

## Zuordnung von 3D Gebäudemodellen und IR-Videsequenzen mit Linienverfolgung

DOROTA IWASZCZUK<sup>1</sup>, LUDWIG HOEGNER<sup>1</sup> & UWE STILLA<sup>1</sup>

*Zusammenfassung: Die Position und Orientierung eines Luftfahrzeuges im Raum wird durch GPS, oft mit einer IMU, bestimmt. Oft ist die mit GPS/IMU aufgenommene Position und Orientierung nicht genau genug um die in den aufgenommenen Daten detektierten Objekte besser zu georeferenzieren und um die Daten aus verschiedenen Quellen zu fusionieren. In diesem Beitrag wird eine Methode zur Verbesserung der Georeferenzierung eines mit einer Infrarotkamera aufgenommenen Videostromes mittels Verfolgung markanter Geradenstücke durch die Bildsequenz vorgestellt. Es wird zwischen normalen Frames und Key-Frames unterschieden. In einem Key-Frame werden für jede Modelllinie innerhalb eines Puffers passende Bildlinien gesucht und als Kandidaten für Korrespondenzen weiter bearbeitet. Die richtigen Korrespondenzen werden mit einem RANSAC-Verfahren selektiert. Dann werden die Position und Orientierung der Kamera neu berechnet. Es werden dabei Unsicherheiten sowohl der äußeren Orientierung wie auch des Gebäudemodells berücksichtigt. Einmal im Bild zugeordnete Modellkanten werden mittels Korrelation im Bildraum durch die Bildsequenz verfolgt. Die Korrektheit der Linienverfolgung wird überprüft, indem in einem regelmäßigen Abstand „Key-Frames“ eingeführt werden.*

### 1 Einleitung

Die Wärmetexturen, die der Detektion und Dokumentation von Wärmeleckagen in der Gebäudehülle dienen, werden immer häufiger aus Infrarotdaten (IR) abgeleitet und den Gebäudemodellen zugeordnet. Die Außenfassaden können von einem Fahrzeug aufgenommen werden (HOEGNER et al, 2007). Von dem Fahrzeug sind jedoch die Dächer nicht sichtbar und die Innenhoffassaden sind oft für die Messfahrzeuge nicht zugänglich. Um diese Flächen abbilden zu können, werde fliegende Plattformen, wie z.B. Hubschrauber (IWASZCZUK, et al., 2012), Flugzeuge FRUEH et al. (2004) oder UAVs, eingesetzt. Um die Bilddaten mit den 3D-Gebäudemodellen zu kombinieren, werden die Parameter der äußeren und inneren Orientierung benötigt. Die Parameter der äußeren Orientierung des Luftfahrzeuges werden durch ein GPS, oft mit einer IMU (Eng. inertial measurement unit) kombiniert, aufgezeichnet. Die innere Orientierung des Kamerasystems wird durch den Selbstkalibrierungsprozess oder im Labor bestimmt. Oft ist die mit GPS und IMU aufgenommene Position und Orientierung nicht genau genug um die in den Daten detektierten Objekte mit hoher Genauigkeit zu georeferenzieren und um die Daten aus verschiedenen Quellen zu fusionieren. Um die äußere Orientierung der Kamera zu verbessern wird oft Model-to-Image Matching durchgeführt.

1) Photogrammetrie & Fernerkundung, Technische Universität München (TUM), Arcisstraße 21, 80333 München, <http://www.pf.bv.tum.de>

FRUEH et al. (2004) projizierte das 3D-Gebäudemodell mit verschiedenen, zufälligen Parametern ins Bild und korrelierte die Modelllinien mit den extrahierten Liniensegmenten. Aus mehreren Versuchen wurde dazu die Kameraposition mit der besten Anpassung ausgewählt. In jedem Bild der Sequenz verlief der Vorgang gleich. Diese Methode nutzt keine Zusammenhänge zwischen den Frames. Hu et al. (2006) verbinden die Frames mit markanten Punkten, die jedoch keinen Zusammenhang mit den Gebäudestrukturen aufweisen. Unter Annahme sog. „Manhattan“ oder „Legoland scenes“ mit dominanten horizontalen und vertikalen Richtungen kann die Ausrichtung der Kamera mithilfe der Fluchtpunkte berechnet werden (DING & ZAKHOR, 2008; CHAM et al., 2010; FOERSTNER, 2010). Dieses Verfahren hilft bei schlechten Navigationsparametern, die Fehler der Drehwinkel zu reduzieren. Um die Texturen präzise zuordnen zu können, wird eine Methode benötigt, die für das Matching bei Gebäuden charakteristische Bildmerkmale verwendet und diese von Frame zu Frame verfolgt.

## 2 Methode

Die in diesem Beitrag vorgestellte Methode dient der Verbesserung der Georeferenzierung eines mit der Infrarotkamera aufgenommenen Videostromes mittels Verfolgung markanter Liniensegmente durch die Bildsequenz. Es wird dabei angenommen, dass sowohl die Lage der extrahierten Bildkanten wie auch die Lage der 3D-Gebäudemodelle unsicher sind. Die Fortpflanzung der Unsicherheit wird schematisch in Abb. 1 dargestellt. Es wird ebenso angenommen, dass die Größe der Unsicherheit der 3D-Modellpunkte bekannt ist, und dadurch die Unsicherheit der 3D-Kante berechnet werden kann. Diese Unsicherheit wird während der Projektion in 2D fortgepflanzt. Im 2D-Bild werden die Kanten extrahiert und die Unsicherheiten der Kanten bestimmt. Dann werden die projizierten Kanten des 3D-Gebäudemodells mit den Bildkanten gematcht. Dabei werden die fortgepflanzten Unsicherheiten des Gebäudemodells und die Unsicherheiten der Bildkanten berücksichtigt.

Für die Projektion werden die mit dem GPS bzw. der IMU ermittelten Parameter der äußeren Orientierung sowie die durch Kalibrierung bekannten Parameter der inneren Orientierung verwendet. Da die mit GPS/IMU aufgezeichneten Parameter der äußeren Orientierung nur als Näherungswerte betrachtet werden können, sollen diese im Matching-Prozess so verbessert werden, dass das projizierte Modell und die Bildstrukturen möglichst gut aufeinander angepasst werden. Dazu werden zwei Typen von Frames definiert: Einerseits die sog. Key-Frames, in denen das vollständige Matching durchgeführt wird (Kap. 2.1), andererseits die normalen Frames in denen das vereinfachte Matching mittels Linienverfolgung stattfindet (Kap. 2.2). Die verfolgten Linien und deren Zuordnung zu den Gebäudekanten wird in einer Chronik gespeichert. Die hier präsentierte Methode setzt voraus, dass die Bewegungen zwischen den Frames gering sind.

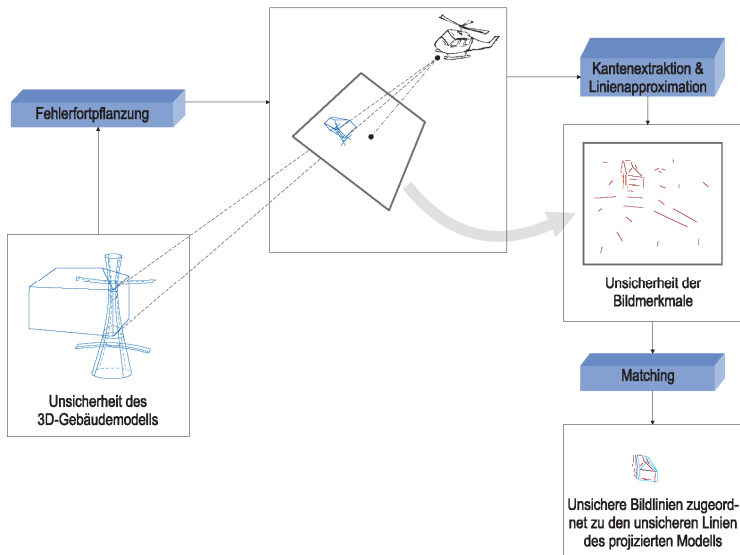


Abb. 1: Schematische Darstellung der Fehlerfortpflanzung und des Matchings

## 2.1 Key-Frame - Einzelbildfall

Um den Ablauf zu starten wird der erste Frame immer als ein Key-Frame definiert. Weitere Key-Frames werden in regelmäßigen Abständen eingesetzt. Ein Key-Frame wird als Einzelbild betrachtet. Zunächst werden die Kanten des 3D-Gebäudemodells ins Bild projiziert und für jede Kante wird in der lokalen Umgebung nach passenden Liniensegmenten im Bild gesucht. Die Größe des Suchbereiches wird durch die Fehlerfortpflanzung der Unsicherheit der Kameraposition und -Orientierung und der Unsicherheit der Modellkante berechnet. Die richtigen Korrespondenzen werden mit einem RANSAC-Verfahren selektiert. Dann werden die Position und Orientierung der Kamera neu berechnet. Um die Unsicherheiten sowohl der äußeren Orientierung wie auch des Gebäudemodells als Konsequenz der ungenauen Erfassung und der Generalisierung zu berücksichtigen, wird ein stochastisches Modell in der Ausgleichung formuliert (IWASZCZUK, et al., 2012). Die neu berechnete Position und Orientierung des Sensors werden auf die weiteren Frames übertragen, da die Verschiebung und Rotation des Sensors zwischen den Frames relativ genau mit der IMU bestimmt werden kann. Daher werden nach jedem Key Frame als Ergebnis die Korrekturen der Position und Orientierung für die nächste Aufnahme weitergegeben. Außerdem werden die Korrespondenzen gespeichert.

## 2.2 Linienverfolgung

Jede Linie, die in einem Key-Frame einer Modelllinie zugeordnet wird, wird im nächsten Frame gesucht. Da in einem Videostrom mehrere Frames pro Sekunde aufgenommen werden, sind die Unterschiede in der Erscheinung der Objekte in benachbarten Frames sehr gering. Deswegen eignet sich die Kreuzkorrelation für die Linienverfolgung, da die geometrischen Verzerrungen von Frame zu Frame sehr gering sind und die Ergebnisse der Kreuzkorrelation kaum beeinflussen. Die neue Position der Linie im nächsten Frame kann aus dem Pfad des Luftfahrzeuges geschätzt werden. Dafür wird um jede zugeordnete Linie ein Bildbereich gewählt und mit dem Bildbereich, wo dieselbe Linie erwartet ist, korreliert. Dadurch ergibt sich für jede Linie ein  $\text{offset}_i = [\Delta \text{row}_i, \Delta \text{col}_i]$ , aus dem eine Verschiebung zwischen den Frames als  $\text{median}(\text{offset}_i)$  berechnet wird. Dann wird in der Umgebung der verfolgten Linie überprüft, ob sich dort eine passende Linie befindet, die in dem neuen Frame als dieselbe Linie erkannt werden kann. Dafür werden statistische Tests nach HEUEL (2002) verwendet. HEUEL (2002) definiert die algebraische Bedingung für die Identität zweier Linien als

$$S(\mathbf{l})\mathbf{m} = -S(\mathbf{m})\mathbf{l} = \mathbf{0}, \quad (1)$$

wo

$$S(\mathbf{a}) = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}, \quad (2)$$

$\mathbf{m} = [m_1, m_2, m_3]^T$ ,  $\mathbf{l} = [l_1, l_2, l_3]^T$  sind Geraden in projektiven Koordinaten.

So kann eine Hypothese  $H_0: \mathbf{d} = S(\mathbf{l})\mathbf{m} = -S(\mathbf{m})\mathbf{l} = \mathbf{0}$  aufgestellt werden. Die Kovarianzmatrix für den Distanzvektor  $\mathbf{d}$  wird als

$$\Sigma_{\mathbf{d}\mathbf{d}} = S(\mathbf{l})\Sigma_{\mathbf{m}\mathbf{m}}S(\mathbf{l})^T + (-S(\mathbf{m}))\Sigma_{\mathbf{l}\mathbf{l}}(-S(\mathbf{m}))^T \quad (3)$$

berechnet.  $H_0$  kann mit dem Signifikanzniveau  $\alpha$  abgelehnt werden, wenn

$$T = \underline{\mathbf{d}}^T \Sigma_{\mathbf{d}\mathbf{d}}^{-1} \underline{\mathbf{d}} > \varepsilon_H = \chi_{1-\alpha, r}^2, \quad (4)$$

wo  $r = \text{rank}(\Sigma_{\mathbf{d}\mathbf{d}})$ .

Die Linienverfolgung wird nur für die Linien durchgeführt, die im nächsten Frame sichtbar sind. Es wird kontinuierlich geprüft, ob durch die Bewegung des Sensors zusätzliche Kanten des Gebäudemodells im Bild sichtbar werden oder verschwinden. Diese Information wird aus der bekannten Geometrie des 3D Gebäudemodells mithilfe des Hidden-Line-Algorithmus berechnet. Dafür werden Ereignisse und Zustände für jede Modelllinie definiert. Prinzipiell kann eine Linie folgende Zustände aufweisen: *lebend vollkommen sichtbar*, *lebend verletzt (teilsichtbar)*, *verdeckt*, *tot*. *lebend vollkommen sichtbar* ist eine Linie, die nach der Projektion vollkommen im Bild erscheint und von keinen Objekten verdeckt wird. *lebend verletzt* ist dagegen eine Linie,

die nur Teilweise im Bild erscheint oder zum Teil von einem Objekt verdeckt wird. Im *verdeckten* Zustand befindet sich eine Linie die zwar ins Bild projiziert wird, aber von einem anderen Objekt vollkommen verdeckt wird. Fällt eine Linie aus dem Bildbereich, so wird sie als *tot* gekennzeichnet. Der Zustand der Modelllinie definiert, wie die korrespondierenden Bildlinien verfolgt werden. Die *lebenden* Linien werden mit den extrahierten Bildlinien abgeglichen, während die Verfolgung der *verdeckten* Linien unterbrochen wird, bis diese wieder erscheinen. Die toten Linien werden nicht mehr verfolgt. Die Tab. 1 definiert sechs Ereignisse (Events), die den Zustand der Linie verändern. Es werden mögliche Zustände vor und nach jedem Ereignis aufgeführt.

Ereignis	Mögliche Zustände vor dem Ereignis	Mögliche Zustände nach dem Ereignis
Geburt	∅	Lebend vollkommen sichtbar Lebend verletzt
Erscheinung	Verdeckt	Lebend vollkommen sichtbar Lebend verletzt
Verletzung	Lebend vollkommen sichtbar	Lebend verletzt
Heilung	Lebend verletzt	Lebend vollkommen sichtbar
Verschwenden	Lebend vollkommen sichtbar Lebend verletzt	Verdeckt
Tod	Lebend vollkommen sichtbar Lebend verletzt Verdeckt	Tot

Tab. 1: Ereignisse und Zustände, die für eine Linie auftreten können

### 3 Testdaten

Die IR-Daten wurden während einer Flugkampagne über einem dicht bebauten Bereich mit einer im Hubschrauber montierten IR-Kamera aufgenommen. Die Frequenz des Videostroms lag bei 25 Bildern pro Sekunde. Die Flughöhe betrug ca. 400 m über Grund und die Bilder wurden schräg nach vorne unter einem Pitch-Winkel von ca. 45° aufgenommen.

Der Hubschrauber verfügte über ein GPS/INS Navigationssystem. Das Projektionszentrum der Kamera war im Verhältnis zur Antenne des Navigationsgerätes versetzt. Die innere Orientierung der Kamera war nur ungefähr bekannt. Daher wurde eine erweiterte Systemkalibrierung mit der Berechnung von boresight & leverarm Parametern (KOLECKI et al., 2010) durchgeführt.

Die 3D-Gebäudemodelle wurden im Detailierungsgrad LOD2 aus Luftbildern semiautomatisch erstellt und im *CityGML*-Format gespeichert.

### 4 Ergebnisse

Die ersten Tests wurden mit den oben beschriebenen Daten durchgeführt. Es wurden fünf nacheinander folgende Frames gewählt. Der erste Frame wurde als Key-Frame definiert. Es

wurden automatisch mit RANSAC zum Modell korrespondierende Kanten gewählt und die Parameter der äußere Orientierung neu berechnet. Die Abb. 2 zeigt das mit den neu berechneten Parametern projizierte Modell (links) und die mit RANSAC gewählten Liniensegmente im Bild. Die dem Modell zugeordneten Linien wurden in den folgenden Frames verfolgt. Die Ergebnisse der Verfolgung wurden in Abb.3 dargestellt. In den Frames 2-4 wurde das in den Frame 1 projizierte Modell um den entsprechenden mittleren Offset verschoben. Die Genauigkeit der in dem ersten Frame neu bestimmten Parameter der äußere Orientierung lag bei 3-4 [m] und 0.3-0.5[°].

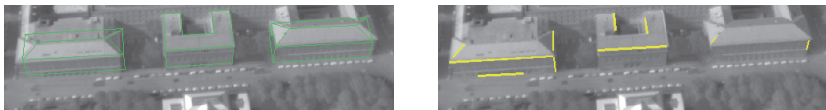


Abb. 2: Ausschnitt aus dem Key-Frame: links - das mit neu berechneten Parametern projizierte 3D-Gebäudemodell, rechts - die mit RANSAC gewählten Liniensegmente, die in den nächsten Frames verfolgt werden

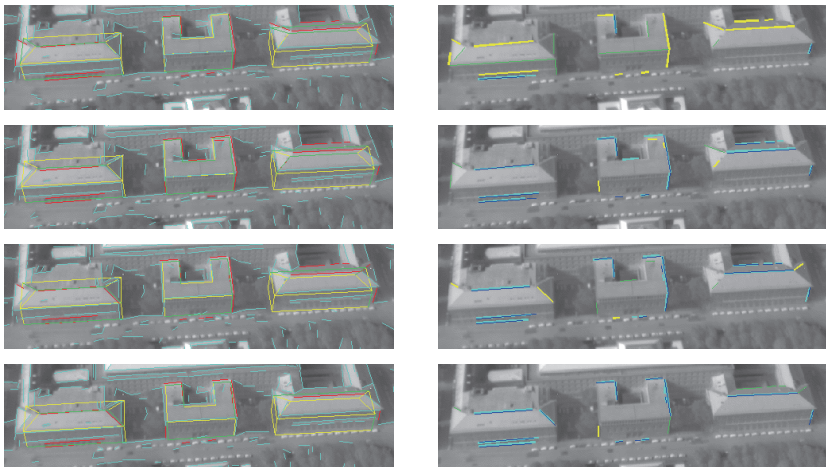


Abb. 3: Ausschnitte aus vier nacheinander folgenden Frames. Links: extrahierte Bildkanten, die dem verfolgten Modell zugeordnet worden sind (Rot), nicht zugeordnete Bildkanten (Cyan), Modellkanten mit zugeordneten Bildkanten (Grün), Modellkanten ohne Zuordnung (Gelb). Rechts: Aus dem vorherigen Frame verfolgte Kanten, die wiedergefunden wurden (dunkel Blau) und Linien die keinen Kanten im aktuellen Bild (Cyan) zugeordnet worden sind (Grün); in Gelb werden die neugefundenen Bildkanten dargestellt, die im vorherigen Bild nicht vorkamen.

## 5 Diskussion

Die Verfolgung der Geradenstücke schränkt den Suchraum für Linienkorrespondenzen ein und beschleunigt dadurch den Prozess. Die Genauigkeit der Neuberechneten Parameter der äußeren Orientierung ist jedoch zu schlecht um ab dem zweiten Frame die Linien im Bild ohne RANSAC verfolgen zu können. Deswegen sollen in Zukunft Experimente mit mehreren Key-Frames am Anfang der Sequenz durchgeführt werden. Die Key-Frames sollen auch nicht am Anfang festgelegt werden, sondern es soll dynamisch entschieden werden, ob der nächste Frame als Key-Frame festgelegt wird. Dafür werden Qualitätsmaße und Kriterien benötigt, die diese Entscheidung automatisch ermöglichen. Es soll untersucht werden, welche Größen (z.B. Genauigkeit des Matchings, Anzahl der gefundenen Korrespondenzen) dafür geeignet sind. Es sollen auch Zuverlässigkeitsmaße für die verfolgten Linien definiert werden. Die Linien, die häufiger gefunden werden, werden als zuverlässiger eingestuft und länger verfolgt, während die nur gelegentlich gefundenen Linien als weniger zuverlässig gesehen werden und die Verfolgung schneller verworfen wird.

Es soll außerdem die Korrespondenzsuche verbessert werden. Insbesondere bei ungenauen GPS/INS Daten können mit RANSAC nicht immer die richtigen Korrespondenzen gefunden werden (IWASZCZUK, et al., 2012). RANSAC lässt auch mehrere Bildlinien die einer Modelllinie zugeordnet sind zu. Deswegen soll in der Zukunft eine Methode eingesetzt werden, die nur 1-zu-1 Korrespondenzen zulässt. Alternativ kann ein Akkumulator-Verfahren angewandt werden, das noch robuster als RANSAC die richtigen Korrespondenzen bestimmt.

Die Analyse der Ereignisse und Zustände hilft die Zuverlässigkeit der Verfolgung zu erhöhen und kann in der Zukunft auch bei der Flugplanung genutzt werden, um die optimale Flugroute des Luftfahrzeuges zu bestimmen.

## 6 Danksagung

Die Autoren danken Fraunhofer IOSB für die Bereitstellung der IR-Daten.

## 7 Literaturverzeichnis

- CHAM T.J.; CIPTADI A.; TAN W.C.; PHAM M.T. & CHIA L.T., 2010: Estimating Camera Pose from a Single Urban Ground-View Omnidirectional Image and a 2D Building Outline Map. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, USA
- DING, M. & ZAKHOR, A., 2008: Automatic registration of aerial imagery with untextured 3D LiDAR models. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- FOERSTNER, W., 2010: Optimal vanishing point detection and rotation estimation of single image from a Legoland scene. The International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences, 38(3B): 157-162
- FRUEH C.; SAMMON R. & ZAKHOR A., 2004: Automated Texture Mapping of 3D City Models With Oblique Aerial Imagery. Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04)

- HEUEL, S., 2002: Statistical Reasoning in Uncertain Projective Geometry for Polyhedral Object Reconstruction. Photogrammetry Department, Institute for Geodesy and Geoinformation, Bonn University, Germany, PhD Thesis
- HOEGNER, L.; KUMKE, H.; MENG, L. & STILLA, U., 2007: Automatic extraction of textures from infrared image sequences and database integration for 3D building models. *Photogrammetrie Fernerkundung Geoinformation (PFG)*, 2007(6): 459-468
- HU, J.; YOU, S. & NEUMANN, U., 2006: Automatic Pose Recovery for High-Quality Textures Generation. *Proceedings of 18<sup>th</sup> International Conference in Pattern Recognition*: 561-565
- IWASZCZUK, D.; HOEGNER, L.; SCHMITT, M. & STILLA, U., 2012: Line based Matching of Uncertain 3D Building Models with IR Image Sequences for Precise Texture Extraction. *Photogrammetrie, Fernerkundung, Geoinformation (PFG)*, 2012(5): 511-521
- KOLECKI, J., IWASZCZUK, D. & STILLA, U., 2010: Calibration of an IR camera system for automatic texturing of 3D building models by direct geo-referenced images. *Proceedings of Eurocow 2010, Castelldefels, Spain*



