Rekonstruktion eben begrenzter Räume aus monokularen Aufnahmen zum Zwecke des berührungslosen Aufmessens

Hempel, Lorenz; Meuser, Klaus; Schmiedel, Roland
Bauhaus-Universität Weimar

Bei der Rekonstruktion von Gebäuden ist es notwendig, bestehende Gebäude in ihrer Struktur und Geometrie zu erfassen. Dabei werden vielfach photogrammetrische Verfahren als eine indirekte Technik zur Erzeugung von 3D-Daten aus fotografischen Aufnahmen eingesetzt. Ein Vorteil ist, daß die geometrischen Daten in den Aufnahmen gespeichert sind und somit sowohl zeitlich, als auch räumlich unabhängig vom Gebäude ausgewertet werden können. Häufig werden dabei hochpräzise arbeitende und damit teuere Geräte eingesetzt, die das Gebäude mit einer hohen. aber für die Aufgabenstellung des Anwenders nicht notwendigen Genauigkeit aufmessen.

Ziel des hier vorgestellten Projektes ist es deshalb, die mathematischen Grundlagen für ein berührungsfreies kostengünstiges Aufmaßsystem zu schaffen, mit dessen Hilfe der Architekt bzw. Bauingenieur in die Lage versetzt wird, mit vergleichsweise geringem gerätetechnischen Aufwand ein Gebäude in einer für ihn hinreichenden Genauigkeit zu rekonstruieren. Traditionelle Vorgehensweise in der Photogrammetrie ist, mit hochgenauen Kameras ein Objekt mit Hilfe exakt vermessener Paßpunkte zu erfassen. Aus den bekannten Koordinaten der Paßpunkte auf dem Objekt und den zugehörigen Bildkoordinaten werden entweder die physikalischen Kameraparameter direkt bestimmt (explizite Kalibrierung) oder aber die zentralprojektiven Parameter (direkte lineare Methode) als Zwischenparameter, aus denen sich in einem zweiten Schritt die Kameraparameter ermitteln lassen (implizite Kalibrierung).

Bei der im Projekt gewählten Vorgehensweise liegt im Gegensatz dazu folgende Situation vor:

- Es werden keine hochgenauen Kameras verwendet. Bei Innenraumaufnahmen werden aufgrund der räumlichen Begrenztheit Kameras mit kurzbrennweitigen Linsensystemen mit entsprechenden Abbildungsfehlern dominieren.
- Als Paßpunkte werden Maßfiguren verwendet, die mit Hilfe von Laserspotstrahlern auf der Objektebene erzeugt werden. Die Genauigkeit der Maßfigur wird durch die Genauigkeit der eigentlichen Laserprojektion, die physikalischen Eigenschaften der Laserspotstrahler und die Lokalisierbarkeit der Laserpunkte auf der Aufnahme beeinflußt. Die Präzision der Paßpunkte wird relativ niedrig sein.
- Die Lage der Paßpunkte in der Objektebene wird nicht absolut, sondern nur relativ zueinander in Abhängigkeit gewisser unbekannter Parameter gegeben sein. Bei gerätetechnischer Kopplung des Laserstrahlprojektors mit der Kamera sind diese Parameter mit den Kameraparametern gekoppelt, andernfalls unabhängig von diesen.
- Weitere Verfälschungen entstehen bei der Digitalisierung der Aufnahmen (digitale Aufnahme oder Scannen von Fotografien).

Die 3D-Rekonstruktion eines polyedrischen Raumes aus monokularen Aufnahmen erfolgt in zwei Schritten. Im ersten Schritt werden aus den monokularen Aufnahmen der den Raum begrenzenden Ebenen maßstabsgerechte Orthobilder erzeugt. Diese werden dann in einem zweiten Schritt zu einem Modell des Raumes zusammengesetzt.

1. Rekonstruktion ebener Objekte aus monokularen Aufnahmen
1.1. Projektion von Paßpunkten mittels Laserspotstrahler

Laserpunkte, wenngleich auf der Objektwand als scheinbar scharf umgrenztes, hinreichend kleines Gebilde erkennbar, erscheinen auf einer Fotografie durchaus nicht als ebenfalls scharf begrenztes Gebilde, dessen Lokalisierung ohne weiteres möglich ist. Um ihrer Aufgabe als Paßpunkte gerecht zu werden, ist es jedoch notwendig, die Laserpunkte im digitalisierten Bild mindestens mit Pixelgenauigkeit, möglichst noch darunter zu lokalisieren. Die Lokalisierbarkeit der Laserpunkte wird beeinflußt durch Lichtbeugung der Laserstrahlen, nicht radialsymmetrische Abstrahlung der Laserstrahlen, Aufweitung des Laserstrahls mit zunehmender Entfernung der Objektwand und die Existenz starker Streulichtumgebungen um die Laserpunkte. In der Wechselwirkung dieser Lasereigenschaften mit den Eigenschaften von Farbfilmen, einschließlich dessen Entwicklung und Digitalisierung des Fotos können erhebliche Unsicherheiten auftreten, die Abweichungen von einigen Millimetern bei der Lokalisierung der Paßpunkte auf der Objektwand entsprechen.

Neben der geometrischen Lokalisierbarkeit der Laserpunkte steht die Frage der automatischen Punkterkennung auf der digitalisierten Aufnahme. Diese hängt unter anderem stark von den Eigenschaften der Objektoberfläche ab, d.h. von Material, Oberflächenstrukturierung, Farbe, Beleuchtung usw.

Diese Einflüsse in ihrer Gesamtheit führen dazu, daß alle Tests auf hypothetische zweidimensionale Verteilungen von Farbwerten, Farbintensitäten oder ähnlichem bei der automatischen Laserpunkterkennung abgelehnt werden. Die Punkterkennung erfolgt deshalb ausschließlich über die Ermittlung des Flächenschwerpunktes (Dichtemittels), wobei dieser wiederum von den verwendeten Farbwerten (Grauwert, Farbanteil) und den gewählten Farbwertschwellen abhängt.

1.2. Mathematisches Modell

Zur Beschreibung der Abbildung einer im dreidimensionalen Raum liegenden Ebene (Objektebene) auf die zweidimensionale Bildebene einer Kamera dient das physikalische Modell einer Lochkamera, das heißt eine Idealisierung der tatsächlichen Abbildungsverhältnisse mittels Linse bzw. Linsensystem. Bei der Lochkamera wird die Abbildung der Objektebene in die Bildebene durch ein Geradenbündel (Lichtstrahlenbündel) vermittelt, dessen Träger, d.h. der allen Projektionsgeraden gemeinsame Punkt, das unendlich kleine Loch der Kamera ist. Die Abbildung läßt sich somit als eine Zentralprojektion eines Punktes im dreidimensionalen Raum auf die Bildebene beschreiben. Im Realfall einer Linse oder eines Linsensystems erfolgt die Abbildung nach der gleichen Gesetzmäßigkeit, allerdings sind dann die durch das Objektiv der verwendeten Kamera festgelegten Abbildungsfehler zu berücksichtigen. Im folgenden werden wir zunächst von solchen Linsenfehlern absehen, eine Begründung für diese Vorgehensweise werden wir später nachholen.

Die Zentralprojektion eines ebenen Objektes auf die Bildebene läßt sich mathematisch als Spezialfall einer projektiven Transformation modellieren. Die projektive Transformation eines Punktes (x, y) der Objektebene in einen Punkt (x', y') der Bildebene wird in diesem Falle durch die Gleichungen

$$x' = \frac{a_{10} + a_{11}x + a_{12}y}{1 + a_{31}x + a_{32}y}, \quad y' = \frac{a_{20} + a_{21}x + a_{22}y}{1 + a_{31}x + a_{32}y} \tag{1}$$

beschrieben, wobei sich für alle Punkte (x, y) des abzubildenden Objektes das Vorzeichen des Nenners nicht ändert. Im weiteren werden wir unterscheiden zwischen Ähnlichkeits-transformationen (Translationen, Skalierungen, Rotationen), affinen Transformationen (hinzu kommen anisotrope Skalierungen, Scherungen) und projektiven Transformationen.

Affine Transformationen sind lineare Abbildungen, d.h. in (1) gilt $a_{31} = a_{32} = 0$. Projektive Transformationen im eigentlichen Sinne besitzen diese Eigenschaft nicht. Allgemeine projektive Transformationen werden zweckmäßigerweise in homogenen Koordinaten beschrieben. Sind $(\hat{x}, \hat{y})$ die kartesischen Koordinaten eines Ebenenpunktes, so heißen die durch die Gleichungen $\hat{x} = \frac{x}{t}, \quad \hat{y} = \frac{y}{t}$ zugeordneten Werte (x, y, t) die homogenen Koordinaten des Ebenenpunktes, der zugehörige Raum heißt zweidimensionaler projektiver Raum $\mathbf{R}P^2$. Die Transformationsgleichung in homogenen Koordinaten lautet

$$\begin{pmatrix} x' \\ y' \\ t' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{10} \\ a_{21} & a_{22} & a_{20} \\ a_{31} & a_{32} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ t \end{pmatrix}. \tag{2}$$

Eine projektive Transformation kann somit als eine lineare Abbildung im $\mathbf{R}P^2$ beschrieben werden. Durch die acht Parameter in der Abbildungsmatrix ist eine projektive Transformation eindeutig gegeben.

Die projektiven Transformationen bilden eine Gruppe im mathematischen Sinne, das Hintereinanderausführen als Gruppenoperation wird durch eine Matrixmultiplikation beschrieben. Von Interesse ist die Frage nach der Möglichkeit zur Separation einer Zentralprojektion, d.h. ihrer Zerlegung in einfache Transformationen. Es läßt sich zeigen, daß sich jede Zentralprojektion als Hintereinanderausführung einer rein projektiven und einer affinen Transformation schreiben läßt. Jede affine Transformation ist ihrerseits eine Verknüpfung von einer homogenen affinen Transformation und einer Translation. Eine homogene affine Transformation läßt sich schließlich als Verknüpfung einer Scherung, einer anisotropen Skalierung und einer Rotation darstellen. Ist **A** die Abbildungsmatrix in (2), so gilt also $\mathbf{A} = \mathbf{R} \cdot \mathbf{S} \cdot \mathbf{W} \cdot \mathbf{T} \cdot \mathbf{P}$ mit

$$\begin{aligned}
\mathbf{P} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_{31} & a_{32} & 1 \end{pmatrix} && \text{(reine projektive Transformation)} \\
\mathbf{T} &= \begin{pmatrix} 1 & 0 & a_{10} \\ 0 & 1 & a_{20} \\ 0 & 0 & 1 \end{pmatrix} && \text{(Translation)} \\
\mathbf{W} &= \begin{pmatrix} 1 & a_{12} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} && \text{(Scherung)} \\
\mathbf{S} &= \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} && \text{(anisotrope Skalierung)}
\end{aligned} \tag{3}$$

$$\mathbf{R} = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{(Rotation)}$$

Unter der Voraussetzung $a_{11} > 0$ ist diese Separation eindeutig bestimmt

1.3. Rekonstruktion mittels mathematischer Optimierung

Auf der Objektebene ist eine Maßfigur aus r Paßpunkten mittels Laserprojektion festgelegt. Durch vier exakt gegebene Paßpunkte, von denen keine drei Punkte kollinear sind, ist eine projektive Abbildung eindeutig bestimmt. Theoretisch kann die Projektionsmatrix durch Lösung eines Gleichungssystems gewonnen werden. Wegen der Ungenauigkeiten bei der Ermittlung der Eingangsdaten und der problembedingten Auswirkung dieser Eingangsfehler auf die Resultate können keine brauchbaren Resultate erzielt werden. Aus diesem Grunde werden mehr als vier Maßpunkte verwendet, die darüber hinaus nur in Abhängigkeit von Parametern bekannt sind.

Sei $X^* = \{\mathbf{x}^i(d_1, d_2, \cdots, d_p),\quad i = 1, \cdots, r\}$ die Menge von r Paßpunkten in der Objektebene E in Abhängigkeit von p Parametern. Die in der Aufnahme ermittelten zugehörigen Bildpunkte seien $Y^* = \{\mathbf{y}^i, i = 1, \cdots, r\}$. Aus verfahrenstechnischen Gründen bestimmen wir eine durch die Matrix **A** festgelegte Abbildung $\varphi_A : \mathbf{R}P^2 \to \mathbf{R}P^2$ mit $\varphi_A(Y^*) = X^*$. Als Umkehrabbildung einer projektiven Transformation ist φ_A ebenfalls eine projektive Transformation. Die Bestimmung der Matrix **A** erfolgt über das nichtlineare, restriktionsfreie Optimierungsproblem

$$z(\mathbf{A}, \mathbf{d}) = \sum_{\mathbf{y}^i \in Y^*} \left\| \mathbf{x}^i(\mathbf{d}) - \varphi_A(\mathbf{y}^i) \right\|^2 \quad \to \quad \min \tag{4}$$

Wie bereits erwähnt, hängt die Genauigkeit der Matrix **A** und der mit ihrer Hilfe rekonstruierten Objektebene wesentlich von der Lokalisierbarkeit der Laserpunkte ab. Insbesondere der rein projektive Anteil der Transformation wird von diesen Ungenauigkeiten beeinflußt. Gebäudeaufnahmen enthalten in der Regel eine Reihe geometrischer Informationen, die wesentlich zur Verbesserung der Rekonstruierbarkeit von Objektebenen aus monokularen Aufnahmen beitragen können. Besondere Bedeutung haben dabei Parallelität und Orthogonalität von Kanten. Parallelen werden durch eine projektive Transformation in ein Geradenbündel durch einen gemeinsamen Punkt, den Fluchtpunkt, abgebildet. Enthält eine Aufnahme das Bild zweier verschiedener Parallelenscharen, so lassen sich daraus zwei Fluchtpunkte und damit die Fluchtgerade ermitteln. Mit Hilfe der Fluchtgeraden als Nenner der projektiven Abbildung (1) kann die Projektion eindeutig in eine reine projektive Transformation mit diesem Nenner und eine affine Transformation separiert werden. Zur Bestimmung der Matrix P des Separationsansatzes können somit Verfahren der automatischen Kantenerkennung eingesetzt werden.

Diese Vorgehensweise hat darüber hinaus den Vorteil, daß die gefundenen Geraden als Grundlage für ein Verfahren dienen, mit dessen Hilfe optische Linsenfehler in Form von radialsymmetrischen Verzerrungen erkannt und ausgeglichen werden können. Durch

Vorschaltung dieses Verfahrens entstehen hinreichend verzeichnungsfreie Aufnahmen, auf die das oben erwähnte Lochkameramodell anwendbar ist.

Durch Kopplung der rein projektiven Transformation mit einer Translation, die sich aus der Festlegung der Objekt- und Bildkoordinatensysteme ergibt, ist die Abbildung bis auf eine homogene affine Transformation eindeutig bestimmt. Diese kann ihrerseits auf zwei Wegen bestimmt werden. Zum einen kann sie wieder über ein Optimierungsproblem (4) bestimmt werden. In diesem Fall ist die Zielfunktion linear in den Elementen der Transformationsmatrix **A** und der Einfluß der Ungenauigkeiten bei der Laserpunktbestimmung ist wesentlich geringer. Eine zweite Möglichkeit, die noch günstigere Werte liefert, setzt voraus, daß die beiden Parallelenscharen senkrecht aufeinander stehen oder zumindest zwei senkrecht zueinander stehende Geraden erkennbar sind. Wiederum mit der im Vergleich zur Laserpunkterkennung genauer arbeitenden automatischen Kantenerkennung kann daraus die unter (3) genannte Scherungsmatrix berechnet werden. Die Laserpunkte werden in diesem Falle nur noch für die anisotrope Skalierung, d.h. die Festlegung des Verhältnisses von Höhe zu Breite, und für die endgültige Festlegung des Maßstabes benötigt.

2. Rekonstruktion von eben begrenzten Räumen

2.1. Rekonstruktion von Ebenen aus Einzelaufnahmen

Die Zusammensetzung von eben begrenzten Räumen aus Einzelebenen (Wänden) erfolgt in zwei Schritten. Zunächst ist in einem ersten Schritt jede der den Raum begrenzenden Ebenen komplett aus Einzelaufnahmen zu rekonstruieren. Das ist notwendig, da in der Praxis eine Wand in der Regel nicht durch eine einzige Aufnahme komplett abgebildet werden kann. Die Erfassung der Wand erfolgt vielmehr durch eine Reihe von Einzelaufnahmen von unterschiedlichen Standorten aus. Die Lagebeziehungen der Kamerastandorte untereinander sind unbekannt, die Kopplung der einzelnen Aufnahmen ist nur über Bildinhalte und über die in mehreren Aufnahmen enthaltene unveränderte laserprojizierte Maßfigur möglich. Liegen die zusammenzusetzenden Aufnahmen nach der im 1. Abschnitt beschriebenen Rekonstruktion in einem maßstabsgerechten Abbild vor, so kann die Zusammensetzung über Paßpunkte (Bildinhalte) oder mittels Pixelmatching erfolgen. Ergebnis ist ein ebenfalls maßstabsgerechtes Abbild der Originalwand. In der Regel werden aber die zusammenzusetzenden Aufnahmen nur fehlerbehaftet vorliegen. Insbesondere handelt es sich bei diesen Abbildungsfehlern im wesentlichen um affine, zum Teil auch projektive Verzerrungen. Aus diesem Grunde wird auch die zusammengesetzte Gesamtwand kein maßstabsgerechtes Abbild der Objektwand sein, sondern nur fehlerbehaftet vorliegen.

2.2. Rekonstruktion von Räumen aus Einzelwänden

Liegen die einen Raum begrenzenden Wände als fehlerbehaftete orthogonale Draufsichten vor, sind diese in einem zweiten Schritt zum Raum zusammenzusetzen. Jedes konvexe Polyeder kann aus seinen Begrenzungsflächen maßstabsgerecht rekonstruiert werden, wenn von den Begrenzungsflächen maßstabsgerechte Abbilder, wenngleich mit unbekannten und zueinander unterschiedlichen Maßstabsfaktoren vorliegen. Liegen die Begrenzungsflächen nur fehlerhaft, also nicht maßstäblich vor, so existiert im allgemeinen kein konvexes Polyeder, das aus diesen maßstäblich vergrößerten bzw. verkleinerten Begrenzungsflächen besteht. Der Test, welcher der beiden Fälle vorliegt erfolgt über ein graphentheoretisches Modell, das das konvexe Polyeder (Raum) auf einen gerichteten Graphen G(V,E) abbildet, wobei die Knotenmenge V den Flächen des Raumes entspricht und eine Kante e_{ik} von V_i

nach V_k genau dann in E enthalten ist, wenn die Flächen V_i und V_k eine gemeinsame Raumkante besitzen. Jede Kante $e_{ik} \in E$ wird bewertet mit einem α_{ik}, das das Verhältnis der Längen der gemeinsamen Raumkante in den Bildern von V_i und V_k darstellt. Die These "Alle Wände liegen in maßstäblicher Form vor" wird abgelehnt, wenn es in diesem Graphen mindestens einen Kreis gibt, für den das Produkt der zugehörigen Bewertungen nicht Eins ist.

Wird die These nicht abgelehnt, besteht die Möglichkeit, den Raum auf direktem Wege zusammenzusetzen. Andernfalls erfolgt die Zusammensetzung über ein Optimierungs-problem. Zu diesem Zweck wird jede Wandfläche in Abhängigkeit von deren räumlichen Parametern und dem Maßstabsparameter in einem gemeinsamen räumlichen Objektmodell erfaßt. Die Koordinaten der Eckpunkte der Wandflächen können in Abhängigkeit von den Parametern bestimmt werden. Jede Raumecke als Schnittpunkt von drei (oder mehr) Wandebenen wird auf diese Weise durch drei (oder mehr) Wandecken repräsentiert, deren Lage von den für die zugehörigen Wände verwendeten Parametern abhängen und damit nicht identisch sind. Die Raumecke selbst wird als Mittelwert dieser Wandecken definiert, als Maß für ihre Güte wird die Summe der Abstandsquadrate über alle Paare der zugehörigen Wandecken verwendet. Das mathematische Optimierungsproblem besteht nun darin, die Summe dieser Gütefunktionen über alle Raumecken zu minimieren.

Dieses Projekt wurde im Rahmen eines Verbundprojektes vom Thüringer Ministerium für Wissenschaft, Forschung und Kultur gefördert.