

Bildoptimierung durch kontrolliertes Aktives Sehen und Bildfusion

Image Optimization via Controlled Active Vision and Image Fusion

Jürgen Beyerer und Fernando Puente León

Mit nur einem Kamerabild alleine lässt sich bei schwierigen Objekten oftmals keine Darstellung visuell ausreichender Güte erzielen, unter Umständen auch dann nicht, wenn man die Aufnahmekonstellation optimal wählen würde. Der Grund liegt in der vielgestaltigen Wechselwirkung zwischen Beleuchtung, Objekt und Aufnahmeoptik einerseits und der eingeschränkten „Wahrnehmungsfähigkeit“ eines gängigen intensitätsempfindlichen 2D-Sensors einer Standardkamera für das informationstragende Lichtfeld der Szene andererseits. Für Bilder höherer Güte werden in diesem Aufsatz Bildserien unter gezielter Variation der Parameter der Aufnahmekonstellation gewonnen. Solche Serien, die wesentlich mehr Information tragen als ein einzelnes Bild, werden zu Resultatbildern mit einer von Einzelbildern nicht erreichbaren Qualität verschmolzen.

Am Beispiel der Gewinnung eines Bildes mit gleichmäßigem Signal-zu-Rausch-Verhältnis auf hohem Niveau wird ein finaler Fusionsansatz demonstriert. Ein solcher Ansatz ist sinnvoll, wenn anhand der Bildserie bereits sehr gut auf die Qualität des zukünftigen Fusionsresultates geschlossen werden kann.

Ein zweites Beispiel zeigt, wie für ein optisch schwieriges biologisches Präparat eine Bildserie durch zweidimensionale Variation der Beleuchtungsrichtung gewonnen und iterativ zu einem hochwertigen Ergebnisbild fusioniert wird. Die Qualität des Resultates der mitlaufenden Fusion bestimmt hier den Abbruch der Serienakquisition. In beiden Beispielen wird Bildinformation in die Bildseriengewinnung rückgekoppelt, und so der Akquisitionsprozess im Hinblick auf einen umfassenden aufgabenspezifischen Informationsgewinn geregelt.

With one camera image only, it is often impossible to achieve a representation of sufficient visual quality – even if an optimal image acquisition constellation was chosen. The reason for this is the manifold interaction between the illumination, the object and the observation optics on one hand, and the limited sensing capabilities of common 2-D intensity sensors – as are used in standard cameras – to measure the information-bearing light field on the other hand. To generate images of superior quality, we record series of images with systematically varying parameters of the image acquisition system. Such series, which contain substantially more information than a single image, are then fused to resulting images featuring a quality that cannot be achieved with a single image.

Based on a method to record an image with a uniform SNR at a high level, a final fusion approach is exemplarily demonstrated. Such an approach is sensible, if the image series already yields a very good insight into the quality of the future fusion result.

A second example shows how – for an optically difficult biologic probe – an image series is obtained by varying the illumination direction two-dimensionally and is iteratively fused to a high-quality image result. The quality of the result of the iterative fusion determines in this case the stop of the data acquisition. In both examples, image information is fed back to the image recording system, and this way the data acquisition process is controlled with regards to a global, task-specific information gain.

Schlagwörter: Datenfusion, iterative Fusion, finale Fusion, Bilder hoher Güte, SNR-Maximierung, Beleuchtungsvariation

Keywords: Data fusion, iterative fusion, final fusion, high-quality images, SNR maximization, illumination variation

1 Einleitung

In der automatischen Sichtprüfung werden meist Grauwert- oder Farbkameras eingesetzt. Ihre Sensorchips sind typischerweise nur für Intensitäten (Strahlungsleistung pro Fläche) empfindlich, registrieren also nur einen Teilaspekt des elektromagnetischen Strahlungsfeldes, das die Information über die zu untersuchende Szene trägt. Die interessierenden Objekte der Szene verändern das eingestrahelte Lichtfeld aufgrund ihrer Gestalt und ihrer optischen Materialeigenschaften in komplizierter Weise. Die spektrale Zusammensetzung, die Ausbreitungsrichtungsverteilung, die Intensität, die Polarisation, der Kohärenzzustand und die Phasenlage stellen gewissermaßen unterschiedliche Kanäle dar, die alle jeweils spezifische Informationen über die Szene zu tragen im Stande sind. Klar, dass hier ein einzelnes Intensitätsbild nur eingeschränkt Aufschluss geben kann über die beobachtete Szene und die Eigenschaften der darin befindlichen Objekte.

Will man am preisgünstigen Intensitätssensor festhalten, bietet sich folgender Ausweg zu einer umfassenderen Informationsgewinnung: anstatt eines einzigen Bildes $g(\mathbf{x})$ wird eine Serie von Bildern

$$g_i(\mathbf{x}) := g(\mathbf{x}, \omega_i), \quad i = 1, \dots, N, \quad (1)$$

aufgenommen, wobei von Bild zu Bild die geometrischen und optischen Bestimmungsstücke $\omega = (\omega_1, \omega_2, \dots, \omega_M)$ der Aufnahmekonstellation geändert werden. Bei geeigneter Variation enthält eine solche Bildserie wesentlich mehr Information über die zu untersuchende Szene als ein einzelnes Bild. Die Fähigkeiten des Sensors werden gewissermaßen virtuell erweitert.

Hier stößt man auf das Problem, dass die Variationsmöglichkeiten sehr vielgestaltig sind. Die Komponenten des Vektors ω enthalten z. B. die Parameter der Freiheitsgrade bei der Bildaufnahme:

- Pose der Kamera relativ zur Szene,
- Spektrale Empfindlichkeit, Filter,
- Polarisationseigenschaften der Aufnahmeoptik,
- Entfernungseinstellung des Objektivs,
- Blendeneinstellung,
- Verstärkung,
- Richtungsverteilung des beleuchtenden Lichtes,
- Spektrum des Lichtes,
- Polarisation,
- Kohärenz,
- Phase,
- Leistung,
- örtliche Intensitätsverteilung des beleuchtenden Lichtes usw.

Allein die große Zahl der Bestimmungsstücke der Aufnahmekonstellation macht klar, dass nur mit zielgerichteter Variation eine sparsame und dennoch aussagekräftige Bildserie gewonnen werden kann.

Sind die Eigenschaften der zu untersuchenden Szene und ihrer Objekte weitgehend unbekannt, so ist es schwierig, die Bildserie mit den zugehörigen Einstellungen $\{\omega_i\}$

a priori vollständig zu planen. In diesem Falle ist eine sequentielle Vorgehensweise sinnvoll, bei der, basierend auf der Auswertung bereits aufgenommener Serienbilder, die Parameter ω für die nächste Aufnahme gezielt berechnet werden, um den Zugewinn an Information möglichst zu maximieren. Diese Vorgehensweise entspricht einer iterativen Exploration der Eigenschaften der Szene durch gezielte Variation von ω .

Die Fusion der Bildserie $g_i(\mathbf{x})$ kann ebenfalls iterativ, d. h. mitlaufend erfolgen, oder man sammelt zunächst die ganze Serie und fusioniert anschließend. Die Skizzen in Bild 1 zeigen die prinzipiellen Abläufe beider Varianten.

An diesen Flussdiagrammen lässt sich auch gut ein Unterschied zu gängigen Regelkreisen bei der Gewinnung höherwertiger Bilder, z. B. Autofokus-, Belichtungszeit- und Verstärkungsregelkreise (AGC, *Automatic Gain Control*), diskutieren. Bei allen drei Beispielen interessiert eigentlich nur das erste Bild, nachdem der Regelkreis eingeschwenkt ist, also gewissermaßen das letzte Bild der Serie. Die Bilder auf dem Weg dorthin sind nur Zwischenergebnisse während des Einschwingvorganges.

Bei den Fusionsschemata in Bild 1 hingegen liefert gegebenenfalls jedes Bild der Serie einen Teil der Nutzinformation bei der Erzeugung des Resultatbildes der Fusion. Die Möglichkeit, aus einer ganzen Serie systematisch erzeugter Bilder das jeweils „Beste“ herauszunehmen und in die Fusion einzubringen, begründet das hohe Potenzial dieser Methodik.

In den folgenden Abschnitten werden zwei Beispiele ausführlich dargestellt und diskutiert. Zunächst geht es um

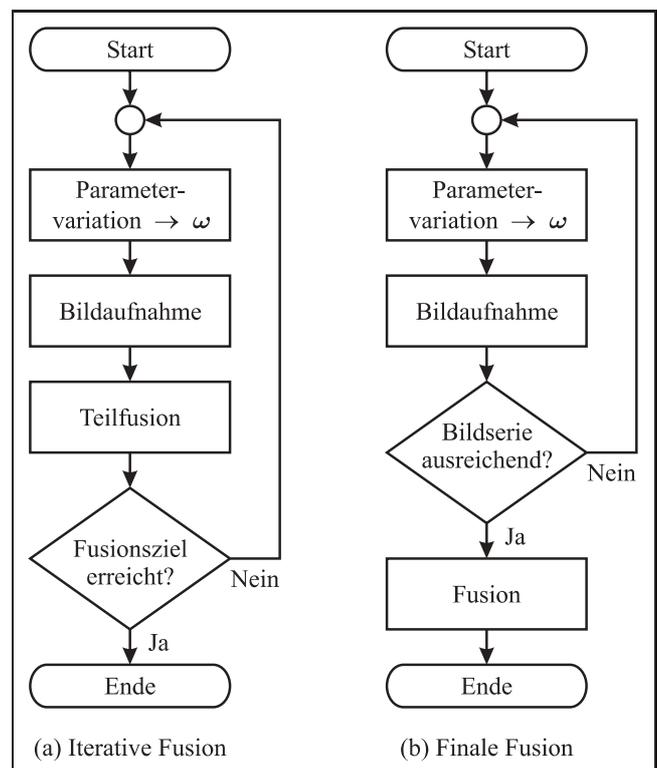


Bild 1: Fusionsvarianten: (a) Iterative Fusion; (b) Finale Fusion.

die Gewinnung eines Bildes mit konstantem Signal-zu-Rauschverhältnis (SNR) auf hohem Niveau. Dazu wird eine Bildserie unter Variation der Belichtungsbedingungen gewonnen, sodass jeder aufgenommene Bildpunkt der Szene in wenigstens einem Bild der Serie ein hohes SNR besitzt.

In Abschnitt 3 wird eine Bildserie unter zweidimensionaler Variation der Beleuchtungsrichtung aufgenommen und iterativ fusioniert. Die Zielsetzung ist, ein Resultatbild mit möglichst optimalem lokalen Kontrast zu erhalten.

2 Bilder mit konstantem SNR

In diesem Abschnitt wird am Beispiel der Gewinnung von Bildern mit konstantem Signal-zu-Rausch-Verhältnis (SNR) gezeigt, wie eine Bildserie unter Variation eines Parameters gewonnen und anschließend zu einem Resultatbild höherer Güte fusioniert wird. Herkömmliche Kamerasensoren haben einen relativ schmalen nutzbaren Dynamikbereich, der typischerweise bei 8–12 Bit liegt. Natürliche und technische Szenen können jedoch eine wesentlich höhere Dynamik aufweisen. Das bedingt, dass sehr helle Szenenbereiche zu einer Übersteuerung des Sensors führen und in dunklen Bereichen die Szeneninformation mit nur sehr geringem SNR erfasst wird oder gar im Rauschen verschwindet. Im Folgenden soll daher eine Bildserie derart gewonnen werden, dass alle Szenenbereiche in mindestens einem Bild der Serie mit maximalem SNR enthalten sind. Bei der anschließenden Fusion werden dann die Regionen mit bestem SNR zu einem Resultatbild verschmolzen, das näherungsweise ein konstantes SNR auf hohem Niveau aufweist.

Um das Verfahren möglichst ohne überflüssigen Ballast zu erklären, wird die Darstellung der physikalischen Details auf das zum Verständnis des Fusionsansatzes Notwendige reduziert. Insbesondere sollen der Einfachheit halber o.B.d.A. nur Grauwertbilder betrachtet werden. Bild 2 illustriert die Entstehung des Intensitätssignales $I(\mathbf{x})$ auf dem Kamerachip und mithin des zugehörigen Grauwertes $g(\mathbf{x})$.

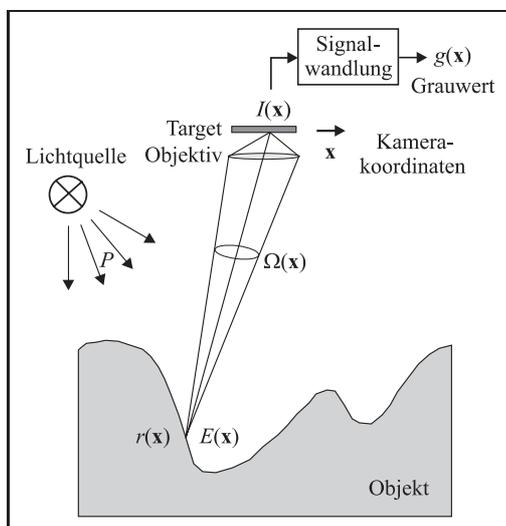


Bild 2: Bildentstehung.

Hierbei bezeichnet P die Leistung der Lichtquelle, $E(\mathbf{x})$ die einfallende Beleuchtungsichte am Ort $\mathbf{x} = (x_1, x_2)^T$ und $r(\mathbf{x})$ die Reflektanz der Oberfläche am Ort \mathbf{x} . Folgende Vereinfachungen und vereinfachende Annahmen werden gemacht:

- (a) Alle ortsabhängigen Größen werden in Abhängigkeit der Kamerakordinaten \mathbf{x} notiert, sodass eine hier irrelevante Unterscheidung zwischen Kamera- und Weltkoordinaten unterbleibt.
- (b) Die Reflektanz $r(\mathbf{x})$ der Objektoberfläche wird als winkelunabhängige Ortsfunktion modelliert. Die physikalisch genauere Beschreibung mit Hilfe der winkelabhängigen *Bidirectional Reflectance Distribution Function* (BRDF) [4] würde hier keine weitergehenden Einsichten liefern.
- (c) Die Variation des Raumwinkels $\Omega(\mathbf{x})$ in Abhängigkeit des Ortes \mathbf{x} und die damit verbundene Intensitätsvariation auf dem Kameratarget bleiben unberücksichtigt.
- (d) Die Abbildung $I \mapsto g$ der Intensitäten auf Grauwerte sei innerhalb $[g_{\min}, g_{\max}]$ näherungsweise linear.

Für die Intensität auf dem Kamerachip gilt in guter Näherung:

$$I(\mathbf{x}) \propto E(\mathbf{x}) r(\mathbf{x}). \tag{2}$$

Die Beleuchtungsichte $E(\mathbf{x})$ am Ort \mathbf{x} hängt von der geometrischen Konstellation von Lichtquelle, Objekt sowie der Reflektanzfunktion $r(\cdot)$ in komplizierter Weise ab, da neben der direkten Einstrahlung auch Licht aus Einfach- und Mehrfachreflexionen die Oberfläche am Ort \mathbf{x} treffen. Sicher lässt sich $E(\mathbf{x})$ aber zerlegen in $E(\mathbf{x}) = P e(\mathbf{x})$, wobei $e(\mathbf{x})$ die Ortsabhängigkeit von $E(\mathbf{x})$ wiedergibt. Die Intensitäten $I(\mathbf{x})$ werden auf elektronischem Wege in Grauwerte konvertiert, für die näherungsweise gilt:

$$g(\mathbf{x}) \propto T P e(\mathbf{x}) r(\mathbf{x}) + n(\mathbf{x}). \tag{3}$$

T bezeichnet die Belichtungszeit der Kamera.

Definiert man $K := T P$ und $s(\mathbf{x}) := e(\mathbf{x}) r(\mathbf{x})$, so erhält man als Signalmodell für die zunächst noch unbeschränkten Grauwerte

$$g(\mathbf{x}) \propto K s(\mathbf{x}) + n(\mathbf{x}). \tag{4}$$

$s(\mathbf{x})$ spielt hierbei die Rolle des Nutzsignals und $n(\mathbf{x})$ sei ein additives Störsignal mit Erwartungswert $E\{n(\mathbf{x})\} = 0$ und Varianz $\text{Var}\{n(\mathbf{x})\} = \sigma^2$.

In der Literatur wird als Nutzsignal oft die Reflektanzfunktion $r(\mathbf{x})$ als „intrinsic Bild“ bezeichnet und mithin als Nutzsignal definiert [2]. Diese Definition ergibt aber nur bei ebenen Objekten einen Sinn, bei denen die Bildstruktur allein von $r(\mathbf{x})$ herrührt. Bei räumlichen Objekten, siehe Bild 2, hängt der visuelle Bildeindruck sowohl von $r(\mathbf{x})$ als auch von $e(\mathbf{x})$ ab. Auch $e(\mathbf{x})$ trägt Information über das Objekt, was man sich leicht anhand der Sichtbarkeit von

räumlichen Strukturen von Objekten mit konstanter Oberflächenreflektanz $r(\mathbf{x}) = \text{konstant}_{\mathbf{x}}$ klar machen kann. Das macht die Definition von $e(\mathbf{x})r(\mathbf{x})$ als Nutzsinal $s(\mathbf{x})$ sinnvoll.

Während die Reflektanz $r(\mathbf{x})$ typischerweise eine Variationsbreite von höchstens 1 : 100 aufweist [2], kann die Variationsbreite unseres Nutzsinales $s(\mathbf{x})$ um mehrere Zehnerpotenzen höher liegen. Man denke hier nur an Szenen mit ausgeprägten Licht- und Schattenregionen. Mit einer solchen Dynamik sind gängige Kamerasensoren überfordert. Die resultierenden Grauwerte sind auf ein Intervall $g \in [g_{\min}, g_{\max}]$ beschränkt. Bild 3 illustriert die damit verbundenen Konsequenzen.

Signalwerte $K s(\mathbf{x})$ unterhalb g_{\min} gehen wegen „Unterbelichtung“ verloren. Liegen die Werte oberhalb von g_{\max} , verliert man das Nutzsinal wegen Übersteuerung. Während in der Nähe von g_{\min} ein niedriges SNR herrscht, liegt nahe bei g_{\max} das maximale SNR vor.

Für die Gewinnung einer Bildserie zur Erzeugung eines Resultatbildes mit konstantem SNR auf hohem Niveau kann man nun K so variieren, dass der in Bild 3 dargestellte Streifen A über die ganze Signalvariationsbreite verschoben wird und so alle Signalbereiche in einem der Bilder der Serie mit hohem SNR enthalten sind. Das Signal-zu-Rausch-Verhältnis beträgt:

$$\text{SNR}(\mathbf{x}) = \left(\frac{K s(\mathbf{x})}{\sigma} \right)^2 \tag{5}$$

Nach der Fusion der Bildserie soll ein Resultat mit

$$\text{SNR}(\mathbf{x}) \stackrel{!}{=} \text{konstant}_{\mathbf{x}} \approx \left(\frac{g_{\max}}{\sigma} \right)^2 \tag{6}$$

entstehen.

Mit dem Parameter α lässt sich die Breite des Streifens A festlegen; siehe Bild 3. Das Ergebnisbild mit maximalem SNR würde man im Extremfall $\alpha = 0$ erhalten. Dann würde man für jeden Punkt \mathbf{x} so belichten, d.h. $K(\mathbf{x})$ so wählen, dass $g(\mathbf{x}) = g_{\max}$ gilt. Das optimale Resultatbild wäre dann schlicht $\propto K^{-1}(\mathbf{x})$. Die Bedingung $g(\mathbf{x}) = g_{\max}$ stellt dabei sicher, dass $K(\mathbf{x})$ bei in diesem Kontext maximalem SNR ermittelt wird. Allerdings müsste man bei dieser

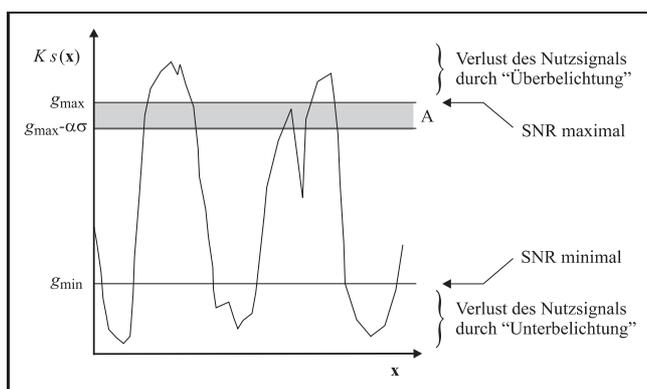


Bild 3: Beschränkung der Grauwerte.

Vorgehensweise für jeden Wert von $s(\cdot) = e(\cdot)r(\cdot)$ einen individuellen Wert K realisieren, was sehr aufwändig wäre.

Praktisch geeigneter ist folgender Ansatz: Beginnend mit einem Bild $g_0(\mathbf{x})$, das mit K_0 aufgenommen wird, werden zwei Bildserien durch „Verschiebung“ des Streifens A gewonnen.

Eine Bildserie $g_i(\mathbf{x})$ „nach unten“ wird gewonnen, indem K solange erhöht wird:

$$K_{i+1} := K_i v, \quad v := \frac{g_{\max}}{g_{\max} - \alpha \sigma} > 1, \quad \alpha > 0 \tag{7}$$

bis $g_i(\mathbf{x}) > g_{\max} - \alpha \sigma$ für alle \mathbf{x} gilt.

Eine Bildserie $g_j(\mathbf{x})$ „nach oben“ wird gewonnen, indem K solange verringert wird:

$$K_{j+1} := K_j v^{-1} \tag{8}$$

bis $g_j(\mathbf{x}) < g_{\max}$ für alle \mathbf{x} gilt.

Mit Hilfe der geometrisch gestuften Variation von K werden alle Szenenbereiche in einem Bild der Serie mit hohem SNR abgebildet. Damit einhergehend werden über- und untersteuerte Szenenbereiche exploriert und für die Sensordynamik überhaupt erst zugänglich gemacht.

Die Zahl N der insgesamt notwendigen Bilder beider Teilserien zusammen ergibt sich aus

$$v^N = D, \tag{9}$$

wobei

$$D := \frac{\max\{s(\mathbf{x})\}}{\min\{s(\mathbf{x})\}} \tag{10}$$

die Dynamik des Nutzsinal $s(\mathbf{x})$ bezeichnet, zu

$$N = \frac{\log D}{\log v}. \tag{11}$$

Ein Resultatbild mit konstantem SNR auf hohem Niveau folgt aus der Fusionsvorschrift:

$$g_{\text{res}}(\mathbf{x}) = \begin{cases} g_i(\mathbf{x}) v^{-i} & \text{für } g_i(\mathbf{x}) \in [g_{\max} - \alpha \sigma, g_{\max}] \\ g_j(\mathbf{x}) v^j & \text{für } g_j(\mathbf{x}) \in [g_{\max} - \alpha \sigma, g_{\max}] \end{cases} \tag{12}$$

Im Bild 4 ist das Ergebnis der Fusion bei einer Szene dargestellt, die eine hell leuchtende LED vor einem davon nicht direkt angeleuchteten Hintergrund zeigt.

Die Bilder 4(d)–(f) zeigen die Bilder 4(a)–(c) nach Durchführung einer Homogenisierung des lokalen mittleren Grauwertes und des lokalen Kontrastes. Der Sinn dieser eigentlich für Texturen entwickelten Homogenisierung [1] ist im vorliegenden Falle nicht, visuell ansprechende Bilder zu erzeugen, sondern das Nutzsinal und das SNR auch in besonders hellen und besonders dunklen Bildbereichen sichtbar zu machen. Das Fusionsresultat Bild 4(c) zeigt klar den Vorteil aufgrund des konstant hohen SNR; siehe Bild 4(f). Man beachte hier insbesondere

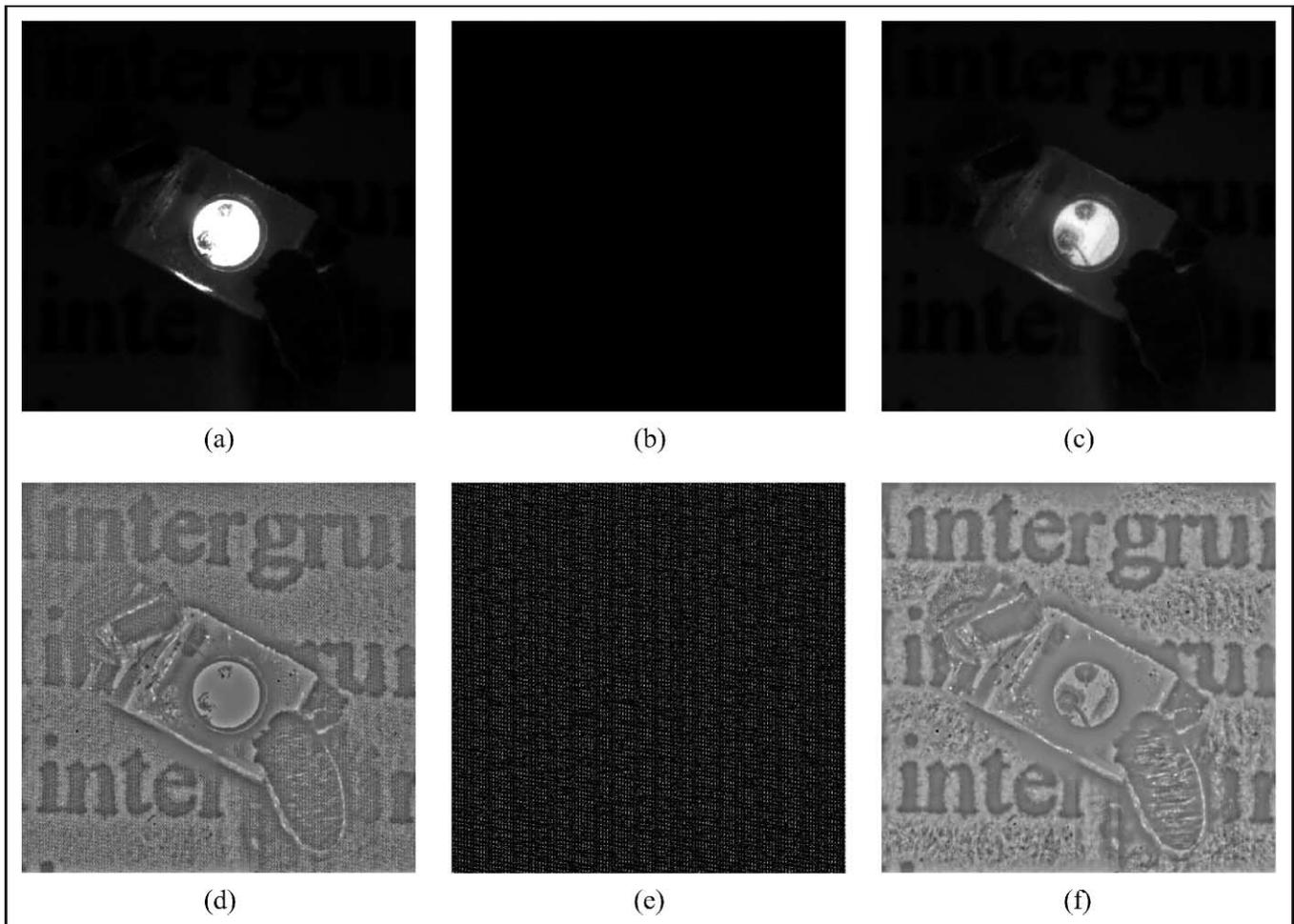


Bild 4: Bilder einer hell leuchtenden LED: (a) konventionell gewonnenes Bild; (b) Aufnahme der Störungen alleine, d. h. für $K=0$ (die Standardabweichung der Störung beträgt $\sigma \approx 0,6$); (c) logarithmiertes Resultatbild der Fusion ($\alpha \approx 200$, $\nu=2$, $N=11$); (d) Ergebnis der Homogenisierung von Bild (a); (e) Ergebnis der Homogenisierung von Bild (b) (durch Spreizung des resultierenden Histogrammes auf den druckbaren Grauwertebereich erkennt man, dass die Störung neben stochastischen Komponenten auch einen periodischen Anteil enthält, der auf ein Netzbrummen zurückzuführen ist); (f) Ergebnis der Homogenisierung von Bild (c).

den Bonding-Draht im Bildzentrum und das Fehlen des Netzbrummens im Hintergrund. Die textuelle Struktur des Hintergrundes in Bild 4(f) ist keine Störung, sondern tatsächlich vorhandene Struktur, die vom Homogenisierungsverfahren hervorgehoben wird. Diese Feinstruktur geht in Bild 4(d) vollständig in der Störung unter.

Nebenbei sei bemerkt, dass die eingeschlagene Vorgehensweise die Linearität der Abbildung $I \mapsto g$ erhöht, da für das Fusionsresultat die Sensorkennlinie nur im Intervall $[g_{\max} - \alpha\sigma, g_{\max}]$ ausgenutzt wird. Die Forderung nach Linearität gemäß Annahme (d) muss demnach nur für dieses Intervall erhoben werden.

3 Beleuchtungsoptimierung

Anders als im vorhergehenden Abschnitt wird hier nicht nur ein Aufnahmeparameter nach einer eindeutig festgelegten Vorschrift variiert; vgl. Gln. (7) und (8). Vielmehr soll eine gleichzeitige Veränderung von zwei Komponenten des Parametervektors ω zugelassen werden – dem Elevations-

winkel θ und dem Azimut ϕ der Beleuchtungsrichtung – um eine möglichst umfassende Exploration beliebig gearteter Prüfobjekte oder Szenen zu erlauben. Vor der Behandlung der Bildoptimierungsmethode sei jedoch zunächst die Aufgabenstellung anhand der Problematik der Aufnahme eines „nichtkooperativen“ Beispielobjektes motiviert.

3.1 Motivation

Bild 5(a) ist eine mit diffusem Licht gewonnene Aufnahme, die einen Ausschnitt eines kleinen Käfers zeigt. Trotz der homogenen Ausleuchtung lässt diese Aufnahme die Struktur der Flügeldecken, die zahlreiche Grübchen aufweisen, nur unzureichend erkennen. Wird hingegen gezieltes Licht verwendet, so gelingt es aufgrund des stark spiegelnden, metallisch anmutenden Oberflächencharakters nicht, den gesamten interessierenden Ausschnitt zufriedenstellend auszuleuchten; siehe Bild 5(b). Lokal wird jedoch auf diese Weise ein besserer Kontrast – und mithin ein höheres SNR – erreicht, der eine detailliertere Erkennung der Oberflächenstruktur zulässt. Offensichtlich rühren bei diesem Objekt die interessierenden Signalanteile überwiegend

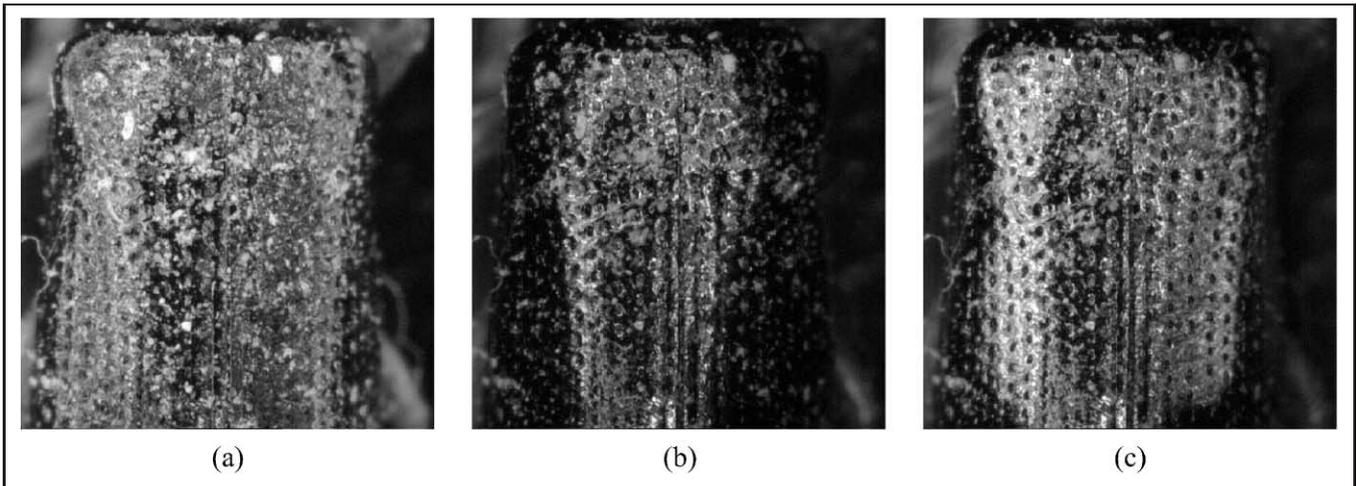


Bild 5: Bilder eines Käfers: (a) Aufnahme mit diffuser Beleuchtung; (b) Aufnahme mit gerichteter Beleuchtung; (c) Ergebnis der Fusion gemäß [6].

vom Oberflächenrelief und nicht von dessen Reflektanzeigenschaften her.

Zwar liefert dieses Beispiel keine überraschenden Erkenntnisse; es trägt jedoch zur Verdeutlichung der folgenden Punkte bei:

- Wird die Beleuchtung als aktive Komponente eines Bildaufnahmesystems betrachtet, so hängt bei gerichteter Beleuchtung der Bildkontrast – und daher auch das SNR – sehr stark von der gewählten Beleuchtungsrichtung ab [5].
- Die Überlagerung mehrerer Lichtquellen (d. h. die Wahl einer „diffuseren“ Beleuchtung) führt aufgrund des Superpositionsprinzips prinzipiell zu einer Kontrastverringern und soll daher im Folgenden vermieden werden.
- Die Wahl einer geeigneten Beleuchtungsrichtung hängt sehr stark vom Oberflächenrelief ab. Generell können daher optimale Beleuchtungsverhältnisse nur lokal erreicht werden. Um eine global gesehen optimale Beleuchtung zu erhalten, müssten verschiedene Bereiche von unterschiedlichen Richtungen aus beleuchtet werden, was in der Praxis extrem schwierig zu realisieren wäre.

Aufgrund dieser Überlegungen wurde in [6] eine Methode vorgestellt, nach welcher eine Bildserie durch systematische Variation der Aufnahmesituation erfasst wird, sodass jedes Objektdetail in mindestens einer der Aufnahmen in ausreichender Qualität enthalten ist. Anschließend erfolgt eine finale Fusion zu einem verbesserten Resultat, welches die Vorteile der Einzelbilder zu vereinen versucht. Dafür wird ein Kriterium $Q\{\cdot\}$ formuliert, das es erlaubt, die Qualität eines Bildes lokal zu spezifizieren. Unter gewissen Voraussetzungen ist es dann möglich, aus allen Bildern der Serie jeweils die besten Bereiche herauszunehmen und zu einem hochwertigen Resultat zu verschmelzen. Dabei geben die lokal optimalen Aufnahmeparameter

$$\omega^*(\mathbf{x}) := \arg \max_{\omega} Q\{g(\mathbf{x}, \omega)\} \quad (13)$$

Auskunft darüber, welche Ausschnitte in das Fusionsresultat eingehen sollten. Um Interferenzen zwischen dem

Nutzsignal und der optimalen Parameterfunktion sowie dadurch verursachte Artefakte zu vermeiden, wird gefordert, dass die verschiedenen Komponenten $\omega_i^*(\mathbf{x})$, $i = 1, \dots, M$, sich örtlich nur „langsam“ ändern dürfen, wohingegen das Nutzsignal $s(\mathbf{x})$ als Bandpasssignal angesetzt wird. Im Idealfall sollten die Fouriertransformierten $\Omega_i^*(\mathbf{f}) = \mathcal{F}\{\omega_i^*(\mathbf{x})\}$ und $S(\mathbf{f}) = \mathcal{F}\{s(\mathbf{x})\}$ disjunkte Träger haben, d. h.:

$$\text{supp}\{\Omega_i^*(\mathbf{f})\} \cap \text{supp}\{S(\mathbf{f})\} = \emptyset \quad \forall i \quad (14)$$

$$\text{mit } \text{supp}\{S(\mathbf{f})\} := \{\mathbf{f} \mid |S(\mathbf{f})| > 0\}. \quad (15)$$

In der Praxis wird diese Glattheitsforderung durch Tiefpassfilterung der Signale $\omega_i^*(\mathbf{x})$ erzwungen.

Bild 5(c) zeigt das Ergebnis dieser Fusionsmethodik für das vorgestellte Beispielobjekt. Um dieses Resultat zu erhalten, wurde die Beleuchtungsrichtung systematisch zweidimensional variiert, d. h. $\omega_1 = \theta \in [-90^\circ, 90^\circ]$ und $\omega_2 = \phi \in [-90^\circ, 90^\circ]$. Der Umfang der Bildserie beträgt hier $N = N_\theta \times N_\phi = 49 \times 49 = 2401$ Bilder.

Verglichen mit den beiden vorangehenden Bildern ist dieses Resultat klar überlegen: es sind wesentlich mehr Details erkennbar als bei Verwendung diffuser oder unidirektionaler Beleuchtung. Dennoch weist die in [6] vorgestellte Fusionsstrategie einige gravierende Nachteile auf, die bei manchen Problemstellungen deren Einsatz in Frage stellen:

- Aus Effizienzgründen wird angestrebt, mit möglichst wenigen Bildern jeden interessierenden Objektbereich in guter Qualität zu erfassen. Dies setzt Wissen über die Objektgeometrie und die Oberflächentextur voraus. Liegt solches Vorwissen nicht vor, so ist im Allgemeinen eine vollständige Abtastung des Parameterraumes erforderlich, was bei hoch dimensionalen Parameterräumen einen unvermeidbaren Aufwand mit sich bringt [3; 5; 8].
- Die berechneten lokal optimalen Aufnahmeparameter als Funktionen des Ortes müssen nicht notwendigerweise „gutartig“ im Sinne von „glatte“ sein. So kann ein optimales Parametersignal $\omega_i^*(\mathbf{x})$ Sprungstellen aufweisen,

die – zusammen mit den erzwungenen Glättungseigenschaften – im Extremfall zu einem Ergebnis führen können, das an keiner Stelle des Bildes zufrieden stellende Aufnahmebedingungen aufweist.

- Schließlich liefert eine finale Fusionsstrategie erst nach der Verarbeitung der gesamten Bildserie ein Resultat – welches trotz der Prüfung auf ausreichenden Informationsgehalt in Bild 1(b) keineswegs den Qualitätserwartungen entsprechen muss. Iterative Verfahren erlauben hingegen eine laufende Beurteilung des Fusionsergebnisses (sowohl visuell als auch maschinell) und können somit besser auf die Bedürfnisse der Aufgabenstellung abgestimmt werden.

3.2 Fusionsansatz

Im letzten Unterabschnitt wurde deutlich gemacht, dass die Selektion der lokal optimalen Aufnahmeparameter und die gleichzeitige Forderung nach örtlich langsam veränderlichen Parameterfunktionen $\omega_i^*(\mathbf{x})$ oftmals kontrovers sind [7]. Im Fusionsresultat Bild 5(c) wird dies vor allem im Randbereich des Käfers deutlich: hier weist die Parameterfunktion $\omega_1^*(\mathbf{x})$ deutliche Sprünge auf, wie man in

Bild 6(a) erkennen kann, die durch die anschließende Glättung zu einem kontrastarmen Ergebnis führen.

Diese Nachteile lassen sich jedoch vermeiden, wenn man die Glättung an solchen Unstetigkeitsstellen „ausschaltet“. Diese Vorgehensweise erfordert eine Kombination der Bildfusion mit einer Segmentierung $\mathcal{R} = \{\mathcal{R}_1, \dots, \mathcal{R}_R\}$ des Bildes in R disjunkte, zusammenhängende Regionen \mathcal{R}_i , die keine Kanten enthalten. Bei diesen Kanten handelt es sich nicht notwendigerweise um Änderungen der Reflektanzfunktion oder um Sprünge der Oberflächentopographie. Vielmehr sind hier so genannte „Parameterkanten“ gemeint, d. h. sprunghafte Änderungen der optimalen Parameterfunktionen $\omega_i^*(\mathbf{x})$. Gewissermaßen entspricht also die optimale Parameterfunktion $\omega^*(\mathbf{x})$ dem Merkmalsvektor, auf dessen Grundlage die Segmentierung erfolgen soll.

Um die Parameterkanten zu detektieren, müsste für sämtliche variierten Parameter ω_i jeweils das optimale Signal $\omega_i^*(\mathbf{x})$ ermittelt werden. Daraufhin könnten Segmentierungsverfahren eingesetzt werden, um die Bereichsgrenzen zu bestimmen. Dies sei am Beispiel des Parametersignals $\omega_1^*(\mathbf{x})$ veranschaulicht; siehe Bild 6(a). Das Ergebnis der

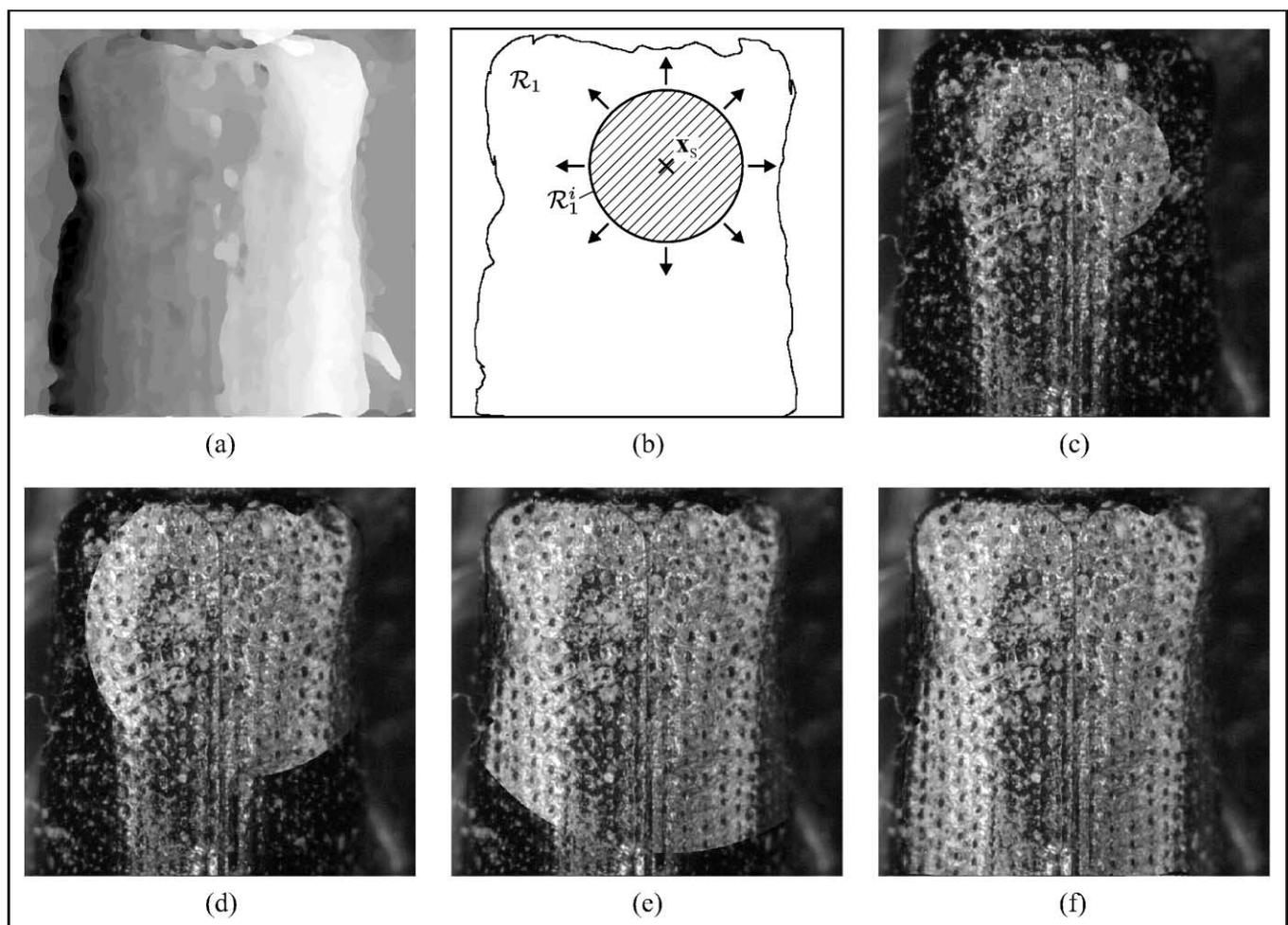


Bild 6: Beispiel zur iterativen Fusion: (a) optimale Parameterkarte $\omega_1^*(\mathbf{x})$; (b) Prinzip der parametergestützten Segmentierung durch Bereichswachstum; (c) Zwischenergebnis der Fusion ($r=100$); (d) Zwischenergebnis der Fusion ($r=200$); (e) Zwischenergebnis der Fusion ($r=300$); (f) Endergebnis der Fusion.

Segmentierung würde in diesem Fall die in Bild 6(b) skizzierten Regionen \mathcal{R}_1 und $\mathcal{R} \setminus \mathcal{R}_1$ ergeben.

Da aber die extrem aufwändige Berechnung des Signals $\omega^*(\mathbf{x})$ zur Segmentierung vermieden werden soll (im Beispiel aus Bild 6(a) war das Signal $\omega_1^*(\mathbf{x})$ ein Zwischenergebnis auf dem Weg zur finalen Fusion), bietet sich eine iterative Methode an, um die Bereichsgrenzen aufzufinden.

3.3 Algorithmische Realisierung

Im Folgenden soll der Einfachheit halber angenommen werden, dass der Parametervektor ω nur die veränderlichen Parameter – hier also den Elevationswinkel θ und das Azimut ϕ der Beleuchtungsrichtung – beinhaltet:

$$\omega := (\theta, \phi)^T. \quad (16)$$

Zur iterativen Exploration der Szene wird zunächst als Startpunkt ein Ort \mathbf{x}_S gewählt und für diesen die Beleuchtung optimiert:

$$\omega^*(\mathbf{x}_S) := \arg \max_{\omega} Q\{g(\mathbf{x}_S, \omega)\}. \quad (17)$$

Zwar weist die zu optimierende Gütefunktion typischerweise mehrere lokale Extrema auf, was die Auffindung des globalen Maximums aufwändig gestaltet. Ist jedoch ein geeigneter Startwert ω_S gegeben, so können klassische Optimierungsverfahren eingesetzt werden, um das interessierende Maximum $\omega^*(\mathbf{x}_S)$ zu finden.

Ausgehend vom Initialpunkt \mathbf{x}_S , welcher der ersten Region \mathcal{R}_k , $k = 1$, zugeordnet wird, erfolgt ein Wachstum dieser Region durch iteratives Hinzufügen von Nachbarpunkten, die mit einem bestimmten Einheitlichkeitskriterium $\mathcal{E}(\mathcal{R}_k)$ für diese Region konform sind.

Für diesen Zweck wird ein Kriterium $\mathcal{E}(\mathcal{R}_k)$ gewählt, das genau dann erfüllt ist, wenn die optimalen Parametervektoren $\omega^*(\mathbf{x}_i)$, $\omega^*(\mathbf{x}_j)$ aller benachbarten Punktepaare

$$\mathbf{x}_i, \mathbf{x}_j \in \mathcal{R}_k, \quad \mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i), \quad (18)$$

der zugehörigen Region \mathcal{R}_k ungefähr gleich sind:

$$\omega^*(\mathbf{x}_i) \approx \omega^*(\mathbf{x}_j), \quad (19)$$

wobei $\mathcal{N}(\mathbf{x})$ die 4er-Nachbarschaft des Punktes \mathbf{x} bezeichnet:

$$\mathcal{N}(\mathbf{x}) := \{\xi \mid \xi = \mathbf{x} + \epsilon, \|\epsilon\| \leq 1\}. \quad (20)$$

Die Prüfung auf Erfüllung der Wachstumsbedingung Gl. (19) erfordert, dass die optimale Parameterfunktion $\omega^*(\mathbf{x})$ zumindest für den Ort des überprüften Nachbarn $\mathbf{x} = \mathbf{x}_j$ bekannt ist. Da aber die aufwändige Berechnung dieser Funktion vermieden werden soll, wird statt dessen angenommen, dass diese Funktion an der untersuchten Stelle glatt ist. Aus dieser Glattheit folgt nämlich, dass das Optimum $\omega^*(\mathbf{x}_j)$ in der Umgebung des Wertes $\omega^*(\mathbf{x}_i)$ zu liegen kommen muss:

$$\omega^*(\mathbf{x}_j) \in \mathcal{U}(\omega^*(\mathbf{x}_i)) \quad (21)$$

mit

$$\mathcal{U}(\omega) := \{\eta \mid \eta = \omega + \epsilon, |\epsilon_1| \leq \Delta\theta, |\epsilon_2| \leq \Delta\phi\}. \quad (22)$$

Befindet sich hingegen kein (lokales) Maximum in $\mathcal{U}(\omega^*(\mathbf{x}_i))$, so kann umgekehrt gefolgert werden, dass die Glattheitsannahme Gl. (21) verletzt ist und somit $\mathbf{x}_j \notin \mathcal{R}_k$ gilt – das Bereichswachstumsverfahren wäre also an die Bereichsgrenze, d. h. eine Parameterkante, gestoßen.

Nach der Segmentierung der ersten Region \mathcal{R}_1 kann entweder ein weiterer Initialpunkt im Bereich $\mathcal{R} \setminus \mathcal{R}_1$ gewählt werden, um die Beleuchtungsoptimierung im restlichen Bildbereich fortzusetzen – oder aber die Exploration abgebrochen werden, falls das Fusionsziel bereits erreicht worden ist.

Die Annahme, dass keine Kanten innerhalb der Regionen vorhanden sind, zusammen mit der schrittweisen Exploration jeder Region sorgen gemeinsam für einen insgesamt gutartigen Verlauf der iterativ ermittelten optimalen Parameterfunktion $\omega^*(\mathbf{x})$. Anders als in der in [6] vorgestellten finalen Fusionsstrategie darf deswegen hier auf eine Glättung der Parameterfunktion verzichtet werden. Somit ergibt sich für die Fusion eine extrem einfache Vorschrift:

$$g_{\text{res}}(\mathbf{x}) = g(\mathbf{x}, \omega^*(\mathbf{x})). \quad (23)$$

Hierbei wurde jedoch vorausgesetzt, dass die optimale Parameterfunktion $\omega^*(\mathbf{x})$ vollständig ermittelt wurde und somit ebenfalls eine vollständige Segmentierung vorliegt. Interessiert man sich jedoch für ein Zwischenergebnis nach der i -ten Iteration, bei welcher die k -te Region auf \mathcal{R}_k^i gewachsen ist, so bietet sich die folgende rekursive Fusionsgleichung an:

$$g_{\text{res}}^i(\mathbf{x}) = \begin{cases} g(\mathbf{x}, \omega^*(\mathbf{x})) & \text{für } \mathbf{x} \in \mathcal{R}_k^i \setminus \mathcal{R}_k^{i-1} \\ g_{\text{res}}^{i-1}(\mathbf{x}) & \text{sonst} \end{cases} \quad (24)$$

mit

$$g_{\text{res}}^0(\mathbf{x}) = g(\mathbf{x}, \omega^*(\mathbf{x}_S)). \quad (25)$$

Damit liefert diese Fusionsmethode nach jeder Iteration ein verbessertes Bild der fortlaufend wachsenden Umgebung des Startpunktes \mathbf{x}_S . Dabei sorgt ein isotropes Wachstum der Region auf einem kreisscheibenförmigen Träger mit Radius r für eine gleichmäßige Exploration der Umgebung des Punktes \mathbf{x}_S ; siehe Bild 6(b). Die rekursive Fusionsvorschrift Gl. (24) erlaubt schließlich eine aufwandsarme Generierung eines laufend verbesserten Resultatbildes.

3.4 Ergebnisse

In diesem Abschnitt werden am Beispiel des vorgestellten biologischen Präparates die Ergebnisse der iterativen Fusionsmethodik präsentiert. Dabei wurde der Startpunkt \mathbf{x}_S für das Bereichswachstum willkürlich gemäß Bild 6(b) gewählt. Das korrespondierende Startbild $g_{\text{res}}^0(\mathbf{x})$ entspricht der in Bild 5(b) dargestellten Aufnahme.

Die Bilder 6(c)–(e) zeigen drei Zwischenergebnisse der iterativen Fusion, wobei der Radius r des Bereichswachstums

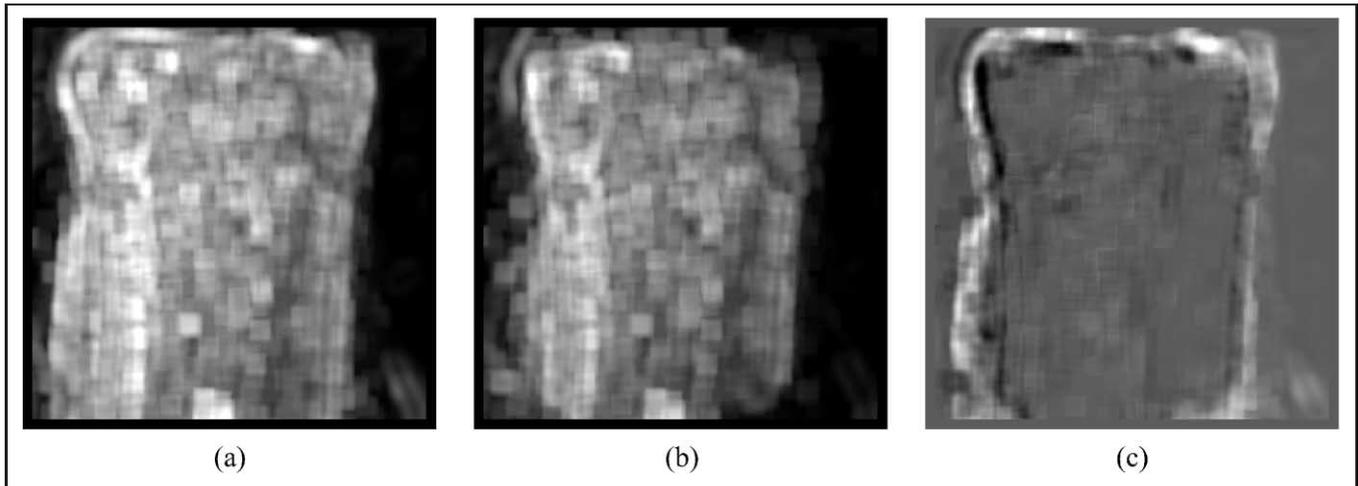


Bild 7: Beurteilung der Fusionsverfahren: (a) Lokaler Kontrast beim Ergebnis der iterativen Fusion; (b) Lokaler Kontrast beim Ergebnis der finalen Fusion; (c) Differenz aus (a) und (b).

jeweils 100, 200 und 300 Pixel beträgt. Als Kriterium $Q\{\cdot\}$ wurde in sämtlichen Fällen die lokale Standardabweichung der Intensitäten gewählt, wenngleich andere Kriterien – wie z. B. die lokale Entropie, der lokale mittlere Gradientenbetrag und sogar der lokale mittlere Grauwert – ebenfalls zu zufrieden stellenden Ergebnissen führen. Je größer die Entfernung zum Initialpunkt \mathbf{x}_S wird, umso deutlicher erscheint der visuelle Qualitätsgewinn. Schließlich ist in Bild 6(f) das Endergebnis der Fusion dargestellt, in welchem die detektierten Grenzen der Region \mathcal{R}_1 gemäß Bild 6(b) verlaufen. In diesem Fall wurde nur diese eine Region herausgearbeitet und auf eine vollständige Segmentierung verzichtet. Der Hintergrund $\mathcal{R} \setminus \mathcal{R}_1$ entspricht somit dem Startbild $g_{\text{res}}^0(\mathbf{x})$.

Verglichen mit dem in Bild 5(c) abgebildeten Resultat der finalen Fusion zeichnet sich dieses Ergebnis durch einen sehr hohen Kontrast aus – insbesondere auch im Randbereich. Um den Qualitätsgewinn quantitativ zu untersuchen, wurde die lokale Standardabweichung der Intensitäten beider Bilder berechnet; siehe 7(a) und (b). Diese ist im iterativen Fusionsergebnis im Mittel um 6% höher als beim finalen Fusionsresultat. Bildet man die Differenz

$$Q\{g_{\text{res,iterativ}}(\mathbf{x})\} - Q\{g_{\text{res,final}}(\mathbf{x})\} \quad (26)$$

beider Bilder, so erkennt man, dass dieser Unterschied im Wesentlichen durch die Kontrastabnahme im Randbereich des finalen Fusionsresultates verursacht wird; siehe Bild 7(c).

Ein weiterer Vorteil der iterativen Vorgehensweise ergibt sich aus der Tatsache, dass dank dem vorgestellten Algorithmus eine vollständige Abtastung des Parameterraumes nicht erforderlich ist. Beim vorgestellten Beispiel mussten – bei gleicher Quantisierung der Aufnahmeparameter – lediglich 614 Bilder aufgenommen und verarbeitet werden. Bei der finalen Fusion waren es rund viermal so viele gewesen.

Freilich darf aus diesem Ergebnis nicht der falsche Rückschluss gezogen werden, dass iterative Fusionsverfahren

grundsätzlich gegenüber einer finalen Fusion zu bevorzugen wären. In diesem Beispiel jedoch konnten durch die vorgeschlagene iterative Vorgehensweise das Problem der Abtastung des Parameterraumes gelöst sowie die unerwünschte Glättung an den Objekt- bzw. Parametergrenzen vermieden werden. Zusätzlich vermag der vorgestellte Fusionsalgorithmus, die Informationssenke laufend mit einem verbesserten Bild der untersuchten Szene auszustatten, was einen zusätzlichen Eingriff in die Datenakquisition – etwa zum Abbruch oder zur Rekonfiguration des Verfahrens – erlaubt.

4 Zusammenfassung

Es wurden Verfahren erörtert, die eine Bildoptimierung durch kontrolliertes Variieren der Aufnahmeparameter eines aktiven Sichtprüfungssystems und den Einsatz von Datenfusionsmethoden erlauben. Bei der Datenakquisition wird eine sequentielle Strategie eingeschlagen. Auf der Grundlage bereits aufgenommener Bilder werden die Einstellungen für die nächsten Aufnahmen gezielt berechnet, um den Zugewinn an Information möglichst zu maximieren. Auf diese Weise wird eine iterative Exploration einer Szene erzielt.

Schließlich werden die gewonnenen Aufnahmen mit Datenfusionsmethoden zu einem hochwertigen Resultatbild kombiniert. Methodisch werden dabei zwei Fusionsvarianten unterschieden. Eine finale Fusion ist opportun, wenn aus den aufgenommenen Bilddaten schon ein Rückschluss auf die Qualität des Fusionsresultates möglich ist. Iterative Fusionsansätze erlauben hingegen eine a posteriori Beurteilung des Fusionsergebnisses und sind mit ihrem laufenden Blick auf das sequentiell entstehende Resultat besonders sparsam bei der Datenakquisition bei gleichzeitigen Qualitätsvorteilen.

Die Vorzüge der vorgeschlagenen Vorgehensweise werden anhand zweier Beispiele – der Erfassung von Bildern mit

konstantem SNR und der zweidimensionalen, örtlich veränderlichen Optimierung der Beleuchtungsrichtung – demonstriert. In beiden Fällen weist das Fusionsresultat einen deutlich höheren nützlichen Informationsgehalt als Einzelaufnahmen auf, was den höheren Aufwand bei der Datenaquisition rechtfertigt.

Literatur

- [1] Beyerer, J.; Puente León, F.: Suppression of inhomogeneities in images of textured surfaces, *Optical Engineering* 36(1) 1997, S. 85–93.
- [2] Fattal, R.; Lischinski, D.; Werman, M.: Gradient domain high dynamic range compression. *ACM Transactions on Graphics* 21(3) 2002, S. 249–256.
- [3] Malz, R.: *Codierte Lichtstrukturen für 3-D-Meßtechnik und Inspektion*, Dissertation, Universität Stuttgart, 1992.
- [4] Nicodemus, F.E.; Richmond, J.C.; Hsia, J.J.; Ginsberg, I.W.; Limperis, T.: *Geometrical Considerations and Nomenclature for Reflectance*, NBS Monograph 160, National Bureau of Standards, U.S. Department of Commerce, Washington, DC, 1977.
- [5] Puente León, F.: Enhanced imaging by fusion of illumination series. In: *Sensors, Sensor Systems, and Sensor Data Processing*, O. Löffel (Hrsg.). *Proc. SPIE* 3100 1997, S. 297–308.
- [6] Puente León, F.; Beyerer, J.: Datenfusion zur Gewinnung hochwertiger Bilder in der automatischen Sichtprüfung. *Automatisierungstechnik* 45(10) 1997, S. 480–489.
- [7] Puente León, F.; Beyerer, J.: Active vision and sensor fusion for inspection of metallic surfaces, In: *Intelligent Robots and*

Computer Vision XVI: Algorithms, Techniques, Active Vision, and Materials Handling (Hrsg.: D.P. Casasent), Proceedings of SPIE 3208, 1997, S. 394–405.

- [8] Yi, S.; Haralick, R.M.; Shapiro, L.G.: Optimal sensor and light source positioning for machine vision, *Computer Vision and Image Understanding* 61(1) 1995, S. 122–137.

Manuskripteingang: 13. Januar 2005.



Prof. Dr.-Ing. habil. Jürgen Beyerer leitet das Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB) sowie den Lehrstuhl für Interaktive Echtzeitsysteme an der Universität Karlsruhe (TH). Hauptarbeitsgebiete: Bildverarbeitung und automatische Sichtprüfung, Automatisierungstechnik, Mustererkennung, Informationstheorie, Robotik und Messtechnik.

Adresse: Fraunhofer-Institut für Informations- und Datenverarbeitung IITB, Fraunhoferstr. 1, 76131 Karlsruhe, Tel.: +49-721-6091 210, Fax: +49-721-6091 556, E-Mail: juergen.beyerer@iitb.fraunhofer.de



Prof. Dr.-Ing. Fernando Puente León leitet das Fachgebiet Verteilte Messsysteme an der Technischen Universität München. Hauptarbeitsgebiete: Messtechnik und Sensorik, Signalverarbeitung, automatische Sichtprüfung und Bildverarbeitung, Mustererkennung, Datenübertragung und Informationsfusion.

Adresse: Lehrstuhl für Messsystem- und Sensortechnik, Technische Universität München, Theresienstr. 90/N5, 80333 München, Tel.: +49-89-289 23350, Fax: +49-89-289 23347, E-Mail: f.puente@tum.de