

## 3D-Aspekte bei der Bildfusion am Beispiel von IR- und VIS-Daten

Dipl.-Ing. Torsten Koch und PD Dr.-Ing. habil. Karl-Heinz Franke  
Zentrum für Bild- und Signalverarbeitung e.V.  
Gustav-Kirchhoff-Strasse 5, D 98693 Ilmenau  
e-Mail: [franke@tu-ilmenau.de](mailto:franke@tu-ilmenau.de), URL: <http://www.zbs-ilmenau.de>

### Abstract

Die Fusion von Bilddaten, die aus unterschiedlichen spektralen Bereichen stammen (z.B. sichtbares Licht und Infrarot) und die mit unterschiedlichen Sensorsystemen von unterschiedlichen Kamerastandorten aus erzeugt wurden, eröffnet neue Möglichkeiten für die Interpretation dieser Informationen, insbesondere bei der Schadensfrüherkennung an elektrischen Einrichtungen und an Bauwerken. Die bei der Zentralprojektion durch die unterschiedlichen Kameraorientierungen verursachten perspektivischen Verzerrungen müssen neben den Unterschieden in Maßstab und Auflösung ausgeglichen werden, um die Überdeckung der Bilder zu erreichen. Der Beitrag beschreibt zwei Verfahren zur Ermittlung der Korrespondenzparameter, die vom dreidimensionalen Charakter der Szene ausgehen und keinen Kalibriereschritt benötigen. Das erste Verfahren beruht auf der inversen View-Interpolation, das zweite basiert auf dem Einpassen eines angepassten Objektmodells in die Szene.

### 1 Einleitung

Das Problem der Fusion von Bilddaten, die aus unterschiedlichen spektralen Bereichen stammen, die zu unterschiedlichen Zeiten nach völlig unterschiedlichen Wirkprinzipien mit unterschiedlichen Sensorsystemen an unterschiedlichen Orten erzeugt wurden, oder auch die Fusion von verschiedenen Merkmalskarten mit unterschiedlichen Auflösungen und variablen Gewichten gewinnt im Zuge des rascher wachsenden technologischen Fortschritts immer mehr an Bedeutung. Aufgabe einer solchen Bildfusion ist es, Pixelattribute aus mehreren Bildern miteinander zu kombinieren. Ziel ist dabei eine möglichst pixelgenaue Registrierung der Bilddaten.

Ein potentielles Anwendungsgebiet einer solchen Bildfusion stellt die Thermografie dar. Hier kann die Bildfusion eingesetzt werden, um nur im Infraroten "sichtbare" Effekte mit dem visuellen Strahlungsbereich zu assoziieren.

Jedoch werden am Beispiel der Thermografie besondere Probleme deutlich, die es zu lösen gilt. Bedingt durch ihr Transmissionsverhalten können für den thermischen IR-Kanal keine optische Gläser eingesetzt werden. Statt dessen werden Materialien, wie z.B. Germanium, Silizium oder spezielle Legierungen, angewandt. Da keine Materialien bekannt sind, die für beide Spektralbereiche durchlässig und zugleich für die optische Abbildung geeignet sind, muss eine Aufteilung des Strahlengangs vorgenommen werden. Eine Möglichkeit ist die Verwendung zweier Kameras, deren optische Achsen geeignet ausgerichtet oder auch „beliebig“ orientiert sind.

Die bei Zentralprojektion (Standardfall) durch die unterschiedlichen Kameraorientierungen verursachten perspektivischen Verzerrungen müssen neben Unterschieden in Maßstab und Auflösung ausgeglichen werden, wenn eine Bildfusion angestrebt ist. Dabei spielt die Dreidimensionalität der Szene eine wesentliche Rolle. Es muss die vollständige 3D-Information extrahiert werden oder es sind angepasste 3D-Objektmodelle in die Szene einzupassen.

## **1.1 Lösungsansätze**

Für die Fusion von (thermischen) IR- und VIS-Bildmaterial sind zwei Aspekte von besonderer Bedeutung.

Einerseits erzeugen, wie einleitend geschildert, global getrennte Strahlengänge unterschiedliche Szenenansichten, und andererseits ist das IR- und VIS-Bildmaterial oft nur schwach miteinander korreliert. Letzteres erschwert merkmals- oder intensitätsbasierte Stereoansätze oder macht sie gänzlich unmöglich.

Zur Lösung des Ansichtenproblems beschreibt der Beitrag zunächst einen Ansatz auf der Basis einer inversen View-Interpolation. Das zweite, hier vorgestellte Verfahren arbeitet modellbasiert. In beiden Fällen werden Lochkameramodelle mit bekannter innerer Orientierung (Brennweite, Hauptpunkt usw.) vorausgesetzt. Radialsymmetrische Verzerrungen der Ausgangsbilder werden in einem Vorverarbeitungsschritt (z.B. nach Tsai) beseitigt.

## **1.2 Inverse View-Interpolation**

Die Technik der View-Synthese wurde entwickelt, um eine Zentralansicht bei Telekonferenzanwendungen zu erzeugen (vgl. [Chen93][Shar99]). Mit ihrer Hilfe wird aus den Abbildungen zweier Kameras eine neue virtuelle Kameraansicht generiert. Voraussetzung hierfür ist die genaue Kenntnis der gegenseitigen Kameraanordnung.

Das Prinzip basiert auf einer Korrespondenzanalyse, welche zur Bestimmung homologer Punkte in beiden Kamerabildern eingesetzt wird. Zu jedem gefundenen homologen Punktpaar kann unter Kenntnis der o.g. Kameraparameter die Position des entsprechenden Punktes im (virtuellen) Zielbild<sup>1</sup> bestimmt werden.

Sind die Projektionszentren der drei Kameras kollinear, wird von einer *View-Interpolation* gesprochen. In [Chen93] wird darauf hingewiesen, dass die lineare View-Interpolation eine korrekte perspektivische Ansicht erzeugt, wenn die Basislinie parallel zu den Abbildungsebenen liegt. Für den hier geschilderten Anwendungsfall soll zusätzlich Parallelität sowohl der optischen Achsen als auch der Sensorzeilen vorausgesetzt werden. Eine solche Kameraanordnung wird auch als Standardgeometrieordnung bezeichnet. Diese gewährleistet neben einer Optimierung der Suche nach homologen Punktpaaren auch die Anwendbarkeit in einem weiten Tiefenbereich.

---

<sup>1</sup> z.B. an einer Kameraposition zwischen den beiden Ausgangskameras

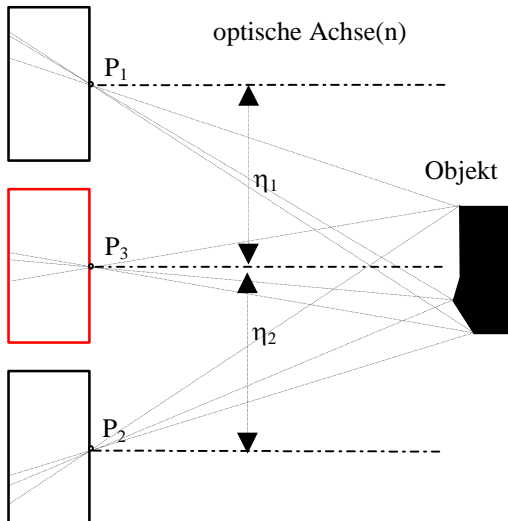


Abb. 1 Kameraanordnung (IR-Kamera in der Mitte;  $P_1, P_2, P_3$  Projektionszentren)

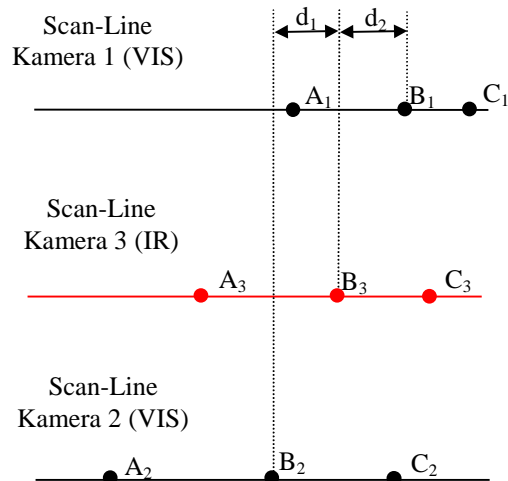


Abb. 2 Beispiel jeweils einer bestimmten Scan-Line aus den drei Kameras mit 3 homologen Punkt-Tripeln

Um eine pixelgenaue Registrierung für VIS und IR zu realisieren, wird im Kontext der Bildfusion die Umkehrung des Prinzips der View-Interpolation verwendet.

Wie im Falle der View-Interpolation werden auch hier zwei VIS-Kameras (Sicherung einer guten Bildkorrelation) vorausgesetzt. Den Platz der virtuellen Kamera bei der View-Interpolation nimmt allerdings nun der IR-Kanal ein.

Die Fusion des IR-Kamerabildes mit denen der VIS-Kameras wird dadurch erleichtert, dass homologe Punkte auf den Epipolarlinien [Klett96] zu suchen sind. Diese sind bei Anwendung einer Standardgeometrieordnung identisch mit den Scan-Lines der CCD-Sensoren. Der jeweils zugehörige Punkt des IR-Bildes befindet sich demzufolge auf der gleichen Scan-Line wie die entsprechenden homologen Punkte der beiden VIS-Kameras. Dies gilt allerdings nur, solange die Kammerkonstanten aller drei Kameras identisch sind. Wenn dies nicht der Fall ist, müssen die Kamerabilder mittels Skalierung zusätzlich angepasst werden. Die Suche nach homologen Punkten wird z.B. mit Hilfe eines Block-Matching-Verfahrens realisiert [Klett96].

Jedem homologen Punktpaar aus den beiden VIS-Bildern kann ein homologer Punkt im IR-Bild zugeordnet werden (Registrierung). Die Koordinate dieses Bildpunktes wird aus einer Interpolation zwischen den Koordinaten der beiden anderen homologen Bildpunkte im Verhältnis  $d_1/d_2 = \eta_1/\eta_2$  bestimmt (siehe Abb. 1 und Abb. 2).

Können keine homologen Punkte in den visuellen Kanälen gefunden werden, läßt sich keine Aussage über die Position des homologen Punktes (sofern vorhanden) im IR-Kanal treffen. Dieses Problem tritt beispielsweise dann auf, wenn aus perspektivischen Gründen Abschattungen oder Verdeckungen entstehen (sogenannte Löcher). Sind hingegen homogene Bildinhalte vorhanden, die es nicht gestatten, mit Hilfe der Korrelation homologe Punkte sicher zu finden, wird von sogenannten Lücken gesprochen. In letztgenanntem Fall schlägt z.B. [Schar99] eine Interpolation der Disparitäten zwischen den beiden nächstgelegenen<sup>1</sup> homologen Punktpaaren vor, um die wahrscheinlichsten Punktkorrespondenzen zu finden.

Nach erfolgter Feststellung der Korrespondenzen zwischen IR- und einem der beiden VIS-Bilder kann nun ein neues Bild<sup>2</sup> erzeugt werden, in dem die Intensitätswerte aus

<sup>1</sup> auf der gleichen Scan-Line

<sup>2</sup> aus der Perspektive des betreffenden VIS-Bildes

dem IR-Bild an den Positionen der zugehörigen Pixel des VIS-Bildes eingetragen sind. Nicht zuzuordnende Pixel müssen auf einen Standardwert gesetzt werden.

### 1.3 Modellbasierte Bildfusion

Mit Hilfe einer Objektmodellierung ist es ebenfalls möglich, eine pixelweise Registrierung zweier Kamerabilder aus unterschiedlichen spektralen Bereichen durchzuführen.

Das hier beschriebene Verfahren realisiert eine selektive<sup>1</sup> Fusion auf Basis eines Polyedermodells des Objektes. Es setzt voraus, dass das betreffende Objekt nicht deformierbar, durch Polyeder approximierbar und möglichst bildfüllend dargestellt ist. Weiterhin müssen mehrere gerade Objektkanten  $l_j$  aus dem Bild extrahierbar und durch ein externes Verfahren den entsprechenden Modellkanten  $k_j$  zuzuordnen sein. Diese Zuordnungen (Kantenkorrespondenzen) können z.B. durch geeignete Objekterkennungsverfahren ermittelt oder interaktiv festgelegt werden. Die Extraktion der Bildkanten wird mit einem geeigneten Kantendetektionsverfahren realisiert.

Basis der modellbasierten Bildfusion ist die Idee, dass mit Hilfe einer Rückprojektion über ein Modell (siehe Abschnitt 1.3.1) der Bildinhalt einer zweiten Kamera mit dem der ersten gemappt wird. Hierbei ist es zwingend erforderlich, die Lage und Orientierung des Objekts (Aspekt) in beiden Kamerakoordinatensystemen<sup>2</sup>  $K_1$  und  $K_2$  zu ermitteln. Hierzu wird eine Aspektbestimmung angewandt (siehe Abschnitt 1.3.2).

Das Ergebnis dieser Bildfusion stellt ein generiertes Zwischenbild (aus Sicht der VIS-Kamera) dar (siehe Abb. 12), in welchem die Intensitätswerte der objektspezifischen Bildpunkte der IR-Kamera an die Positionen der jeweiligen homologen Bildpunkte der VIS-Kamera versetzt sind. Alle anderen nicht zuzuordnenden Bildpunkte werden auf einen definierten Hintergrundwert gesetzt.

Im Sinne einer besseren Visualisierung kann dieses Zwischenbild als Falschfarbdarstellung dem visuellen Bild überblendet werden (siehe Abb. 14).

#### 1.3.1 Rückprojektion

Die Rückprojektion realisiert die Registrierung der Bildpunkte der beiden Kamerabilder. Das zugrundeliegende Prinzip beruht auf der Überlegung, dass ausgehend von jedem Bildpunkt der Kamera 1  $p_{1,n}$  ein Sehstrahl durch das Projektionszentrum dieser Kamera gelegt und mit den Oberflächen eines Objektmodells geschnitten wird (siehe Abb. 3). Im allgemeinen entstehen dabei mehrere Schnittpunkte, wovon nur der zu Kamera 1 nächstgelegene Schnittpunkt  $s_n$  betrachtet wird. Ausgehend von diesem Schnittpunkt wird durch das Projektionszentrum von Kamera 2, ein weiterer Sehstrahl geführt und mit deren Abbildungsebene geschnitten. Dieser Schnittpunkt  $p_{2,n}$  stellt den zu  $p_{1,n}$  homologen Bildpunkt dar.

Der Intensitätswert an der (i.d.R.) nicht-ganzzahligen Position  $p_{2,n}$  wird durch eine bilineare Interpolation zwischen den vier Nachbarpixeln bestimmt.

Die notwendige Information über die Lage und Orientierung des Objektes bezüglich der beiden Kameras wird durch das Aspektbestimmungsverfahren (s.u.) ermittelt.

Bei fester Kameraanordnung kann die Orientierung der zweiten Kamera aus der Orientierung der ersten mit Hilfe des Aspektunterschieds zwischen beiden Kameras gewonnen werden. Dieser Aspektunterschied wird zuvor mit Hilfe einer Kalibrierung bestimmt und stellt eine Koordinatentransformation von  $K_1$  nach  $K_2$  dar.

---

<sup>1</sup> nur das Objekt betreffend - keine Fusion des Hintergrundes

<sup>2</sup> math. positives 3D-Koordinatensystem; Ursprung ist Projektionszentrum; z-Achse = optische Achse

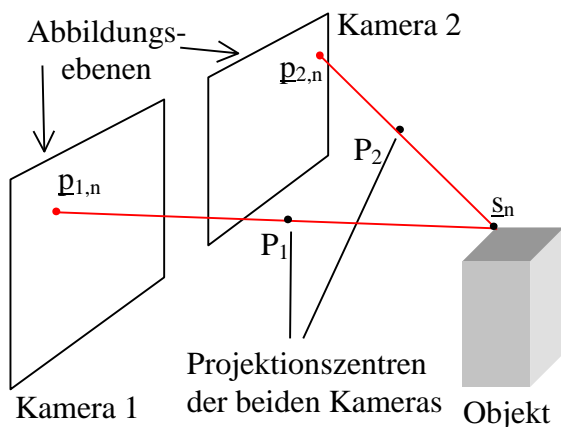


Abb. 3 Prinzip der Rückprojektion

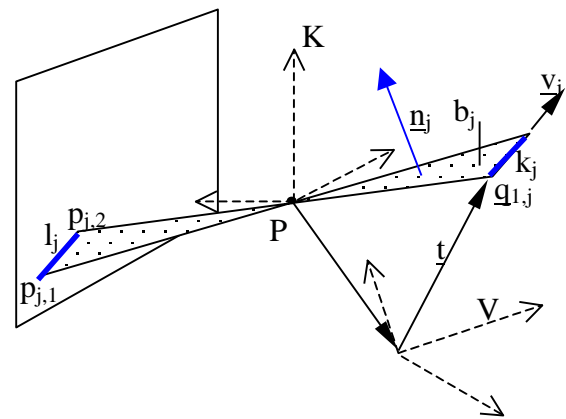


Abb. 4 Beziehung zwischen Modelllinie  $k_j$  und Bildlinie  $l_j$  bei Aspektbestimmung

### 1.3.2 Aspektbestimmung

Als Aspekt (bzgl. Kamera  $x$ ) sei die Transformationsvorschrift Gl. (1) definiert, welche Punkte aus dem Koordinatensystem des Modells ( $V$ ) in das der Kamera  $x$  ( $K_x$ ) überführt.

Das hier beschriebene Verfahren zur Aspektbestimmung basiert auf dem Rekonstruktionsverfahren für allgemeine Kanten-Konfigurationen nach [Voss95]. Dieses Verfahren zeichnet sich insbesondere dadurch aus, dass beliebig viele Kanten verwendet werden können. Hierdurch können Genauigkeitsfehler bei der Kantenextraktion ausgeglichen werden.

Es stellt im Wesentlichen ein zweistufiges Verfahren dar, welches im ersten Schritt den Rotationsanteil des Objektes (beschrieben durch die Rotationsmatrix  $\underline{R}$ ) iterativ bestimmt und im zweiten Schritt den Translationsvektor  $\underline{t}$  mit Hilfe einer linearen Ausgleichsrechnung ermittelt. Allerdings setzt dieses Verfahren einen Startwert für die iterative Bestimmung von  $\underline{R}$  voraus. Sollte dieser ungünstig gewählt sein, läuft dieses Verfahren auf eine suboptimale Lösung von  $\underline{R}$  zu.

$$f_x(\underline{o}) := \underline{R} \cdot \underline{o} + \underline{t}; \quad \underline{o} \in V, \underline{t} \in R^3, f_x(\underline{o}) \in K \quad (1)$$

$V$  Modellkoordinatensystem,  $K$  Kamerakoordinatensystem,  $\underline{R}$  Rotationsmatrix(3x3),  
 $\underline{t}$  Translationsvektor,  $\underline{o}$  Objektvertices

Nach der Bestimmung von  $\underline{R}$  und  $\underline{t}$  können die Modellkoordinaten mittels Gl. (1) in das Kamerakoordinatensystem transformiert werden. Damit ist der Aspekt vollständig definiert.

## 2 Ergebnisse

Zu Demonstrationszwecken wurde für jedes Verfahren eine Beispielszene bearbeitet. In Abb. 5 bis Abb. 9 wird dargestellt, welche Fusionsergebnisse durch eine inverse View-Interpolation zu erwarten sind. Auf die Interpolation der Disparitäten bei auftretenden Lücken<sup>1</sup> wurde aus Aufwandsgründen genauso verzichtet wie auf den Einsatz

<sup>1</sup> aufgrund homogenen Bildinhaltes in beiden VIS-Bildern

einer Heuristik zur Vermeidung von mehrfachen Zuordnungen eines IR-Bildpunktes (siehe Dopplungen in Abb. 8, unten links). Weiterhin kann es zu Fehlzusordnungen (oben rechts) kommen.

Das Beispiel zur modellbasierten Bildfusion stellt Abb. 12 dar. Die entsprechenden Ausgangsbilder für VIS- und IR-Kanal sind in Abb. 10 und Abb. 11 dargestellt. Wie bereits erwähnt, wird eine selektive Bildfusion realisiert, d.h. nur die Pixel, die gemäß Modell zum Objekt gehören, werden fusioniert. Mögliche Fusionsfehler werden in Abb. 13 und Abb. 14 (Überblendung von VIS-Bild und Fusionsergebnis) deutlich gemacht. Die Fusionsfehler in Abb. 13 sind durch eine ungenauere Aspektbestimmung entstanden (es wurden weniger Kantenkorrespondenzen verwendet als in Abb. 14). Andere Fusionsfehler sind bei Abweichungen zwischen Modell und realem Objekt zu erwarten.

### **3 Zusammenfassung und Ausblick**

Im Rahmen dieses Beitrages wurden zwei Verfahren vorgestellt, die es gestatten, Bildmaterial aus unterschiedlichen spektralen Bereichen zu fusionieren. Hierbei wurde das Problem der perspektivischen Verzerrungen zwischen IR- und VIS-Bild auf unterschiedliche Weise gelöst.

Im Falle der inversen View-Interpolation zeigten die Ergebnisse, dass nicht von einer vollständig fehlerfreien Fusion ausgegangen werden kann. Daher eignet sich dieses Verfahren vor allem für allgemeine, nicht sicherheitskritische Einsatzfälle.

Der Vorteil dieses Verfahrens gegenüber dem modellbasierten Ansatz ist vor allem darin zu sehen, dass keine Objektmodellierung vorausgesetzt werden muß. Als nachteilig erweist sich jedoch der relativ hohe Rechenaufwand, der auf die Korrespondenzanalyse zurückzuführen ist.

Die modellbasierte Bildfusion eignet sich - abgesehen von der Voraussetzung der Modellierbarkeit - für Einsatzfälle, in denen eine feste und kalibrierte IR-VIS-Kameraanordnung vorliegt. Es kann sogar nachträglich Bildmaterial - ohne Kenntnis der Kameraanordnungen - selektiv fusioniert werden, wenn extrahierbare Kanten im IR-Bild vorhanden sind. Da nur zwei Kameras zum Einsatz kommen, ist der gerätetechnische Aufwand geringer als bei der inversen View-Interpolation.

Als Nebenprodukt können beide Verfahren Tiefeninformation zu jedem registrierten Bildpunkt liefern. Dies ist insbesondere für die Thermografie von Bedeutung, da zur genauen Temperaturbestimmung neben der Intensität u.a. auch die Entfernung und Oberflächeneigenschaften, wie z.B. Emissionskoeffizienten, bekannt sein müssen. Eine Modellierung bietet in diesen Zusammenhang die zusätzliche Möglichkeit, auch laterale Oberflächeneigenschaften des Objekts zu speichern, die im Bedarfsfall für jeden einzelnen objektspezifischen Bildpunkt abgerufen werden können.

Für die Fusion von infraroten und visuellen Bildern ergeben sich vielfältige Anwendungsmöglichkeiten. Dazu gehört die Analyse der thermischen Isolation bei Industrieprodukten und im Bauwesen. Ein weiteres interessantes Anwendungsfeld ist die Qualitäts- und Funktionskontrolle von elektrotechnischen oder elektronischen Geräten im Produktionsprozess sowie die Überwachung und Diagnostik von elektrischen Anlagen. Die Fusion von IR und VIS erlaubt die einfache und eindeutige Zuordnung von defekten Baugruppen und Anlagenteilen. Die Ferndiagnostik ist in diesem Zusammenhang bei Betriebsmitteln der Hochspannungstechnik von besonderer Bedeutung.

## 4 Abbildungen

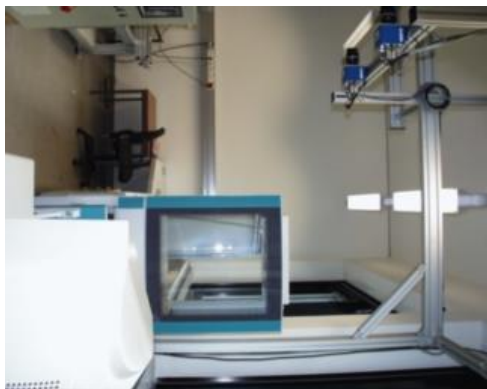


Abb. 5 linkes VIS-Kamerabild

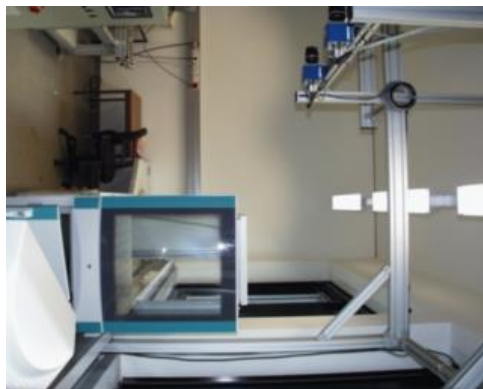


Abb. 6 rechtes VIS-Kamerabild

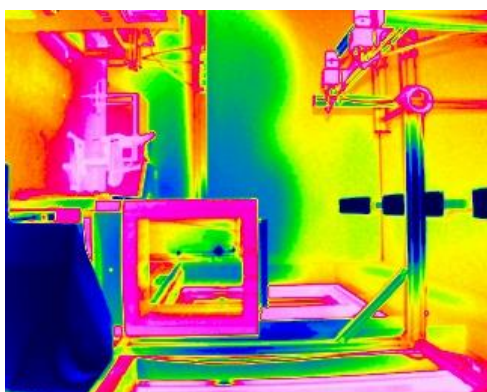


Abb. 7 mittleres (IR-) Kamerabild

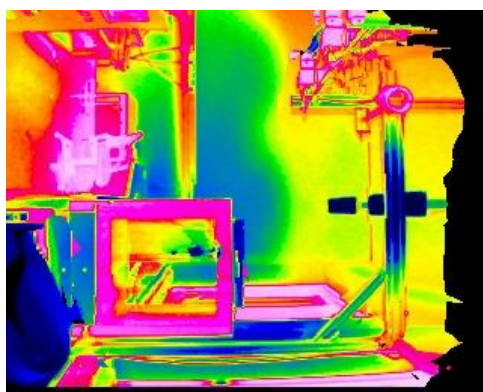


Abb. 8 Ergebnis (IR- auf rechtes VIS-Bild gemappt)

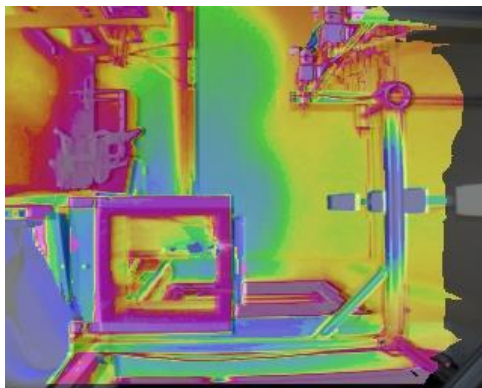


Abb. 9 Ergebnis (Abb. 8) mit rechtem VIS-Bild überblendet

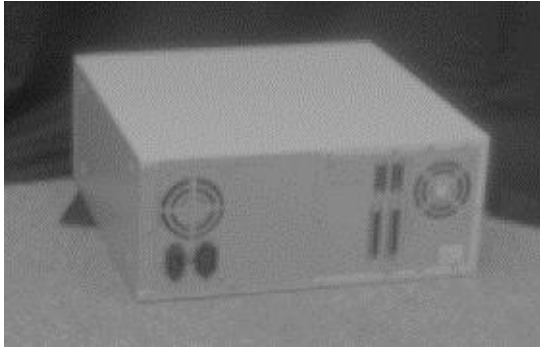


Abb. 10 VIS-Bild

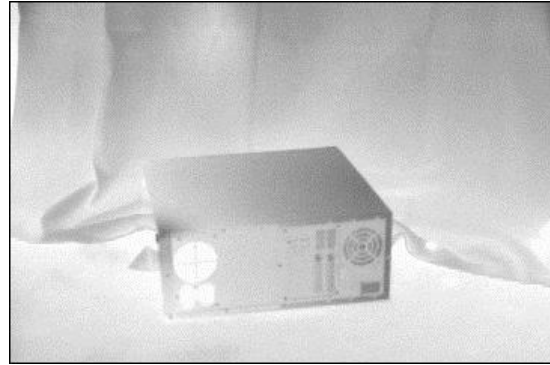


Abb. 11 IR-Bild

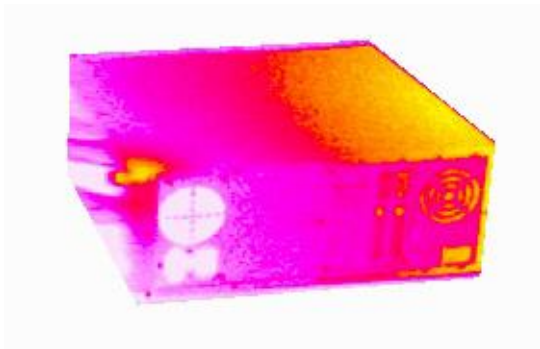


Abb. 12 Ergebnis der Bildfusion (in Falschfarbdarstellung; IR- auf VIS-Bild gemappt)

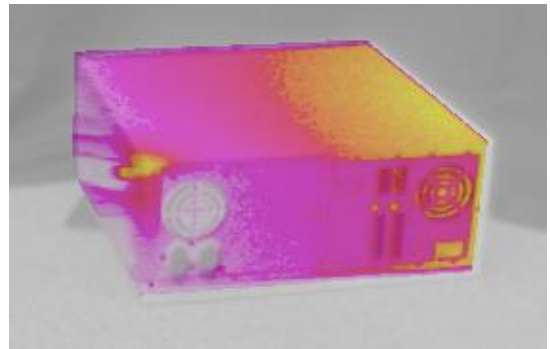


Abb. 13 Ergebnis (nach ungenauer Aspektbestimmung) mit VIS-Bild überblendet

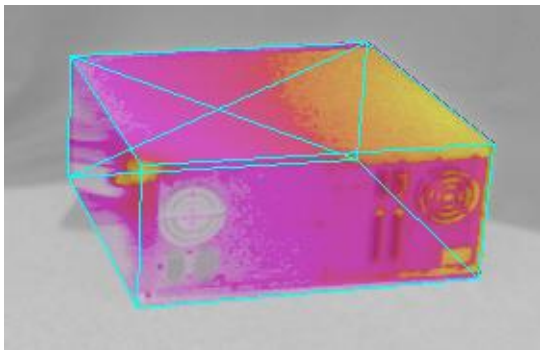


Abb. 14 Ergebnis mit VIS-Bild überblendet (zur Veranschaulichung wurde das Modell eingezeichnet)

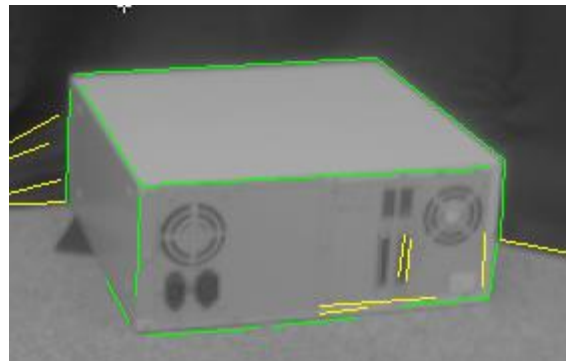


Abb. 15 extrahierte Bildkanten (grüne Kanten korrespondieren mit Modellkanten im Fall von Abb. 14; gelbe wurden nicht berücksichtigt)

## 5 Literatur

- [Chen93] S. E. Chen, L. Williams: View interpolation for image synthesis. In Computer Graphics (SIGGRAPH'93), S.279 – 283; 1993
- [Klett96] Klette, Reinhard; Koschan, Andreas; Schlüns, Karsten; Computer Vision - Räumliche Information aus digitalen Bildern; Vieweg Verlag ; 1996
- [Schar99] Daniel Scharstein: View Synthesis Using Stereo Vision; Springer; 1999
- [Voss95] Klaus Voss; Ralf Neubauer; Michael Schubert; Monokulare Rekonstruktion für Robotvision; Aachen: Shaker 1995