



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Studienarbeit

Nikolaus Rusitska

nikolaus.rusitska [at] haw-hamburg.de

Scene Reconstruction
with Multiple View Geometry

Related Work

Inhalt

1	Einleitung	3
2	Techniken	6
2.1	Szenengeometrie	6
2.2	Raumgeometrie	7
3	Related Work	9
3.1	Merkmale und Korrespondenzen	9
3.1.1	SIFT	9
3.1.2	LESH	10
3.1.3	GLOH	11
3.1.4	SURF	11
3.2	Szenengeometrie und 3D-Punkte	13
3.2.1	Epipolar Geometry Toolbox Project	13
3.2.2	Multiple View Geometry Toolbox	14
3.2.3	BEEM	14
3.3	Triangulation und Meshing	15
3.3.1	Meshlab	15
3.3.2	Gyroviz	15
3.3.3	CGAL	16
4	Fazit	17
4.1	Merkmale und Korrespondenzen	17
4.2	Szenengeometrie und 3D-Punkte	17
4.3	Triangulation und Meshing	17
5	Referenzen	19

1 Einleitung

Die Rekonstruktion aufgenommener Szenen findet in mehreren Bereichen Anwendung und kann für verschiedene Zwecke eingesetzt werden.

Wird eine Szene anhand von mehreren Bildern analysiert – seien dies nun einzelne Fotografien oder die Einzelbilder eines Videos – spricht man auch von Multiple View Geometry.

Die Problematik dabei ist die Rekonstruktion dreidimensionaler Objekte aus zweidimensionalen Quellen. Ziel ist es jeweils die Tiefeninformationen zurückzugewinnen, die bei der zweidimensionalen Bildaufnahme verloren wurden.

Das Hauptaugenmerk bei dieser Arbeit liegt auf der Rekonstruktion anhand von Videodaten.

Dabei gibt es verschiedene Verfahren mit denen dieses Erreicht werden kann und verschiedene Schwierigkeiten, die bei den einzelnen Verfahren gehandhabt werden müssen.

Eine Möglichkeit ist die Aufnahme einer Szene mit einer zuvor kalibrierten Kamera. Dabei werden vor der Aufnahme Messungen durchgeführt um Verzerrung, Brennweite und andere Parameter der Kamera zu bestimmen. Sind die Parameter der Kamera bekannt, lassen sich so die Tiefeninformationen wieder rekonstruieren.

Allerdings lässt sich deswegen nur mit Material arbeiten, bei dem auch die Daten der verwendeten Kamera bestimmt wurden. Szenen, die mit einer unbekanntem Kamera gefilmt wurden, lassen sich auf diese Weise nicht mehr rekonstruieren.

Ein weiterer Weg ist die Aufnahme von Videomaterial mit darauf spezialisierten Kameras. So gibt es beispielsweise stereoskopische Kameras, bei denen zwei Kameras auf einmal verwendet werden und die menschliche Sicht mit zwei Augen imitiert wird.

Hierbei vergrößern sich die Einschränkungen noch im Vergleich zur vorher genannten Technik, da nur Material dieser speziellen Kameras verwendet werden kann.

Es gibt weitere Techniken mit vergleichbaren Einschränkungen, bei denen zum Beispiel eine Lichtquelle in einer bekannten Position zur Kamera angebracht ist und dabei durch den Lichteinfall Tiefeninformationen gewonnen werden. Auch gibt es Technologien bei denen moduliertes infrarotes Licht verwendet wird. Dabei kann durch die Phasenverschiebung des wieder zurück kommenden Lichtes die Entfernung gemessen werden.

Auch diese Techniken haben allerdings gemeinsam, dass nur Material von speziellen Kameras verwendet werden kann. Es ist dabei nicht möglich mit Videomaterial zu arbeiten, das nicht speziell vorbereitet wurde.

Hier sollen allerdings Möglichkeiten behandelt werden, mit denen Videomaterial unbekannter und ‚normaler‘ Kameras verarbeitet werden kann.

Anwendungen, die dafür denkbar sind, ist die Integration von computergenerierten Objekten in reales Filmmaterial. Dieses kommt beispielsweise bei der Postproduktion von Filmen zu Einsatz, oder bei realitätsinteraktiver Software (Augmented Reality)



Ein animiertes Objekt, das in eine reale Szene eingefügt wurde

Das Ziel ist also ein dreidimensionales Modell aus gewöhnlichem Videomaterial zu erzeugen und die Positionen der Kamera während der Aufnahme zu bestimmen. Durch die Verfolgung der Kamera können die künstlichen Objekte in der gefilmten Szene positioniert werden. Durch ein 3D-Modell der Szene wird die benötigte Verdeckung der künstlichen Objekte durch Objekte in der Szene erreicht.

Hierzu werden Übereinstimmungen (Korrespondenzen) zwischen Einzelbildern gesucht. Anhand der Korrespondenzen wird dann die räumliche Beziehung dieser Einzelbilder ermittelt.

Es wird dann die Geometrie beider Kamerasichten ermittelt und so kann aus den zweidimensionalen Positionen der Übereinstimmungen in den Kamerabildern eine dreidimensionale Position des Punktes im Bezug auf die zugrunde gelegten Kamerapositionen berechnet werden.

Daraus ergibt sich dann eine Sammlung von Punkten in einem dreidimensionalen Raum. Zusätzlich können auch die Kamerapositionen in diesem Raum bestimmt werden. Um aus diesen Punkten ein 3D-Modell zu ermitteln müssen durch Triangulation die Punkte in (zusammenhängende) Dreiecke zusammengefasst werden. So erhält das Modell eine Oberfläche.

Ein weiteres Thema bei der dreidimensionalen Rekonstruktion ist die Texturierung des erstellten Modells. Hierbei werden Teile der gefilmten Bilder wieder auf die Dreiecke aufgebracht. Für die oben beschriebene Anwendung ist dieses allerdings unwichtig, das im resultierenden Videomaterial wieder das Ausgangsmaterial benutzt werden soll. Das

Modell ist nur dafür da, die Teile der generierten Objekte auszublenden, die vom Modell in der Szene verdeckt würden.

Daher wird dieses Thema hier hintenangestellt und nicht speziell behandelt.

2 Techniken

Die Schritte der Rekonstruktion einer Szene lassen sich in zwei Themengebiete einteilen. Die sich vor allem durch die verwendeten Dimensionen unterscheiden.

Die Geometrie der Szene, in der die Beziehungen zwischen den Einzelbildern beschrieben werden.

Die Geometrie des dargestellten dreidimensionalen Raumes

2.1 Szenengeometrie

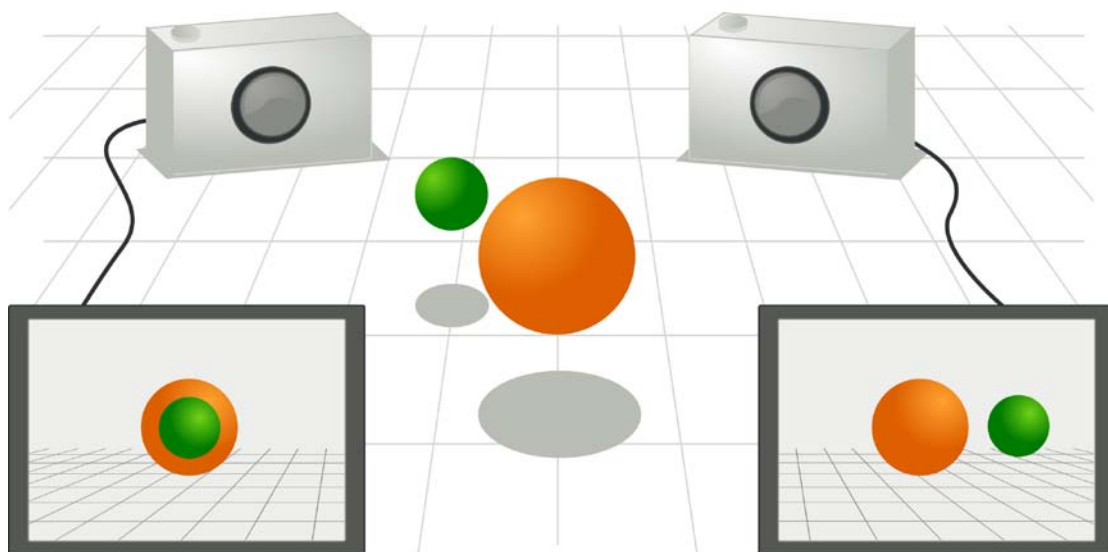
Die Geometrie der abgebildeten Szene bezieht sich auf die zweidimensionalen Abbildungen der Szenen. Es werden stets zweidimensionale Räume zugrunde gelegt. (Allerdings können auch homogene Koordinatensysteme verwendet werden, die nutzbaren Koordinaten bleiben dabei zweidimensional.)

Es werden Beziehungen zwischen den Einzelbildern hergestellt.

Ein Weg um dieses zu erreichen ist das Verknüpfen von zwei Einzelbildern durch die Suche nach Korrespondenzen der beiden Bilder. Dabei werden auffällige Punkte im ersten Bild gesucht und mit Merkmalen beschrieben. Dieses wird beim folgenden Bild ebenfalls durchgeführt; danach werden die Beschreibungen der Punkte beider Bilder miteinander verglichen um Punkte zu finden, die in beiden Bildern vorhanden sind.

Aus den Unterschieden der Positionen der korrespondierenden Punkte wird versucht die Lage der beiden Abbildungen zueinander zu ermitteln.

Stellt man sich ein Einzelbild als Ebene im Raum vor, so wird dann versucht zwei Ebenen so zueinander zu positionieren, dass die Positionierung passend zu den sich wiederholenden Punkten ist.



Aufnahme einer Szene aus zwei verschiedenen Blickwinkeln

Daraus lassen sich die äußeren Parameter der beiden Kameras gewinnen. Hierbei handelt es sich um die Lage im Raum und die Positionierung. Dadurch ist nun die Geometrie bekannt, die beide Bilder der Szene miteinander verbindet.

Diese Geometrie beschreibt das Verhältnis beider Bilder in drei Dimensionen. Dennoch wird sie hier mit in die zweidimensionale Kategorie aufgenommen, da sie aus zweidimensionalen resultiert und lediglich eine dreidimensionale Transformation verkörpert. Dreidimensionale Daten werden erst im nächsten Schritt daraus gewonnen, der dann in den dreidimensionalen Bereich der Raumgeometrie eingeordnet wird. Bei der Geometrie der Szene wird von nur einer dreidimensionalen Transformation ausgegangen, da angenommen wird, dass die erste Abbildung beziehungsweise die erste Kamera sich untransformiert im Koordinatenursprung befindet. Es braucht nur die Transformation der zweiten Abbildung im Bezug auf die Erste angegeben zu werden. Will man mehrere Bilder im Bezug zueinander setzen, berechnet man danach die Geometrie des zweiten Bildes zum dritten Bild und transformiert danach beides mit der Transformation vom ersten zum zweiten Bild. Dieses kann dann für beliebig viele Bilder so fortgesetzt werden.

2.2 Raumgeometrie

Die Raumgeometrie beschreibt die Geometrie der realen aufgenommenen Szene beziehungsweise der rekonstruierten Szene, bei der es sich lediglich um ein Modell der aufgenommenen Szene handelt.

Mithilfe der Szenengeometrie kann aus zwei korrespondierenden Punkten der dazugehörige dreidimensionale Raumpunkt abgeschätzt werden.

Wird dieses für die Korrespondenzen aller zu analysierenden Einzelbilder durchgeführt, erhält man eine Wolke von dreidimensionalen Punkten. Diese Punkte liegen auf der Oberfläche der Objekte in der Szene.

Aus diesen Punkten soll ein 3D-Modell der Szene erzeugt werden. Ein 3D-Modell besteht aus zweidimensionalen Figuren, die im Raum angeordnet sind und so die Oberfläche eines dreidimensionalen Objektes ergeben. Die gefundenen Punkte müssen demnach in geometrischen Figuren zusammengefasst werden.

Die einfachste geometrische Figur (für diesen Anwendungsbereich) ist das Dreieck, da es mit den wenigsten Punkten eine Figur aus geraden Linien und mit einer Oberfläche bildet.

Bei der Zusammenfassung von Punkten zu einem Dreieck wird auch von Triangulation gesprochen. Dabei muss ein Triangulationsalgorithmus nicht zwingend genau drei Punkte zu einem Dreieck verbinden. Es können auch mehr Punkte zusammengefasst werden, weil auf Punkte verzichtet werden kann, die auf der Oberfläche des Dreiecks, oder sehr dicht daran liegen.

Würde man ausschließlich ein solches Verfahren anwenden, würde eine große, verbundene Oberfläche für die gesamte Szene resultieren. Soll eine Szene möglichst

originalgetreu rekonstruiert werden, sollte versucht werden Objekte passend zur Originalszene zu gruppieren.

Da allerdings – wie einleitend beschrieben – das 3D-Modell vor allem zur Verdeckung gesucht wird, wird dieses vorerst nicht beachtet.

Wurde eine Oberfläche (oder gegebenenfalls auch ganze Objekte) geformt, kann diese mit einer Textur versehen werden. Um die Oberfläche eines Dreiecks zu texturieren, können die Punkte im Originalbild benutzt werden, die zu den Punkten des Dreiecks gehören. Dabei ist allerdings zu beachten, dass diese Fläche mindestens zweimal verfügbar ist, nämlich in beiden Bildern, in denen die Korrespondenzen gefunden wurden. Die Fläche kann zusätzlich auch in mehreren Quellbildern enthalten sein. Im Idealfall sollten alle Kandidaten für die Textur miteinander verknüpft werden, damit später an den Kanten der Dreiecke keine Absätze entstehen.

Wie eingangs beschrieben, wird auch dieses Teilthema nur am Rande behandelt, da es für die geplanten Anwendungsfälle kaum Bedeutung hat.

3 Related Work

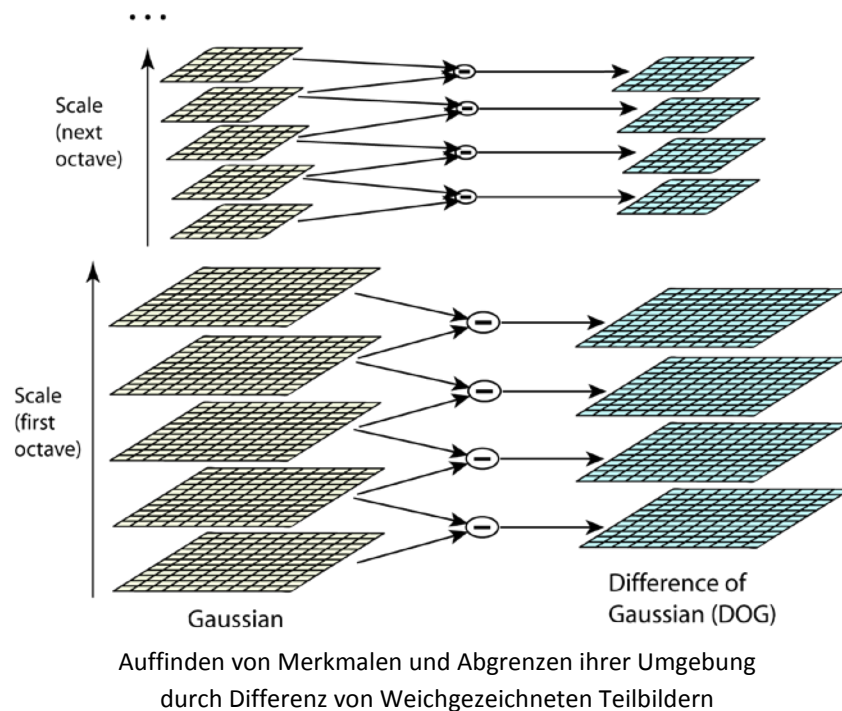
3.1 Merkmale und Korrespondenzen

Im Folgenden werden einige Projekte vorgestellt, mithilfe deren Algorithmen Merkmale in Bildern gefunden werden können. Die gefundenen Merkmale werden durch Datenvektoren (auch Deskriptoren genannt) beschrieben. Vergleicht man die Deskriptoren der Merkmale von zwei Bildern können Korrespondenzen zwischen beiden Bildern gefunden werden, indem anhand der Daten die Ähnlichkeit von jeweils zwei Vektoren untersucht wird. Die beiden Merkmale, die die größte Ähnlichkeit besitzen werden als Korrespondenz angesehen. Dabei kann teilweise auch ein Mindestmaß für die Ähnlichkeit festgelegt werden.

3.1.1 SIFT

In der Arbeit „Distinctive Image Features from Scale-Invariant Keypoints“ wird ein Algorithmus zum Extrahieren invarianter Merkmale aus Bildern beschrieben. (Lowe, 2004) Dieser Algorithmus wird SIFT genannt und steht für „Scale-Invariant Feature Transform“. Ziel dabei ist es robuste Merkmale zu erkennen, die auch trotz starker Transformation wiedererkannt werden können. Der Algorithmus ist stabil gegenüber Transformationen wie Translation, Rotation (zweidimensional), Skalierung und Änderung des dreidimensionalen Ausgangspunktes. Zudem werden Merkmale auch trotz Änderungen der Beleuchtung erkannt. Hierbei können sich sowohl die Helligkeit als auch die Farbgebung des Bereiches des Merkmals verändern.

Um skalierungsinvariante Merkmale zu finden, wird das Eingangsbild in kleinere Bereiche unterteilt, die durch einen Gauß'schen Weichzeichner geglättet werden. Ein geglättetes Teilbild und seine Nachbarn werden jeweils von einander abgezogen. Die geglätteten Bildteile werden danach um den Faktor zwei herunter skaliert und es wird eine weitere Differenz der geglätteten Bereiche gebildet. Dieses kann in beliebiger Tiefe (begrenzt durch die Auflösung des Bildes) weitergeführt werden. Das Ziel ist hierbei Merkmale zu finden, die sich von dem um sie herum liegenden Hintergrund unterscheiden. Dieses wird durch die Subtraktion der einzelnen Bereiche von seinen Nachbarn erreicht.



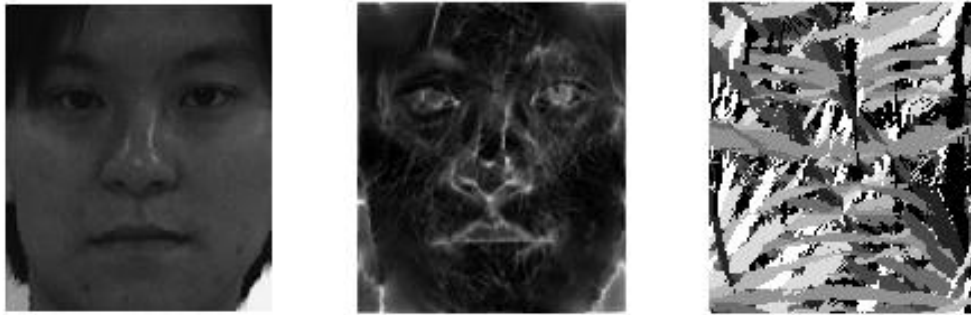
Ein Deskriptor beschreibt die lokale Umgebung eines gefundenen Merkmals. Er wird erstellt indem mehrere Histogramme um das gefundene Merkmal herum gebildet werden. Es werden dabei mehrere Histogramme einzelner Teilbereiche des gefundenen Merkmals erstellt. Den Teilbereichen wird eine Orientierung in Relation zum Merkmal zugeordnet. Diese ist dann im Vergleich mehrerer Merkmale rotationsinvariant. Der Deskriptor eines SIFT-Merkmals enthält insgesamt 128 Elemente.

Bei dem Histogramm eines Bereiches spielen Farb- und Helligkeitsverläufe eine große Rolle. Es wird vermutet, dass auch die menschliche Wahrnehmung sich an gerichteten Verläufen orientiert um Merkmale wiederzuerkennen.

SIFT ist oft der Ausgangspunkt anderer ähnlicher Algorithmen, die nach einem vergleichbaren Schema arbeiten.

3.1.2 LESH

Das Local Energy based Shape Histogram (LESH) geht aus einer Projektarbeit an der Technischen Universität Berlin hervor (Sarfraz & Hellwich, 2008). Es handelt sich hierbei um einen SIFT-ähnlichen Algorithmus, der auf die Erkennung von Merkmalen in Gesichtern spezialisiert ist. Es wird die lokale Energie von Bildbereichen betrachtet. Dies bedeutet, dass die Frequenz der Helligkeit und Farbe betrachtet wird, um eine Kantenerkennung zu erreichen. Dadurch kann die Form von Objekten betrachtet werden um Merkmale zu erkennen.



Kantenerkennung und Erstellung von Formen anhand der Kanten bei LESH

Der Beschreibungsvektor von LESH wird ähnlich dem von SIFT gebildet. Der Unterschied zu SIFT besteht darin, dass bei der (relativen) Orientierung des Deskriptors die ermittelte Form, an der der Marker liegt mit einbezogen wird.

3.1.3 GLOH

GLOH (Gradient location-orientation histogram) ist ein SIFT-ähnlicher Algorithmus. Um die Auffindung von Merkmalen zu verbessern werden bei der Verknüpfung der Teilbilder, die mit einem Gaußfilter geglättet wurden, neben der Subtraktion (wie bei SIFT) die Teilbilder zusätzlich mit dem Laplace-Operator (einem Differentialoperator) verknüpft. Hierdurch soll eine höhere Genauigkeit bei der Festlegung der Umgebung des Merkmals erreicht werden, über die das Histogramm für den Deskriptor gebildet wird. (Mikolajczyk & Schmid, 2005)

GLOH verwendet mehr Einzelabschnitte als SIFT um die Histogramme zu berechnen und kommt so auf 272 Elemente. Diese werden dann allerdings wiederum reduziert. Es werden nur die Histogramme mit den größten Eigenvektoren verwendet, wodurch nur die 128 ausdrucksstärksten Elemente verwendet werden.

3.1.4 SURF

Bei diesem Projekt wurde ein Verfahren entwickelt, das auf der Vorgehensweise von SIFT beruht. Bei dem Speeded Up Robust Features (SURF) Projekt wurde versucht die Geschwindigkeit und Robustheit von SIFT zu verbessern. (Bay, Ess, Tuytelaars, & Gool, 2008)

Während bei SIFT (und vergleichbaren Algorithmen) die Bildbereiche mit einem Gaußfilter geglättet werden, verwendet SURF einen Medianfilter (auch Mittelwertfilter genannt), bei diesem Filter wird nicht für jeden Pixel der Gauß'sche Mittelwert seiner Umgebung berechnet, sondern lediglich der (wertemäßig) mittlere Wert der Umgebung gewählt. Dieses Verfahren ist dadurch um einiges schneller, als Algorithmen die eine Gaußfilterung durchführen. Wie bei SIFT werden danach die Bildteile von einander abgezogen um die Umgebung eines Markers zu erkennen.



Merkmale sowie deren Umgebung und Orientierung bei SURF

Auch bei der Bildung des Deskriptors geht SURF einen anderen Weg. Bei SIFT und anderen ähnlichen Algorithmen werden Histogramme (teilweise mit Verbesserungen, wie bei GLOH) der Umgebung des Merkmals gebildet um das Merkmal zu beschreiben. Bei SURF werden stattdessen Wavelet-Transformationen der Umgebung des Merkmals gebildet. Diese haben den Vorteil, dass sie eine bessere Verteilung im Frequenzraum als Histogramme haben. Dadurch sinkt die Verwechslungsgefahr und es genügen weniger Daten als bei einem Histogramm. Somit wird der SURF-Deskriptor auf 64 Elemente reduziert, was auch den Vergleich von Deskriptoren beschleunigt.

3.2 Szenengeometrie und 3D-Punkte

Dieser Abschnitt beschäftigt sich mit Projekten und Algorithmen, mit deren Hilfe die Geometrie einer Szene abgeschätzt werden kann und mit Hilfe einer bekannten Geometrie und Korrespondenzen entsprechende Raumpunkte berechnet werden können.

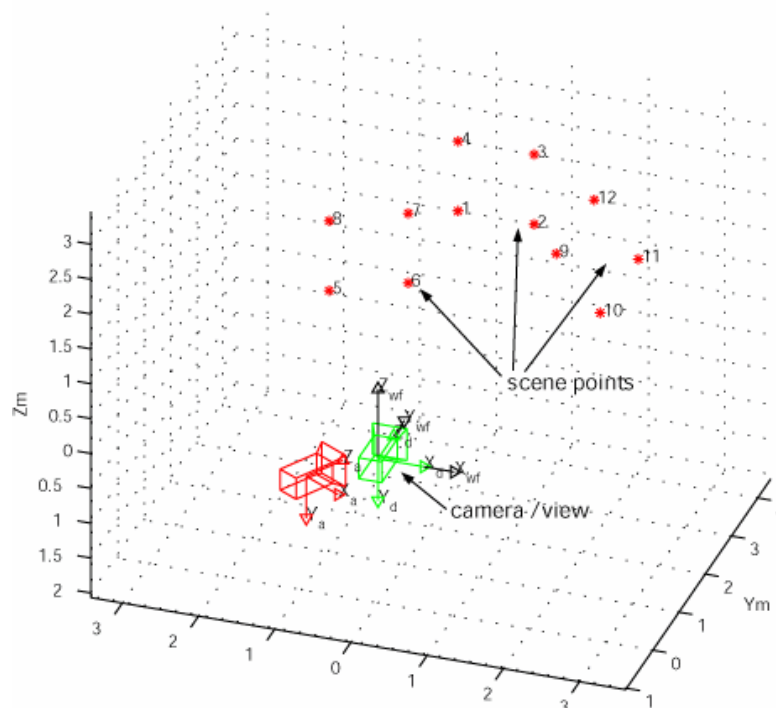
3.2.1 Epipolar Geometry Toolbox Project

Dieses Projekt der Universität Siena bietet eine Umgebung in MATLAB, mit der diverse Verarbeitungen von mehrfach gefilmten Szenen möglich sind. (Mariottini & Prattichizzo, 2006)

Es sind dabei Funktionen zur Verarbeitung und Visualisierung von dreidimensionalen Daten enthalten. Auch sind mehrere Algorithmen zur Abschätzung der Szenengeometrie enthalten. Funktionen zur Berechnung von Epipolarlinien und andere Funktionen der epipolaren Geometrie ermöglichen neben der Erstellung auch die Validierung von ermittelten Szenengeometrien.

Neben dem Umgang mit Lochkamerabildern – der typischen Modellierung von Bildern einer gewöhnlichen Kamera – enthält die Toolbox auch Funktionen zur Verarbeitung von Bildern aus Panoramakameras.

Während zwar die Verarbeitung von zwei- und dreidimensionalen Punkten möglich ist, bietet die Toolbox keine Möglichkeiten an, um bei bekannter Szenengeometrie aus zweidimensionalen Korrespondenzen dreidimensionale Raumpunkte zu berechnen.



Darstellung einer 3D-Szene durch die Epipolar Geometry Toolbox

3.2.2 Multiple View Geometry Toolbox

Die Visual Geometry Group der Universität Oxford hat – im Zusammenhang der Arbeit *Multiple View Geometry* (Hartley & Zisserman, 2003) – eine Toolbox mit Funktionen für MATLAB zusammengestellt. Diese Toolbox bietet diverse zwei- und dreidimensionale geometrische Funktionen, sowie Funktionen zur Verarbeitung von verschiedenen Parametern einer Szene (z.B. Projektionsmatrizen und intrinsische Parameter von Kameras).

Dabei ist auch die Möglichkeit enthalten, bei einer bekannten Szenengeometrie (bekannter Fundamentalmatrix) aus korrespondierenden, zweidimensionalen Bildpunkten die entsprechenden Raumpunkte zu bestimmen.

Desweiteren sind auch eigene Funktionen enthalten um die Szenengeometrie aus Korrespondenzen abzuschätzen. Dabei gibt es auch die Möglichkeit direkt die Projektion der Kameras sowie Raumpunkte aus Korrespondenzen von drei Bildern abzuleiten. Dieses Verfahren wird Trifokal-Tensor genannt und ermöglicht die direkte Bestimmung der Projektionsmatrizen der Kameras ohne eine Fundamentalmatrix¹ benutzen zu müssen.

3.2.3 BEEM

Das Balanced Exploration and Exploitation Model (BEEM) geht aus einer Projektarbeit an der Universität Haifa (Israel) hervor (Goshen & Shimshoni, 2006).

Bei BEEM handelt es sich um einen Algorithmus zur Abschätzung der Szenengeometrie. Das besondere an diesem Algorithmus ist, dass nicht versucht wird in einem Schritt unter Berücksichtigung möglichst vieler Korrespondenzen ein möglichst gutes Modell der Szenengeometrie zu erstellen. Stattdessen wird mit einem bewusst schwachen Modell angefangen.

Der Algorithmus geht von SIFT-(artigen-)Korrespondenzen aus und bildet zunächst aus zwei Korrespondenzen² ein schwaches Modell der Szenengeometrie, das danach in mehreren Schritten verbessert wird. Hierbei ist es möglich die Genauigkeit gegenüber dem Aufwand zu skalieren.

¹Bei vielen Verfahren wird zuerst aus korrespondierenden Bildpunkten die Fundamentalmatrix hergeleitet, die die Geometrie der Szene abbildet. Danach wird anhand der Fundamentalmatrix die Projektionsmatrix errechnet, die die zweidimensionale Abbildung der Kamera modelliert. (Anhand dieser können dann aus Bildpunkten die Raumpunkte bestimmt werden.)

² Der Algorithmus schätzt eine Fundamentalmatrix, die gewöhnlich mit mindestens acht Korrespondenzen geschätzt wird, mit zwei SIFT-Korrespondenzen ab, indem mittels der Orientierung der Deskriptoren aus einer Korrespondenz und deren Deskriptoren vier Korrespondenzen geschätzt werden.

3.3 Triangulation und Meshing

Wurden dreidimensionale Raumpunkte gefunden, sollen diese durch Triangulation in ein 3D-Modell überführt werden. Dieser Abschnitt befasst sich mit Projekten in denen Algorithmen und Verfahren zur Triangulation von dreidimensionalen Punkten verwendet werden. Das Erstellen eines 3D-Modells wird auch Meshing³ genannt.

3.3.1 Meshlab

Meshlab ist ein Programm, das in einem Projekt an der Universität Pisa im Jahr 2005 entstand. Es ist dafür ausgelegt Daten von 3D-Scannern in dreidimensionale Polygonmodelle zu überführen.

Das Programm bietet Funktionen um dreidimensionale Punktwolken in Meshes umzuformen. Zusätzlich sind auch Funktionen und Filter enthalten, mit denen 3D-Modelle nachverarbeitet und verbessert werden können.

Dabei sind auch Säuberungsfilter enthalten, mit denen erstellte Polygonmodelle gesäubert werden können. Dabei wird die Anzahl der Polygone so reduziert, dass die Reduzierung möglichst wenig Einfluss auf das optische Erscheinungsbild des Objektes hat.

Meshlab ist keine dedizierte Bibliothek, sondern ein einzelstehendes Programm. Da es allerdings Quelloffen ist, ist es dennoch möglich einzelne Funktionen daraus zu verwenden.

3.3.2 Gyroviz

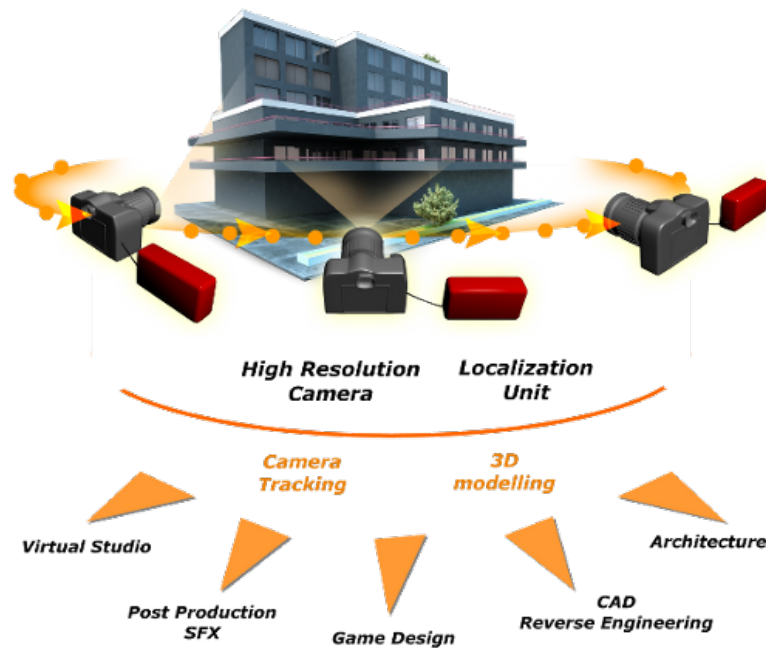
Gyroviz ist ein französisches Forschungsprojekt, das noch bis Ende 2010 läuft. Es startete im Dezember 2007 und hat das gesetzte Ziel eine automatische Modellierung von 3D-Szenen zu bieten, ohne dass eine Benutzerinteraktion nötig ist. Dabei sollen bisherige Umsetzungen im Bereich der Multiple View Geometry in Robustheit und Effizienz verbessert werden.

Bei dem Ansatz von Gyroviz wird allerdings von lokalisierten Kameras ausgegangen. Dieses bedeutet in der angestrebten Umsetzung von Gyroviz, dass Kameras mit Lokalisierungseinheiten verbunden sind, die Gyroskope und Beschleunigungssensoren enthalten um die Position und Ausrichtung der Kamera bei der Aufnahme des Bildes messen.

Dieser Ansatz von Gyroviz ist zwar für die angestrebten Einsatzziele dieser Arbeit nicht von Belang, allerdings werden auch im Gyroviz-Projekt Verfahren verwendet um die dadurch gemessenen Punkte in Polygone umzuwandeln.

Bei Gyroviz wird desweiteren auch eine Texturierung der erstellten Polygone angestrebt. (Auch wenn dieses vorerst hier nicht Gegenstand des Interesses ist.)

³ Mesh (engl. Netz) bezeichnet das Gitternetz eines dreidimensionalen Modells.



Aufnahme einer Szene bei Gyroviz: die Kameras sind mit Lokalisierungseinheiten verbunden

3.3.3 CGAL

Die Computational Geometry Algorithms Library ist eine Bibliothek die mehrere Algorithmen zur Verarbeitung und zur Erstellung dreidimensionaler Objekte beinhaltet.

Es handelt sich dabei um ein Projekt mehrerer europäischer Firmen und Hochschulen (sowie der Universität Tel Aviv), das seit 1996 besteht und seit dem stetig erweitert wird.

CGAL ist eine Open Source Bibliothek, die darauf abzielt möglichst effiziente und robuste Algorithmen zur 3D-Verarbeitung zu bieten.

Beispielsweise sind Algorithmen enthalten, um durch Triangulation aus drei oder mehreren Punkten⁴ ein dreidimensionales Dreieck zu erstellen.

Desweiteren sind Funktionen enthalten um für eine Menge von Raumpunkten eine konvexe Hülle zu berechnen. Solche Algorithmen liefern eine dreidimensionale (und möglichst kleine) Oberfläche für eine übergebene Menge von Punkten.

Respektive der geschilderten Anwendung könnte allerdings nur ein Teil der Punkte an einen solchen Algorithmus übergeben werden, da in der Regel eine gesamte Oberfläche für alle Punkte berechnet wird, und dadurch ein einzelnes Objekt resultiert, das die Oberfläche der Objekte in der Szene überschreitet.

Desweiteren bietet die Bibliothek Algorithmen zur Filterung und Weiterverarbeitung von Polygonen und Meshes.

Einige der Funktionen der CGAL Bibliothek sind auch in MATLAB verfügbar.

⁴ Vgl. 2.2

4 Fazit

In dieser Ausarbeitung wurden Projekte vorgestellt, die verschiedene Möglichkeiten bieten, die die Problemstellungen behandeln, die eingangs beschrieben wurden.

Die erste Projektumsetzung zielt darauf ab, herauszufinden ob und inwiefern sich die verschiedenen Schritte mithilfe der Algorithmen der vorgestellten Projekte in einem Projekt zusammenfassen lassen.

Eine erste Umsetzung wird in MATLAB durchgeführt, deswegen werden an dieser Stelle Verfahren bevorzugt, deren Algorithmen für MATLAB verfügbar sind.

4.1 Merkmale und Korrespondenzen

Als wichtige Eigenschaften bei Verfahren zum Auffinden und der Beschreibung von Merkmalen werden hier Geschwindigkeit und Einfachheit gewählt.

Sollen die Bilder einer Kamera analysiert werden, muss eine große Menge von Bildern verarbeitet werden. Deswegen sollte die Geschwindigkeit des Verfahrens eine Rolle spielen. Die Einfachheit des Verfahrens (beispielsweise auch die Einfachheit des Deskriptors) reduziert den Aufwand der Umsetzung und trägt damit auch zur Geschwindigkeit bei.

Als das am besten geeignete Verfahren wird an dieser Stelle SURF gewählt, da das Verfahren eine einfache und schnelle Merkmalerkennung und -beschreibung bietet. Zudem ist eine Implementierung des Verfahrens für MATLAB verfügbar.

4.2 Szenengeometrie und 3D-Punkte

Zur Ermittlung der Szenengeometrie und der Extraktion der Raumpunkte wird versucht werden die Epipolar Geometry Toolbox und die Multiple View Geometry Toolbox zu kombinieren. Dabei soll er die Aufgabe der EGT sein, die Szenengeometrie abzuschätzen, während die MVG-Toolbox anhand der Szenengeometrie die Raumpunkte berechnen soll.

BEEM als Alternative für die Abschätzung der Szenengeometrie wird hier nicht ausgewählt, da sie (im Gegensatz zu den anderen beiden Bibliotheken) bisher nicht für MATLAB verfügbar ist. Zudem ist es nicht sicher, wie sich die Expansion von SIFT-Merkmalen verhält, wenn andere Marker (zum Beispiel SURF) benutzt werden.

4.3 Triangulation und Meshing

Für die Triangulation wird die CGAL Bibliothek als am geeignetsten angesehen. Sie wird seit längerer Zeit weiterentwickelt und verspricht deswegen Korrektheit und Robustheit. Die Algorithmen der Bibliothek sind zudem für MATLAB verfügbar.

Die Entwicklung von Gyroviz ist noch nicht abgeschlossen, weswegen die Verwendbarkeit nicht abschließend bewertet werden kann. Auch ist dementsprechend (zumindest bisher) noch keine Implementierung der Algorithmen für MATLAB verfügbar.

Meshlab enthält viele nützliche Funktionen, allerdings sind diese nicht für MATLAB erhältlich und müssten zuvor aus den Quellen von Meshlab extrahiert werden, da die Funktionen von Meshlab nicht in dedizierten Bibliotheken verfügbar sind.

5 Referenzen

Rusitska, N. (2010, Februar). Multiple View Geometry 3D-Umgebungsmodell anhand einer bewegten Kamera. Hamburg.

Bay, H., Ess, A., Tuytelaars, T., & Gool, L. V. (2008, September 10). Speeded-Up Robust Features (SURF). Zurich, Leuven.

Sarfraz, M. S., & Hellwich, O. (2008). Head Pose Estimation In Face Recognition Across Pose Scenarios. *International Conference on Computer Vision Theory and Applications*, (pp. 235-242). Berlin.

Thormählen, T., & Seidel, H.-P. (2008). 3D-Modeling by Ortho-Image Generation from Image Sequences. In *ACM Transactions on Graphics, Vol. 27, No. 3, Article 86*.

Wan, C., & Sato, J. (2008). Computing Multiple View Geometry in Space-Time from Mutual Projections of Multiple Cameras. *19th International Conference on Pattern Recognition (ICPR) 2008*, (pp. 1-4).

Goshen, L., & Shimshoni, I. (2006). Balanced Exploration and Exploitation Model Search for Efficient Epipolar Geometry Estimations (BEEM). In *9th European Conference on Computer Vision (ECCV 2006)* (pp. 151-164).

Mariottini, G. L., & Prattichizzo, D. (2006). The Epipolar Geometry Toolbox: multiple view geometry and visual servoing for MATLAB. Siena, Italien.

Gaßmann, F., & Franke, K.-H. (2005, Oktober 27). Verfahren zur Triangulation ungeordneter 3D-Punktmengen. Ilmenau.

Mikolajczyk, K., & Schmid, C. (2005, Februar 23). A performance evaluation of local descriptors. Oxford, Montbonnot.

Heyden, A., & Pollefeys, M. (2004). Multiple View Geometry. In G. Medioni, & S. B. Kang, *Emerging Topics in Computer Vision* (pp. 45-107). Prentice Hall.

Košecká, J. (2004). Multiple-View Geometry for Image-Based Modeling. *Siggraph Course #23*. Los Angeles.

Lowe, D. (2004). Distinctive Image Features from Scale-Invariant Keypoints. In *International Journal of Computer Vision* 60 (pp. 91-110).

Hartley, R., & Zisserman, A. (2003). *Multiple View Geometry in Computer Vision* (2. Ausg.). Cambridge: Cambridge University Press.