil-Clustering x wird einem Label L_a eines Ausgangs-Clusterings a zugeordnet, bei dem Ähnlichkeit der Cluster-Zentren am größten ist. Als Ähnlichkeit wird die euklidische Distanz verwendet. Eine kleinere Distanz ergibt eine höhere Ähnlichkeit. So werden sequentiell alle Label angepasst. Als Ausgangs-Clustering wird das Teil-Clustering mit den meisten Cluster-Zentren verwendet. Aufbauend auf dieser Zuordnung können die Cluster-Zentren noch einmal angepasst werden. Hierfür werden die Durchschnitts-Cluster-Zentren berechnet.

Anschließend werden alle Datenpunkte aus dem Ursprungsbild einem Cluster aufgrund der euklidischen Distanz zugeordnet. Diese Skalierung wird auf alle Varianten der anderen Vorverarbeitungsmethoden angewendet. Für die Clustering-Analyse werden die Clustering-Verfahren K-Means, HDBSCAN und Meanshift verwendet.

4 Ergebnisse

Um verschiedene Gebiete der multispektralen Bilduntersuchung zu nutzen, werden die zwei in Kapitel 3 beschriebenen Datensätze verwendet. Bei dem ersten Datensatz handelt es sich um Aufnahmen von verschiedenen Objekten aus dem Sperrmüll nach einer groben Zerkleinerung, die sich im Material unterscheiden. Die Objekte im zweiten Datensatz sind unterschiedlich reife geflückte Erdbeeren, von unreif bis überreif. So werden die Sektoren Recycling und Landwirtschaft in den Datensätzen abgedeckt. Pro Datensatz werden verschiedene Szenen benutzt. Ausgewählte Ergebnisse aus der Registrierung werden anschließend für die Clustering-Analyse verwendet.

4.1 Registrierung

Das in Kapitel 3 beschriebene zweistufige Verfahren für die multispektrale Bildregistrierung wird mit verschiedenen Parametern und Einstellungen auf die verschiedenen Szenen der beiden Datensätze angewendet. Folgende Metriken und Statistiken werden für die Evaluation berechnet.

1. RMSE von den paarweisen Differenzbildern
2. RMSE von den paarweisen Positionen der Matches
3. Anzahl an Matches vor Filterung
4. Anzahl an Matches nach Filtern

Die erste Metrik ist eine Methode, um Bildähnlichkeit zu messen. Hierfür werden zunächst die paarweisen Differenzbilder berechnet. Dadurch erhält man die Intensitätsdifferenz von jedem korrespondierenden Pixelpaar. Diese werden anschließend quadriert, summiert und durch die Gesamtanzahl geteilt. Als letzter Schritt wird die Quadratwurzel der mittleren

quadratischen Abweichung (MSE) berechnet. Formel (4.1) zeigt die Berechnung, mit $n =$ Anzahl der Pixelpaare und $y', y =$ Intensitätswerte des ersten und zweiten Bildes.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{t=n} (y' - y)^2} \quad (4.1)$$

Bei der zweiten Metrik wird auch ein RMSE berechnet. Hier wird die Differenz $y' - y$ aus (4.1) zwischen den Positionen der gematchten Merkmalspunkte eines Bildpaares berechnet. Dieser Fehler kann als pixelweise Abweichung interpretiert werden.

Da in der Literatur die Anzahl der gefundenen und korrekten Matches wichtig ist, werden diese als Metriken genutzt. Bei der dritten Metrik handelt es sich um alle gefundenen Matches, während die vierte Metrik die Anzahl der Matches nach dem Ratio-Test von Lowe und einem möglichen Differenzfilter beinhaltet. Durchschnittswerte werden für jede Kamera-Kombination sowie der Gesamtszene berechnet. Für die Metriken 2-4 werden die bildweisen Matches unter Verwendung von SIFT und einem flannbased Matcher berechnet. Zusätzlich wird die in [15] vorgestellte zielkanals-basierte Methode als Vergleich herangezogen. Bei der Methode wird ein Bild als Zielbild für alle anderen Bilder ausgewählt. Um einen weiteren Vergleichswert zu berechnen, werden die Evaluationsmetriken auch auf den nicht registrierten Bildern berechnet.

4.1.1 Sperrmülldatensatz

Für die erste Untersuchung wird der Datensatz mit verschiedenen, zerkleinerten Sperrmüll-Objekten verwendet. Hierfür wurden zehn verschiedene Szenen ausgewählt und die Registrierung durchgeführt. Anschließend wurden auf den registrierten Bildern die Merkmalspunkte und Matches erneut berechnet. Tabelle 4.1 zeigt die durchschnittliche Anzahl an gefundenen Matches, gegenübergestellt mit der Anzahl an Matches nach der Filterung mit dem Ratio-Test von Lowe von den einzelnen Registrierungsmethoden. Für jede Kamera-Kombination wurde der Durchschnitt für alle Kanal-Kombinationen berechnet. Kanal-Kombinationen innerhalb einer Kamera sind in Zeile 1-3 zu sehen, während die Kombinationen der Kameras in Zeile 4-6 gezeigt wird. Zusätzlich zeigt die Tabelle den Gesamtdurchschnitt der einzelnen Methoden. Die erste Spalte zeigt die Ergebnisse für die Berechnung auf den nicht registrierten Bilddaten (roh). Es werden durchschnittlich mehr Merkmalspunkte gefunden. Jedoch ist hier die größte Filterung zu beobachten. Somit ergeben sich am wenigsten nutzbare Merkmalspunkte. Die Registrierung durch die

Anzahl Matches	roh		zielbasiert		zwei stufig zwei Filter		zwei stufig ein Filter	
avg.	gef.	Filter	gef.	Filter	gef.	Filter	gef.	Filter
UV	758	337	706	406	702	404	703	405
VIS_NIR	391	189	420	108	424	160	431	162
SWIR	863	313	781	229	731	299	739	300
UV-VIS_NIR	437	4	467	81	475	82	482	87
UV-SWIR	800	37	731	75	690	74	696	75
VIS_NIR-SWIR	800	11	730	62	690	76	696	72
total	675	148	639	160	619	183	624	184

Tabelle 4.1: Durchschnittliche Anzahl an Matches vor und nach Ratio-Test für die Müll-
daten

zielbasierte Methode ergeben insgesamt weniger Matches. Lediglich innerhalb der VIS_NIR-Kamera und in Kombination dieser mit der UV-Kamera werden mehr Matches gefunden. Allerdings werden bei der ersteren mehr Matches aussortiert, sodass nur circa 26% nutzbar sind. Allgemein werden weniger Matches im Vergleich zu *nicht registriert* aussortiert, sodass eine größere Anzahl an Merkmalspunkten nutzbar ist. Bei den Spalten drei (zweistufig, zwei Filter) und vier (zweistufig, ein Filter) handelt es sich um Varianten von dem in Kapitel 3 vorgestellten zweistufigen Verfahren. In Spalte drei wird ein Distanzfilter in beiden Stufen eingesetzt. Hierdurch wurden bei der Registrierung Matches aussortiert, bei der die Distanz in X - und Y -Richtung zwischen den Merkmalspunkten eines Matches einen Schwellwert überschritt. Ziel dieses Filters ist es, eine starke Rotation zu verhindern. Bei der Registrierung für Spalte vier wurde der Distanzfilter nur für die erste Stufe eingesetzt. Beide Varianten des zweistufigen Verfahrens erzeugen eine ähnliche durchschnittliche Anzahl an Matches. Auch ist die Anzahl an nutzbaren Matches sehr ähnlich. Im Vergleich werden bei diesen Registrierungsmethoden zwar die wenigsten Matches gefunden, jedoch werden am wenigsten aussortiert. Das führt zu den meisten nutzbaren (korrekten) Matches. Auffällig ist, dass für alle Varianten deutlich weniger Matches für die Kombinationen akzeptiert werden, sodass das Verhältnis zwischen gefunden und gefilterten Matches sehr gering ist. Innerhalb der Kameras werden mehr Matches für korrekt befunden und behalten.

Tabelle 4.2 ist ähnlich aufgebaut, wie Tabelle 4.1. In dieser werden die durchschnittlichen Ergebnisse der RMSE-Berechnungen dargestellt. Metrik 1 findet sich in den Spalten *Bild*. Wie zu erwarten, ist diese zwischen Bildern unterschiedlicher Kameras höher, als zwischen Bildern einer Kamera. Je kleiner der RMSE, desto ähnlicher sind die Bilder. Im

RMSE	nicht registriert		zielbasiert		zwei stufig zwei Filter		zwei stufig ein Filter	
	Bild	Matches	Bild	Matches	Bild	Matches	Bild	Matches
avg.								
UV	4.66	4.77	3.27	1.14	3.16	1.09	3.16	1.09
VIS_NIR	4.00	1.78	5.88	4.26	5.31	2.03	5.48	2.12
SWIR	3.64	2.73	4.22	4.17	3.42	1.55	3.43	1.53
UV-VIS_NIR	6.78	9.14	5.61	3.63	5.49	3.88	5.48	3.59
UV-SWIR	5.57	10.35	5.51	3.56	5.32	3.54	5.28	3.38
VIS_NIR-SWIR	5.96	10.89	6.72	5.05	6.47	5.07	6.42	5.05
total	5.10	6.61	5.20	3.64	4.86	2.86	4.87	2.79

Tabelle 4.2: Durchschnittlicher RMSE Bild- und Matchesweise für den Mülldatensatz

Vergleich ist der Bildweise-RMSE bei den zweistufigen Methoden am niedrigsten. Bei der zielbasierten Methode liegt er höher, als bei den nicht registrierten Bildern *nicht registriert*. Vorallem Kombinationen mit Bildern der VIS_NIR-Kamera haben einen höheren RMSE. Jedoch ist der Unterschied des Gesamtdurchschnitts für die einzelnen Registrierungsmethoden ehr gering.

Metrik 2, also der RMSE zwischen den Matches, findet sich unter *Matches*. Bei der Berechnung wurden nur die korrekten Matches genutzt. Auch ist es vorgekommen, dass für einige Bildpaare, keine korrekten Matches gefunden wurden. Die Durchschnittswerte wurden ohne diese Bildpaare berechnet. Die besten Ergebnisse, also möglichst kleine Werte, können bei den Varianten der zweistufigen Registrierungsmethode beobachtet werden. Das sich diese stärker in den Kombinationen von Bildern unterschiedlicher Kameras unterscheiden, ist zu erwarten gewesen, da sich die Berechnungen der Registrierung innerhalb einer Kamera nicht unterscheiden. Durch den zusätzlichen Filter in der zweiten Stufe ergibt sich ein größerer Fehler. Auch die zielbasierte Registrierung führt zu niedrigeren Werten, im Vergleich zu den Ergebnissen auf den nicht registrierten Bilddaten. Allgemein lässt sich ein größerer Unterschied zwischen den einzelnen Methoden feststellen.

Um eine optische Einschätzung der verschiedenen Registrierungsmethoden zu erhalten, zeigt Abbildung 4.1 beispielhaft Falschfarbenbilder, wie sie auch in [15] benutzt wurden. Hierbei werden alle einzelnen Kanäle in drei Kanäle fusiniert. Man erhält einen Eindruck von der Überlagerung der einzelnen Kanäle. Es ist zu sehen, dass das nicht registrierte Bild eine starke Verzerrung zwischen den Farbkanälen aufweist. Vor allem sind diese Unterschiede an Kanten der Objekte sichtbar. Bei den registrierten Bildern sind lediglich



(a) keine Registrierung (b) zielbasiert (c) 2-stufig-zwei Filter (d) 2-stufig-ein Filter

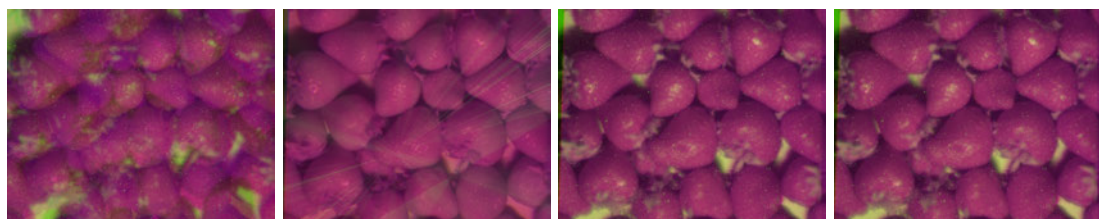
Abbildung 4.1: Falschfarbenbilder der Registrierungsmethoden von einer Szene aus dem Sperrmülldatensatz

leichte Verschiebungen zwischen den Farbkanälen zu sehen. Es lassen sich optisch keine deutlichen Unterschiede zwischen den Registrierungsmethoden feststellen.

4.1.2 Erdbeerdatensatz

Für die zweite Untersuchung werden zehn verschiedene Szenen mit unterschiedlich reifen Erdbeeren ausgewählt und registriert. Abbildung 4.2 zeigt die Falschfarbenbilder der Registrierungsmethoden. Auch hier kann an Bild (a) erkannt werden, dass eine Registrierung notwendig ist. Die einzelnen Farbkanäle sind stark verschoben. Auffällig in den Bildern der Registrierungsmethoden (b-d) ist die obere linke Bildkante. Hier ist eine deutliche Verschiebungen zu erkennen. In (b) sind feine Linien im grünen Farbkanal zu erkennen. Diese lassen auf stark verdrehte Einzelbilder schließen, sodass diese Registrierung zu einem unerwünschten Ergebnis führt. Andere Auffälligkeiten sind nicht zu erkennen.

Bei der Anzahl an Matches, dargestellt in Tabelle 4.3, ist eine große Anzahl an gefundenen Matches in Bildern der SWIR-Kamera zu beobachten. Jedoch werden sehr viele der Matches aussortiert. Auf den nicht registrierten Bildern werden zwischen den Bildern der



(a) keine Registrierung (b) zielbasiert (c) 2-stufig-zwei Filter (d) 2-stufig-ein Filter

Abbildung 4.2: Falschfarbenbilder der Registrierungsmethoden von einer Szene aus dem Erdbeerdatensatz

Anzahl Matches	nicht registriert		zielbasiert		zweistufig zwei Filter		zweistufig ein Filter	
	gef.	Filter	gef.	Filter	gef.	Filter	gef.	Filter
avg.								
UV	1314	250	1130	271	1140	272	1140	272
VIS_NIR	374	173	225	96	220	113	216	113
SWIR	6172	222	573	20	4450	142	4446	142
UV-VIS_NIR	397	1	235	74	229	56	227	75
UV-SWIR	4188	1	586	11	3029	12	3029	12
VIS_NIR-SWIR	4188	0	586	8	3029	6	3029	9
total	2772	108	556	80	2016	100	2014	104

Tabelle 4.3: Durchschnittliche Anzahl an Matches vor und nach Ratio-Test für die Erdbeerdaten

VIS_NIR-Kamera und der SWIR-Kamera alle Matches aussortiert. Die meisten Matches werden auf den nicht registrierten Daten gefunden. Auch die Gesamtanzahl der gefilterten Matches ist hier am höchsten. Jedoch werden insgesamt mehr Matches aussortiert und das Verhältnis zwischen den gefundenen und gefilterten ist mit 3,9% am niedrigsten. Auch bei den zweistufigen Methoden ist das Verhältnis mit 5,0% und 5,2% nicht sehr hoch. Lediglich die zielbasierte Methode kann mehr als 14% erreichen. Hier werden jedoch nur circa ein Viertel der Matches gefunden. Bei der Kamerakombination *UV-VIS_NIR* werden durchschnittlich am wenigsten Matches detektiert. Tabelle 4.4 zeigt die durchschnittlichen RMSE Ergebnisse für den Erdbeerdatensatz. Die Tabelle ist analog zu Tabelle 4.1 aufgebaut. Am niedrigsten ist der bildweise RMSE bei den Methoden *zielbasiert* und *zweistufig - ein Filter*. Mit zwei Filtern in der zweistufigen Methode er-

RMSE	nicht registriert		zielbasiert		zweistufig zwei Filter		zweistufig ein Filter	
	Bild	Matches	Bild	Matches	Bild	Matches	Bild	Matches
avg.								
UV	2.57	3.35	1.98	1.62	2.09	1.57	2.09	1.57
VIS_NIR	5.75	5.54	4.11	4.13	3.45	2.09	3.38	2.13
SWIR	9.33	1.73	8.64	0.68	9.35	2.37	9.35	2.37
UV-VIS_NIR	5.96	14.01	4.06	4.18	4.11	7.38	3.90	3.65
UV-SWIR	7.93	15.93	7.77	2.18	7.75	4.85	7.74	5.12
VIS_NIR-SWIR	8.59	1.25	9.28	3.14	9.33	8.11	9.37	6.27
total	6.69	7.35	5.97	2.65	6.02	4.39	5.97	3.52

Tabelle 4.4: Durchschnittlicher RMSE Bild- und Matchesweise für Erdbeerdatensatz

hört sich der bildweise Fehler leicht. Am größten ist er bei den nicht registrierten Daten. Jedoch liegt die Differenz zwischen dem größten und kleinsten Wert bei 0.72, sodass der Unterschied sehr gering ausfällt. Bei Kombinationen mit Bildern der SWIR-Kamera ist der bildweise RMSE größer, im Vergleich zu Kombinationen mit Bildern der anderen Kameras. Die kleinsten Werte treten bei den Bildpaaren von der UV-Kamera auf.

Der auf den gefilterten Matches berechnete RMSE zeigt einen deutlich größeren Unterschied. Am größten ist er bei den nicht registrierten Bildern. Jedoch beruhen die Berechnungen der angegebenen Durchschnittswerte auf wenigen Daten, wie in Tabelle 4.3 zu sehen ist. Der geringste Fehler kann bei *zielbasiert* festgestellt werden. Auch die zweistufigen Methoden liegen mit ihren Werten unterhalb von dem der nicht registrierten Bilddaten. Auf den Erdbeerdaten ist der Unterschied zwischen den zweistufigen Methoden jedoch größer, als bei den Sperrmüll Daten. Dieser Unterschied resultiert aus den Fehlern zwischen den Bildern unterschiedlicher Kameras, da die Registrierung in der erste Stufe, pro Kamera gleich ist. Die berechneten Fehler zwischen Bildern der einzelnen Kameras sind kleiner, als die der Kamera-Kombinationen.

4.1.3 Zusammenfassung

In diesem Unterkapitel werden vier Metriken eingesetzt, um verschiedene Registrierungsverfahren zu vergleichen. Die zweistufige Methode mit einem Filter erzeugt die meisten nutzbaren Matches und hat auch den niedrigsten RMSE für beide Datensätze. Allgemein lässt sich feststellen, dass die Registrierungen zwischen Bildern einer Kamera bessere Ergebnisse liefern als Bilder verschiedener Kameras. Hier werden zum Teil keine nutzbaren Matches gefunden, was eine Registrierung nicht möglich macht.

4.2 Clustering-Verfahren

Um eine Einschätzung über die in Kapitel 3 vorgestellten Vorverarbeitungsschritte und Clustering-Verfahren zu erhalten, wird für den Vergleich eine Szene pro Datensatz ausgewählt, um Daten mit Objekten aus verschiedenen Sektoren zu untersuchen. Diese Szenen wurden im Vorfeld durch die zweistufige Registrierung mit einem Filter zu einem 15-dimensionalen Array umgewandelt. Anschließend werden die einzelnen Variationen der Daten durch die verschiedenen Vorverarbeitungsmethoden erstellt. Dabei beschreibt *roh* die unveränderten Daten. *skaliert* beschreibt die Daten mit einer Vorverarbeitung durch

eine Standardisierung. *PCA_99* und *PCA_95* bezeichnet die Daten nach einer PCA und mit einer Auswahl an Hauptkomponenten, dass die entsprechende Zahl mindestens dem kumulativen erklärten Varianzverhältnis entspricht. Für die Evaluation werden folgende Metriken verwendet.

1. Calinski-Harabasz-Index
2. Davies-Bouldin-Index

Der in Kapitel 2 beschriebene Silhouetten Index wird nicht benutzt, da hierfür die paarweisen Distanzen für die unskalierten Bilder berechnet werden müssen, was ein zu hoher Rechenaufwand bedeutet. Da für die Daten keine ground truth Label bekannt sind, werden keine anderen, darauf beruhenden Metriken verwendet.

Um einen Vergleich der Skalierungsmethoden zu erhalten, werden beide in Kapitel 3 vorgestellten Varianten evaluiert. *C.-Avg* bezeichnet die Berechnungen der Cluster aus dem Durchschnitt der gefundenen Cluster-Zentren in den Teilbildern, während *sortiert* die Ergebnisse nach dem Label-Mapping beschreibt. Zusätzlich wird ein Teilbild der Zerlegung als Stichprobe verwendet, welches die Daten möglichst gut repräsentiert *Probe*. So kann eine Einschätzung gegeben werden, ob die Cluster-Analysen auf den vollen Daten zu besseren Ergebnissen führt.

4.2.1 Sperrmülldatensatz

Für den Sperrmülldatensatz wurden k (Anzahl der Cluster) für K-Means auf 9 gesetzt, da sich in der Szene acht verschiedene Objekte vorhanden sind. Der Calinski-Harabasz Index für den Sperrmülldatensatz ist in Tabelle 4.5 dargestellt. Die einzelnen Vorverarbeitungsmethoden sind spaltenweise aufgeführt, während zeilenweise die verschiedene Clustering-Verfahren und die verwendete Mapping-Methode für die Skalierung aufgetragen sind. Für den Calinski-Harabasz Index bedeuten große Werte eine höhere Ähnlichkeit innerhalb der Cluster und einen größeren Unterschied zwischen den Clustern. Die besten Ergebnisse erzielt der K-Means mit der Mapping-Methode *C.-Avg* auf den *PCA_95*. Insgesamt bringt diese Vorverarbeitung die höchsten Werte über alle Clustering-Verfahren. *PCA_99* erzielt durchschnittlich den niedrigsten Index. Etwas höher sind die Werte bei *roh. skaliert* liegt durchschnittlich im Mittelfeld. Die Ergebnisse über alle Mapping-Methoden mit K-Means sind durchschnittlich am höchsten, während von Mean-Shift einen fast doppelt so hohen durchschnittlichen Wert im Vergleich zu HDBSCAN erreicht. Zwischen den

Mapping-Methoden ergeben sich deutliche Unterschiede in den durchschnittlichen Werten. *C.Avg* erzeugt einen fast 80-fach höheren Index als *Probe*. Auch *sortiert* ist um das 1.6-fache niedriger. Dieser große Unterschied in den durchschnittlichen Werten kommt vor allem von den Ergebnissen von HDBSCAN. Hier werden zum Teil nur zweistellige Indizes erreicht.

Für die zweite Metrik, dem Davies-Bouldin Index, werden die Ergebnisse in Tabelle 4.6 gezeigt. Die Tabelle ist wie Tabelle 4.5 aufgebaut. Bei diesem Index sind niedrigere Werte um null ein Zeichen für eine gute Seperierung der Cluster. Der niedrigste Index wird hier für *Mean-Shift - Probe* mit der Vorverarbeitungsmethode *PCA_95* berechnet, während der höchste Wert ebenfalls mit Mean-Shift, allerdings auf den *skalierten* Daten und mit der Mapping-Methode *sortiert* berechnet wird. Diese beiden Vorverarbeitungsmethoden bilden durchschnittlich auch die niedrigsten (*PCA_95*) und höchsten (*skaliert*) Werte. Die weit niedrigsten Werte werden auf den nicht vorverarbeiteten Daten (*roh*) berechnet. Durchschnittlich werden die niedrigsten Ergebnisse mit K-Means erreicht. Keiner der einzelnen berechneten Indizes liegt über 1.16. HDBSCAN liegt bei diesem Vergleich im Mittelfeld wobei Werte über fünf für die Mapping-Methode *sortiert* vorkommen, während Mean-Shift, durch den sehr hohen Wert, durchschnittlich die höchsten Werte liefern. Durch den selben hohen Wert bei *skaliert* und Mean-Shift *sortiert* und durch HDBSCAN, wird der durchschnittliche Index für diese Mapping-Methode stark in die höhe geschoben. Die anderen beiden Mapping-Methoden liegen mit einem durchschnittlichen Unterschied von 0.49 relativ dicht bei einander. *C.-Avg* liegt etwas tiefer.

Nach diesen metrischen Vergleichen folgt nun eine optische Einschätzung. Hierfür werden die Cluster-Ergebnisse als Bild mit der Bildauflösung dargestellt. Hierdurch werden die einzelnen Pixel in verschiedene Farben, entsprechend ihrer zugewiesenen Label

Calinski-Harabasz Index	roh	skaliert	PCA_99	PCA_95
HDBSCAN - C.-Avg	1,150,280.59	1,102,799.26	840,850.00	1,602,341.23
HDBSCAN - sortiert	40,527.61	30,648.56	16,732.31	32,530.12
HDBSCAN - Probe	62.65	45.11	61.49	1269.21
K-Means - C.-Avg	1,767,005.98	1,639,034.00	1,871,708.34	2,389,663.57
K-Means - sortiert	1,613,405.48	1,637,589.08	1,591,256.94	1,545,763.58
K-Means - Probe	26,809.56	25,039.93	28,179.41	36,632.00
Mean-Shift - C.-Avg	923,748.37	1,622,652.29	925,606.74	1,514,973.61
Mean-Shift - sortiert	838,415.94	1,144,571.30	862,022.02	1,321,299.33
Mean-Shift - Probe	19,757.05	18,290.31	19,995.24	40,862.18

Tabelle 4.5: Calinski-Harabasz Index Sperrmüll Datensatz

gezeigt. Zusätzlich wird das RGB-Bild als Vergleich mit abgebildet. Angezeigt werden die Ergebnisse für HDBSCAN in Abbildung 4.3. Bei den Labels der Mapping-Methode *Center-Avg* können nach allen Vorverarbeitungsmethoden grobe Strukturen erkannt werden. Die Objekte sind zum Teil homogen geclustert, eine eindeutige Unterscheidung der einzelnen Objekte kann jedoch nicht beobachtet werden. Für *sortiert* können auf den rohen und standardisierten Daten zum Teil noch die Strukturen der Zerlegungsmethode gesehen werden. Die Bilder *PCA_99* und *PCA_95* zeigen nur zwei Farben, was dafür spricht, dass ein Cluster und Rauschen detektiert wurde. Jedoch sind die einzelnen Umrisse der Objekte gut zu erkennen. Auf den Stichprobendaten scheint ein Großteil der Bilder einem Cluster zugeordnet zu werden. Lediglich bei *PCA_95* können die Umrisse von den Objekten erkannt werden. In Abbildung 4.4 werden die Ergebnisse von K-Means dargestellt. Hier sind ähnliche Ergebnisse für alle Varianten zu beobachten. Umrisse einzelner Objekte können erkannt werden. Objekte werden zum Teil homogen geclustert. Die Cluster-Ergebnisse von Meanshift sind in Abbildung 4.5 zu sehen. Auch hier sind keine großen Unterschiede zwischen den einzelnen Varianten der Vorverarbeitungsmethoden und Mapping-Methoden zu erkennen. Einige Objekte werden einheitlich einem Cluster zugewiesen. Allgemein werden wenige Cluster erkannt.

4.2.2 Erdbeerdatensatz

Für den Erdbeerdatensatz wurde k auf fünf gesetzt. Die Tabellen für dieses Unterkapitel sind analog zu dem vorherigen Unterkapitel aufgebaut. Tabelle 4.7 zeigt den Calinski-Harabasz Index für eine Szene aus dem Erdbeerdatensatz. Die höchsten Werte können bei K-Means und *PCA_95* erkannt werden. Im Allgemeinen sind die Werte für K-Means

Davies-Bouldin Index	roh	skaliert	PCA_99	PCA_95
HDBSCAN - C.-Avg	1.34	1.50	1.45	1.03
HDBSCAN - sortiert	3.26	4.49	7.88	6.95
HDBSCAN - Probe	3.41	4.68	3.47	1.88
K-Means - C.-Avg	1.06	1.12	1.03	0.97
K-Means - sortiert	1.12	1.12	1.12	1.16
K-Means - Probe	1.09	1.12	1.06	0.96
Mean-Shift - C.-Avg	1.62	0.89	1.21	1.67
Mean-Shift - sortiert	3.10	90.62	0.84	0.99
Mean-Shift - Probe	0.82	0.92	0.83	0.58

Tabelle 4.6: Davies-Bouldin Index, Sperrmülldatensatz

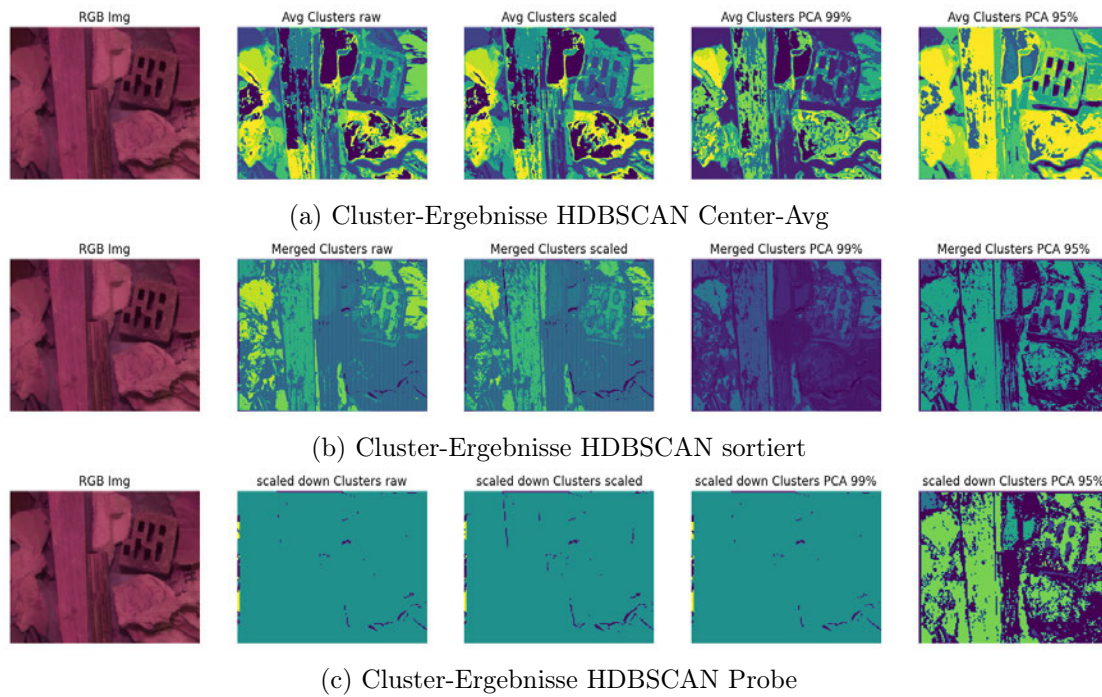


Abbildung 4.3: HDBSCAN Cluster-Ergebnisse für eine Szene aus dem Sperrmülldatensatz

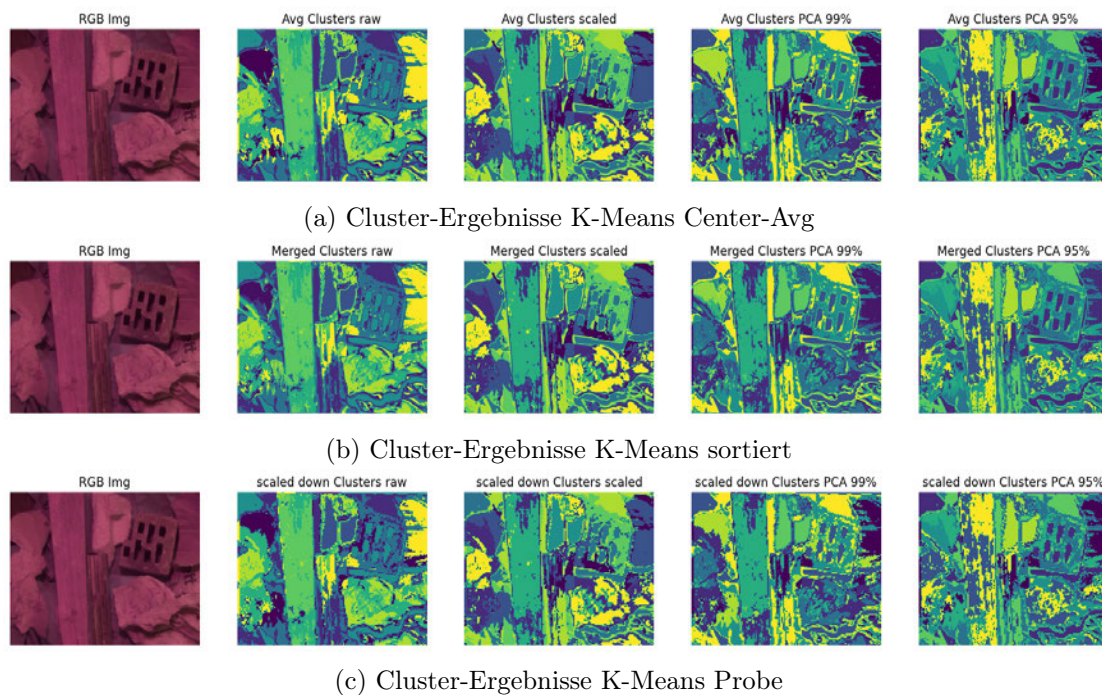


Abbildung 4.4: K-Means Cluster-Ergebnisse für eine Szene aus dem Sperrmülldatensatz

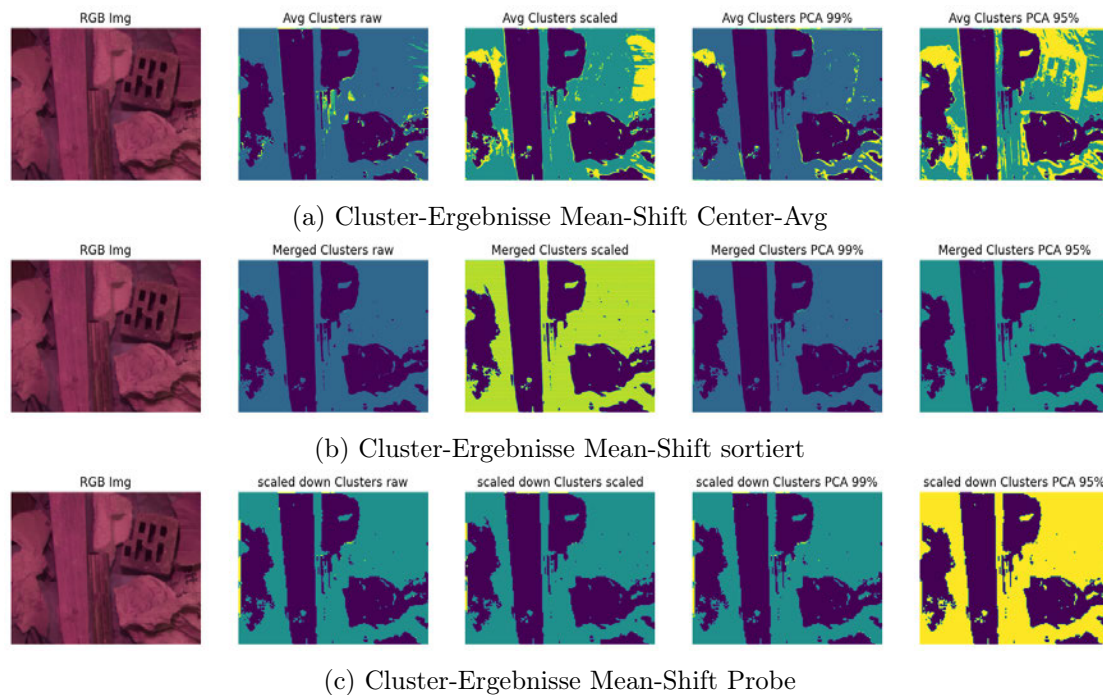


Abbildung 4.5: Mean-Shift Cluster-Ergebnisse für eine Szene aus dem Sperrmülldatensatz

durchschnittlich höher, im Vergleich zu den anderen Clustering-Verfahren. Diese liegen um mehr als den Faktor fünf niedriger und unterscheiden sich durchschnittlich um circa 4000 voneinander. Die einzelnen Vorverarbeitungsmethoden liegen durchschnittlich relativ eng bei einander. *PCA_95* und *PCA_99* liegen dabei etwas höher. Der durchschnitts Wert für *skaliert* liegt unter dem von *roh*. Zwischen den Mapping-Methoden liegen größere Unterschiede. *C.-Avg* liegt um das 1.4-fache höher als *sortiert* und sogar um das 75-fache höher als *Probe*.

Davies-Bouldin Index für die Szene des Erdbeerdatensatzes ist in Tabelle 4.1 dargestellt. Der kleinste Wert findet sich bei *HDBSCAN - C.-Avg* mit der Vorverarbeitungsmethode *PCA_95*. Auch die höchsten Einzelwerte können bei *HDBSCAN* festgestellt werden. Diese treten bei der Mapping-Methode *sortiert* bei *roh* und *PCA_95* auf. Diese Werte bilden mit 61.47 und 48.47 große Ausreißer, da alle anderen Werte kleiner als 3.8 liegen. Hierdurch ist der Durchschnittswert für *HDBSCAN* am höchsten. Durchschnittlich erzeugt *K-Means* die kleinsten Werte, während *Mean-Shift* im Mittel um 0.44 höher liegt. Im Vergleich zwischen den Vorverarbeitungsmethoden liegen *skaliert* und *PCA_99* am niedrigsten, während *roh* und *PCA_95* Werte über sechs haben. Die Mapping-Methoden

Calinski-Harabasz Index	roh	skaliert	PCA_99	PCA_95
HDBSCAN - C.-Avg	428,788.22	359,534.96	274,789.74	259,471.24
HDBSCAN - sortiert	31,708.68	32,910.12	68,574.50	29,153.58
HDBSCAN - Probe	1,642.83	708.54	1,976.28	2,073.75
K-Means - C.-Avg	976,793.54	944,507.70	1,010,273.36	1,145,418.56
K-Means - sortiert	976,091.49	943,853.82	1,009,488.03	1,144,793.69
K-Means - Probe	15,444.69	14,509.38	15,937.47	18,038.98
Mean-Shift - C.-Avg	122,821.21	177,306.63	287,617.45	504,514.33
Mean-Shift - sortiert	70,177.87	24,853.37	172,801.75	65,045.25
Mean-Shift - Probe	5,437.65	364.23	4,897.54	5,401.79

Tabelle 4.7: Calinski-Harabasz Index, Erdbeerdatensatz

Davies-Bouldin Index	roh	skaliert	PCA_99	PCA_95
HDBSCAN - C.-Avg	1.14	1.26	1.15	0.85
HDBSCAN - sortiert	61.47	1.33	2.40	48.47
HDBSCAN - Probe	1.00	1.33	1.01	1.01
K-Means - C.-Avg	0.96	1.00	0.95	0.88
K-Means - sortiert	0.96	1.00	0.95	0.88
K-Means - Probe	0.96	1.01	0.94	0.87
Mean-Shift - C.-Avg	0.99	2.42	0.98	0.91
Mean-Shift - sortiert	1.35	0.87	1.55	3.73
Mean-Shift - Probe	0.94	1.01	0.95	0.92

Tabelle 4.8: Davies-Bouldin Index, Erdbeerdatensatz

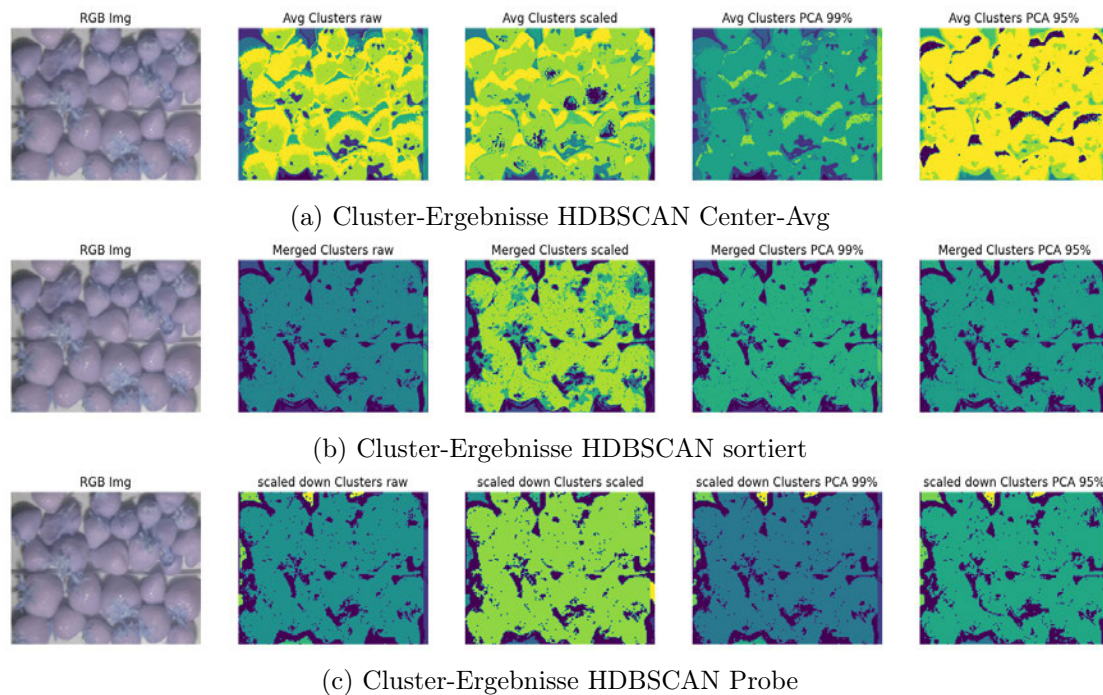


Abbildung 4.6: HDBSCAN Cluster-Ergebnisse für eine Szene aus dem Erdbeerdatensatz

C.-Avg und *Probe* liegen mit durchschnittlichen Werten von 1.12 und 1.00 eng bei einander, während *sortiert* mit einem durchschnittlichen Wert von 10.41 deutlich höher liegt. Die Abbildungen 4.6, 4.7, 4.8 zeigen die Ergebnisse der Clustering-Verfahren optisch. Bei HDBSCAN können nur leichte Unterschiede beobachtet werden. Einzelne Früchte sind nur bei *Center-Avg* zu vermuten, eine deutliche Erkennung ist nicht möglich. Noch weniger ist bei den Ergebnissen von Mean-shift zu erkennen. Auf den *skalierten* Daten wird ein großes Cluster für die Mapping-Methoden *sortiert* und *Probe* gefunden. Aber auch in den anderen Ergebnissen sind keine Früchte zu erkennen. Bei K-Means kann deutlich mehr erkannt werden. Auf *roh* und *skaliert* ist die Erkennung von einzelnen Früchten möglich, obwohl einzelnen Datenpunkte innerhalb einer Frucht zum Teil zwei verschiedenen Clustern zugeordnet werden. Die Kelche und Blätter der einzelnen Erdbeeren werden erkannt und bilden ein eigenes Cluster. Auf den PCA-Daten können diese nicht erkannt werden. Große Unterschiede zwischen den Mapping-Methoden sind nicht zu sehen.

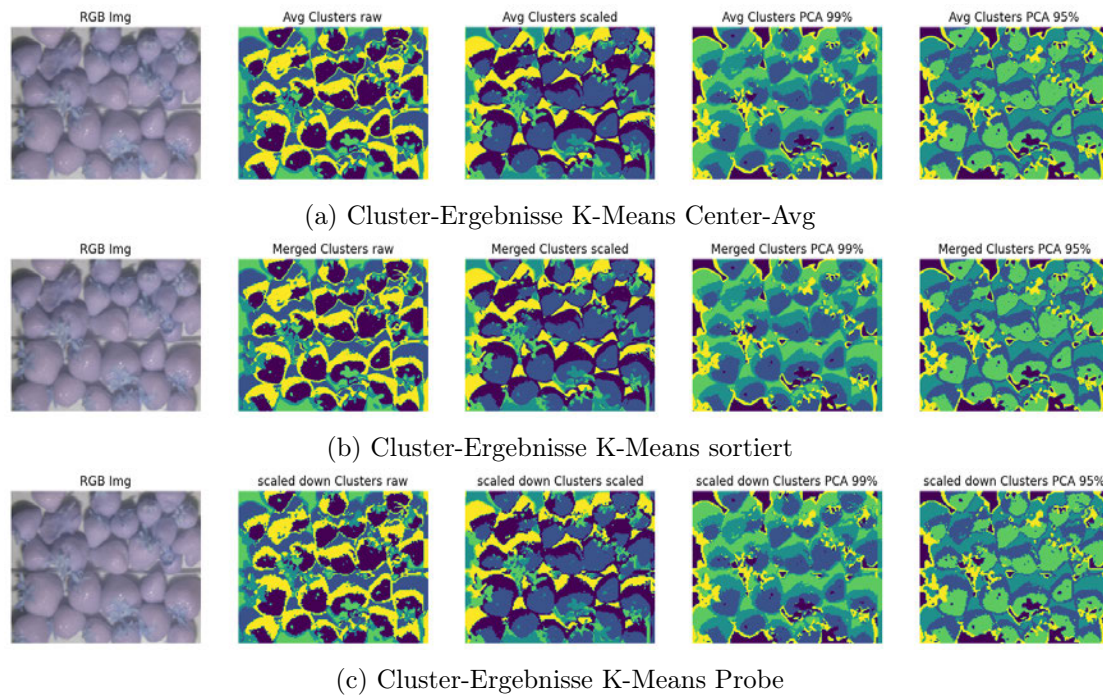


Abbildung 4.7: K-Means Cluster-Ergebnisse für eine Szene aus dem Erdbeerdatensatz

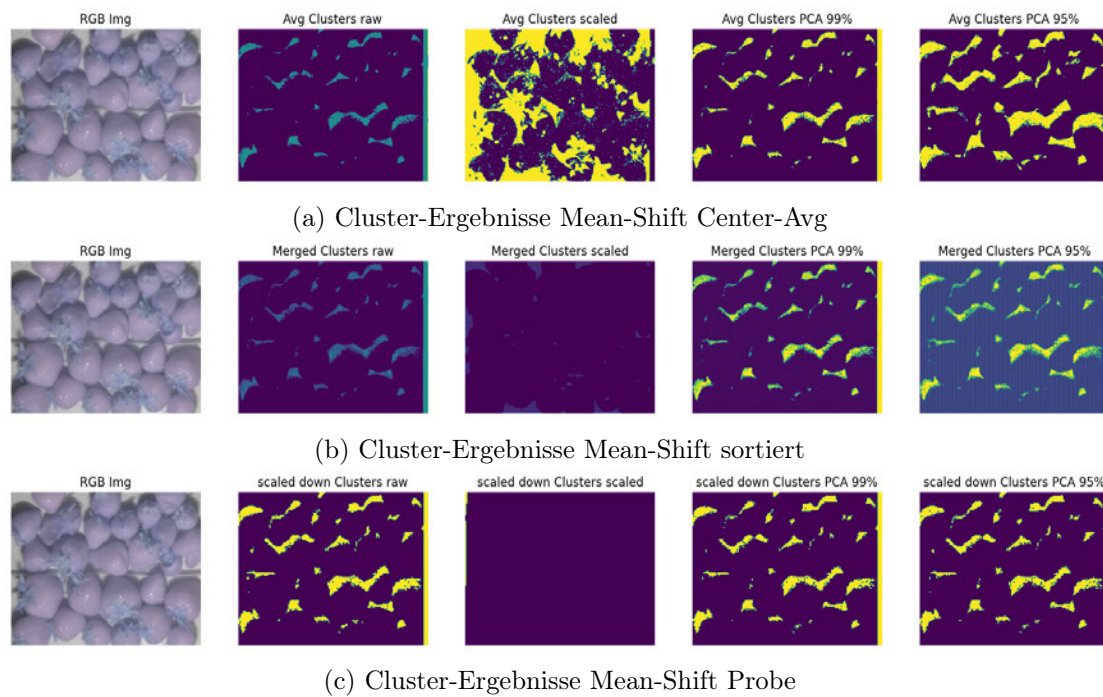


Abbildung 4.8: Mean-Shift Cluster-Ergebnisse für eine Szene aus dem Erdbeerdatensatz

5 Diskussion

5.1 Multispektrale Registrierung

Die Untersuchungen in Kapitel 4 zeigen, dass auf beiden Datensätzen eine große Anzahl an Matches gefunden werden können, viele jedoch unbrauchbar sind, da sie durch den Ratio-Test aussortiert werden. Vorallem zwischen den verschiedenen Kameras werden viele Matches aussortiert. Dies kann ein Grund dafür sein, dass die Registrierungen innerhalb einer Kamera besser funktionieren, als zwischen den Kameras. Im Allgemeinen zeigen die Registrierungsmethoden bessere Ergebnisse als, die Rohdaten. Eine Verbesserung ist somit durch alle Methoden möglich. Ein Ausreißer bildet hierbei der durchschnittliche RMSE auf den Matches des Erdbeerdatensatzes. Jedoch wurden für die Berechnung nur sehr wenig Matches benutzt, sodass eine Interpretation schwierig ist. Um solche Fälle besser zu beurteilen, wäre eine Anpassung der Metrik nötig gewesen. Diese schneiden im Durchschnitt in den Metriken besser ab, was auf ein besseres Registrierungsergebnis hinweist. Durch die Einschränkungen der Transformation auf Rotation, Skalierung und Translation können bessere Registrierung in der ersten Stufe, also auf den Bildern innerhalb einer Kamera, gesehen werden. Diese Ergebnisse decken sich mit der Vermutung, dass sich die Verzerrung zwischen diesen Bildern nur aus den genannten Transformation revidieren lassen. Auch die Einschränkung durch den Distanzfilter scheint in dieser Stufe sinnvoll, sodass die Rotation wenige Grad beschränkt werden konnte. Die Ergebnisse zeigen, dass dieser Distanzfilter in der zweiten Stufe zu schlechteren Ergebnissen führt. Eine mögliche Erklärung dafür ist, dass damit mehr Matches aussortiert werden und weniger Matches zur Berechnung der Transformation zur Verfügung stehen. Aus der Literatur ist bekannt, dass mehr Matches bessere Registrierungen bilden können.

Die erste Metrik, also der bildweise RMSE, muss vorsichtig interpretiert werden, da es sich bei den Daten um multispektrale Bilder handelt. Hier unterscheiden sich die Intensitätswerte zwischen den einzelnen spektralen Kanälen, sodass ein kleiner Wert gegen null nur schwer zu erreichen ist. Trotzdem kann dieser Wert als Vergleichsmetrik genutzt

werden, da Verschiebungen der Bilder im Differenzbild zu größeren Werten führen. Die Nutzung einer optischen Einschätzung der Registrierungsergebnisse kann sinnvoll sein. Die RMSE-Metriken können als pixelweiser Fehler interpretiert werden. Er gibt somit an, um wie viele Pixel sich die einzelnen Bilder oder Positionen der Matches unterscheiden. Da eine möglichst pixelgenaue Registrierung für eine Weiterverarbeitung notwendig ist, sollte dieser möglichst klein gehalten werden. Wenn allerdings nur die RMSE-Metriken genutzt werden würden, wäre die zielbasierte Methode für einige Szenen des Erdbeerdatensatzes der zweistufigen Methoden vorzuziehen. Die optische Einschätzung zeigt jedoch, zum Teil stark verdrehte Einzelbilder, für die zielbasierte Methode, sodass das Ergebnisse ohne weitere Verarbeitung nicht zu gebrauchen sind. Ein Vorteil der zweistufigen Registrierung ist, das in der zweiten Stufe, also der Registrierung zwischen Bildern der Kameras durch die Auswahl, mehr Möglichkeiten zu einer geeignete Transformation zu finden sind. Da die Bilder innerhalb der Kameras bereits registriert sind, können verschiedene Paare genutzt werden. Dies ist gegenüber der zielbasierten Methoden ein Vorteil, da hier nur ein Bild als Zielbild genutzt wird. Wenn für eine Kombination mit diesem nicht genug Matches erkannt werden können, kann keine Transformation berechnet und somit keine Registrierung durchgeführt werden. Gleichzeitig erfordert dieser Schritt eine manuelle Bildauswahl, es sei denn, man nimmt beispielsweise die Anzahl an nutzbaren Matches als Auswahlkriterium. Bei der optischen Auswärtung fallen zum Teil starke Unterschiede am Rand der Bilder auf. Diese sind auf Transitionen zurückzuführen, die das Bild weiter verschieben als Pixel an der Seite vorhanden sind. Das kann durch eine schlechte Auswahl der Bildausschnitte im Vorfeld der Registrierung entstehen. Es lassen sich keine Signifikanten Unterschiede in den Registrierungsmethoden zwischen den beiden Datensätzen feststellen.

5.2 Clustering-Verfahren

Für den Vergleich der Clustering-Verfahren K-Means, HDBSCAN und Mean-Shift wurden in Kapitel 4 die Metriken Calinski-Harabasz Index und Davies-Bouldin Index verwendet. Verglichen wurden zusätzlich verschiedene Vorverarbeitungsmethoden, unter anderem auch eine Skalierungsmethode mit unterschiedlichen Mapping-Methoden für die Übertragung der Ergebnisse von den Teilbildern auf das Gesamtbild. Die Skalierungsmethode wurde eingesetzt, um die multispektralen Daten mit weniger Speicher und Rechenaufwand zu Clustern. Verglichen wird diese mit einem Stichprobendatensatz, der aus einem Teilbild des ursprünglichen multispektralen Bildes besteht. Die Ergebnisse

zeigen, dass die vorgestellte Methode der Zerlegung des Gesamtbildes einer Stichprobe vorzuziehen ist. Die hier gefundenen Cluster unterscheiden sich zwar stark, was ein niedriger Davies-Bouldin Index aussagt. Jedoch ist auch der Calinski-Harabasz Index im Verhältnis zu den anderen Mapping-Methoden sehr klein. Je größer dieser ist, desto ähnlicher sind sich die Daten innerhalb eines Clusters und desto verschiedener zwischen den Clustern. Es zeigt sich somit, dass entweder die Probe größer gefasst werden muss oder eine Nutzung von allen Datenpunkten zu besseren Ergebnissen führt. Auch zwischen den Mapping-Methoden gibt es große Unterschiede in den Ergebnissen. Die zusätzliche Berechnung der Durchschnitts-Cluster-Zentren und die anschließende Zuweisung der Daten zu diesen, verbessert beide Metriken, vor allem für HDBSCAN und Mean-Shift deutlich. Gerade bei dem Calinski-Harabasz Index ist der Unterschied stärker. Hieraus kann geschlossen werden, dass sich vor allem die Ähnlichkeit der Daten innerhalb der Cluster vergrößert, da sich der Davies-Bouldin Index nur in wenigen Fällen stark verändert. Für K-Means sind die Unterschiede in beiden Indizes und für beide Datensätze minimal, sodass für dieses Clustering-Verfahren kein Favorit bestimmt werden kann.

Auch aus dem Vergleich der Vorverarbeitungsmethoden lassen sich Erkenntnisse gewinnen. Hierbei wurden die Daten ohne Vorverarbeitung mit den Daten nach einer Standardisierung und den Daten nach einer Dimensionsreduzierung durch eine PCA mit einer erklärten Varianz von 99% und 95% verglichen. Hierbei wurden die Daten auf sechs bzw. drei Dimensionen reduziert. Die Ergebnisse von dem Calinski-Harabasz Index sind für beide Datensätze *PCA_95* am besten. Jedoch liefert die Vorverarbeitungsmethode für den Davies-Bouldin Index nur auf den Sperrmülldaten das beste Ergebnis. Dieser Unterschied lässt sich auf einen Ausreißer bei HDBSCAN *sortiert* zurückführen. Denn vor allem für K-Means liefert diese Vorverarbeitung, die beste Ergebnisse. Auffällig in diesem Vergleich ist, dass der Calinski-Harabasz Index für *sortiert* kleiner ist als für die rohen Daten. Hieraus lässt sich schließen, dass eine Standardisierung der Werte bei beiden Datensätzen zu einem schlechteren Clustering-Ergebnis führt. Es lässt sich sagen, dass eine Dimensionsreduzierung bessere Ergebnisse auf den Daten liefert, da die durchschnittliche Indizes auch bei *PCA_99* im Vergleich bessere Ergebnisse liefern. Dieses Ergebnis wird jedoch durch die visuelle Untersuchung von den K-Means Ergebnissen auf dem Erdbeerdatsatz skeptisch hinterfragt. Hier können deutlich mehr Detail bei *roh* und *skaliert* erkannt werden. Allerdings kann diese Einschätzung durch die verwendete Farben trüben.

Als bestes Clustering-Verfahren schneidet K-Means ab. In allen Variationen liegt der Davies-Bouldin Index unter 1.2. Daraus ergibt sich, dass sich die gefundenen Cluster stark unterscheiden. Auch der Calinski-Harabasz Index ist durchschnittlich am höch-

ten, sodass auch von einer hohen Ähnlichkeit der Daten innerhalb der Cluster ausgegangen werden kann. Diese Ergebnisse gelten für beide Datensätze. Bei den anderen Clustering-Verfahren unterscheiden sich die Ergebnisse zwischen den Datensätzen. Auf den Sperrmülldaten hat HDBSCAN einen kleineren Calinski-Harabasz Index, aber auch der Davies-Bouldin Index ist kleiner, als auf den Erdbeerdaten. Für Mean-Shift ist dies genau umgekehrt. Hier ist der Davies-Bouldin Index auf den Erdbeerdaten besser, als der von HDBSCAN, jedoch ist der Calinski-Harabasz Index auf diesem Datensatz für Mean-Shift besser. Für den Erdbeerdatensatz sind die Indizes für Mean-Shift vertauscht. Hieraus lässt sich schließen, dass HDBSCAN im Vergleich mit Mean-Shift auf den Sperrmülldaten weniger Ähnlichkeit innerhalb der Cluster, dafür jedoch größere Unterschiede zwischen den Clustern hat. Für den Erdbeerdatensatz ist dieser Sachverhalt genau umgekehrt.

6 Fazit und Ausblick

Diese Arbeit verglich verschiedene Vorverarbeitungsmethoden, die für eine Clustering-Analyse von multimodalen multispektralen Bilddaten nötig sein können. Ein besonderes Augenmerk wurde dabei auf die Bildregistrierung gelegt, bei der die einzelnen spektralen Bildkanäle oder Einzelbilder möglichst Pixelgenau übereinander gelegt werden müssen. Dieser Schritt ist wichtig, da so für jedes Pixel eine Art Spektrum gebildet wird. Das in dieser Arbeit vorgestellte Verfahren ist zweistufig aufgebaut und nutzt Vorwissen der Datenaufnahme für die Auswahl der Transformationen zwischen den einzelnen Bildern. So wird eine affine Transformation für Bilder einer Kamera gewählt, während für die Transformation zwischen Bildern verschiedener Kameras eine projektive Transformation als Model genutzt wird. Die Ergebnisse zeigen, dass dieses Verfahren auf unterschiedlichen Datensätzen besser funktioniert, als ein zielbasiertes Verfahren. Dabei wird für die erste Stufe zusätzlich ein Differenzfilter eingesetzt, welcher hier starke Rotationen verhindert, welche im Vorfeld bei der zielbasierten Methoden aufgetreten sind. Der Einsatz des Differentfilters in der zweiten Stufe verschlechtert die Registrierung, da so mehr Matches aussortiert werden, die für eine gute Berechnungen erforderlich sind. Vorallem da in der zweiten Stufe, also der Registrierung zwischen den Kameras, bereits weniger Matches vorhanden sind.

Weiterhin wurden verschiedene Vorverarbeitungsmethoden für eine Cluster-Analyse verglichen. Auch wurde ein Zerlegungsmethode genutzt, welche die Auflösung des ursprünglichen Bildes reduziert und das Clustering auf den skalierten Bildern durchführt. Hierdurch kann Speicher- und Rechenaufwand gespart werden. Um die Ergebnisse der gesamten Daten zu nutzen wurden verschiedene Mapping-Methoden vorgestellt, die die Ergebnisse der skalierten Teilbilder auf das ursprüngliche Bild abbildet. Besonders gute Ergebnisse wurden mit der Berechnung von Durchschnitts-Cluster-Zentren und der anschließenden Zuteilung der Daten gemacht. Diese Ergebnisse waren deutlich besser, als die von einer kleiner Stichprobe. Weiterhin ist eine Dimmensionsreduktion durch eine PCA sinnvoll. Hierdurch müssen verschiedene Funktionen, wie die Distanz mit weniger Werten berechnet werden, sodass die Komplexität sinkt. Zum anderen werden hiermit bessere

Clustering-Ergebnisse erzielt. Eine Standardisierung auf den Daten ist nicht sinnvoll und führt zu schlechten oder ähnlichen Ergebnissen wie die unverarbeiteten Daten, sodass sich diese Berechnung gespart werden kann.

Als bestes Clustering-Verfahren schneidet K-Means ab. Es erzeugt im Vergleich zu HDBSCAN und Mean-Shift Cluster, die sich mehr unterscheiden, während sich die Daten innerhalb der Cluster ähnlicher sind. Als Gesamtergebnis lässt sich festhalten, dass Clustering-Verfahren eine gute erste Einschätzung von multispektralen Bilddaten geben können. Vorallem auf den Erdbeerdaten können bereits Frucht und Kelch unterschieden werden. Auch einzelne Objekte im Sperrmüll können segmentiert werden, wobei verschiedene Materialklassen nicht gefunden wurden.

Für die Registrierung wurde nur ein Matcher und ein Merkmalsdetektor untersucht. Für anschließende Arbeiten wäre ein Vergleich mit weiteren Algorithmen in diesen Feldern sinnvoll, um den Fehler noch kleiner werden zu lassen. Weiterhin sollten weitere Daten und Metriken für die Analyse genutzt werden.

Auch der Vergleich der Vorverarbeitungsmethoden und vorallem der Zerlegungsmethode ist nicht abgeschlossen und eine genaue Analyse über Speicherbedarf und Rechenaufwand sollte durchgeführt werden. Zudem sollten andere Dimensionsreduktionsmethoden in Betracht gezogen werden, um auch nicht-lineare Zusammenhänge abbilden zu können. Abschließend sollte die Liste der Clustering-Verfahren erweitert werden um einen größeren Einblick in die Möglichkeiten der multispektralen Bildanalyse zu gewinnen.

Literaturverzeichnis

- [1] BAY, Herbert ; ESS, Andreas ; TUYTELAARS, Tinne ; GOOL, Luc V.: Speeded-Up Robust Features (SURF). In: *Computer Vision and Image Understanding* 110 (2008), jun, Nr. 3, S. 346–359
- [2] BRAUERS, J. ; SCHULTE, N. ; AACH, T.: Multispectral Filter-Wheel Cameras: Geometric Distortion Model and Compensation Algorithms. In: *IEEE Transactions on Image Processing* 17 (2008), Nr. 12, S. 2368–2380
- [3] CALINSKI, T. ; HARABASZ, J.: A dendrite method for cluster analysis. In: *Communications in Statistics - Theory and Methods* 3 (1974), Nr. 1, S. 1–27
- [4] CHEN, Runzi ; ZHAO, Shuliang ; LIANG, Meishe: A Fast Multiscale Clustering Approach Based on DBSCAN. In: *Wireless Communications and Mobile Computing* 2021 (2021), jul, S. 1–11
- [5] COMANICIU, D. ; MEER, P.: Mean shift: a robust approach toward feature space analysis. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2002), may, Nr. 5, S. 603–619
- [6] CORCEL, Mathias ; BARRON, Cecile ; GUILLON, Fabienne ; DEVAUX, Marie F.: *Multi-scale k-means clustering of multispectral images*. 2016. – URL <https://hal.science/hal-01602999>
- [7] DAVIES, David L. ; BOULDIN, Donald W.: A Cluster Separation Measure. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1 (1979), apr, Nr. 2, S. 224–227
- [8] ESTER, Martin ; KRIEGEL, Hans-Peter ; SANDER, Jörg ; XU, Xiaowei: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Portland, Oregon : AAAI Press, 1996 (KDD'96), S. 226–231

- [9] HALKIDI, Maria ; BATISTAKIS, Yannis ; VAZIRGIANNIS, Michalis: On Clustering Validation Techniques. In: *Journal of Intelligent Information Systems* 17 (2001), Nr. 2/3, S. 107–145
- [10] JHAN, Jyun-Ping ; RAU, Jiann-Yeou: A Generalized Tool for Accurate and Efficient Image Registration of UAV Multi-lens Multispectral Cameras by N-SURF Matching. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), S. 6353–6362
- [11] JOHANNES JORDAN: *Interactive Analysis of Multispectral and Hyperspectral Image Data: Interaktive Analyse von multispektralen und hyperspektralen Bilddaten*. Erlangen-Nürnberg, Friedrich-Alexander-Universität Erlangen-Nürnberg, Dissertation, 2017. – URL [urn:nbn:de:bvb:29-opus4-96456](https://nbn-resolving.org/urn:nbn:de:bvb:29-opus4-96456). – Zugriffsdatum: 10.04.23
- [12] JORDAN, Johannes ; ANGELOPOULOU, Elli: Mean-shift clustering for interactive multispectral image analysis. In: *2013 IEEE International Conference on Image Processing*, IEEE, sep 2013
- [13] JOURNAUX, Ludovic ; TIZON, Xavier ; FOUCHEROT, Irene ; GOUTON, Pierre: Dimensionality Reduction Techniques: An Operational Comparison On Multispectral Satellite Images Using Unsupervised Clustering. In: *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006*, IEEE, jun 2006
- [14] KEPPNER, Matthis: *Clustering multispektraler Bilddaten*. 2023
- [15] KEPPNER, Matthis: *Multispektrale Bildregistrierung: in einem multimodalen Setup (UV - VIS - IR)*. 2023
- [16] LOWE, D. G.: Object recognition from local scale-invariant features. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*, IEEE, 1999
- [17] LOWE, David G.: Distinctive Image Features from Scale-Invariant Keypoints. In: *International Journal of Computer Vision* 60 (2004), nov, Nr. 2, S. 91–110
- [18] MACQUEEN, J.: Some Methods for Classification and Analysis of Multivariate Observations. In: *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, S. 281–297.
- [19] MCINNES, Leland ; HEALY, John ; ASTELS, Steve: hdbscan: Hierarchical density based clustering. In: *The Journal of Open Source Software* 2 (2017), Nr. 11, S. 205

- [20] ROUSSEEUW, Peter J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. In: *Journal of Computational and Applied Mathematics* 20 (1987), nov, S. 53–65
- [21] SHAHAROM, M. F. M. ; TAHAR, K. N.: Multispectral Image Matching Using SIFT and SURF Algorithm: A Review. In: *International Journal of Geoinformatics* (2023), jan, S. 13–21
- [22] STEINBACH, Michael ; ERTÖZ, Levent ; KUMAR, Vipin: The Challenges of Clustering High Dimensional Data. In: WILLE, Luc T. (Hrsg.): *New Directions in Statistical Physics*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2004, S. 273–309. – ISBN 978-3-642-07739-5
- [23] TAREEN, Shaharyar Ahmed K. ; SALEEM, Zahra: A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In: *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, IEEE, mar 2018
- [24] TREBUNA, Peter ; HALCINOVA, Jana ; FIL'O, Milan ; MARKOVIC, Jaromir: The importance of normalization and standardization in the process of clustering. In: *2014 IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMi)*, IEEE, jan 2014
- [25] WANG, Xuezhi ; YANG, Weiping ; WHEATON, Ashley ; COOLEY, Nicola ; MORAN, Bill: Efficient registration of optical and IR images for automatic plant water stress assessment. In: *Computers and Electronics in Agriculture* 74 (2010), nov, Nr. 2, S. 230–237
- [26] YASIR, Rahat: *Data Driven Multispectral Image Registration Framework*, University of Saskatchewan, Diplomarbeit, 2018
- [27] ZITOVÁ, Barbara ; FLUSSER, Jan: Image registration methods: a survey. In: *Image and Vision Computing* 21 (2003), oct, Nr. 11, S. 977–1000

Erklärung zur selbstständigen Bearbeitung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

Ort

Datum

Unterschrift im Original