

Multivariate Kerndichteschätzung zur Filterung automatischer Punktzuordnungen

ALESSANDRO CEFALU¹, DIETER FRITSCH² & NORBERT HAALA³

Die automatische Orientierung von Bildverbänden hat in den vergangenen Jahren viele Fortschritte gemacht. Die verschiedenen Lösungsstrategien stützen sich in hohem Maße auf die korrekte Aufdeckung von Merkmalskorrespondenzen zwischen Bildpaaren. Hierbei entstehen jedoch häufig große Mengen an Fehlzuordnungen. Aus diesem Grund werden zur Orientierung meist robuste, iterative RANSAC-basierte Verfahren verwendet, welche für eine spätere klassische Ausgleichung sowohl Ausreißer eliminieren als auch Näherungslösungen liefern. Trotz ihrer Robustheit ist die Effektivität dieser Verfahren weiterhin in sehr vom Anteil der Fehlzuordnungen abhängig. Es bleibt also erstrebenswert diese bereits im Vorfeld zu entfernen. Wir stellen eine auf Kerndichteschätzung basierende Methode zur Filterung der Korrespondenzen vor, welche auf Analyse der Parallaxen und der relativen Deskriptororientierung beruht.

1 Einleitung

Die Forschung im Bereich der automatischen Orientierung von Bildverbänden (Structure-from-Motion) hat in den letzten Jahren viele Erfolge verzeichnet. Dies zeigt sich unter anderem in der steigenden Zahl verfügbarer, sowohl freier als auch kommerzieller, Softwarelösungen. Diese werden immer robuster, schneller und anwenderfreundlicher. Auch wenn die verschiedenen Lösungsstrategien sich in den Details unterscheiden, so lässt sich die übliche Vorgehensweise in wenige Einzelschritte untergliedern. Zunächst werden homologe Punkte in den Bildern identifiziert. Beginnend mit einem Bildpaar werden sukzessiv Einzelbilder dem Verband hinzugefügt, indem bereits triangulierte Punkte zur Orientierung der neuen Bilder verwendet werden. Eine gemeinsame Ausgleichung kann in definierten Intervallen erfolgen, wird aber in jedem Fall als Abschluss des Prozesses durchgeführt. Im ersten Schritt werden zunächst markante Punkte detektiert und mittels eines Merkmalsvektors (Deskriptor) beschrieben. Die Zuordnung findet rein merkmalsbasiert statt, also ohne Überprüfung der geometrischen Korrektheit der Zuordnung. Auf Basis dieser vorläufigen Zuordnungen wird anschließend versucht die geometrische Kamerakonfiguration abzuleiten und dabei Fehlzuordnungen zu eliminieren. Aufgrund des, gerade bei größeren Perspektivunterschieden, häufig hohen Anteils an Fehlzuordnungen, wird hierbei meist auf robuste, iterative Monte-Carlo-Methoden, i.d.R. Varianten des RANSAC (FISCHLER & BOLLES, 1981), zurückgegriffen. Es ergeben sich diverse Fragestellungen welche die Effektivität des Prozesses entscheidend beeinflussen.

1) Alessandro Cefalu, Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24D, 70174 Stuttgart, E-Mail: alessandro.cefalu@ifp.uni-stuttgart.de

2) Dieter Fritsch, Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24D, 70174 Stuttgart, E-Mail: dieter.fritsch@ifp.uni-stuttgart.de

3) Norbert Haala, Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24D, 70174 Stuttgart, E-Mail: norbert.haala@ifp.uni-stuttgart.de

Bei einer Anzahl von m Bildern ergibt sich die Anzahl der möglichen Bildpaare zu $m(m-1)/2$. Sie nimmt also nahezu quadratisch zu. Hiervon ist jedoch meist nur ein Bruchteil wirklich verknüpft, sodass eine frühzeitige Entscheidung, ob ein Bildpaar weiterverarbeitet werden soll, in hohem Maße sinnvoll ist. Zu diesem Zweck werden meist Ähnlichkeitsmaße verwendet. Im Wesentlichen kann hier zwischen Verfahren unterschieden werden, welche Deskriptoren, wie z.B. *GIST* (OLIVA & TORRALBA, 2001), für ein gesamtes Bild erzeugen und Verfahren, welche auf dem Bag-of-Words Modell aufbauen (CSURKA et al., 2004).

Des Weiteren stellt der Aufwand des *RANSAC* eine Hürde dar. Die nötige Anzahl der Iterationen¹ ergibt sich zu $\log(1-p)/\log(1-I^s)$. Hierbei entspricht p der benutzerdefinierten Wahrscheinlichkeit, mit der eine Lösung gefunden werden soll, s der Größe der minimalen Stichprobe² (minimales Sample) und I (für Inlier-Rate) dem Anteil an korrekten Zuordnungen. Die Stichprobengröße s hat sehr großen Einfluss, sodass die Verwendung eines Algorithmus zur Lösung der relativen Orientierung mit möglichst kleiner Samplegröße eine naheliegende Maßnahme darstellt. Ideal ist der 5-Punkte-Algorithmus (NISTER, 2004). Da dieser nur bei Kenntnis der inneren Orientierung verwendbar ist, kommt häufig auch der 7-Punkte-Algorithmus zur Anwendung (HARTLEY, 1997).

Es existieren verschiedenste Ansätze, welche sich als Varianten des *RANSAC* auf einen möglichst effizienten Umgang mit den Zuordnungen konzentrieren. So wird bei Verfahren wie *Guided MLESAC* (TORDOFF & MURRAY, 2002) oder *PROSAC* (CHUM & MATAS, 2005) auf ein Maß zurückgegriffen, welches die Wahrscheinlichkeit einer korrekten Zuordnung repräsentiert, um die Auswahl der Stichproben zu steuern. Man versucht die Stichproben aus möglichst vielversprechenden Zuordnungen zu bilden. Oft wird hier auf die Distanz im Deskriptorraum zurückgegriffen, welche jedoch insbesondere bei großer Perspektive ein ungenügendes Maß darstellt. Andere Ansätze versuchen im Laufe der Iterationen gefundene Lösungen weiter zu nutzen. *GASAC* (RODEHORST & HELLWICH, 2006) z.B. verwendet einen genetischen Algorithmus, also neben Selektion auch Mechanismen wie Rekombination und Mutation, zur Erzeugung neuer minimaler Samples. Wieder andere Verfahren versuchen den Aufwand der Einzeliterationen zu minimieren. Beispielsweise verwendet *WALDSAC* (CHUM & MATAS, 2008) Walds sequentiellen Hypothesentest zum vorzeitigen Abbruch der Einzeliterationen.

In allen Fällen bleibt es jedoch erstrebenswert den Anteil der Fehlzuordnungen bereits im Vorfeld zu reduzieren. Dies senkt die nötige Iterationsanzahl und reduziert den Aufwand der Einzeliterationen, da die Anzahl an durchzuführenden Tests pro Iteration gesenkt wird. Häufig wird das Verhältnis der Deskriptordistanzen zwischen bester und zweitbesten Zuordnung gebildet. Man geht von einer sicheren Zuordnung aus, wenn der zweitnächste Nachbar deutlich

¹ Die Abschätzung der Iterationszahl ist i.d.R. überoptimistisch. Sie geht davon aus, dass ein Sample welches aus korrekten Zuordnungen besteht immer zu einer gültigen Lösung führt, was nicht der Fall ist.

² minimale Stichprobe: Aus der Gesamtmenge der Zuordnungen gewählte Untermenge der kleinstmöglichen Größe, d.h. sodass eine Lösung für die relative Orientierung gefunden werden kann.

weiter entfernt ist als der nächste Nachbar. Eine weitere Möglichkeit stellt die Vorwärts-Rückwärtszuordnung dar. Zuordnungen finden nur Verwendung, wenn sie in beiden Suchrichtungen bestätigt werden, d.h. sich in beiden Richtungen das gleiche Paar als nächste Nachbarn ergibt³. In beiden Fällen bleiben unter Umständen jedoch weiterhin größere Mengen an Fehlzuordnungen erhalten. Das gemeinsame Ziel dieser und ähnlicher Verfahren ist eine verbesserte Nachbarschaftsanalyse im Merkmalsraum des Deskriptors.

Andere Ansätze der Vorfilterung stützen sich hingegen auf Charakteristika der Zuordnungen, welche systematische Effekte bzw. Zusammenhänge aufweisen, wenn alle Korrespondenzen eines Bildpaares betrachtet werden. Das *ICF*-Verfahren (LI & HU, 2010) definiert bidirektionale Korrespondenzfunktionen, welche für jedes Bildpaar mittels Support-Vector-Regression robust geschätzt werden. Zuordnungen, welche sich nicht konform mit diesen Funktionen verhalten werden als Fehlzuordnungen klassifiziert. Histogramme der Parallaxen können verwendet werden, um schwach überlappende Luftbildaufnahmen zu registrieren (MIZOTIN et al., 2010). Die Modalwerte dienen zur näherungsweise Bestimmung der Bewegung des Flugzeugs. Diese wird wiederum verwendet, um die überlappenden Bereiche der Bilder zu bestimmen und anschließend eine erneute Punktzuordnung in nur diesen Bereichen durchzuführen.

Der im Folgenden vorgestellte Ansatz ist letzterem recht ähnlich, wobei wir die Systematiken für eine direkte Filterung der Zuordnungen heranziehen und auch kein Bewegungsmodell zugrunde gelegt werden muss. Wir zeigen, dass eine Anwendung des Verfahrens auch bei ungünstigen Kamerakonfigurationen oder bei sich wiederholenden Strukturen (z. B. bei Fassaden) zu einer Steigerung des Anteils korrekter Zuordnungen führt. Das zur Klassifizierung verwendete Maß kann des Weiteren weiterverwendet werden, um zu einem Ähnlichkeitsmaß für das gesamte Bildpaar zu gelangen. Außerdem eignet es sich auch zur weiteren Verwendung in *RANSAC*-Verfahren mit gesteuerter Sample-Auswahl.

2 Ansatz zur Korrespondenzfilterung

2.1 Systematische Effekte

Stellt man Punktkorrespondenzen eines Bildpaares dar, fällt dem Betrachter sofort auf, dass korrekte Zuordnungen ein gemeinsames Verhalten aufweisen, die Richtung und Länge der Parallaxen ist ähnlich. Das Gegenteil ist bei Fehlzuordnungen der Fall. Dieser Umstand lässt sich leicht über Histogramme dieser Charakteristika darstellen (Abb. 1), bei denen sich klare Häufungen in gewissen Wertebereichen erkennen lassen. Es lässt sich auch feststellen, dass diese selbst bei geringem Überlapp oder starker Perspektive erkennbar bleiben. Auch bleibt der Effekt bei Reduktion der Bildauflösung weitestgehend erhalten. Es liegt nahe zu versuchen, diese Systematiken zu nutzen, um eine Filterung der Fehlzuordnungen vorzunehmen.

³ Dieses Verfahren ist in allen folgenden Beispielen bereits als Teil der Zuordnung enthalten.

Ein Histogramm der Deskriptordistanzen hingegen veranschaulicht, dass korrekte und falsche Zuordnungen jeweils in etwa der χ^2 -Verteilung folgen und sich dabei häufig ungünstig überlagern. Dies ist auch der Grund, weshalb sich die Deskriptordistanz nur bedingt zur Filterung oder für gesteuertes Sampling eignet. Wir verwenden hingegen zusätzlich zur horizontalen Parallaxe p_x und vertikalen Parallaxe p_y die Deskriptororientierung. Diese ist bei rotationsinvarianten Deskriptoren wie z.B. SIFT (LOWE, 2004) oder dem binären Deskriptor ORB (RUBLEE et al., 2011) i.d.R. verfügbar. Die Differenz der Nullrichtungen eines (korrekten) korrespondierenden Punktpaares kann näherungsweise als Maß für die Rotation der Kamera um ihre optische Achse angesehen werden. Wir bezeichnen diese im Folgenden als rotatorische Parallaxe p_r .

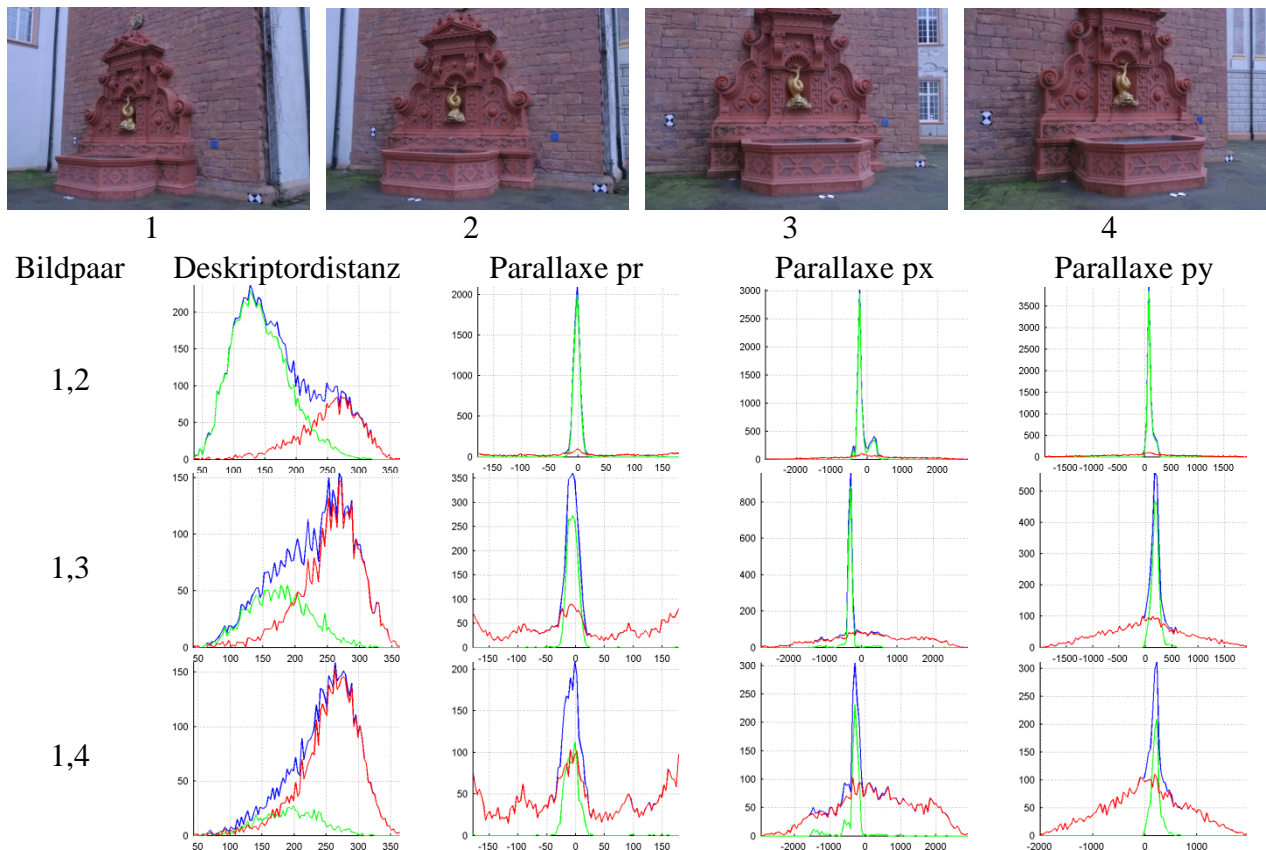


Abb. 1: Oben: Vier Bilder des *FountainP11* Datensatzes. SIFT-Punkte des ersten Bildes wurden rein merkmalsbasiert denen der anderen Bilder zugeordnet, sodass sich drei Paare ergeben. Unten: Die Graphen stellen die Histogramme der Deskriptordistanzen, sowie der Parallaxen p_r , p_x und p_y dar. Dabei ist das Histogramm über alle Zuordnungen in blau dargestellt. Die Histogramme der korrekten Zuordnungen sind in grün, die der Fehlzusordnungen in rot dargestellt. Man erkennt, dass die Deskriptordistanz mit zunehmend schlechter Bildkonfiguration an Aussagekraft verliert. Systematische Effekte in den Parallaxen bleiben hingegen weitestgehend erhalten.

Für die drei Merkmale p_r , p_x und p_y definieren wir einen dreidimensionalen Raum. Um diesen Raum vom Merkmalsraum des Deskriptors zu unterscheiden, bezeichnen wir ihn im weiteren Verlauf als Korrespondenzraum. Auch im vollständigen Korrespondenzraum sind die zuvor beschriebenen systematischen Häufungen zu erkennen (Abb. 2).

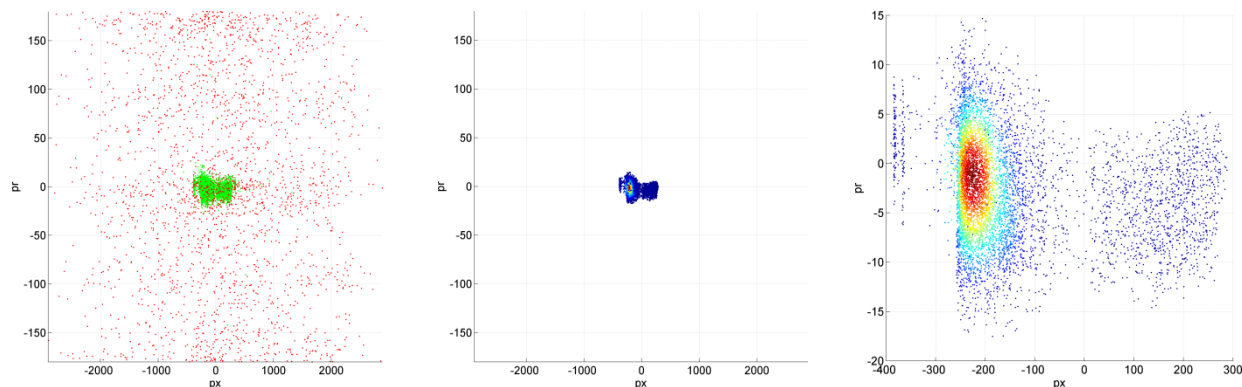


Abb. 2: Darstellung zweier Dimensionen des Korrespondenzraums für Bildpaar 1,2 des vorherigen Beispiels. Links: Mittels bekannter Orientierung gefilterte Zuordnungen (korrekte Zuordnungen in grün dargestellt, falsche in rot). Mitte: Mittels Kerndichteschätzung positiv klassifizierte Zuordnungen. Die Farbe indiziert die geschätzte Kerndichte (steigend von blau nach rot). Rechts: Mittlere Darstellung zur Verdeutlichung vergrößert.

Man kann jedoch auch vermuten, dass eine Bewertung über Histogramme zu einer eher undifferenzierten Klassifizierung führen könnte. Die Häufungen können beliebige Konturen annehmen, welche durch die Diskretisierung des Korrespondenzraums nur ungenügend wiedergespiegelt würden. Des Weiteren führte die diskrete Quantisierung eines Histogramms, gerade bei Vorhandensein nur weniger korrekter Zuordnungen, zu einem erschwerten Vergleich der Zuordnungen untereinander. Dies wäre für eine spätere Verwendung im Rahmen eines gesteuerten Samplings unbefriedigend. Aus diesen Gründen greifen wir auf Kerndichteschätzung als Alternative zu Histogrammen zurück.

2.2 Multivariate Kerndichteschätzung

Die Kerndichteschätzung (PARZEN, 1962) ist, wie Histogramme auch, eine Möglichkeit zur Schätzung nichtparametrischer Wahrscheinlichkeitsdichtefunktionen und wird in der Mustererkennung häufig für Klassifizierungsaufgaben verwendet (BISHOP, 2006). Im Gegensatz zu Histogrammen liefert sie allerdings eine stetige Funktion, welche an jedem beliebigen Punkt t berechnet werden kann. Sie basiert auf der Überlagerung von Kernfunktionen k (wir wählen den Gaußkern), welche mit einer gewissen Bandbreite h um die Stichproben⁴ x_i gebildet werden. Die Summe aller Funktionswerte der Kerne am betrachteten Ort ergibt die entsprechende Kerndichte f . Die Erweiterung auf mehrdimensionale Probleme wird als multivariate Kerndichteschätzung (1) bezeichnet. Die Anzahl der Dimensionen wird mit d bezeichnet. Die Matrix S dient zur Skalierung des Korrespondenzraums auf den Einheitskubus, enthält also auf der Hauptdiagonalen die Kehrwerte der jeweiligen Wertebereiche. In Versuchen ermittelten wir, dass Werte zwischen $1/200$ bis $1/150$ für die Bandbreite h gute Ergebnisse liefern, sodass wir standardmäßig $h = 1/175$ verwenden.

⁴ In diesem Zusammenhang entspricht die Stichprobe einer einzelnen Zuordnung.

$$f_n(t) = \frac{1}{nh^d} \sum_{i=1}^n k(u_i) \quad (1)$$

mit

$$k(u_i) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}u_i' S u_i} \quad (2)$$

und

$$u_i = \frac{1}{h}(t - x_i) \quad (3)$$

2.3 Filterung

Wir berechnen die Kerndichte an den Orten der Zuordnungen und sortieren die Ergebnisse nach aufsteigender Kerndichte. Wir tasten den Wertebereich in festen Intervallen ab und ermitteln jeweils die Anzahl der Zuordnungen mit höherer Kerndichte. Für die sich so ergebende Kurve ermitteln wir den Ort (Kerndichtewert) an dem die Krümmung maximal ist. Wir gelangen somit auf einfache Weise zu einem adaptiven Schwellwert T (Abb. 3). Wir klassifizieren alle Zuordnungen mit höherer Kerndichte als positiv (Inlier), alle übrigen als negativ und verwerfen diese.

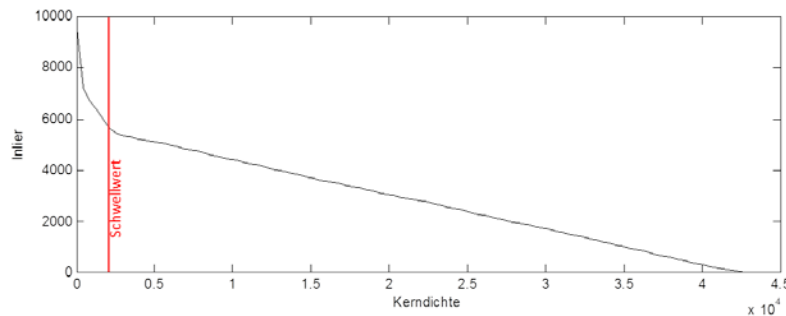


Abb. 3: Adaptive Schwellwertbestimmung: Mit steigendem Kerndichteschwellwert sinkt die Zahl der höher bewerteten Zuordnungen (Inlier, schwarze Kurve). Als endgültiger Schwellwert dient die Kerndichte, bei der die Krümmung der Kurve maximal ist (rot markiert).

2.4 Bildähnlichkeit

Mit der Anzahl der Zuordnungen vor und nach der Filterung, nv und nn , können wir mit nn/nv ein Maß für die Reduktion der Anzahl der Zuordnungen bestimmen. Bei Bildpaaren ähnlichen Inhalts bleiben viele Zuordnungen erhalten, sodass das Maß hier höhere Werte annimmt. Bei unähnlichen Bildpaaren geschieht das Gegenteil. Des Weiteren kann durch die Anpassung des Schwellwertes während der Filterung eine Aussage über die Intensität der Filterung getroffen werden. Hierzu setzen wir den Schwellwert T ins Verhältnis zur maximalen Kerndichte $\max(f)$. Wir kombinieren diese beiden Faktoren zu einem Maß für die **Bildähnlichkeit** sim (4).

$$sim = \frac{nn}{nv} \left(1 - \frac{T}{\max(f)}\right) \quad (4)$$

3 Evaluation

Um das Verfahren zu evaluieren, verwenden wir Bilddatensätze mit bekannter innerer und äußerer Orientierung. Die Datensätze *FountainP11* (Abb. 1) und *CastleP30* (Abb. 7) wurden als Benchmark-Datensätze veröffentlicht (STRECHA et al., 2008) und stehen online⁵ zur Verfügung. Alle weiteren verwendeten Datensätze wurden mittels *Agisoft PhotoScan* orientiert, um zu Referenzlösungen zu gelangen. Wir ermitteln zunächst die wahren korrekten sowie die wahren falschen Zuordnungen über einen Schwellwert von zwei Pixeln für die Distanz der Punkte zur Epipolarlinie. Wir klassifizieren die Zuordnungen des Weiteren mit dem in Kapitel 2 beschriebenen Verfahren und bezeichnen die beiden Klassen wie zuvor als positiv und negativ. Wir können diese dann in vier Gruppen einteilen: korrekte positive kp , falsche negative fn , korrekte negative kn und falsche positive fp . Hieraus stellen wir folgende Kenngrößen auf, um eine Aussage über die Qualität der Filterung zu treffen. Die **Erhaltungsrate** $kp/(kp + fn)$ beschreibt, wie viele der korrekten Zuordnungen nach der Filterung erhalten bleiben. Die **Eliminationsrate** $kn/(kn + fp)$ beschreibt hingegen, wie viele der Fehlzusammenordnungen entfernt wurden. Sind beide Werte hoch, konnte eine klare Trennung der Fehlzusammenordnungen von den Inliers erfolgen. Des Weiteren stellen wir die **ungefilterte Inlierrate** $(kp + fn)/nv$ sowie die **gefilterte Inlierrate** kp/nn dar.

Zur Darstellung der vier Kenngrößen, sowie der Bildähnlichkeit (Kapitel 2.4), aller möglichen Bildpaare eines gesamten Datensatzes bietet sich eine Matrixdarstellung an. Zeilen- und Spaltenindex der Matrix entsprechen den Bildnummern des jeweiligen Bildpaares. Die Kenngröße wird an der jeweiligen Stelle farbig angezeigt. Aufgrund der Vertauschbarkeit der Indizes besteht eine Symmetrie gegenüber der Hauptdiagonalen. Alle im Folgenden dargestellten Größen haben einen Wertebereich von 0 bis 1 und werden mit identischer Skala farbcodiert.

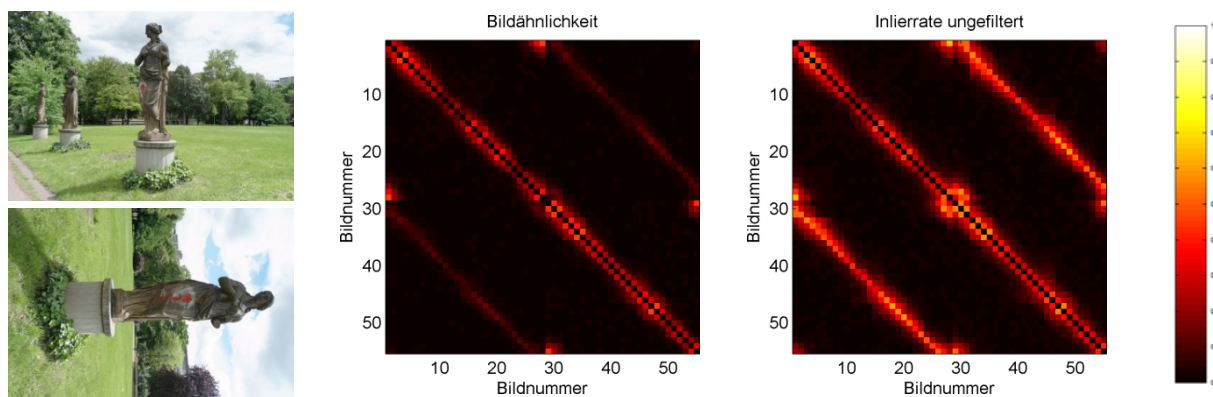


Abb. 4: 55 Bilder (Pyramidenstufe 3, 536x356pix, 2 Ringverbände) einer Statue. Zwei Beispielbilder sind auf der linken Seite dargestellt. Mittig ist die Bildähnlichkeit für den gesamten Datensatz farbcodiert in Matrixform dargestellt. Sie korreliert mit der rechtsseitig dargestellten Inlierrate, was als Indiz für ihre Korrektheit angesehen werden kann. Der große Vegetationsanteil führt zu hoher Selbstähnlichkeit. Außerdem führt der Abstandsunterschied zwischen Vorder- und Hintergrund zu Multimodalität der Kerndichte.

⁵ <http://cvlabwww.epfl.ch/~strecha/multiview/denseMVS.html>

Abb. 4 stellt die ermittelten Bildähnlichkeiten für einen Datensatz von 55 Aufnahmen einer Statue in einem Park dar. Die zwei erfassten Ringverbände bilden sich als zwei diagonale Streifen ab. Zur groben Verifizierung der Bildähnlichkeit kann die strukturelle Übereinstimmung mit der bekannten Inlierrate herangezogen werden. Der Datensatz weist zwei Hauptprobleme auf. Zum einen besteht der Bildinhalt zu großen Teilen aus Vegetation. Vegetation führt bei merkmalsbasierter Zuordnung meist zu einem sehr hohen Anteil an Fehlzuordnungen, da die Textur von Vegetation nur ungenügend eindeutig ist. Man spricht hier auch von hoher Selbstähnlichkeit. Des Weiteren besteht ein großer Abstandsunterschied zwischen Vorder- und Hintergrund, was zu zwei getrennt auftretenden Häufungsbereichen in den lateralen Parallaxen führt. Die sich ergebende Kerndichtefunktion ist also multimodal.

Bei dem in Abb. 5 verwendeten Datensatz handelt es sich um eine Auswahl von 188 Bildern eines Tympanons, welche mit einer monochromen Industriekamera erfasst wurden (ABDELWAHAB, 2012). Die ausgewählten Bilder wurden senkrecht zur Hauptebene des Reliefs gemacht, sodass die Bildkonfiguration der eines Luftbildverbandes ähnelt. Die Oberfläche des Granitreliefs ist komplex und weist große Tiefenunterschiede auf. Im Gegensatz zum vorherigen Datensatz handelt es sich nicht zwingend um Tiefensprünge, sondern häufig eher um Tiefenverläufe. Dies führt zu einer Streckung der dichten Bereiche im Korrespondenzraum. Zusätzlich existieren viele texturschwache Regionen, welche wie im vorherigen Beispiel zu hoher Selbstähnlichkeit führen. In Abb. 6 stellen wir für diesen Datensatz die weiteren Kenngrößen dar.

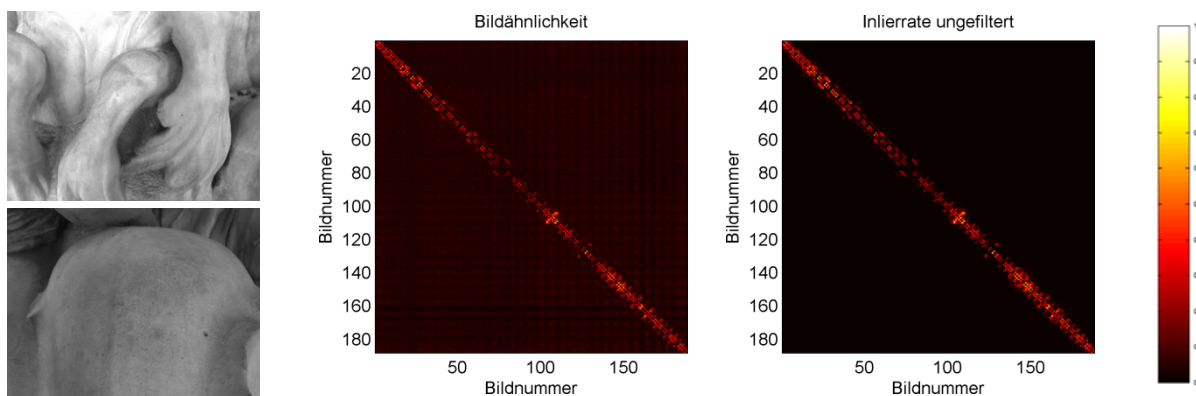


Abb. 5: 188 monochrome Nahaufnahmen (1600x1200pix) eines Granitreliefs. Die Bildkonfiguration ähnelt der eines klassischen Luftbildverbandes. Auch hier korreliert die ermittelte Bildähnlichkeit mit der Inlierrate. Die komplexe Oberfläche weist größere Tiefenunterschiede auf, welche zu einer Streckung der Verteilung der positiven Zuordnungen im Korrespondenzraum führt.

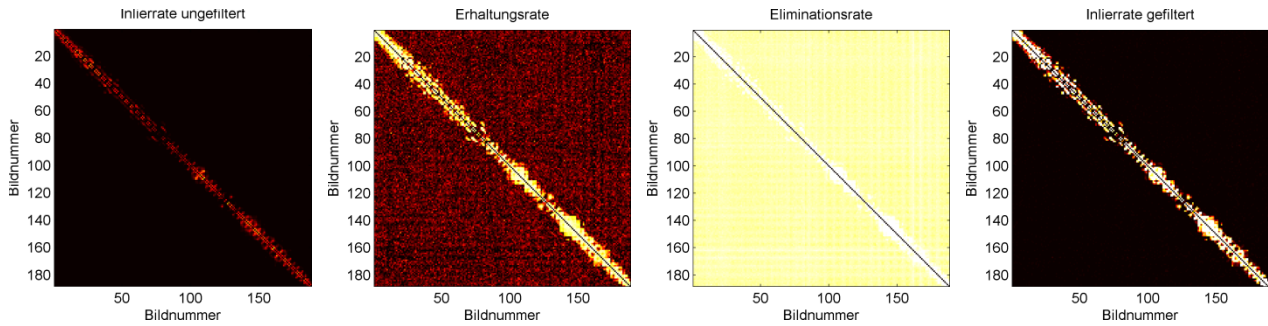


Abb. 6: Weitere Kenngrößen für den Datensatz aus Abb. 5 (identische Farbskala). Links ist erneut die Inlierrate vor der Filterung dargestellt. Die Erhaltungsrate ist bei tatsächlich verknüpften Bildern hoch, was bedeutet, dass Inlier bei der Filterung weitestgehend erhalten blieben. Die Eliminationsrate ist beim gesamten Datensatz sehr hoch. Fehlzusammenhänge wurden also nahezu vollständig entfernt. Dementsprechend hoch liegt die Inlierrate nach der Filterung, welche rechts abgebildet ist.

Abb. 7 stellt vier Beispielbilder aus dem zuvor erwähnten Datensatz *CastleP30* dar. Es handelt sich um 30 Aufnahmen eines Innenhofes, welcher sich vor allem durch wiederholte Fassadenelemente auszeichnet. Erneut führt dies zu einer hohen Selbstähnlichkeit der Bilder. Im Vergleich zu den vorherigen Beispielen kommt hier erschwerend hinzu, dass die sich wiederholenden Elemente eine regelmäßige Verteilung aufweisen und somit auch Fehlzusammenhänge ein gewisses Maß an Systematik im Korrespondenzraum erzeugen. In Abb. 8 stellen wir für den gesamten Datensatz die Kenngrößen wie in Abb. 6 dar. Auch hier kann eine deutliche Trennung von Inliern und Fehlzusammenhänge erreicht werden.



Abb. 7: Vier Bilder aus dem *CastleP30* Datensatz. Viele wiederholt auftretende, regelmäßig angeordnete Fassadenelemente bestimmen das Erscheinungsbild. Dies führt zu hoher Selbstähnlichkeit. Die Regelmäßigkeit der Fassadenelemente erzeugt in gewissem Maße systematische Effekte in den Fehlzusammenhänge.

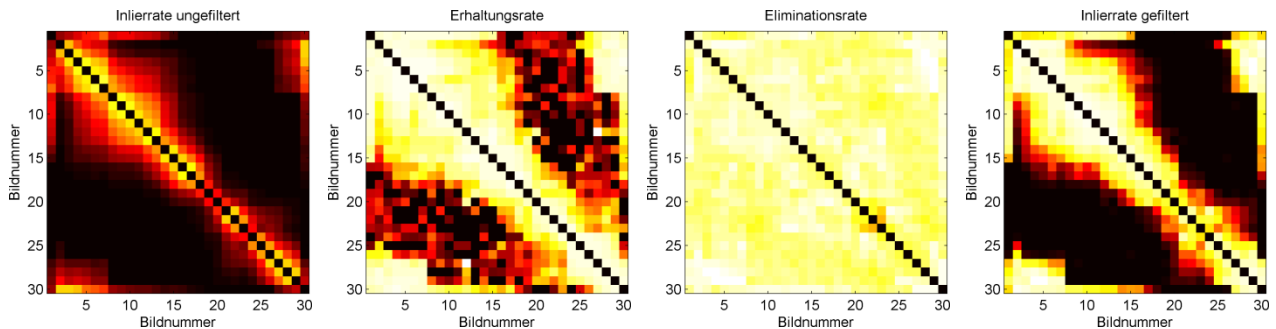


Abb. 8: Kenngrößen für den gesamten Datensatz *CastleP30*. Auch hier wird durch die Filterung eine deutliche Steigerung der Inlierrate erreicht. Hohe Erhaltungsrate und Eliminationsrate indizieren eine gute Trennung der Fehlzusammenhänge von den korrekten Zusammenhänge.

In Tab. 1 stellen wir Ergebnisse numerisch dar, welche sich für die sechs Bildpaare ergeben, die aus den Bildern in Abb. 7 gebildet werden können. Zusätzlich zur Bildähnlichkeit sowie der Inlierate vor und nach der Filterung geben wir sowohl für den 5-Punkte-Algorithmus (5PA), als auch für den 7-Punkte-Algorithmus (7PA), die resultierenden RANSAC-Iterationen an (bei einer Zuverlässigkeit $p=0.999$). Zur Veranschaulichung der Robustheit des Verfahrens gegenüber Auflösungsänderungen sind die Ergebnisse zusätzlich zur vollen Auflösung (Pyr 0) auch für die dritte Pyramidenstufe (Pyr 3) angegeben.

Tab. 1: Eckdaten der Filterung für die Bilder aus Abb. 7

	Paar	Bild- ähnlichkeit %	Inlier ungefiltert %	Inlier gefiltert %	Iterationen ungefiltert 5PA	Iterationen gefiltert 5PA	Iterationen ungefiltert 7PA	Iterationen gefiltert 7PA
Pyr 0	1,2	46,80	47,98	96,70	269	4	1177	5
	1,3	30,56	31,84	96,35	2109	4	20837	5
	1,4	10,92	8,36	58,77	1687110	96	241138760	282
	2,3	65,94	66,18	95,47	51	5	121	6
	2,4	41,33	36,82	83,77	1018	13	7531	21
	3,4	54,83	57,86	95,17	104	5	315	6
Pyr 3	1,2	44,02	53,38	96,24	156	4	556	5
	1,3	27,14	40,09	88,57	664	9	4149	13
	1,4	9,40	20,10	70,37	21062	37	521505	78
	2,3	54,78	62,83	96,13	68	5	176	5
	2,4	33,68	45,96	85,83	334	12	1591	17
	3,4	47,63	71,82	100,00	33	1	67	1

4 Diskussion und Ausblick

Die Beispiele zeigen, dass der vorgestellte Ansatz für verschiedenste Szenarien erfolgreich anwendbar ist. Er zeigt sich robust gegenüber Auflösungsänderungen und ist in der Lage, Zuordnungen korrekt zu klassifizieren, deren Verteilung im Korrespondenzraum starker Ausdehnung unterliegt. Auch multimodale Verteilungen können aufgrund der nichtparametrischen Natur der Kerndichteschätzung verarbeitet werden. Die Elimination der Fehlzuordnungen liegt in den meisten Fällen bei über 90% und führt zu einer deutlichen Zunahme der Inlier-Rate. Die in Tab. 1 angegebenen Iterationszahlen sind lediglich für die Basisvariante des RANSAC berechnet und somit eher in ihrer Tendenz zu interpretieren. Sie heben jedoch deutlich die exponentielle Natur des zugrundeliegenden Verfahrens hervor. Umso geringer die Inlierate vor der Filterung ausfällt, desto größer ist der Nutzen der Vorfilterung, selbst wenn die Inlierate nach der Filterung nur mittelmäßig ausfällt. Ein Extremfall ist mit Bildpaar 1,4 in Tab. 1 gegeben. Natürlich bleibt fraglich, ob ein derart schlecht konfiguriertes

Bildpaar in die Auswertung der Bündeltriangulation aufgenommen werden sollte. Es wäre jedoch nach der Vorfilterung mit recht geringem Aufwand möglich.

Das von uns vorgeschlagene, a priori einfach zu berechnende, Maß der Bildähnlichkeit ist gut zur Beurteilung der Verwendbarkeit eines Bildpaares geeignet. Wir stellen allerdings fest, dass wir für Bilddatensätze mit hoher Selbstähnlichkeit eindeutigere Ergebnisse erzielen, wenn wir diese auf höherer Pyramidenstufe bearbeiten. Ein Vorgehen, welches ohnehin üblich ist, da die Bearbeitungszeit hier naturgemäß wesentlich geringer ausfällt.

Eine optimalere Trennung der Fehlzuordnungen von den Inliern könnte durch Hinzunahme weiterer Charakteristika, wie z.B. die bei skaleninvarianter Zuordnung ermittelte Pyramidenstufe oder ggf. verfügbare Merkmalsradien, erreicht werden. Wir gehen außerdem davon aus, dass auch ein Training der Bandbreite h sowie des Schwellwertes T hierzu einen positiven Beitrag leisten könnte. Die zu erwartenden Verbesserungen lägen in einem stabileren Verhalten des Bildähnlichkeitsmaßes und einer höheren Erhaltungsrate. Grundsätzlich könnten auch Verfahren der Kerndichteschätzung mit adaptiver Bandbreite hinsichtlich ihrer Eignung untersucht werden. Für die Zukunft planen wir außerdem, die ermittelte Kerndichte innerhalb eines gesteuerten RANSAC-Verfahrens weiterzuverwenden.

5 Literaturverzeichnis

- ABDEL-WAHAB, M., WENZEL, K., FRITSCH, D., 2012: Automated and Accurate Orientation of Large Unordered Image Datasets for Close-Range Cultural Heritage Data Recording. Photogrammetrie - Fernerkundung - Geoinformation (PFG). Heft 6, S. 679-690.
- BISHOP, C., M., 2006: Pattern Recognition and Machine Learning. Springer Verlag
- CHUM, O. & MATAS, J., 2005: Matching with PROSAC – Progressive Sample Consensus. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), S. 220-226.
- CHUM, O. & MATAS, J., 2008: Optimal Randomized RANSAC. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), S. 1472-1482.
- CURSKA, G., DANCE, C., R., FAN, L., WILLAMOWSKI, J., BRAY, C., 2004: Visual Categorization with Bags of Keypoints. ECCV International Workshop on Statistical Learning in Computer Vision.
- FISCHLER, M. & BOLLES, R., 1981: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the Association for Computing Machinery (ACM), S. 381-395.
- HARTLEY, R., I., 1997: Projective Reconstruction and Invariants from Multiple Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 16, S. 1036–1041.
- LI, X. & HU, Z., 2010: Rejecting Mismatches by Correspondence Function. International Journal of Computer Vision, S. 1-17.
- LOWE, D., G., 2004: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, Volume 60, S. 91-110.
- MIZOTIN, M., KRIVOVYAZ, G., VELIZHEV, A., CHERNYAVSKIY, A. & SECHIN, A., 2010: Robust Matching of Aerial Images with Low Overlap. ISPRS Archives, Volume, S. 13 -18.
- NISTER, D., 2004: An Efficient Solution to the Five-Point Relative Pose Problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 26, S. 756 -770.

- OLIVA, A. & TORRALBA, A., 2001: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, Volume 42, S. 145-175.
- PARZEN, E., 1962: On Estimation of a Probability Density Function and Mode. *Annals of Mathematical Statistics*, Volume 33, S. 1065-1076.
- RODEHORST, V. & HELLWICH, O., 2006: Genetic Algorithm Sample Consensus (GASAC) - A Parallel Strategy for Robust Parameter Estimation. *IEEE Workshop on Computer Vision and Pattern Recognition (CVPRW)*.
- RUBLEE, E., RABAUD, V., KONOLIGE, K., BRADSKI, G., 2011: ORB: an Efficient Alternative to SIFT and SURF. *IEEE International Conference on Computer Vision (ICCV)*, S. 2564-2571.
- STRECHA, C., VON HANSEN, W., VAN GOOL, L., FUA, P., THOENNESSEN, U., 2008: On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, S. 1-8.
- TORDOFF, B. & MURRAY, D., W., 2002: Guided Sampling and Consensus for Motion Estimation. *European Conference on Computer Vision (ECCV)*, S. 82 – 96.