Kann KI den Fotografen ersetzen

Die nachträgliche Gestaltung realer Aufnahmen

Mario Bock
Positionspapier und lokaler Proof of Concept
Version 1.0 · 15. Oktober 2026

Zusammenfassung

Mit KI und algorithmischen Kameramodellen lassen sich Perspektive und optische Bildwirkung nach der Aufnahme verändern. Daraus ergibt sich mein Vorschlag für eine Kamera-App: Ein multimodales Sprachmodell unterstützt unerfahrene Nutzer dabei, eine Person zu positionieren, eine Pose zu wählen und die Umgebung einzubeziehen. Das anschließend aufgenommene Foto wird dreidimensional analysiert. Erst danach werden der gewünschte Kamerastandpunkt und die Charakteristiken von Sensor und Objektiv bestimmt. Apples Spatial Reframing zeigt bereits einen Teil dieses Prinzips, nämlich die nachträgliche Änderung der scheinbaren Aufnahmeperspektive [2]. Mein lokal aufgesetzter Prototyp untersucht zusätzlich, wie sich Quellbilddaten räumlich reprojizieren und mit einer veränderten virtuellen Kamera aufnehmen lassen.

Vorhandene Bildinformationen werden anhand ihrer Tiefe mathematisch neu angeordnet; generative KI ergänzt anschließend freigelegte Flächen, Lücken zwischen projizierten Bildproben und ausgewählte Konturen. Eine Animation vom Original zur neuen Ansicht kann sichtbar machen, welche Details aus der Aufnahme stammen und wo Ergänzungen stattfinden. Das Ziel ist ein hochwertiges Bild desselben realen Moments, nicht eine frei erfundene Szene. Im vorgeschlagenen Verfahren bleiben dieselben Quellpixel und Farben erhalten: Sie werden an andere Bildpositionen projiziert, während entstandene Lücken ergänzt werden. Ein Strickmuster auf einem Kleid bleibt damit dieselbe aufgenommene Struktur und wird nicht neu entworfen.

Meine These lautet, dass die Verbindung aus Aufnahmehilfe, nachträglicher Kameragestaltung, übertragbaren Farbstilen und digitaler Beleuchtung große Teile alltäglicher Foto- und Eventdienstleistungen wirtschaftlich ersetzen kann.

1 Personenführung und automatische Aufnahmeauswahl

Ein Teil fotografischer Erfahrung besteht darin, eine Person im Raum zu platzieren und ihr verständliche Hinweise zu geben. Eine Kamera-App kann diese Aufgabe bereits vor der Aufnahme unterstützen. Ein multimodales Large Language Model, kurz LLM, erhält dafür das Kamerabild und einen Bildwunsch. Es kann vorschlagen, den Oberkörper leicht zu drehen, einen Schritt von einer Wand wegzugehen, sich an ein Geländer anzulehnen oder mit einem Gegenstand zu interagieren. Person und Handlung werden dabei tatsächlich vor der Kamera arrangiert.

Samsung führte mit dem Galaxy S10 im Jahr 2019 Shot Suggestion ein. Die Funktion gibt Hinweise zu Komposition, Positionierung und Proportionen [1]. Eine situationsbezogene Anleitung zu Posen und zur Interaktion mit der Umgebung führt diesen Ansatz weiter. Ein LLM kann solche Hinweise im Dialog formulieren und auf Veränderungen im Kamerabild reagieren. Eigene Kameramodi für Porträts, Gruppenbilder und Veranstaltungen könnten den gewünschten Stil vorgeben.

Praktisch zu untersuchen ist dabei insbesondere, wie Nutzer diese Unterstützung annehmen und wie sie sich in eine natürliche Aufnahmesituation einfügt.

Für die Auswahl echter Momente existiert mit Samsungs Single Take bereits ein weiterer Baustein. Die mit der Galaxy-S20-Reihe eingeführte Funktion erstellt aus einer kurzen Aufnahme mehrere Fotos und Videoclips. Für das Galaxy S21 beschreibt Samsung fünf bis fünfzehn Sekunden Aufnahmezeit und eine automatische Auswahl anhand von Bildqualität und Gesichtsausdruck. Geschlossene Augen werden berücksichtigt; weitere Analysen beziehen Körperhaltung und Bewegungen ein [15]. Der Nutzer muss damit nicht mehr jeden geeigneten Einzelmoment selbst treffen.

In meinen lokalen Versuchen mit leistungsfähigen multimodalen LLMs ließen sich mehrere Aufnahmen anhand von offenen Augen, Lächeln, freundlicher Ausstrahlung, Pose und Belichtung vergleichen und eine bevorzugte Aufnahme auswählen. Das ist eine qualitative Beobachtung aus meinen Versuchen. Sie legt nahe, Aufnahmehilfe und Bildauswahl in einer App zu verbinden. Mit leistungsfähigeren, effizienteren Modellen und mehr lokal ausführbarer Bildanalyse kann diese Unterstützung schrittweise ausgebaut werden. Die App würde dann sowohl bei der Gestaltung einer Szene helfen als auch geeignete tatsächlich aufgezeichnete Momente vorschlagen.

2 Die Aufnahme als räumlich bearbeitbarer Bildzustand

Die Perspektive beim Auslösen muss nicht die endgültige Bildperspektive bleiben. Apple nennt die entsprechende Funktion in der Fotos-App Spatial Reframing. Nach Apples Beschreibung lässt sich eine Aufnahme durch Berühren und Ziehen räumlich neu ausrichten, während die Vorschau eine veränderte Kameraposition zeigt. Neu benötigte Inhalte ergänzt ein Bildmodell [2]. Damit ist ein Teil des hier diskutierten Prinzips bereits als Produktfunktion beschrieben.

Der hier vorgeschlagene Ansatz führt diesen Gedanken weiter. Eine gewöhnliche RAW-Datei bewahrt Sensordaten für spätere Farb- und Tonwertentscheidungen. Der hier vorgeschlagene Bildzustand ergänzt Quellbild und Tiefeninformation um eine räumliche Darstellung, aus der sich eine neue virtuelle Kamera aufnehmen lässt. Sensorgröße, Brennweite, Motivabstand und Objektivcharakter werden so zu nachträglichen Gestaltungsparametern.

2.1 Quellpixel als Splats im Raum

Die sichtbaren Bildproben werden anhand ihrer Tiefe in den 3D-Raum zurückprojiziert. Jede Probe erhält eine räumliche Position und bleibt mit ihrer Quellkoordinate und Farbe verbunden. Gerendert werden diese Proben als Splats, also kleine Flächenelemente mit einer projizierten Ausdehnung. Gaussian Splatting verwendet dafür räumliche Gauß-Funktionen mit Größe, Orientierung und weiteren Attributen [4]. Anschaulich wird das Foto zu einer farbigen, räumlich aufgefalteten Oberfläche. Das Ausgangsbild kann von jeder Kamera stammen; das Handy ist hier ein zugängliches Beispiel.

Mein lokaler Prototyp schätzt Tiefe und Ausgangsbrennweite mit Apple Depth Pro [3]. Er verarbeitet das vollständige Quellraster mit einer globalen Schätzung und überlappenden Detailkacheln. Jeder Quellpixel wird zunächst als Kandidat für einen Splat berücksichtigt. Die aktuelle Darstellung verwendet vereinfachte runde Punktsplats. Sie macht das Prinzip der quellgebundenen Projektion praktisch nutzbar; eine detailliertere Gaussian-Repräsentation kann darauf aufbauen.

Die virtuelle Kamera betrachtet diese Szene aus einer neu gewählten Position und Blickrichtung. Eine andere Brennweite oder Sensorgröße verändert den Bildwinkel; ein anderer Abstand verändert zusätzlich Perspektive und Parallaxe. Die ursprünglichen Bildproben bleiben dabei dieselben. Ihre Positionen im Szenenkoordinatensystem bleiben fest, während sich ihre Anordnung und Größe im Zielbild ändern. Bei extremen Blickpunktänderungen werden zusätzliche Flächen sichtbar. Sie können bevorzugt aus weiteren vorhandenen Aufnahmen übernommen oder andernfalls generativ ergänzt werden.

Algorithmische Kameramodelle ergänzen Fokus, Schärfentiefe, Blendenzahl und Blendenform, Verzeichnung sowie anamorphotische Stauchung und Unschärfecharakteristik. So entsteht ein anderer Bildlook aus derselben Aufnahme. Die Quellfarben bleiben als Ausgangsdaten erhalten; die Ausgabe kann sie durch Unschärfe, Rasterung oder eine bewusst gewählte Farb- und Lichtgestaltung anders abbilden. Die Kameramodelle berechnen dafür Projektion und optische Wirkung der Zielkamera aus den gespeicherten Bild- und Tiefendaten.

2.2 Überdeckungen und Lücken geometrisch bestimmen

Eine neue Kamera verändert die Abstände zwischen projizierten Splats. Manche liegen weiter auseinander, andere überdecken dieselben Zielpixel. Der lokale Prototyp entscheidet ihre Sichtbarkeit mit einem strikten Tiefenpuffer. Pro Zielpixel trägt nur das nächstgelegene akzeptierte Fragment bei. Dahinterliegende Fragmente werden verdeckt, statt transparent durchzuscheinen. Ein tatsächlich neben dem Vordergrund sichtbarer Hintergrund bleibt erhalten.

Zusammenhängende Proben werden deckend gerendert. Ihre Größe skaliert mit dem Verhältnis von Quelltiefe zu Zielkameratiefe, sodass ihre projizierte Fläche zum neuen Abstand passt. Eine Bereinigung vergleicht jede Probe mit ihren acht Nachbarn. In der Voreinstellung wird eine Probe bei drei verfügbaren Nachbarn innerhalb einer relativen Tiefendifferenz von 18 Prozent beibehalten. Dadurch werden räumlich zusammenhängende Bereiche gestützt und einzelne unpassende Proben ausgesondert. Die Regel verwendet am Bildrand die tatsächlich verfügbaren Nachbarn.

An starken Tiefensprüngen, etwa zwischen Person und Hintergrund, können ausgewählte Randbereiche entfernt werden. Das verhindert, dass ein fehlerhafter Übergang als gestreckte Fläche in der neuen Ansicht erscheint. Differenziertere Kanten- und Oberflächenmodelle können diese Entscheidung schrittweise präzisieren. Die geometrische Stufe ordnet dabei vorhandene Bildinformation neu an; sie generiert keine zusätzlichen Szeneninhalte.

Nach der Kameraauswahl werden die verbleibenden Freiflächen ergänzt. Dazu gehören freigelegter Hintergrund, Zwischenräume zwischen Splats und feine Konturen wie Haarkanten. Das Originalfoto, der Zielrender und die Tiefenansicht geben dem Bildmodell Struktur und Erscheinung vor. Für den beschriebenen Ablauf ist vorgesehen, die belegten Bildanteile gezielt zu übernehmen und den generativen Schritt auf Freiflächen und ausgewählte Übergänge zu begrenzen. So bleiben beispielsweise Textilmuster oder Schuhstrukturen aus der tatsächlichen Aufnahme erhalten.

Mein Prototyp übergibt derzeit FLUX.2 Klein 4B das Original, einen Zielrender mit transparenten Lücken und halbdeckenden bekannten Bereichen sowie eine Tiefenreferenz. Diese Referenzen steuern eine Ausgabe des gesamten Bildes; die Generierung erfolgt mit ungefähr vier Megapixeln und wird bei größeren Exporten hochskaliert.

Die gezielte Übernahme geschützter Bildanteile und eine maskierte Ergänzung sind die entsprechende Weiterentwicklung zum beschriebenen Verfahren.

2.3 Kameraentscheidungen durch ein LLM unterstützen

Im lokalen Prototyp wähle ich Kameraposition und Objektivparameter manuell. Eine spätere App kann diese Bedienung durch ein multimodales LLM ergänzen. Ein Wunsch wie ein engeres Porträt mit ruhigem Hintergrund lässt sich dann in Vorschläge für Bildausschnitt, Abstand, Brennweite, Sensorformat und Blende übersetzen. Die eigentliche Projektion wird weiterhin vom Kameramodell berechnet. Das LLM steuert die Parameter und bewertet die Vorschauen.

Der Nutzer könnte mehrere Varianten vergleichen oder die Auswahl dem gewünschten Kameramodus überlassen. Die Bildgestaltung wäre damit weniger an die Erfahrung beim Auslösen gebunden. Apples Spatial Reframing und mein lokaler Versuch zeigen verwandte Teilfunktionen; die hier beschriebene Verbindung mit einer sprachlich gesteuerten Kameraauswahl ist die vorgeschlagene Erweiterung. Für das Ergebnis, in dem sich Kamera und Objektiv nachträglich verändern lassen, verwende ich scherzhaft die eigene Bezeichnung "Super RAW".

2.4 Mehr reale Ansichten als Grundlage der Ergänzung

Die räumliche Darstellung lässt sich durch zusätzliche reale Beobachtungen erweitern. Statt nur ein Foto aufzunehmen, kann eine kurze Videosequenz beim leichten Bewegen der Kamera weitere Seitenflächen und Hintergrundbereiche erfassen. Passende Frames werden räumlich zugeordnet und in einer gemeinsamen Szene verbunden. Wo eine neue Zielansicht bislang verdeckte Flächen zeigt, können diese aus tatsächlich aufgenommenen Nachbaransichten stammen, statt neu generiert zu werden.

Die Zeit liefert dabei zusätzliche Beobachtungen. Dreht sich eine Person, werden zuvor verdeckte Seitenflächen sichtbar. Bewegt sie sich weiter, kann der dahinterliegende Hintergrund in späteren Frames aufgenommen werden. Eine bewegte Kamera erschließt weitere Blickrichtungen. Zeitliche Korrespondenzen und Bewegungsmodelle ordnen solche Beobachtungen den betreffenden Oberflächen zu. Mehrere synchronisierte Kameras ergänzen gleichzeitig weitere Ansichten desselben Moments. Auch ältere Aufnahmen derselben Umgebung können Hintergrundinformation beitragen, wenn Kamerageometrie, Beleuchtung und Szenenzustand passend zugeordnet werden.

Tiefensensoren liefern weitere Daten. Apple ARKit stellt auf unterstützten LiDAR-Geräten beispielsweise sceneDepth bereit [5]. Die Kombination aus Bildsequenzen, Tiefenmessungen und lernenden Modellen verbessert die räumliche Grundlage schrittweise. Wo geometrische Verfahren keine ausreichende Beobachtung finden, unterstützt generative KI, möglichst konditioniert auf existierende Bilder. Bei kleinen Kameraänderungen und guter Abdeckung kann der ergänzte Anteil so gering bleiben, dass er für die Bildwirkung kaum ins Gewicht fällt, während die kennzeichnenden Details des Originals erhalten bleiben.

3 Die Entstehung des Bildes visuell nachvollziehbar machen

Für die Akzeptanz eines Fotos, das anders aussieht als unmittelbar nach dem Auslösen, ist eine verständliche visuelle Erklärung seiner Entstehung wichtig. Eine Animation kann zeigen, wie sich die projizierten Quellpixel vom ursprünglichen Foto in die neue Ansicht bewegen. Der Betrachter erkennt dabei im Detail, welche Bildstrukturen erhalten geblieben sind. Das Strickmuster eines Kleides lässt sich beispielsweise während der Kamerabewegung verfolgen, statt nur zwei fertige Bilder miteinander zu vergleichen.

Aus Quellkoordinaten und Kameramatrizen lässt sich ein solcher Übergang direkt ableiten. Dieselbe Quellkennung bleibt während der Projektion erhalten. Freigelegte Bereiche erscheinen zunächst leer und werden anschließend als Ergänzung eingeblendet. Vorhandene Daten und neu erzeugte Bildanteile bleiben so unterscheidbar. Dieser Herkunftsviewer ist als Ergänzung des Prototyps vorgesehen.

Die Erklärung macht den Grundgedanken sichtbar: Eine echte Aufnahme wird räumlich neu betrachtet und gezielt vervollständigt. Gerade für Menschen, die keine frei erfundenen Erinnerungsbilder möchten, kann die nachvollziehbare Verbindung zum Ausgangsfoto die Akzeptanz fördern. Auch gewählte Farb- oder Lichtänderungen lassen sich als eigener Bearbeitungsschritt anzeigen.

4 Farbe und Beleuchtung als nachträgliche Gestaltung

Die Gestaltung von Farbe und Tonwerten ist bereits mit verfügbaren Werkzeugen weitgehend automatisierbar. Voreinstellungen, Farbprofile und Lookup-Tabellen bieten vorbereitete Looks. Photoshop überträgt mit Match Color Farbeigenschaften zwischen Bildern [6]; adaptive Profile in Lightroom passen Farbe, Tonwerte und Kontrast bildabhängig an [7]. Eine App kann daraus auswählbare Stile machen oder eine Referenzaufnahme für den gewünschten Look verwenden. Die wiederholte technische Anwendung eines vorhandenen Farbstils erfordert dann keine individuelle Bearbeitung jedes Fotos.

Auch digitale Lichtgestaltung hat etablierte Vorläufer. Apples Porträtlicht bietet auf unterstützten Geräten verschiedene virtuelle Beleuchtungseffekte, die sich nach der Aufnahme anpassen lassen [8]. Generative Verfahren erweitern diesen Ansatz. IC-Light verändert Bildbeleuchtung anhand von Text oder Hintergrundbedingungen [9]. Für bewegte Bilder untersucht RelightVid zeitlich konsistente Lichtänderungen [10]; Runways Relight Scene bietet promptgesteuerte Änderungen von Videolicht als Produktfunktion an [11].

Diese Modelle werden schrittweise leistungsfähiger. Geometrische Berechnungen, geschätzte Oberflächen und vorhandene Referenzbilder können die Lichtgestaltung führen; generative KI unterstützt dort, wo zusätzliche Bildinformation nötig ist. Eine entsprechende App kann so Perspektive, Kameralook, Farbgestaltung und Beleuchtung in einem Ablauf zusammenführen. Ein Nutzer wählt beispielsweise weicheres Gesichtslicht oder eine wärmere Stimmung, ohne selbst eine Studiobeleuchtung aufbauen zu müssen.

Farb- und Lichtgestaltung sind dabei bewusst gewählte Änderungen der Darstellung. Die Quellaufnahme bleibt gespeichert, während der gewünschte Stil auf die Ausgabe angewendet wird. Umfangreiches generatives Relighting bearbeitet auch bereits sichtbare Bereiche und ist deshalb ein eigener Modus neben der geometrischen Neuordnung mit Lückenfüllung. Die Herkunftsansicht kann diese Bearbeitung getrennt kenntlich machen. Das Ziel bleibt, Identität und kennzeichnende Strukturen der Aufnahme zu erhalten, auch wenn sich ihre fotografische Wirkung verändert.

5 Die wirtschaftliche Folge für alltägliche Fotografie

Die Kombination verlagert bisher bezahlte Entscheidungen in Software. Personenführung wird zur Aufnahmehilfe, geeignete Momente werden aus Sequenzen ausgewählt, Perspektive und Objektivwirkung nachträglich bestimmt. Farbstile und Lichtgestaltung werden auswählbar. Ein unerfahrener Nutzer kann damit aus einer brauchbaren Handyaufnahme ein Bild gestalten, dessen Erscheinung bislang mit professioneller Ausrüstung und anschließender Bearbeitung verbunden wurde.

Meine Zukunftsthese ist, dass dieser Ablauf den typischen, überwiegend standardisierten Eventfotografen in vielen alltäglichen Aufträgen wirtschaftlich verdrängen wird. Wenn Kunden den tatsächlichen Moment mit den gewünschten Details und einer hochwertigen Bildwirkung erhalten, verliert der Aufpreis für eine separat beauftragte Kameraaufnahme seine bisherige Begründung. Der Wunsch nach echten Erinnerungen ist mit diesem Ansatz vereinbar: Er beginnt mit einer realen Aufnahme und bewahrt deren Bezug zur beobachteten Szene.

Ich erwarte eine stärkere Trennung zwischen automatisierbaren Standardleistungen und eigenständiger fotografischer Kunst. Für eine persönliche Bildsprache, besondere Inszenierung und außergewöhnliche Ergebnisse können spezialisierte Fotografen weiterhin gezielt beauftragt werden und höhere Preise erzielen. Die alltägliche Anwendung eines bekannten Looks wird dagegen zunehmend zu einer Softwarefunktion. Das ist die hier vertretene wirtschaftliche Prognose.

6 Ausblick auf Videografie und räumliche Ereignisaufnahmen

Der nächste Schritt ist eine räumliche Ereignisaufnahme, bei der mehrere Kameras über längere Zeit dieselbe Szene beobachten. Meine Zukunftsannahme ist, dass bereits wenige Kameras, beispielsweise drei auf einer Hochzeit, genügend reale Anhaltspunkte für eine überzeugende 4D-Aufzeichnung liefern können, wenn zeitliche Beobachtungen, räumliche Rekonstruktion und generative Ergänzung gemeinsam genutzt werden. Kamerastandpunkte und Fahrten würden dann erst nach der Aufnahme aus dieser Szene gewählt. Die Zahl drei beschreibt hier das angestrebte einfache Aufnahmesystem.

Die Informationsgrundlage besteht aus mehr als den drei Blickrichtungen eines einzelnen Augenblicks. Gäste drehen sich, gehen durch den Raum und geben Hintergrundflächen frei. Kameras können sich ebenfalls bewegen und zusätzliche Seitenansichten aufnehmen. Ein Bereich, der zum gewünschten Zeitpunkt verdeckt ist, kann kurz davor oder danach tatsächlich sichtbar sein. Durch zeitliche Zuordnung und Bewegungsmodelle lassen sich solche Beobachtungen in die Rekonstruktion einbeziehen. Bei bewegten Oberflächen wird dabei ihre Veränderung über die Zeit modelliert, statt verschiedene Körperhaltungen wie eine starre Szene zusammenzusetzen.

4D Gaussian Splatting ergänzt die räumliche Darstellung um die Zeit. Wu und Kollegen zeigen dynamische Szenen mit zeitabhängiger Verformung von Gaussian-Primitiven und Echtzeitdarstellung [12]. Yang und Kollegen beschreiben eine Repräsentation in Raum und Zeit, aus der neue Blickpunkte und Bullet-Time-artige Darstellungen entstehen [13]. Anschaulich lässt sich ein rekonstruierter Moment wie bei Matrix anhalten und aus einer anderen Richtung betrachten. Ebenso kann die virtuelle Kamera stehen bleiben, während die erfasste Bewegung weiterläuft.

Die verbleibenden unbeobachteten Flächen werden durch generative KI ergänzt, geführt von den tatsächlich erfassten Ansichten und dem zeitlichen Verlauf. Dass sich eine geringe Kamerazahl mit generativer Unterstützung verbinden lässt, ist bereits eine konkrete Forschungsrichtung. SparseCam4D untersucht eine 4D-Rekonstruktion aus wenigen Kameraansichten und nutzt dafür zusätzliche generative Beobachtungen [16]. Das stützt die technische Richtung des Ausblicks: Die Aufnahme liefert die Identität und Struktur der Szene, weitere Frames erweitern die beobachteten Flächen, und generative Modelle vervollständigen die verbleibenden Bereiche.

Heute erfordern Erzeugung und Optimierung dynamischer Szenen erheblichen Rechenaufwand; die Abdeckung hängt vom Material ab. Die Forschung entwickelt zugleich effizientere Repräsentationen. Mobile-4DGS reduziert Speicher- und Darstellungsaufwand und zeigt interaktive WebGL-Beispiele mit Abspielen, Anhalten und zeitlichem Durchfahren [14]. Die Browserwiedergabe nutzt vorbereitete Rekonstruktionen. Kürzere Erstellungszeiten und einfachere Verarbeitung sind weitere Entwicklungsziele.

Daraus ergibt sich mein praktischer Ausblick: Wenige sinnvoll platzierte Kameras erfassen einen realen Anlass fortlaufend. Die künftige Software verbindet räumliche und zeitliche Beobachtungen zu einer frei gestaltbaren Aufnahme, deren generative Ergänzung eng an vorhandenes Material gebunden ist. Überlappende Originalaufnahmen aufzubewahren sichert dafür die benötigte Datengrundlage. Zusätzliche Kameras und eine längere Aufzeichnung schaffen so einen konkreten Weg zu nachträglicher Kameraführung.

Als Erweiterung des räumlich bearbeitbaren Bildzustands würde damit auch der Zeitpunkt einer Szene zusammen mit dem virtuellen Kamerastandpunkt wählbar. Meine weitergehende These ist, dass Kameraführung und Einstellungsgestaltung dadurch zunehmend in die spätere Bearbeitung wandern. Zusammen mit automatisierter Auswahl und Schnitt kann dies standardisierte Videodienstleistungen ebenso verändern und wirtschaftlich verdrängen wie die beschriebenen Fotodienstleistungen.

7 Fazit

Der vorgeschlagene Bildzustand hält die Entscheidung über die fotografische Wirkung über den Auslösezeitpunkt hinaus offen. Vorhandene Quellpixel werden räumlich zugeordnet und aus einer anderen virtuellen Kamera projiziert. Zusätzliche reale Aufnahmen füllen die Szene mit beobachteter Information; generative KI ergänzt verbleibende Flächen. Der lokale Prototyp verbindet dafür bereits Tiefenschätzung, Punktsplats, Sichtbarkeitsberechnung und virtuelle Kameramodelle.

Mit Aufnahmehilfe, automatischer Momentwahl und auswählbarer Farb- und Lichtgestaltung kann dieser Ansatz hochwertige Bilder für unerfahrene Nutzer zugänglich machen. Eine visuelle Erklärung zeigt dabei den Zusammenhang mit dem Original. Meine These ist, dass diese Verbindung einen großen Teil alltäglicher Foto- und Eventdienstleistungen ersetzen wird, während eigenständige fotografische Kunst als spezialisierte Leistung bestehen bleibt.

Quellen und Projektlinks

Quellenstand 15. Oktober 2026

[1] Samsung. How Samsung Developed the Galaxy S10's Share-Friendly, AI-Powered Camera. 3. April 2019.

[2] Apple. Apple Intelligence brings powerful AI capabilities into everyday experiences. Abschnitt zu Spatial Reframing. 8. Juni 2026.

[3] Bochkovskii et al. Depth Pro: Sharp Monocular Metric Depth in Less Than a Second. ICLR 2025. Offizielle Implementierung.

[4] Kerbl et al. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Transactions on Graphics, 2023.

[5] Apple. ARFrame sceneDepth. ARKit-Dokumentation zu Tiefendaten auf unterstützten LiDAR-Geräten.

[6] Adobe. Match color between two images. Photoshop-Dokumentation, 2. Dezember 2025.

[7] Adobe. Use Adaptive Profiles. Lightroom-Dokumentation, 28. April 2026.

[8] Apple. Take portraits with your iPhone camera. Benutzerhandbuch zu Porträtlicht und nachträglicher Anpassung.

[9] Zhang, Rao und Agrawala. Scaling In-the-Wild Training for Diffusion-based Illumination Harmonization and Editing by Imposing Consistent Light Transport. ICLR 2025. IC-Light-Projekt.

[10] Fang et al. RelightVid: Temporal-Consistent Diffusion Model for Video Relighting. arXiv:2501.16330, 2025.

[11] Runway. Adjust Video Lighting with AI. Produktbeschreibung von Relight Scene.

[12] Wu et al. 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering. CVPR 2024.

[13] Yang et al. Real-time Photorealistic Dynamic Scene Representation and Rendering with 4D Gaussian Splatting. ICLR 2024.

[14] Du et al. Mobile-4DGS: Unified Static-Dynamic Real-time Mobile Gaussian Splatting. WebGL-Demonstrationen, 2026.

[15] Samsung. Behind the Snapshot: How the Galaxy S21's AI Improves Your Photos in the Blink of an Eye, Single Take. 17. Februar 2021.

[16] Pan et al. SparseCam4D: Spatio-Temporally Consistent 4D Reconstruction from Sparse Cameras. CVPR 2026.