Baden-Württemberg Stiftung

WIR STIFTEN ZUKUNFT

Forschungsergebnisse

Optische 3D-Sensorsysteme für mobile Anwendungen

der Baden-Württemberg Stiftung gGmbH

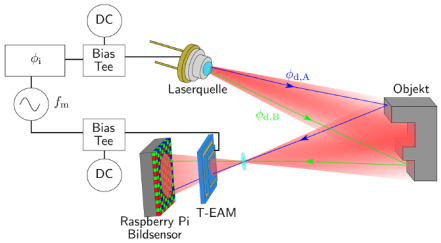
FORSCHUNGSTAG 2022

3D-Kamera mit optischem Modulator und maschinellem Lernen

Markus Miller[1], Michael Schelling[2], Timo Ropinski[2] und Rainer Michalzik[1]
Universität Ulm, [1]Institut für Funktionelle Nanosysteme / [2]Institut für Medieninformatik

Wir präsentieren Ergebnisse des aktuell laufenden, von der Baden-Württemberg Stiftung finanzierten Forschungsprojekts *DREAM – 3D-Kamera mit ElektroAbsorptionsModulator*. Hierin demonstrieren wir eine 3D-Kamera basierend auf der Verwendung eines konventionellen CMOS-Bildsensors in Verbindung mit einem anwendungsspezifischen optischen Modulator. Damit kann von der hohen Pixeldichte gewöhnlicher optimierter Bildsensoren profitiert werden. Die hohe Integrationsfähigkeit erlaubt die Anwendung des Systems selbst in mobilen Geräten. Ein Prototyp der Kamera wird charakterisiert, das Lichtlaufzeit-Prinzip wird simuliert, und Fehlerkorrektur erfolgt mit neuen Methoden maschinellen Lernens.

Kamera-Aufbau und Funktionsprinzip



Komponenten

- Beleuchtungsquelle (VCSEL)
- Gewöhnlicher Bildsensor
- Resonanter Transmissions-Elektroabsorptionsmodulator (T-EAM) mit p-i-n Diodenstruktur auf GaAs-Basis mit InGaAs-Quantenfilmen

- Synchrone Modulation von Quelle und T-EAM mit einer Frequenz f_m
- Indirekte Bestimmung der Distanz d durch Messung der Phasenverschiebung

$$\phi_d = \frac{2\pi \cdot 2d \cdot f_m}{c}$$

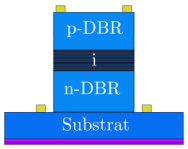
- Signalmischung von Modulator- und Beleuchtungssignal mit anschließender Integration durch den Bildsensor → Pixelenergie bei sinusförmiger Modulation:

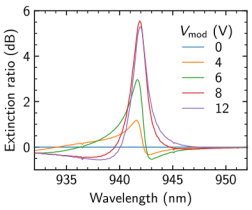
$$E_i \approx C_1 + C_2 \cos(\phi_d + \phi_i)$$

- Berechnung der Phasenverschiebung aus vier Einzelbildern:

$$\phi_d = \operatorname{atan} \frac{E_1 - E_3}{E_0 - E_2}$$

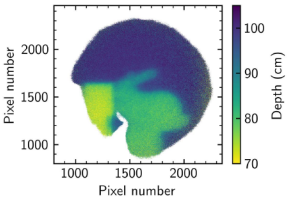
Herstellung und Charakterisierung der Modulatoren

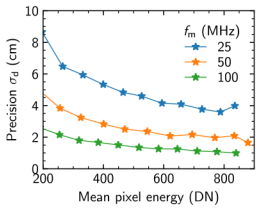




- Extinktionsverhältnis: $ER \approx 6$ dB
- Resonanzwellenlänge: $\lambda_{res} \approx 940$ nm
- Modulationsbandbreite: $f_c \approx 75$ MHz
- Modulatorgröße: $A_{mod} = 0.16\ mm^2$

3D-Kamera-Prototyp



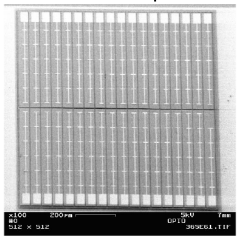


Aufzeichnung von 3D-Tiefenkarten mit einer Standardabweichung der Entfernung bzw. **Präzision von 1 cm bei einer Frequenz von 100 MHz und 1 m Objektdistanz.**

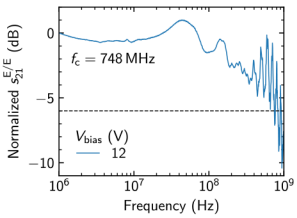
Verbesserung der Modulatoren

Ziel: Erhöhung der Modulationsbandbreite zur Verbesserung der Tiefenauflösung durch Segmentierung der Modulatoren mittels Ätzung → $f_{3dB} > 500$ MHz

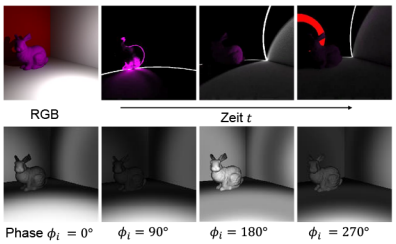
36 Modulatoren pro 1 mm²



Frequenzverhalten



Datensynthese

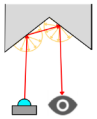


Bestimmung der zeitlich aufgelösten Impulsantwort der Szene durch *Path-Tracing*.

Berechnung der Lichtintegration für verschiedene Modulationen des *VCSELs* und des *T-EAMs* sowie Phasenverschiebungen ϕ_i.

Beispielszenen

Fehlerkorrektur mittels Deep Learning in 3D

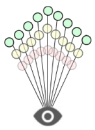


Fehlerquellen

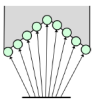
- *Multi-Path-Interference*, durch Mehrfachreflexion (siehe links)
- *Shot Noise*, hardwarebedingtes gaußsches Rauschen
- *Mixed Pixel Effect*, Signalmischung an Objektkanten

Fehlerkorrektur

- Extraktion zusätzlicher Informationen durch ein 2D-Netzwerk
- Projektion der Tiefenkarte eine 3D-Punktwolke



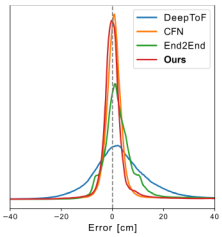
- Iterative Korrektur durch ein 3D-Punktwolken-Netzwerk
- Berücksichtigung der Kamerastrahlen
- Assoziation zwischen Punkt und Kamerapixel bleibt erhalten



- Abgleich mit tatsächlicher Geometrie
- Rückprojektion in eine Tiefenkarte

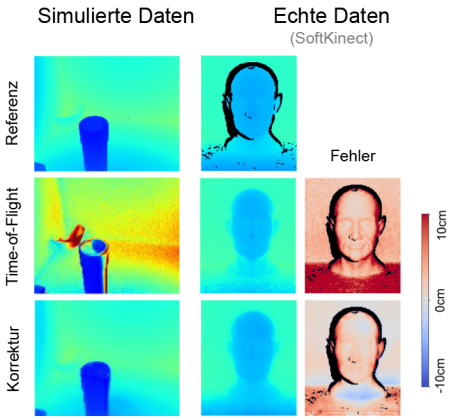
Fehlerreduktion um 67 % auf simulierten Daten des Prototyps und echten Daten einer *SoftKinect* Time-of-Flight-Kamera

Resultate



Method	Median	σ
DeepToF	2.27	15.61
CFN	0.42	8.80
End2End	1.72	14.97
Ours	**-0.28**	**7.64**

Fehlerverteilung



Baden-Württemberg Stiftung
WIR STIFTEN ZUKUNFT

Baden-Württemberg Stiftung

3D-Kamera mit ElektroAbsorptionsModulator – DREAM

Die Bedeutung von optischer dreidimensionaler (3D) Bilderfassung ist sowohl im Industrie- als auch im Konsumenten-Bereich in den letzten Jahren enorm gestiegen. Beispiele hierfür sind Umgebungsüberwachung, Objekt-Lageerkennung und -Tracking sowie Gesichts- oder Gestenerkennung. Unter den verschiedenen 3D-Messverfahren spielt die Lichtlaufzeitmessung (Time of Flight, ToF) eine große Rolle. Im Projekt 3D-Kamera mit ElektroAbsorptionsModulator – DREAM sollte ein vergleichsweise einfaches, prinzipiell echtzeitfähiges ToF-Verfahren zur 3D-Bilderfassung experimentell demonstriert werden. Es beinhaltet aktive Objektbeleuchtung durch Laserlicht, eine einzelne konventionelle Monochrom-(CMOS-)Kamera sowie einen synchron zum Laser hochfrequent modulierten globalen Verschluss (shutter), der direkt oder durch eine Abbildungsoptik getrennt vor den Bildsensor platziert wird. Hierfür geeignete optoelektronische Elektroabsorptionsmodulatoren (EAMs) wurden für den gegebenen Zweck entwickelt, optimiert und als Halbleiterkomponente unter Reinraumbedingungen selbst hergestellt. Der Arbeitsbereich der Kamera reicht von etwa zehn Zentimetern bis zu mehreren Metern. Der Hardwareaufbau wurde im Projekt durch moderne Softwareansätze zur Datenauswertung komplettiert.

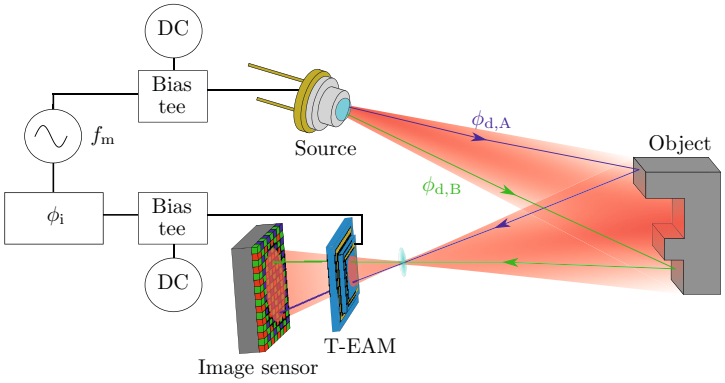


Bild 1: Schematischer Aufbau der DREAM-Kamera (aus [1]). DC: Vorstrom des Lasers bzw. Vorspannung des EAMs. Bias tee: Zusammenführung von statischen und modulierten Signalen.

Bild 1 illustriert das Prinzip der Kamera. Das Messobjekt wird mit modulierter Strahlung eines Vertikallaserdioden-(VCSEL-) Arrays mit etwa 940 nm Emissionswellenlänge beleuchtet. Je nach Tiefenunterschied erreicht das modulierte und vom Objekt rückgestreute Licht den vor den Bildsensor platzierten Transmissions-EAM (T-EAM) mit einer bestimmten Phasenverschiebung Φ_d. Im EAM werden alle auftreffenden Strahlen nochmals gleichartig mit derselben Frequenz f_m in ihrer Intensität moduliert. Für eine Tiefenmessung werden typischerweise vom Sensor vier Bilder mit unterschiedlichen relativen Modulationsphasen Φ_i aufgenommen. Aus den vier Graustufen jedes Pixels des Bildsensors lässt sich unmittelbar Φ_d ermitteln, und der Objektabstand ist schließlich direkt proportional zu Φ_d und umgekehrt proportional zu f_m [1]. Im Messprinzip wird die Modulationsphase ausgewertet. Es handelt sich also um ein indirektes ToF-Verfahren (iToF) im Gegensatz zu direktem ToF, in dem periodisch kurze Pulse ausgesendet werden. Der T-EAM beruht auf dem quantum-confined Stark effect (QCSE), so dass sich eine Absorptionsänderung in den Quantenfilmen des Halbleitermaterials bei variierter elektrischer Spannung ergibt. Mehrere Generationen von EAMs wurden im Projekt hergestellt, mit dem Ziel einer Optimierung des Modulationskontrasts sowie der elektrischen Bandbreite im dynamischen Betrieb. Darauf basierend erfolgte der Aufbau und die umfassende Charakterisierung von Kamera-Prototypen. Die ermittelte maximale Präzision (Standardabweichung der Entfernung) beträgt 5 mm bei 70 cm Objektabstand und 100 MHz Modulationsfrequenz.

Bild 2 stellt Tiefenkarten einer beispielhaften Szene 3D-gedruckter Objekte in einer Cornell-Box mit 50 cm Kantenlänge gegenüber, links gemessen (und vom System bearbeitet) mit einer kommerziellen iToF-Kamera und rechts mit der DREAM-Kamera, wobei hier für besseren Vergleich 4 × 4 Pixel durch

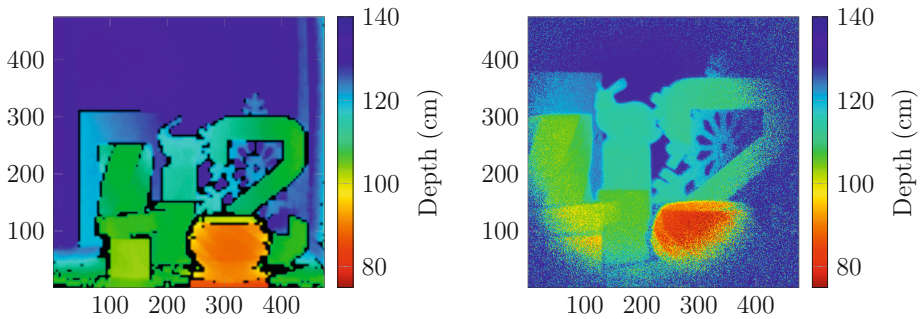


Bild 2: Tiefenkarten von Aufnahmen mit einer kommerziellen iToF-Kamera (links) und der DREAM-Kamera bei f_m = 100 MHz (rechts). Die horizontalen und vertikalen Achsen kennzeichnen die Bildpixel.

Mittelung zusammengefasst wurden, um eine ähnliche Auflösung zu erzielen. Beide Karten zeigen die einzelnen Objektentfernungen deutlich, jedoch ist Rauschen in der linken Aufnahme unterdrückt. Zudem sind Pixel mit geringer Intensität und mehrdeutigen Werten (z.B. an Objektkanten) schwarz und ohne Tiefeninformation. Die laterale Auflösung der Szene ist jedoch im T-EAM-Kamerabild deutlich höher, so dass selbst die Ohren des standardisierten, etwa 10 cm hohen Stanford Bunnys noch klar erkennbar sind. Insgesamt zeigt der Vergleich das große Potenzial des DREAM-Kamerasystems. Auch ohne aufwendige Modulationsverfahren oder Bildverarbeitung lässt sich eine mit einem kommerziellen System vergleichbare Tiefenauflösung bei merklich verbesserter lateraler Auflösung erzielen.

Der softwareseitige Teil des Projektes befasste sich mit der Verbesserung der Rekonstruktion der Tiefenkarten der iToF-Kamera, was durch maschinelles Lernen unter Verwendung von neuronalen Netzen umgesetzt wurde. Dies beinhaltete mehrere Arbeitspunkte, wobei die Hauptkomponenten die Simulation des Kamera-Prototyps und des Versuchsaufbaus, die Erzeugung eines synthetischen Datensatzes, die Konzeption der Netzwerkarchitektur, die Adaption von synthetischen zu realen Messungen, sowie die abschließende Auswertung auf synthetischen und realen Daten waren.

Zunächst wurde zur Vorbereitung das Funktionsprinzip der (amplitudenmodulierte Signale emittierenden) Kamera simuliert, indem mittels Transient Rendering eine zeitlich aufgelöste Impulsantwort der belichteten Szene bestimmt wurde, woraus anschließend mittels signaltheoretischer Verarbeitung eine Simulation des modulierten, kontinuierlichen Signals möglich ist. Durch Verfolgung mehrerer Trajektorien der Lichtausbreitung, sogenanntes Path Tracing, wurde eine realistische Simulation inklusive Multi-Path Interference (MPI) erzielt. In Kombination mit einem automatisierten Szenengenerator wurde hiermit ein annotierter Datensatz mit über 20.000 Datenpaaren erzeugt, welcher die Simulation diverser Fehlerquellen erlaubt und verschiedene Materialien, Objekte und Messdistanzen umfasst.

Zur Korrektur von Fehlern in der rekonstruierten Tiefenkarte wurden zwei Netzwerkarchitekturen entwickelt, welche speziell auf Fehler durch MPI und Bewegungsartefakte zugeschnitten sind. Die erste Architektur projiziert die zweidimensionale Tiefenkarte intern in eine dreidimensionale Punktwolke, welche mit Hilfe neu entwickelter Ray-Aligned Depth Update (RADU) Faltungen iterativ korrigiert wird, wobei Informationen über die Kamerablickrichtung integriert werden, wodurch eine verbesserte Kompensation von MPI-Effekten erzielt werden konnte. Durch eine Verbindung von zweidimensionalen Netzwerkschichten mit den beschriebenen dreidimensionalen RADU-Schichten und eine kombinierte Überwachung der 2D- und 3D-Tiefenrekonstruktionen des Netzwerkes, wie in Bild 3 dargestellt, wurde eine Verbesserung der Rekonstruktion im Vergleich zu existierenden Methoden auf mehreren Datensätzen erzielt.

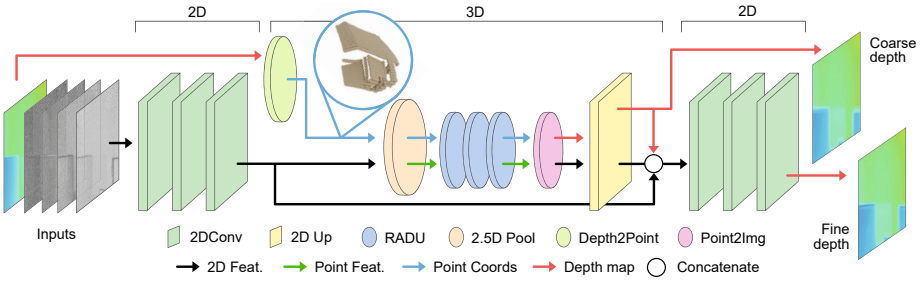


Bild 3: Die 2D-Tiefenkarte wird intern als eine 3D-Punktwolke repräsentiert und von 3D-RADU-Punktfaltungen verarbeitet (blau). Durch Einbettung in 2D-Schichten (grün) und eine zweistufige Überwachung (rechts) können hohe Fehlerkorrekturraten durch die 3D-Schichten erzielt werden.

Zur Kompensation von Bewegungsartefakten wurden Netzwerkarchitekturen verwendet, welche den optischen Fluss zwischen den Einzelmessungen bestimmen. Um ein Training dieser Netzwerke auf Tiefenkarten des iToF-Sensors zu ermöglichen, wurde ein effizient imple-

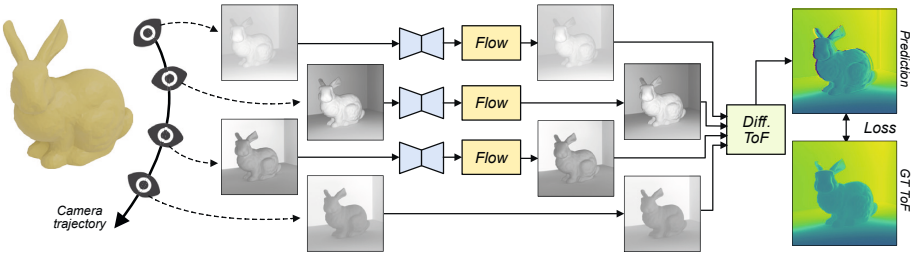


Bild 4: Um Bewegungen zwischen den Messungen (links) zu kompensieren, werden neuronale Netzwerke (blau) darauf trainiert, den optischen Fluss zu berechnen. Durch eine differenzierbare Tiefenberechnung (grün) wird eine Überwachung durch die rekonstruierten Tiefen ermöglicht.

mentierbares Gradientenkorrekturverfahren der Tiefenberechnung entwickelt, wodurch ein schwach überwachtes Training zur Bewegungskompensation realisiert werden konnte, wie es in Bild 4 veranschaulicht ist. Unter Verwendung mehrerer Regularisierungen des Netzwerkes konnten hiermit existierende Verfahren zur Kompensation von Bewegungsartefakten für verschiedene Funktionsmodi des iToF-Sensors übertroffen werden [2].

Schließlich wurde die Leistung der entwickelten Methode zur MPI-Korrektur auf realen Daten des Kamera-Prototyps ausgewertet. Hierfür wurde eine Phantom-Evaluation mittels 3D-gedruckter Modelle durchgeführt, wodurch synthetische und reale Daten abgeglichen werden konnten und die Simulation durch die Charakterisierung der Kamera weiter anpassbar war. Zur Adaption des Netzwerks an reale Messdaten, sogenannte Domain Adaptation, wurde eine unüberwachte Trainingsmethode entwickelt, welche von Student-Teacher-Methoden abgeleitet wurde und mit Hilfe von zyklisch generierten Pseudo-Zielwerten eine iterative Anpassung der internen Gewichte des Netzwerkes an die statistischen Verteilungen realer Daten erlaubt. Durch Verwendung dieser Methode konnten hohe Fehlerkorrekturraten auf realen Messdaten des Kamera-Prototyps erzielt werden. Des Weiteren konnten durch diesen Ansatz auf einem existierenden Benchmark-Datensatz niedrigere Restfehler erreicht werden als sie von bisher existierenden Methoden erzielt wurden.

Literatur:

[1] M. Miller, B. Sassiya, S. Menzel, K.J. Ebeling, and R. Michalzik, „3D indirect time-of-flight imaging employing transmission electroabsorption modulators“, Proc. SPIE, vol. 12136, 2022, DOI: 10.1117/12.2620458.

[2] M. Schelling, P. Hermosilla, and T. Ropinski, „Weakly-supervised optical flow estimation for time-of-flight“, 2023 IEEE/CVF Winter Conf. on Applications of Computer Vision (WACV), DOI: 10.1109/WACV56688.2023.00217.

Ansprechpartner:

Prof. Dr. Rainer Michalzik
Universität Ulm, Institut für Funktionelle Nanosysteme
rainer.michalzik@uni-ulm.de

Prof. Dr. Timo Ropinski
Universität Ulm,
Institut für Medieninformatik
timo.ropinski@uni-ulm.de