

Zuverlässiges Structure-From-Motion für Bildpaare

Sergey Cheremukhin

Einführung

Kameras sind allgegenwärtig, ob in Form ausgefeilter Spiegelreflexkameras, als Kompaktkameras oder integriert im Mobiltelefon. Diese weite Verbreitung führt dazu, dass eine unermessliche Zahl von Bildern aufgenommen wird. Insbesondere wägt ein Fotograf nicht mehr ab, ob es sich im gegebenen Moment wohl lohnen könnte, auf den Auslöser zu drücken. Stattdessen schießt dieser gleich mehrere Bilder desselben Motivs.

Mehrere Bilder desselben Motivs lassen aber auch Rückschlüsse auf dessen räumliche Struktur zu. Vereinfacht ausgedrückt, könnte man die vielen redundanten Aufnahmen dazu verwenden, das Motiv anstatt nur in zwei Dimensionen, in drei Dimensionen zu betrachten.

Prinzipiell ist es einfach, aus mehreren Bildern auf den Ort des aufgenommenen Motivs zu schließen. Wenn bekannt ist, wo die Kameras sich zum Zeitpunkt der Aufnahme befanden und in welche Richtung diese blickten, müssten sich die zum Motiv gehörenden Objektpunkte doch einfach triangulieren lassen.

Jedoch ist bei Schnappschüssen üblicherweise weder der Ort noch die Orientierung der Kameras bekannt. Neben diesen sogenannten extrinsischen Parametern müssen zudem noch die intrinsischen Parameter bekannt sein. Zu den intrinsischen Kameraparametern gehören zum Beispiel die Brennweite und ob der Bildsensor exakt mittig zur Bildachse liegt. Es muss also nicht nur die Struktur des Motivs trianguliert, sondern es müssen gleichzeitig auch die Kameraparameter geschätzt werden.

Wir orientieren uns zunächst an dem üblichen [6, 10, 13, 15] Ansatz, eine Lösung dieses

Problems umzusetzen, wenden uns ab Abschn. „Autokalibrierung aus zwei Bildern“ jedoch dem von uns entwickelten, vereinfachten Verfahren zu.

Das Kameramodell

Nahezu alle folgenden Ausführungen sind davon abhängig, wie das Kameramodell definiert wird. Wir verwenden eine lineare projektive Kamera, die im Folgenden kurz vorgestellt wird.

Generell unterscheidet man bei Kameras zwischen extrinsischen und intrinsischen Parametern. Die extrinsischen Parameter beschreiben die Position und Orientierung im Raum, während die intrinsischen Parameter die Eigenschaften der eigentlichen Abbildung auf die Bildebene modellieren. Mathematisch kann eine Kamera als 3×4 -Matrix P aufgefasst werden, die sich wie folgt faktorisieren lässt:

$$P = KR [I_{3 \times 3} | -C]. \quad (1)$$

Die extrinsischen Parameter werden dabei durch R und C modelliert, wobei R als Rotationsmatrix die Orientierung der Kamera und C als Vektor die Position im Raum beschreibt. $I_{3 \times 3}$ sei hier die 3×3 -Identitätsmatrix.

Die intrinsischen Parameter werden durch die obere Dreiecksmatrix K beschrieben. Diese hat die Form

DOI 10.1007/s00287-012-0637-0
© Springer-Verlag 2012

Sergey Cheremukhin
FernUniversität in Hagen, Fakultät für Mathematik
und Informatik, Lehrgebiet Mensch-Computer-Interaktion,
Universitätsstr. 1, 58097 Hagen
E-Mail: Sergey.Cheremukhin@FernUni-Hagen.de

Zusammenfassung

Diese Arbeit stellt einen neuen Ansatz vor, eine Rekonstruktion von 3D-Geometrie aus einfachen Fotos durchzuführen. Diese Fotos wurden von Kameras aufgenommen, von denen weder die relative Position und Blickrichtung im Raum, noch die internen Abbildungsparameter bekannt sind. Das Grundelement dieses Ansatzes sind Bildpaare, aus denen bereits eine metrische Rekonstruktion gewonnen wird. Daraus ergibt sich eine Vereinfachung des Rekonstruktionsprozesses, da bisher üblicherweise mindestens Bildtripel für diesen Schritt notwendig waren. Zudem wird gezeigt, wie die aus Bildpaaren gewonnenen Teilrekonstruktionen zu einer vollständigen Rekonstruktion, die alle Kamerapositionen gleichzeitig umfasst, zusammengefügt werden können.

$$\mathbf{K} = \begin{bmatrix} f & s & x_p \\ & af & y_p \\ & & 1 \end{bmatrix}. \quad (2)$$

Dabei ist f die Brennweite (focal length), a das Seitenverhältnis (aspect ratio), s ein Scherungsparameter (skew) sowie $(x_p, y_p)^T$ der Bildmittelpunkt (principal point). $\mathbf{K}$ wird auch Kalibrierungsmatrix genannt, da eine Kalibrierung üblicherweise genau die Elemente von $\mathbf{K}$ festlegt.

Beziehungen zwischen Bildern

Das Ziel ist also, sowohl die Geometrie des fotografierten Motivs in Form vom Objektpunkten im dreidimensionalen Raum als auch die beteiligten Kameras in Form ihrer Kameramatrizen zu bestimmen.

Der erste Schritt dahin besteht üblicherweise darin, die Aufnahmen derart zueinander in Beziehung zu setzen, dass einzelne Bildpunkte, die eine Projektion desselben Objektpunktes darstellen, als solche identifiziert werden. Abbildung 1 zeigt dies anhand eines konkreten Bildpaars. Der Prozess, der zu einer solchen Zuordnung führt, soll nun in aller Kürze dargestellt werden.

Zunächst werden interessante Punkte in den Bildern identifiziert. Interessant sind Punkte, wenn sie sich trotz eventueller perspektivischer Verzerrungen leicht in anderen Bildern wiederfinden lassen. Diese Aufgabe übernimmt üblicherweise ein Interest-Point-Detektor. Wir nutzen dazu den Detektor von Harris und Stephens [5] in einem Scale-Space-Rahmenwerk [7].

Die so gefundenen Bildpunkte werden dann mithilfe sie umgebenden Pixel beschrieben. Dies geschieht mit dem Ziel, die Bildpunkte zwischen Bildern abgleichen zu können, also herauszufinden, welche Bildpunkte eine Abbildung desselben Objektpunktes sind. Wir beschreiben die Merkmale mit SIFT-Deskriptoren [9]. Diese nutzen die Gradientenverteilung im Umfeld eines Bildpunktes. Ein Deskriptor ist dabei nichts weiter als ein Vektor von Zahlen.

Der Abgleich von Bildpunkten zwischen Bildern ist damit auf die Zuordnung dieser Zahlenvektoren

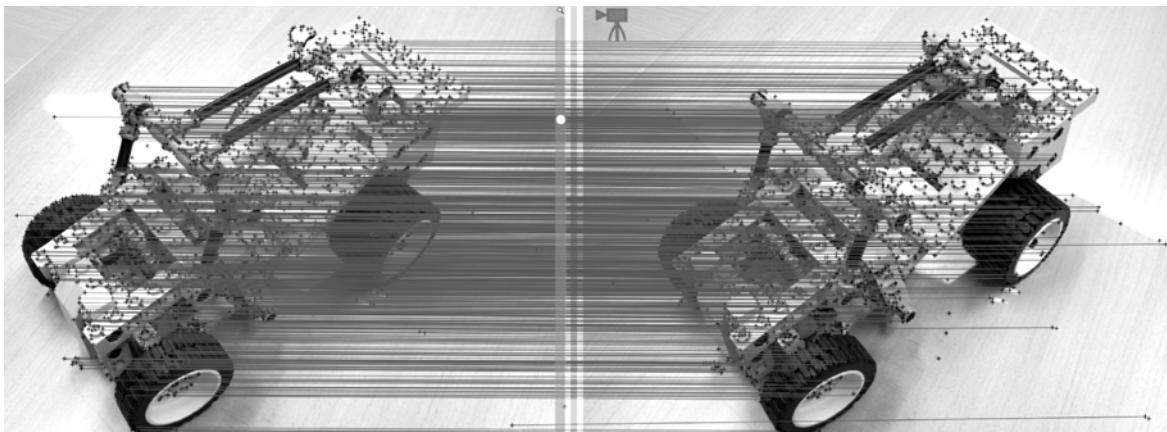


Abb. 1 Bildpunktkorrespondenzen in einem Bildpaar

Abstract

We present a novel approach towards a 3D-reconstruction from images which were taken with uncalibrated cameras. Here, uncalibrated means that there is no information available about external camera parameters (relative position and rotation) or about the internal ones (e. g. focal length). Our basic building block is the ability to create a metric reconstruction from image pairs. This contrasts the usual approach in which at least three images are needed. This results in a simplified reconstruction pipeline. We also show how to proceed from image pairs and subsequently create a reconstruction from longer image sequences.

ren reduziert. Wir nutzen dafür einen kd-Baum mit Best-Bin-First-Heuristik [2]. Die dabei unvermeidbaren Falschzuordnungen werden dabei mit einem RANSAC-Verfahren [3] herausgefiltert. Dabei handelt es sich um eine Schleife, in der in jedem Schritt aus einer Teilmenge der Daten eine Modellhypothese aufgestellt wird. Alle weiteren Daten werden anschließend daraufhin überprüft, ob sie diese Modellhypothese unterstützen. Die Idee ist, dass jenes Modell zu wählen ist, das die größte Menge an Unterstützern auf sich vereint. Konkret verwenden wir eine Kombination aus LO-RANSAC und PRO-SAC [4], zweier Weiterentwicklungen dieser Idee. Das Modell ist in diesem Fall durch eine Fundamentalmatrix gegeben, die weiter unten noch vorgestellt wird.

Am Ende dieses ersten Schrittes haben wir nun also einige Bildpunkte identifiziert, die Abbildungen desselben Objektpunkts darstellen. Seien nun für ein konkretes Bildpaar I und I' solch korrespondierende Bildpunkte mit x_i und x'_i bezeichnet. Dann ist für die Rekonstruktion der Kameras und Objektpunkte im Wesentlichen der folgende Fehler zu minimieren:

$$\varepsilon_r = \sum_{i=0}^n d^2(x_i, P X_i) + d^2(x'_i, P' X_i). \quad (3)$$

Die Distanzfunktion $d(\cdot, \cdot)$ berechnet dabei die euklidische Distanz in der Bildebene. Dieser sogenannte Reprojektionsfehler ε_r sagt also etwas darüber aus, wie exakt die rekonstruierten Kameras die re-

konstruierten Objektpunkte auf die gemessenen Bildpunktkorrespondenzen abbilden.

Den Reprojektionsfehler direkt zu minimieren führt zu einem nichtlinearen und damit iterativen Verfahren, dem Bundle Adjustment [14]. Dieses hängt von guten Startwerten sowohl für die Objektpunkte X_i als auch für die Kameras P und P' ab. Bei lediglich zwei Bildern ist ein direkter Lösungsversuch nicht Erfolg versprechend.

Berechnung vorläufiger Kameras

Wir nähern uns der Lösung an, indem wir zunächst vorläufige Kameras berechnen. Dabei hilft uns die geometrische Beziehung zwischen den Bildpunktkorrespondenzen, die sich daraus ergibt, dass diese Abbildungen desselben Objektpunkts sind. Die Beziehung wird durch die Fundamentalmatrix F ausgedrückt. Für einen Punkt x_i in Bild I bildet F diesen Punkt auf die Linie

$$l_i = F x_i \quad (4)$$

ab. Diese Linie hat die Eigenschaft, dass sie durch x'_i verläuft. Abbildung 2 zeigt grafisch, wie es zu diesem Zusammenhang kommt. Da außerdem $x_i^T l'_i = 0$, folgt

$$x_i^T F x'_i = 0. \quad (5)$$

Diesen Zusammenhang kann man nutzen, um ein Gleichungssystem aufzustellen und F aus den Bildpunktkorrespondenzen zu berechnen. Dazu reichen acht Bildpunktkorrespondenzen aus.

Ist F bekannt, so lässt sich unter dem Postulat, dass die erste Kamera $P = [I_{3 \times 3} | 0]$ sein soll, die

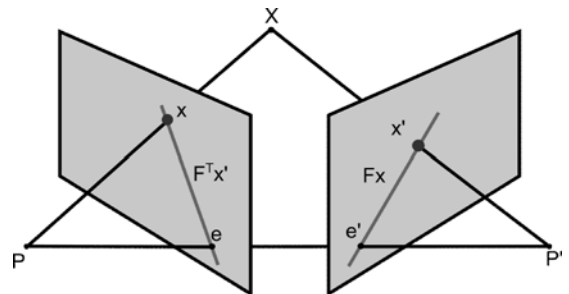


Abb. 2 Epipolare Geometrie. Die Punkte x und x' stehen in einer durch die Fundamentalmatrix F ausgedrückten Beziehung, die sich daraus ergibt, dass beide Punkte Abbildungen desselben Objektpunkts X sind. Dabei wird x auf die Linie Fx und x' auf die Linie $F^T x'$ abgebildet

zweite Kamera P' auf $P' = [[e']_{\times} F|e']$ festlegen. Dabei ist e' der sogenannte Epipol des zweiten Bilds. Im Epipol liegt nicht nur der Schnittpunkt aller Linien Fx , sondern auch der Schnittpunkt zwischen der Geraden, die die Kamerazentren verbindet und der Bildebene der zweiten Kamera. Entsprechend gibt es auch einen Epipol e auf der Bildebene der ersten Kamera. Beide Epipole sind in Abb. 2 eingezeichnet.

Der hier verwendete Operator $[e]_{\times}$ weist einem dreidimensionalen Vektor e eine schiefsymmetrische Matrix zu, sodass $\forall y \in \mathbb{R}^3 : e \times y = [e]_{\times} y$. Damit wird also das Kreuzprodukt in eine äquivalente Matrixschreibweise überführt.

Diese beiden Kameras, P und P' , entsprechen natürlich nicht den Kameras, mit denen I und I' ursprünglich aufgenommen wurden. Diese lassen sich auch nicht exakt rekonstruieren, da sowohl die Objektpunkte als auch die Kameras immer einer Ähnlichkeitsabbildung (bestehend aus Drehung, Translation und Skalierung) unterworfen werden können, ohne die Abbildung zu beeinflussen. Jedoch lassen sich P und P' auf zwei Kameras $\hat{P}$ und $\hat{P}'$ abbilden, die sich von den ursprünglichen Kameras in der Tat *nur* noch durch eine Ähnlichkeitsabbildung unterscheiden. Dazu ist eine 3D-Homographie, also eine 4×4 -Matrix H zu bestimmen, sodass:

$$\hat{P} = PH^{-1} \text{ und } \hat{P}' = P'H^{-1}, \quad (6)$$

$$\hat{X}_i = HX_i. \quad (7)$$

Dies ist das Ziel der Autokalibrierung.

Die Grundidee der Autokalibrierung

Das Überführen von P und P' in $\hat{P}$ und $\hat{P}'$ geschieht durch einen Prozess, der Autokalibrierung genannt wird. Ein häufig verwendeter Ansatz basiert auf der oben erwähnten Faktorisierung einer metrischen Kamera $\hat{P}$ zu

$$\hat{P} = KR[I_{3 \times 3} | -C] = K[R|t], \quad (8)$$

wobei wir $t = -RC$ gesetzt haben. Damit ergibt sich

$$\hat{P}\hat{P}^T = K[R|t][R|t]^T K^T. \quad (9)$$

Wenn wir den Translationsanteil aus $[R|t]$ entfernen, indem wir die Matrix

$$\hat{\Omega} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (10)$$

einfügen, so heben sich die Rotationsmatrizen auf, weil für diese $R^T = R^{-1}$ gilt. Konkret ergibt sich

$$\hat{P}\hat{\Omega}\hat{P}^T = K[R|t]\hat{\Omega}[R|t]^T K^T = KK^T. \quad (11)$$

Im Falle einer projektiv verzerrten Rekonstruktion, also mit $P = \hat{P}H$ und $P' = \hat{P}'H^{-1}$, haben wir jedoch

$$P \underbrace{H^{-1}\hat{\Omega}H^{-T}}_{\Omega} P^T = KK^T, \quad (12)$$

sodass Ω nicht mehr die in (10) gezeigte Form hat. Jedoch gilt auch $P\Omega P^T = KK^T$ immer noch. Falls wir also auf Ω schließen könnten, so wären wir in der Lage, durch eine Faktorisierung

$$\Omega = H^{-1}\hat{\Omega}H^{-T} \quad (13)$$

auf H zu schließen.

Ist eine genügend lange Bildsequenz vorhanden, so erlaubt die Minimierung von

$$\varepsilon_A = \sum_i \|KK^T - P\Omega P^T\|_F^2 \quad (14)$$

genau das. Die Matrix H genügt anschließend, um die vorläufige projektive Rekonstruktion in eine metrische umzuwandeln.

Autokalibrierung aus zwei Bildern

Könnte man die Autokalibrierung bereits auf der Basis von lediglich zwei Bildern erfolgreich durchführen, so ließen sich die dann kalibrierten Teilrekonstruktionen einfach zusammenführen, da lediglich eine Ähnlichkeitsabbildung bestimmt werden müsste. Dazu reichen lediglich drei in beiden Rekonstruktionen vorhandene Punkte aus. Üblicherweise werden jedoch zunächst längere Sequenzen im projektiven Raum zusammengeführt, bevor eine metrische Rekonstruktion durchgeführt wird [13], was daran liegt, dass die Autokalibrierung bisher auf kurzen Bildsequenzen eben nicht zuverlässig funktioniert hat.

Unser Ansatz, den wir im Folgenden vorstellen, basiert auf dem linearen Algorithmus aus [12]. Dabei wird die im Abschn. „Die Grundidee der Autokalibrierung“ erwähnte 4×4 -Matrix Ω zunächst bestimmt und dann gemäß Gl. 13 zerlegt, um H zu bestimmen. Bei Ω handelt es sich um eine symmetrische Matrix mit Rang 3. Im Fall zweier Bilder ist die Lösung des linearen Ansatzes jedoch unterbestimmt. Das Ergebnis ist ein Lösungsraum, der

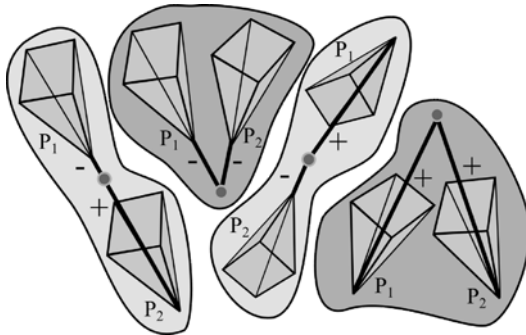


Abb. 3 Die mathematische Modellierung projektiver Kameras verbietet es nicht, dass Bildpunkte Projektionen von Objektpunkten sind, die sich eigentlich hinter der Kamera befinden. Aus diesem Grund können die vorläufigen Kameras aus Abschn. „Berechnung vorläufiger Kameras“ vier verschiedene Kamera-Konstellationen beschreiben

durch $\Omega_a + \gamma \Omega_b$ aufgespannt wird. Um die Rang-3-Bedingung zu erfüllen, werden nun solche Werte für γ gesucht, die die Determinante von $\Omega_a + \gamma \Omega_b$ verschwinden lassen [11, Abschn. „Linear Approach“].

Dazu müssen die Nullstellen eines Polynoms vom Grad vier ermittelt werden, was vier mögliche Lösungen ergibt. Zudem können die beiden vorläufigen Kameras P und P' eine von vier möglichen Konstellationen beschreiben, die in Abb. 3 dargestellt sind.

Wir haben also aus 16 möglichen Lösungen (bis zu 4 Lösungen für γ für jede von 4 möglichen Kamera-Konstellationen) die beste auszuwählen. Dazu wird aus den möglichen Ω jeweils eine Homographie H errechnet. Die beste Lösung führt zu einer H , bei der sich die meisten transformierten Objektpunkte HX_i vor den transformierten Kameras PH^{-1} und $P'H^{-1}$ befinden.

Dieser Kalibrierungsansatz 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 Falschzuordnungen enthalten oder die internen Abbildungsparameter der Kameras sich zu stark unterscheiden. In diesen Fällen versuchen wir, ein Bild im betrachteten Bildpaar durch ein anderes passendes Bild zu ersetzen (sofern genug Bildpunkt-Korrespondenzen gefunden werden können) und erneut eine Rekonstruktion zu finden. Diese Versuche werden so lange wiederholt, bis eine Kalibrierung gelingt oder bis keine weiteren passenden Bilder vorhanden sind.

Zusammenfügen metrischer Teilrekonstruktionen

Die Kamera-paare, bei denen die Autokalibrierung erfolgreich durchgeführt werden konnte, haben wir nun in einen metrischen Raum überführt. Das bedeutet, dass Abstände zwischen Objektpunkten in diesem Raum eine sinnvolle Größe darstellen. Das liegt daran, dass relative Abstände von projektiven Abbildungen nicht, von Ähnlichkeitsabbildungen aber sehr wohl erhalten werden.

Dieser Fakt erlaubt es uns, die Teilrekonstruktionen mithilfe der Objektpunkte auf einfache Weise zusammenzusetzen. Wie in Abb. 4 zu sehen ist, ist die Idee, nach und nach eine vollständige Rekonstruktion aufzubauen. Dazu wird eine Transformation berechnet, die die Kamerazentren aufeinander abbildet, die zu einer Ansicht gehören, die in beiden betrachteten Teilrekonstruktionen vorhanden ist.

Dazu entfernen wir zunächst den Einfluss der Kalibrierungsmatrix. Konkret hätte wir gerne eine Homographie, die Objektpunkte $\hat{X}_i$, für die

$$x_i = KR [I_{3 \times 3} | -C] \hat{X}_i \quad (15)$$

gilt, auf Objektpunkte X_i abbildet, für die

$$x_i = R [I_{3 \times 3} | -C] X_i \quad (16)$$

gilt. Die X_i sollen also ohne den Einfluss einer Kalibrierungsmatrix auf die 2D-Punkte abgebildet werden.

Wie sich leicht überprüfen lässt, ist durch

$$T = \begin{bmatrix} R & -RC \\ e_4^T & 1 \end{bmatrix}^{-1} \begin{bmatrix} P \\ e_4^T \end{bmatrix} \quad (17)$$

mit $e_4^T = (0 \ 0 \ 0 \ 1)$ eine solche Homographie gegeben, sodass $X_i = T\hat{X}_i$.

Basierend auf diesen transformierten Objektpunkten können die Teilrekonstruktionen zusammengeführt werden. Die Zusammenführung geschieht über die Berechnung einer Ähnlichkeitsabbildung. Eine solche lässt sich aus drei Objektpunkt-Korrespondenzen berechnen. Die beiden Kamerazentren, die zur gemeinsamen Ansicht gehören, stehen dabei bereits als Objektpunkt-Korrespondenz zur Verfügung. So müssen aus den Objektpunkten lediglich zwei weitere Korrespondenzen identifiziert werden. Bei der Auswahl dieser Objektpunkt-Korrespondenzen hilft die gemein-

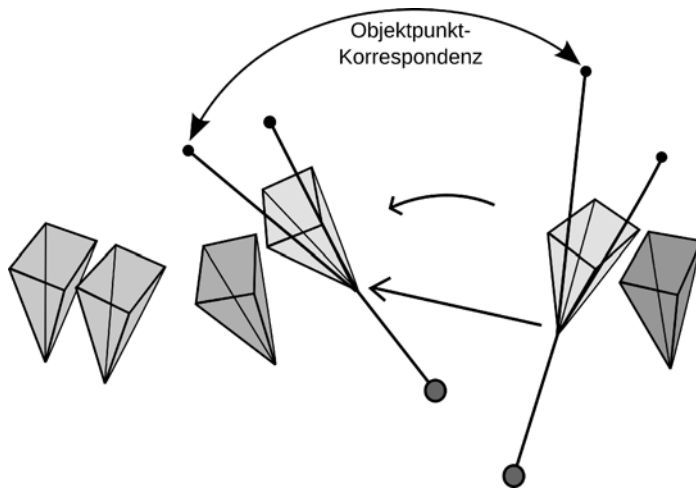


Abb. 4 Zusammenfügen von Teilrekonstruktionen. Auf der linken Seite sei eine bereits rekonstruierte Sequenz von Kamerapositionen angedeutet, deren letzte Position zur selben Ansicht gehört, wie die linke Kameraposition des gerade betrachteten Bildpaares, das auf der rechten Seite dargestellt ist. Diese beiden sollen mithilfe der Objektpunkt-Korrespondenzen zur Deckung gebracht werden



Abb. 5 Rekonstruktionsbeispiel. Gezeigt ist eine Rekonstruktion aus einer Bildfolge, von der einige Bilder links abgebildet sind

same Ansicht, da jeder Bildpunkt-Korrespondenz durch die oben genannte Triangulierung ein Objektpunkt zugeordnet ist. Da die triangulierten Objektpunkte potenziell fehlerbehaftet sind, wird der Auswahlprozess wieder in einen RANSAC-Rahmen eingebettet.

Sind nun drei Objektpunkte ausgewählt, so werden diese jedoch nicht direkt verwendet. Täte man dies, so erhielte man eine Ähnlichkeitsabbildung, die lediglich die jeweiligen Schwerpunkte der drei Objektpunkt-Tripel exakt aufeinander abbildet. Insbesondere kämen die beiden Kamerazentren aufgrund der immer mit Fehlern behafteten Rekonstruktion nicht zur Deckung. Um Letzteres zu erzwingen, berechnen wir aus den zwei Objektpunkten und dem Kamerapunkt jedes Tri-

pels jeweils einen virtuellen vierten Punkt, den wir statt des Kamerazentrums verwenden. Dieser virtuelle Punkt wird gerade so gewählt, dass das Kamerazentrum im Schwerpunkt des modifizierten Objektpunkte-Tripels zu finden ist.

Für die anschließende Berechnung der Ähnlichkeitsabbildung wird der SVD-Algorithmus eingesetzt [1]. Dieser Ansatz erweist sich in empirischen Versuchen als stabil und wird auch in der Literatur [8] als stabil angesehen.

Im letzten Schritt führen wir eine metrische Variante des oben erwähnten Bundle-Adjustment-Algorithmus aus, um die Rekonstruktion weiter zu optimieren.

Wir verwenden hier eine kompakte Parametrisierung durch sieben Parameter für die Kameras

(drei für die Drehung, drei für die Verschiebung und eine für die Brennweite) und drei Parameter für die 3D-Punkte. Abbildung 5 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 zur Begutachtung zu visualisieren sowie den Rekonstruktionsprozess zu verfolgen und eventuell interaktiv zu kontrollieren. Zudem erlaubt der Verzicht auf die Zusammenfügung im projektiven Raum vor der Autokalibrierung die Verwendung von sowohl in der Laufzeit als auch in der Umsetzung weniger aufwändigen Algorithmen.

Danksagung

Diese Arbeit entstand unter der Betreuung von Dr. Klaus Häming und Prof. Dr. Gabriele Peters und wurde unterstützt von der Deutschen Forschungsgemeinschaft (DFG) unter der Fördernummer PE 887/3-3.

Literatur

1. Arun KS, Huang TS, Blostein SD (1987) Least-squares fitting of two 3-d point sets. *IEEE Trans Pattern Anal Mach Intell* 9:698–700
2. Beis JS, Lowe DG (1997) Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: *Proc IEEE Conf Comp Vision Patt Recog*, San Juan, Puerto Rico, 17–19 June 1997, pp 1000–1006
3. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 24(6):381–395
4. Häming K, Peters G (2010) Structure-from-motion reconstruction pipeline – a survey with focus on short image sequences. *Kybernetika* 46(5):926–937
5. Harris C, Stephens M (1988) A combined corner and edge detector. In: *4th ALVEY Vision Conference*, University of Manchester, Manchester, England, 31 August–2 September 1988, pp 147–151
6. Hartley RI, Zisserman A (2003) *Multiple View Geometry in Computer Vision*, 2. Aufl. Cambridge University Press
7. Lindeberg T (1994) Scale-space theory: a basic tool for analyzing structures at different scales. *J Appl Statistics* 21(2):224–270
8. Lorusso A, Eggert DW, Fisher RB (1995) A comparison of four algorithms for estimating 3-d rigid transformations. In: *Proceedings of the 1995 British conference on Machine vision*, University of Birmingham, Birmingham, England, 11–14 September 1995, Bd 1, pp 237–246, BMVA Press
9. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vision* 60(2):91–110
10. Peters G, Häming K (2010) Fast freehand acquisition of 3d objects and their visualization. *J Commun Comput* 7:2–3
11. Pollefeys M, Koch R, van Gool LJ (1998) Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters, in: *Proceedings of the Sixth International Conference on Computer Vision (ICCV-98)*, Bombay, India, 4–7 January 1998, IEEE Computer Society 1998, pp 90–95
12. Pollefeys M, Verbiest F, Gool LV (2002) Surviving dominant planes in uncalibrated structure and motion recovery. In: *Johansen (ed) Computer Vision – ECCV 2002*, 7th European Conference on Computer Vision, Lecture Notes in Computer Science, Bd 2351, pp 837–851, Springer
13. Snavely N, Simon I, Goesele M, Szeliski R, Seitz SM (2010) Scene reconstruction and visualization from community photo collections. In: *Proceedings of the IEEE* 98(8):1370–1390
14. Triggs B, McLauchlan PF, Hartley RI, Fitzgibbon AW (2000) Bundle adjustment – a modern synthesis. In: *ICCV '99: Proceedings of the International Workshop on Vision Algorithms*, 20–25 September 1999, Corfu, Greece, Springer, London, pp 298–372
15. Vergauwen M, Gool LV (2006) Web-based 3d reconstruction service. *Mach Vision Appl* 17(6):411–426