# Komponenten und Methoden für die multimodale Gefahrenanalyse in öffentlichen Räumen

David Reese[1], Rico Nestler[1,2], Karl-Heinz Franke[1]

[1]Zentrum für Bild- und Signalverarbeitung e. V.,
Werner-von-Siemens-Straße 12, 98693 Ilmenau
[2]TU Ilmenau, Fakultät für Maschinenbau, FG QBV,
Postfach 10 05 65, 98684 Ilmenau
eMail: david.reese@zbs-ilmenau.de, rico.nestler@zbs-ilmenau.de,
karl-heinz.franke@zbs-ilmenau.de
URL: https://www.zbs-ilmenau.de

**Zusammenfassung.** Gegenstand des Beitrags ist ein multimodales bildgebendes System, das als bildsensorisches Add-On einen passiven THz-Scanner zur automatischen Gefahrenanalyse in öffentlichen Räumen ergänzt. Die mit diesem Sensorverbund realisierte multimodale Bildgebung erfordert zur Ausschöpfung der erfassten Szeneninformationen in der semantischen Datenanalyse eine Fusion aller Bildmodalitäten. Daten der 3D-Bildgebung sind zur geometrisch korrekten Überlagerung aller Modalitäten von entscheidender Bedeutung. Weiterhin unterstützen sie die korrekte Vordergrund-Hintergrund-Trennung im Rahmen der semantischen Segmentierungsschritte, das robuste Tracking von Personen sowie die Steuerung der THz-Bildgebung. Der Beitrag diskutiert ein hierfür entwickeltes Verfahren zur Bilddatenfusion und hierarchischen Auswertung. Zuvor werden geeignete Methoden zur 3D-Datenerzeugung (Stereo-Matching) im Anwendungskontext verglichen. Die Tiefenbildrekonstruktion und die Bilddatenfusion erfordern eine geometrische Kalibrierung des gesamten Sensorverbunds. Hierzu wird ein neuartiges Kalibriertarget für die gemeinsame Kalibrierung von VIS- und THz-Bildmodalitäten vorgestellt. Die im Beitrag diskutierten Methoden und Verfahren der Verarbeitung von Daten aus multimodaler, multisensorieller Bildgebung sind in einem breiten Spektrum von Applikationen aus dem Bereich der Sicherheits- und Überwachungstechnik anwendbar.

## 1 Einleitung und Motivation

Die Gewährleistung der Sicherheit an öffentlichen Orten mit hohem Andrang, wie z.B. Flughäfen, ist eine wichtige gesellschaftliche Fragestellung [1] [2]. Bildauflösende Systeme zur passiven Submillimeterwellenabtastung (im Folgenden THz-Scanner genannt) sind in der Lage, auf größere Distanzen und ohne gesundheitliche Beeinträchtigung von Personen Signale in Videoechtzeit zu extrahieren, die je nach Signaleigenschaften und lateralen Zusammenhängen Hinweise auf verborgene Gefahrenobjekte liefern können (vgl. Abb. 1 links). Aktive bildgebende THz-Systeme werden bereits heute in der Sicherheitskontrolle in speziellen Messkabinen für Einzelpersonen eingesetzt.

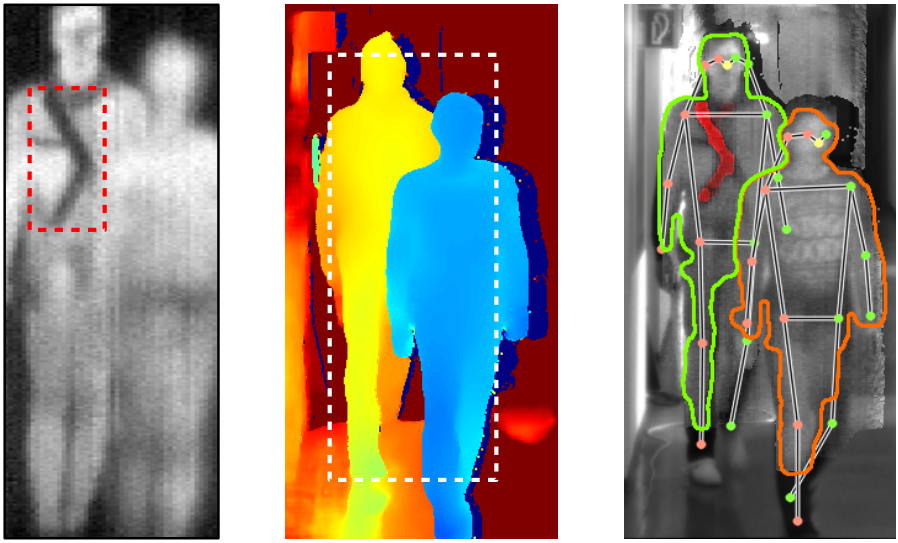

**Abb. 1:** Links: THz-Bildgebung einer Szene mit 2 Personen. Die hintere Person trägt verdeckt eine Gewehrattrappe aus Holz an der Brust. Mitte: Farbcodiertes Tiefenbild (3D-Bildmodalität) derselben Szene (größerer Bildausschnitt als links). Rechts: Ergebnisse der Bilddatenfusion und der hierarchischen, multimodalen Auswertung. Personen und segmentierte Gefahrobjekte sind überlagert dargestellt. Quelle: Supracon AG, Leibniz IPHT, ZBS e. V.

Wegen des prinzipbedingt niedrigen Signal-Rausch-Verhältnisses, der niedrigen Auflösung sowie der charakteristischen Inhomogenitäten und Artefakte in den THz-Rohdaten ist nach [2] und [3] die Detektion von Objekten ab einem bestimmten Ausprägungsgrad und nur bei gutmütigen optischen Verdeckungen monomodal möglich. Zur robusten Auswertung realer Szenen sind jedoch weitere Bildmodalitäten sehr bedeutsam, beispielsweise zur Bestärkung von Objekthypothesen durch Ergänzung mit Bilddaten aus Millimeterwellenabtastungen [4] oder, wie in diesem Beitrag dargestellt, durch die Ergänzung mit RGB- und Tiefenbilddaten zur Unterstützung der semantischen Szenenanalyse (siehe Abb. 1).

Die nachfolgend beschriebenen Komponenten und Methoden greifen Projektergebnisse aus dem BMBF-Verbundprojekt HITD (Multimodale Fernerkennung verborgener Gefahrenpotentiale in der Personenkontrolle – FKZ 13N14715) auf. Gegenstand des Projekts waren neuartige photonische Systeme zur Gewährleistung der Sicherheit in urbanen Räumen und öffentlichen Einrichtungen, bei Großveranstaltungen sowie in Verkehrsmitteln. Gesamtziel war die Schaffung eines praxisrelevanten Gefahrenerkennungskonzepts aus sensorischen Daten, das die Beurteilung des Gefährdungspotenzials von Personen oder Personengruppen aus der Distanz erlaubt und auf rechtlich zulässige Weise eingesetzt werden kann. Beim ZBS e. V. wurden hierfür Beiträge zur Erzeugung und Nutzung multimodaler, fusionierter Sensordaten sowie zur Au-

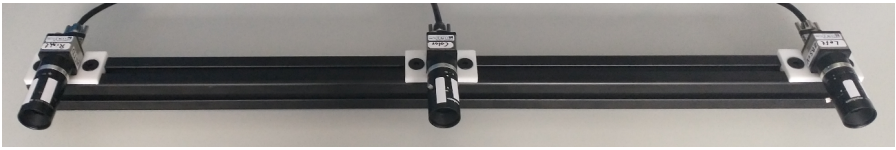

**Abb. 2:** Vorgeschlagenes Sensormodul mit zentrierter RGB-Kamera sowie zwei Grauwert-Kameras zur passiven Stereo-Rekonstruktion bei großen Arbeitsabständen.

tomatisierung der Gefahrenerkennung geleistet, um robuste Analyseergebnisse mit geringen Fehlalarmraten zu gewährleisten und diese Systeme praktisch handhabbar zu machen.

Der Beitrag stellt in Abschnitt 2 zunächst das bildsensorische Add-On für den THz-Scanner vor. In Abschnitt 3 werden Komponenten und Methoden für die geometrische Systemkalibrierung, die Tiefenbildrekonstruktion sowie die 3D-gestützte Sensordatenfusion behandelt. Abschnitt 4 stellt beispielhaft ein Verfahren zur hierarchischen multimodalen Datenauswertung vor. Der Beitrag schließt mit einer Diskussion über weitere Verarbeitungsschritte zur Sensordatenfusion und Gefahrenerkennung, die im Kontext der von uns entwickelten Softwareapplikation SEFUGA erfolgt.

## 2 Bildsensorisches Add-On für THz-Scanner

Das in Abb. 2 dargestellte Sensormodul dient der multimodalen sensorischen Ergänzung eines THz-Scanner-Demonstrators durch RGB- und Tiefenbilddaten. Das Sensormodul ist für Arbeitsabstände > 15 m ausgelegt. Aufgrund des passiven Charakters des THz-Scanners ohne aktive Szenenbeleuchtung sowie aus Gründen der Augensicherheit wurden passive Systeme und Verfahren für die Tiefenbildrekonstruktion gewählt. Um trotz ungünstiger Lichtverhältnisse eine hinreichend gute Datengrundlage für die Tiefenbildrekonstruktion zu gewährleisten, wurden Mono-Kameras mit höherer Lichtempfindlichkeit als RGB-Kameras gewählt. Neben den physikalisch-energetischen Überlegungen wurde die Kameraauswahl dadurch begründet, dass der laterale Auflösungsverlust bei klassischen Verfahren der Tiefenbildrekonstruktion mit lokalem Stereo-Matching (z.B. [5]) bei monochromen Bildpaaren geringer ist, da die Auflösung der primären Bildinformation der Sensorauflösung entspricht und nicht interpoliert ist. Eine zusätzliche Farbkamera ergänzt stattdessen eine RGB-Bildmodalität, die insbesondere für die Detektion von Personen verwendet werden kann. Diese stellt bei unserer Umsetzung der multimodalen Bildgebung die Referenz-Bildmodalität dar, d.h. bei der Bilddatenfusion werden die Modalitäten 3D (bzw. Tiefe) und THz geometrisch auf die RGB-Modalität gemappt. Die RGB-Modalität bleibt somit abgesehen von der Vorverarbeitung (Verzeichnungskorrektur) geometrisch unverändert.

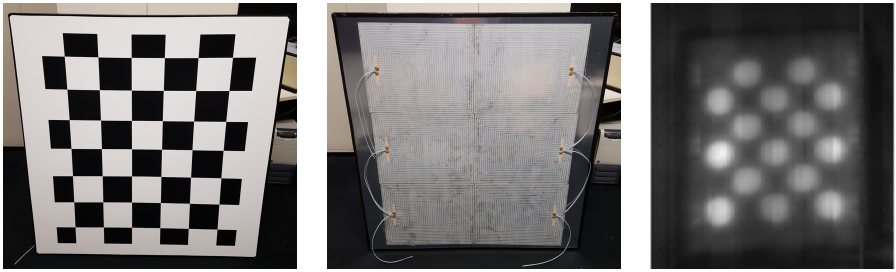

**Abb. 3:** Beheizbares Kalibriertarget für VIS- und THz-Bildmodalitäten. Links: Schachbrettmuster zur Kalibrierung der VIS-Bildmodalität. Mitte: Heizelemente auf Rückseite des Targets. Rechts: Abbildung des beheizten Targets in der THz-Bildmodalität.

## 3 Bilddatenfusion

Ziel der Bilddatenfusion als wichtigem Bestandteil der Informationsfusion ist die Ausrichtung und Überlagerung von Bilddaten der verschiedenen Sensoren, sodass ein pixelregistrierter multimodaler Bildstapel für die weitere Verarbeitung entsteht. Die wesentliche Herausforderung neben der geometrischen Anpassung ist es hierbei, trotz der oft fehlenden Korrelation der modenspezifischen Szenenabbilder eine automatische und genaue Registrierung zu realisieren.

In [4] erfolgt die Informationsfusion basierend auf korrelierenden Informationen in Millimeter- und Submillimeterbildmodalitäten, die aufgrund der ähnlichen Wellenlängenbereiche oft vorhanden sind. Bei den im Beitrag behandelten Bildmodalitäten kann jedoch nicht davon ausgegangen werden, dass hinreichende Datenkorrelationen vorliegen. Stattdessen wird die geometrische Kalibrierung aller bildgebenden Elemente des Gesamtsystems genutzt. Auf Grundlage der Kalibrierung wird auch die Tiefenbildrekonstruktion durchgeführt.

### 3.1 Multimodale geometrische Kalibrierung

Um die geometrische Kalibrierung des Gesamtsystems zu ermöglichen, ist ein spezielles Kalibriertarget zum Referenzieren aller Bildmodalitäten erforderlich. Dazu wurde ein beheizbares Target (Abb. 3) konzipiert, das ein Schachbrettmuster für die VIS-Modalitäten (Grauwert und RGB) und ein Muster aus Kreisen für die THz-Modalität bereitstellt. Bei dem Aufbau wird die hohe Absorptionseigenschaft von Holz im THz-Wellenlängenbereich und die hohe Transmissionseigenschaft der mit dem Schachbrettmuster bedruckten Kunststofffolie sowie von Styrodur als temperaturisolierende Zwischenschicht ausgenutzt.

Die Kalibrierung erfolgte nach der in [6] vorgestellten etablierten Methode. Die Detektion der Schnittpunktkoordinaten des Schachbrettmusters sowie der Koordinaten der Kreismittelpunkte erfolgte in der Demonstratoranwendung weitestgehend[1] auto-

[1] Aufgrund der Datenqualität der THz-Bilder ist vereinzelt eine manuelle Unterstützung notwendig. Zukünftig erfolgt der Prozess vollautomatisch.

**Abb. 4:** Geometrische Kalibrierung des THz-Scanners und der ergänzenden Sensorik im ZBS-Kalibriertool 3D-EasyCalib™ [7].

matisch. Durch Extrapolation wurden aus den Kreismittelpunkten die virtuellen Koordinaten der entsprechenden Schnittpunktkoordinaten für die THz-Modalität abgeleitet. Die Kalibrierung der intrinsischen und extrinsischen Kameraparameter erfolgte mit dem ZBS-Kalibriertool 3D-EasyCalib™ [7] (Abb. 4).

## 3.2 Passive Tiefenbildrekonstruktion

Die Tiefenbildrekonstruktion wurde nach der üblichen Schrittfolge für stereoskopische Systeme umgesetzt. Für die Korrespondenzsuche wurde eine GPU-Implementierung des semiglobalen Blockmatching-Verfahrens SGM nach [5] in unsere Softwareapplikation integriert. Zusätzlich wurde die Einsatzmöglichkeit neuronaler Ansätze getestet, wobei HITNET [8] als vielversprechendste Kompromisslösung in Hinblick auf Qualität und Laufzeit integriert wurde. In unseren Tests wurden die von den Autoren bereitgestellten Netzwerkgewichte verwendet, die mit dem Middlebury Dataset V3 [9] gelernt wurden.

Abb. 4 zeigt Disparitätsbilder des SGM-Verfahrens und der HITNET-Inferenz im Vergleich an einem Beispiel mit ungünstigen Beleuchtungs- (schwach beleuchteter Gang) und Texturbedingungen (schwierige Abgrenzung der Personen durch dunkele Kleidungen). Im Allgemeinen liefert HITNET wesentlich schärfer abgegrenzte Kanten und eine höhere Disparitätsauflösung als SGM, hier mit aktivierter Subpixel-Disparitätsanalyse. HITNET liefert im Gegensatz zu SGM ein vollständiges Disparitätsbild, schätzt also im Rahmen einer integrierten Disparitätskartenverbesserung auch Werte in Abschattungsbereichen und außerhalb des Überlappungsbereichs der Sensoren. Allerdings können für die weiteren Verarbeitungsschritte falsche Schätzun-

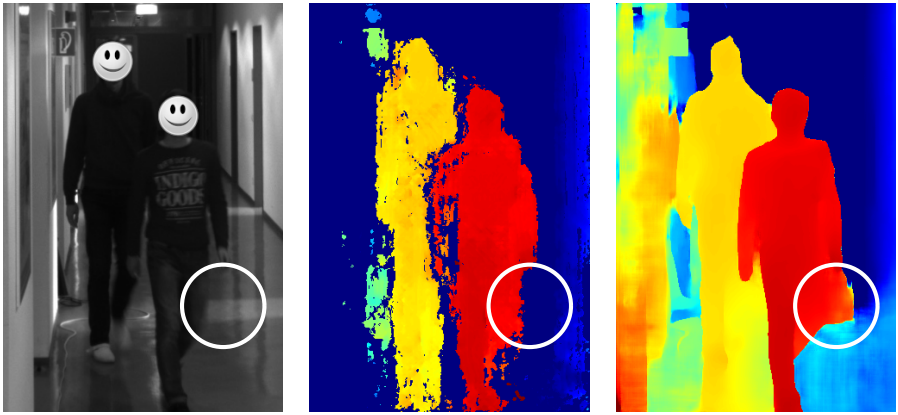

**Abb. 5:** Links: Ausschnitt einer rektifizierten Aufnahme der linken Mono-Kamera. Mitte: Aus SGM resultierende Disparitätskarte. Rechts: Aus HITNET- Inferenz resultierende Disparitätskarte.

gen aus der HITNET-Inferenz nachteilig gegenüber den durch SGM eindeutig ungültig klassifizierten Disparitäten in Bereichen ohne Texturierung bzw. 3D-Information sein (vgl. durch Kreis markierten Bereich in Abb. 4).

Beide Verfahren sind grundsätzlich auch bei den dargestellten ungünstigen Bedingungen für die Abstandsabschätzung von Personen geeignet. Insbesondere für texturarme, sehr dunkle und spiegelnde Bereiche des Hintergrunds resultieren zwar viele falsche oder ungültige Disparitätswerte, für die Verarbeitung ist aber die Abgrenzbarkeit zwischen Personen und der Personen vom Hintergrund entscheidender.

Für den praktischen Einsatz ist die zeitliche Verfügbarkeit von 3D-Informationen von wesentlicher Bedeutung. Zur Bewertung wurden unter Einsatz einer aktuellen GPU (Nvidia GeForce RTX 3070 Mobile mit 140W Max GPU Power) Vergleichsexperimente durchgeführt. Für SGM (640x480 Pixel, Disparitätsbereich 256 Pixel, Subpixel-Präzision) wurde eine durchschnittliche Verarbeitungsrate von 245 Hz ermittelt, für HITNET (640x480 Pixel, Disparitätsbereich 400 Pixel) 6,5 Hz. Die anvisierte Verarbeitungsrate der gesamten Verarbeitungskette entspricht der Bildwiderholrate des eingesetzten THz-Scanners (je nach Modell und Konfiguration ca. 10-25 Hz). Diese Verarbeitungsraten können bei der aktuellen Netzauswahl und HITNET-Konfiguration aktuell nur mit sehr leistungsstarker GPU-Hardware erreicht werden. SGM bietet dagegen die Möglichkeit für größere und variable Eingangsauflösungen.

Für den im nächsten Abschnitt dargestellten Schritt der Bilddatenfusion werden die Disparitätsbilder unter Einsatz der Kalibrierparameter zu Tiefenbildern konvertiert.

### 3.3 Adaptive geometrische Überlagerung der multimodalen Bilddaten

Die Informationsfusion durch geometrische Bildregistrierung kann einerseits dadurch geschehen, dass ein Texturbild (z.B. RGB oder THz) auf ein Tiefenbild abgebildet

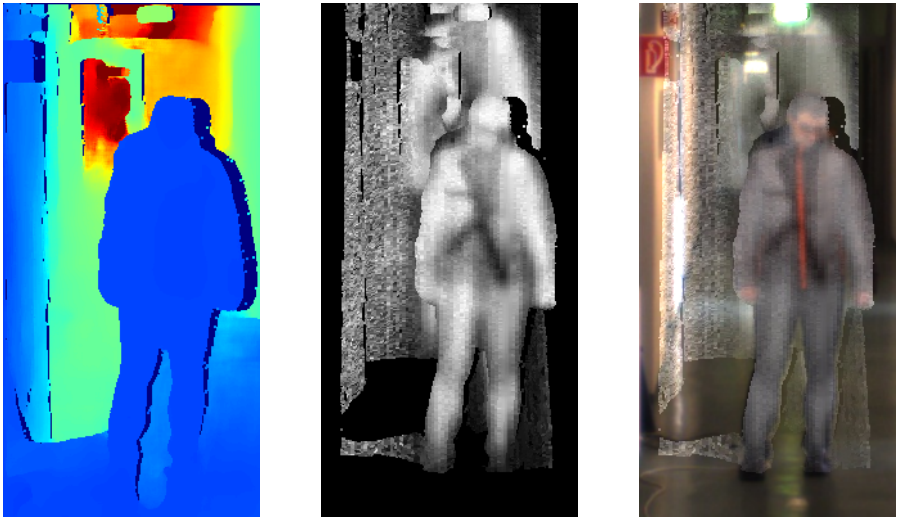

**Abb. 6:** Links: Tiefenbild aus HITNET-Inferenz nach geometrischer Registrierung mit der RGB-Referenzbildmodalität. Mitte: THz-Bild nach geometrischer Registrierung mit der RGB-Referenzbildmodalität. Rechts: Gemeinsame Darstellung (Überlagerung) von RGB- und THz-Bildmodalitäten.

wird. Andererseits kann ein Tiefenbild auf ein optisches Texturbild abgebildet werden. Diese beiden Abbildungsrichtungen erfordern separate Implementierungen, die aufgrund der Datenmengen zeiteffizient umgesetzt werden müssen. Da RGB als Referenzbildmodalität festgelegt wurde, sind beide Richtungen erforderlich. Zunächst muss das Tiefenbild auf die RGB-Modalität abgebildet werden (vgl. Abb. 6 links), erst dann können weitere Texturdaten (hier THz) auf die fusionierte RGB-D-Bildmodalität abgebildet werden (vgl. Abb. 6 Mitte bzw. rechts). Die Umsetzung erfolgte durch effiziente vektorielle CPU-Implementierungen beider Abbildungsrichtungen unter Verwendung einer SIMD-Parallelisierung (Single Instruction, Multiple Data) zur Verfahrensbeschleunigung.

Die Sensorgrößen der ergänzenden Sensoriken (1x Farbe, 2x Grauwert, insgesamt 5 Kanäle) beträgt jeweils 2456 x 2054 Pixel. Zum Erreichen der anvisierten Verarbeitungsrate ist eine frühe Datenreduktion in der Verarbeitungskette notwendig. Auch in Hinblick auf Variabilität der Arbeitsabstände bei dynamischen Szenen erfolgt ein digitales Zoom sowie die Skalierung auf eine Zielauflösung in einem effizienten Einzelschritt. Neben der RGB-Referenzbildmodalität erfolgt dies auch für die Tiefenbildmodalität. Die Parameter der geometrischen Kalibrierung werden hierbei berücksichtigt, sodass unabhängig von gewählten Zoomfaktoren und Zielauflösungen stets die korrekte Registrierung aller Bildmodalitäten erfolgt.

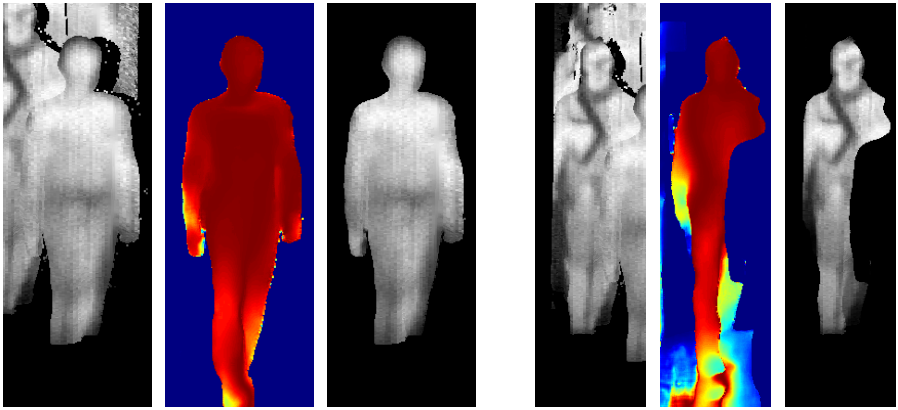

**Abb. 7:** Linke Bildergruppe: Robuste Abgrenzung einer vorderen Person von einer hinteren im THz-Bild durch Verrechnung mit dem farbig codiert dargestellten Abstandsmerkmalbild. Rechte Bildergruppe: Abgrenzung einer hinteren Person von einer vorderen erfolgt in gleicher Art und Weise.

## 4 Hierarchische Auswertung

Auf Basis des multimodalen Bildstapels soll eine Bildanalyse in verschiedenen Semantikebenen erfolgen, um beispielsweise Personen vom Hintergrund zu trennen oder verborgen getragene Objekte zu segmentieren und zukünftig zu identifizieren. Die korrekte Zuordnung von Objekten zu Personen ist in der Anwendung besonders wichtig. Steht nur die THz-Bildmodalität zur Verfügung, ist die semantische Trennung von überlappenden Personen nicht einfach möglich (vgl. Abb. 1 links). In diesem Abschnitt wird ein Beispielverfahren aus der gesamten multimodalen hierarchischen Datenauswertung beschrieben, welches diese Trennung robust ermöglicht. Ausgenutzt wird hierbei, dass schwierige Überlappungsfälle im Regelfall durch hintereinander versetzt befindliche Personen wie im Bildbeispiel (Abb. 7) begründet sind.

Zunächst wird ein vom ZBS e. V. bereits entwickeltes Verfahren zur Detektion mehrgliedriger 2D-Skelettmodelle (vgl. Abb. 1 rechts) aus der RGB-Bildmodalität angewendet. Aus der Tiefenbildmodalität und dem überlagerten Skelettmodell wird für jede Person ausgewertet, in welchem Abstand sich die Person zum Sensorsystem befindet. Anschließend wird für jede Person im relevanten Bildausschnitt ein Abstandsmerkmalbild (farbcodierte Bilder in Abb. 7) generiert, das hohe Intensitäten für Pixel mit gerigem Tiefenabstand zur jeweiligen Person aufweist. Durch multiplikative Verrechnung der personenbezogenen Abstandsmerkmalbilder mit dem THz-Bild werden angrenzende Personen ausgeblendet.

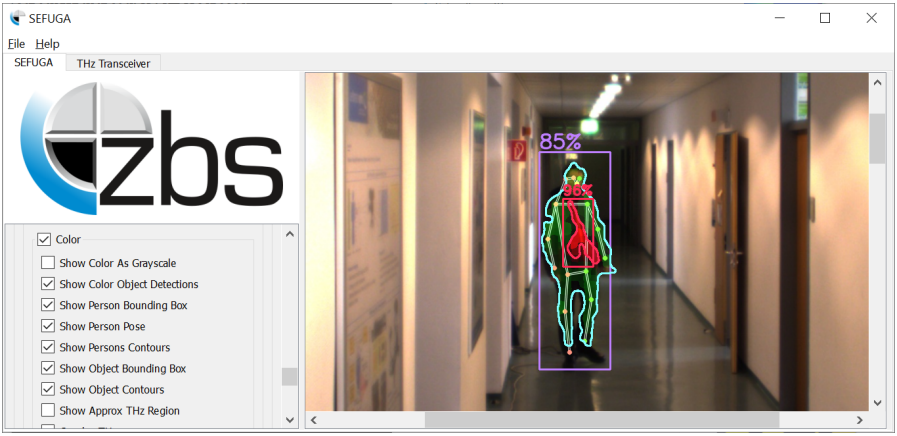


**Abb. 8:** Ergebnisdarstellung in der SEFUGA-Softwareapplikation. Für Objekte und Personen werden Kennzahlen für das ermittelte Gefahrenpotential angezeigt.

## 5 Zusammenfassung und Ausblick

Im Hinblick auf die Gefahrenerkennung sind weitere höhere Analyseschritte erforderlich, die in diesem Beitrag nicht ausführlich behandelt werden. So wird neben der Abgrenzung von Personen zum Hintergrund und untereinander auch die Erkennung von Objekten sowie eine Klassifikation von Gefahrenobjekten erfolgen.

Im beschriebenen Vorhaben wurde von ZBS e. V. die SEFUGA-Softwareapplikation (siehe Abb. 8) entwickelt, die eine komplette Verarbeitungskette zur Sensordatenfusion und Gefahrenerkennung integriert. Diese umfasst weitere, hier nicht im Detail beschriebene Verfahren auf Grundlage multimodaler Daten. Zur Segmentierung von Objekten wurde ein Verfahren basierend auf multimodaler Klassifikation von Pixelgruppen umgesetzt. Insbesondere werden hierdurch auch Problemfälle behandelt, z.B. Abbildung von Objekten am Rand von Personen in der THz-Modalität. Dies ist häufig der Fall, wenn Personen nicht frontal zum THz-Scanner ausgerichtet erfasst werden. Zur Bestärkung von Objekthypothesen sowie zur Objektklassifikation wurde ein neuronales Detektionsnetzwerk ergänzt. Weitere Robustheit gegenüber Fehldetektionen wird durch Auswertung des zeitlichen Verhaltens (zeitliche Modalität aus Datenverarbeitung) erzielt. Hier erfolgt ein getrenntes Tracking von Personen und Objekten sowie die Zuordnung der Objekte einschließlich Gefahrenpotential zu den Personen. Durch Einsatz der Tiefenbilder erfolgt eine Plausibilitätskontrolle beim Tracking und es kann zukünfigt eine Fokusnachführung der THz-Bildgebung umgesetzt werden.

Es ist abzusehen, dass sich neuronale Ansätze für einzelne Teilaspekte der Verarbeitung weiter durchsetzen werden, beispielsweise zur Tiefenbildrekonstruktion, zur Personendetektion- und Segmentierung und zur Objektdetektion. Insbesondere in Hinblick auf die Erkennung von Gefahrenobjekten ist hierbei die Verfügbarkeit geeigneter Lernstichproben eine wichtige Fragestellung für die Einsatzfähigkeit in der Praxis. Es ist zu erwarten, dass gute Belehrungsdaten auch im Rahmen des Feldeinsat-

zes der entwickelten Lösung gewonnen werden können. In Hinblick auf die neuronale Tiefenbildrekonstruktion stehen zwar generische Netzwerkgewichte zur Verfügung, jedoch könnten auch diese durch anwendungsspezifische Trainingsdaten optimiert werden. Da die hier verwendete HITNET-Architektur auch RGB-Daten verarbeiten kann, bietet sich außerdem eine praktische Evaluierung der Verwendung von zwei RGB-Kameras gegenüber zwei Mono-Kameras und einer RGB-Kamera an.

Die im Beitrag diskutierten Methoden und Verfahren der Verarbeitung von Daten aus multimodaler, multisensorieller Bildgebung sind in einem breiten Spektrum von Applikationen aus dem Bereich der Sicherheits- und Überwachungstechnik anwendbar. Mit den vorgestellten Methoden zur Bilddatenfusion lassen sich beispielsweise einfach Bildmodalitäten zur Infrarotabtastung einer Szene ergänzen, die für den gezeigten Anwendungsfall weitere nutzbare Szeneninformationen bereitstellen könnten. Für Konzeption und Aufbau ggf. erforderlicher spezifischer Kalibriertargets bestehen am ZBS weitreichende Erfahrungen.

## Literatur


[1] M. Kowalski, M. Kastek, N. Palka, H. Polakowski, M. Szustakowski und M. Piszczek, „Investigation of concealed objects detection in visible, infrared and terahertz ranges of radiation“, *Photonics Letters of Poland*, Jg. 5, Nr. 4, S. 167–169, 2013.

[2] S. L. Tapia, R. Molina und N. P. de la Blanca, „Detection and localization of objects in passive millimeter wave images“, in *2016 24th European Signal Processing Conference (EUSIPCO)*, IEEE, 2016, S. 2101–2105.

[3] X. Shen, C. R. Dietlein, E. Grossman, Z. Popovic und F. G. Meyer, „Detection and segmentation of concealed objects in terahertz images“, *IEEE transactions on Image Processing*, Jg. 17, Nr. 12, S. 2465–2475, 2008.

[4] T. Li, Y. Shen, W. Jin, G. Zhao und Y. Cai, „Processing and fusion of passively acquired, millimeter and terahertz images of the human body“, *Optical Engineering*, Jg. 56, Nr. 4, S. 043 102, 2017.

[5] H. Hirschmüller, „Stereo processing by semiglobal matching and mutual information“, *IEEE Transactions on pattern analysis and machine intelligence*, Jg. 30, Nr. 2, S. 328–341, 2007.

[6] Z. Zhang, „A flexible new technique for camera calibration“, *IEEE Transactions on pattern analysis and machine intelligence*, Jg. 22, Nr. 11, S. 1330–1334, 2000.

[7] D. Vehar, R. Nestler und K.-H. Franke, „3D-EasyCalib™-Toolkit zur geometrischen Kalibrierung von Kameras und Robotern“, *22. Anwendungsbezogener Workshop zur Erfassung, Modellierung, Verarbeitung und Auswertung von 3D-Daten, 3D-NordOst*, S. 15–26, Dez. 2019.

[8] V. Tankovich, C. Hane, Y. Zhang, A. Kowdle, S. Fanello und S. Bouaziz, „Hitnet: Hierarchical iterative tile refinement network for real-time stereo matching“, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, S. 14 362–14 372.

[9] D. Scharstein u. a., „High-resolution stereo datasets with subpixel-accurate ground truth“, in *German conference on pattern recognition*, Springer, 2014, S. 31–42.