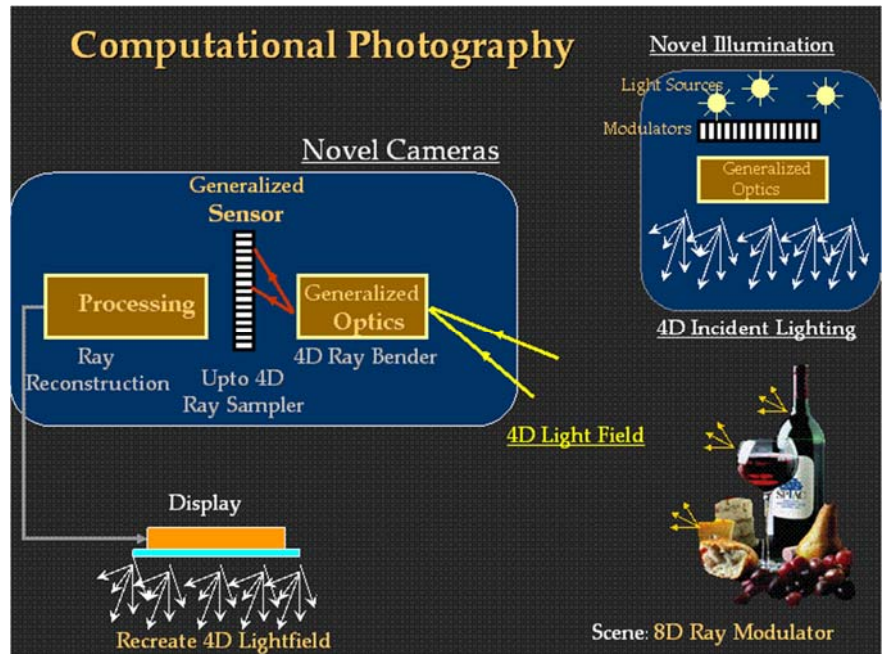


27. Computational Photography

Computational Photography

Computational Photography ist ein Forschungsgebiet, das durch die Verbreitung von Digitalkameras zunehmend die alltägliche Fotografie verändert. Durch neuartige Aufnahmetechniken und Bearbeitungs-algorithmen lassen sich Bilder erzeugen, die mit normaler Fotografie nicht möglich wären. Dazu gehören beispielsweise Aufnahmen mit extrem hohem Kontrastumfang, nachträgliches Verbessern verwackelter Bilder, Komposition mehrerer Personen oder Objekte zu einem neuen Gesamtbild, Entfernen störender Personen oder ähnliches. Ziel ist es, durch das Zusammenspiel von Bildverarbeitung, Computer Vision und Computergrafik neuartige Funktionalitäten und Anwendungsgebiete für die Photographie zu erschließen. Dabei sollen mittels algorithmischer („computational“) Ansätze die physikalischen Grenzen der traditionellen projektiven Kamera, wie begrenzte Tiefenschärfe, Auflösung und Verlust der dritten Dimension, überwunden werden. Damit können Bilder erzeugt werden, die mit einer herkömmlichen Kamera nicht erfassbar gewesen wären. Durch die Kombination von Software, digitalen Sensoren, moderner Optik sowie intelligenter Beleuchtung geht Computational Photography über die Grenzen der traditionellen Fotografie hinaus.



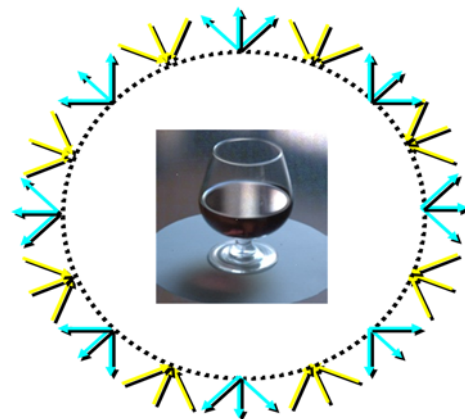
Moderne Kameras sind hoch komplexe Produkte, die eine über hundertjährige technische Evolution durchlaufen haben. Dennoch blieb ihr grundlegender optischer Aufbau aus Linse(n), Blende und Sensor praktisch unverändert. Aktuelle Systeme sind in der Lage, die Leistungen ihres natürlichen Vorbildes, des menschlichen Auges, zu übertreffen. Andererseits teilen „traditionelle“ Kameras aber auch die meisten Nachteile des menschlichen Auges. Die Bildaufnahme mit traditionellen Kameras führt neben dem Verlust einer Dimension der Szenengeometrie zu weiteren Verlusten an physikalisch verfügbarer Informationen: Der Sensor summiert einfallende Lichtstrahlen aus allen Einfallswinkeln zu einem Intensitätswert – dies führt zum Verlust von zwei Dimensionen der Information über die Reflektionseigenschaften von Szenenoberflächen (charakterisiert durch die *Bidirectional Radiance Distribution Function*, kurz BRDF). Generell lässt sich aus der Bestrahlungsstärke auf die Bildebene nur die Strahlungsdichte eines Objekts bestimmen, nicht aber die Reflektivität des Objekts, da sich die Strahlungsdichte des Objekts multiplikativ aus der Reflektivität und der unbekanntenen Bestrahlungsstärke des Objekts zusammensetzt. Das ist eines der Grundprobleme der Bildverarbeitung, was die Identifizierung und Klassifizierung von Objekten aus ihrer „Helligkeit“ im Bild schwierig macht.

Die Abbildung durch eine Linse in Kombination mit einer Blende führt zur Reduktion der Tiefenschärfe, was de facto zum Verlust brauchbarer Informationen für weite Tiefenbereiche der Szene führt und besonders problematisch ist, wenn wenig Licht zur Verfügung steht. Verfügbare Sensoren integrieren über ein breites Spektrum der Wellenlängen des einfallenden Lichtes – dies führt zum örtlichen Verlust der spektralen Dimension (multispektrale Sensoren messen unterschiedliche Wellenlängen an unterschiedlichen Orten). Auch über die Zeit wird das einfallende Licht integriert – somit ist auch die temporale Auflösung begrenzt. Zuletzt führt die Abtastung im Ortsbereich zu einer begrenzten räumlichen Auflösung und die Quantisierung zu einer begrenzten Helligkeitsauflösung und dadurch zu einer Reduktion des Signalumfangs (*dynamic range*).

Lichtfeld

Das Lichtfeld ist eine Funktion, welche die Lichtmenge beschreibt, die an jedem Punkt des dreidimensionalen Raums aus allen Richtungen fällt. Im Wesentlichen betrachten wir bei der Computational Photography das Lichtfeld, das auf

einen Punkt fällt, d.h. welche Strahlen an diesen Punkt reflektiert werden. Dabei werden photometrisch Objekte als Volumen betrachtet, welche ein an jedem Punkt einfallendes Beleuchtungsfeld in ein Beleuchtungsfeld umwandeln, welches vom Objekt wieder reflektiert wird. Einfallende Beleuchtung kann mittels eines 4D-Lichtfelds parametrisiert werden. Dazu stellen wir uns um das Objekt eine Kugel vor, die das Objekt einschließt. Wir verwenden (u_i, v_i) um die Position auf der Kugel zu beschreiben, auf der das Licht auftrifft, und die Winkel (θ_i, φ_i) beschreiben die Richtung des eintreffenden Lichtes. Daraus ergibt sich das einfallende Lichtfeld aus: $R_i(u_i, v_i, \theta_i, \varphi_i)$. Das ausgehende Licht kann auf gleiche Weise mittels eines 4D Lichtfelds parametrisiert werden, der Unterschied ist nur, dass wir jetzt betrachten, wie das Licht die objektumgebende Oberfläche verlässt: $R_r(u_r, v_r, \theta_r, \varphi_r)$. Daher können wir mittels einer achtdimensionalen Funktion - Reflexionsfeld genannt - die Reflexionseigenschaften eines Objektes charakterisieren. Das Reflexionsfeld codiert für jeden einfallenden Lichtstrahl die 4D Reflexion des Lichtstrahls. Das Reflexionsfeld beinhaltet daher auch die nötigen Informationen, um das Objekt unter anderen Beleuchtungsgegebenheiten und von verschiedenen Blickwinkeln aus zu rendern.



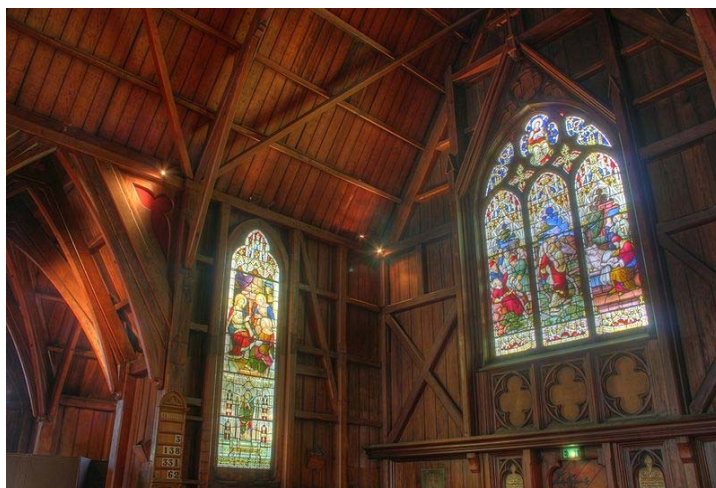
4D Lichtfeld Kameras, in der Literatur teilweise auch als „Plenoptische Kameras“ bekannt, sind daher nicht nur in der Lage, Ort und Intensität der Projektion der von einer Szene ausgehenden Lichtstrahlen zu detektieren, sondern diese auch nach deren Einfallswinkel auf der Kamera-Ebene zu quantifizieren. Die so gewonnene 4D Abtastung des Lichtfelds einer Szene ermöglicht einen vielfältigen Informationsgewinn. 4D Lichtfeldkameras verwenden ein Mikrolinsen-Array, das direkt auf dem Sensor aufgebracht wird. Dadurch lassen sich die eintreffenden Lichtstrahlen durch



die einzelnen Pixel im Sensorbereich unter jeder Mikrolinse nach ihrem Einfallswinkel quantifizieren und somit das 4D Lichtfeld einer Szene rekonstruieren. Diese Methode hat allerdings den Nachteil, dass die direkt erreichbare Ortsauflösung der Anzahl der Mikrolinsen entspricht, wobei deren Durchmesser (in Pixeln) wiederum die Winkelauflösung bestimmt: je größer die Winkelauflösung, desto geringer ist die Ortsauflösung und umgekehrt.

High Dynamic Range (HDR)

Unter *High Dynamic Range* (HDR) Bildgebung versteht man eine Fülle an Techniken, die es, im Vergleich zu herkömmlichen Techniken für digitale Bildgebung oder fotografischen Methoden, ermöglichen, einen größeren Bereich zwischen den hellsten und dunkelsten Regionen eines Bildes darzustellen. Dieser große dynamische Bereich ermöglicht HDR-Bildern, den Helligkeitsbereich in realen Szenen genauer zu repräsentieren. Der mögliche Helligkeitsbereich erstreckt sich von direktem Sonnenlicht bis hin zu schwachem Sternenlicht und wird durch die Aufnahme von mehreren Bildern erzielt, die die gleiche Szene mit unterschiedlicher Belichtung darstellen. Einfach ausgedrückt: HDR beschreibt eine Fülle an Techniken, die aufeinander abgestimmt wurden, um Bilder mit mehr Kontrast repräsentieren zu können. Nicht-HDR Kameras nehmen ein Bild mit begrenztem Kontrastbereich mit nur einem einzigen Belichtungslevel auf. Als Ergebnis erhält man ein Bild, das Detailverluste in hellen oder dunklen Bildbereichen aufweist, abhängig davon, ob die Kamera eine hohe oder niedrige Belichtungseinstellung aufweist. HDR kompensiert diesen Detailverlust durch die Aufnahme mehrerer Bilder mit verschiedenen Belichtungsstufen und ordnet diese intelligent an, sodass wir ein Bild erhalten, dessen dunkle als auch helle Bildbereiche im Detail darstellbar sind.



Um die Anzeige von HDR-Bildern auf Geräten mit einem kleinen dynamischen Bereich zu ermöglichen, werden Techniken zur Abbildung von Farbtönen (engl. *tone mapping*) benötigt, die den gesamten Kontrast reduzieren. Ausdrucke, CRT- oder LCD-Monitore und Projektoren haben alle einen begrenzten dynamischen Bereich, der für die Darstellung der vollen Bandbreite der Lichtintensitäten, ungeeignet ist. Tone Mapping behandelt die Kontrastreduk-

tion von den Lichtintensitäten der Szene zu den anzeigbaren Wertebereichen, wobei Bilddetails und die Farbwirkung möglichst erhalten bleiben, um den originalen Szeneninhalte wahrnehmen zu können.

Fotomontage/Komposition

Fotomontage (auch bekannt als "Komposition", engl. *image compositing*) bezeichnet das Schneiden und Zusammenfügen von verschiedenen Fotos zu einem zusammengesetzten Bild. Dies kann heutzutage mittels moderner Bildbearbeitungssoftware realisiert werden. Fotomontage kombiniert visuelle Elemente von unterschiedlichen Quellen in einem einzigen Bild, meist um die Illusion zu schaffen, dass all diese Elemente Teil derselben Szene sind. Bei Filmaufnahmen ist Komposition unter verschiedenen Bezeichnungen bekannt: „chroma key“, „blue screen“, „green screen“ und andere Bezeichnungen.



Image Inpainting

Inpainting wird jener Bildbearbeitungsprozess genannt, der bei der Rekonstruktion von schlecht erhaltenen oder nicht vorhandenen Bild- oder Videoteilen Anwendung findet. Im Bereich der Malerei wäre das die Arbeit, die von einem ausgebildeten Bildrestaurator ausgeführt wird. In der digitalen Welt bezieht sich das Inpainting (auch bekannt als Bild- oder Videointerpolation) auf die Anwendung hochentwickelter Algorithmen, um verlorene oder fehlerhafte Teile von Bilddaten zu ersetzen. Die Anwendung des Inpaintings stellt z.B. die Entfernung eines Objektes aus einer Szene dar, um ein beschädigtes Bild zu retuschieren. Für Fotografie und Film kann Inpainting verwendet werden, um Zerfallsmerkmale zu beseitigen (z.B. Risse in Fotografien, Kratzer und Staubflecken im Film) oder um Bildteile hinzuzufügen oder zu entfernen (z.B. Aufnahmedatum). Ziel ist es generell, ein verändertes Bild zu produzieren, in dem die Inpaint-Region mit dem Bild so verschmolzen wird, dass ein gewöhnlicher Betrachter nicht bemerkt, dass das Bild bearbeitet wurde.



Warping

Image Warping ist der Prozess der Bildmanipulation, der jede im Bild vorkommende Form räumlich signifikant verändert. Warping wird verwendet, um die Verzerrung der Bilder zu korrigieren, aber auch für kreative Zwecke (z.B. für Morphing). Während ein Bild auf verschiedenste Weise transformiert werden kann, wird beim reinen Warping jeder Punkt auf einen anderen Punkt abgebildet, ohne die Farbe des Punktes zu ändern. Warping kann auf jeder Funktion basieren, die von (einem Teil) einer Ebene auf eine Ebene abbildet. Bei einer injektiven Funktion kann das Original wieder rekonstruiert werden. Ist die Funktion bijektiv, kann jedes Bild invers transformiert werden. Mathematisch kann das Problem folgendermaßen beschrieben werden: Gegeben sei ein gesampeltes Bild $I(x)$. Warping ist die räumliche (geometrische) Deformation des Quellbildes $I_s(x)$. Das Ergebnis $I_d(x)$ wird Zielbild genannt. Das Warping wird mittels der Warpingfunktion $W(x;p)$ mit den Parametern p bestimmt. Für jedes Pixel x in I_d gibt uns die Warpingfunktion an, woher dieses Pixel im Quellbild I_s stammt: $I_d(x) = I_s(W(x; p))$. Generell gibt es zwei Formen von Transformationen: affine Transformationen und projektive Transformationen (auch perspektivische Transformation oder Homographie genannt).

Identity	$W(x) = x$	
Translation	$W(x; t) = x - t$	
Rigid	$W(x; R, t) = Rx - t$	
Similarity	$W(x; R, a, t) = aRx - t$	
Affine	$W(x; A, t) = Ax - t$	
where $\alpha \in \mathcal{R}$, $A \in \mathcal{R}^{2 \times 2}$, $t \in \mathcal{R}^2$, $R \in \mathcal{R}^{2 \times 2}$, $ R = 1$		

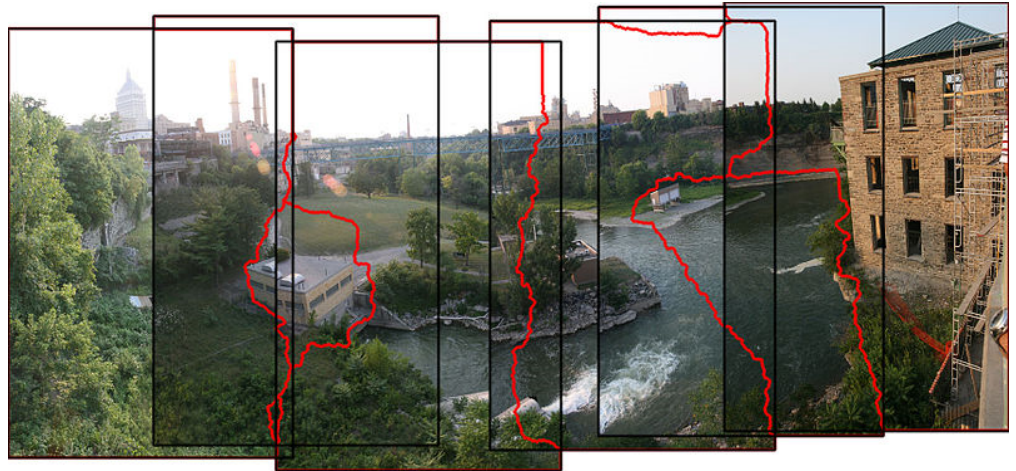
Affine Transformation:
$$\begin{bmatrix} x' \\ y' \\ w \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix}$$

Projektive Transformation:
$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix}$$

Die affine Transformation ist eine Transformation, die gerade Linien und Distanzverhältnisse zwischen Punkten, die auf einer Linie liegen, erhält (z.B. der Mittelpunkt eines Liniensegments bleibt auch nach der Transformation dessen Mittelpunkt). Jedoch werden nicht zwingend Längen und Winkel erhalten. Translation, Rotation, Scherung und ähnliche Transformationen sind alle affin, sowie deren Kombinationen. Die projektive Transformation einer Ebene ist eine Transformation, die in der projektiven Geometrie verwendet wird. Projektive Transformationen erhalten keine Größen oder Winkel. Jedes Bildpaar, das dieselbe ebene Fläche im Raum zeigt, steht mittels einer projektiven Transformation zueinander in Beziehung. Dies ermöglicht viele praktische Anwendungen wie Bildrektifizierung, Bildregistrierung oder die Berechnung der Kamerabewegung zwischen zwei Bildern.

Bildmosaik

Der Prozess zur Erstellung eines Bildmosaiks (engl. *image mosaicing*), auch oftmals als *Stitching* (engl. *to stitch* „nähen“, „heften“) bezeichnet, kombiniert mehrere Bilder mit überlappenden Bereichen zu einem hochauflösenden Panoramabild. Um die Bilder korrekt übereinander legen zu können, ohne einen Parallaxenfehler entstehen zu lassen, muss die Kamera um den Fokuspunkt rotiert werden. Der Prozess kann in drei



Bestandteile aufgeteilt werden: *Bildregistrierung*, *Kalibrierung* und *Blending*. Bei der Bildregistrierung werden korrespondierende lokale Merkmale zwischen den Bildern bestimmt, um die Bilder korrekt "übereinander legen" zu können (siehe Kap. 17). Die Kalibrierung versucht, die Differenzen zwischen einem idealen Linsenmodell und dem realen Linsensystem der Kamera(s) zu minimieren, d.h. optische Defekte (wie Verzerrungen), unterschiedliche Belichtungszeiten der Bilder, Vignettierung, sowie chromatische Aberrationen zu korrigieren. Das Blending ("Verschmelzen") korrigiert die im Kalibrierungsschritt festgemachten Abweichungen und bildet die einzelnen Aufnahmen auf ein gemeinsames Ausgabebild ab. Farben werden zwischen den Bildern angepasst, um die Beleuchtungsunterschiede auszugleichen. Die Bilder müssen so miteinander verschmolzen werden, dass keine Übergänge (engl. *seams*) von einem Einzelbild zum nächsten sichtbar sind.

Morphing

Image Morphing ist ein spezieller Bildeffekt, der nahtlos ein Bild zu einem anderen verändert. Morphing wird oft in surrealen Sequenzen oder im Fantasyfilm-Genre verwendet, um z.B. eine Person in eine andere zu verwandeln. Traditionell wurde solch ein Effekt durch Aus- und Einblendetechniken (engl. *fading*) im Film erzielt. In den frühen 1990ern wurden immer häufiger Computertechniken verwendet, die überzeugendere Ergebnisse lieferten. Hierfür werden beim Überblenden die Bilder auch gleichzeitig anhand von markierten korrespondierenden Punkten verzerrt. Ein Beispiel wäre die Verwandlung eines Gesichts in ein anderes, indem Keypoints im ersten Gesicht (beispielsweise die Kontur der Nase oder Position der Augen) markiert werden, die auch im zweiten Bild existieren. Beim Morphing wird anschließend das erste Gesicht verzerrt, um die Form des zweiten Gesichtes zu erhalten, und zum gleichen Zeitpunkt dieses ausgeblendet und das zweite Gesicht eingeblendet. Zu einem späteren Zeitpunkt wurden höher entwickelte Überblendungstechniken angewendet, die verschiedene Teile eines Bildes auf das andere graduell überblendeten, anstatt das gesamte Bild auf einmal zu wandeln.

