Das fliegende Auge – Bild- und Videoauswertung für UAVs

Von Dr.-Ing. Tobias Schuchert, Fraunhofer IOSB, Karlsruhe

Der Film „Das fliegende Auge“ von 1983 hat die Wahrnehmung der Öffentlichkeit für die Gefahren der Überwachung aus der Luft geprägt. Der Film handelt von einem bemannten Superhubschrauber mit diversen Überwachungs- und Bekämpfungsfunktionen. Heute, 30 Jahre später, ist man dem Szenario des Films näher denn je. Die in dem Film gezeigten Überwachungsmöglichkeiten scheinen keine Science-Fiction mehr zu sein. Videokameras und Richtmikrofone mit hoher Genauigkeit sind genauso verfügbar wie Wärmebildkameras, mit denen man auch begrenzt durch Wände schauen kann. Allerdings werden heutzutage keine Superhubschrauber benötigt, sondern vielmehr können selbst Privatleute günstige miniaturisierte Fluggeräte (auch Drohnen genannt) erwerben, die über Bild- und Videokameras oder sogar Wärmebildgeräte verfügen. Aufgrund von gesetzlichen Bestimmungen und um dem negativen Image des Wortes "Drohne" entgegenzuwirken, existieren unterschiedliche Synonyme für unbemannte Luftfahrzeuge.

Neben UAV (Unmanned/Uninhabited Aerial Vehicle, deutsch: unbemanntes Luftfahrzeug) wird in letzter Zeit das Synonym RPAS (Remotely Piloted Aircraft System) verwendet. Dabei wird einerseits das Gesamtsystem (Fluggerät, Pilot, Bodenkontrollstation und Kommunikation) betrachtet, als auch betont, dass das Luftfahrzeug von einem Piloten gesteuert wird, d.h. dass das Luftfahrzeug gar nicht oder nur begrenzt autonom fliegen kann und der Pilot stets die Kontrolle hat. RPAS umfasst somit eine Untergruppe der UAVs und schließt insbesondere autonome Systeme aus. Die Bild- und Videoauswertung ist relativ unabhängig von der Art der Steuerung des Fluggerätes, so dass im Verlauf dieses Artikels der Begriff UAV verwendet wird. Vielmehr sind die hier vorgestellten Verfahren auch für bemannte Systeme nutzbar.
UAV werden heutzutage bereits in vielen Bereichen eingesetzt. Die meisten UAVs sind in privater Hand und werden für private Zwecke verwendet, z.B. für Bild- oder Videoaufnahmen aus der Luft. Weiterhin existieren viele kommerzielle Anwendungen, z.B. Inspektions-, Vermessungs- oder Dokumentationsaufgaben, sowie militärischen Anwendungen wie Aufklärung und Überwachung. Auch bei Polizeibehörden sind UAVs bereits im Einsatz bzw. werden getestet, vor allem in den Bereichen Schutz und Sicherheit. Dazu zählen Aufgaben wie z.B. Straßen- und Verkehrsüberwachung, Überwachung/Kontrolle bei Großveranstaltungen oder Strafverfolgung, Tatortaufnahme oder Tracking / Verfolgung von verdächtigen Personen oder Objekten.

Sensorik

So vielfältig die Anwendungen sind, so vielfältig sind auch die verfügbaren UAVs. Diese lassen sich in unterschiedliche Kategorien wie z.B. Größe, Reichweite, Ausdauer, Flughöhe oder Nutzlast einteilen. Neben den rechtlichen Vorschriften und den Kosten zum Erwerb und Betrieb einer Drohne, sind die Sensorik und die anschließende Signalauswertung ein wesentlicher Aspekt für das Erreichen des gewünschten Ergebnisses. Dabei wird zwischen Sensorik zur Steuerung der Drohne, z.B. GPS, INS (inertiales Navigationssystem) und Videokameras, und Sensorik für die eigentliche Anwendung unterschieden. Bei der Anwendungssensorik spielen besonders die Nutzlast, die Kommunikation und die Stromversorgung des UAV eine Rolle, so dass es sich dabei in den meisten Fällen um Foto- oder Videokameras handelt. Die fortschreitende Miniaturisierung hat dazu geführt dass heutzutage kleine, leichte und hochaufgelöste Kameras zu Verfügung stehen.
Hier ist vor allem interessant welche Anforderungen die jeweilige Anwendung an die Sensorik und damit auch an das UAV stellt. Ist ein Livebild nötig oder reicht es, wenn die Daten nachträglich ausgewertet werden? Sind Videos nötig oder reichen Einzelbilder? Wie ist die Beleuchtungssituation (Tag/Dämmerung/Nacht)? Wie hoch muss die Auflösung auf welche Entfernung sein? Wird eine 3D Rekonstruktion der Umgebung ggfs. in Echtzeit benötigt?

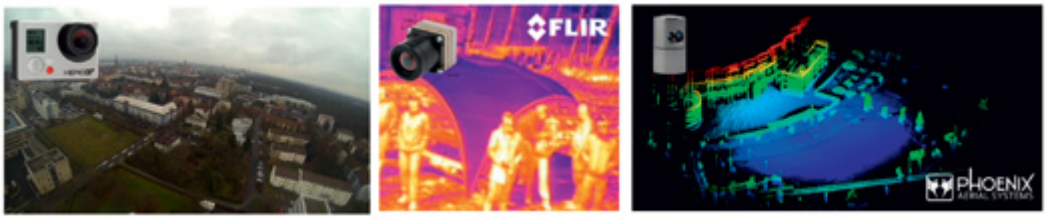

Abbildung 1: Sensorik für UAV und Beispielaufnahmen. Links: GoPro-Actionkamera [1], Mitte: FLIR Wärmebildgerät [2], Rechts: VeloDyne LiDAR-Sensor [3,4].

Modulare Systeme ermöglichen es, Sensorik und Trägersysteme untereinander auszutauschen, so dass mit einer begrenzten Auswahl an Komponenten eine Vielzahl an Anwendungsfeldern abgedeckt werden kann, allerdings nicht gleichzeitig. Bei der Sensorik wird vor allem zwischen der bildgebenden Sensorik (2D) und Sensorik zur Generierung von 3D Modellen (Tiefenbildkameras oder Laserscanner) unterschieden. Die bildgebende Sensorik lässt sich weiterhin hinsichtlich ihrer spektralen Eigenschaften charakterisieren. Standardkameras decken einen Teil des sichtbaren Lichtspektrums ab. Im Vergleich dazu erlauben andere Kameras den Bereich im nahen oder fernen Infrarot (IR-Kamera) oder sogar mehrere Bereiche gleichzeitig (Multi- bzw. Hyperspektralkamera) abzubilden. Multi- oder Hyperspektralkameras bieten die Möglichkeit unterschiedliche Spektralbereiche simultan zu erfassen, z.B. im sichtbaren und im unsichtbaren Lichtspektrum. Dies ist insbesondere für die Landwirtschaft interessant um Informationen über den Zustand von Agrarflächen zu erhalten und wird daher hier nicht näher betrachtet [5].
Abbildung 1 zeigt Beispiele für unterschiedliche Sensorik, die auf aktuellen (Mikro-)UAV, d.h. bis zu einem Gewicht von ca. 5 kg, eingesetzt wird. Auf der linken Seite sieht man eine GoPro HD Hero 3+ -Actionkamera [1], die eine weite Verbreitung bei zivilen UAVs hat. Ursprünglich für Aufnahmen von Surfern und anderen Freizeitsportlern entwickelt, wird sie auf Grund ihres Gewichts, Robustheit und Einfachheit geschätzt. Die Kamera hat standardmäßig ein relativ weitwinkliges Objektiv, was sich gut für Übersichtsbilder eignet. Des Weiteren bietet die aktuelle HD Version der GoPro-Kamera Videos in sehr hoher Auflösung (4096x2160 Pixel (4K) bei 15 Bilder/s) und Fotos mit 12 MP.
Bei Anwendungen, die auf der Messung von Wärmeverteilungen basieren, z.B. zur Suche/Verfolgung von Personen oder zur Untersuchung/Dokumentation von Wärmedämmungen, können thermische Infrarotkameras verwendet werden. Abbildung 1 Mitte zeigt das ungekühlte Mikrobolometer Quark 640 der Firma FLIR [2], sowie ein Bildbeispiel. Die Auflösung der Kamera beträgt 640 x 512 Pixel bei 30 Hz, d.h. deutlich niedriger als die Auflösung der GoPro. Im Gegensatz zu Kameras für den sichtbaren Bereich, ermöglichen es thermische Infrarotkameras die Wärmeverteilung auch in dunklen Szenarien darzustellen, z.B. bei Nacht oder in Höhlen.
Werden 3D-Informationen der Umgebung benötigt, z.B. um Gebäude zu vermessen, können unterschiedliche Varianten verwendet werden. Einerseits können aus den Bildern von normalen Kameras Informationen gewonnen werden, um 3D Modelle der Umgebung zu erstellen. Dies hat den Vorteil, dass keine spezielle Hardware benötigt wird. Allerdings ist die Genauigkeit der 3D Rekonstruktion begrenzt und muss normalerweise im Nachhinein berechnet werden. Andererseits können spezielle Sensoren zur Erstellung von 3D-rekonstruktionen verwendet werden. Neben Stereokameras, und Tiefenkameras, wie der Microsoft Kinect, ist es auch möglich sogenannte LiDAR-Sensoren mit UAVs zu verwenden. LiDAR steht für „Light detection and ranging“ und ermöglicht es, ähnlich eines Radars, die Entfernung von Objekten mit Laserstrahlen zu bestimmen. In dem Fall des in Abbildung 1 dargestellten VeloDyne-Sensors HDL-32E [3], rotiert ein Spiegelsystem sehr schnell 360° um die eigene Achse und tastet auf diese Weise die Umgebung mit Hilfe von 32 Lasern ab. Das Beispielbild zeigt die für den Sensor charakteristischen Kreisstrukturen, die durch die vertikale Anordnung der Laser erzeugt wird. Der Sensor liefert 3D Informationen bis zu einer Entfernung von 70 Metern mit einer Genauigkeit von 2 cm bei einem Gewicht von unter 2 kg. Die geringen Ausmaße erlauben die Nutzung dieses Sensors an UAVs, was Phoenix Aerial Systems [4] in einem Testprojekt nachgewiesen hat. Da LiDAR-Sensoren relativ teuer sind, wird in der wissenschaftlichen Gemeinde eher mit günstigeren Sensoren zur Messung der Entfernung eines Objektes gearbeitet, wie der Microsoft Kinect oder ToF-Kameras. Diese sind allerdings auf Grund ihrer speziellen Eigenschaften (Beleuchtungsabhängigkeit/Auflösung, Reichweite) nur bedingt für den realen Einsatz mit UAVs geeignet.
Im Hinblick auf Sensorik für UAVs ist es auch wichtig, wie die Sensorik an der Drohne befestigt ist. Ist die Kamera fixiert oder lässt sich die Kamera von dem UAV getrennt steuern? Ist die Aufhängung des Sensors stabilisiert?

Bild- und Videoauswertung

Neben der Sensorik spielt die Auswertesoftware eine wesentliche Rolle bei der (Live-) Auswertung der Daten. Der beste Sensor kann nutzlos sein, wenn eine schlechte Auswertesoftware genutzt wird. Neben Verfahren zum Tracking von Personen und Objekten, wie z.B. Autos, sind bei Bildern und Videos von UAVs besonders Verfahren zur Verbesserung des Bildes erforderlich. Das liegt daran, dass einerseits das UAV sich bewegt und diese Bewegung in Verbindung mit einer zu großen Belichtungszeit zu einer gewissen Unschärfe der Bilder kommen kann. Andererseits ist der Speicherplatz für Daten begrenzt bzw. müssen die Daten per Funk übertragen werden. Um dennoch eine hohe Auflösung zu haben, bzw. ein Livebild über die Funkverbindung zu versenden, werden die Daten stark komprimiert, z.B. mit JPEG oder MPEG. Dadurch entstehen Komprimierungsartefakte, die abhängig vom Kompressionsgrad die Analyse des Bildes mehr oder weniger erschweren können. Um dennoch wichtige

Informationen aus den Daten zu gewinnen, wird Bild- und Videoauswertesoftware verwendet, die mit Hilfe intelligenter Verfahren versucht die verlorene Information wieder zurückzugewinnen. Neben den erwähnten Bildstörungen, den sogenannten Artefakten, gibt es noch eine Vielzahl an weiteren Störungen, die auch von der gewählten Sensorik abhängen können. Bei mobilen Kameras mit CMOS-Sensor-Technik, die in vielen aktuellen Videokameras und Handys verwendet werden (auch in der Actionkamera GoPro), tritt insbesondere der sogenannte Rolling-Shutter-Effekt auf. Dies hängt damit zusammen, dass die Daten vom CMOS-Sensor zeilenweise ausgelesen werden und es dadurch bei bewegten Objekten oder bewegter Kamera zu Bildverzerrungen kommen kann. Kameras mit CCD-Sensor haben auf Grund der unterschiedlichen Funktionsweise kein Rolling-Shutter Problem. Alternativ gibt es auch Global-Shutter CMOS-Kameras, die allerdings preislich höher angesiedelt sind. Schließlich spielt auch die Beleuchtung eine wichtige Rolle. Überbelichtung, d.h. die Kamera wird geblendet und ein Bereich des Bildes wird weiß, kann Bildmaterial unbrauchbar machen, starke Beleuchtungsänderungen erschweren die Auswertung der Daten, z.B. zum Tracking von Objekten, und eine geringe Beleuchtung führt zu nicht sichtbaren/verfolgbaren Objekten, bzw. zu einer längeren Belichtungszeit und dadurch wieder zu mehr Bewegungsunschärfe. Die meisten Bildartefakte (starke Kompression, Beleuchtungsänderungen, Bewegungsunschärfe und niedrige Auflösung) sind unabhängig von der Anwendung bei UAVs, sondern gelten allgemein für Bild- und Videoaufnahmen. Im Allgemeinen existieren drei Arten von Verfahren zur Behebung bzw. zur Minimierung dieser Artefakte: (1) Verfahren, die die Informationen im Bild nutzen. Hierbei wird z.B. eine Kante detektiert und entsprechend bearbeitet. (2) Verfahren, die Informationen aus einem Video nutzen. Dabei wird u.a. die im Video sichtbare Bewegung zwischen Bildern berechnet und genutzt. (3) Lernbasierte Verfahren, die mit ähnlichen Bildern trainiert werden, damit sie „lernen“, wie sich Artefakte auf Bilder auswirken. Mit diesem Wissen lassen sich dann theoretisch umgekehrt bei bekannten Artefakten die ungestörten Bildinhalte wieder rekonstruieren.

Distante Biometrie

Bei der sog. distanten Biometrie geht es darum Personen in großer Entfernung zu detektieren, (wieder)zuerkennen und zu identifizieren. Detektieren bedeutet in dem Fall, dass eine Person in einem Bild erkannt wird, d.h. Kopf, Augen oder Mund werden gefunden. Dies ist im Allgemeinen ab einer Kopfgröße von ca. 6x6 Pixeln möglich. Die (Wieder-)Erkennung ist ab einer Kopfgröße von ca. 25x25 Pixeln möglich, und bedeutet, dass eine Person in mehreren Bildern an charakteristischen Merkmalen erkannt werden kann, z.B. sogenannte „soft“-biometrische Merkmale wie Frisur oder Brille. Für die Identifikation ist schließlich eine Kopfgröße von ca. 120 x 120 Pixel nötig. Die Identifikation erlaubt eine dauerhafte eindeutige Zuordnung, z.B. durch Gesichtserkennungsverfahren.

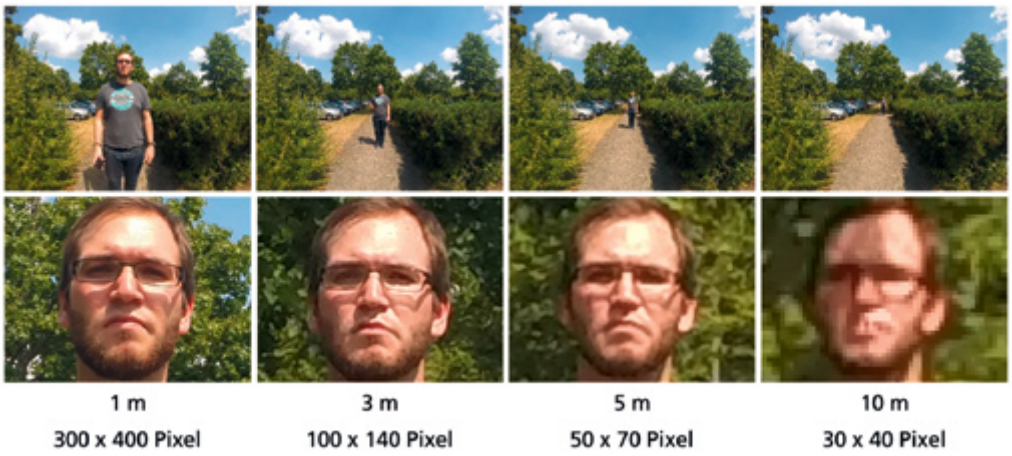

Abbildung 2: Darstellung der Gesichtsgröße in Pixeln bei unterschiedlichen Entfernungen zur Kamera bei 12 MP Auflösung. Oben: Originalbild, unten: Vergrößerung des Bildausschnittes mit Gesicht. Die Werte unter den Bildern geben die Entfernung in Metern und die Größe des Gesichtes in Pixeln an.

Die Größenangaben sind nur Richtwerte und hängen insbesondere von dem gewählten Sensor und der verwendeten Optik ab. Abbildung 2 zeigt beispielhaft die Auflösung eines Gesichtes aus unterschiedlichen Entfernungen. Die Aufnahmen wurden mit einer GoPro Hero3+ und einer Auflösung von 4000 x 3000 Pixel und einem Sichtfeld von ca. 120° aufgenommen. Hierbei ist zu berücksichtigen, dass das Sichtfeld, auch FoV (Field-of-View) genannt, bei der GoPro verhältnismäßig groß ist. Bei einem kleineren FoV von z.B. 60° würde sich die Anzahl der Pixel die das Gesicht abdecken mehr als verdoppeln. In Abbildung 2 lässt sich erkennen, dass mit den gewählten Einstellungen eine Person bis ca. 3m identifizierbar und bis ca. 10m wiederzuerkennen ist. Befindet sich die Person in einer größeren Entfernung, kann man zwar noch eine Person detektieren, aber nicht mehr einer Identität zuordnen. Bei der Verwendung von UAVs ist noch darauf zu achten, dass im Normalfall von oben bzw. aus einem steilen Winkel auf die Personen geschaut wird. Dadurch ist das Gesicht einer Person deutlich schlechter zu erkennen, so dass eine Identifizierung einer Person mit dieser Ausstattung praktisch nicht möglich ist. Im Falle einer Wiedererkennungsaufgabe lassen sich in dem Fall allerdings noch statische und dynamische Merkmale der kompletten Person nutzen, z.B. Kleidung/Rucksack (statische Merkmale) oder Gangart/Verhalten (dynamische Merkmale), wobei dynamische Merkmale meist nur in Videos detektiert werden können.

Bewegungsunschärfe

Bewegungsunschärfe (engl. Blur) von Bildbereichen kann durch längere Aufnahmebelichtungszeiten, (schnelle) Bewegung von Objekten oder Kamerabewegung hervorgerufen werden. Beim sog. Deblurring wird versucht diese Bewegungsunschärfe in Bildern zu reduzieren. Abbildung 3 zeigt eine verwackelte Aufnahme des Colloseums in Rom. Rechts daneben ist das gleiche Bild nach der Durchführung des Deblurrings nach [6] dargestellt. Beim deblurren von Bildern wird versucht die Bewegung der Kamera bzw. des Objektes bei der Aufnahme zu schätzen, um anschließend die inverse Bewegung anzuwenden, um ein schärferes Bild zu erhalten. Allerdings ist das Schätzen dieser Bewegung äußerst komplex und daher ist der Rechenaufwand enorm und dazu scheitern (noch) die meisten Methoden bei komplexeren realen Szenen.

Abbildung 3: Bildbeispiel: Deblurring eines Bildes [6].

Rauschen

Denoising (Entrauschung) bezeichnet den Prozess, Rauschen in Bild- oder Videodaten zu minimieren bzw. im Idealfall ganz zu entfernen. Die Ursachen von Rauschen bei Bildern und Videos sind vielfältig. Neben dem Rauschen des Sensors, welches bei schlechter Beleuchtung verstärkt sichtbar wird, sind vor allem verlustbehaftete Komprimierungsverfahren, wie z.B. JPEG, die Quelle für schlechte Bildqualität. Abbildung 4 zeigt beispielhaft, wie sich diese beiden Rauscharten auf ein Bild auswirken und wie die Bildausschnitte nach der Bearbeitung mittels eines aktuellen Verfahrens [7] aussehen. Die beiden Ausschnitte der bearbeiteten Bilder zeigen, dass das Rauschen bzw. die Kompressionsartefakte deutlich reduziert werden. Allerdings erscheinen die bearbeiteten Bilder auch weicher, d.h. mehr geglättet, im Gegensatz zum Originalbild. Wie auch zum Deblurring existiert eine Vielzahl an wissenschaftlichen Arbeiten zum Denoising von Bildern. Im Gegensatz zu den Deblurring-Verfahren, zeigen aktuelle Verfahren zur Rauschreduktion teils
erstaunlich gute Ergebnisse. Ein Kernproblem der Verfahren ist weiterhin, die feinen Details im Bild von Rauschen zu unterscheiden und somit nicht zu unterdrücken.

Abbildung 4: Links: Originales Bild bzw. vergrößerter Bildausschnitt. Mitte: Bild mit simuliertem Sensorrauschen und mit [7] entrauschtes Bild. Rechts: Bild mit Kompressionsartefakten und mit [7] entrauschtes Bild.

Niedrige Auflösung

In bestimmten Fällen ist es gewünscht eine Person oder ein Objekt aus großer Entfernung aufzunehmen. Dies ist besonders bei der Verwendung von UAVs von Interesse, da diese einerseits in einigen Metern Höhe fliegen und andererseits auf Grund des Gewichts und der Ausmaße keine großen Objektive verwenden. Wie bereits erwähnt, ist mit einer aktuellen Action-Kamera schon aus einer Entfernung von knapp 10 Metern eine (Wieder)erkennung von Personen nur noch schwer möglich. In diesen Fällen sind Verfahren zur Auflösungserhöhung interessant. Dabei wird versucht, die im Bild fehlende Information zu schätzen. Abbildung 5 zeigt, wie sich die Erhöhung der Auflösung mit dem Verfahren [8] auf eine Gesichtsaufnahme auswirkt. Bei diesem Verfahren werden Bild- und Bewegungsinformationen aus einem niedrig aufgelösten Video genutzt, um ein höher aufgelöstes Bild des Gesichtes zu berechnen. Die Herausforderung insbesondere bei Gesichtern ist, dass z.B. Sprechen oder Lachen zu komplexen Muskelbewegungen führt, die die Rekonstruktion erschweren.

Abbildung 5: Links: Niedrig aufgelöstes Bild (100 x 120 Pixel). Mitte: Aus Video berechnetes hochaufgelöstes Bild nach [8]. Rechts: Vergleich zweier Bildausschnitte (oben: Originalbild, unten: verbessertes Bild).

Fazit

Schnelle Fortschritte in der Technologie ermöglichen immer neue Anwendungsfelder für UAVs. Diese Entwicklungen führen auch zu sinkenden Preisen, so dass „hochwertige“ UAVs bereits heute für Privatleute verfügbar sind. In den meisten Fällen werden Bild- oder Videokameras genutzt, so dass die Probleme bei der Auswertung der Daten denen von Standard-Kameras ähneln. Zu nennen sind insbesondere Artefakte durch Bewegungsunschärfe oder durch Rauschen. Aktuelle Kameras erlauben bereits „distante Biometrie“, wobei die Entfernungen auf denen Personen und Objekte (wieder-)erkannt bzw. identifiziert werden können stetig größer werden. Neben der Hardware spielt die Software eine wesentliche Rolle bei der Bild- und Videoauswertung. Die hier vorgestellten Verfahren ermöglichen es zusätzliche Informationen aus Bildern und Videos zu gewinnen und damit die Bildqualität zu verbessern. Diese Verfahren werden am Fraunhofer IOSB erforscht und auf ihre Praxistauglichkeit untersucht.

Literatur

[1] http://gopro.com/
[2] http://www.flir.com/
[3] http://www.velodynelidar.com/
[4] http://www.phoenix-aerial.com/
[5] http://www.tetracam.com/Products_Hardware_Accessories.htm#Books
[6] Xu, L., Jia, J., “Two-Phase Kernel Estimation for Robust Motion Deblurring”, ECCV 2010.
[7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. “Image denoising by sparse 3d transform-domain collaborative ?ltering”. IEEE TIP, 16(82):3736–3745, 2007.
[8] Schuchert, T., Oser, F., “Optical flow estimation with confidence measures for super-resolution based on recursive robust total least squares”, ICPRAM 2012.

Autor

Dr. Tobias Schuchert ist Leiter der Gruppe „HPC“-Videoauswertung am Fraunhofer IOSB in Karlsruhe. Nach dem Studium der Informationstechnik an der TU Dortmund forschte er von 2005 bis 2009 am Forschungszentrum Jülich, wo er zum Thema: 3D

Rekonstruktion und Bewegungsschätzung promovierte. Seit 2009 ist er am Fraunhofer Institut in Karlsruhe angestellt. Der Fokus von Dr. Tobias Schuchert liegt auf der Videoauswertung von großen Datenmengen, insbesondere von Videoaufnahmen von UAVs. Weitere Themenfelder sind Bildverbesserung, Aktivitätserkennung, Georegistrierung und 3D Rekonstruktion. Er ist Autor und Co-Autor von mehr als 20 wissenschaftlichen Artikeln.

© Verlag Deutsche Polizeiliteratur