Vereinfachte 3D-Rekonstruktion aus Sequenzen unkalibrierter Bilder

Sergey Cheremukhin

FernUniversität in Hagen
Mensch-Computer-Interaktion
Sergey.Cheremukhin@FernUni-Hagen.de

Art der Arbeit: Studienarbeit
Betreuer der Arbeit: Dr. Klaus Häming, Prof Dr. Gabriele Peters

Abstract: Diese Arbeit stellt einen neuen Ansatz vor, eine Rekonstruktion von 3D-Geometrie aus einfachen Fotos durchzuführen. Diese Fotos wurden von Kameras aufgenommen, deren interne Abbildungsparameter unbekannt sind (d.h. nicht kalibriert sind). Das Grundelement dieses Ansatzes sind Bildpaare, im Gegensatz zu den im unkalibrierten Fall sonst üblichen Bildtripel. Daraus ergibt sich eine Vereinfachung des Rekonstruktionsprozesses. Es wird gezeigt, dass bereits Korrespondenzen zwischen Bildpaaren ausreichen, um von diesen zunächst hinreichend genau auf die internen Kameraparameter zu schließen. Anschließend wird das Zusammenfügen der aus einer Vielzahl an Bildpaaren gewonnenen Teilrekonstruktionen beschrieben.

1 Einführung

Die Rekonstruktion von 3D-Geometrie aus einfachen Fotos verspricht einen stark vereinfachten Zugang zur Generierung von 3D-Modellen für jedefrau und jedermann. Die Aufnahmen können dabei von einfachen, unkalibrierten Kameras stammen.

Die grundlegende Vorgehensweise bei der Rekonstruktion macht sich die geometrischen Beziehungen zunutze, die zwischen Bildpunkten, 3D-Szene und den verwendeten Kameras gelten. Wie in [HZ03] beschrieben, lässt sich daraus ein Modell ableiten, dessen Parameter sich schrittweise bestimmen lassen. Eine konkrete Umsetzung dieser Vorgehensweise wurde bspw. in [PH10] vorgestellt. Diese Umsetzung diente als Ausgangspunkt für das hier beschriebene System.

Das verwendete Kameramodell ist eine projektive Kamera der Form $\mathbf{P} = \mathbf{KR}\left[\mathbf{I_{3x3}} | - \mathbf{C}\right]$. Die Matrix $\mathbf{K} := \begin{bmatrix} f & s & x_0 \\ & af & y_0 \\ & & 1 \end{bmatrix}$ ist dabei die sogenannte Kalibrierungsmatrix und stellt die internen Abbildungsparameter dar. Dabei ist f die Brennweite (focal length), a das Seitenverhältnis (aspect ratio), s ein Scherungsparameter, sowie $(x_0\, y_0)^T$ der Bildmittelpunkt. $\mathbf{R}$ ist eine 3×3 Drehmatrix und $\mathbf{C}$ ist das Kamerazentrum.

Die Eingabedaten sind lediglich Bildmerkmale, die wir mittels eines Harris-Merkmaldetektors [HS88] in einem Scale-Space-Rahmenwerk [Lin94] ermitteln. Um Beziehungen zwischen unterschiedlichen Bildern herzustellen, werden Merkmale, die das gleiche Szenenelement abbilden zu Korrespondenzen zusammengefasst. Dazu beschreiben wir die Merkmale mittels SIFT-Deskriptoren [Low04] und gleichen diese in einem kd-Baum mit Best-Bin-First-Heuristik [BL97] ab. Die dabei unvermeidbaren Falschzuordnungen werden in einer robusten Kombination aus LO-RANSAC und PROSAC [HP10] herausgefiltert.

2 Eigene Entwicklungen

Wir verfolgen den Ansatz, aus jedem Bildpaar eine metrische Rekonstruktion zu gewinnen, also die zugehörigen Kameras automatisch zu kalibrieren. Der übliche Ansatz wäre, nach der Berechnung von Bildpaaren zunächst zu Bildtripel [SSG+10] überzugehen. Wir sparen also einen Schritt im Rekonstruktionsprozess ein. Jedoch hat die Kalibrierung von Bildpaaren allein auf der Grundlage von Korrespondenzen im allgemeinen keine eindeutige Lösung. Insbesondere ergeben sich aus der rein mathematisch vorhandenen Möglichkeit, dass die Szenenstruktur, die zu einem abgebildeten Merkmal gehört, sowohl vor als auch hinter einer Kamera liegen kann, bereits vier mögliche Konstellationen.

Im ersten Schritt werden also Bildmerkmale in beiden Bildern eines Bildpaares gefunden und zu Korrespondenzen zusammengefasst. Anschließend werden fehlerhafte Korrespondenzen herausgefiltert, indem nur solche akzeptiert werden, die tatsächlich zu ein und derselben Kamerakonstellation gehören können. Eine solche Konstellation wird zu diesem Zeitpunkt durch eine Fundamentalmatrix beschrieben, die aus den Korrespondenzen berechnet werden kann.

Im zweiten Schritt wird der lineare Autokalibrierungsalgorithmus aus [PVG02] verwendet. Es werden zwei konkrete Kameramatrizen gewählt, die zu der Fundamentalmatrix aus dem ersten Schritt passen. Diese werden genutzt, um aus den Korrespondenzen vorläufige 3D-Punkte zu triangulieren. Danach wird eine symmetrische 4×4 Matrix $\mathbf{\Omega}^*$ mit Rang 3 gesucht, so dass die Gleichung $\mathbf{KK^T} \sim \mathbf{P}\mathbf{\Omega}^*\mathbf{P^T}$ erfüllt ist. Im Fall lediglich zweier Bilder ist die Lösung unterbestimmt, und der Lösungsraum wird durch $\mathbf{\Omega_a^*} + \gamma\mathbf{\Omega_b^*}$ aufgespannt. Um die Rang-3-Bedingung zu erfüllen, werden nun solche Werte für γ gesucht, die die Determinante von $\mathbf{\Omega_a^*} + \gamma\mathbf{\Omega_b^*}$ verschwinden lassen [PKv98]. Wir wählen dann aus max. 16 möglichen Lösungen (bis zu 4 Lösungen für γ für jede von 4 möglichen Kamerakonstellationen) die beste aus. Dazu wird aus aus den möglichen $\mathbf{\Omega}^*$ jeweils eine 3D-Homographie $\mathbf{H}$ errechnet. Die beste Lösung führt zu einem $\mathbf{H}$, bei dem sich die meisten transformierten 3D-Punkte $\mathbf{H^{-1}X_n}$ vor den transformierten Kameras $\mathbf{P_iH}$ befinden. Dieser Kalibrierungsalgorithmus schlägt fehl, wenn die Bildpunkt-Korrespondenzen nicht allgemein genug, also bspw. lediglich Abbildungen von Punkten ein und derselben Ebene im Raum sind, die Bildpunkt-Korrespondenzen zu viele Falsch-Zuordnungen enthalten oder die internen Abbildungsparameter der Kameras sich zu stark unterscheiden. In diesen Fällen wird versucht, ein Bild im betrachteten Bildpaar durch ein anderes passendes Bild zu ersetzen (sofern genug Bildpunkt-Korrespondenzen gefunden werden können) und er-

Abbildung 1: Rekonstruktionsbeispiel. Gezeigt ist eine Rekonstruktion aus einer Bildfolge, von der einige Bilder links abgebildet sind.

neut eine Rekonstruktion zu finden. Diese Versuche werden solange wiederholt, bis eine Kalibrierung gelingt oder bis keine weiteren passenden Bilder vorhanden sind.

Im dritten Schritt wird das zusammenfügen der Teilrekonstruktionen zu einer größeren Szene vorbereitet, indem der Einfluss der Kalibrierungsmatrizen entfernt wird. Dazu werden die rekonstruierten 3D-Punkte $\mathbf{X_n}$ mit einer geeigneten 3D-Homographie

$$\mathbf{T_i} := \left[\begin{array}{cc} \mathbf{R_i} & -\mathbf{R_i C_i} \\ \multicolumn{2}{c}{\mathbf{e_4}^{\mathbf{T}}} \end{array} \right]^{-1} \left[\begin{array}{c} \mathbf{P_i} \\ \mathbf{e_4}^{\mathbf{T}} \end{array} \right], \text{ mit } \mathbf{e_4}^{\mathbf{T}} = (0\,0\,0\,1)$$

transformiert, so dass diese die neue Position $\mathbf{T_i X_n}$ annehmen.

Basierend auf diesen transformierten Punkten werden immer zwei Teilrekonstruktionen betrachtet, die sich mindestens eine Ansicht teilen. Zudem müssen sich über diese gemeinsame Ansicht mindestens zwei Korrespondenzenpaare einander zuordnen lassen. Da jeder Korrespondenz durch die oben genannte Triangulierung ein 3D-Punkt zugeordnet ist, bedeutet dies, dass zwei 3D-Punkte gefunden worden sind, die jeweils den selben Punkt der betrachteten Szene repräsentieren. Gibt es mehr als zwei solcher Punktepaare, werden in einem PROSAC-Sampling die besten ausgewählt. Zu jedem 3D-Punktepaar wird ein weiterer, virtueller, Punkt hinzugefügt, so dass der Schwerpunkt dieser Punkttripel im Kamerazentrum liegt. Dadurch wird die später berechnete relative Drehung eine Drehung um das Kamerazentrum. Anschließend wird die relative Orientierung (Drehung, Translation und Skalierung) zwischen diesen zwei Punkttripel berechnet. Für die Berechnung wird der SVD-Algorithmus eingesetzt [AHB87]. Dieser Algorithmus erweist sich in empirischen Versuchen als stabil und wird auch in der Literatur [LEF95] als stabil angesehen.

Im letzten Schritt führen wir eine metrische Variante des Bundle-Adjustment-Algorithmus [TMHF00] aus, um die Rekonstruktion weiter zu optimieren. Wir verwenden hier eine kompakte Parametrisierung durch 7 Parameter für die Kameras (3 für die Drehung, 3 für die Verschiebung und 1 für die Brennweite) und 3 Parameter für die 3D-Punkte. Abbildung 1 zeigt das Ergebnis für eine konkrete Bildsequenz.

Das vorgestellte Verfahren zur 3D-Rekonstruktion eignet sich gut zur schnellen Berechnung von 3D-Geometrie aus Bilddaten. Das inkrementelle Zusammenfügen von kalibrierten Teilrekonstruktionen ermöglicht es zudem, bereits Zwischenergebnisse zu visualisieren sowie den Rekonstruktionsprozess zu verfolgen und interaktiv zu kontrollieren.

Literatur

[AHB87] K. S. Arun, T. S. Huang und S. D. Blostein. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:698–700, Mai 1987.

[BL97] Jeffrey S. Beis und David G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *In Proc. IEEE Conf. Comp. Vision Patt. Recog*, Seiten 1000–1006, 1997.

[HP10] Klaus Häming und Gabriele Peters. Structure-from-Motion Reconstruction Pipeline - A Survey with Focus on Short Image Sequences. *Kybernetika*, 46(5):926–937, 2010.

[HS88] Chris Harris und Mike Stephens. A Combined Corner and Edge Detector. In *4th ALVEY Vision Conference*, Seiten 147–151, 1988.

[HZ03] Richard I. Hartley und Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second. Auflage, 2003.

[LEF95] A. Lorusso, D. W. Eggert und R. B. Fisher. A comparison of four algorithms for estimating 3-D rigid transformations. In *Proceedings of the 1995 British conference on Machine vision (Vol. 1)*, BMVC '95, Seiten 237–246, Surrey, UK, UK, 1995. BMVA Press.

[Lin94] Tony Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2):224–270, 1994.

[Low04] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[PH10] Gabriele Peters und Klaus Häming. Fast Freehand Acquisition of 3D Objects and their Visualization. *Journal of Communication and Computer*, 7(2-3), 2010.

[PKv98] Marc Pollefeys, Reinhard Koch und Luc J. van Gool. Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters. In *ICCV*, Seiten 90–95, 1998.

[PVG02] Marc Pollefeys, Frank Verbiest und Luc Van Gool. Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery. In *Johansen (Eds.) Computer Vision - ECCV 2002, 7th European Conference on Computer Vision, Lecture Notes in Computer Science, Vol.2351*, Seiten 837–851. Springer-Verlag, 2002.

[SSG+10] Noah Snavely, Ian Simon, Michael Goesele, Richard Szeliski und Steven M Seitz. Scene Reconstruction and Visualization From Community Photo Collections. *Proceedings of the IEEE*, 98(8):1370–1390, 2010.

[TMHF00] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley und Andrew W. Fitzgibbon. Bundle Adjustment - A Modern Synthesis. In *ICCV '99: Proceedings of the International Workshop on Vision Algorithms*, Seiten 298–372, London, UK, 2000. Springer.