

# Überwachung von Industrierobotern mittels Nahinfrarot-Kamerasystem

Sebastian Sporrer, Norbert Jung

Hochschule Bonn-Rhein-Sieg, Institut für Sicherheitsforschung

## Zusammenfassung

Diese Arbeit präsentiert eine Methode zur zuverlässigen Personendetektion für die Absicherung des Arbeitsbereichs von Industrierobotern. Hierzu wird ein im Nahinfrarotbereich (NIR) arbeitendes aktives Kamerasystem eingesetzt, das durch erweiterte und robuste Hauterkennungseigenschaften besonders dazu geeignet ist, zwischen verschiedensten Materialoberflächen und menschlicher Haut zu unterscheiden. So soll zum einen die Erkennungsleistung gegenüber handelsüblichen, im visuellen Bereich arbeitenden RGB-Kamerasystemen gesteigert werden und gleichzeitig eine „intelligente“ Form des *Mutings* realisiert werden. Die im Rahmen des Projekts „Sichere Personendetektion im Arbeitsbereich von Industrierobotern durch ein aktives NIR-Kamerasystem (SPAI)“ entwickelte und hier vorgestellte Methode erreicht in einer ersten Variante eine pixelweise Personenerkennungsrate von ca. 98,16%.

## 1 Einleitung

Der Arbeitsbereich eines Industrieroboters birgt für Menschen, die diesen bewusst oder unbewusst betreten, große Risiken. Die Kraft und Geschwindigkeit solcher Roboter kann durch Kollisionen oder Quetschungen zu schlimmstenfalls tödlichen Verletzungen führen. Im Jahre 2015 sorgte der tödliche Unfall in einem Werk von Volkswagen für mediale Aufmerksamkeit (Kühling, 2015). Nach Angaben der DGUV wurden in den Jahren 2005 bis 2012 insgesamt 2269 Arbeitsunfällen mit Industrierobotern registriert, von denen 59 als „schwer“ - z.B. Verlust von Gliedmaßen, Tod - eingestuft wurden (DGUV, 2015). Es ist daher wünschenswert einen Schutzraum um den Roboter zu definieren und auf das Betreten durch Personen zu überwachen, um durch angemessene Reaktionen einen sicheren Zustand herzustellen. Dieser Schutzraum soll besonders mit Hinblick auf zukünftige Szenarien inspiriert durch die *Smart Factory* flexibel zu gestalten sein. Die bisher obligatorischen Zäune und Lichtgitter sind besonders bei dieser Anforderung einem Kamerasystem unterlegen, bei dem der Schutzraum idealerweise unmittelbar durch eine zugehörige Software interaktiv definiert wird. Auch eine weiterführende Objekterkennung spielt in diesem Szenario eine große Rolle: Am Markt befindliche Kamerasysteme detektieren im definierten Schutzbereich alle bewegten Objekte und

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2016 in  
B. Weyers, A. Dittmar (Hrsg.):  
Mensch und Computer 2016 – Workshopbeiträge, 4. - 7. September 2016, Aachen.  
Copyright © 2016 bei den Autoren.  
<http://dx.doi.org/10.18420/muc2016-ws01-0008>

interpretieren diese als Sicherheitsrisiko. Aufgrund der fehlenden Unterscheidung von Personen und anderen Objekten führt dieses Verhalten zu unerwünschten Effekten, wie zum Beispiel dem Erkennen des Roboters selbst. Durch aktive Nahinfrarot-Kamerasysteme ist es möglich mittels robuster Hauterkennung sicher zwischen Personen und anderen Objekten zu unterscheiden und so eine intelligente Form des *Mutings* zu realisieren.

## 2 Grundlagen

Das in dieser Arbeit vorgestellte Verfahren nutzt ein robustes Verfahren zur pixelweisen Hauterkennung mittels „spektraler Signaturen“ im Nahinfrarotbereich. Zur Akquise der zugehörigen multispektralen Bilddaten wird ein spezielles aktives Kamerasystem verwendet. Beide Themen werden in diesem Abschnitt im Detail vorgestellt.

### 2.1 Hauterkennung im nahen Infrarot

Menschliche Haut kann in sechs Hauttypen eingeteilt werden (Fitzpatrick 1988), die sich insbesondere im sichtbaren Licht durch ihre Farbe unterscheiden. Dies wird durch verschiedene Konzentrationen von Pigmenten (Melaninen) in der Haut verursacht. Im nahinfraroten Bereich des elektromagnetischen Spektrums verlieren Melanine zunehmend an Einfluss auf die Remissionseigenschaften der Haut, wodurch Haut und andere Materialoberflächen in diesem spektralen Bereich weitestgehend entfernungs- und intensitätsunabhängig robust zu unterscheiden sind. Hierzu sind für die meisten Anwendungsfälle lediglich einige wenige schmale Bänder aus dem Bereich nötig, die als Stützstellen des Remissionsspektrums verstanden werden können und als „spektrale Signatur“ bezeichnet werden. In Abbildung 1 werden verschiedene Remissionsspektren zusammen mit einem von den verschiedenen Hauttypen aufgespannten Bereich dargestellt. Während Haut besonders im sichtbaren Bereich pigmentbedingte Remissionscharakteristika aufweist und von anderen Materialien wie beispielsweise Holz nur schwer unterschieden werden kann, verjüngt sich der Korridor im Nah-Infrarotbereich und weist eine Treppenstufen-Charakteristik auf, mit der eine robuste Unterscheidung möglich ist. Diese Charakteristik ist im Wesentlichen auf den Wassergehalt und die spezifische Reflektion von Nahinfrarotlicht durch die Haut zurückzuführen. Sie ist von Hauttyp, Alter und Geschlecht unabhängig und wurde bereits von (Jacquez et al. 1955) untersucht. Diese Eigenschaft der Haut wurde für technische Systeme bereits in mehreren Arbeiten auf verschiedene Arten umgesetzt (Hacskaylo 1985; Kilgore & Whillock 2008; Determan & Wunderlin 2008, Zhang et al. 2011). Ein Vergleich zwischen Hautdetektionssystemen basierend auf Erkennung im visuellen Spektrum und nahem Infrarot wurde jüngst von (Mendenhall et al. 2015) veröffentlicht.

Robuste Hauterkennung mittels „spektraler Signaturen“ wird durch die Forschungsgruppe an der Hochschule Bonn-Rhein-Sieg bereits seit über 10 Jahren erforscht. So wurden in dieser Zeit verschiedene Forschungsprojekte durchgeführt, in deren Rahmen diese Technik für verschiedene Anwendungsfälle – meist mit direktem Bezug zur Arbeitssicherheit – untersucht wurde. Im Zuge der Projektarbeiten wurden dabei mehrere Veröffentlichungen zum Fortschritt verfasst, die in (Schwaneberg 2013) zusammengefasst sind.

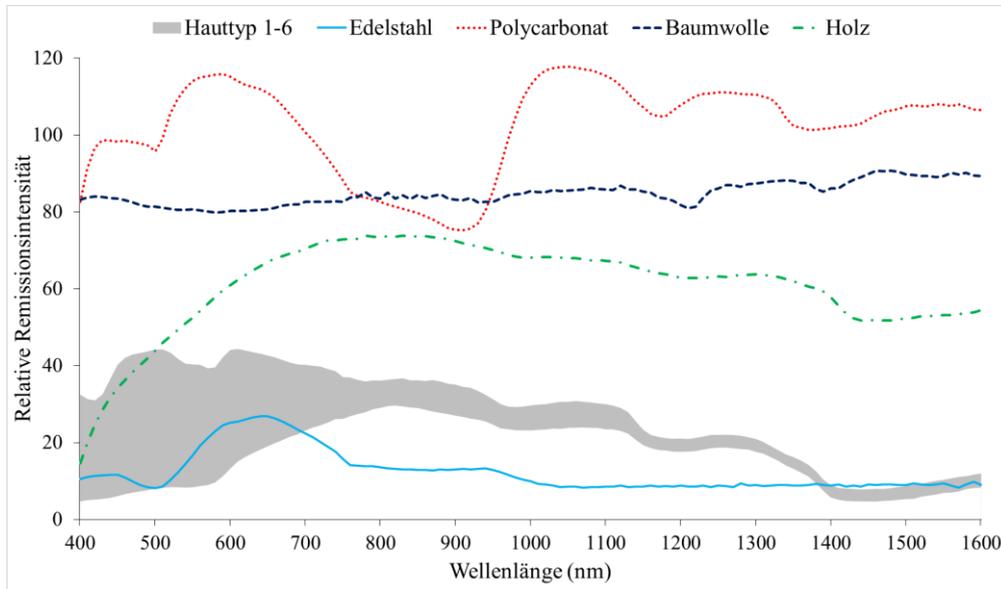


Abbildung 1: Remissionsspektren verschiedener Materialien und von sechs Hauttypen aufgespannter spektraler Hautkorridor im sichtbaren und nahinfraroten Spektrum (Sporrer et al. 2015).

## 2.2 Aktives Nahinfrarot-Kamerasystem

Auch bildgebende Verfahren wurden für den NIR-Bereich vorgeschlagen. Verschiedene Forscher nutzen mehrere NIR-Kameras mit vorgeschalteten optischen Bandpassfiltern und halbdurchlässigen Spiegeln um multispektrale Bilder zu erzeugen (Pavlidis et al. 2000; Dowdall et al. 2003; Peskosky 2010). Einen weiteren Ansatz verfolgten Bourlai et al. 2012) mit der Verwendung eines Filtrrades, das mit unterschiedlichen optischen Bandpassfiltern ausgestattet wurde und erzeugen die einzelnen spektralen Kanäle sequenziell. (Bertozzi et al. 2013) beschreiben die Verwendung eines Bildsensors, der ähnlich zu dem im visuellen Bereich etablierten *Bayer-Filter* mit einem Muster pixelweise verschiedener NIR-Bandpassfilter ausgestattet ist. In der Forschungsgruppe an der Hochschule Bonn-Rhein-Sieg wurde ein aktives NIR-Kamerasystem entwickelt (Steiner et al. 2016a), welches für die Bildakquise im Rahmen der Evaluation dieser Arbeit verwendet wurde und im Folgenden kurz beschrieben wird.

Das Kamerasystem umfasst eine im Nahinfrarotbereich empfindliche digitale Kamera mit einem InGaAs-Bildsensor. Um den Signal-zu-Rausch-Abstand zu optimieren, wird der Sensor thermoelektrisch gekühlt und bietet im Gegensatz zu verbreiteten Silizium-Bildsensoren lediglich eine geringe Auflösung von  $640 \times 512$  Pixeln. Die weitere wichtige Komponente stellt ein Blitzlicht bestehend aus möglichst geeignet verteilten verschiedenen NIR-LEDs dar, das für die aktive Beleuchtung der Szene mit verschiedenen definierten Wellenlängenbändern eingesetzt wird. Sonstige Unterschiede optischer und leistungstechnischer Natur werden bestmöglich ausgeglichen und auf ein Minimum reduziert, um eine gleichmäßige Ausleuchtung der Szene zu ermöglichen. Abbildung 2 zeigt ein solches Kamerasystem schematisch. Die

zentrale Steuerung des Messablaufs wird durch einen in das Blitzlicht eingebetteten Mikrocontroller koordiniert. Nach dem Startsignal, welches durch den auswertenden Computer initiiert wird, werden für jeden LED-Typus die zugehörigen LEDs eingeschaltet und der Auslöser der Kamera über eine externe Schnittstelle betätigt. Das aufgenommene Bildmaterial wird dann zum Computer gesendet und mit der Wellenlängenbandinformation des momentan aktiven Typus markiert. Dies wird für alle Kanäle wiederholt. Insbesondere wird auch ein Bild ohne aktive Beleuchtung zur Identifikation des Umgebungslichtes aufgenommen. Um ein multispektrales Bild mit allen Kanälen zu akquirieren wird damit die Bildwiederholrate der Kamera bei  $N$  Wellenlängenkanälen um den Faktor  $1/(N+1)$  verlangsamt.

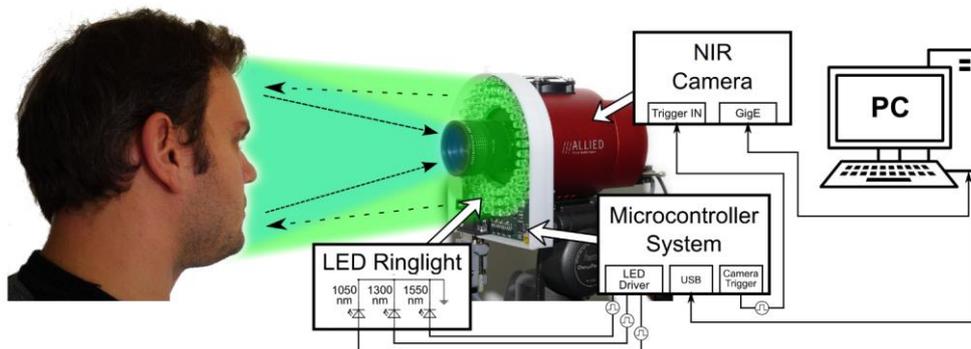


Abbildung 2: Schematische Darstellung des NIR-Kamerasystems (Steiner et al. 2016a).

Zur Tiefenerfassung, die für das vorgestellte Szenario zur dreidimensionalen Verortung von erkannten Objekten nötig ist, wird das System als stereoskopisch ausgelegt, so dass mindestens zwei der erwähnten NIR-Kameras Verwendung finden. Zur Disparitätsbestimmung zwischen den Bildern der beiden Kameras werden gängige und bewährte Verfahren benutzt, die auch im sichtbaren Spektralbereich Verwendung finden. Näheres zur Untersuchung des stereoskopischen Aufbaus kann (Sporrer et al. 2015) entnommen werden.

Eine weitere Herausforderung bei der sequenziellen Multispektralbildakquise stellen Bewegungsartefakte dar, die an Rändern von sich bewegenden Objekten durch den räumlichen Versatz der Pixel zwischen zwei Kanälen auftauchen. Eine umfassende Untersuchung zu diesem Thema wurde in (Steiner et al. 2016b) durchgeführt.

Die Trennungsleistung der pixelweisen Hauterkennung wurde im Rahmen von (Steiner et al. 2016a) analysiert. Auch mit einem aktiven Nahinfrarot-Kamerasystem ist die robuste Hauterkennung nach dem in 2.1 vorgestellten Prinzip möglich. Für diese Untersuchung wurde neben verschiedenen Materialien auch typische Berufsbekleidung in die Datenbasis integriert. Abbildung 3 zeigt Porträtaufnahmen der sechs Hauttypen nach (Fitzpatrick 1988). Im oberen Bereich werden Bilder aus dem sichtbaren Bereich dargestellt, im unteren Bereich NIR-Falschfarbenbilder. Für diese Falschfarbenbilder wurden die Nahinfrarotkanäle als Rot, Grün und Blau interpretiert.



Abbildung 3: RGB- (o.) und NIR-Falschfarbenporträts von Individuen der Hauttypen 1 bis 6 (Steiner et al. 2016a).

### 3 Pixelweise Personenerkennung

Die pixelweise Hautinformation wird im folgenden Personendetektionsalgorithmus verwendet, um die Zugehörigkeit von Objekten zur Klasse „Person“ zuverlässig zu erkennen. Nach einer Aufbereitung der aufgenommenen multispektralen Bilddaten, die neben verschiedenen Verfahren zur Kalibrierung der intrinsischen Parameter der Kamera auch einen Weißabgleich bietet, werden drei Operationen parallel durchgeführt. Zum einen wird mit dem in 2 beschriebenen Verfahren zur Hauterkennung eine Hautmaske basierend auf den aktuellen Bilddaten erzeugt. Bei diesen Masken handelt es sich um binäre Bilder, bei denen Pixel mit dem Wert 0 (Schwarz) keine Haut darstellen und Pixel mit dem Wert 1 (Weiß) auf Haut hindeuten. Bei der zweiten parallelen Operation handelt es sich um eine Hintergrund-/Vordergrundsegmentierung, mit der eine Vordergrundmaske erstellt wird, die Objekte im Vordergrund pixelweise



Abbildung 4: Zwischenergebnisse der Personendetektion an zwei Beispielen - Ausgangsbild, Hautmaske, Vordergrundmaske, Tiefendaten, Personenmaske (v.l.)

markiert. In der dritten parallelen Operation werden Tiefendaten aufgrund der beiden Kamerabilder berechnet und jedem Pixel ein Tiefenwert zugeordnet. Dies erfolgt über die Zuweisung eines Grauwertes, dessen Intensität mit der Entfernung abnimmt. Mit diesen drei Komponenten wird dann eine Personenmaske berechnet, indem das Vorhandensein von Hautpixeln in zusammenhängenden Vordergrundobjekten geprüft wird. Diese Bereiche werden dann mit der minimalen und maximalen Tiefe, repräsentiert durch die jeweiligen Graustufen, eingefärbt. Abbildung 4 (rechts) zeigt solche Personenmasken (minimale Entfernung) als Resultat der Zwischenschritte (Mitte) und der Ausgangsbilder (links) an zwei Beispielen. Insbesondere ist im oberen Beispiel zu erkennen, dass der Stuhl in der Mitte zwar als Objekt im Vordergrund entdeckt wird, jedoch nicht als Person interpretiert wird. Die Hintergrundsubtraktion erlaubt außerdem das Erkennen von Bereichen im Bild als Person, die aufgrund ungewöhnlicher Posen durch semantische Beziehungen schwer zu erkennen wären.

## 4 Evaluation

Zur Evaluation der erreichten Zuverlässigkeit des vorgestellten Personendetektionsverfahrens dienten 50 multispektrale Bilder, die in einer typischen Umgebung inklusive eines exemplarischen Industrieroboters während eines Feldversuchs erhoben wurden. Die gewählten Szenen beinhalten einzelne sowie mehrere Personen mit Alltagsbekleidung. Eine standardisierte Datenbank mit erforderlichen multispektralen NIR-Bilddaten ist den Autoren nicht bekannt. Zur Auswertung der pixelgenauen Erkennungsleistung wurden manuell solche Pixel in den Bildern markiert, die vom Algorithmus als personenzugehörig erkannt werden müssen. Zur Markierung der Nicht-Personenpixel wurde diese Auswahl invertiert. Um Probleme bei der manuellen Markierung am Rand der Personen und Überschneidungen mit dem Hintergrund oder anderen Objekten vorzubeugen, wurde ein kleiner Abstand zum Rand gelassen. Dabei wurde darauf geachtet, dass alle Extremitäten erfasst wurden (siehe Abbildung 5). Zur Bewertung der Erkennungsleistung wird das Verhältnis der durch das System erkannten Pixel zu allen markierten Pixeln ermittelt. Bei diesem Versuch wurden insgesamt ungefähr 98,16% aller Personenpixel als solche erkannt. Nicht-Personenpixel wurden zu ungefähr 1,91% fehlerhaft als Personenpixel erkannt (Falsch-Positive).



Abbildung 5: Beispiel händische Markierung von Personenpixeln.

## 5 Diskussion & Ausblick

In der Norm EN ISO 10218 wird für sicherheitsbezogene Teile zur Abstandsüberwachung bei kollaborierenden Robotern ein Sicherheits-Integritätslevel (SIL) von 2 gefordert. Die Leistung des gemessenen Systems auf der vorliegenden Datenbasis liegt mit 1,84% Fehlerrate über dem in der IEC 61508 definierten Grenzwert dieser Stufe von 1,0% und erfüllt so lediglich SIL 1. Dies ist einerseits auf technische Limitierungen, wie die Ausprägung der aktiven Beleuchtung,

und andererseits auf algorithmische Probleme, wie beispielsweise die Erkennung von einem Objekt als zwei getrennte durch eine fehlerhafte Hintergrundsubtraktion, zurückzuführen. Während die Beleuchtung durch entsprechenden Aufwand auf die jeweilige Anwendung optimiert werden kann, resultieren die Probleme der Hintergrundsubtraktion vermutlich aus dem eingeschränkten „Farbraum“ der multispektralen Bilddaten. Das Einbeziehen der Tiefendaten und anderer Informationen könnte hier Abhilfe schaffen.

Für die Falsch-Positiv-Rate von 1,91% zeigen sich hauptsächlich Situationen verantwortlich, in denen sich Personen und andere Objekte optisch überlappen oder berühren. Im vorgestellten Verfahren werden die Pixel des Objekts dabei der Person zugeordnet. Unter welchen Umständen eine getrennte Betrachtung von Person und Objekt in diesen Fällen möglich und gewünscht ist, bleibt zu spezifizieren. Ein weiteres Problem stellen die Auswirkungen von übersteuerten Pixeln auf benachbarte Pixel dar. Das sogenannte „Ausfransen“ (engl. „Blooming“) kann durch sehr starke Fremdlichtquellen, wie direkte Sonneneinstrahlung oder die direkte Reflexion der kamerasystemeigenen Lichtpulse, entstehen. Der Effekt kann möglicherweise durch geeignete Verfahren auf Bildsensorebene durch den Hersteller kompensiert werden.

Zur abschließenden Bewertung des hier vorgestellten Verfahrens muss, auch mit Hinblick auf zukünftige Verfahren, eine standardisierte Datenbasis erhoben werden, die für Leistungsvergleiche pixelweiser Personendetektionsverfahren im sichtbaren und nahinfraroten Bereich herangezogen werden kann. Eine weitere zentrale Fragestellung ist die Auswahl eines standardisierten Bewertungsverfahrens, ähnlich dem hier vorgestellten pixelweisen Ansatz, um die Qualität eines Personendetektionsverfahrens zu messen und darauf aufbauend normative Richtlinien bestimmen zu können.

### Danksagung

Wir danken der Deutschen Gesetzlichen Unfallversicherung (DGUV) für die finanzielle und fachliche Unterstützung bei der Durchführung des Projekts FF-FP 0357 „Sichere Personendetektion im Arbeitsbereich von Industrierobotern durch ein aktives NIR-Kamerasystem (SPAI)“.

### Literaturverzeichnis

- Bertozi, M., Fedriga, R., Miron, A. & Reverchon, J. L. (2013). Pedestrian detection in poor visibility conditions: Would SWIR help?. In Petrosino, A. (Hrsg.): *Image Analysis and Processing (ICIAP)*, ser. *Lecture Notes in Computer Science*. Berlin Heidelberg: Springer, Band 8157, S. 229-238.
- Bourlai, T., Narang, N., Cukic, B. & Hornak, L. (2012). On designing a SWIR multiwavelength facial-based acquisition system. In *Infrared Technology and Applications XXXVIII*, ser. *Lecture Notes in Computer Science*. Band 8353, S. 83 530R–83 530R–14.
- Determan, G. E. & Wunderlin, D. J. (2008). *Encoded binary liveness detector*. US Patent 2008/0 203 307 A1.
- Deutsche Gesetzliche Unfallversicherung e.V. (2015). Industrieroboter. *DGUV Information 209-074*.
- Dowdall, J., Pavlidis, I. & Bebis, G. (2003). Face detection in the near-IR spectrum. *Image and Vision Computing*, 7(21), 565 - 578.

- Fitzpatrick, T. (1988). The validity and practicality of sun-reactive skin types I through VI. *Archives of Dermatology*, 6(124), 869.
- HacsKaylo, M. (1985). *Automatic human body detector*. US Patent 4,500,784.
- Jacquez, J., Huss, J., McKeehan, W., Dimitroff, J. & Kuppenheim, H. (1955). Spectral reflectance of human skin in the region 0.7-2.6 $\mu\text{m}$ . *J. of Applied Physiology*, 3(8), 397.
- Kilgore, G. & Whillock, P. (2008). *Skin detection sensor*. US Patent 11/264,654.
- Kühling, S. (2015). Unglück im VW-Werk Baunatal: Roboter tötet Arbeiter. In Seidenfaden, H. (Hrsg.): <http://www.hna.de/kassel/kreis-kassel/baunatal-ort312516/roboter-toetet-arbeiter-vw-werk-5191637.html> [01.07.15], zuletzt geprüft: 27.6.2016. Kassel: Dierichs.
- Jacquez, J., Huss, J., McKeehan, W., Dimitroff, J. & Kuppenheim, H. (1955). Spectral reflectance of human skin in the region 0.7-2.6 $\mu\text{m}$ . *J. of Applied Physiology*, 3(8), 397.
- Mendenhall, M. J., Nunez, A. S. & Martin, R. K. (2015). Human skin detection in the visible and near infrared. *Applied Optics*, 35(54), 10 559 - 10 570.
- Pavlidis, I., Symosek, P., Fritz, B., Bazakos, M. & Papanikolopoulos, N. (2000). Automatic detection of vehicle occupants: the imaging problem and its solution. *Machine Vision and Applications*, (54), 313 - 320.
- Peskoski, K. R. (2010). *Design of a monocular multi-spectral skin detection, melanin estimation, and false-alarm suppression system*. Master's thesis, Air Force Inst of Tech Wright-Patterson AFB OH.
- Schwaneberg, O. (2013). *Concept, system design, evaluation and safety requirements for a multispectral sensor*. Ph.D. Dissertation, University of Siegen, Fakultät IV: Naturwissenschaftlich-Technische Fakultät.
- Sporrer, S., Steiner, H., Velte, M. & Jung, N. (2015). NIR camera based person detection in the working range of industrial robots. In *Proc. Int. Conf. on Safety of Industrial Automated Systems (SIAS)*. S. 147 - 152.
- Steiner, H., Sporrer, S., Kolb, A. & Jung, N. (2016). Design of an active multispectral SWIR camera system for skin detection and face verification. *Journal of Sensors - Special Issue on Multispectral, Hyperspectral, and Polarimetric Imaging Technology*, 1(2016), article ID 9682453.
- Steiner, H., Kolb, A. & Jung, N. (2016). Real-time motion compensation for field sequential multispectral imaging. In *Proc. Europ. Conf. on ComputerVision (ECCV)*. Springer International Publishing, (submitted).

### Kontaktinformationen

Sebastian Sporrer<sup>1</sup>, Norbert Jung<sup>2</sup>  
Hochschule Bonn-Rhein-Sieg, Institut für Sicherheitsforschung  
Grantham-Allee 20  
53757 Sankt Augustin  
Tel.: +49 2241 / 865 – (456<sup>1</sup>|211<sup>2</sup>)  
E-Mail: (sebastian.sporrer<sup>1</sup>|norbert.jung<sup>2</sup>)@h-brs.de