

Abschlussbericht zum Vorhaben
„SPAI - Sichere Personendetektion im
Arbeitsbereich von Industrierobotern
durch ein aktives NIR-Kamerasystem“
(FP357)

Laufzeit
01.06.2013 – 31.03.2016

Bericht vom 30.06.2016

Autoren
Sebastian Sporrer, Holger Steiner,
Norbert Jung

Inhaltsverzeichnis

Inhaltsverzeichnis

Kurzfassung deutsch

Kurzfassung englisch

1. Problemstellung

Literatur

2. Forschungszweck/-ziel

3. Methodik

Arbeitspakete

Zeitlicher Ablauf

4. Ergebnisse des Gesamtvorhabens

Spezifikation

Initiale Untersuchung

Funktionsmuster

Funktionale Sicherheit & Strahlungssicherheit

Tiefendaten

Bewegungskompensation

Personenerkennung

Evaluation der Personenerkennung

Falschpositive durch „Blooming“

Bewegungskompensation schneller Bewegungen

Grenzen der aktiven Beleuchtung und Auflösung

Hintergrundsubtraktion

Veröffentlichungen

Schutzrechte

5. Auflistung der für das Vorhaben relevanten Veröffentlichungen, Schutzrechtsanmeldungen und erteilten Schutzrechte von nicht am Vorhaben beteiligten Forschungsstellen

6. Bewertung der Ergebnisse hinsichtlich des Forschungszwecks/-ziels, Schlussfolgerungen

7. Aktueller Umsetzungs- und Verwertungsplan

8. Anhang/Anhänge

Kurzfassung deutsch

Im Rahmen dieses Projekts wurden die Machbarkeit und die Vorteile einer Berührungslos Wirkenden Schutzeinrichtung (BWS) auf Basis eines aktiven Nahinfrarot-Kamerasystems untersucht. Solche Systeme ermöglichen eine robuste, pixelweise Erkennung von Haut anhand ihres spektralen Fingerabdrucks im nahen Infrarotbereich. Diese Technik kann genutzt werden, um automatisch zwischen Personen und anderen Objekten im Bild zu unterscheiden. Besonders im Arbeitsumfeld von Industrierobotern bietet dieses System damit einen Mehrwert in Bezug auf die Sicherheit für die Mitarbeiter und Verfügbarkeit eines solchen Roboters. Es folgt eine kurze Zusammenfassung der Ergebnisse:

- **Robuste Hauterkennung mit aktivem Nahinfrarot-Kamerasystem**
Hauptpixel können, bezogen auf das für die Anwendung exemplarisch ausgewählte Bildmaterial, mit einer Sensitivität von 0,999974 von anderen Pixeln unterschieden werden. Die Daten wurden mit einem Funktionsmuster erhoben, das ebenfalls die Machbarkeit des Ansatzes beweist.
- **Personenerkennungsleistung**
98,1 % aller vorgegebenen, zu einer Person gehörenden Pixel können in einer Stichprobe der Bilder automatisch erkannt werden.
- **Unterscheidung zwischen Personen und anderen Objekten**
Ermöglicht eine „intelligente“ Form von Muting, bei dem Werkstoffe und der Roboter selbst ignoriert, Menschen jedoch im gleichen Abschnitt zuverlässig detektiert werden.
- **Echtzeitfähigkeit**
Die Bildakquise ist mit aktuellen Nahinfrarot-Kameramodellen mit einer Geschwindigkeit von 25 Multispektralbildern pro Sekunde in hinreichender Zeit möglich. Die Verarbeitungsgeschwindigkeit skaliert mit dem Hardware-Aufwand.
- **Tiefendaten**
Mit einem stereoskopischen Aufbau von zwei NIR-Kameras ist die dreidimensionale Akquise von Bildern mit etablierten Algorithmen möglich. Die Kombination von NIR-Kameras und solchen für den sichtbaren Bereich reicht unseren Untersuchungen nach nicht aus.
- **Kompensation von Bewegungsartefakten**
Bewegungsartefakte können in großen Teilen algorithmisch kompensiert werden. Sie entstehen durch die prinzipbedingte Einzelbildakquise der Nah-Infrarotkanäle. Die untersuchten Algorithmen stoßen jedoch bei schnellen Bewegungen und in Kombination mit weiteren Artefakten an ihre Grenzen.

Zusammenfassend kann die Machbarkeit und die Tauglichkeit dieser Art von Schutzeinrichtung zur Überwachung von gefährlichen Bereichen bestätigt werden. Die Größe des Überwachungsbereichs ist jedoch stärker limitiert, als die eines heute gängigen Kamerasystems für den sichtbaren Bereich. Das liegt zum einen an der im Vergleich zu Siliziumbildsensoren recht jungen Entwicklung im Bereich der Nahinfrarot-Bildsensoren und zum anderen an der Leistungsgrenze der aktiven Beleuchtung. Diese und andere Probleme, insbesondere eine gegen Bewegungsartefakte robustere Bildakquise, werden während des Nachfolgeprojekts „beyondSPAI“ in den Fokus der Arbeiten gerückt. Im Rahmen dessen soll die Anwendung eines Nahinfrarot-Kamerasystems im Verbund mit zusätzlicher Sensorik in einem multimodalen Sensorsystem zur Absicherung von kollaborierenden Robotern untersucht werden.

Kurzfassung englisch

This project covered the research about feasibility and benefits of a Vision Based Protective Device (VBPD) based on an active near-infrared camera system. This kind of system allows a robust pixel-by-pixel detection of skin using its spectral fingerprint in the near-infrared region. This technique can be used for an automatic distinction between persons and other object within the image data. Especially in the working area of industrial robots, the system offers an additional benefit for the safety of employees and the availability of the robot. The following is a short summary of the results:

- **Robust skin detection using NIR camera system**
Pixels of human skin can be distinguished from other pixels with a sensitivity of 0.999974 according to the selected generic footage for the intended application. The data was acquired using a functional prototype, which also shows the feasibility of this kind of a system.
- **Person detection**
98.1 % of all picked pixels of persons in a sample of images are recognized by the system automatically as person-pixels.
- **Distinction between persons and other objects**
Allows a "smart" way of Muting, in which materials and the robot are ignored, but persons in the same area are detected safely.
- **Real-time capability**
Image acquisition can be performed with a speed of 25 multispectral images per second in a sufficient amount of time using currently available near-infrared cameras. The processing time scales with hardware-effort.
- **Depth information**
A stereoscopic setup with two near-infrared cameras allows the acquisition of three-dimensional images with state-of-the-art methods. The combination of a NIR-camera and a camera for the visible spectrum is not suitable for this purpose, regarding our investigation.
- **Compensation of motion artifacts**
A large amount of motion artifacts, caused by the necessary single image acquisition of each near-infrared channel, can be compensated with state-of-the-art algorithms. Nevertheless, fast movements and the combination of other artifacts show the performance limits of those algorithms.

Summarizing, the feasibility and suitability of this kind of protective device for the observation of dangerous areas can be confirmed. The size of the observable area is more restricted compared to common camera systems working in the visible range. This is due to the relatively young development of near-infrared images sensors compared to silicon-based image sensors, as well as the performance limits concerning the active light source. These and other problems, in particular an image acquisition method more robust against motion artifacts, are addressed in the follow-up research project "beyondSPAI". In the context of this, the application of a NIR-camera system as part of a multimodal sensor system including different kinds of sensors to safeguard collaborative robots will be investigated.

1. Problemstellung

Das Ziel dieses Projekts war die Untersuchung und exemplarische Implementierung einer sicheren Personenerkennung und -Ortung in Echtzeit mit Hilfe von speziellen Kamerasystemen, die den nahinfraroten (NIR) Spektralbereich nutzen.

In vorangegangenen Projekten sind die charakteristischen Eigenschaften von menschlicher Haut im nahinfraroten Spektralbereich für sicherheitskritische Anwendungen bereits seit mehreren Jahren an der Hochschule Bonn-Rhein-Sieg untersucht worden. Die Eigenschaften von Haut erlauben es, sie mit sehr hoher Zuverlässigkeit zu erkennen und von anderen Materialien zu unterscheiden. Fehlauflösungen durch fälschlicherweise als gefährlich erkannte, tatsächlich aber unkritische Betriebssituationen können so vermieden werden. Dies macht ein solches Kamerasystem besonders für den Bereich der Maschinensicherheit, beispielsweise bei der Absicherung des Gefahrenbereichs von Industrierobotern, interessant. Bisher untersuchte Ansätze mit Stereo- und 3D-Kamerasystemen wie solche in [1] und [2], die bereits am Institut für Arbeitsschutz (IFA) der Deutschen Gesetzlichen Unfallversicherung untersucht wurden, konnten keine hinreichend zuverlässige Personendetektion gewährleisten.

Die berührungslose Erkennung von Personen und Gliedmaßen im Gefahrenbereich von Verarbeitungsmaschinen oder Roboterarbeitsplätzen ist ein wichtiger Aspekt für Sicherheitsanwendungen, da diese Maschinen oder Roboter in aller Regel mit potenziell gefährlichen, beweglichen Teilen ausgestattet sind, die für praxistaugliche Auslegungen nicht ohne Weiteres von Bedienern abgeschirmt werden können. In vielen Fällen ist ein direkter Eingriff in den Verarbeitungsablauf durch den Menschen sogar notwendig, um effektives Arbeiten zu ermöglichen. Nach aktuellem Stand der Technik verwenden solche Schutzeinrichtungen häufig eine Methode bekannt als Muting, die Werkstücken oder bewegten Maschinenteilen, bzw. Roboterarmen erlaubt, in den Schutzbereich einzudringen, während das Eindringen anderer Objekte, wie z.B. Personen, zu einem sofortigen Stopp der Anlage führt. Das größte Problem dieser Technik ist jedoch, dass sie ein detailliertes Modellwissen über die Anwendung voraussetzt, beispielsweise exakte Angaben über Zeitpunkt, Position, Form, Größe und Ausrichtung des zugeführten Werkstücks. Dies erschwert sowohl die Verwendung an handbeschickten Maschinen, als auch die Zusammenarbeit von Mensch und Maschine an Roboterarbeitsplätzen. Auch die in jüngerer Zeit vorgeschlagenen kamerabasierten Schutzeinrichtungen, die durch die Absicherung größerer Bereiche und eine dreidimensionale Lokalisierung Vorteile gegenüber den rein zweidimensionalen Schutzeinrichtungen bieten, können dieses Problem nicht lösen.

Das Projekt baut auf den bisher erzielten Forschungsergebnissen und Entwicklungen im Bereich der Hauterkennung mit Hilfe von aktiven Nahinfrarot-Sensoren auf. Die Arbeitsgruppe der H-BRS hat innerhalb der letzten Jahre in enger Kooperation mit dem Institut für Arbeitsschutz (IFA) der Deutschen Gesetzlichen Unfallversicherung (DGUV) eine Reihe von gemeinsamen Forschungs- und Entwicklungsprojekten mit weiteren Forschungspartnern und Industrieunternehmen durchgeführt. Dabei wurden verschiedene praxistaugliche Demonstratoren von optischen Sensorsystemen als applikationsspezifische Schutzeinrichtungen für Personen an Be- und Verarbeitungsmaschinen sowie für biometrische Anwendungen entwickelt und in Feldstudien untersucht (siehe [3-9]). In diesem Zusammenhang hat die DGUV-Forschungsförderung mit dem Projekt LBIS die Anwendung der Sensorik zur Absicherung von Baukreissägen und Knochenbandsägen gefördert. (Projekt FF-FP0289)

Bei den verwendeten Sensoren handelte es sich bisher stets um Punktsensoren, mit denen die Intensität des von einzelnen, abhängig von der verwendeten Optik mehr oder weniger stark begrenzten Bereichen reflektierten Lichts ermittelt wird. Dies kann mit einem Bildsensor mit einer Auflösung von nur einem einzelnen Pixel verglichen werden. Durch die Verwendung von einer oder auch mehrerer solcher Sensorzellen kann eine Vielzahl von Anwendungsfällen, insbesondere bei der Absicherung von gefährlichen Maschinen, abgedeckt werden.

Es liegt nahe, dass die Übertragung des bei den bisherigen Prototypen bewährten Sensorprinzips auf einen Bildsensor eine Vielzahl weiterer Anwendungsgebiete eröffnet. So ließe sich beispielsweise bei der Gesichts- und

Hand-, bzw. Gestenerkennung durch die zusätzliche Information eines hauterkennenden Kamerasystems die Robustheit gegenüber falsch-positiven Erkennungen deutlich steigern, da z.B. auf T-Shirts oder Plakate aufgedruckte Gesichter oder Hände im Blickfeld der Kamera schon aufgrund des Oberflächenmaterials ausgeblendet werden könnten. Auch durch Beleuchtungsänderung hervorgerufene Probleme bei der Erkennung und Verfolgung von Personen könnten durch die aktive NIR-Beleuchtung besser gelöst werden. Gerade im Maschinensicherheitsumfeld erweisen sich Kamera-basierte Systeme bisher für viele Anwendungsfälle noch als wesentlich zu unzuverlässig. [10,11]

Die Entwicklung und eingehende Untersuchung eines speziellen aktiven Bildsensorsystems im Nahinfrarotbereich basierend auf einer NIR-Kamera in Kombination mit schmalbandiger Beleuchtung, sowie die Bilddatenaufbereitung und die robuste Hautflächenerkennung anhand der gewonnenen Sensordaten standen daher im Fokus der Arbeiten für dieses Forschungsprojekt.

Bei dem hier beschriebenen Vorhaben handelt es sich um ein anwendungsnahes Forschungsprojekt, dessen Ergebnisse in einer innovativen Sicherheitsapplikation Verwendung finden sollen. Es existiert nach unseren Recherchen bisher kein vergleichbares System am Markt.

Literatur

- [1] J. Pilger, „Machbarkeitsstudie zur Personendetektion mit Kamerasystemen im Arbeitsumfeld von Industrierobotern“, Bachelorarbeit, Hochschule Bonn-Rhein-Sieg / Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung, Betreuer: Dr. M. Schaefer, Prof. N. Jung, 2005.
- [2] J. v.d. Lippe, „Person Detection and Tracking based on Range Data for Safety Applications“, Masterarbeit, Hochschule Bonn-Rhein-Sieg / Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung, Betreuer: Dr. M. Huelke, Prof. N. Jung. 2008.
- [3] D. Reinert, N. Jung, O. Schwaneberg, et al., „Finger and hand protection on circular table and panel saws“, Elsevier, SAFETY Vol. 47 Issue 8, 2009.
- [4] O. Schwaneberg, N. Jung, H. Steiner, et al., „Push-buttons with Material Classification based on Spectral Signatures“, SIAS-2010 Conference, Tampere, Finland.
- [5] N. Jung O. Schwaneberg, H. Steiner, et al., „Field Study Results of a skin detecting Safety Sensor on Circular Saws“, SIAS-2010 Conference, Tampere, Finland.
- [6] A. Asteroth, H. Steiner, N. Jung, et al., „Safety Assessment of a Material Classifying Safety Sensor using ISO 13849-1 supplemented by Model Checking“, 9th International TÜV-Symposium "Functional Safety in Industrial Application", Köln, 2010.
- [7] O. Schwaneberg, U. Köckemann, H. Steiner, und N. Jung: „A Near-Infrared LED-based Material Classification Sensor System“, in Optical Sensors, OSA Technical Digest, Optical Society of America, 2011.
- [8] O. Schwaneberg, H. Steiner, H. P. Bolívar und Jung, N., „Design of an LED-based sensor system to distinguish human skin from workpieces in safety applications“, Applied Optics, Band 51, Nr. 12: pp. 1865-1871, 04/2012. DOI:10.1364/AO.51.001865.
- [9] H. Steiner, O. Schwaneberg und N. Jung: „Advances in Active Near-Infrared Sensor Systems for Material Classification“, in Imaging Systems and Applications, OSA Technical Digest (Optical Society of America, Washington, DC, 2012), ITu2C.2.
- [10] O. Zilken: „Modellbasierte Erkennung von Handgliedmaßen im Gefahrenbereich von Maschinen mit Handvorschub“, Masterarbeit, Hochschule Bonn-Rhein-Sieg, 2007.
- [11] M. Vieth: „Reliable Recognition and Tracking of Multiple Persons in Work Safety Relevant Environments“, Masterarbeit, Hochschule Bonn-Rhein-Sieg, 2007.

2. Forschungszweck/-ziel

Das Ziel des Projektes war die Entwicklung und Evaluation eines hauterkennenden Kamerasystems, das sich für den Einsatz als Schutzeinrichtung eignet. Besonderes Augenmerk lag auf dem Bereich der Maschinensicherheit, beispielsweise bei der Absicherung des Gefahrenbereichs von Industrierobotern. Kamerabasierte Systeme erweisen sich hier bisher für viele Anwendungsfälle noch als wesentlich zu unzuverlässig. Durch den innovativen Lösungsansatz, spektrale Signaturen im Nahinfrarotbereich auszuwerten, kann die nötige Zuverlässigkeit von dem vorgeschlagenen Kamerasystem erreicht und entsprechende Anwendungsfälle abgedeckt werden.

Die Zielgruppe sind einerseits Unternehmen, die entsprechende Kamerasysteme herstellen und vertreiben können, sowie andererseits Unternehmen, die solche Systeme als Schutzeinrichtung beispielsweise an Roboterarbeitsplätzen einsetzen können, um dort die Sicherheit der Anlage zu erhöhen.

In diesem Projekt wurde daher ein aktives Multispektral-Kamerasystem entwickelt, das verschiedene Materialien und menschliche Haut voneinander unterscheiden kann. Das Kamerasystem baut auf dem in vorherigen Forschungsprojekten entwickelten Konzept eines schnellen und zuverlässigen LED-basierten Sensorsystems mit mehreren unterschiedlichen Wellenlängenbändern im Nahinfrarotspektrum auf. Die LEDs generieren sequentiell kurze Lichtpulse, um die im Blickfeld der Kamera befindlichen Objektoberflächen zu beleuchten. Die Kamera erfasst dabei jeweils ein Bild pro Wellenlängenband. Anschließend können die Helligkeitsunterschiede des im jeweiligen Band von den Objektoberflächen reflektierten Lichts miteinander verglichen und eine spektrale Signatur der Oberflächen errechnet werden. Das Kamerasystem soll ebenfalls als Stereo-Vision-System ausgelegt werden, um erkannte Personen dreidimensional im Raum zu orten. Hierbei wurde untersucht, ob als zweite Kamera auch eine herkömmliche, preisgünstigere Farbbildkamera verwendet werden kann.

3. Methodik

In vorherigen Arbeiten ist ein aktiver Punktsensor für den Nahinfrarot-Bereich (NIR) entwickelt worden, der unter anderem zur Absicherung von Be- und Verarbeitungsmaschinen bereits sehr gute Leistungen zeigte. Diese Ergebnisse sind auf ein Kamerasystem aber nicht direkt übertragbar, da hier neben dem größeren Arbeitsbereich auch die Auswirkungen störender Einflüsse wie Haare, Bart, Brille, Schminke, Tätowierungen etc. beachtet werden müssen. Deshalb zielte das Projekt darauf ab, das Problem der Hauterkennung mittels Bildverarbeitung systematisch unter Verwendung eines aktiven NIR-Kamerasystems zu untersuchen und einen entsprechenden Demonstrator zu entwickeln. Die Entwicklung und eingehende Untersuchung des NIR-Kamerasystems sowie die Bilddatenaufbereitung und die robuste Hautflächenerkennung standen daher im Fokus der Arbeiten.

Zunächst wurde eine detaillierte und systematische Untersuchung der Anforderungen erforderlich, denen ein System gerecht werden muss, das das beschriebene Verfahren verwendet. Darauf aufbauend erfolgte eine detaillierte Spezifikation des Systems. Dazu zählen die eingehende Analyse von möglichen Anwendungsfällen und -szenarien, sowie die laufende Betrachtung des zum Zeitpunkt aktuellen Stands der Technik im Bereich der Sicherheitssensorik. Daraus ergaben sich die konkreten Anforderungen für die Auslegung der Hauterkennung und eine detaillierte Spezifikation des zu entwickelnden Systems.

Als nächstes war die Auswahl und eingehende Evaluation des kommerziellen NIR-Kamerasystems erforderlich sowie Auswahl oder das Design der gepulsten NIR-Lichtquelle, für die Parameter wie die notwendige optische Ausgangsleistung und der Öffnungswinkel des Lichtkegels definiert werden mussten. Hierzu zählt neben der Auswahl geeigneter Wellenlängenbänder zur Erfassung der spektralen Signaturen und dem detaillierten Entwurf eines flexiblen Steuerungskonzepts für das Gesamtsystem auch die Analyse der relevanten Umgebungseinflüsse auf die Sensorik.

Die Bestimmung einer für die Erkennung von Haut optimierten spektralen Signatur wurde durch die spektrale Vermessung von Hautproben und typischen Werkstoffen, sowie anschließende Datenanalyse mit in Vorläuferprojekten speziell entwickelten Software-Werkzeugen erreicht. Im Detail wurden dazu zunächst die Remissionsspektren der zu untersuchenden Materialien mit Spektrometern aufgenommen. Auf Grundlage dieser Daten können vorab Zielklassen für die Materialunterscheidung spezifiziert und Software-Filter, die eine Klassifikation der Materialien erlauben, erstellt werden. Dazu wurden die Remissionsintensitäten aus allen möglichen Kombinationen einer begrenzten Anzahl von Wellenlängenbändern durch einen Brute-Force-Algorithmus miteinander verglichen, jeweils Klassifikationsfilter trainiert und evaluiert, sowie abschließend die beste Kombination ausgewählt. Dabei wurde in dem Verfahren auch berücksichtigt, dass das Kamerasystem LEDs zur Beleuchtung verwendet, deren typisches Emissionsspektrum je nach Peak-Wellenlänge und Modell näherungsweise einer Gaußverteilung mit einer Halbwertsbreite zwischen 30 und 130nm entspricht. Die Remissionsintensitäten der jeweiligen Wellenlängenbänder wurden dazu dem passenden LED-Modell entsprechend gefiltert, um ein möglichst realitätsnahes Ergebnis zu erzielen. Mit Hilfe der so bestimmten bestmöglichen Kombination von Wellenlängenbändern konnten anschließend geeignete LEDs für die Entwicklung der Beleuchtungseinheit des aktiven Kamerasystems ausgewählt werden.

Nach Implementierung von Hard- und Software eines entsprechenden Ringlichts mit eingebettetem Steuersystem und einer Schnittstelle zur Kamera wurde die Software zur Bildakquise und -verarbeitung entwickelt. Hierbei mussten zu erwartende Probleme wie optische Verzerrungen, Störungen oder durch die Bewegung von Objekten im Bild während der Aufnahmen entstehende Artefakte behandelt und geeignete Lösungen gefunden werden. Dies beinhaltet Methoden zur Kalibration des Kamerasystems, Algorithmen zur Korrektur von nicht kalibrierbaren Fehlern und Bildrauschen, sowie Algorithmen zur Bewegungskompensation. Letztere war besonders wichtig, da das Kamerasystem prinzipbedingt die Bilder der unterschiedlichen Wellenlängenbänder sequentiell nacheinander aufnimmt. Hierdurch entsteht bei bewegten Objekten ein räumlicher Versatz in den zusammengehörigen Bildern, der hauptsächlich an Kanten und Objektbegrenzungen zu starken Verfälschungen der spektralen Signaturen führen kann. Um eine robuste Unterscheidung der Materialien unter Berücksichtigung

von Einflüssen zu ermöglichen, die sich durch verschiedene Entfernungen und Winkel zur Objektoberfläche ergeben, wurden mit dem entwickelten Kamerasystem erneut Messreihen an den betrachteten Materialien durchgeführt und Techniken aus dem Bereich des maschinellen Lernens zum Training von komplexeren Klassifikationsalgorithmen, wie beispielsweise Support Vector Machines (SVM) oder Entscheidungsbäumen, eingesetzt. Hierbei wurden automatisiert Klassifikationsmerkmale und -Regeln aus den Trainingsdaten abgeleitet, die anschließend zur Bewertung und Klassifizierung ähnlicher Daten im Feld verwendet wurden. Schließlich wurde ein Haut-Filter implementiert, der zuverlässig und eindeutig erkannte Region mit menschlicher Haut in den Bilddaten pixelgenau hervorhebt.

Auf Grundlage der NIR-Multispektralbilddaten und der so gewonnenen Hautflächen-Information wurden anschließend die Algorithmen für die Hauptfunktion „Personenerkennung“ implementiert. Hier sollten nach Möglichkeit keine völlig neuen Algorithmen entwickelt werden, sondern bewährte Verfahren an die veränderten Rahmenbedingungen angepasst werden. Die entwickelten Erkennungs- und Klassifikationsalgorithmen wurden schließlich unter realen Bedingungen an Testobjekten und Testpersonen getestet, evaluiert und ggf. optimiert.

Die Datenauswertung des Systems wurde nach entsprechender Zuverlässigkeitserüchtigung gemäß etablierter Standards exemplarisch auf einer Embedded-Plattform getestet, die üblicherweise aus Kostengründen und wegen begrenzter elektrischer Leistungsaufnahme nur begrenzte Speicherressourcen und Rechenleistungen bieten. Dabei wurde zu Testzwecken eine Portierung der Algorithmen zur optimalen Ausnutzung der Ressourcen, z.B. durch Verwendung von Hardwarebeschleunigung durch dedizierte Grafikverarbeitende Prozessoren (GPUs) durchgeführt. Nach erfolgter Systemintegration und den Abschlusstests des kompletten Systems fand die Erprobung in einer praxisnahen Feldstudie statt.

Arbeitspakete

- AP I** Anforderungsanalyse und Erstellen einer entsprechenden Systemspezifikation
- AP II** Spektrometrische Vermessung von Hautproben, typischen Werkstoffen und Umgebungsmaterialien sowie Integration in die bestehende Materialdatenbank.
- AP III** Multispektrale Bildaufnahme typischer Materialoberflächen und Hautproben mit der NIR-Kamera im Labor unter Verwendung der vorhandenen spektral programmierbaren Lichtquelle
- AP IV** Bestimmung der für die Anwendung optimalen spektralen Signaturen auf Grundlage der gewonnenen Spektrometer- und Multispektralbilddaten
- AP V** Aufbau eines geeigneten aktiven NIR-Kamerasystems mit Entwicklung einer NIR-Lichtquelle, sicherheitserüchtigter Systemsteuerung und Modifikation der NIR-Kamera, falls erforderlich
- AP VI** Entwicklung und Implementierung von Methoden zur Kalibration des Kamerasystems, sowie Algorithmen und Filtern zur Berechnung von Haut-Bildern
- AP VII** Entwicklung und Optimierung der Hauterkennungsalgorithmen und anwendungsspezifische Anpassung bewährter Algorithmen zur Gesichts- und Gliedmaßenenerkennung
- AP VIII** Evaluation und Zuverlässigkeitserüchtigung des Funktionsmusters unter realen Bedingungen an Testobjekten und –personen
- AP IX** Exemplarische Portierung der Auswertung auf eine Embedded-Plattform
- AP X** Konzeption, Durchführung und Auswertung von Feldtests
- AP XI** Dokumentation der wissenschaftlichen Ergebnisse, Vorbereitung der wissenschaftlichen und wirtschaftlichen Verwertung (Anmeldung von Schutzrechten)

Zeitlicher Ablauf

	2013			2014									2015												2016			
	6	7	12	1	2	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	
AP I																												
AP II																												
AP III																												
AP IV																												
AP V																												
AP VI																												
AP VII																												
AP VIII																												
AP IX																												
AP X																												
AP XI																												

Abweichungen

In den Zeiträumen August bis November 2013 und März bis Mai 2014 ruhten die Projektarbeiten aufgrund vakanter Stellen. Die Arbeiten des letzten Quartals wurden aufgrund von Lieferschwierigkeiten von Schlüsselkomponenten gestreckt. Insgesamt wurde eine zuwendungsneutrale Verlängerung des Projekts von 10 Monaten beantragt und gewährt. Zu Beginn des Projekts wurden Teile der Arbeitspakete V und VI vorgezogen, um während der Arbeiten in AP III ein entsprechendes aktives NIR Kamerasystem zur Datenakquise nutzen zu können. Die APs VIII und X (Evaluation und Feldtests) wurden aufgrund ihrer inhaltlichen Nähe parallel bearbeitet und unter Nutzung von Synergieeffekten gestaucht. Auch Paket IX konnte aufgrund bereits früh absehbarer Ergebnisse gestaucht werden. Die freigewordenen Ressourcen wurden auf die Optimierung des Funktionsmusters in APs V VI und VII verwendet, um die noch nicht erfüllten selbstgesetzten Anforderungen zu erfüllen und für potentiell folgende Anschlussprojekte einen optimalen Ausgangspunkt zu bieten. Die geplante Dokumentation der Ergebnisse war am Projektende mit der Evaluation verzahnt und ist nicht konkret zeitlich zu trennen.

4. Ergebnisse des Gesamtvorhabens

Spezifikation

Im Rahmen des ersten Arbeitspaketes wurde ein Dokument zur Spezifikation der berührungslos wirkenden Schutz Einrichtung erstellt. Wichtige Aspekte sind in der folgenden Liste zusammengefasst.

- Gemäß DIN EN ISO 10218-2 wird das System als Kategorie 3-System mit einem geforderten Performancelevel von „d“ entworfen.
- Sofern bei der Entwicklung des Funktionsmusters angebracht, werden Richtlinien der IEC/TS 61496-4-3 angewandt.
- Funktionale Anforderungen
 - Räumliche Detektion und Verfolgung von Personen (inkl. Aller Gliedmaßen) im Bild mit Hilfe robuster Hauterkennung anhand spektraler Signaturen; Definition per Software
 - Maximaler Überwachungsbereich: Pyramide; Grundfläche 9,8 x 7,4 m², Höhe 7,5 m
- Technische Anforderungen
 - Reaktionszeit 165 ms
 - NIR-Kamera: Sensitiv bei $900 \leq \lambda \leq 1600$ nm
 - Bezogen auf den Überwachungsbereich hinreichende, wellenlängenspezifische und aktive Beleuchtung

Initiale Untersuchung

In einer Messreihe wurden insgesamt 135 Individuen spektral vermessen. Hierzu wurden NIR-Porträtaufnahmen der Gesichter in verschiedener Stellung und der Arme angefertigt, um eine multispektrale Bilddatenbank zu erstellen, die zur weiteren Untersuchung als Grundlage diente. Des Weiteren wurden verschiedene Punkte der Haut (siehe Abbildung 1) und der getragenen Kleidung mit Hilfe eines NIR-Spektrometers vermessen. Für 15 der 135 Individuen wurden keine Daten mit dem Spektrometer erhoben. Auch die Spektraldaten der 120 verbleibenden Individuen wurden in einer Datenbank gesammelt. Zusätzlich wurden verschiedene Materialien spektral vermessen. So wurden etwa typische Metalle (Kupfer, Edelstahl, Aluminium), Kunststoffe (z.B. PVC) und sonstige Werkstoffe (Holz in verschiedenen Feuchtgraden) in die Datenbank aufgenommen.

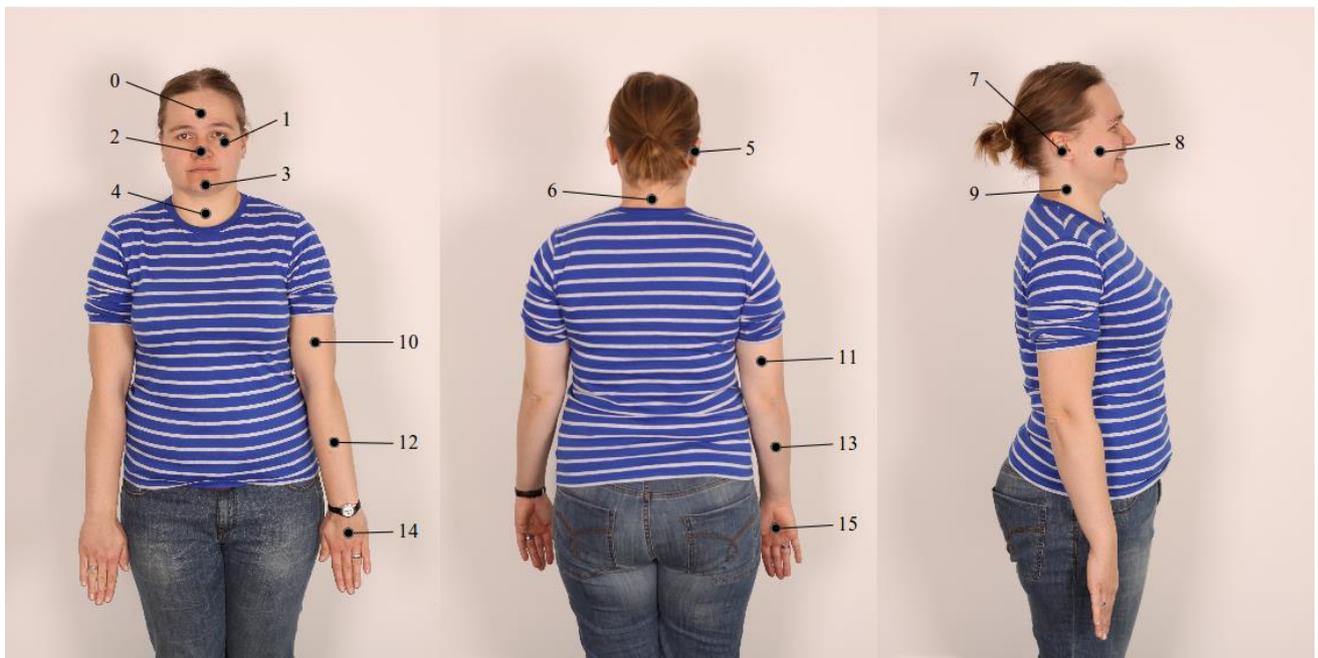


Abbildung 1: Punkte zur spektralen Vermessung mit einem NIR-Spektrometer.

Die Möglichkeit zur Trennung von Haut und anderen Materialien durch spektrale Signaturen wird in einem ersten Auswertungsverfahren für die mit dem Spektrometer gesammelten Messdaten untersucht. Bei 19 von den 120 Individuen müssen die Spektrometer-Messdaten aufgrund von Fehlern bei der Durchführung verworfen werden. Von den verbleibenden 101 Individuen werden die Messpunkte 0, 4, 6, 8, 9, 10, 11, 12, 13, 14 und 15 zur Auswertung herangezogen. Die Messpunkte 1, 2, 5 und 7 haben sich nach der Begutachtung der erhobenen Daten als ungeeignete Messpunkte erwiesen. Aufgrund mangelnder Auflagefläche und Andruckproblemen konnten dort mehrmals nur schwache und verzerrte Signale gemessen werden. Messpunkt 3 wird vernachlässigt, da viele der männlichen Probanden zum Zeitpunkt der Messung Bartträger waren und an diesem Messpunkt daher mehrfach das Spektrum von Haaren statt von Haut erfasst wurde. Mittels Software wird für die Wellenlängenbereiche um 1050 nm, 1300 nm und 1550 nm die schmale Abstrahlcharakteristik von typischen LEDs dieser Wellenlängenbänder simuliert, um die Spektren an zu erwartende Remissionswerte anzunähern und auf die ausgewählten Bereiche zu reduzieren. In einem weiteren Schritt werden auf Grundlage dieser Werte $(g_1, g_2$ und $g_3)^1$ normierte Differenzen gebildet:

$$d[g_a, g_b] = \left(\frac{g_a - g_b}{g_a + g_b} \right) \text{ with } 1 \leq a < b < 3$$

Die drei resultierenden Kombinationen für normierte Differenzen dienen als Attribute für eine Klassifikation mit Modellen aus dem Bereich des maschinellen Lernens. Sowohl mit einem neuronalen Netz, als auch einem einfachen Entscheidungsbaum lassen sich die Trainingsdaten scharf trennen, so dass keine Fehlklassifikationen

¹ entsprechend der oben genannten aufsteigenden Reihenfolge der Wellenlängenbänder

entstehen. Auch bei einer 10-fachen Kreuzvalidierung ist die Trennungsleistung beider Verfahren nahezu perfekt. Die Ergebnisse sind in Tabelle 1 als Wahrheitsmatrizen der Klassifikatoren zusammengefasst. Die Daten werden anhand $d[g_1, g_2]$ und $d[g_2, g_3]$ in Abbildung 2 grafisch dargestellt.

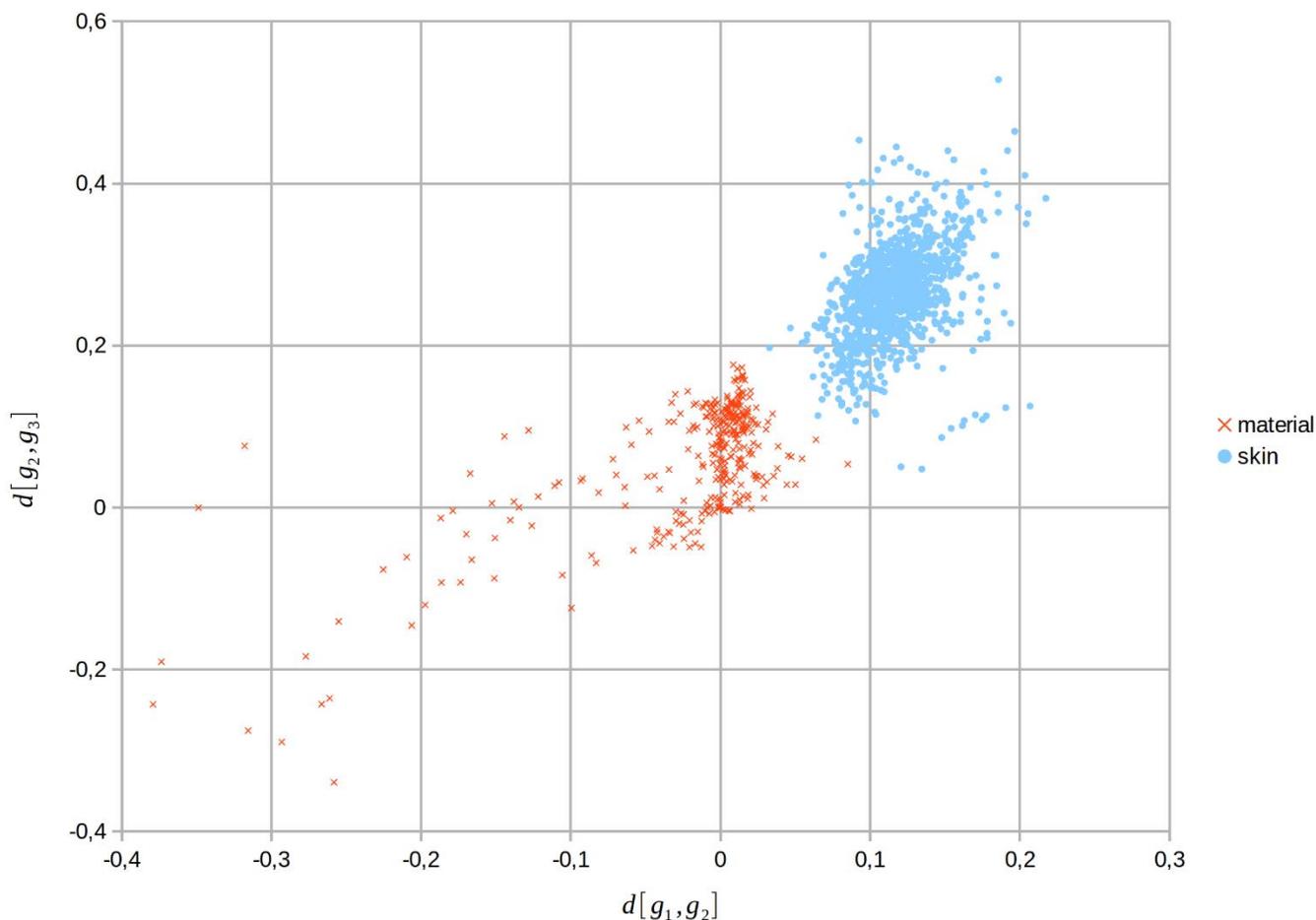


Abbildung 2: Lage von Material- und Hautproben im reduzierten Merkmalsraum (Spektrometermessungen).

Tabelle 1: Wahrheitsmatrizen einer 10-fachen Kreuzvalidierung der beiden Klassifikatoren für die Spektrometerdaten.

	Neuronales Netz		C4.5 Entscheidungsbaum	
	Haut erkannt	Material erkannt	Haut erkannt	Material erkannt
Haut	1110	1	1110	1
Material	0	335	2	333

Die Trennungsleistung des Klassifikators zur pixelweisen Hauterkennung auf Basis von multispektralen Bildern wird in [14] ausführlich anhand der während dieser Messreihe akquirierten multispektralen Bilddaten untersucht und bewertet. Zusammenfassend kann mit einer Genauigkeit von 100% eine sehr gute Klassifikatorleistung anhand der erhobenen Merkmale attestiert werden. Das für dieses Projekt definierte Ziel mit Hinblick auf Verarbeitungsgeschwindigkeit kann allerdings mit dem vielversprechenden Klassifikatormodell der Support-Vektor-Maschine und der Akquise von insgesamt 4 Wellenlängenbändern, wie in [14] dargestellt, nur schwer eingehalten werden. Da sich die verbesserte Erkennungsleistung der SVM hauptsächlich auf haut-ähnliche Materialien bezieht, die ihre Anwendung in der Nachbildung menschlicher Haut finden, kann die Erkennungsleistung kosteneffizienter Verfahren wie den oben genannten einfachen Entscheidungsbäumen und die Reduktion auf nur 3 Wellenlängenbänder für dieses Projekt als hinreichend betrachtet werden. Für den vorliegenden Anwendungsfall der Personenerkennung ist die Erkennung von hautähnlichen Stoffen, insbesondere bei Prothesen als Haut sogar gewünscht. Die Ergebnisse aus [14] lassen sich auf diese reduzierte Variante übertragen. Abbildung 3 zeigt die Lage der verschiedenen Pixel im gleichen Merkmalsraum wie in Abbildung 2.

Auch für die Pixelvariante kann eine scharfe Trennung der Trainingsdaten durch einen Entscheidungsbaum erwirkt werden. Eine Verallgemeinerung dieser Trennungseigenschaften wird erneut anhand einer 10-fachen Kreuzvalidierung gezeigt, bei der insgesamt 2 Pixel einer jeden Klasse fehlklassifiziert werden. Bei insgesamt 77771 Hautpixeln und 101334 Materialpixeln ergibt dies eine **Sensitivität** (Sicherheit) von ca. **0,999974** und eine **Spezifität** (Verfügbarkeit) von ca. **0,99998**.

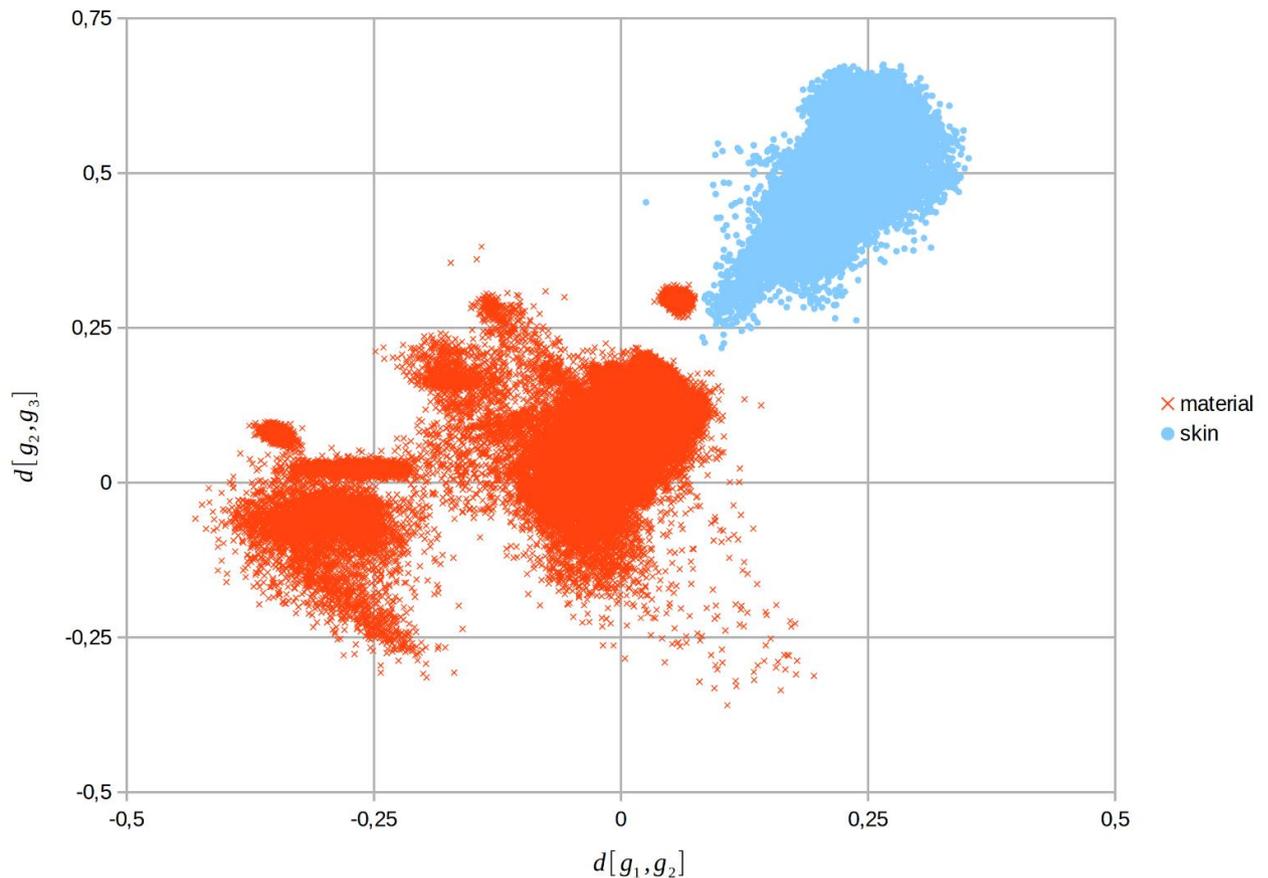


Abbildung 3: Lage von Haut- und Materialproben im reduzierten Merkmalsraum (Alle markierten Pixel der multispektralen Bildaufnahmen).

Funktionsmuster

Das finale Funktionsmuster (Abbildung 4) besteht aus einem aktiven Kamerasystem und der Anwendungssoftware für die auswertende Logik. Das aktive Kamerasystem umfasst zwei NIR-Kameras, die als Stereo-Vision System zur Ermittlung von Tiefendaten genutzt werden. Die aktive Beleuchtung der Szene in ausgewählten Wellenlängenbändern erfolgt durch insgesamt 80 Hochleistungs-LEDs. Ein eingebettetes System synchronisiert Beleuchtungsphasen und Auslöser der Kameras und ermöglicht so der auswertenden Logik die Akquise multispektraler Bilddaten.

Da eine der beiden verwendeten Kameras eine Bildrate von lediglich 30 Hz bietet, muss das Gesamtsystem ebenfalls mit dieser Geschwindigkeit arbeiten. Die in der Spezifikation errechnete Forderung einer Mindestbildrate von 38 Hz wird mit dieser Konfiguration nicht erfüllt. Jedoch sind zum Zeitpunkt des Verfassens dieses Abschlussberichts vergleichbare Kameramodelle mit 300 Hz Bildwiederholrate verfügbar, so dass die Anforderung in der Praxis erfüllt werden kann. Bereits die zweite verwendete Kamera ist mit einer Bildrate von 100 Hz hinreichend schnell.

Damit sowohl die Kameras den spezifizierten Bereich von $9,8 \times 7,4 \text{ m}^2$ bei 7,5 m Entfernung erfassen und die aktive Beleuchtung einen hinreichenden Lichtkegel erzeugen können, muss der Winkel α berechnet werden, der für die Kamera den minimalen Bildwinkel und die Leuchtquelle den minimalen Öffnungswinkel darstellt.

$$\alpha = 2 \times \tan^{-1}\left(\frac{d}{2h}\right) = 2 \times \tan^{-1}\left(\frac{\sqrt{9,8^2 + 7,8^2}}{2 \times 7,5}\right) < 80^\circ$$

Für die Kameras wurde aufgrund dieser Berechnung ein Objektiv mit einer Brennweite von 8 mm ausgewählt. Bei einer Sensorgröße von 20,35 mm (Diagonale) der verwendeten Kamera wird damit ein Bildwinkel $\alpha_{\text{Bild}} > 103^\circ$ erreicht. Dieser ist für den spezifizierten Bereich hinreichend und bietet auch die Möglichkeit zur Erfassung von Bereichen außerhalb der maximalen Sicherheitszone. Analog wurde für die Leuchtmittel der aktiven Lichtquelle eine Linse mit einem Öffnungswinkel von $\alpha_{\text{LED}} = 60^\circ$ ausgewählt. Linsen mit einem größeren Öffnungswinkel waren für den verwendeten LED-Typ zum Zeitpunkt der Implementation nicht verfügbar. Eine weitere Option besteht im Weglassen dieser Linsen. Damit würde zwar ein Öffnungswinkel von $\alpha_{\text{LED}} = 120^\circ$ erreicht, jedoch die Beleuchtungsstärke in der maximal spezifizierten Distanz auf ein unzureichendes Niveau sinken.

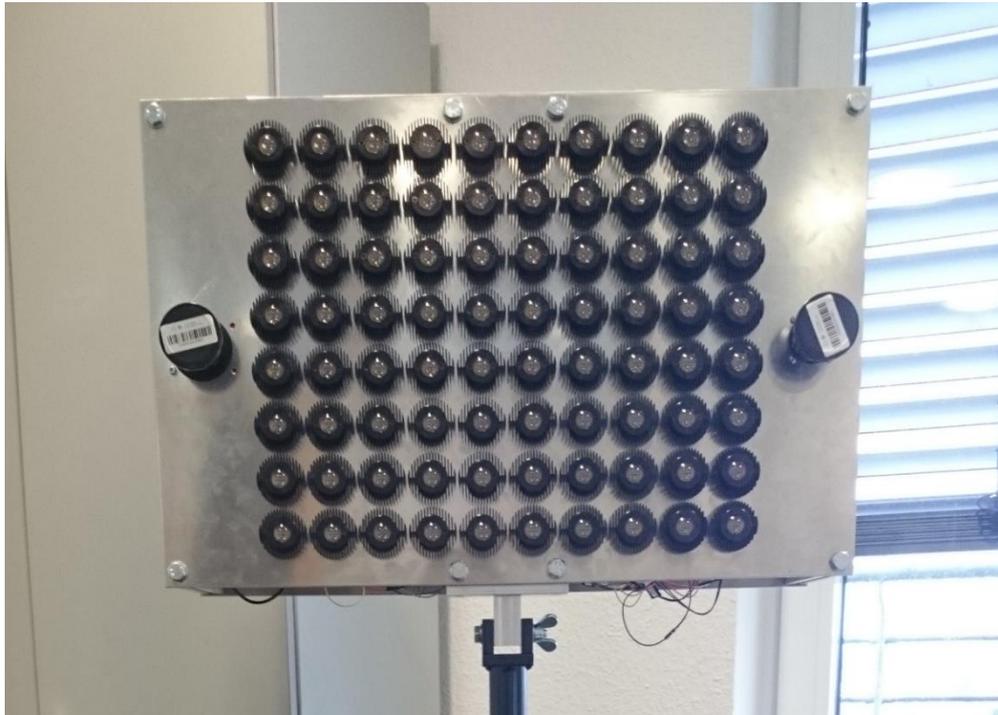


Abbildung 4: Foto des finalen Funktionsmusters.

Die Erzeugung der Bilder unter der Beleuchtung eines der drei verwendeten NIR-Kanäle und des Dunkelbildes wird über ein eingebettetes System synchronisiert. Dieses sorgt nach Freigabe durch die auswertende Logik (im Beispiel ein PC) dafür, dass die LEDs der zugehörigen Kanäle sukzessive aktiviert werden und simultan die Auslöser der Kameras betätigt werden (siehe Abbildung 5). Diese Bilder werden dann an die auswertende Logik übertragen. Dort wird mittels Software aus den Einzelbildern ein Multispektralbild erstellt. In weiteren Verarbeitungsschritten werden anschließend Haut- und Personenerkennung auf Grundlage dieser Daten durchgeführt. Die Leistungselektronik wird zudem durch Spannungs- und Strommesspfade überwacht, so dass der Ausfall einer einzelnen LED vom eingebetteten System erkannt wird und eine Sicherheitsfunktion ausgelöst werden kann.

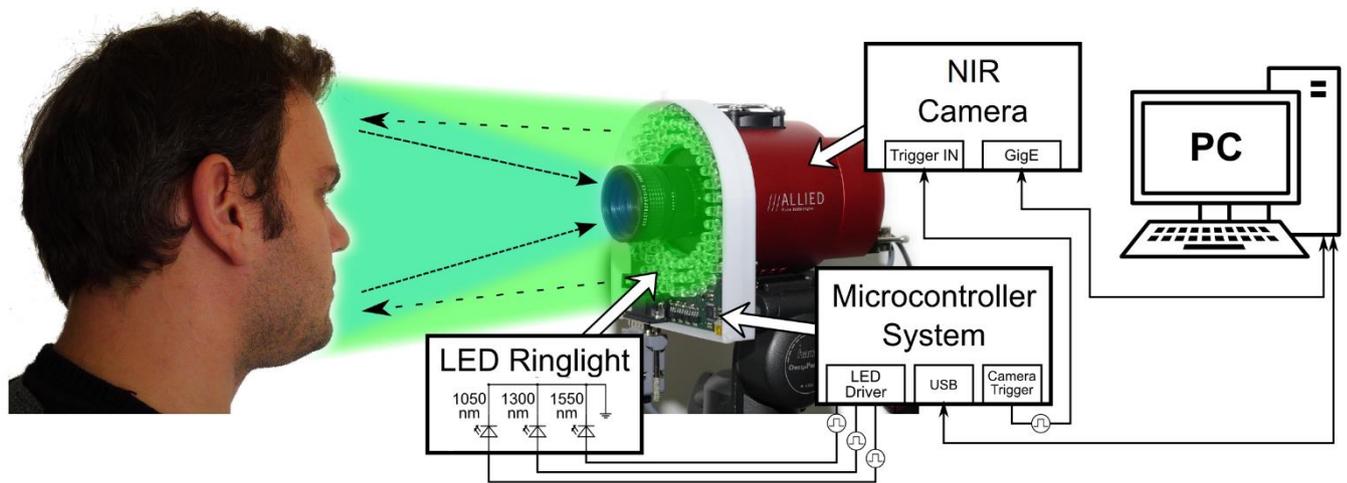


Abbildung 5: Schematische Darstellung des NIR-Kamerasystems.

Begleitend zur Hardware wurden zwei Software-Applikationen für die auswertende Logik erstellt. Die Demonstrations-Applikation „SkinCam“ wird verwendet, um das Funktionsmuster zu steuern und die ankommenden Daten zu speichern oder in Echtzeit anzuzeigen. Die Software beinhaltet außerdem die komplette Signalverarbeitungskette vom Zusammenfügen der Einzelbilder zu NIR-Multispektralbildern, bis zur Personenerkennung. Zusätzlich wurde eine Funktion implementiert, die Definition und Überwachung von Sicherheitszonen direkt im Bild erlaubt und visualisiert. Eine Evaluations-Applikation wird durch „SPAI-Eval“ bereitgestellt. Diese Software erlaubt das Ausführen der Signalverarbeitungskette auf bestehendem Bildmaterial und bietet neben der Möglichkeit zur Ausgabe von Zwischenergebnissen die Option, einzelne Schritte der Verarbeitung untereinander abzustimmen und evaluationsrelevante Daten zu exportieren.

Funktionale Sicherheit & Strahlungssicherheit

Die funktionale Sicherheit des Systems kann nur in Teilen evaluiert werden, da Angaben zum Kamerasystem hinsichtlich ihrer Ausfallrate nicht bekannt sind. Das eingebettete System zur Steuerung des Messablaufs und der Kommunikation zur auswertenden Logik wurde mittels Parts-Count Verfahren mit einer $MTTF_d > 96$ Jahren bewertet, die laut DIN EN ISO 13849 die Eingruppierung in die Klasse „Hoch“ zulässt. Die Berechnung der Ausfallrate befindet sich in tabellarischer Form im Anhang dieses Berichts. Der Diagnosedeckungsgrad kann aufgrund der implementierten Selbsttest-Routinen ebenfalls als hoch angesehen werden. Durch dieses Subsystem wird also die Eingruppierung in das erforderliche Performance Level $PL_d = d$ nicht verhindert.

Durch die vergleichsweise hohe benötigte Ausgangsleistung der Beleuchtung muss die Strahlungssicherheit für Personen im erfassten Bereich gewährleistet werden. Hierzu wurde an einem Vorläufer des hier dargestellten Funktionsmusters durch sachkundiges Personal eine Vermessung der Beleuchtungsstärke nach DIN EN 62471:2008-03 durchgeführt. Die im untersuchten Funktionsmuster verwendete Anzahl betrug insgesamt 32 LEDs². Die aktive Beleuchtung wurde aufgrund der Messergebnisse in die Freie Gruppe eingestuft und ist daher unbedenklich. Die Ergebnisse lassen sich auch auf das oben beschriebene Funktionsmuster übertragen, da auch bei einer linearen Interpolation³ der Messergebnisse auf die erhöhte Anzahl der LEDs der erlaubte Grenzwert für die freie Gruppe nicht überschritten wird.

Tiefendaten

Ein zentrales Thema dieses Projekts ist die Akquise von Tiefendaten mittels Stereo Vision. Insbesondere ist die Frage zu beantworten, ob eine der beiden Kameras für einen solchen Aufbau durch eine handelsübliche Kamera

² Je 8 x Typ 1050 nm und Typ 1300 nm, 16 x Typ 1550 nm.

³ Die gemessenen Werte werden mit dem Faktor 2,5 versehen, was dem Unterschied der verwendeten LED-Anzahl entspricht. Aufgrund der Verteilung kann dieser Wert als Abschätzung in die sichere Richtung angesehen werden.

für den sichtbaren Spektralbereich ersetzt werden kann. Diese Variante wurde im Zuge einer Master-Arbeit [12] untersucht und für nicht anwendbar befunden. Abbildung 6 zeigt rechts ein Tiefenbild, welches mit diesem Verfahren berechnet wurde, und links zur Kontrolle ein Tiefenbild, das mit einer Microsoft Kinect berechnet wurde. Diese Bilder stellen die Entfernung von Bildbereichen zur Kameraebene anhand der Pixelintensitäten dar. Es gilt: Je heller der Pixel, desto näher befindet sich dieser am Kamerasystem. Im Tiefenbild, das zum Vergleich mit der Kinect erstellt wurde, ist die abgebildete Szenerie ohne Vorwissen für den menschlichen Beobachter gut zu erkennen, während dies im rechten Bild nicht möglich ist. Die verwendeten Stereo-Vision-Verfahren ermitteln die Entfernung anhand des Versatzes einzelner Bildpunkte im linken und rechten Kamerabild. So ist das Wiedererkennen von selben Merkmalen im linken und rechten Bild essentiell für diese Verfahren. Der aus den NIR-Daten künstlich generierte „Farbraum“ unterscheidet sich jedoch vom sichtbaren Farbraum zu stark, so dass solche Merkmale nicht, oder nur vereinzelt wiedergefunden werden.

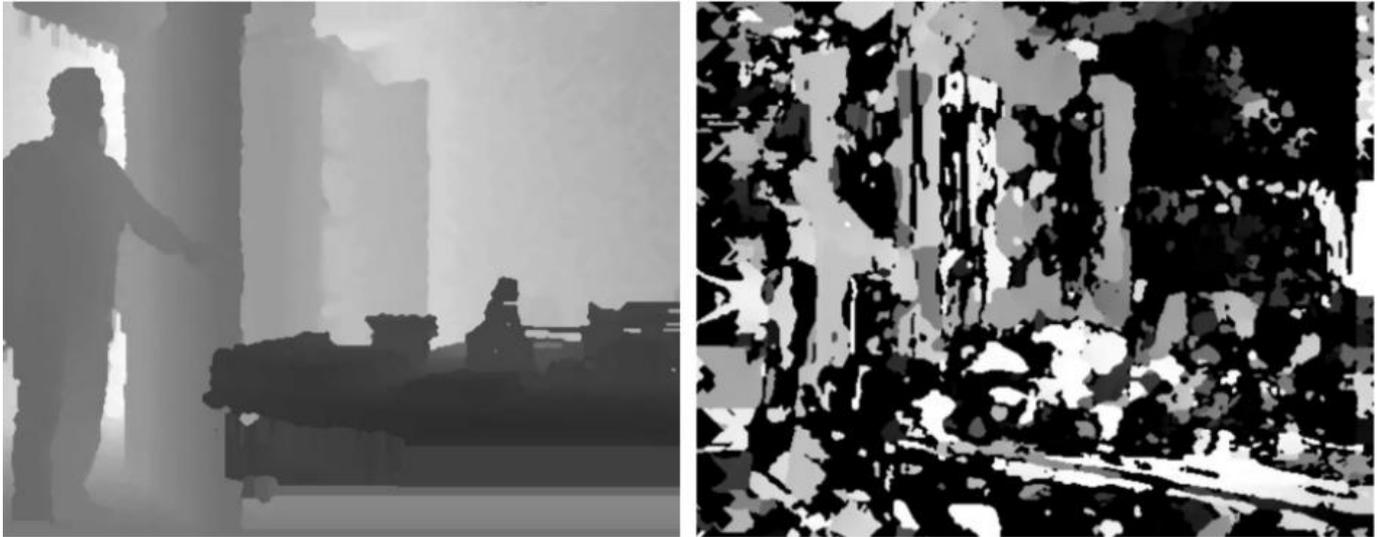


Abbildung 6: Tiefenbild der gleichen Szene erstellt mit der Microsoft Kinect (links) und einem NIR-/RGB-Stereo-Vision-System [12].

Werden nun anstelle der Kombination aus NIR-Kamera und konventioneller Kamera zwei NIR-Kameras verwendet, funktioniert die Erkennung von Merkmalen deutlich besser, so dass in dieser Konstellation brauchbare Ergebnisse mit den bewährten Verfahren erzielt werden können. Abbildung 7 zeigt die Tiefenbilder verschiedener Verfahren, auf Basis von Bilddaten zweier NIR-Kameras.

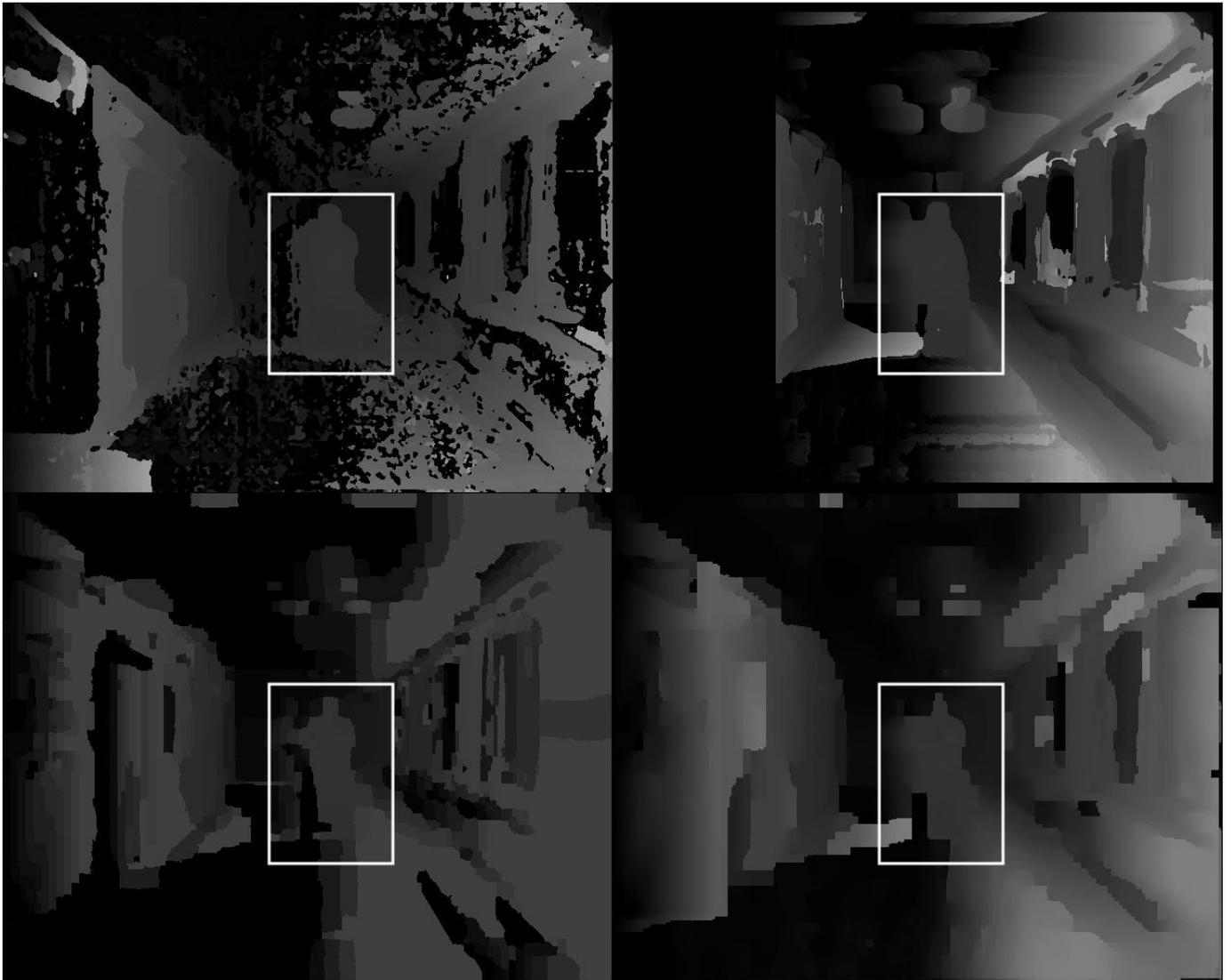


Abbildung 7: Vergleich Stereo Vision Verfahren SGBM (l.o.), BM (r.o.), BP (l.u.), CSBP (r.u.), 4 m Abstand.

Die Tiefenauflösung eines Stereo Vision Systems lässt sich mit folgender Formel⁴ berechnen:

$$\Delta Z = \frac{Z^2}{fT} \Delta d$$

Dabei ist $f = 8 \text{ mm}$ die Brennweite des verwendeten Objektivs, $\Delta d = 0,025 \text{ mm}$ die minimale Disparität, definiert durch die Pixelbreite des Bildsensors, und $T = 600 \text{ mm}$ der Abstand zwischen den beiden Kameras. Die Tiefenauflösung ΔZ kann dann in Abhängigkeit von der Entfernung Z des Objekts berechnet werden. In Tabelle 2 sind einige der Wertepaare gelistet.

Tabelle 2: Tiefenauflösung des Funktionsmusters bei verschiedenen Entfernungen.

Z (m)	1,5	2	3	4	5	6	7	7,5
ΔZ (mm)	11,72	20,83	46,88	83,33	130,21	187,50	255,21	292,97

Bewegungskompensation

Die Veröffentlichung dieses Abschnitts erfolgt zu einem späteren Zeitpunkt.

⁴ Entnommen von Seite 442 aus G. Bradski and A. Kaehler, *Learning OpenCV*, 1st ed. Sebastopol: O'Reilly, 2008, pp. 405-458.

Personenerkennung

Für die Erkennung von Personen im Bild wurden zwei verschiedene Verfahren untersucht. Das erste Verfahren nutzt eine Methode aus dem Bereich des maschinellen Lernens, die sogenannten „Convolutional Neural Networks“ und erlaubt eine Klassifikation auf Pixelebene. Es ist also mit diesem Verfahren möglich, jeden Pixel einer definierten Klasse von Objekten, wie zum Beispiel „Person“ oder „Tisch“, zuzuordnen. Das System lernt die Unterscheidung der gewünschten Klassen aufgrund eines Trainings mit bereits manuell klassifizierten Bildern, der sogenannten „Trainingsmenge“. Diese Menge dient durch geschickte Unterteilung und wiederholte Anwendung sowohl dem Training, als auch der Validierung des Klassifikators in dieser Phase. Dieser Prozess nennt sich „Kreuzvalidation“ und wird angewendet um die Allgemeingültigkeit eines trainierten Modells besser abschätzen zu können. Die Versuche zu diesem ersten Verfahren können anhand verschiedener Trainingsmengen unterschieden werden. Die Kennzahl zur Bewertung des Klassifikators ist die Klassengenauigkeit (CA) in Abhängigkeit einer Klasse K definiert als:

$$CA(K) = \frac{\sum_{\text{Bilder}} \text{Anzahl erkannter Pixel aus } K \text{ im Bild}}{\sum_{\text{Bilder}} \text{Anzahl aller Pixel aus } K \text{ im Bild}}$$

Die Ergebnisse der Versuche können Tabelle 3 entnommen werden. Dieses Verfahren eignet sich, um die These zu stützen, dass das Bereitstellen von Hautinformationen bei der Erkennung von Personen einen immensen Sprung in der Genauigkeit bedeutet. Jedoch ist eine Genauigkeit von ca. 65 % für eine Anwendung in der funktionalen Sicherheit keineswegs hinreichend. Hinzu kommt, dass dieses Verfahren sehr rechenintensiv ist und daher nur mit erhöhtem Aufwand in Echtzeitumgebungen eingesetzt werden kann. Aus diesen Gründen wurde ein weiteres Verfahren untersucht.

Tabelle 3: Genauigkeiten der CNN Klassifikatoren, trainiert mit verschiedenen Datensätzen [12].

Trainings-Daten	CA (%) Person	CA (%) Objekt
NIR	12,3	66,9
NIR + RGB	39,1	71,2
NIR + RGB + Tiefe	42,4	72,5
NIR + Tiefe + Haut	65,6	71,1

Das zweite und im Funktionsmuster implementierte Verfahren beruht auf einer Erkennung von allgemeinen Objekten im Bild, um dann in einem zweiten Schritt durch das Vorhandensein von echten Hautpixeln für jedes Objekt zu entscheiden, ob es sich um eine Person handelt oder nicht. Abbildung 8 zeigt das Konzept des Verfahrens als Blockdiagramm inklusive aller anderen Prozessschritte. Die Visualisierung einzelner Zwischenschritte und das Resultat der implementierten Personenerkennung werden in Abbildung 9 und Abbildung 10 veranschaulicht.

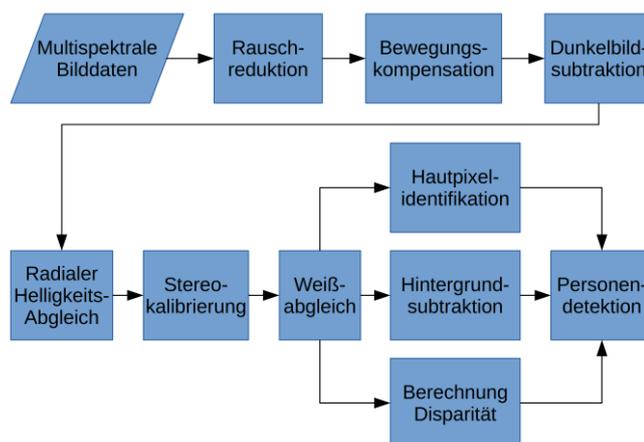


Abbildung 8: Blockschaltbild des gewählten Personenerkennungsprozesses.

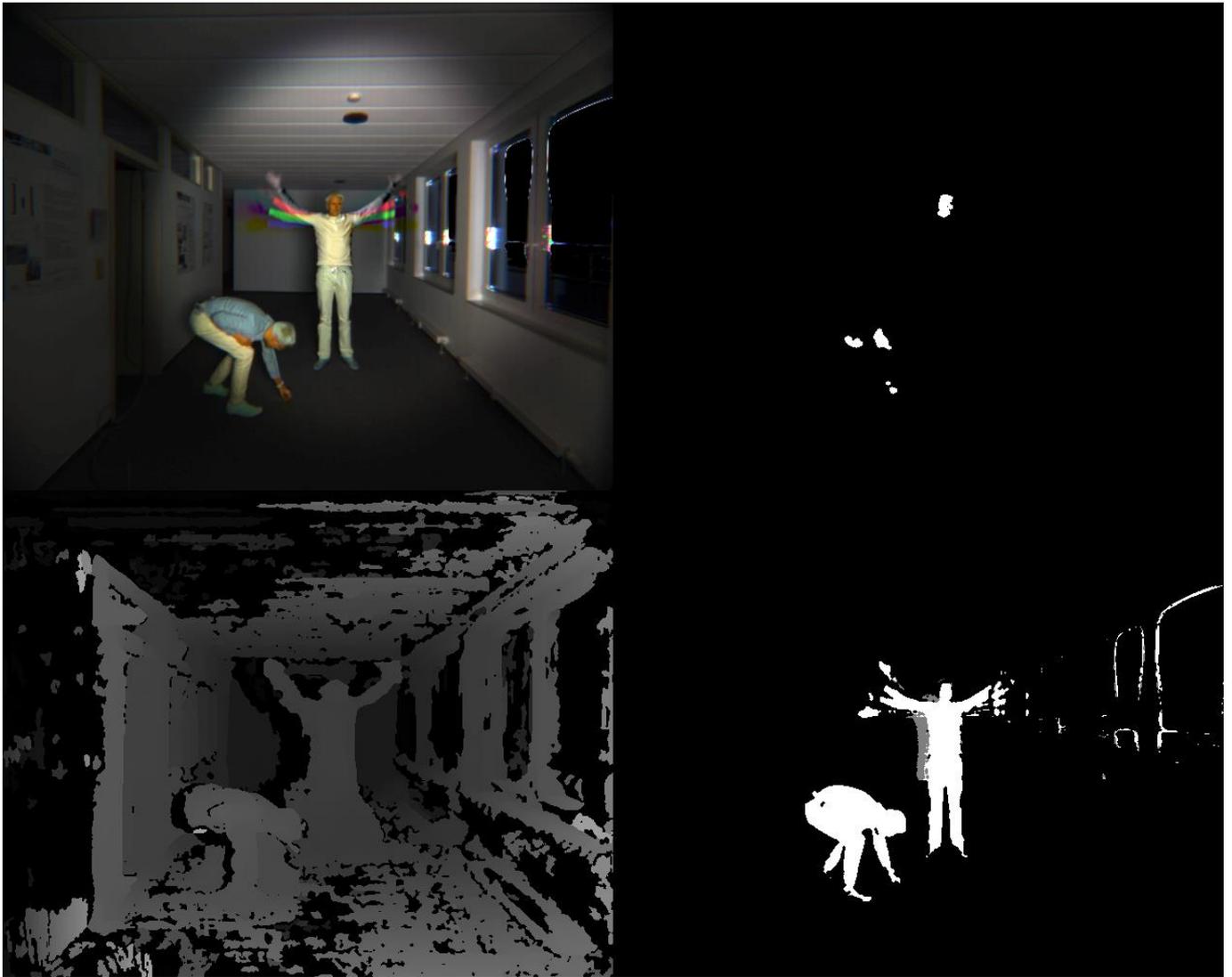


Abbildung 9: Zwischenergebnisse des gewählten Personenerkennungsalgorithmus: Kalibrierung (o.l.), Hauterkennung (o.r.), Tiefe (u.l.), Vordergrund (u.r.).

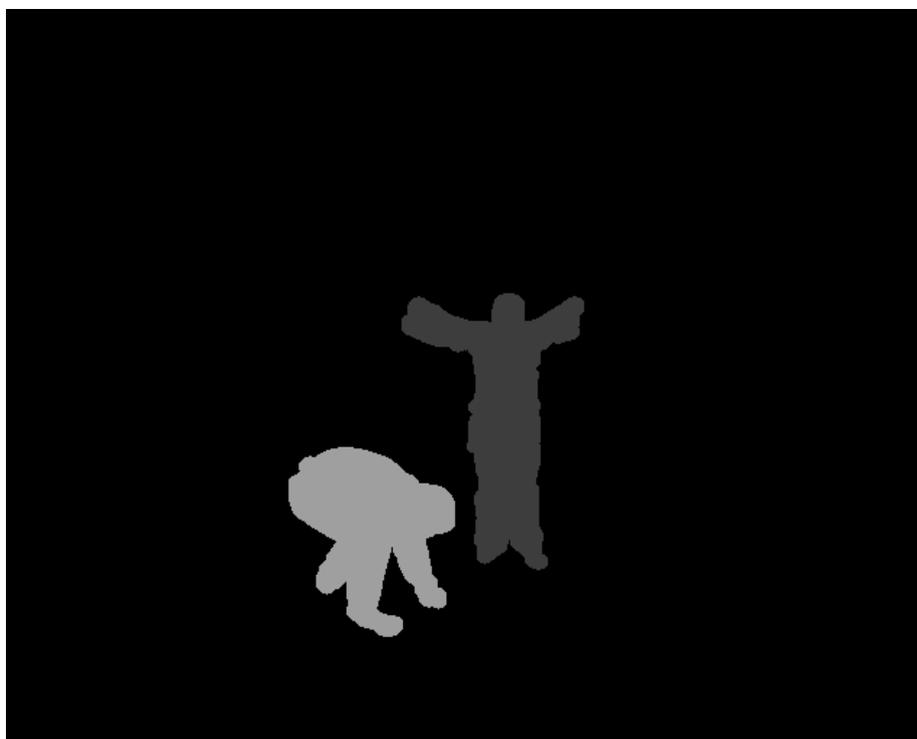


Abbildung 10: Ergebnis der Personenerkennung (Personenmaske mit Tiefendaten).

Evaluation der Personenerkennung

Zur Evaluation des implementierten Verfahrens wird ebenfalls die im letzten Abschnitt eingeführte Klassengenauigkeit als Kennzahl verwendet. Es wurde anhand von Stichproben eine Menge von Bilddaten erstellt, die zusammen mit manuell erstellten Personenmasken zur Auswertung dienen. Die Personenmasken sind Bilder gleicher Größe, in denen weiße Pixel das Vorhandensein von Pixeln im Ausgangsbild markieren, die zu einer Person gehören. Die restlichen Pixel sind schwarz. Sie wurden unter zwei Paradigmen erstellt: Um Probleme bei der Markierung speziell an Übergängen von Person und Hintergrund oder anderen Objekten zu vermeiden, werden nur Bereiche einer Person mit einigen Pixeln Abstand zum sichtbaren Rand erstellt. Allerdings ist darauf zu achten, dass jede Extremität erfasst wird. Durch konventionelle Bildbearbeitungsalgorithmen lässt sich das bildliche Objekt einer Person recht einfach vergrößern, so dass die Erkennung von Randpixeln für diese Auswertung keine Rolle spielt, so lange jede Extremität erkannt wird. Sie wird daher in der Quantifizierung der Erkennungsgenauigkeit nicht berücksichtigt.

Tabelle 4 fasst die Ergebnisse der Evaluation zusammen. Die Versuche 1-3 wurden basierend auf ersten Bilddaten mit verschiedenen Bewegungsabläufen und Objekt-/Personenkonstellationen durchgeführt. 50 Bilder wurden zufällig für Versuch 1 ausgewählt. In Versuch 2 wurden gegenüber Versuch 1 die Bilder aus der Testmenge entfernt, welche Personen zeigen, die weiter als der maximale Abstand von 7 m positioniert sind. In Versuch 3 wurden die Parameter zur einfachen Expansion der gefundenen Objekte variiert. Versuch 4 wurde auf 50 zufällig ausgewählten Bildern aus Bildmaterial eines Feldtests in der Umgebung eines Industrieroboters. Auch dieses Bildmaterial umfasst verschiedenen Objekt-/Personenkonstellationen und insbesondere den Industrieroboter selbst (siehe Abbildung 11 als Beispiel). Auch, wenn die Methode bereits bessere Gesamt-Ergebnisse liefert, als das erste auf Convolutional Neural Networks basierende Verfahren, so bleiben doch Probleme, die auch durch diesen Ansatz nicht gelöst werden. Auf diese Probleme wird im Folgenden detailliert eingegangen.

Tabelle 4: Ergebnisse der Personenklassifikation (erkannte Pixel).

Testdaten	Erkennungsrate (%)
Alle	94,2
Ohne $d > 7$ m	96
Ohne $d > 7$ m, geänderte Parameter	98,1
Feldversuch Industrieroboter	98,2 (1,9% Falschpositive)



Abbildung 11: NIR-Falschfarbenbild und automatisch generierte Personenmaske einer Person, die unter einem Industrieroboter hockt.

Falschpositive durch „Blooming“

Die in der gewählten Kamera verwendeten Indiumgalliumarsenid-Bildsensoren verfügen, unseren praktischen Erfahrungen nach, bisher noch über unzureichende Anti-Blooming-Vorkehrungen. Beim sogenannten „Blooming“ handelt es sich um die räumliche Ausdehnung der Energie von überbelichteten Pixeln auf benachbarte (siehe Abbildung 12). Die Ausprägung unterscheidet sich von Bild zu Bild leicht und kann zu Falschpositiven Klassifikationen von Hautpixeln im Bild führen. In Abbildung 13 wird das Problem veranschaulicht. Im oberen Bereich ist ein Ausschnitt eines NIR-Falschfarbenbildes zu sehen. Die zugehörige Hautmaske, in denen erkannte Hautpixel weiß markiert sind. Die an der Decke befindlichen Lampen und die indirekte Sonneneinstrahlung durch die Fenster an der rechten Seite führen vereinzelt an den Rändern zu Falschpositiv Klassifikationen.



Abbildung 12: Blooming bei starker Sonneneinstrahlung durch Fenster im Bild ohne aktive Beleuchtung (links) und mit eingeschalteter aktiver Beleuchtung als NIR-Falschfarbenbild (rechts) [12].



Abbildung 13: Auswirkungen des Bloomings auf die Hauterkennung im NIR-Falschfarbenbild (oben) und der zugehörigen Hautmaske (unten).

Bewegungskompensation schneller Bewegungen

Die Veröffentlichung dieses Abschnitts erfolgt zu einem späteren Zeitpunkt.

Grenzen der aktiven Beleuchtung und Auflösung

Ab einem Abstand von $d > 7$ m ist eine Hauterkennung nur mit Einschränkungen möglich. Die Beleuchtungsstärke nimmt mit $1/d^2$ ab, was zu einem nicht hinreichenden Signal-zu-Rausch-Abstand führt und die Robustheit der Klassifikation einschränkt. Ein weiteres Problem bei weit entfernten Objekten ist die Auflösung, bedingt durch die im Vergleich zu konventionellen Siliziumbildsensoren geringe Auflösung des verwendeten Indiumgalliumarsenid-Bildsensors. Beide Effekte sind in Abbildung 14 zu erkennen. Die linke Seite des Bilds stellt einen Ausschnitt einer NIR-Falschfarbenaufnahme dar. Auf der rechten Seite ist erneut die zugehörige Hautmaske abgebildet.

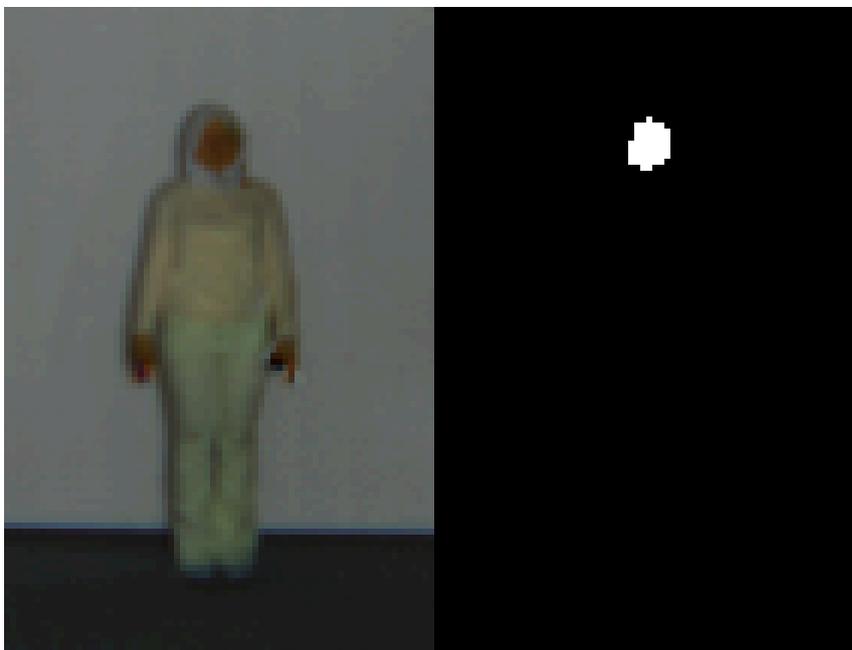


Abbildung 14: Vergrößerter Ausschnitt einer Person bei $d > 7$ m im NIR-Falschfarbenaufnahme (links) und Hautmaske (rechts).

Hintergrundsubtraktion

Bei der Vordergrund-/Hintergrundseparierung sind Artefakte festzustellen, die sich negativ auf die Personenerkennung auswirken. Abbildung 15 zeigt zwei dieser Artefakte anhand der Vordergrundmaske. Weiße Pixel geben hier die Zugehörigkeit zum Vordergrund, also zu einem Objekt an, während graue Pixel als Indikator für einen möglichen Schatten gelten. Auf der linken Seite des Bildes ist zu erkennen, dass nahezu der gesamte Oberkörper nicht als vordergründiges Objekt erkannt wird und weitestgehend zum Hintergrund gezählt wird. Auf der rechten Seite wird die Person als zwei Objekte (Getrennte Ober- und Unterkörper) wahrgenommen. Das Resultat für die Personenerkennung ist das Nicht-Erkennen des Unterkörpers, da keine Hautpixel zugehörig zu diesem abgegrenzten Objekt erkannt werden. Alle sichtbaren Hautstellen befinden sich nur im oberen Körperbereich. Eine mögliche Ursache ist das statistische Hintergrundmodell des verwendeten Verfahrens. Die farblichen Ausprägungen im NIR-Falschfarbenraum sind vergleichsweise eingeschränkt. Dies ist an der überwiegenden Färbung in Graustufen bei NIR-Falschfarbenaufnahmen zu erkennen.

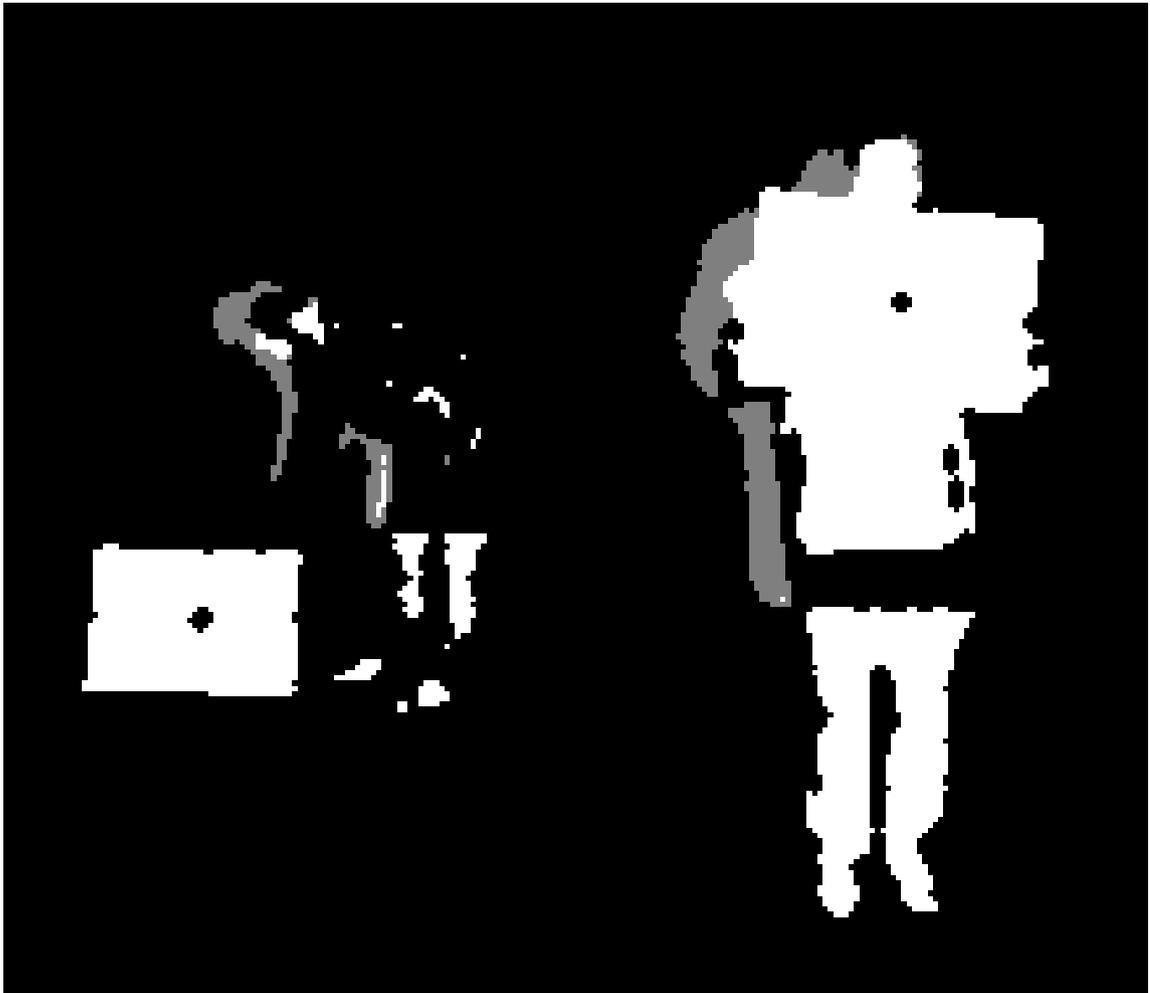


Abbildung 15: Vordergrundmasken zweier Szenen (links: fehlender Oberkörper; rechts: Ungewollte Unterteilung in Unter- und Oberkörper).

Veröffentlichungen

Die obigen Ergebnisse wurden/werden in Teilen durch folgende Beiträge der Öffentlichkeit zugänglich gemacht:

- [12]M. Velte, "Semantic Image Segmentation Combining Visible and Near-Infrared Channels with Depth Information," Master's thesis, Dept. Comput. Sci., Bonn-Rhein-Sieg Univ. Appl. Sci., Sankt Augustin, DE, 2015.
- [13]S. Sporrer, H. Steiner, M. Velte, and N. Jung, "NIR Camera Based Person Detection in the Working Range of Industrial Robots," in *Proc. Int. Conf. Safety of Industrial Automated Systems (SIAS)*, Königswinter, DE, 2015, pp. 147-152.
- [14]H. Steiner, S. Sporrer, A. Kolb, and N. Jung, "Design of an Active Multispectral SWIR Camera System for Skin Detection and Face Verification," *Journal of Sensors*, vol. 2016, Article ID 9682453, 16 pages, 2016. doi:10.1155/2016/9682453
- [15]H. Steiner, A. Kolb, and N. Jung, "Real-Time Motion Compensation for Field Sequential Multispectral Imaging," in *Computer Vision – ECCV 2016*. Springer International Publishing, 2016, submitted for publication.

Schutzrechte

Bisher wurden im Rahmen des Projekts keine Schutzrechte beantragt. Es bestehen bereits die folgenden Schutzrechte:

- N. Jung, D. Reinert, and O. Schwaneberg, "A manually fed machine for working on materials, objects and the like, and protective means for such a machine," EP Patents EP 2 054 193 B1 / 2 193 878 B1, U.S. Patents US 8 539 868 B2 / US 8 171 831 B2, 2006.

- N. Jung, D. Reinert, and O. Schwaneberg, "Apparatus for authenticating a person of at least one biometric parameter", U.S. Patent US 8 058 616 B2, 2007.
- O. Schwaneberg, N. Jung, and D. Reinert, "Apparatus For Authenticating A Person On The Basis Of At Least One Biometric Parameter", WO Patent WO/2008/125 491, 2008.
- R. Ischganeit, N. Jung, F. Neerhut, D. Reinert, O. Schwaneberg, H. Steiner, and M. Tilkorn, "Sicherheitsvorrichtung für Fahrtreppen oder Fahrsteige sowie Fahrtreppe oder Fahrsteig", EP Patent 2 404 860 B1, 2010.

5. Auflistung der für das Vorhaben relevanten Veröffentlichungen, Schutzrechtsanmeldungen und erteilten Schutzrechte von nicht am Vorhaben beteiligten Forschungsstellen

Bis auf die folgende Patentanmeldung des Unternehmens Pilz GmbH & Co. KG sind den Autoren zum Zeitpunkt der Verfassung des Abschlussberichts keine weiteren Patente oder Veröffentlichungen Dritter bekannt, die einen Bezug zum vorliegenden Forschungsprojekt haben.

- D. Döttling, T. Pilz, R. Frank, M. Hufnagel, "Verfahren und Vorrichtung zum Absichern eines gefährlichen Arbeitsbereichs einer automatisiert arbeitenden Maschine," DE Patent 102 012 102 236 A1, EP Patent EP 2 825 812 A1, US Patent US 20 150 049 911 A1, CN Patent CN 104 169 630 A, WO 2 013 135 608 A1, 2012.

6. Bewertung der Ergebnisse hinsichtlich des Forschungszwecks/-ziels, Schlussfolgerungen

Durch die oben beschriebenen Ergebnisse wurde gezeigt, dass durch die Verwendung eines aktiven NIR-Kamerasystem die Erkennung von Personen auf digitalen Bilddaten signifikant verbessert werden kann. Die robuste Erkennung von Haut anhand ihrer spektralen Signatur kann also auf ein Kamerasystem übertragen werden und bietet entscheidende Mehrinformation für Klassifikatoren. Die Tiefenwahrnehmung mittels Stereo Vision konnte nur mit zwei NIR-Kameras umgesetzt werden, was die Herstellungskosten für ein solches System im Gegensatz zur gemischten RGB/NIR-Variante maximiert. Die Praxistauglichkeit wird ebenfalls durch die relativ geringe Auflösung der am Markt vorhandenen Bildsensoren eingeschränkt. Für einen etwas geringeren Überwachungsbereich, als den in der Spezifikation festgelegten, stellt ein solches System jedoch eine taugliche Alternative dar. Die Machbarkeit konnte durch die Erstellung des Funktionsmusters nachgewiesen werden. Ein weiterer Nachteil besteht in den aufwändigen Algorithmen, die zur Datenverarbeitung, insbesondere der Personendetektion und Stereo Vision nötig sind. Diese verhindern bisher die Umsetzung der Funktionalität auf eingebetteten Rechnersystemen. In diesem Zuge könnten alternative Varianten von Prozessoren, wie zum Beispiel FPGAs, Abhilfe schaffen. Zusammenfassend lässt sich sagen, dass sich die untersuchte Technik in Form einer Schutzeinrichtung für bestimmte Anwendungen praxistauglich umsetzen lässt, jedoch durch die geringe Produktionsmarge der verwendeten Bauteile und Kameras zum derzeitigen Stand ein hoher finanzieller Aufwand für ein solches System nötig ist, der zu Lasten der Akzeptanz eines solchen Systems gehen wird.

7. Aktueller Umsetzungs- und Verwertungsplan

Die Verwertung der obigen Ergebnisse ist durch die Akquise und Durchführung des Projekts „beyondSPAI“ gesichert. Insgesamt steht ein Budget von 490.000 Euro zu Verfügung. Das Projekt wird als Teil des Programms FHprofUnt durch das Bundesministerium für Bildung und Forschung hauptfinanziert. Neben fachbereichsübergreifenden Forschungspartnerschaften innerhalb der Hochschule Bonn-Rhein-Sieg konnten weitere starke Partner für die Umsetzung des Projekts gewonnen werden. So beteiligt sich das Unternehmen K.A. Schmersal GmbH & Co. KG mit 45.000 Euro an den Entwicklungskosten und stellt den praktischen Bezug zur Forschung her. Weitere Partner sind das Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung, die Technische Hochschule Köln und die Universität Siegen.

Der Fokus dieses neuen Forschungsprojektes liegt auf der Weiterentwicklung des im Projekt SPAI untersuchten aktiven NIR-Kamerasystems. Um konkret die Probleme bei schnellen Bewegungen zu adressieren werden alternative Bildakquise-Verfahren untersucht. Diese umfassen ein passives optisches Filtermuster, das nebst breitbandiger Beleuchtung ähnlich wie ein Bayer Muster zur multispektralen Bilderzeugung genutzt werden soll. Neben der Weiterentwicklung der Personendetektion soll das NIR-Kamerasystem zusätzlich zu weiteren Sensoren in einen multimodalen Sensorverbund integriert werden, der sich für die Überwachung von kollaborierenden Robotern eignet. In Bezug auf dieses Gesamtsystem wird sich der Fragestellung gewidmet, ob eine holistische Auswertung der gesamten Sensordaten eine hinreichende Möglichkeit zur Erkennung oder gar Vorhersage von gefährlichen Situationen ermöglicht.

8. Anhang/Anhänge

Bauteil	Bezeichnung / Bauart	Ausfallrate λ [FIT] nach SN 29500	Anzahl	Gesamtausfallrate λ [FIT]	Gesamtrate gefahrbringender Ausfälle λ_d [FIT]	MTTFd in Jahren
Widerstände (R0603)	Metallschicht	0,2	6	1,2	0,6	190258,752
Kondensatoren (C0603)	Keramik, HDK/MDC (X7R,X5R)	2	9	18	9	12683,9168
BLM18HE	HF-Drossel	5	3	15	7,5	15220,7002
MI0805K400R	HF-Drossel	5	1	5	2,5	45662,1005
Leuchtdioden (LED0805)	Sichtbares Licht - Plastik	2	5	10	5	22831,0502
ATMEGA8	Mikrocontroller	100	1	100	50	2283,10502
PRQC14.74SR5010X000	Schwingquarz	15	1	15	7,5	15220,7002
FTDI (FT232RL)	USB-RS232-Bridge	40	1	40	20	5707,76256
USB-Stecker	Gold, Mehrfachkontakte	0,1	1	0,1	0,05	2283105,02
D-Sub-Stecker	Gold, Mehrfachkontakte	0,1	1	0,1	0,05	2283105,02
Lötstellen Widerstände	Maschinell	0,03	12	0,36	0,18	634195,84
Lötstellen Kondensatoren	Maschinell	0,03	18	0,54	0,27	422797,226
Lötstellen Induktivitäten	Maschinell	0,03	8	0,24	0,12	951293,76
Lötstellen Dioden	Maschinell	0,03	10	0,3	0,15	761035,008
Lötstellen ICs	Maschinell	0,03	63	1,89	0,945	120799,208
Lötstellen Stecker	Maschinell	0,03	14	0,42	0,21	543596,434
Lötstellen Stiftleisten	Manuell	0,5	22	11	5,5	20755,5002

Summe:	109,575	1041,80015
--------	---------	------------

Tabelle 5: Berechnung der mittleren Zeit bis zum gefahrbringenden Ausfall (MTTFd) nach Parts-Count Methode für das Logikmodul.

Tabelle 6: Berechnung der mittleren Zeit bis zum gefahrbringenden Ausfall (MTTFd) nach Parts-Count Methode für das Leistungselektronikmodul.

Bauteil	Bezeichnung / Bauart	Ausfallrate λ [FIT] nach SN 29500	Anzahl	Gesamtausfallrate λ [FIT]	Gesamtrate gefahrbringender Ausfälle λ_d [FIT]	MTTFd in Jahren
Widerstände:						
R2512	Metallschicht	0,2	24	4,8	2,4	47564,688
R1206	Metallschicht	0,2	1	0,2	0,1	1141552,51
R0805	Metallschicht	0,2	24	4,8	2,4	47564,688
R0603	Metallschicht	0,2	125	25	12,5	9132,42009
Kondensatoren:						
CPOL-EU153CLV-0605	Al-ELKO, flüssiger Elektrolyt	5	24	120	60	1902,58752
EMVA500ARA471MKG5S	Al-ELKO, flüssiger Elektrolyt	5	12	60	30	3805,17504

C1206	Keramik, HDK/MDC (X7R,X5R)	2	2	4	2	57077,6256
C1005	Keramik, HDK/MDC (X7R,X5R)	2	2	4	2	57077,6256
C0603	Keramik, HDK/MDC (X7R,X5R)	2	157	314	157	727,10351
Induktivitäten:						
BLM18HE	HF-Drossel	5	12	60	30	3805,17504
M1206 - Ferrit	HF-Drossel	5	24	120	60	1902,58752
SRR1280	Netztransformatoren und Überträger für Schaltnetzteile	10	24	240	120	951,29376
Dioden:						
Leuchtdioden (LED0805)	Sichtbares Licht - Plastik	2	10	20	10	11415,5251
SS26	Schottky Diode	1	26	26	13	8781,17316
LSM115JE3	Schottky Diode	1	6	6	3	38051,7504
Sonstige:						
HMHA281	Optokoppler (bipolar)	15	3	45	22,5	5073,56672
ISO1541	Digitaler Isolator	3	3	9	4,5	25367,8336
SI8660	Digitaler Isolator	3	1	3	1,5	76103,5008
IS31LT3360	Schaltregler	10	24	240	120	951,29376
LTC2945	OPV, Komparator, Spannungsüberwachung	4	24	96	48	2378,2344
ATMEGA8	Mikrocontroller	100	4	400	200	570,776256
PRQC16.00SR5010X000	Schwingquarz	15	4	60	30	3805,17504
TSR-1-2450	Schaltregler	10	1	10	5	22831,0502
PES1-M	Schaltregler	10	1	10	5	22831,0502
Lötstellen / Stecker / Verbindungen:						
Lötstellen Widerstände	Maschinell	0,03	384	11,52	5,76	19818,62
Lötstellen Kondensatoren	Maschinell	0,03	394	11,82	5,91	19315,6093
Lötstellen Induktivitäten	Maschinell	0,03	60	1,8	0,9	126839,168
Lötstellen Dioden	Maschinell	0,03	84	2,52	1,26	90599,4057
Lötstellen ICs	Maschinell	0,03	605	18,15	9,075	12579,091
Lötstellen TSR-1-2450	Manuell	0,5	3	1,5	0,75	152207,002
Lötstellen Stiftleisten	Manuell	0,5	364	182	91	1254,45331
Jumper Kanalauswahl	Klemmen(Federkraft), Leiterquerschnitt 0,5 - 16 mm	0,5	48	24	12	9512,9376
Steckverbindung Logic-Power	Klemmen(Federkraft), Leiterquerschnitt 0,5 - 16 mm	0,5	22	11	5,5	20755,5002

Summe:	1073,055	106,383411
--------	----------	------------

Tabelle 7: Berechnung der mittleren Zeit bis zum gefährbringenden Ausfall (MTTFd) nach Parts-Count Methode für das eingebettete System.

Platine	Gesamtrate gefährbringender Ausfälle λ_d [FIT]	MTTFd in Jahren
Logikmodul	109,575	1041,800147
Leistungselektronikmodul	1073,055	106,383411
<i>Summe</i>	1182,63	96,52659846