

Seminar Auswertung von 3D-Daten

Vortrag über

Omnivergent Stereo

von Heiko Käßler

1. Einleitung
2. Omnivergent Imaging
 - 2.1 Omnivergente Bilder
3. Kreisförmiges Omnivergent Stereo
 - 3.1 Stereo auf kreisförmigen omnivergenten Bildern
4. Center-strip Stereo
5. Dual-strip Stereo
6. Zusammenfassung

1. Einführung

- lange Zeit waren planare (ebene) perspektivische Bilder das primäre Ergebnis von Computervision-Algorithmen, da sie leicht von herkömmlichen Kameras zu bekommen waren
 - in den letzten Jahren gab es großes Interesse an neuen Bildpräsentationen für Anwendungen der Computervision-Algorithmen und Computergrafik
 - ➔ enthalten Mosaik, die Strahleninformationen aller Winkel entschlüsseln, die in einem optischen Zentrum umgerechnet werden
 - ➔ bekannt als nichtperspektivische Repräsentation (z.B. Lichtfelder, multiperspektivische Panoramas)
 - diese können von einem oder mehreren normalen Bildern oder von neuen Bilderstellungsgeräten abgeleitet werden (catadioptric Kamera)
 - für den letzteren Fall kann man sich die Bilder als Ausgabe eines virtuellen Sensors mit bestimmten Eigenschaften vorstellen
 - ➔ diese Annäherung hat einen großen Vorteil: Performance des Visionsalgorithmus wird wesentlich erhöht, wenn man bessere Inputdaten bereitstellt
 - die Motivation für diese Annäherung begründet sich darin, dass perspektivisch richtige Bilder sich schlecht für die 3D-Rekonstruktion eignen
- Gründe:
1. Bildsequenzen sind sehr redundant und repräsentieren Szenen, die schwer zu verarbeiten sind
 2. Speichern, Übertragen und Verarbeiten von langen Bildsequenzen ist sehr hinderlich
 3. bei traditionellen Stereo-Algorithmen ist es schwierig, Hunderte oder Tausende von Bildern zu verwalten

- im Gegensatz dazu würde ein idealer Sensor Beispieldstrahlen generieren, die zur Optimierung einer Szenenrekonstruktion hergestellt werden
- um die Annäherung zu demonstrieren, definiert man den Typ eines Virtuellen Sensors, der generiert wird, um die Performance eines Stereo-Rekonstruktionsalgorithmus zu verbessern
- um diese Objekte zu bekommen, wird ein virtueller Sensor eingeführt, dieser heißt "omnivergente Kamera" und hat die folgenden Eigenschaften:

- * omnivergence: - jeder Punkt in einer Szene wird von zwei Kameras abgebildet, die zu diesem Punkt "vergent" sind und einen maximalen Vergenzwinkel haben (Bild 1)
 - diese Strategie bewirkt die gleichzeitige Optimierung der Rekonstruktionsgenauigkeit von jedem Punkt der Szene

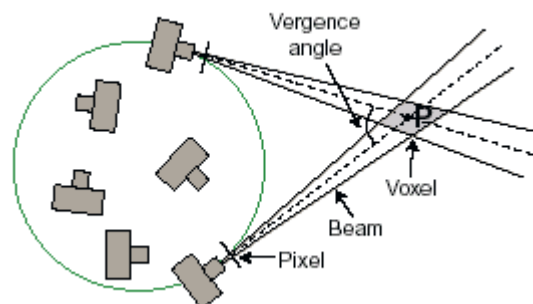


Bild 1

- * einheitliche Genauigkeit: - Szenenstrahlen gehen in alle Richtungen und liefern isotrope (aus allen Richtungen in gleicher Stärke) Tiefenaufösungen
- * Kompaktheit: - Strahleninformationen von hundert oder tausend Inputbildern werden in zwei oder mehr zusammengesetzten Bildern konzentriert, dadurch reduziert sich ein N-Sichten-Konstruktionsproblem auf eine "binocular stereo matching"-Aufgabe mit minimalem Genauigkeitsverlust

- * verstärkende Eigenschaft: - wegen ihrer linearen Epipolargeometrie werden omnivergente Bilder direkt als Input für existierende Stereoalgorithmen genutzt, da sie einen algorithmusunabhängigen Mechanismus zur Optimierung der Rekonstruktionsgenauigkeit bereitstellen

Wozu dient das eigentlich?

- die Idee hinter der omnivergenten Kamera ist, zwei Bilder als Input für einen binocular matching-Algorithmus bereitzustellen, mit dem Ziel, die bestmögliche Szenenrekonstruktion zu erhalten
 - eine weitere Annahme ist, dass man so viele Bilder aus so vielen Standpunkten wie man will, haben kann, aber nur ein zweidimensionaler Vorrat an Pixeln vorhanden ist, aus dem die Bilder konstruiert werden
 - als letztes muss man sicherstellen, dass die konstruierten Bilder den Anforderungen der meisten Stereo-Algorithmen genügen
- Epipolarlinien sollten mit den Horizontallinien übereinstimmen

Welche Standpunkte und Betrachtungsweisen soll man für ein Bild nehmen?

- für jeden 3D-Szenenpunkt P suchen wir das beste Paar an Blickwinkeln, von dem wir P rekonstruieren
- dabei gehen wir davon aus, dass diese Blickwinkel mit den zwei Kameras, die zu P vergent sind, übereinstimmen und deren Vergenzwinkel maximal ist
- mit dieser Strategie kann man zeigen, wie man aus einer 2D-Strahlenvielfalt, die in zwei Bildern angeordnet sind, die bestmögliche Stereorekonstruktion herstellen kann

- ➔ diese Bilder sind ein omnivergentes Stereo-Paar (Output eines virtuellen Sensors, der gleichzeitig zu jedem Punkt der Szene vergent ist)
- mit Rundblick-Stereo-Verfahren ist es also nicht nur möglich, binokulare Stereo-Rekonstruktionen durchzuführen, sondern auch 360°-Rekonstruktionen, einfach durch Veränderung des Inputs für den Algorithmus
 - Stereo auf Rundblicke kann auch Mängel aufweisen:
 1. Die Epipolargeometrie für zylindrische Rundblicke ist nicht linear, was zu Komplikationen bei den Stereo-Suchprozeduren führt
 2. Stereo auf Rundblickbilder produziert keine genaue Rekonstruktion (in manchen Regionen ist sie gut, aber in anderen extrem schlecht)
- ➔ diese beiden Mängel werden durch die folgenden omnivergenten Annäherungen abgestellt

2. Omnivergent Imaging

- **Annahme:** Man kann eine Kamera innerhalb einer festen Region überall hin bewegen, mit dem Ziel, einer so genau wie möglichen Rekonstruktion der Szene außerhalb der Region.
Man kann so viele Bilder wie man will, von jedem Blickpunkt aus erzeugen, den man will, aber nur eine festgelegte Anzahl von Pixeln einer begrenzten Ressource zuschreiben.

Welches Pixel speichert man und von welcher Kamera nimmt man es?

- es wird davon ausgegangen, dass die Bewegung einer Kamera in einer kreisförmigen Region der Ebene stattfindet (Bild 1)

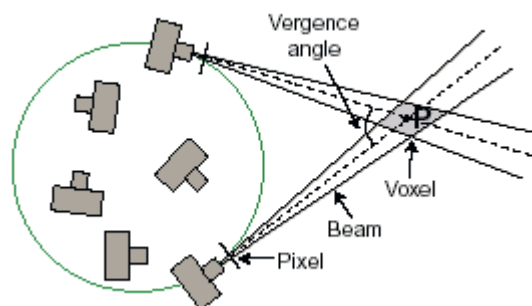


Bild 1

- zur genauest möglichen Lokalisierung eines Punktes P ist es das beste, ein Paar von Kameras auszusuchen, die vergent zum Punkt P sind und einen maximalen Vergenzwinkel haben
- es wird ein omnivergenter virtueller Sensor eingeführt, der den Vergenzwinkel für alle Punkte der Szene maximiert und seine Leistung zu anderen Sensoren für einen spezifischen Kamerablickpunkt innerhalb des Kreises vergleicht
- ein Pixel fängt ein Bündel von Strahlen ein und die Richtung des Bündels ist dadurch definiert, dass es seine Winkelhalbierende ist
- der Schnittpunkt von zwei Bündeln ist ein Voxel
- ein virtueller Sensor mit der Auflösung N bedeutet eine Ansammlung von N Bündeln

- dieses Voxel repräsentiert die Unsicherheit der Rekonstruktion, da der Punkt irgendwo in dieser Voxelregion liegt
 - zu Visualisierung ist es praktisch, anzunehmen, dass der Punkt P in der Mittel des Voxels liegt
 - limitiert man die Anzahl der Bündel (Pixel) auf N , ist die Anzahl der möglichen Bündelschnittstellen endlich, also ist die Anzahl der Voxel auch endlich
 - bei der traditionellen binocular stereo-Konfiguration (zwei Kameras ausgerichtet in einer parallelen oder vergentem Konfiguration) gibt es den Nachteil, dass nur die Rekonstruktion von Untergruppen einer Szene, die vor der Kamera ist, ermöglicht wird
- eine Lösung ist, zwei oder mehrere Rundblickkameras einzusetzen

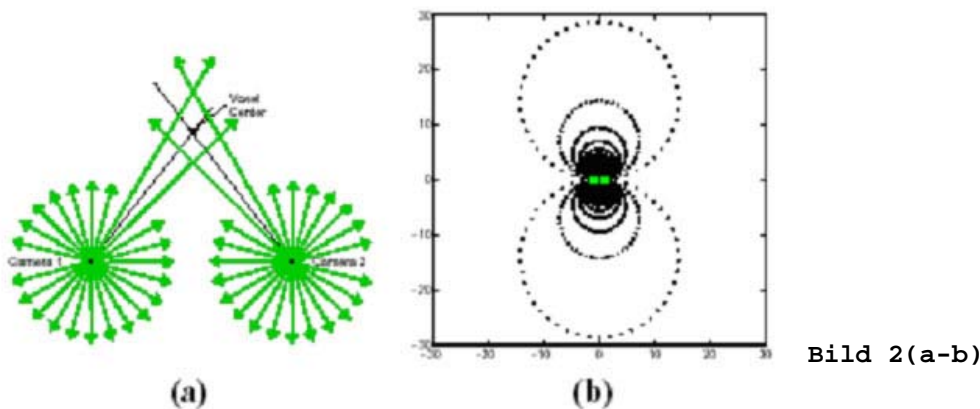
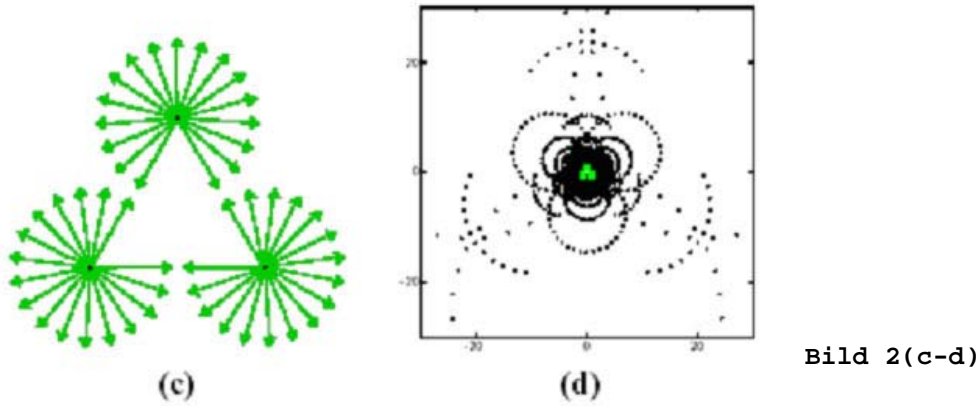
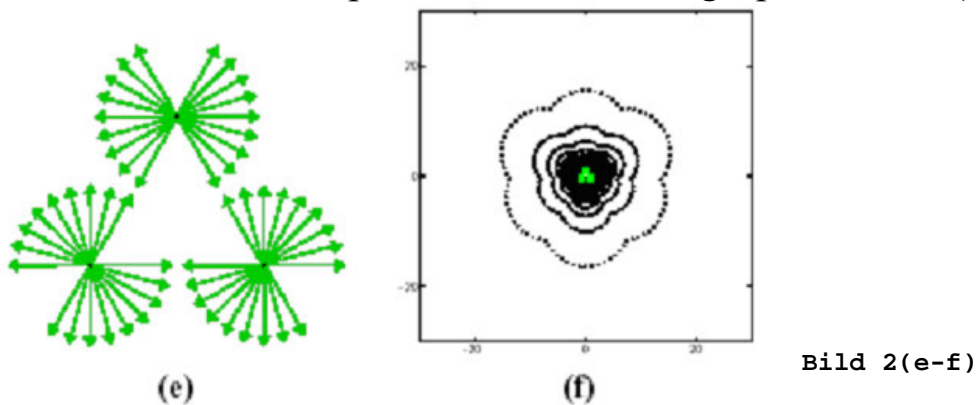


Bild 2b zeigt das Voxelzentrum eines binokularen Rundblicksensors mit 180 Bündeln

- ungewöhnlich dabei ist, dass der Anteil der Szene, die nah der Linie zwischen den beiden Kamerazentren liegt, wo die Distanz zwischen den Voxelzentren viel größer als in anderen Teilen der Szene ist, mit relativ schlechter Genauigkeit rekonstruiert ist
- Einheitlichkeit kann man mit drei Kameras annähern (Bild 2c-d)

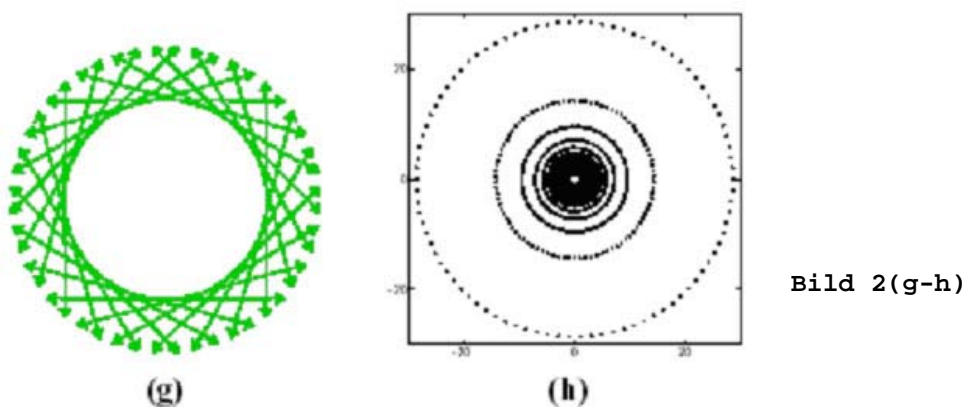


- die Verteilung ist jetzt einheitlicher, einige Teile der Ebene sind jetzt bedeckt durch drei Bündel und andere durch zwei, was eine ungleiche Verteilung der Voxel liefert
- eine weitere Vereinheitlichung kann man erreichen, indem man nur die Strahlen der Außenwinkel des Dreiecks erfasst, welches durch die drei optischen Zentren aufgespannt wird (Bild 2(e-f))



➔ jeder Punkt der Szene wird durch exakt zwei Bündel bedeckt und die Voxelzentren sind gleichmäßiger verteilt

➔ wenn man diese Idee zu einem regulären n-Eck erweitert ($n = N$), erhält man den omnivergenten Sensor (Bild 2(g-h))



2.1 Omnivergente Bilder

- der omnivergente Sensor mit der Auflösung N erfasst zwei Mengen von $N/2$ Bündeln, gleichmäßig verteilt um den Kreis
- die eine Hälfte sind Vorwärtstangenten und die andere Hälfte Rückwärtstangenten (Bild 2 g)



(g)

Bild 2(g)

- speichert man diese Vorwärtstangenten in einem und die Rückwärtstangenten in einem anderen Bild, erhält man zwei eindimensionale omnivergente Bilder

- eine Anpassung zwischen Pixeln in den beiden Bildern kann durch Standard-stereo matching-Algorithmen erreicht werden
- der omnivergente Sensor hat drei Hauptvorteile:
 1. er gibt eine einheitliche Genauigkeit über der ganzen Szene
 2. er reduziert ein N -Sichten Problem auf ein binokulares Stereo-Problem, das den konventionellen Stereo-Algorithmen genügt
 3. er bietet einen Mechanismus zur Maximierung des Vergenzwinkels in jedem Punkt der Szene

3. Spherical Omnivergent Stereo

- Verallgemeinerung der Analyse von einer 2D Kamerabewegung zu einer 3D Kamerabewegung in einem kugelförmigen Körper
- **Annahme:** Man kann eine Kamera innerhalb einer geschlossenen Kugel überall hin bewegen und so viele Bilder machen wie man will.
- für jeden gegebenen Punkt P in der Szene sucht man sich ein Paar Strahlen aus, die zu P vergent sind, so dass der Vergenzwinkel maximal ist
- der Vergenzwinkel ist definiert durch $\angle PXY$ für alle Punktepaare (X, Y) auf der Kugel
- für jeden gegebenen Punkt P wird durch die Menge aller Strahlen von P, die die Kugel streifen, ein Kreis K_p definiert, wo diese Strahlen die Kugel schneiden
- der Vergenzwinkel ist bei der Auswahl von Paaren an gegenüberliegenden Polen von K_p maximal
- kugelförmige Vergenz-Eigenschaft:
 - S ist eine Kugel mit dem Mittelpunkt C und P ist ein Punkt außerhalb von S.
 - $\angle PXY$ ist maximal für alle Punkte X und Y in S, wenn X und Y auf K_p sind und X, Y, P und C zusammen eine Ebene aufspannen
- wenn man zulässt, dass P sich gegenüber der Kugel verändert, reicht es aus, jeden Punkt der Szene mit zwei Strahlen, die einen maximalen Vergenzwinkel haben, abzubilden
- die Menge der Tangentenstrahlen ist aber dreidimensional, deshalb ist es möglich, ein 3D-omnivergentes Bild durch Speichern von Tangentenstrahlen in einer räumlichen Datenstruktur zu erstellen
- solange es Stereoalgorithmen gibt, die 3D Bilder bearbeiten, ist das Speichern und Bearbeiten eine grauenvolle Aufgabe
 - es ist aber auch noch möglich die Dimensionalität auf 2D zu reduzieren

Wie geht das?

- die Kugel kann durch einen Kreis, der um 180° um eine Achse L rotiert, generiert werden
- die resultierende Menge an Kreisen entspricht den Längengraden auf der Kugel mit der Polarachse L (Bild 3a)

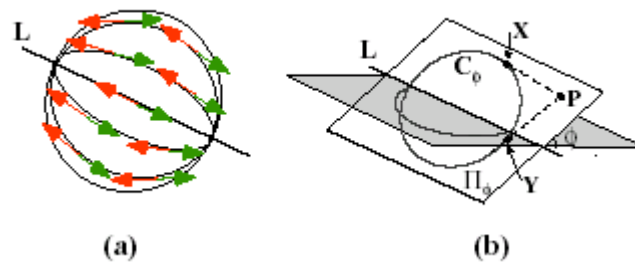
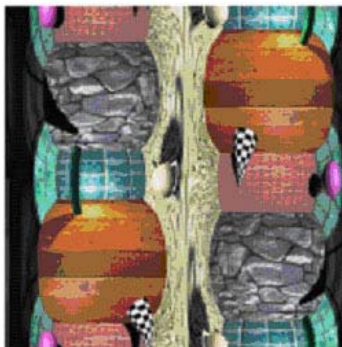


Bild 3

- jeder Kreis C_ϕ liegt in einer Ebene Π_ϕ , und die Menge all solcher Ebenen definiert eine Ein-Parameter-Familie von Kreisen über L , die „pencil“ genannt wird
- betrachtet man nun einen Punkt P außerhalb der Kugel
→ P muss mindestens in einer der Ebenen des Pencils liegen
(Punkte auf L liegen in allen Ebenen)
- betrachtet man nun die zwei Strahlen von P , die zu C_ϕ tangential sind
→ X und Y sind die zwei Tangentialpunkte auf C_ϕ
→ X und Y liegen auch auf der Kugel K_p , da die Strahlen auch Tangenten zur Kugel sind
- da X, Y und P auf Π_ϕ liegen, folgt aus der Vergenz-Eigenschaft, dass diese beiden Strahlen den maximalmöglichen Vergenzwinkel für alle Strahlen in der Kugel liefern
- Konsequenz daraus ist, dass es genügt, den 2D-Raum der Strahlen zu erfassen, die mit den Tangenten von allen Kreisen C_ϕ übereinstimmen

- diese Konstruktion garantiert also, dass jeder Punkt der Szene durch ein Paar von Strahlen mit dem maximalen Vergenzwinkel erfasst wird
- diese Strahlen werden in (Φ, θ) -Bilder durch Bewegung der Kamera und Erhöhung eines Winkels θ entlang C_Φ in einem Zeilenvektor gespeichert
- das Aufstellen von Zeilenvektoren für nachträgliche Werte von Φ liefert ein kugelförmiges omnivergentes Bild (spherical omnivergent image (SOI))
- die gegenüberliegenden Seiten dieses Bildes stimmen für $\theta = 0,360$ und $\Phi = 0,360$ überein, so dass das Bild die Form einer Wulst hat (Bild 6b)

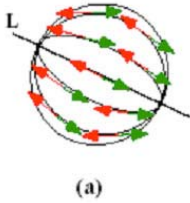


(b)

Bild 6(b): kugelförmiges omnivergentes Bild

3.1 Stereo auf Spherical Omnivergent Images

Bild 3(a)



- man kann nun alle longitudinalen Tangenten (Bild 3a) nur durch das Erfassen der Vorwärtstangenten jedes Kreises C_Φ und einer Drehung um 360° in Φ erfassen

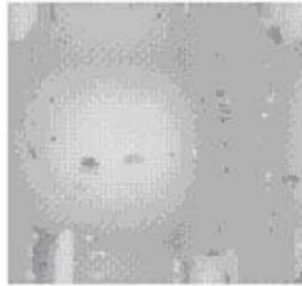
- das Ergebnis ist ein einziges (Φ, θ) -Bild, in dem jeder Punkt doppelt dargestellt ist (eine 360° -Rotation von C_Φ über L bedeckt die Kugel zweimal, und die Tangenten drehen sich in der zweiten Phase um): Bild 6(b)
- Alternativ ist es auch möglich, ein SOI-Stereopaar durch zwei halbkugelförmige omnivergente Bilder zu produzieren
 → eines durch die Vorwärtstangenten und eines durch die Rückwärtstangenten für eine 180° -Rotation
- für das Stereo matching ist die zweite Konstruktion nützlicher, weil sie direkt zwei Bilder produziert (Bild 7(a-b))



Bild 7(a-b)

- eine Schlüsseleigenschaft des SOI ist, dass seine Epipolargeometrie identisch zu dem eines Paares der verbesserten ebenenperspektivischen Bilder ist
- das besondere daran ist, dass parallele Punkte in derselben horizontalen Scanlinie der beiden Bilder eines kugelförmigen Stereo-Paares erscheinen
 → diese Eigenschaft ermöglicht es, kugelförmige omnivergente Bilder direkt von vorhandenen Stereo matching-Algorithmen zu verarbeiten

- Bild 7c zeigt das Anwenden eines konventionellen flächenbasierten Zuordnungs-Stereo-Algorithmus auf ein Stereo-Paar



(c)

4. Center-Strip Stereo

- es gibt auch noch eine weitere Variante einer omnivergenten Kamera, da es schwierig ist, eine Kamera über eine kugelförmige Oberfläche zu bewegen
→ center-strip omnivergenter Sensor
- es gibt nur eine einzige Kamerabewegung entlang eines kreisförmigen Pfades
- auf jedem Punkt im Kreis erfasst die Kamera einfallendes Licht entlang des 1D-Raumes der Strahlen, die orthogonal zum Kreisradius sind (Bild 4)

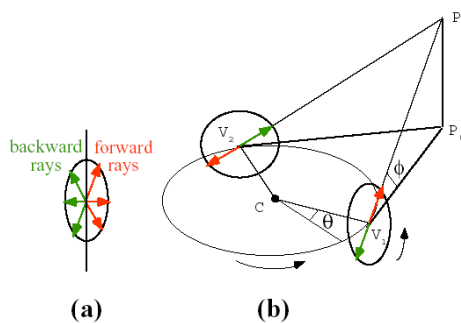
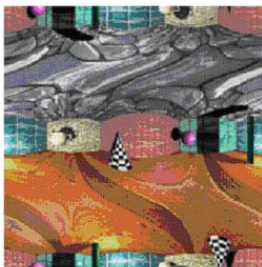


Bild 4

- wenn die optische Achse der Kamera tangential zum Kreis ist und die Punkte der Y-Achse der Kamera normal zur Bewegungsebene sind, dann werden diese Strahlen zum center-strip des Bildes projiziert
- center-strips von aufeinander folgenden Positionen entlang des Kreises werden in aufeinander folgenden Spalten eines neues (Φ, θ) -Mosaikbildes gesetzt

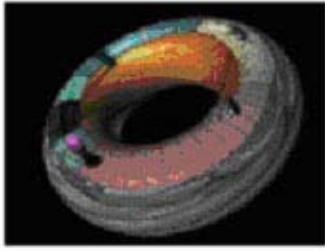
→ center-strip omnivergentes Bild (COI) (Bild 6c)



(c)

- COI's sind Beispiele für multiperspektivische Rundblicke

- es gibt viele Gemeinsamkeiten zwischen COI's und SOI's
- wie im kugelförmigen Fall haben die COI's ringförmige Topologie (Bild 7 f)

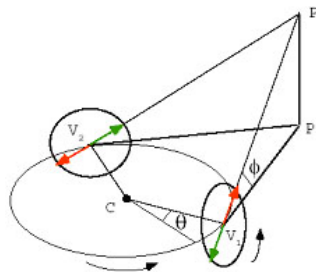


(f)

Bild 7(f)

→ man kann also ein COI in ein Stereopaar, durch das Setzen von Vorwärtstangenten in ein Bild und Rückwärtstangenten in ein anderes Bild, zerlegen

- die COI's haben also horizontale Epipolarlinien (Bild 4b)



(b)

Bild 4

- projiziert man den Szenenpunkt P senkrecht zu P_0 , der tangential zum Kreis ist, wird er von zwei Punkten V_1 und V_2 auf dem Kreis abgebildet
- P liegt auf Strahlen von P_0 die tangential zum Kreis sind
- weil $\|P_0V_1\| = \|P_0V_2\|$, ist es leicht zu sehen, dass $\angle PV_1P_0 = \angle PV_2P_0$
- da Strahlen in derselben Zeile eines COI mit demselben Wert von Φ übereinstimmen, folgt daraus, dass Szenenpunkte in derselben Zeile von Vorwärts- und Rückwärts-COI Bildern erscheinen
- die COI's maximieren also nicht den Vergenzwinkel für jeden 3D-Szenenpunkt, vielmehr wird $\angle P_0V_1V_2$ (horizontaler Vergenzwinkel) maximiert, was aber eine gute Annäherung ist, wenn das vertikale Feld der Sichten beschränkt ist

5. Dual-Strip Stereo

- die center strip-omnivergente Kamera fordert das Erfassen eines ganzen Kreises von Strahlen
- spezielle optische Geräte werden gebraucht, um Vorwärts- und Rückwärtsstrahlen in einem Punkt einzufangen
- es gibt eine Alternative
→ dual-strip Sensor
- wie der center-strip Sensor fängt der dual-strip Sensor Strahlen von Blickpunkten entlang eines Kreises ein, aber er ist aus zwei symmetrischen nichtzentralen Schnittbildern geformt (Bild 5a)
- beide Strahlen der dual-strip-Kamera in der Ebene sind tangential zu einem inneren Kreis $R' = R \cdot \sin(\alpha)$, wobei R der Radius des äußeren Kreises ist, auf dem die Kamera entlang läuft
- α ist der Winkel der dual-strip-Kamera bei der normalen Richtung
- (Bild 5b) die epipolaren Begrenzungen der dual-strip-Bilder sind wieder horizontale Linien
- weil $\|P_0V_1'\| = \|P_0V_2'\|$ (tangential zu R') und $\|V_1V_1'\| = \|V_2V_2'\|$ (Dualkameras sind symmetrisch) ist, ist auch $\angle PV_1P_0 = \angle PV_2P_0$

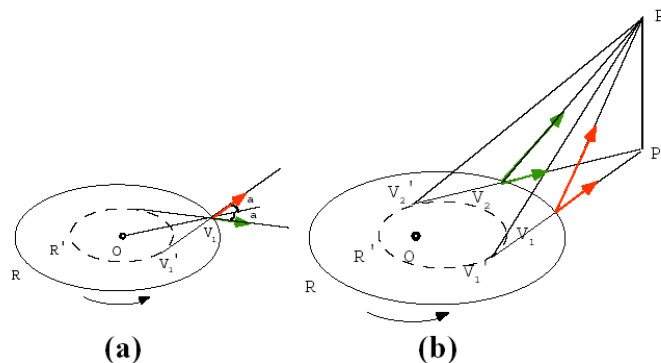


Bild 5

- für dual-strip-Kameras sind horizontale Vergenzwinkel relativ zum inneren Kreis R' maximiert

- das besondere an dual-strip-Bildern ist, dass sie von regulären Perspektivkameras, die um einen Kreis laufen, erstellt werden können
- zwei Spalten von Pixeln, symmetrisch über einem Zentrum, werden von jedem Bild gewählt
- ein Mosaikbild wird von der linken Spalte zur nachfolgenden Position gebildet und ein anderes von der rechten Seite (Bild 8)

Bild 8: 60igste Spalte oben, 260igste Spalte unten



- diese Bilder werden durch das Stapeln der 60igsten und 260igsten Spalte, bzw. durch eine Sequenz von 1350 Frames (mit einer Auflösung von 320*240), die von einer normalen Digitalkamera aufgenommen werden, gebildet

6. Zusammenfassung

- das Konzept des virtuellen Sensors wird eingeführt, da traditionelle perspektivische Bilder nicht gut für das Lösen von Computervision-Aufgaben geeignet sind
- Grund: Optimierung des Inputs von Computervision-Algorithmen zur Erreichung besserer Resultate
- es werden die omnivergenten Sensoren eingeführt, die gleichzeitig zu jedem Punkt der Szene vergent sind und einen maximalen Vergenzwinkel haben
- es wird beschrieben, wie man omnivergente Bilder durch Verarbeitung einer Sequenz von perspektivischen Inputbildern erzeugt
- omnivergente Bilder repräsentieren simultanes Szenenerscheinen von vielen Blickwinkeln aus und sind so konstruiert, dass der Vergenzwinkel in jedem 3D-Punkt maximal ist
 - ➔ omnivergentes Stereo produziert genaue 3D-Szenenmodelle mit minimalem Rekonstruktionsfehler
- weiterhin haben omnivergente Bilder horizontale Epipolargeometrie, was die direkte Anwendung von konventionellen binokularen Stereoalgorithmen zur Bestimmung von Pixelübereinstimmung ermöglicht
- es wurden drei verschiedene Varianten des omnivergenten Sensors vorgestellt
 - kugelförmiger omnivergenter Sensor, der die Eigenschaft des maximalen Vergenzwinkels einhält
 - center-strip Sensor, der nur den horizontalen Vergenzwinkel maximiert, aber leichter in der Praxis zu implementieren ist

der center-strip Sensor ist eine gute Wahl, wenn die Szene relativ weit weg ist und das vertikale Blickfeld beschränkt ist

- der dual-strip Sensor kann leicht realisiert werden durch das Bewegen einer normalen Kamera um einen Kreis
- es bleiben mehrere wichtige Themen, die zusätzlich bearbeitet werden müssen:
 1. Es wird nur das Problem des binokularen Stereo betrachtet (man bekommt bessere Resultate in der Praxis durch Formulierung eines N-okularen Stereo-Problems).
Weil omnivergente Kameras einen maximalen Vergenzwinkel haben, ist Stereo matching mit omnivergenten Bildern bei Lichtveränderungen empfindlicher als bei reduzierten baseline-Bildern
Zusätzliche Bilder können durch das Auswählen von zusätzlichen nichttangentialen Strahlen und/oder Auswählen von verschiedenen Mengen von Inputblickpunkten generiert werden
 2. Die Verallgemeinerung der Kreis- und Kugelfälle zu beliebigen Kamerapfaden und Flächen ist ein wichtiges Thema der künftigen Arbeit
Einige der Untersuchungen verallgemeinern zu einer glatten Oberfläche, d.h. Speichern von Tangentenstrahlen ist ausreichend, um den Vergenzwinkel zu maximieren
Die Wahl des Kamerapfades sollte idealer Weise abhängig sein von einer a priori-Information über die Szenenbelegung
 3. Man erforscht, wie man reale Geräte bauen kann, die die omnivergenten Sensoren realisieren können