

Variations- und Deep-Learning-Ansätze bei der intrinsischen Zerlegung von Lichtfeldern ¹

Anna Alperovich²

Abstract: Bei der intrinsischen Bildzerlegung geht es darum, ein beleuchtungsinvariantes Reflexionsbild von einem Eingangsfarbbild zu trennen, was nach wie vor noch eines der grundlegenden Probleme im Bereich der Computer Vision darstellt. Diese Zerlegungsart wird häufig bei der Bearbeitung von Fotos und Materialien, der Bildsegmentierung sowie der Formschatzung eingesetzt. Im Fokus dieser Arbeit liegt die intrinsische 4D-Zerlegung eines Lichtfelds. Im Rahmen dessen soll das Problem in Bezug auf die folgenden drei Variablen formuliert und gelöst werden soll: Albedo, Schattierung und Spekularität. Dadurch wird es wiederum möglich, sich mit nicht-Lambertschen Szenen auseinanderzusetzen. Dem Problem soll sich mit Variations- und Deep-Learning-Ansätzen angenähert, ihre Leistung verglichen und die Stärken und Schwächen beider Techniken diskutiert werden. Es soll nachgewiesen werden, dass der in dieser Arbeit vorgestellte Deep-Learning-Ansatz eine generische Lösung für Lichtfelder darstellt und bei vier zeitgenössischen Computer-Vision-Problemstellungen eingesetzt werden kann: Disparitätsschatzung, Reflexionstrennung, intrinsische Bildgebung und bildverarbeitende Ultrahochauflösung. Umfangreiche Auswertungen auf der Grundlage mehrerer öffentlich zugänglicher, synthetischer und realer Datensätze belegen die Fruchtbarkeit der im Rahmen dieser Arbeit vorgestellten Methodik. Im Ergebnis werden die Vorteile der Verwendung von Lichtfeldern gegenüber anderen Datenstrukturen aufgezeigt.

1 Einführung

Der Fokus liegt dabei auf inversen Problemen, bei denen bestimmte Komponenten einer Szene aufgrund ihrer fotografischen Bilder wiederhergestellt werden sollen. Es werden zwei Ansätze untersucht: Bei dem ersten Ansatz wird die Modellierung des Problems auf der Grundlage physikalischer Prinzipien der Bilderzeugung untersucht. Bei dem zweiten Ansatz kommt Deep Learning zum Einsatz.

Die vorliegende Arbeit unterscheidet sich in vielerlei Hinsicht von vergleichbaren Forschungsvorhaben auf diesem Gebiet. Erstens wird in dieser Untersuchung anstelle eines Einzelbildes einer Szene eine Serie von Bildern, die aus einer etwas anderen Perspektive (Lichtfelder, siehe Abb. 1) aufgenommen wurden, als Eingabe für die verwendeten Algorithmen verwendet. In früheren Untersuchungen wurde aufgezeigt, dass vom Lichtfeld übernommene umfangreiche Informationen verwendet werden können, um die Geometrie zuverlässig abzuschätzen und verschiedene physikalische Eigenschaften einer Szene wiederherstellen zu können. Zweitens liegt der Fokus in dieser Arbeit auf den nicht-Lambertschen Phänomenen, die dank der Lichtfelddaten von der diffusen Reflexion getrennt betrachtet werden können.

¹ Englischer Titel der Dissertation: "Variational and Deep Learning Approaches for Intrinsic Light Field Decomposition"

² Universität Konstanz, ann.alperovich@gmail.com

Diese Arbeit besteht aus zwei Teilen, wobei im ersten Teil Variationsmethoden für intrinsische Lichtfelder diskutiert werden, bei denen das Eingangslichtfeld in drei Rendering-Komponenten zerlegt wird: Albedo, Schattierung und Spekularität. Um diese Komponenten zu modellieren, ist die exakte Geometrie einer Szene erforderlich. Zunächst jedoch soll in dieser Untersuchung die Methode zur Disparitätsschätzung vorgestellt werden, wodurch nicht ausschließlich Disparitätsmarker wiederhergestellt, sondern auch stückweise glatte Normalen-Maps ausgegeben werden können. Im Ergebnis kann die 3D-Darstellung einer Szene berechnet und zur Modellierung der intrinsischen Komponenten verwendet werden. Der größte Nachteil dieser Methode besteht darin, dass zur Berechnung der Disparität davon ausgegangen wird, dass die Szene rein Lambertsch ist, was in der Realität so gut wie nie der Fall ist. Somit kann der vorgeschlagene Algorithmus keine genauen Disparitätsmarker in spiegelnden Bereichen wiederherstellen. Durch die Regularisierung der Normalen kann dieses Problem teilweise für kleine hervorgehobene Bereiche gelöst werden, die von nahezu diffusen Bereichen umgeben sind, aber das Problem bleibt für große spiegelnde Oberflächen bestehen.

Im zweiten Teil wird ein Deep-Learning-Ansatz beschrieben, bei dem die Möglichkeit untersucht wird, ein physikalisches Modell zu beschreiben, indem das neuronale Netz ausschließlich mit Trainingsbeispielen gespeist wird. Hierzu wurde ein 3D-Convolutional-Neural-Network (CNN) entwickelt, um die winkelabhängigen Daten im Lichtfeld zu nutzen und den Mehrwert dieses Netzes im Zusammenhang mit den zeitgenössischen Computer-Vision-Problemen aufzuzeigen. Da im Rahmen dieser Untersuchung das Netz auf die nicht-Lambertschen Daten trainiert wird, wird es dazu gezwungen, spiegelnde Bereiche zu verarbeiten und als Ergebnis die Einschränkung des Modellierungsansatzes zu überwinden. In zahlreichen Versuchen wird die Leistung beider Ansätze verglichen und der Vorteil der

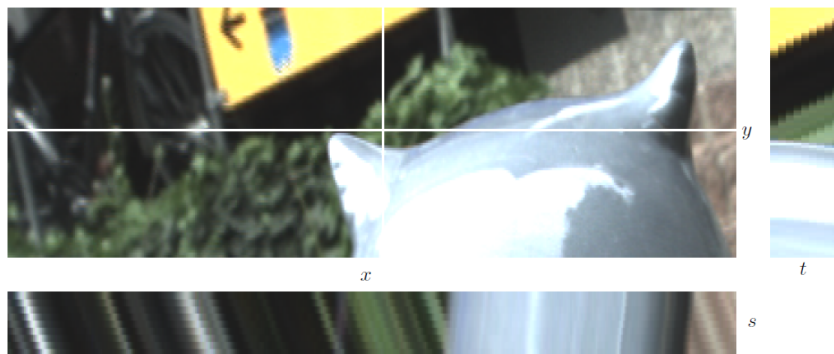


Abb. 1: In der Abbildung oben links ist die Mittelansicht eines Lichtfelds zu sehen, das durch die Bildkoordinaten x und y parametrisiert ist. Unten und rechts sind die Abbildungen der Epipolarebenen (EPIs) für die weißen Linien in der Mittelansicht zu sehen, wobei s und t die Koordinaten der Ansichtspunkte beschreiben. Während sich die Kamera bewegt, zeichnen 3D-Szenenpunkte gerade Linien auf den EPIs, deren Neigung der Disparität entspricht. Jede Zuweisung einer Eigenschaft eines Szenenpunktes zu den entsprechenden Strahlen sollte entlang dieser Linien konstant sein, was zur kohärenten Regulierung genutzt werden kann.

Verwendung eines Deep Networks für die Disparitätsschätzung und die intrinsische Bildgebung veranschaulicht.

Dadurch soll aufgezeigt werden, dass mit dem Lichtfeld als Eingabe sowohl Variations- als auch Deep-Learning-Methoden den Einzelbildmethoden und Bild + tiefen Methoden überlegen sind. Insbesondere soll der Vorteil der Verwendung von Lichtfeldern für spiegelnde Szenen aufgezeigt werden, bei denen die meisten Algorithmen aufgrund der Lambertschen Annahme versagen. Dies würde ein breites Spektrum von Anwendungen und Impulsen für zukünftige Forschungen zum Verständnis nicht-Lambertscher Szenen mit Lichtfeldern eröffnen.

2 Variationsmethoden

2.1 Disparitätsschätzung

Eine häufige Anwendung des Lichtfelds besteht darin, die Tiefe der aufgenommenen Szene zu schätzen, um darüber hinaus ihre 3D-Struktur rekonstruieren zu können. Aktuelle Algorithmen zur Disparitätsschätzung funktionieren außerordentlich gut, wenn die Tiefe der Lichtfeldbilder geschätzt wird. Diese Methoden sind jedoch normalerweise nicht auf die Normalenschätzung ausgelegt. Daher sind die Tiefenschätzungen von Algorithmen, die auf Kostenvolumina basieren, selbst wenn sie mit Sublabel-Genauigkeit [Mo16] optimiert wurden, häufig stückweise flach und können daher keine genauen Normalen-Maps bestimmen. Häufig ist ihre Genauigkeit naturgemäß auch um die Okklusionsgrenzen herum begrenzt. Das Ziel des in dieser Arbeit beschriebenen Disparitätsschätzungsalgorithmus besteht darin, zu einer Beseitigung dieser Nachteile beizutragen, siehe Abb. 2 für eine Beispielausgabe der vorgeschlagenen Methode.

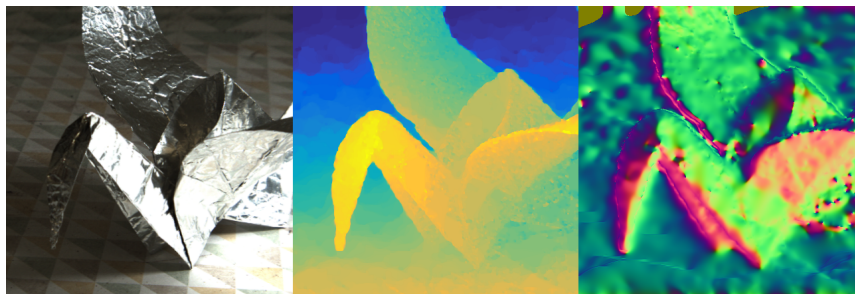


Abb. 2: Es soll ein neuartiger Ansatz zur Berechnung von Disparität-Kostenvolumina, der auf dem Konzept der okklusionsbewussten Fokalstapelsymmetrie basiert, vorgestellt. Mit dem vorgeschlagenen Konzept können die Tiefe als auch die Normalen gemeinsam optimiert werden, um herausfordernde reale Szenen rekonstruieren zu können, die mit einer plenoptischen Kamera (Lytro Illum) aufgenommen werden. In der Abbildung links ist die Mittelsicht des Lichtfelds dargestellt, die obere rechte Abbildung zeigt die Disparitätsmarker und in der unteren rechten Abbildung ist die Normalen-Map zu sehen

Zunächst soll eine neuartige Methode vorgestellt werden, wie mit Okklusionen umgegangen werden kann, wenn Kostenvolumina auf der Idee der Fokalstapelsymmetrie beruhen [Li15]. Mit diesem neuartigen Datenanteil können wesentlich genauere Ergebnisse als mit der vorherigen Methode [Li15] erzielt werden, wenn ein globales Optimum mithilfe der Sublabel-genauen Relaxierung berechnet wird [Mo16]. Zweitens soll eine Nachbearbeitung unter Verwendung einer gemeinsamen Regulierung von Tiefe und Normalen vorgeschlagen werden, um eine glatte Normalen-Map zu erhalten, die mit der Tiefenschätzung übereinstimmt. Hierfür sollen die Ansätze nach Graber et al. [Gr15] zur linearen Verbindung von der Tiefe und den Normalen sowie das Konzept der Relaxation nach Zeisl et al. [ZZP14] Anwendung finden, um sich der Nichtkonvexität der Einheitslängenbeschränkung auf der Normalen-Map anzunähern. Die daraus resultierenden Teilprobleme in Bezug auf Regulierung der Tiefe und den Normalen können mit dem Primal-Dual-Algorithmus effizient gelöst werden. Zum Zeitpunkt der Einreichung dieser Arbeit konnten wesentlich bessere Ergebnisse im Vergleich zu früheren Abhandlungen erzielt werden, in denen neuere Benchmark-Verfahren zur Disparitätsschätzung von Lichtfeldern in Bezug auf die Genauigkeit von Disparitätenkarten und Normal-Maps und mehrere andere Metriken behandelt wurden.

2.2 Intrinsische Lichtfelder

Es soll ein neuartiges Variationsmodell für die intrinsische Zerlegung eines Lichtfelds vorgestellt werden, das anstatt auf herkömmlichen Bildern auf dem 4D-Strahlenraum definiert wird. Hierzu wird ein intrinsisches Lichtfeld als Funktion modelliert $L(\mathbf{r}) = A(\mathbf{r})S(\mathbf{r}) +$

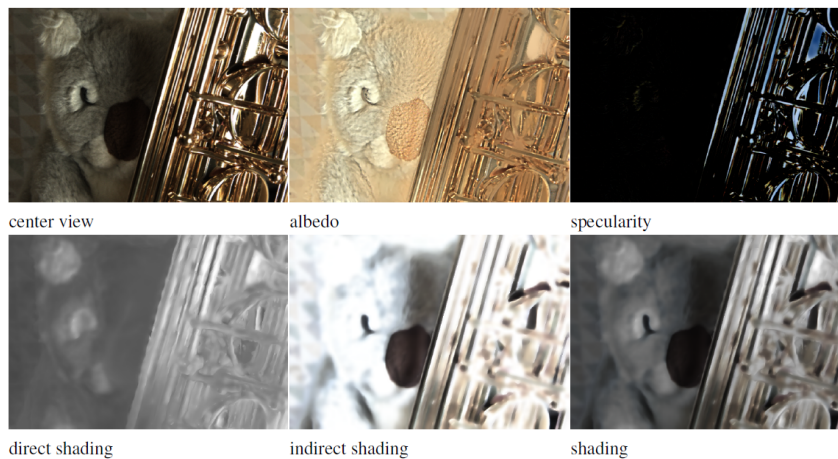


Abb.3: Ein variationsbasierter Ansatz zur Zerlegung der Lichtfeldstrahlung in die intrinsischen Komponenten Albedo, Schattierung und Spekularität. Durch die weitere Aufteilung der Schattierung in direkte und indirekte Anteile wird im Vergleich zu früheren Arbeiten eine herausragende Leistung erzielt.

$H(\mathbf{r})$, wo die Strahlung L jedes Strahls \mathbf{r} in Albedo A , Schattierung S und Glanzfarben-Komponente H zerlegt wird. Die Funktionen $L, A, S, H : \mathcal{R} \rightarrow \mathbb{R}^3$ ordnen dem Strahlenraum entsprechende RGB-Farbwerte zu. Albedo steht für die Farbe eines Objekts, unabhängig von der Beleuchtung und der Kameraposition. Die Schattierung beschreibt Intensitätsänderungen aufgrund von Beleuchtung, Interreflexionen und der Objektgeometrie. Schließlich stellt die Spekularitätskomponente Glanzfarben dar, die an nicht-Lambertschen Oberflächen auftreten, was von der Beleuchtung, der Objektgeometrie sowie der Kameraposition abhängt.

Da die meisten vorhandenen intrinsischen Bildalgorithmen für Lambertsche Szenen entwickelt wurden, wird ihre Leistung dann beeinträchtigt, wenn Szenen betrachtet werden, die glänzende Oberflächen aufweisen. Im Gegensatz dazu wird es durch die reichhaltige Struktur des Lichtfelds mit vielen dicht abgetasteten Ansichten möglich, sich nicht-Lambertschen Oberflächen zu widmen, indem ein zusätzlicher Zerlegungsterm eingeführt wird, der die Spekularität modelliert. Die Regulierung entlang der Bilder in der Epipolarebene fördert die Albedo- und Schattierungsbeständigkeit in allen Ansichten.

Das ursprüngliche Modell soll erweitert werden, indem an diesem gewisse Neuronen vorgenommen werden, bevor sich der Schattenbildung und Zwischenreflexionen gewidmet wird. Somit besteht das neue Zerlegungsmodell $L(\mathbf{r}) = A(\mathbf{r})s_d(\mathbf{r})S_i(\mathbf{r}) + H(\mathbf{r})$ aus den Komponenten s_d und S_i , die die direkte und indirekte Schattierung beschreiben, siehe Abb. 3. Unter direkter Schattierung kann die Schattierung verstanden werden, die ein Objekt hätte, wenn es sich allein in einer Szene befinden würde, d. h. ohne das Vorhandensein von Schatten oder reflektiertem Licht. Durch die zweite Komponente S_i werden Reflexionen und Schattenbildungen modelliert.

Im Gegensatz zum ersten Ansatz, bei dem die Interreflexion nur auf der Grundlage der Geometrie modelliert wird, wird in dieser Abhandlung die indirekte Schattierung mittels einer Kombination aus Geometrie- und Farbinformationen modelliert. Es wird ein Konfidenzmaß zur Beschattung für das Lichtfeld berechnet, welches zur Regulierung herangezogen wird. In umfangreichen Experimenten konnte nachgewiesen werden, dass das hier ausgewählte Forschungsdesign die Schätzung der Schattierungskomponente signifikant verbessert. Eine weitere Verbesserung besteht in dem Hinzufügen einer Frequenzanalyse bei der Spekularitätsschätzung, die auf der Annahme basiert, dass hohe Variationen der Pixelintensitäten eines 3D-Punkts in Ansichten mit Subapertur durch Spekularität verursacht werden. Die neuen Schwerpunkte münden somit in einen neueren Ansatz, um das Einganglichtfeld in die Komponenten Albedo, Schattierung und Spekularität zerlegen zu können.

3 Deep-CNN für Lichtfelder

Lichtfelder weisen eine komplexe, stark redundante Struktur auf, siehe Abb. 1 Für Szenen mit rein diffuser Reflexion weisen EPIs Muster orientierter Linien mit konstanter Farbe auf. Die Linien entsprechen der Projektion eines einzelnen 3D-Punkts im Raum, und ihre Neigung oder Disparität ist umgekehrt proportional zur Entfernung des Punkts zum Beob-

achter. Diskontinuitäten im Muster werden durch Okklusionen verursacht, die Übergänge zwischen mehreren Orientierungspunkten an der Okklusionskante erzeugen.

Die Situation wird zudem weniger eindeutig, wenn Reflexion oder glänzende, nicht-Lambertsche Oberflächen ins Spiel kommen, weil die EPIs dann überlagerte Muster aufweisen [JSG16]. Die Ausrichtung der Muster, die der Spiegelreflexion entsprechen, entspricht nicht der Disparität, sondern der Bewegungsrichtung der Spekularität, die von der intrinsischen Oberflächengeometrie abhängt. Um zwischen diesen beiden Fällen zu unterscheiden, ist zu eruieren, ob ein Punkt eine diffuse oder spiegelnde Reflexion aufweist. Mit einer bekannten Geometrie kann die Bewegungsrichtung der Spekularität direkt geschätzt und Reflexionskomponenten voneinander getrennt werden [Su16]. Wenn sowohl Form als auch Reflexionsvermögen unbekannt sind, ist es kaum möglich zu bestimmen, welche Phänomene zu einem bestimmten EPI geführt haben.

Dennoch weisen EPIs aus natürlichen Lichtfeldern eine insgesamt regelmäßige Struktur auf. Zudem ist es wahrscheinlich, dass sie im gesamten Bildraum der Epipolarebene eine vergleichsweise niedrigdimensionale Mannigfaltigkeit bilden. Darüber hinaus hängt eine entsprechende Codierung eines EPIs mit nur wenigen Parametern mit den komplexen miteinander verzahnten Aufgaben zusammen, wie z. B. der Disparitätsschätzung oder der Trennung von Albedo-, Schattierungs- und Spekularitätskomponenten. Die Vermutung liegt nahe, dass Sie die anderen Aufgaben besser meistern können, wenn Sie eruieren, wie man eine Komprimierung erfolgreich ausführt. Gegenstand dieses Teils der Arbeit ist es daher, eine niedrigdimensionale Darstellung von EPIs aus beliebigen Beispiellichtfeldern zu

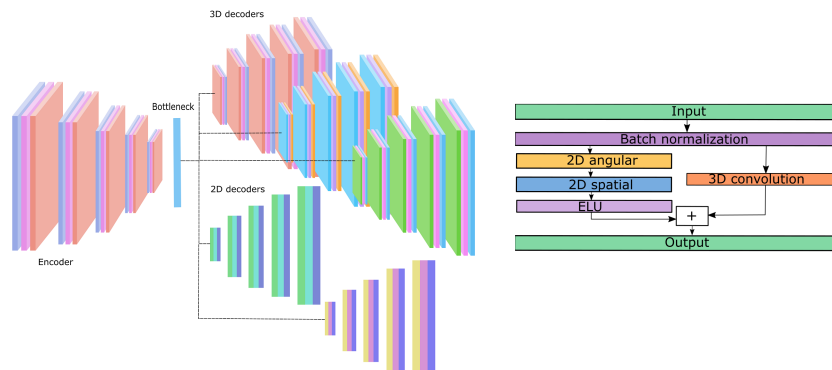


Abb. 4: **Links:** Übersicht über die vorgeschlagene Architektur. Abhängig von der verfügbaren Aufgabe und dem verfügbaren GPU-Speicher können Anzahl und Inhalt der Gruppen geändert werden. Die Decoderpfade stellen exakte Spiegelbilder dieser Kette dar. Die Abmessung der Ansichtspunkte der Form in den 2D-Decodern wird entfernt. Von dem Netzwerk wird eine beliebige Anzahl von 2D- und 3D-Decodern unterstützt. **Recht:** Beispiel eines residualen Block des Netzes. Nach der Batch-Normalisierung wird der Ausgangstensor über zwei Wege übergeben. Der rechte behält den Eingangstensor bei oder führt eine 3D-Faltungsberechnung durch, wenn er erneut abgetastet werden muss. Die linke führt winkelige und räumliche 2D-Faltungsberechnung durch, gefolgt von der ELU-Schicht. Der Ausgangstensor ergibt sich aus der Summe dieser beiden Pfade.

entwerfen, jedoch in einer Art und Weise, dass die latenten Variablen gemeinsam verwendet werden können, um verschiedene angeleitete Aufgaben in der Lichtfeldanalyse genau lösen zu können. Zu diesem Zweck wird in dieser Abhandlung ein neuronales Encoder-Decoder-Netz vorgeschlagen, das auf dem Konzept des Deep Autoencoders [HS06] basiert, welches in jüngster Zeit sehr erfolgreich aussagekräftige mannigfaltige Darstellungen erzeugen konnte.

Es soll die erste Netzwerkarchitektur für das Multitasking-Lernen in Lichtfeldern vorgestellt werden, siehe Abb. 4. Der Grundgedanke ist es, das Netzwerk um einen Autoencoder herum aufzubauen, damit es ohne Anleitung nur mit rohen Lichtfelddaten trainiert werden kann. Hierbei werden jedoch mehrere Pfade hinzugefügt, um die latente Darstellung zu dekodieren, die gemeinsam mit dem Autoencoder unter Anleitung trainiert werden können, je nachdem, welche Daten im aktuellen Trainingsbeispiel verfügbar sind. Durch die Kombination von angeleitetem und unbeaufsichtigtem Training kann sichergestellt werden, dass die latente Darstellung nicht ausschließlich auf die gewünschten Aufgaben wie Tiefenrekonstruktion oder die intrinsische Komponentendarstellung angewendet werden kann, sondern sich auch gut auf Datensätze übertragen lässt, bei denen für die auszuführenden Aufgaben keine Trainingsinformationen verfügbar sind. Sobald das Netz bereitgestellt wird, können alle Decoderketten nur anhand der Lichtfelddaten ausgewertet werden. In dieser Abhandlung liegt der Fokus auf den vier zeitgenössischen Computer-Vision-Problemen, Disparitätsschätzung, intrinsische Bildgebung, Reflexionstrennung und bildverarbeitende Ultraauflösung.

CNN zur Reflexionstrennung und Disparitätsschätzung. Gemäß dem dichromatischen Reflexionsmodell verfügt das von einem Szenenpunkt reflektierte Licht über zwei unabhängige Komponenten: Licht, das vom Oberflächenkörper und an der Grenzfläche reflektiert wird. Die Objektreflexion ist als diffuse Komponente bekannt und unabhängig von der Blickrichtung, während es sich bei der Grenzflächenreflexion um Folgendes handelt

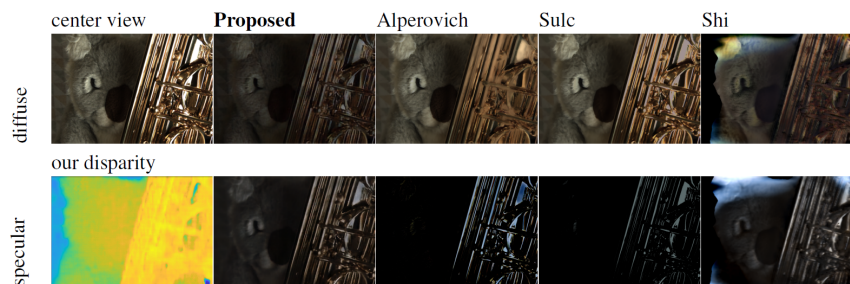


Abb. 5: Mit der plenoptischen Kamera (Lytro Illum) aufgenommene Lichtfelder. Die Szene besteht aus einem hochglänzenden Saxofon und einem fast lambertianischen Koala. Das in dieser Arbeit vorgestellte Netz erkennt im Vergleich zu anderen Methoden mit großem Erfolg mehr spiegelnde Oberflächen des Saxofons. Während der Koala als ein spiegelndes Objekt ähnlich zu [Sh17] fehlinterpretiert wird, ist die hier vorgestellte Methode die einzige, bei der der diffuse Teil hinter dem großen spiegelnden Fleck auf dem Saxofon nicht unscharf dargestellt wird.

die spiegelnde Komponente, die ansichtsabhängig ist. Bei der Trennung der spiegelnden und diffusen Reflexionsanteile handelt es sich um ein inverses Problem, das innerhalb der Computer-Vision-Community nach wie vor ein reges Forschungsgebiet darstellt. Unter Bezugnahme der Arbeit von Sulc et al. [Su16] wird dargestellt, dass das Auftreten von Spekularitäten in einem Lichtfeld von der Szenengeometrie abhängt und entlang der Strömungsrichtungen der Spekularität konstant bleibt, während die Lambertsche Komponente lediglich in der Disparitätsrichtung konstant bleibt. Daher bietet es sich an, eine im Lichtfeld codierte geometrische Information zu verwenden, um eine Reflexionstrennung durchzuführen. Es wird die in Abb. 4 beschriebene vorgeschlagene Architektur übernommen, um gemeinsam Aufgaben in den Bereichen Disparitätsschätzung und Reflexionstrennung auszuführen. In der Abb. 5 werden die im Rahmen dieser Untersuchung erzielten Ergebnisse und Vergleiche zum konkreten Beispiel dargestellt.

Intrinsische Zerlegung eines Lichtfelds auf der Grundlage von Deep-CNN. Im Gegensatz zum vorherigen Netz führen wird im Rahmen dieser Arbeit eine vollständige intrinsische Zerlegung durchgeführt, eine Architektur entworfen, mit der es ermöglicht wird, doppelt größere Patches als Eingabe zu verarbeiten und Sprungverbindungen vom Codierer zu entsprechenden Decoderteilen eingeführt, durch die die Rekonstruktionsqualität von Decodern verbessert wird. 3D-Falten werden durch eine 2D-Sequenz ersetzt, die auf die Orts- und Winkelverteilung einwirken. Eine schematische Beschreibung des Restblocks finden Sie in Abb. 4. Diese Entwurfsauswahl verringert die Anzahl der Parameter im Netz und beschleunigt den Trainingsprozess. Wie im vorherigen Fall kann dieses Netz Lichtfelder verarbeiten, für die keine Ground-Truth-Informationen verfügbar sind.

Generative Adversarial Networks für hochauflösende Lichtfelder. Schließlich soll aufgezeigt werden, dass die vorgeschlagene Encoder-Decoder-Architektur (Abb. 4) für eine völlig andere Aufgabe verwendet werden kann. Mit einigen notwendigen Modifikationen und gemäß den jüngsten Forschungstrends soll ein 3D-Generative-Adversarial-

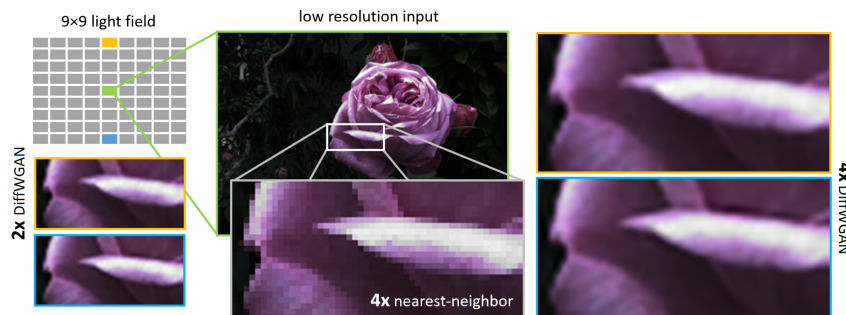


Abb. 6: Ausgabe des ultrahochauflösenden Netzwerks. Bei einer relativ bescheidenen Menge von Eingabeansichten mit Subapertur erzeugt das Netz dennoch ein ultrahochauflösendes Lichtfeld mit den Vergrößerungsfaktoren $\times 2$ und $\times 4$. Es wurde das Interpolations-Verfahren für das computererzeugte Bild mittels des Nächster-Nachbar-Interpolations-Verfahrens durchgeführt und das Ergebnis für den visuellen Vergleich präsentiert.

Autoencoder-Netzwerk vorgestellt werden, um das hochauflösende Lichtfeld aus einem niedrigauflösenden Lichtfeld mit einer geringen Anzahl von Ansichtspunkten wiederherzustellen.

Es werden lediglich drei Ansichten entlang der horizontalen und vertikalen Achse benötigt, um die Winkelauflösung um den Faktor drei zu erhöhen, während gleichzeitig die räumliche Auflösung in jeder Richtung um den Faktor zwei oder vier erhöht wird. In der Abb. 6 sind die Ein- und Ausgänge des Netzes dargestellt.

4 Zusammenfassung

In dieser Arbeit [A120] werden verschiedene Forschungsfragen innerhalb der Computer Vision, insbesondere im Bereich der Lichtfelder, diskutiert, die allesamt eng miteinander verzahnt sind. Zunächst wäre da die Disparitätsschätzung, bei der es sich um ein bewährtes Verfahren im Bereich von Lichtfeldern handelt. Es wurden zahlreiche Algorithmen vorgeschlagen, um Disparitätsmarker zu berechnen, aber nur wenige von ihnen bewerten ihre Qualität für einige Anwendungen, bei denen geometrische Informationen benötigt werden. Es wurde ein Disparitätsschätzungsalgorithmus entworfen, der sich besonders zur Modellierung intrinsischer Komponenten eignet, indem Disparitätsmarker und Normalen-Maps gemeinsam optimiert werden. Die generierte Disparitätskarte wurde dazu verwendet, um eine 3D-Darstellung einer Szene zu rekonstruieren, die wiederum als Grundlage für die Modellierung intrinsischer Komponenten dient.

Im Rahmen dieser Abhandlung wurde ein variierender intrinsischer Zerlegungsmodus für Lichtfelder vorgeschlagen. Um die Lichtfeldstruktur also nutzbar zu machen, wurden die Komponenten Albedo und Schattierung dazu gebracht, entlang der Disparitätsrichtungen konstant zu bleiben. Um spiegelnde Oberflächen zu identifizieren, wird auf die Glanzfarbeigenschaften zurückgegriffen, um ansichtsabhängig operieren zu können. Es wurden die Projektionen eines 3D-Punkts in Subapertur-Ansichten analysiert und die Beschaffenheit der Reflexion ausgewertet. Dabei wurde das Variationsmodell optimiert, indem eine bessere Variante für die Schattierungskomponente eingeführt wurde, die sich durch eine feinere Aufgliederung in direkte und indirekte Anteile auszeichnet. Die Vorteile des neuen Modells wurden in einer Vielzahl von natürlichen und synthetischen Lichtfeldern veranschaulicht.

Es wurde eine neuronale Netzarchitektur entwickelt, die speziell für Lichtfelder konzipiert ist. Um die reichhaltige, aber redundante Lichtfeldstruktur nutzbar zu machen, wurde ein Encoder-Decoder-Netz entworfen, in dem ein kleiner Satz von Funktionen aus den Winkel- und Raumdimensionen extrahiert wurde und anschließend Verwenden Sie diese Funktionen, um verschiedene Aufgaben gleichzeitig zu auszuführen. Unter der Annahme, dass diese Aufgaben eng miteinander verbunden sind, werden die extrahierten Funktionen in separate Decodierungspfade unterteilt. Da es komplex ist, eine intrinsische Komponente mit Ground-Truth-Informationen oder Disparitätswerte für die realen Szenen zu erhalten, wurde das Netz in dieser Arbeit so konzipiert, dass es sowohl angeleitet als auch unbeaufsichtigt trainiert werden kann. Wenn Ground-Truth-Informationen verfügbar sind (für

synthetische Szenen), bestraft die Standardverlustfunktion die Abweichung zwischen der Ausgabe des Netzes und dem Ground-Truth-Beispiel. Für die natürlichen Szenen ohne Ground-Truth-Informationen stellt das Netz sicher, dass die Ausgabekomponenten dem zugrunde liegenden physikalischen Modell der Bilderzeugung entsprechen.

Literaturverzeichnis

- [Al20] Alperovich, Anna: Variational and Deep Learning Approaches for Intrinsic Light Field Decomposition. 2020.
- [Gr15] Graber, G.; Balzer, J.; Soatto, S.; Pock, T.: Efficient Minimal-Surface Regularization of Perspective Depth Maps in Variational Stereo. In: Proc. CVPR. 2015.
- [HS06] Hinton, G.; Salakhutdinov, R.: Reducing the dimensionality of data with neural networks. Science, 313(5786):504–507, 2006.
- [JSG16] Johannsen, O.; Sulc, A.; Goldluecke, B.: What sparse light field coding reveals about scene structure. In: Proc. CVPR. 2016.
- [Li15] Lin, H.; Chen, C.; Kang, S.-B.; Yu, J.: Depth Recovery from Light Field Using Focal Stack Symmetry. In: Proc. ICCV. 2015.
- [Mo16] Moellenhoff, T.; Laude, E.; Moeller, M.; Lellmann, J.; Cremers, D.: Sublabel-Accurate Relaxation of Nonconvex Energies. In: Proc. CVPR. 2016.
- [Sh17] Shi, J.; Dong, Y.; Su, H.; Yu, S.: Learning Non-Lambertian Object Intrinsic across ShapeNet Categories. In: Proc. CVPR. 2017.
- [Su16] Sulc, A.; Alperovich, A.; Marniok, N.; Goldluecke, B.: Reflection Separation in Light Fields based on Sparse Coding and Specular Flow. In: VMV. 2016.
- [ZZP14] Zeisl, B.; Zach, C.; Pollefeys, M.: Variational Regularization and Fusion of Surface Normal Maps. In: Proc. 3DV. 2014.



Anna Alperovich wurde am 17. Juni 1989 in Nowosibirsk, Russland, geboren. Sie bekam einen Bachelor- und einen Masterabschluss von der Staatlichen Universität Nowosibirsk, wo sie Angewandte Mathematik studierte. 2012 verteidigte sie ihre Masterarbeit zum Thema der inversen Probleme. Anschließend wechselte sie an die Technische Universität Eindhoven in den Niederlanden, wo sie das Postgraduiertenprogramm Mathematics for Industry (= Mathematik in der Industrie) absolvierte. 2015 promovierte sie dort in Ingenieurwissenschaften (PDEng). Anschließend zog sie nach Deutschland, Konstanz, wo sie in Computer Vision promovierte. Sie genoss ihren Aufenthalt in Konstanz; in der Nähe der wunderschönen Berge und einer grandiosen Aussicht auf den Bodensee. Gegenwärtig arbeitet sie als Forschungsingenieurin für maschinelles Lernen bei der Carl Zeiss AG, wo sie erfolgreich eine Deep-Learning-Lösung für verschiedene industrielle Problemstellungen entwickelt.