

Bildbasierte Geolokalisierung für UAVs

Michael Schleiss

Fraunhofer FKIE,
Fraunhoferstr. 20, 53343 Wachtberg

Zusammenfassung When unmanned aerial vehicles (UAVs) fly autonomous missions, they typically rely on global satellite navigation systems (GNSS) like GPS for global position estimation. However, GNSS signals can be easily jammed. We propose a camera-based method that uses onboard imagery and data from OpenStreetMap as a backup system for GNSS. First, the aerial imagery from the onboard camera is translated into a map-like representation. Then we match it with a reference map to infer the vehicle's position. Experiments over a typically sized mission area are performed and exhibit localization accuracy close to 6 m. Our results show that the proposed method can serve as a backup to GNSS systems where suitable landmarks like buildings and roads are available.

Keywords Image-based navigation, geolocalisation, GPS-denied, UAV

1 Einleitung

Wenn selbstfahrende Autos durch Tunnel oder tiefe Hochhaus-schluchten fahren, dann benötigen diese beim Navigieren einen Ersatz für die Satellitennavigation, denn GPS und Co stehen in diesen Situationen nicht zur Verfügung. Ähnlich sieht es beim Einsatz von autonomen UAVs in geschlossenen Räumen aus. Daher wurden für diese Einsatzzwecke unter anderem visuelle Methoden zur Lokalisierung erforscht [1,2], die bei fehlendem Signal von Satellitennavigationssystemen (GNSS) als Ersatz fungieren können.

DOI: 10.58895/ksp/1000124383-32 erschienen in:

Forum Bildverarbeitung 2020

DOI: 10.5445/KSP/1000124383 | <https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000124383/>

Anders sieht es beim Außeneinsatz von autonomen Drohnen aus. Hoch in der Luft geht man bisher von einem sehr guten Empfang von GNSS-Signalen und einer hohen Genauigkeit der Eigenpositionsbestimmung (1-5 m) aus [3]. Mit Hilfe im Handel frei verfügbarer Technik kann man jedoch GNSS-Signale blockieren (jamming) oder fälschen (spoofing) [4]. UAVs, die zur Lokalisierung nur auf GNSS setzen, werden so zum Landen gezwungen und können von böswilligen Akteuren gekapert oder zerstört werden. Im schlimmsten Fall droht sogar der Absturz des Vehikels.

Wenn man jedoch bedenkt, dass autonome Luftfahrzeuge in Zukunft ein integraler Bestandteil der Logistik werden und den Transport von wertvollen Gütern, wie Medikamente [5] und Organe [6], oder sogar Personen [7] automatisieren sollen, wird schnell klar, dass man auch beim Einsatz von UAVs unter freiem Himmel eine Backup-Strategie für den Ausfall der Satellitennavigation benötigt. Auch Polizei- und Rettungskräfte werden in der Zukunft vermehrt auf den Einsatz von UAVs zurückgreifen, zum Beispiel zur Bewachung von kritischer Infrastruktur, der Verschaffung eines Überblicks bei Katastrophen oder der Suche nach Vermissten. Auch hier liegt eine Verletzbarkeit vor, die durch Kriminelle und Terroristen ausgenutzt werden könnte.

Ziel der Forschungstätigkeit ist es daher eine Methode zur visuellen Bestimmung der Eigenposition für UAVs unter freiem Himmel vorzustellen, um GNSS-Jamming und Spoofing umgehen zu können (siehe Abb. 1.1).

2 Verwandte Arbeiten

Wir unterscheiden zunächst einmal zwei Arten der Lokalisierung. Die relative Lokalisierung gibt die Position in Bezug auf einen Startpunkt an, bei der absoluten Lokalisierung erhält man eine georeferenzierte Position in Latitude und Longitude, so wie bei einem Satellitennavigationsempfänger.

Als Beispiel für die relative Lokalisierung sei die visuelle Odometrie genannt mit Hilfe des optischen Flusses genannt [8]. Es werden aufeinanderfolgende Bildpaare verglichen und, vorausgesetzt man kennt die Flughöhe, aus dem Versatz eine Bewegungsrichtung und

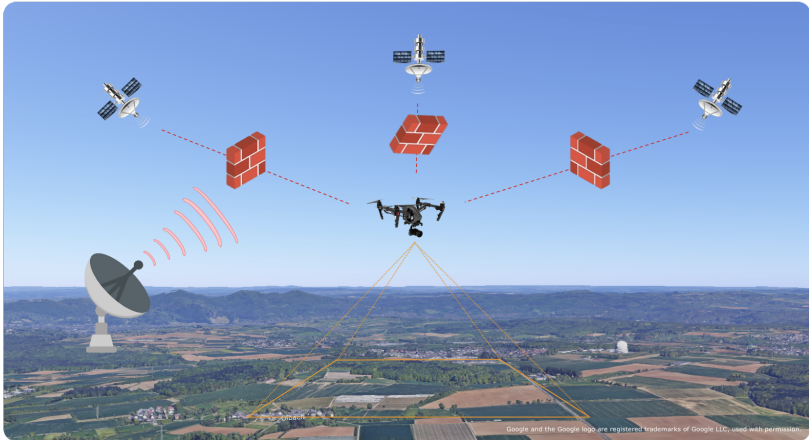


Abbildung 1.1: Aufgrund der hohen Distanz zwischen UAV und den Navigationssatelliten ist es ein Leichtes, das Signal mit Hilfe eines Störers zu stören. Die Nutzung der vorhandenen Bordsensorik soll in diesem Fall eine autarke Lokalisierung ermöglichen.

Geschwindigkeit berechnet. Bei der absoluten Lokalisierung kommen zum Beispiel im militärischen Umfeld traditionell terrainbasierte Lokalisierungsverfahren oder Scene Matching-Verfahren, wie TERCOM [9] und DSMAC [10], zum Einsatz.

Die weite Verbreitung von kommerziell erhältlichen UAVs in der jüngeren Vergangenheit führte auch zu vermehrter Forschung zu visueller Geolokalisierung in diesem Bereich. Einer der ersten Versuche eine Geolokalisierung aus den Bildern der Bordkamera eines leichtgewichtigen UAVs zu gewinnen, wurde von Conte und Doherty vorgeschlagen [8]. Diese kombinieren eine visuelle Odometrie mit einem Algorithmus, der die Bordbilder mit einer Datenbank aus georeferenzierten Luftaufnahmen abgleicht, um den Drift zu reduzieren. Der Abgleich basiert auf der normalisierten Kreuzkorrelation der Bildintensitäten. Conte und Doherty berichten von brauchbaren Ergebnissen, dies ist jedoch überwiegend auf die Leistung der visuellen Odometrie zurückzuführen. Das Modul für die Driftkorrektur liefert nur verhältnismäßig seltene Ausgaben, da die meisten „Matches“ wegen hoher Unsicherheit zurückgewiesen werden. In ih-

rem Experiment konnten nur an zwei Positionen eine Driftkorrektur durchgeführt werden [8].

Im Gegensatz dazu nutzen Cesetti et al. Feature Deskriptoren, nämlich SIFT, für die Georeferenzierung der Bordbilder [11]. Vorausgesetzt werden große Flughöhen, um sinnvolle Merkmale aus natürlichen Landmarken extrahieren zu können. In geringen Flughöhen können verrauschte Muster von Bäumen, Wiesen etc. im Bildmaterial dominieren. In ihren Experimenten berücksichtigen Cesetti et al. daher nur Bordbilder mit einem Fußabdruck am Boden von mindestens einem Quadratkilometer. Dadurch ist der Einsatz dieser Methode nur auf spezifische Szenarien eingeschränkt.

Grönwall et al. erweitern [8] indem Sie Lidar-gestützte Messungen für die visuelle Odometrie heranziehen [12]. Jedoch bleibt das grundlegende Problem von seltenen Matches weiterhin bestehen. Shan et al. übersetzen Bilder und Referenzmaterial in eine HOG-basierte (Histogram of oriented Gradients) Repräsentation.

Lindsten et al. segmentieren das Bild anhand verschiedener Klassen wie Straßen, Gebäude, Wiesen und Gewässern und vergleichen dann mit Hilfe eines Histogramms der Pixelhäufigkeiten pro Klasse die Bildaufnahmen mit einer entsprechenden Referenzkarte [13]. Zur Segmentierung kommt ein Clusteringalgorithmus in Form von Superpixeln zum Einsatz. Durch die Nutzung eines Histogramms gehen jedoch geometrische Informationen verloren und führen zu uneindeutigen Positionsschätzungen in Arealen mit ähnlichen Klassenverteilungen.

Mannberg und Savvaris nutzen Objekt Detektoren, um die Position von Gebäuden in Luftaufnahmen zu bestimmen und reduzieren die Detektionen in eine Repräsentation, in der jedes Gebäude von einem Punkt auf einer Karte dargestellt wird [14]. Ein Fingerabdruck, der die geometrische Anordnung der Punkte berücksichtigt, wird berechnet und in einer Referenzdatenbank abgeglichen. Die Autoren berichten, dass ihr Framework auch auf andere Typen von Landmarken ausgeweitet werden kann. Es ist aber unklar, inwiefern Landmarken, die man nicht zu Punkten reduzieren kann, wie Straßen und Flüsse, eingebunden werden sollen.

Wir nutzen den gleichen Template Matching Ansatz wie Conte und Doherty. Aber wir erhalten höhere Matchingraten, indem wir die Bordbilder segmentieren und dadurch in eine robuste Re-

präsentation überführen. Der Segmentierungsprozess ist vergleichbar mit dem Ansatz von Lindsten, allerdings nutzen wir ein neuronales Netz und gleichen die segmentierten Bilder mit einer Referenzkarte statt einem Histogramm. Dadurch bleibt die geometrische Anordnung der Landmarken erhalten. Unsere Methode funktioniert in verschiedenen Flughöhen typisch für kommerzielle UAVs und kann mehrere Arten von Landmarken (Häuser, Gebäude, Wälder, Flüsse, etc.) berücksichtigen.

3 Methodik

Ähnlich dem Scene Matching und dem Verfahren von Conte et al. haben wir ein Verfahren entwickelt, das das Bild einer zum Boden gerichteten Bordkamera mit einer Referenzdatenbank vergleicht und zur absoluten Lokalisierung nutzt, also georeferenzierte Positionen ausgibt. Das Verfahren lässt sich in drei Schritte einteilen (siehe auch Abb. 3.1).

- Die Bordkamera erstellt eine Luftaufnahme (Nadir).
- Ein neuronales Netz segmentiert die Luftaufnahme und übersetzt sie damit in eine straßenkartenähnliche Repräsentation.
- Das segmentierte Bild wird mit einer Referenzkarte bestehend aus Straßen und Hausgrundrissen per Template Matching abgeglichen.

Im Rahmen der visuellen Geolokalisierung ist es in unserem Fall ausreichend zwei Freiheitsgrade, nämlich Latitude und Longitude zu bestimmen, denn sowohl die Orientierung als auch die Höhe über Grund können driftfrei per inertialer Messeinheit, Magnetometer und Altimeter bestimmt werden. Entsprechende Sensorik ist für kommerziell erhältliche UAVs verfügbar und kann vorausgesetzt werden. Wir erwarten zudem, dass die Bordbilder nach Norden und in Lotrichtung zum Boden ausgerichtet sind. Dies kann durch ein 3-DoF Gimbal problemlos sichergestellt werden. Alternativ werden die Bilder mit Hilfe von Informationen der inertialen Messeinheit und eines magnetischen Kompasses perspektivisch korrigiert.

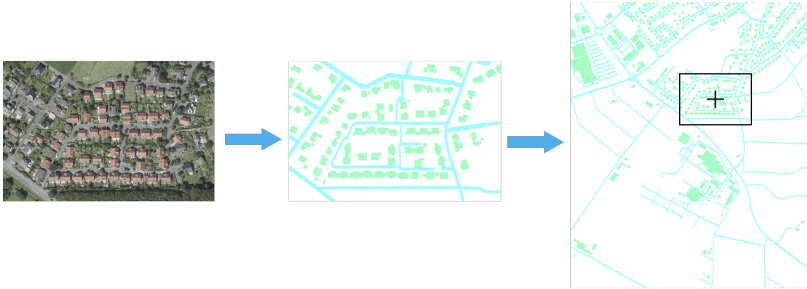


Abbildung 3.1: Das Verfahren besteht aus drei Schritten (v.l.n.r.): Aufnahme eines Luftbildes, Segmentierung von Straßen (blau) und Häusern (grün), Abgleich mit einer Referenzkarte.

Für die Segmentierung kommt ein neuronales Netz zum Einsatz, das darauf trainiert wurde, bestimmte Landmarken zu erkennen. Wir haben uns für Gebäude und Straßen entschieden, da diese bereits eine große Abdeckung in vielen Einsatzszenarien ermöglichen und entsprechendes Referenzmaterial öffentlich und kostenlos verfügbar ist, zum Beispiel über OpenStreetMap. Wir greifen dabei auf die U-Net Architektur zurück, die ursprünglich aus dem Bereich der medizinischen Bildverarbeitung stammt [15], mittlerweile aber in vielen anderen Szenarien, wie zum Beispiel der pixelweisen Segmentierung von Straßenszenen [16], zur Anwendung kommt. In unserem Fall empfängt das neuronale Netz die Bordbilder einer Tageslichtkamera und weist jedem Pixel eine Klasse, zum Beispiel Haus, Hintergrund oder Straße zu.

Wir nutzen, ähnlich wie Conte und Doherty [8] ein Template Matching Verfahren. Dabei wird das segmentierte Bild in Sliding-Window-Manier über die Referenzkarte geschoben. An jeder Position wird die Summe der Quadrate der Grauwertunterschiede bestimmt. Die Position mit der geringsten Abweichung stellt für uns einen Match dar und wird für die Ausgabe der Positionsschätzung herangezogen. Voraussetzung sind, dass der Maßstab, in dem die Bordbilder aufgenommen wurden, bekannt ist. Dieser wird mit Hilfe des Altimeters bestimmt und die Bordbilder dementsprechend skaliert, sodass sie mit der Bodenauflösung der Referenzkarte übereinstimmt.

Wir treffen in unserem Verfahren folgende Annahmen. Wir gehen davon aus, dass die Erdoberfläche lokal in unserem Missionsgebiet durch eine Ebene approximiert werden kann. Wir gehen auch davon aus, dass die grobe Position zu Beginn des Fluges bekannt ist, sodass das Flugvehikel mit einer geeigneten Referenzkarte des Missionsgebiets ausgestattet werden kann.

4 Training des Bildsegmentierers

Der Bildsegmentierer dient dazu, Landmarken wie Häuser und Gebäude aus den Bordbildern zu extrahieren. Dafür trainieren wir diesen mit Hilfe einer großen Sammlung an öffentlich verfügbaren Luftaufnahmen¹ und Daten aus OpenStreetMap. Wir haben dafür ein 125km² großes Gebiet um Bonn gewählt, das sowohl urbane auch ländliche Komponenten enthält. Das Gebiet wurde in Patches, der Größe 512 x 512 Pixel mit einer Bodenauflösung von 0,1 m pro Pixel aufgeteilt. Dies entspricht ca. 90.000 Trainingsbildern.

Die Trainingsmasken wurden mit Hilfe von Gebäudeumrissen und Straßenlinien aus OpenStreetMap erstellt. Da für Straßen nur die Mittellinie vermerkt ist, wurde die Breite anhand des Typs der Straße (Autobahn, Bundesstraße, Wohnstraße, etc.) geschätzt. Für die weiteren Details des Trainingsprozedere verweisen wir auf [17].

Die Luftaufnahmen bilden ein Gebiet über 100km² aus dem Stadtgebiet von Bonn ab. Es enthält den Stadtkern, aber auch Randgebiete, landwirtschaftliche Flächen und Wälder. Einmal trainiert lässt sich der Bildsegmentierer auch über anderen Gebieten anwenden, solange diese in ihrer Erscheinung dem Trainingsdatensatz ähneln [18,19].

5 Evaluierung

Das Lokalisierungsexperiment wird auf einem separaten Datensatz evaluiert. Anstatt der öffentlich verfügbaren Luftbilder, nutzen wir hier Daten, die wir selbst auf einem Gebiet südlich von Bonn aufgenommen haben.

¹ Digitale Orthophotos NRW: https://www.bezreg-koeln.nrw.de/brk_internet/geobasis/luftbildinformationen/aktuell/digitale_orthophotos

Die Nutzlast des Flugvehikels besteht aus einer Tageslichtkamera und einem INS. Die Kamera nimmt fünf Bilder pro Sekunde mit einer Auflösung von 3.280×2.464 Pixeln, einem Öffnungswinkel von $39,1^\circ$ und einer durchschnittlichen Flughöhe von circa 300 m auf. Die Bodenfläche beträgt ca. 216 m auf 144 m bei einer Auflösung von unter 0,1 m pro Pixel.

Das inertialen Navigationssystem (INS) besteht aus Gyroskop, Accelerometer und Magnetometer. Ein Altimeter zur Höhenmessung steht in dieser Messreihe nicht zur Verfügung. Stattdessen wird die Höheninformation des GPS-Moduls genutzt. Die Kamera und das INS sind fest am Flügel fixiert und nicht an einem Gimbal angebracht. Neben Kamera und INS befindet sich ein GPS+RTK Modul in der Nutzlast, mit dessen Hilfe eine zentimetergenaue Position als Referenz für die Evaluierung aufgezeichnet wird.

Das Verfahren wird auf einer Flugbahn von 1,61km Länge evaluiert. Die Referenzkarte umfasst ein Gebiet von circa einem Quadratkilometer.

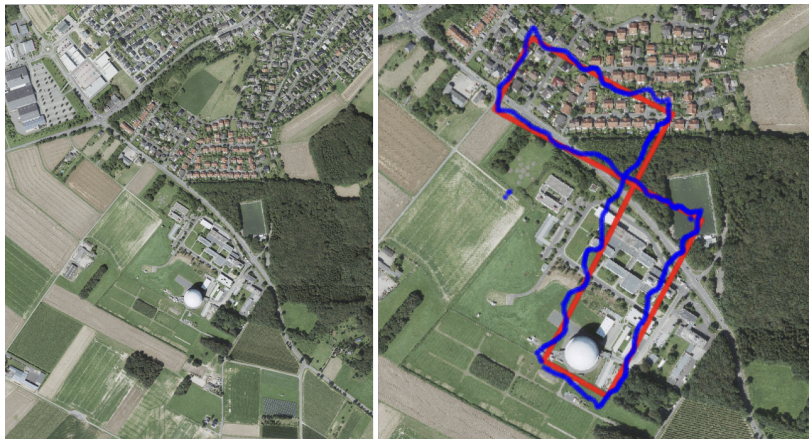


Abbildung 5.1: Links ist eine Luftaufnahme des Referenzgebiets zu sehen, in dem die Position gesucht wurde. Rechts sieht man die ermittelten Positionen (blau) und die Referenzpositionen, die durch das GPS+RTK ermittelt wurden (rot). Die drei blauen Punkte über dem Acker (links, Mitte) stellen Fehllokalisierungen dar.

Die Trajektorie besteht aus 471 Einzelbildern der Bordkamera. Jedes dieser Bilder wurde unabhängig in der Referenzkarte lokalisiert, um den Fokus auf die Qualität der Geolokalisierung zu legen. Das heißt, es wurde keine Informationen eines Bewegungsmodells berücksichtigt und keine Filterschritte durchgeführt.

Das Ergebnis der Evaluierung wird in Abb. 5.1 dargestellt. Die mittlere Abweichung von der Referenzposition beträgt 5,7m bei einer Standardabweichung von 7,4m.

6 Fazit

Bezogen auf die Positionierungsgenauigkeit ist das beschriebene Verfahren damit in der Lage, in diesem Anwendungsfall, als Ersatz für Satellitennavigationssystemen zu fungieren. Kritisch ist zu sehen, dass das Verfahren nur in Gebieten funktioniert, wo es auch ausreichend menschliche Bebauung in Form von Straßen und Häusern gibt. Ziel der weiteren Forschung wird es sein, diese Limitation aufzuheben. Zudem sollen weitere Datensätze gesammelt werden, die eine Vielzahl von Einsatzszenarien, wie zum Beispiel unterschiedliche Jahreszeiten, Wetterbedingungen oder Landbedeckungen abdecken und als Benchmark für die entwickelten Verfahren dienen sollen.

Literatur

1. J. V. Carroll, "Vulnerability assessment of the us transportation infrastructure that relies on the global positioning system," *The Journal of Navigation*, vol. 56, no. 2, p. 185, 2003.
2. A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.
3. H. Taira, M. Okutomi, T. Sattler, M. Cimpoi, M. Pollefeys, J. Sivic, T. Pajdla, and A. Torii, "Inloc: Indoor visual localization with dense matching and view synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7199–7209.

4. A. Ruegamer, D. Kowalewski *et al.*, "Jamming and spoofing of gnss signals—an underestimated risk?!" *Proc. Wisdom Ages Challenges Modern World*, vol. 3, pp. 17–21, 2015.
5. B. McCall, "Sub-saharan africa leads the way in medical drones," *The Lancet*, vol. 393, pp. 17–18, 2019.
6. M. Francisco, "Organ delivery by 1,000 drones," *Nature Biotechnology*, vol. 34, p. 684, 2016.
7. D. Lenton, "The measure of volocopter flying taxi," *Engineering & Technology*, vol. 13, no. 7/8, pp. 10–11, 2018.
8. G. Conte and P. Doherty, "Vision-based unmanned aerial vehicle navigation using geo-referenced information," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1–18, 2009.
9. J. P. Golden, "Terrain contour matching (tercom): a cruise missile guidance aid," in *Image processing for missile guidance*, vol. 238. International Society for Optics and Photonics, 1980, pp. 10–18.
10. J. R. Carr and J. S. Sobek, "Digital scene matching area correlator (ds-mac)," in *Image Processing For Missile Guidance*, vol. 238. International Society for Optics and Photonics, 1980, pp. 36–41.
11. A. Cesetti, E. Frontoni, A. Mancini, A. Ascani, P. Zingaretti, and S. Longhi, "A visual global positioning system for unmanned aerial vehicles used in photogrammetric applications," *Journal of intelligent & robotic systems*, vol. 61, no. 1-4, pp. 157–168, 2011.
12. C. Grönwall, J. Rydell, M. Tulldahl, E. Zhang, F. Bissmarck, and E. Bilock, "Two imaging systems for positioning and navigation," in *2017 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS)*. IEEE, 2017, pp. 120–125.
13. F. Lindsten, J. Callmer, H. Ohlsson, D. Törnqvist, T. B. Schön, and F. Gustafsson, "Geo-referencing for uav navigation using environmental classification," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 1420–1425.
14. M. Mannberg and A. Savvaris, "Landmark fingerprinting and matching for aerial positioning systems," *Journal of Aerospace Information Systems*, vol. 11, no. 3, pp. 131–139, 2014.
15. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

16. M. Siam, M. Gamal, M. Abdel-Razek, S. Yogamani, and M. Jagersand, "Rtseg: Real-time semantic segmentation comparative study," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 1603–1607.
17. D. J. H. Olivier Courtin, *RoboSat.pink Computer Vision framework for GeoSpatial Imagery*, DataPink, 2019. [Online]. Available: <http://RoboSat.pink>
18. P. Kaiser, J. D. Wegner, A. Lucchi, M. Jaggi, T. Hofmann, and K. Schindler, "Learning Aerial Image Segmentation from Online Maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6054–6068, 2017.
19. E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark," *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2017-July, pp. 3226–3229, 2017.