

# **Objektive Bewertung von Deblocking-Verfahren**

von der Fakultät

Elektrotechnik und Informationstechnik

der Technischen Universität Dortmund

genehmigte

**Dissertation**

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften

von

**Patrick René Steve Piastowski**

Dortmund, 2008

Tag der mündlichen Prüfung: 16.09.2008

Hauptreferent: Prof. Dr.-Ing. Hartmut Schröder

Korreferent: Prof. Dr.-Ing. André Kaup



# Vorwort

Die vorliegende Arbeit entstand zu großen Teilen während meiner Tätigkeit als wissenschaftlicher Angestellter am Arbeitsgebiet Schaltungen der Informationsverarbeitung der Technischen Universität Dortmund. Ich möchte Herrn Prof. Dr.-Ing. Hartmut Schröder, dem Leiter des Arbeitsgebietes, einen besonderen Dank für die Ermöglichung, Betreuung und Begutachtung dieser Arbeit aussprechen. Durch konstruktive Kritik und fruchtbare Diskussionen hat er erheblich zum Gelingen der Arbeit beigetragen.

Ebenso möchte ich mich bei meinen ehemaligen Kollegen für die ständige Diskussionsbereitschaft und die Anregungen bedanken, die im Dialog mit ihnen entstehen konnten. Mein Dank gilt auch den Studenten, die in irgendeiner Form zu dieser Arbeit beigetragen haben.

Herrn Prof. Dr.-Ing. André Kaup, dem Leiter des Lehrstuhls für Multimediakommunikation und Signalverarbeitung an der Universität Erlangen-Nürnberg, danke ich für sein Interesse an der Arbeit und für die Übernahme des Korreferats.

Ein ganz besonderer Dank gilt meiner Familie. Speziell meine Frau Helena und meine Eltern Artur und Gisela, aber auch andere Familienmitglieder hatten sicherlich des Öfteren unter meiner zeitweise wechselhaften Laune zu leiden. Ich bedanke mich für das Verständnis, die Geduld und die Unterstützung, die mir stets entgegengebracht wurden. Herrn Dipl.-Ing. Sebastian Lenke wie auch meiner Frau gilt darüber hinaus mein Dank für die Durchsicht des Manuskripts. Nicht zuletzt möchte ich Herrn Dipl.-Stat. Martin Gebel danken, der mir mehrmals ein kompetenter Diskussionspartner war.

Patrick René Steve Piastowski



# Inhaltsverzeichnis

<b>Vorwort</b>	<b>I</b>
<b>Notation</b>	<b>VII</b>
<b>Kurzfassung/Abstract</b>	<b>XI</b>
<b>1 Einleitung und Übersicht</b>	<b>1</b>
<b>2 Ausgewählte Grundlagen der Bild- und Videokompression</b>	<b>5</b>
2.1 Transformationscodierverfahren . . . . .	7
2.2 Hybride Codierverfahren . . . . .	10
2.3 Artefakte der digitalen Codierung und Übertragung . . . . .	18
<b>3 Grundlagen der Regressionsanalyse</b>	<b>33</b>
3.1 Einfache lineare Regression . . . . .	33
3.2 Multiple lineare Regression . . . . .	35
3.3 Voraussetzungen der linearen Regression . . . . .	39
3.4 Auswahl von Regressoren und Multikollinearität . . . . .	42
3.5 Schrittweise Regression . . . . .	43
3.6 Abschließende Bemerkungen zur Regressionsanalyse . . . . .	45
<b>4 Objektive Verfahren zur Bildqualitätsbewertung</b>	<b>47</b>
4.1 MSE und PSNR . . . . .	49
4.2 Gradientenbasierte Blocking-Messung . . . . .	51
4.3 Ein- und Mehrkanalmodelle der visuellen Wahrnehmung . . . . .	55
4.4 Faktorenmodelle . . . . .	58
4.5 Bewertung der Wirksamkeit von Deblocking-Verfahren . . . . .	62
<b>5 Stand der Technik in der Artefaktreduktion</b>	<b>65</b>
5.1 Restoration-Verfahren . . . . .	66
5.2 Enhancement-Verfahren . . . . .	68
5.2.1 Lineare Filterung von Blockgrenzen . . . . .	69
5.2.2 Subtraktion von Fehlersignalen . . . . .	71
5.2.3 Adaptive Filterung von Blockgrenzen mit zwei Filtermodi . . . . .	73

5.2.4	Filterung von Wavelet-Koeffizienten . . . . .	78
<b>6</b>	<b>Adaptive Reduktion von Blockartefakten mit lokaler Bildanalyse</b>	<b>81</b>
6.1	Lokale Bildanalyse . . . . .	81
6.2	Filter zur Störungsreduktion . . . . .	86
6.3	Varianten des Verfahrens . . . . .	89
6.4	Erweiterung des Verfahrens . . . . .	92
<b>7</b>	<b>Subjektive Tests</b>	<b>99</b>
7.1	Randbedingungen . . . . .	100
7.2	Bildmaterial . . . . .	103
7.3	Test 1 . . . . .	106
7.4	Test 2 . . . . .	107
7.5	Test 3 . . . . .	108
7.6	Auswertung der subjektiven Testdaten . . . . .	109
7.6.1	Ausschluss unzuverlässiger Daten . . . . .	109
7.6.2	Datenanalyse . . . . .	111
<b>8</b>	<b>Neuer Ansatz zur Bewertung von Deblocking-Verfahren</b>	<b>127</b>
8.1	Allgemeines Faktorenmodell . . . . .	128
8.2	Randbedingungen des neuen Ansatzes zur Bewertung von Deblocking- Verfahren . . . . .	130
8.3	Teilkomponenten/Qualitätsmaße . . . . .	132
8.3.1	Diskontinuitätsmaß . . . . .	133
8.3.2	Glättungsmaß . . . . .	137
8.3.3	Konturdegradationsmaß . . . . .	140
8.3.4	Degradationsmaß . . . . .	142
8.3.5	Wechselwirkung aus Diskontinuitäts- und Glättungsmaß . . . . .	143
8.3.6	Weitere Maße . . . . .	144
8.4	Kombination von Qualitätsmaßen . . . . .	146
8.4.1	Variante 1 . . . . .	148
8.4.2	Variante 2 . . . . .	151
8.4.3	Variante 3 . . . . .	155
8.4.4	Variante 4 . . . . .	157
8.4.5	Anmerkungen zur Entwicklung und Kombination von Qualitätsmaßen	157
<b>9</b>	<b>Schätzgüte des neuen Ansatzes und Vergleich mit anderen Verfahren</b>	<b>161</b>
9.1	Schätzgüte in den ersten beiden subjektiven Tests . . . . .	161
9.2	Schätzgüte im dritten subjektiven Test . . . . .	168
9.3	Plausibilität und Stabilität . . . . .	171

9.4 Zusammenfassung und Interpretation der Ergebnisse . . . . .	176
<b>10 Zusammenfassung und Ausblick</b>	<b>183</b>
Literaturverzeichnis	189
Abbildungsverzeichnis	201
Tabellenverzeichnis	205
A Playlists	207
B Ergänzende Ergebnisse	211
C Anmerkungen zur Software-Implementierung	215
D Messung der Wirksamkeit von Deringing-Verfahren	219
E Verwendetes Bildmaterial	221
Lebenslauf	227



# Notation

## Abkürzungen

AC	Alternating Current, Wechselstrom/Wechselanteil
BLUE	Best Linear Unbiased Estimator
CLS	Constrained Least Squares
Codec	Einheit bestehend aus (En)Coder und Decoder
DC	Direct Current, Gleichstrom/Gleichanteil
DCT	Diskrete Cosinus Transformation
DPSNR	Differential Peak Signal to Noise Ratio
DMOS	Differential Mean Opinion Score
DMSE	Differential Mean Squared Error
DVB	Digital Video Broadcast
DVD	Digital Versatile Disc
GOP	Group Of Pictures
HDTV	High Definition Television
HVS	Human Visual System
IDCT	Inverse Diskrete Cosinus Transformation
JPEG	Joint Photographic Experts Group, auch Verfahren zur Kompression von Einzelbildern
MAP	Maximum A Posteriori
MOS	Mean Opinion Score

MPEG	Motion Picture Experts Group, auch Verfahren zur Kompression von Bildsequenzen
MSE	Mean Squared Error, quadratischer Fehler
NOP	No Operation, keine Operation
POCS	Projections Onto Convex Sets
PSNR	Peak Signal to Noise Ratio
VO	Video Object
VOP	Video Object Plane

## Operatoren

$\ \cdot\ _2$	$l_2$ -Norm
$\arg \max_{x \in X} f(x)$	entspricht dem Wert $x \in X$ , für den $f(x)$ maximal wird
$\arg \min_{x \in X} f(x)$	entspricht dem Wert $x \in X$ , für den $f(x)$ minimal wird
$cov(\cdot)$	Kovarianz
$diag[\cdot]$	Diagonalmatrix
$E\{\cdot\}$	Erwartungswert
$p(x)$	Wahrscheinlichkeitsdichtefunktion, auch Wahrscheinlichkeit des Ereignisses $x$
$p(x y)$	Wahrscheinlichkeit des Ereignisses $x$ unter der Nebenbedingung $y$
$round(\cdot)$	Rundungsoperator
$var(\cdot)$	Varianz

---

## Formelzeichen

$\alpha$	Vertrauensniveau
$\beta_j$	Regressionskoeffizient
$\Delta$	Quantisierungsstufenhöhe
$\varphi_{k,l}$	DCT-Basisfunktion
$\sigma$	Standardabweichung
$\sigma^2$	Fehlervarianz
$\hat{\sigma}^2$	geschätzte Fehlervarianz
$\sigma_q^2$	Quantisierungsfehlervarianz
$b_j$	geschätzter Regressionskoeffizient
$B$	Bestimmtheitsmaß
$C_{u,v}$	(DCT-)Koeffizient
$C_{Q_{u,v}}$	quantisierter (DCT-)Koeffizient
$d_i$	Residuum
$e_q$	Quantisierungsfehler
$F_{n_1, n_2; 1-\alpha}$	Quantil der F-Verteilung
$H_0$	Nullhypothese
$H_A$	Alternativhypothese
$HG$	Summe der Hauptgradienten
$MG$	mittlerer Gradient
$NG$	Summe der Nebengradienten
$P_q$	Quantisierungsfehlerleistung
$PH$	Bildhöhe in Zeilen
$PW$	Bildbreite in Spalten

$Q_{u,v}$	Quantisierungsparameter aus einer Quantisierungsmatrix
$QP$	Quantisierungsparameter
$r$	Korrelationskoeffizient
$s(x, y)$	Bildsignal (auch $s(m, n)$ )
$s_{ref}(x, y)$	Bildsignal (Referenzbild)
$S(k, l)$	transformiertes Bildsignal
$S_{DCT,i}$	DCT-Koeffizient
$s_i$	bestimmter einzelner Bildpunkt
$s'_i$	bestimmter einzelner gefilterter Bildpunkt
$S_i$	bestimmter einzelner transformierter Bildpunkt
$t_{n;1-\alpha/2}$	Quantil der Studentischen t-Verteilung
$u_i$	Realisation der Störvariablen $U_i$
$U_i$	Störvariable im Rahmen der Regressionsanalyse
$x_i$	bestimmter Wert der Einflussgröße $X$
$x_{ji}$	bestimmter Wert der Einflussgröße $X_j$
$X$	Einflussgröße im Rahmen der Regressionsanalyse
$X_j$	Einflussgröße im Rahmen der Regressionsanalyse
$y$	mittlerer von menschlichen Betrachtern wahrgenommener Qualitätswert
$y_e$	geschätzter Qualitätswert
$y_i$	Realisation von $Y_i$
$\hat{y}_i$	Schätzung des deterministischen Teils der Zielgröße im Rahmen der Regressionsanalyse
$Y$	Zielgröße im Rahmen der Regressionsanalyse
$Y_i$	potenzielle Beobachtung der Zielgröße $Y$ für bestimmte Werte der Einflussgrößen

# Kurzfassung

In der modernen Bildsignalverarbeitung und -übertragung sind nahezu sämtliche beteiligten Komponenten digital. Die sehr großen digitalen Datenmengen können in den meisten Fällen nur durch den Einsatz von Kompressionstechniken gehandhabt werden. Insbesondere bei hohen Kompressions- bzw. niedrigen Datenraten sind damit Informationsverluste und die Erzeugung sichtbarer Codierartefakte, die sowohl für menschliche Betrachter als auch für nachfolgende Verarbeitungsschritte störend sein können, verbunden.

In dieser Arbeit werden verschiedene Ansätze zur Reduktion von Codierartefakten in Bildern analysiert und miteinander verglichen. Ein neues Postprocessing-Verfahren zur Reduktion von Blockartefakten in Endgeräten wie z. B. Fernsehgeräten wird vorgestellt. Es weist eine hohe Wirksamkeit auf und benötigt außer den bereits decodierten Bildern keinerlei weitere Daten.

Das Hauptaugenmerk dieser Arbeit liegt auf der Evaluation der Wirksamkeit verschiedener Artefaktreduktionsverfahren bezogen auf die Verbesserung der Bildqualität. Zwar kann eine solche Evaluation auf Basis subjektiver Tests durchgeführt werden, allerdings sind subjektive Tests sehr zeitaufwändig, kostenintensiv und daher oft nicht durchführbar. Objektive Qualitätsmaße (wie z. B. PSNR oder MSE) könnten eine Alternative darstellen. Sie eignen sich zur raschen Bewertung unterschiedlicher Verfahren der Bildsignalverarbeitung, wie z. B. zum Vergleich von Rauschreduktionsverfahren oder auch zur Evaluation der Qualität von Video-Codecs. Es wird gezeigt, dass häufig verwendete Qualitätsmaße jedoch zum Vergleich der Wirksamkeit von Artefaktreduktionsverfahren ungeeignet sind. Zu diesem Zweck wird ein neues Qualitätsmessverfahren entwickelt. Es basiert auf einem Faktorenmodell und verwendet verschiedene Qualitätsmaße, die an die Eigenschaften von Störungen, welche im Zusammenhang mit blockbasierter hybrider Transformationscodierung auftreten, angepasst sind. Die hohe Schätzgüte des neuen Qualitätsmessverfahrens bei der Wirksamkeitsbewertung von Deblocking-Verfahren wird gezeigt.

# Abstract

Today almost all components for processing and distribution of still images and video sequences are digital. The large amounts of digital data can often only be handled by use of data compression methods. Especially in case of high compression ratios and low bitrates respectively this results in information loss and appearance of visible coding artifacts, which are annoying for human observers and for succeeding image processing algorithms.

In this thesis different approaches for reducing coding artifacts in image data are analysed and compared to each other. A new postprocessing method for reducing blocking artifacts in end devices like television sets is introduced. This algorithm has a good performance and does not need any information besides already decoded images.

The main point of interest in the thesis is the performance evaluation of different artifact reduction algorithms relating to the improvement of image quality. Of course an evaluation can be done based on subjective tests, but such tests are very time-consuming, expensive and often unpracticable. Objective quality metrics (e. g. PSNR, MSE) could be an alternative. They are suitable for rapid assessment of different image processing methods, for example comparison of noise reduction systems or evaluation of video codecs. However it is shown, that commonly used quality metrics are unsuitable for comparing the performance of artifact reduction systems. For this purpose a new quality measure is developed. It is based on a factor model and uses different quality metrics, which are adapted to the properties of artifacts appearing in the context of block-based hybrid transform image coding. A high reliability of the new measure for the purpose of evaluating the effectiveness of blocking reduction techniques is shown.

# 1 Einleitung und Übersicht

Moderne Bildsignalverarbeitung ist geprägt durch den Einsatz digitaler Komponenten sowohl bei der Vorverarbeitung, der Übertragung, der Speicherung und der Nachverarbeitung von Bildern und Bildsequenzen. Die Zahl der analogen Systeme und Komponenten nimmt stetig ab und es ist absehbar, dass ihr Einsatz in Zukunft nur noch auf wenige Spezialfälle beschränkt sein wird. Die Digitalisierung von Bilddaten geht jedoch mit der Erzeugung von sehr großen Datenmengen einher. Deshalb ist es aus Kostengründen, Kapazitäts- und Bandbreitebeschränkungen unumgänglich, mit geeigneten Mitteln für eine Reduzierung der Datenmengen zu sorgen. Zu diesem Zweck wurden Kompressionsverfahren für Bilder (JPEG, JPEG2000, TIFF-LZW) und Bildsequenzen (MPEG-1, MPEG-2, MPEG-4, H.264) entwickelt, die örtliche und zeitliche Korrelationen in Bildsignalen sowie bestimmte Eigenschaften des menschlichen visuellen Systems zur Datenreduktion ausnutzen. Die erzielbaren Gewinne bei der Verringerung der Datenmengen lassen sich in Gewinne durch Redundanzreduktion und in Gewinne durch Irrelevanzreduktion einteilen. Durch Redundanzreduktion gehen keine Informationen verloren, jedoch ist der Grad der erreichbaren Datenreduktion verhältnismäßig klein. Größere Gewinne ermöglicht die Irrelevanzreduktion, die jedoch mit einem unwiederbringlichen Datenverlust (meist durch Quantisierung) einhergeht. Wird der Datenverlust auf geeignete Art und Weise beeinflusst, wie es in typischen Bildkompressionsverfahren der Fall ist, so ist er bis zu einem begrenzten Maß für den menschlichen Betrachter nicht wahrnehmbar. Starke Datenverluste bzw. hohe Kompressionsgrade, wie sie insbesondere im Consumerbereich nicht unüblich sind, äußern sich jedoch durch Störungen - Artefakte genannt - bei der Bildwiedergabe. Bekannte bzw. häufige Artefakte sind Block- bzw. Blocking-, Ringing- und Blurring-Artefakte. Insbesondere Blockartefakte, die sich durch horizontal oder vertikal verlaufende Diskontinuitäten an bestimmten Positionen in Bildern äußern, sind für den Menschen deutlich wahrnehmbar.

Die zunehmende Durchdringung des Marktes für Fernsehgeräte mit neuartigen Displays vergrößert das Problem weiter, da viele moderne Displays, insbesondere LC-Displays, Codierartefakte besonders deutlich sichtbar werden lassen. Dafür sind einerseits die Eigenschaften solcher Displays verantwortlich [DKL<sup>+</sup>05], andererseits aber auch der durch die Größe der Displays bedingte kleiner werdende relative Betrachtungsabstand in typischen Betrachtungssituationen [Win05]. Hinzu kommt, dass auch zunehmend neuartige Anwendungen, z. B. die Übertragung von Videos auf mobile Geräte über niederratige Kanäle, verfügbar werden, die durch sehr geringe Betrachtungsabstände und ein vergleichsweise hohes Maß an Störungen geprägt sind [BBK05].

Eine Verringerung der genannten Probleme kann durch die Verwendung von Verfahren zur Reduktion von Codierartefakten erreicht werden. Neueste Kompressionsverfahren wie z. B. H.264 beinhalten bereits standardisierte Deblocking-Filter. Aber auch für die nach wie vor sehr bedeutsamen älteren Kompressionsverfahren wie z. B. MPEG-2, die in Systemen wie DVB oder DVD eingesetzt werden, existiert eine Vielzahl von Ansätzen zur Verringerung von Artefakten. Aufgrund ihrer hohen Flexibilität sind für den Einsatz in der Praxis besonders so genannte Enhancement-Verfahren relevant [TCS05, SK98], deren Ziel es ist, bereits decodierte gestörte Bilder in geeigneter Art und Weise zu bearbeiten bzw. zu filtern, um die Sichtbarkeit von Störungen zu reduzieren. Mehrere solcher Verfahren werden im Rahmen dieser Arbeit vorgestellt und bezüglich ihrer Eigenschaften untersucht. Dazu zählt auch ein neues adaptives Verfahren mit mehreren Filtermodi, von dem unterschiedliche Varianten betrachtet werden, deren Leistungsfähigkeit demonstriert wird. Eine erste Vorstellung des Verfahrens erfolgte in [Pia05a].

Die Vielzahl der in der Literatur vorgestellten Enhancement-Ansätze unterscheidet sich in der Komplexität und Funktionsweise, besonders aber in der Wirksamkeit bezüglich einer Verbesserung der Bildqualität in unterschiedlichem Bildmaterial und bei unterschiedlich starker Ausprägung von Störungen. Für einen gegebenen Anwendungszweck stellt sich daher die Frage, welches Verfahren eine hinreichende oder besonders hohe Wirksamkeit bietet, oder wie die Wirksamkeit mehrerer zur Auswahl stehender Verfahren im Vergleich zu beurteilen ist. Zur Beantwortung solcher Fragen können subjektive Tests durchgeführt werden, welche jedoch mit erheblichem Aufwand und hohen Kosten verbunden sind. Wünschenswert sind daher einfach anwendbare Tools zur objektiven Bewertung von Artefaktreduktionsverfahren.

Oft werden zu diesem Zweck sehr einfache Tools wie der MSE oder der PSNR eingesetzt [TCS05, KIL04, RK00]. In der Arbeit wird gezeigt, dass damit mögliche Wirksamkeitsabschätzungen von sehr geringem Wert sind, da eine niedrige Korrelation mit den subjektiven Eindrücken menschlicher Betrachter gegeben ist (siehe auch [BBK05, ASS02, VQE03]). Auch andere - zum Teil sehr komplexe - Tools, die maßgeblich zum Zwecke der Prädiktion von absoluter Bildqualität im Kontext der Bildcodierung entwickelt oder erfolgreich dazu eingesetzt wurden (z. B. Qualitätsmetriken auf Basis von visuellen Wahrnehmungsmodellen), eignen sich mitunter kaum für das beschriebene Problem. Selbst Tools, die in der Lage sind, die Stärke von Blockartefakten abzuschätzen (z. B. gradientenbasierte Blocking-Maße), eignen sich zwar gut zur Prädiktion der Qualität unterschiedlich stark komprimierter Bilder, jedoch nur begrenzt zur Bewertung bzw. zum Vergleich der Wirksamkeit von Artefaktreduktionsverfahren bezogen auf die Verbesserung der Gesamtbildqualität einzelner gestörter Bilder. Ein bedeutendes Problem ist in diesem Zusammenhang, dass viele Qualitätsbewertungsverfahren auf die Erfassung recht großer Bildqualitätsunterschiede ausgelegt sind, beim Vergleich von Artefaktreduktionsverfahren aber eher feine Veränderungen in Bildern auf einem bestimmten Qualitätsniveau erfasst werden müssen.

---

In dieser Arbeit wird ein neuer Ansatz zur objektiven Bewertung der Wirksamkeit von Deblocking-Verfahren vorgestellt, der auf einem Faktorenmodell basiert. Darunter ist zu verstehen, dass mehrere Störungs- oder Qualitätsmaße in geeigneter Weise miteinander kombiniert werden, so dass sich die Wirksamkeit eines Verfahrens bzgl. der Verbesserung der Bildqualität mit einem einzigen Wert ausdrücken lässt. Faktorenmodelle wurden bereits mehrfach zur Beurteilung der absoluten Qualität komprimierter Bilder und Bildfolgen vorgeschlagen [CG02, Fec00, Frä98]. Aber auch sie lassen sich aufgrund einer Optimierung auf die Erfassung recht großer Bildqualitätsunterschiede nicht ohne Weiteres zur Bewertung der Wirksamkeit von Artefaktreduktionsverfahren einsetzen. Die im Rahmen dieser Arbeit entwickelten Faktorenmodelle wurden speziell zu diesem Zweck - und somit auf die Erfassung kleiner Qualitätsunterschiede - optimiert und enthalten verhältnismäßig einfache Qualitätsmaße, die an die Merkmale von Störungen, welche im Zusammenhang mit blockbasierter Bildcodierung und deren Beseitigung stehen, in hohem Maße angepasst sind. Solche Merkmale sind z. B. Diskontinuitäten an Blockgrenzen, die Streuung von Pixelintensitäten im Bereich von Blockgrenzen oder die Veränderung von Kantenverläufen durch Bildkompression und Filterung. Der Fokus liegt auf örtlichen Merkmalen. Zeitliche Effekte werden aktuell nicht betrachtet, sollten aber in Zukunft durchaus einbezogen werden. Die Abbildung der einzelnen Qualitätsmaße auf eine Vergleichsskala (DMOS - Differential Mean Opinion Score) wird über einen linearen Regressionsansatz vorgenommen. Es wurden mehrere Varianten des Verfahrens implementiert, welche in der Arbeit vorgestellt und untersucht werden. Zur Datengewinnung für die durchgeführten Regressionsanalysen und weitere Untersuchungen wurden geeignete subjektive Tests durchgeführt.

Schon in [Fec99] wurde die Verwendung spezialisierter Maße für unterschiedliche Aufgaben im Umfeld der Bildqualitätsbewertung als ein besonders vorteilhafter Ansatz betrachtet. Dies kann im Rahmen der Arbeit bestätigt werden, denn mit dem neuen Qualitätsbewertungsansatz ist speziell eine Prädiktion der Wirksamkeit von Deblocking-Verfahren in Einzelbildern bezogen auf eine Gesamtqualitätsverbesserung möglich, die weit besser mit den subjektiven Eindrücken menschlicher Betrachter korreliert als eine Prädiktion auf Basis oft eingesetzter Qualitätsmaße wie PSNR oder MSE. Es wird gezeigt, dass die Aussagekraft von Schätzungen, die mit Hilfe des neuen Verfahrens ermittelt werden, deutlich höher ist als die Aussagekraft von Schätzungen mehrerer Vergleichsverfahren. Der Aufbau des Verfahrens in Form eines Faktorenmodells bietet trotz der eingesetzten spezifischen Qualitätsmaße einen hohen Grad an Flexibilität, denn es ist eine Anpassung an erweiterte oder andere Aufgabenstellungen sowohl durch den Austausch oder die Hinzunahme von Einzelmaßen als auch durch eine Veränderung des Mappings der einzelnen Qualitätsmaße möglich.

In Kapitel 2 werden zunächst für die Arbeit relevante Grundlagen der Bild- und Videokompression sowie der Entstehung von Störungen in Bildern erarbeitet. Kapitel 3 behandelt relevante Aspekte und Hintergründe der Regressionsanalyse. In Kapitel 4 wird eine Übersicht über Verfahren zur Bewertung von Bildqualität im Kontext der Bildco-

dierung gegeben. Es folgt in Kapitel 5 eine Darstellung des Stands der Technik in der Reduktion von Codierartefakten, bevor in Kapitel 6 ein neues Postprocessing-Verfahren zur Verringerung von Blockartefakten vorgestellt wird. Kapitel 7 enthält die Beschreibung und Auswertung der im Rahmen der Arbeit durchgeführten subjektiven Tests. In Kapitel 8 werden die Einzelmaße eines neuen Ansatzes zur Bewertung der Wirksamkeit von Deblocking-Verfahren sowie mehrere Kombinationen dieser Maße in Form von Regressionsmodellen vorgestellt. Die Modellvarianten werden schließlich in Kapitel 9 bezüglich ihrer Schätzgüte untersucht und mit anderen Verfahren verglichen. In Kapitel 10 werden die Ergebnisse und Erkenntnisse aus dieser Arbeit zusammengefasst und durch einen Ausblick abgerundet.

## 2 Ausgewählte Grundlagen der Bild- und Videokompression

Die Akquisition von Bilddaten durch Kameras oder Filmabtaster erzeugt insbesondere bei hochauflösten Bewegtbildsequenzen sehr große Datenmengen. Beispielsweise benötigt man zur Speicherung einer Sekunde HDTV-Bildmaterials mit der Auflösung von 1280x720 Bildpunkten bei 50 Vollbildern pro Sekunde ca. 132 Megabytes Speicherkapazität (RGB Farbdarstellung, 24 Bit Farbtiefe). Das entspricht einem Speicherbedarf von mehr als 460 Gigabytes pro Stunde. Da solche Datenmengen mit vertretbarem Aufwand weder übertragen noch archiviert werden können, sind Kompressionsverfahren notwendig, auf deren für diese Arbeit relevante Grundlagen in den folgenden Abschnitten eingegangen wird. Die Begriffe Kompression und Codierung werden hier synonym verwendet. Codierung bedeutet zwar primär die Umwandlung von Daten in eine bestimmte Form, im Bereich der Bildverarbeitung zielt eine Bildcodierung jedoch in aller Regel auf eine Kompression bzw. Datenreduktion ab.

Zunächst soll ein allgemeines Bildverarbeitungssystem betrachtet werden. Zweck eines solchen Systems ist entweder die Bildkommunikation mit Bildaufnahme und Bildverarbeitung zur Darstellung auf einem Display oder die Bildaufnahme und -verarbeitung zur Weiterverarbeitung in maschinellen Systemen [Sch98]. Typischerweise enthält ein Bildkommunikationssystem die in Abbildung 2.1 dargestellten bildverarbeitenden Komponenten in Verbindung mit der Bildübertragung oder Speicherung [Ric03].

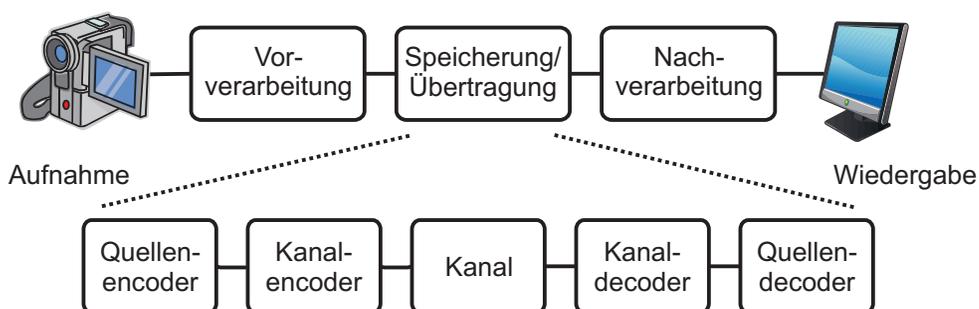


Abbildung 2.1: Typisches Bildverarbeitungssystem

Die Vorverarbeitung dient der Restauration von bereits bei der Aufnahme gestörten Bildern oder der Anpassung an weitere Verarbeitungsschritte. In diesem Zusammenhang

sind beispielsweise Rauschreduktions- oder Deblurring-Verfahren sowie die Kontrastanpassung zu nennen.

Die Speicherung bzw. Übertragung von Bilddaten lässt sich in mehrere Komponenten aufteilen. Die Kanalcodierung dient der Vorbereitung eines Signals zur Übertragung über einen Kanal. Mit der Übertragung über einen Kanal kann auch die Speicherung auf einem Medium gemeint sein. Dem Signal werden in der Regel Zusatzinformationen (Fehlerschutz) hinzugefügt, die dem Kanaldecoder die korrekte Decodierung ermöglichen sollen, auch wenn das Signal auf dem Kanal gestört wird. Die Kanal(de)codierung ist nicht Thema dieser Arbeit.

Die Quellencodierung hat das Ziel, für eine Datenkompression bzw. Datenreduktion zu sorgen, d. h. die Überführung eines Signals in eine Form vorzunehmen, die sich mit weniger Datenbits als die ursprüngliche Form darstellen lässt. Dazu können aus dem Signal redundante und irrelevante Anteile entfernt werden [Ohm95]. Unter redundanten Anteilen sind Informationen zu verstehen, die mehrmals im Signal vorhanden sind oder sich ohne Informationsverlust kürzer bzw. einfacher darstellen lassen. Dies können beispielsweise mehrfach gleiche Bilder in (unbewegten) Bildsequenzen sein. Unter irrelevanten Anteilen sind Informationen zu verstehen, die von der empfängerseitigen Signalsinke nicht dargestellt bzw. wahrgenommen werden und somit unbedeutend sind. Beispielsweise ist es nicht sinnvoll, Signalanteile über ein bandbegrenztes System zu übertragen, die außerhalb der entsprechenden Bandgrenzen liegen.

Die Nachverarbeitung dient schließlich zur Verringerung von Störungen, mit denen das Nutzsignal versehen wurde, oder zur Anpassung des Signals an die Anforderungen des Displays, des menschlichen visuellen Systems oder eventueller nachfolgender Verfahren zur Auswertung des Signals [Sch98]. Dies ist auch das Thema der vorliegenden Arbeit: die Verringerung von Störungen, die im Rahmen der Quellencodierung/-decodierung entstehen sowie die Bewertung solcher Störungen bzw. von Verfahren zur Verringerung der Störungen.

Abbildung 2.1 unterscheidet nicht zwischen monochromen Bildverarbeitungssystemen und solchen, die Farbbilder verarbeiten können, da deren grundsätzlicher Aufbau identisch ist. Unterschiede weist selbstverständlich die Darstellung bzw. Speicherung von monochromen und farbigen Bildinformationen auf. Zur digitalen Darstellung von monochromen Bildern reicht es aus, für jeden Bildpunkt einen Zahlenwert zu verwenden, der die Helligkeit des entsprechenden Punktes repräsentiert. Die Auflösung solcher Helligkeitswerte in digitalen Systemen beträgt in den meisten Fällen 8 Bit, in fortgeschrittenen Systemen auch 10 oder 12 Bit. Für die Darstellung von Farbsignalen werden mehrere Komponenten benötigt. Der insbesondere im Umfeld der Akquisition, Speicherung, Verarbeitung und Darstellung von Bildern in Computersystemen verbreitete RGB-Farbraum basiert beispielsweise auf den drei Primärfarben Rot, Grün und Blau. Durch additive Mischung der Grundfarben sind alle weiteren Farben des RGB-Farbraums darstellbar. Dabei werden die drei Farbanteile (R, G, B) typischerweise mit 8 Bit Auflösung bzw. Farbtiefe verarbeitet.

In dem eng mit dem PAL-Farbfernsehsystem verbundenen YUV-Farbraum werden die Farben ebenfalls durch drei Komponenten Y, U und V dargestellt. Dabei ist Y die Helligkeitskomponente (Luminanz) eines Bildpunktes. U und V sind so genannte Farbdifferenzkomponenten (Chrominanz). Da das menschliche Auge für Informationen in den Farbdifferenzkanälen weniger empfindlich ist als für Informationen im Helligkeitskanal [Ric03], wird im PAL-Farbfernsehsystem die Chrominanz mit geringerer Bandbreite übertragen als die Luminanz. Die Y-Komponente allein ermöglicht eine monochrome Bilddarstellung mit korrektem Bildeindruck. Der YCrCb-Farbraum ist eng mit dem YUV-Farbraum verwandt, wird jedoch in digitalen Fernsehsystemen eingesetzt. In den meisten Fällen werden die einzelnen Komponenten mit 8 Bit Auflösung gespeichert und verarbeitet, wobei zu beachten ist, dass die Chrominanzkomponenten auch negative Werte annehmen können und dementsprechend ein geeignetes Zahlenformat verwendet werden muss. Die Chrominanz wird oft unterabgetastet. Auf diese Weise entstehen mehrere YCrCb-Formate [Ric03]. Die nicht unterabgetastete Variante wird mit YCrCb 4:4:4 bezeichnet. Im Format YCrCb 4:2:2 ist die horizontale Auflösung der Chrominanzkanäle gegenüber dem Luminanzkanal halbiert. Das YCrCb 4:2:0 Format zeichnet sich durch eine Unterabtastung der Chrominanzkanäle um den Faktor zwei in horizontaler und in vertikaler Richtung aus.

Eine Transformation der Farbwerte zwischen unterschiedlichen Farbräumen ist über Transformationsmatrizen möglich. Für weitere Informationen zur Farbdarstellung und zu Farbräumen sei auf [Ric03, Str05, Wen88] verwiesen.

## 2.1 Transformationscodierverfahren

Die Transformationscodierung von Bilddaten basiert auf dem Gedanken, eine mathematische Transformation vom Ortsbereich in einen geeigneten Bildbereich (z. B. Frequenzbereich) vorzunehmen, der die Daten in einer Form repräsentiert, die für eine Kompression besser geeignet ist als die Ausgangsform [NH95]. Da die (spektralen) Eigenschaften von Bildern von Abschnitt zu Abschnitt variieren, wird eine Transformation oft nicht bildweise, sondern blockweise vorgenommen. Dabei ist eine geeignete Blockgröße zu finden. Für Bildabschnitte mit wenigen Bilddetails eignen sich tendenziell eher große Blöcke, für Bildbereiche mit hohem Detailgrad sind hingegen kleine Blöcke sinnvoller [Ohm95]. In der Praxis wird oft eine feste Blockgröße verwendet.

Für die Bilddatenkompression hat sich die diskrete Cosinus-Transformation bewährt. Diese soll zunächst in Gleichung 2.1 für einen  $M \cdot N$  Pixel großen Ausschnitt aus einem Bildsignal  $s(m, n)$  definiert werden [Sch98].

$$S_{DCT}(k, l) = \alpha_k \cdot \alpha_l \cdot \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} s(m, n) \cdot \cos \left[ \frac{\pi (2m+1) k}{2M} \right] \cdot \cos \left[ \frac{\pi (2n+1) l}{2N} \right] \quad (2.1)$$

Es gelte  $m, k = 0, 1, \dots, M - 1$  und  $n, l = 0, 1, \dots, N - 1$ , sowie

$$\alpha_k = \begin{cases} \sqrt{\frac{1}{M}} & k = 0 \\ \sqrt{\frac{2}{M}} & 1 \leq k \leq M - 1 \end{cases}, \alpha_l = \begin{cases} \sqrt{\frac{1}{N}} & l = 0 \\ \sqrt{\frac{2}{N}} & 1 \leq l \leq N - 1 \end{cases}. \quad (2.2)$$

Die entsprechende Rücktransformation lautet:

$$s_{IDCT}(m, n) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \alpha_k \cdot \alpha_l \cdot S(k, l) \cdot \cos \left[ \frac{\pi(2m+1)k}{2M} \right] \cdot \cos \left[ \frac{\pi(2n+1)l}{2N} \right]. \quad (2.3)$$

Die eingeführte 2D-DCT weist eine hohe Energiekonzentration im niederfrequenten Spektralbereich auf, was für eine Datenkompression günstig ist. Daneben sind die rein reelle Hin- und Rücktransformation sehr ähnlich und lassen sich effizient in Software und in Hardware implementieren.

Das bekannte JPEG-Verfahren der *Joint Photographic Experts Group* zur Kompression von Einzelbildern setzt eine Transformationscodierung in der beschriebenen Art ein. In Abbildung 2.2 sind die wichtigsten Komponenten eines JPEG-Encoders als Blockschaltbild dargestellt. Zunächst werden Bilder in 8x8 Pixel große Blöcke eingeteilt. Anschließend werden die Blöcke mit einer 2D-DCT transformiert, bevor die Koeffizienten quantisiert werden.

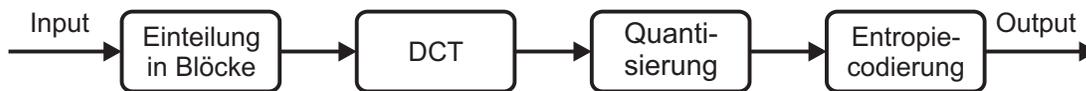


Abbildung 2.2: Blockschaltbild eines JPEG-Encoders

Durch die bereits erwähnte Energiekonzentration im niederfrequenten Spektralbereich werden viele (besonders hochfrequente) Koeffizienten zu Null quantisiert und müssen nicht weiterverarbeitet werden. Dadurch kommt es zu einer Datenreduktion bzw. Kompression der Bilddaten. Je gröber quantisiert wird, umso mehr Koeffizienten fallen weg. Gleichzeitig wird die Auflösung der verbleibenden Koeffizienten reduziert. Die Quantisierung der einzelnen Koeffizienten  $C_{u,v}$  eines Blocks wird über folgende Vorschrift durchgeführt:

$$C_{Q_{u,v}} = \text{round} \left( \frac{C_{u,v}}{Q_{u,v}} \right). \quad (2.4)$$

Es gilt  $u, v \in 0 \dots 7$ . Die Quantisierungsparameter  $Q_{u,v}$  können einer Quantisierungsmatrix entnommen werden, die frei definiert werden kann, aber bei der Decodierung bekannt sein muss, da dort die quantisierten Koeffizienten mit den entsprechenden Werten der

Quantisierungsmatrix zu multiplizieren sind. Der JPEG-Standard [ISO92] schlägt Quantisierungsmatrizen für Luminanz und Chrominanz vor, deren Werte so gewählt sind, dass sie eine recht hohe Datenkompression ermöglichen, den subjektiven Bildeindruck jedoch nicht oder nur wenig verschlechtern. Die Matrizen sind so gestaltet, dass Koeffizienten, die hohen Ortsfrequenzanteilen entsprechen, zunehmend gröber quantisiert werden als solche, die niedrigeren Ortsfrequenzanteilen entsprechen. Eine besonders grobe Quantisierung ist für Koeffizienten vorgesehen, die diagonal hochfrequenten Komponenten entsprechen, da das HVS hierfür unempfindlich ist.

Die Entropiecodierung hat das Ziel, in den Datensymbolen vorhandene Redundanzen zu verringern bzw. zu beseitigen. Dabei werden Symbolen mit unterschiedlicher Auftretswahrscheinlichkeit Codes mit unterschiedlichen Längen zugeordnet. So ist es sinnvoll, häufig auftretenden Symbolen kurze Codes und selten auftretenden entsprechend längere Codes zuzuordnen [Str05]. Auf diese Weise kann die mittlere Datenrate eines Systems minimiert bzw. der Signalentropie, welche ein Maß für den mittleren Informationsgehalt einer Signalquelle ist, angenähert werden.

Im JPEG-Verfahren werden AC- und DC-Koeffizienten unterschiedlich verarbeitet. Grundsätzlich kommen Huffman-Codes (siehe z. B. [PS04]) zum Einsatz, bei den AC-Koeffizienten kombiniert mit einer Lauflängen-Codierung. Die Effizienz der Lauflängen-Codierung wird durch eine vorherige Umsortierung der Koeffizienten in Form eines Zick-Zack-Musters (Zigzag Scan) erhöht. Für detailliertere Informationen hierzu sei auf [Str05, Ric03] verwiesen. Die Decodierung von JPEG-Daten erfolgt im Prinzip in umgekehrter Reihenfolge wie die Encodierung. Die einzelnen Funktionsblöcke sind durch solche mit entsprechend inverser Funktionalität (z. B. IDCT) zu ersetzen.

Es existiert eine Variante von JPEG für die Codierung von Bildsequenzen, die mit Motion JPEG oder MJPEG bezeichnet wird. Das Verfahren sieht eine voneinander unabhängige Verarbeitung aller Einzelbilder nach dem JPEG-Standard vor. Es ist mit überschaubarem Aufwand implementierbar und eignet sich sehr gut für den Videoschnitt, da jedes Bild angesprungen und unabhängig von anderen Bildern decodiert werden kann. Dies bringt auch eine hohe Fehlertoleranz bei der Übertragung von MJPEG-Sequenzen über gestörte Kanäle mit sich, da die Decodierung nach jedem Bild problemlos neu begonnen werden kann. Nachteilig ist vor allem die geringe Kompressionsrate des Verfahrens, da keine zeitlichen Redundanzen in Bildsequenzen ausgenutzt werden.

An dieser Stelle sei auch noch das JPEG-2000-Verfahren erwähnt. Dieses basiert auf einer mehrstufigen Wavelet-Transformation von Bildern und erreicht eine Datenreduktion durch Quantisierung von Wavelet-Koeffizienten. Im Gegensatz zum JPEG-Verfahren erfolgt die Verarbeitung der Bilddaten nicht blockweise, weshalb keine typischen Blockartefakte (siehe weiter unten) auftreten. Daher ist das JPEG-2000-Verfahren für diese Arbeit irrelevant. Für weitere Informationen über JPEG-2000 bzw. die wavelet-basierte Bildkompression sei auf [Str05, Ric03, Mal99] verwiesen.

## 2.2 Hybride Codierverfahren

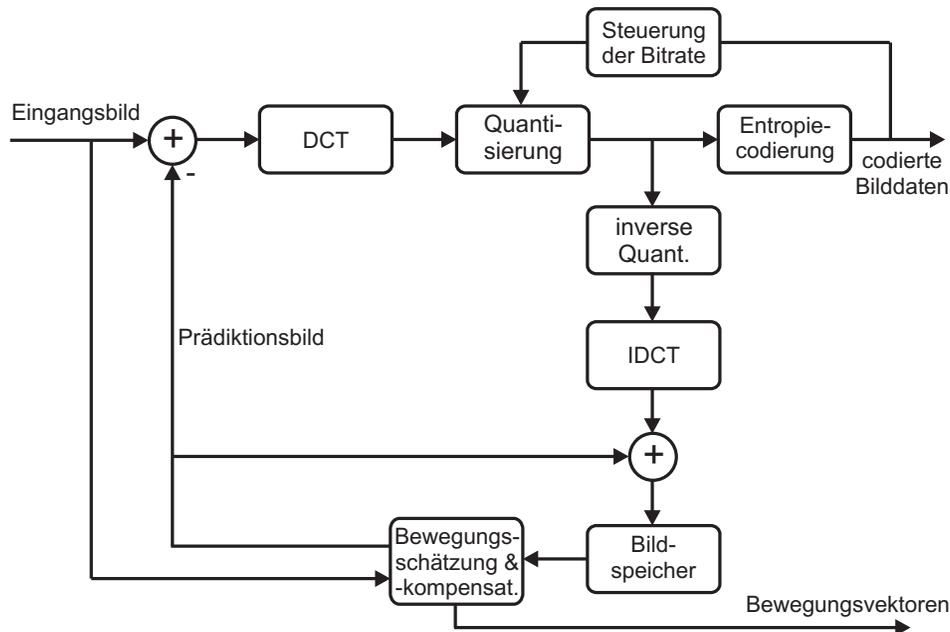
In vielen Verfahren zur Kompression bzw. Codierung von Bildsequenzen werden neben der Transformationscodierung auch Techniken zur Ausnutzung der zeitlichen Korrelationen zwischen aufeinander folgenden Bildern eingesetzt, da sich diese typischerweise nur gering voneinander unterscheiden. Werden auf diese Weise mehrere Dekorrelationstechniken miteinander kombiniert, spricht man von hybrider Bildcodierung [Ohm95].

Hybride Bildcodierverfahren sind z. B. die Bildkompressionsstandards der MPEG-Gruppe (*Motion Picture Experts Group*), also MPEG-1, MPEG-2 und MPEG-4, der neue Standard H.264 (MPEG-4/AVC) und die speziell auf Bildtelefonie bzw. Videokommunikation optimierten Standards H.261 und H.263. Auf die beiden zuletzt genannten Verfahren wird im Folgenden nicht näher eingegangen, da sie in ihrer Funktionsweise den MPEG-Verfahren sehr ähnlich sind, so dass eine gesonderte Behandlung nicht notwendig ist.

Grundsätzlich muss bei Verfahren zur Codierung von Bildsequenzen zwischen Intra- und Inter-Codierung unterschieden werden. Im Fall der Intra-Codierung werden einzelne Bilder unabhängig von anderen Bildern verarbeitet. Die Vorgehensweise entspricht dabei im Wesentlichen dem JPEG-Verfahren (siehe weiter oben). Die Intra-Codierung ist in periodischen Abständen notwendig, um einen Einstieg des Decoders in einen Datenstrom an verschiedenen Stellen zu ermöglichen, was insbesondere bei gestörten Datenkanälen wichtig ist. Werden zeitliche Korrelationen zwischen Bildern ausgenutzt, spricht man von Inter-Codierung. Der Aufbau eines hybriden Video-Encoders ist in Abbildung 2.3 dargestellt.

Im oberen Bereich des Schaltbildes erkennt man die Komponenten eines JPEG-Encoders wieder. Diese werden jedoch nicht auf das aktuelle Eingangsbild, sondern auf die Differenz zwischen Eingangsbild und Prädiktionsbild, also auf ein Prädiktionsfehlerbild, angewendet. Das Prädiktionsbild ist eine Schätzung des aktuellen Eingangsbildes und wird auf Decoderseite aus dem vorherigen Bild mit Hilfe einer Bewegungskompensation erzeugt. Daher müssen auch im Encoder die wichtigsten Komponenten des Decoders (Dequantisierung und IDCT) und ein Bildspeicher vorhanden sein. Die Aufgabe der Bewegungskompensation ist es, mit Hilfe einer Bewegungsschätzung im vorherigen Bild Blöcke zu finden, die sich so verschieben lassen, dass sie möglichst genau das aktuelle Bild repräsentieren. Auf diese Weise kann auf Basis des vorherigen Bildes und von Bewegungsvektoren für einzelne Bildblöcke ein Prädiktionsbild erzeugt werden. Man spricht hier von einer Vorwärtsprädiktion. Zur Bewegungsschätzung werden so genannte Block-Matching-Verfahren eingesetzt. Grundlagen der Bewegungsschätzung finden sich z. B. in [dH00, Hen98].

Die transformierten und quantisierten Prädiktionsfehlerbilder werden - ähnlich wie beim JPEG-Verfahren - einer Entropiecodierung unterzogen. Der dabei entstehende Datenstrom wird bezüglich der Datenrate überwacht. Durch Steuerung der Quantisierung kann eine definierte Datenrate eingehalten werden. Auch die Bewegungsvektoren werden



**Abbildung 2.3:** Aufbau eines hybriden Video-Encoders, Inter-Codierung, angelehnt an [Str05]

geeignet codiert und schließlich zusammen mit den transformierten Bilddaten übertragen oder gespeichert.

In Abbildung 2.4 ist der entsprechende Decoder dargestellt. Zunächst muss die enco-derseitige Entropiecodierung der Bilddaten wieder rückgängig gemacht werden. Es folgt eine inverse Quantisierung und inverse DCT mit dem Zweck, die Bilddaten wieder in den Ortsbereich zu überführen. Für den Fall eines Intra-Bildes ist die Decodierung nun abgeschlossen. Im Falle von Inter-Bildern stellen die Bilddaten bis zu diesem Punkt nur ein Prädiktionsfehlerbild dar. Deshalb ist noch eine aus bereits decodierten Daten und Bewegungsvektoren zu erzeugende Prädiktion für das aktuelle Bild zu addieren.

## MPEG-1

Die Arbeiten am MPEG-1-Standard begannen in der 80er Jahren und die Verabschiedung erfolgte 1993 [ISO93]. Der Standard sollte die Kompression von Video- und Audiodaten für CD-Rom Anwendungen bei Datenraten bis ca. 1,5 Mbit/s ermöglichen. Primär sind Bildauflösungen von 352x288 oder 352x240 Pixeln bei bis zu 30 Bildern bzw. Frames pro Sekunde vorgesehen. Andere Auflösungen sind jedoch möglich. Das verwendete Farbformat ist YCrCb 4:2:0.

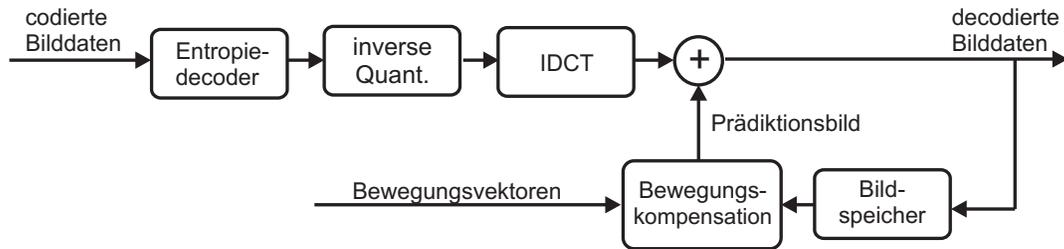


Abbildung 2.4: Aufbau eines hybriden Video-Decoders

Die Syntax eines MPEG-1 Datenstroms lässt sich hierarchisch in die in Abbildung 2.5 dargestellten Bestandteile aufteilen [ISO93, Ric03, Sch04]. Eine Videosequenz besteht aus GOPs (Group Of Pictures / Gruppe von Bildern), in denen mehrere Bilder bzw. Frames zusammengefasst sind. Ein Bild kann in mehrere streifenförmige Slices aufgeteilt werden, die jeweils aus mehreren Makroblöcken bestehen. Ein Makroblock besteht jeweils aus vier Luminanz-Blöcken der Größe 8x8 Pixel und jeweils einem 8x8 Pixel großen Block aus den beiden unterabgetasteten Chrominanz-Kanälen. Pro Makroblock können auch ein oder zwei Bewegungsvektoren übertragen werden.

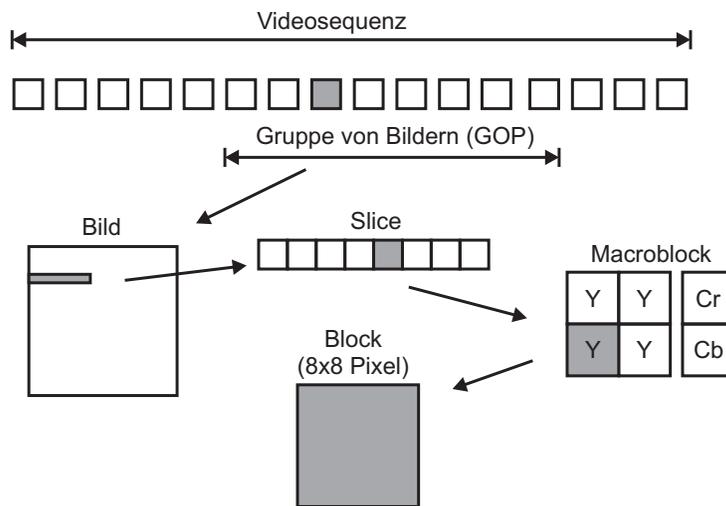


Abbildung 2.5: Struktur eines MPEG-Streams

Innerhalb einer GOP weisen die übertragenen Frames nicht die ursprüngliche (Aufnahme-)Reihenfolge auf. Der Grund dafür ist die im Verfahren vorgesehene Prädiktionsschleife. Im MPEG-Verfahren ist es möglich, dass die Prädiktion für das aktuelle Bild nicht nur auf Basis vergangener, sondern auch auf Basis zukünftiger Bilder erzeugt wird. Man unterscheidet in diesem Zusammenhang drei Frametypen. So genannte I-Frames

werden im Intra-Modus codiert, es kommt also keine zeitliche Prädiktion zum Einsatz. P-Frames werden im Inter-Modus, also mit Hilfe von Bewegungskompensation auf Basis eines vergangenen Frames codiert. B-Frames werden ebenfalls im Inter-Modus codiert, die Bewegungskompensation kann aber auf zwei Frames zugreifen, von denen eines auch in der Zukunft liegen darf. Einzelne Makroblöcke können in P- und B-Frames auch im Intra-Modus codiert werden.

Damit eine Decodierung überhaupt möglich ist, müssen die Frames vor der Übertragung so umsortiert werden, dass zu jeder Zeit alle benötigten Daten im Decoder vorliegen. In Abbildung 2.6 ist ein Beispiel für die Umsortierung der Frames in einer GOP gegeben, so dass jederzeit eine Decodierung auch der B-Frames im Decoder möglich ist. Grundsätzlich ist es zulässig, nur I-Frames zu verwenden. Der Prozess der Encodierung und Decodierung wird dadurch erheblich vereinfacht, jedoch auf Kosten der möglichen Datenreduktion, da keine zeitlichen Redundanzen ausgenutzt werden. Die ausschließliche Verwendung von I-Frames entspricht im Wesentlichen der Motion JPEG Codierung.

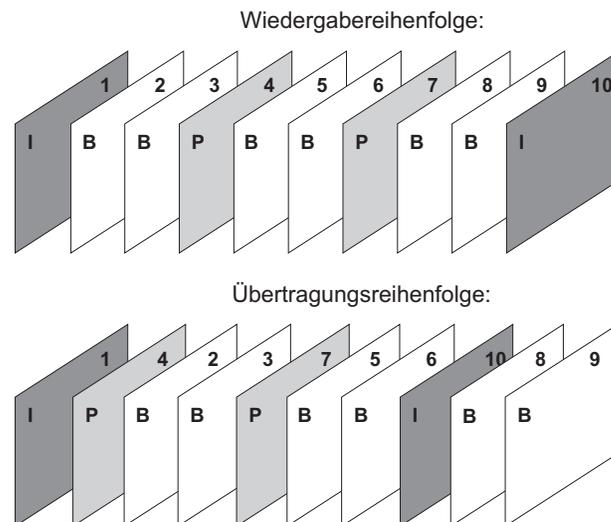


Abbildung 2.6: Bildtypen und Übertragungsreihenfolge

Die Quantisierungsvorschrift des MPEG-1-Standards [ISO93] für AC-Koeffizienten eines Intra-Bildes ausgehend von den unquantisierten Koeffizienten  $C_{u,v}$  gibt Gleichung (2.5) an:

$$C_{Q_{u,v}} = \text{round} \left( 8 \cdot \frac{C_{u,v}}{QP \cdot Q_{u,v}} \right). \quad (2.5)$$

$QP$  ist dabei ein Quantisierungsparameter im Wertebereich  $1 \dots 31$ , der hauptsächlich der Datenratensteuerung dient und jeweils für einen Makroblock konstant ist.  $Q_{u,v}$  ist eine Quantisierungsmatrix vergleichbar mit der im JPEG-Verfahren verwendeten. Der

DC-Koeffizient wird grundsätzlich durch Acht geteilt und auf den nächsten Integer-Wert gerundet. In P- und B-Bildern werden alle Koeffizienten ebenso wie in Gleichung (2.5) quantisiert. Es kommt jedoch eine andere Quantisierungsmatrix zum Einsatz. Für weitere Informationen zu MPEG-1 siehe z. B. [Ric03, Str05, NH95].

## MPEG-2

1996 wurde der MPEG-2-Standard [ISO96] als Nachfolger des MPEG-1-Standards verabschiedet. MPEG-2 wurde primär für die digitale Übertragung und Speicherung von Fernsehbildern entwickelt und so setzen heute Systeme wie DVB und DVD diesen Codierstandard ein. Die Funktionsweise ähnelt sehr stark der von MPEG-1, aber es wurden diverse Erweiterungen und Verbesserungen integriert. MPEG-2 unterstützt niedrige Bildauflösungen ebenso wie typische SD-Fernsehaufösungen (720x480, 720x576 Pixel) und HDTV-Auflösungen (1280x720, 1920x1080 Pixel). Zum Farbformat YCrCb 4:2:0 kamen die Formate YCrCb 4:2:2 und YCrCb 4:4:4 hinzu.

Eine für den Fernsbereich sehr bedeutende Erweiterung stellt die Unterstützung von Zeilensprung-/Interlaced-Bildformaten dar. Dazu können Frames in ein Paar von Fields aufgeteilt werden (Field-Modus), die dann separat codiert werden - also vergleichbar mit der Codierung von progressiven Frames. Dazu war die Definition von I-, P-, und B-Fields notwendig. Ein I-Frame muss mindestens ein I-Field enthalten, das zweite Field kann jedoch auch ein P- oder B-Field sein. Die Prädiktion in P- und B-Fields kann ausgehend von Fields in anderen Frames oder in demselben Frame erfolgen.

Auch im üblichen Frame-Modus besteht die Möglichkeit einer speziellen Behandlung von Interlaced-Bilddaten. Dazu ist es in jedem Makroblock möglich, die Bilddaten so umzusortieren, dass sich in der oberen Hälfte nur Zeilen aus dem einen Field und in der unteren Hälfte nur Zeilen aus dem anderen Field befinden (Field DCT Coding). Abbildung 2.7 veranschaulicht dies.

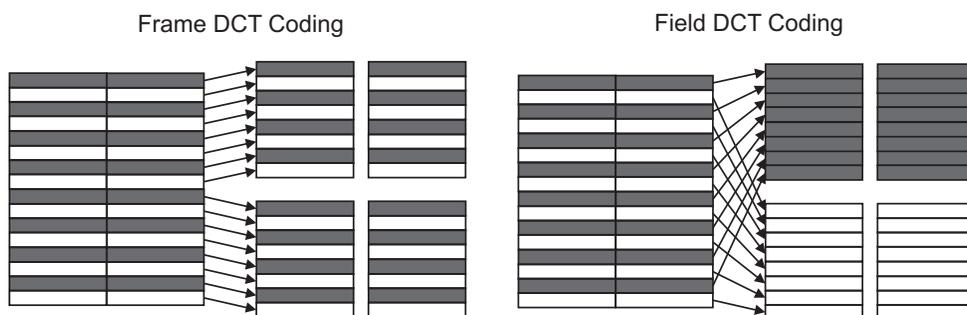


Abbildung 2.7: Frame DCT Coding und Field DCT Coding

Des Weiteren ist zu nennen, dass die Bewegungsschätzung bis auf Halbpixel-Genauigkeit verbessert wurde und dass es als Erfordernis der Interlaced-Unterstützung auch möglich ist, für die obere und untere Hälfte eines Makroblocks getrennte Bewegungsvektoren zu definieren.

Da Funktionalität und Möglichkeiten des MPEG-2-Standards sehr groß sind, wurden so genannte Profiles und Levels eingeführt, die jeweils Teilmengen der Gesamtfunktionalität definieren. Auf diese Weise soll sichergestellt werden, dass Systeme, die nicht die gesamte Funktionalität benötigen, nicht alle Funktionen implementiert haben müssen, aber dennoch zu ähnlichen Systemen kompatibel sind. Wenn in einem System beispielsweise ein bestimmtes Profile und Level implementiert sind, so kann davon ausgegangen werden, dass dieses System mit anderen Systemen, die über entsprechende Implementierungen verfügen, kompatibel ist. Die Profiles bestimmen grundsätzliche Eigenschaften eines MPEG-Systems, beispielsweise die Unterstützung aller oder nur einiger Bild-Typen (I-Bilder, P-Bilder, ...), die Unterstützung des Zeilensprungverfahrens oder das Farbformat (z. B. YCrCb 4:2:0). Die Levels definieren unterstützte zeitliche und örtliche Auflösungsbereiche. Das in DVB- und DVD-Systemen eingesetzte Main Profile @ Main Level unterstützt beispielsweise alle Bild-Typen, das Farbformat YCrCb 4:2:0, Interlaced-Codierung und Auflösungen bis 720x576 Pixel bei bis zu 30 Frames/s. Für eine Aufstellung aller Profiles und Levels sei auf den Standard [ISO96] verwiesen.

In den MPEG-2-Standard wurde als Option auch die skalierbare Codierung integriert. Dadurch wird es möglich, in einem MPEG-2 Datenstrom mehrere Qualitätsstufen (Layer) zu übertragen. Eine Skalierbarkeit ist bezüglich der örtlichen und zeitlichen Auflösung einer Bildsequenz und bezüglich der generellen Bildqualität (SNR-Scalability) vorgesehen. Durch dieses Konzept kann beispielsweise ein leistungsschwacher Decoder nur den Base-Layer eines Datenstroms, der nur eine eingeschränkte Bildqualität bietet, decodieren, während ein leistungsstärkerer Decoder auch einen oder mehrere Enhancement-Layer, die die örtliche und zeitliche Auflösung des decodierten Bildmaterials erhöhen, decodieren kann. Für weitere Informationen zu MPEG-2 siehe z. B. [Ric03, Str05, RH96].

## MPEG4

1998 wurde der MPEG-4-Standard (genauer MPEG-4 Part 2: Visual) [ISO98] eingereicht und 1999 verabschiedet. Die zur Zeit aktuellste Version ist [ISO04]. MPEG-4 stellt in vielerlei Hinsicht eine Erweiterung der Möglichkeiten der früheren MPEG-Standards kombiniert mit Verbesserungen zum Erreichen einer höheren Codiereffizienz dar. Es werden Datenraten von wenigen kbit/s bis zu mehr als 1 Gbit/s bei Auflösungen bis zu 4k x 4k Pixel unterstützt. Die grundsätzliche Vorgehensweise bei der Codierung und Decodierung lässt jedoch viele Ähnlichkeiten mit MPEG-1/MPEG-2 erkennen, beispielsweise die nach wie vor verwendete Transformationscodierung auf Basis einer 8x8 DCT oder der hierarchische Aufbau eines Videostroms.

Eine der größten Neuerungen stellt das Object Based Coding dar. Dadurch wird es möglich, eine Videoszene in mehrere Video-Objekte (VO) einzuteilen, die separat codiert werden. Dies können z. B. ein Hintergrund-VO und mehrere Vordergrund-VOs sein, wie exemplarisch in Abbildung 2.8 anhand von zwei Vordergrund-VOs dargestellt. Ein Abbild eines VO zu einem bestimmten Zeitpunkt nennt sich Video Object Plane (VOP). Wird der ganze Bildausschnitt als rechteckiges VO behandelt, entsprechen die VOPs den Frames in den früheren MPEG-Standards.



**Abbildung 2.8:** Beispiel für die Definition von Video-Objekten

Die Bewegungsschätzung hat bei MPEG-4 eine wählbare Genauigkeit auf bis zu Viertelpixel-Niveau. Für jeden Block können Bewegungsvektoren ermittelt und übertragen werden. Zusätzlich ist eine globale Bewegungskompensation für einzelne Video-Objekte vorgesehen, wodurch die Bewegung ganzer Bildbereiche mit sehr wenigen Parametern möglich wird.

Die funktionalen Teilkomponenten des MPEG-4 Standards werden Tools genannt. Das Hinzufügen neuer Tools nach Verabschiedung des Standards ist vorgesehen, wodurch eine hohe Flexibilität gegeben ist. Kompatibilität wird ähnlich wie bei MPEG-2 durch Profiles sichergestellt. Profiles definieren Teilmengen von MPEG-4 Tools, die mindestens implementiert sein müssen. Dabei gibt es z. B. für mobile Geräte auch Profiles, die die Implementierung nur eines kleinen Satzes an Tools erfordern.

Ein Tool, für das es in früheren Standards keine Entsprechung gab, ist das so genannte Shape Coding. Es stellt Methoden zur Beschreibung der Form von Video-Objekten bereit. Es ist in diesem Zusammenhang auch möglich, jedem einzelnen Pixel eines VO eine einstellbare Transparenz im Wertebereich  $0 \dots 255$  zuzuweisen. Des Weiteren werden Tools zur Fehlerverschleierung (error resilience), die die Robustheit eines MPEG-4-Systems ge-

genüber Übertragungsfehlern steigern sollen, zur Verfügung gestellt. Auch Methoden zur örtlichen und zeitlichen Skalierbarkeit sind in einem Tool zusammengefasst.

Daneben gibt es noch Tools zur Unterstützung der Codierung von 3D-Objekten. Derartige Objekte können in Gittermodelle und Texturen aufgeteilt und getrennt verarbeitet werden. Darüber hinaus sind im MPEG-4-Standard Methoden zur Animation von Gesichtern vorgesehen, die im Decoder gerendert werden. Auf diese Weise lässt sich Videotelefonie mit sehr geringer Datenrate realisieren, da neben der einmaligen Übertragung eines Gesichts nur noch Bewegungsparameter übertragen werden müssen.

Erwähnung finden sollte im Zusammenhang mit MPEG-4 noch ein optionales Verfahren zur adaptiven Reduktion von Blockartefakten, das im Anhang F des Standards vorgeschlagen wurde ([ISO00, KYKR99]) und in Kapitel 5 näher beschrieben wird. Einen guten Überblick über den MPEG-4-Standard liefern [Ric04, ISO02].

## H.264

Der aktuellste Vertreter der hybriden Videocodecs ist H.264, auch bekannt als MPEG-4/AVC oder MPEG-4 Part 10 [ISO03]. Auch dieser Standard basiert grundsätzlich auf MPEG-1 bzw. MPEG-2. Entwicklungsziel war primär eine besonders hohe Codiereffizienz bei typischen Videosequenzen aus rechteckförmigen Bildern. Viele der in MPEG-4 Part 2: Visual eingeführten Features (z. B. Codierung von beliebig geformten Video-Objekten) sind in H.264 nicht enthalten. Auf einige ausgewählte Details von H.264 soll im Folgenden eingegangen werden. Weitere Informationen finden sich z. B. in [OBL<sup>+</sup>04, Ric04].

Die Blockgröße für die Bewegungskompensation ist in mehreren Schritten zwischen 16x16 und 4x4 Pixeln wählbar und die Bewegungsschätzung hat Viertelpixel-Auflösung. Es können zur Verbesserung der Bewegungskompensation mehrere Referenzframes verwendet werden, was im Encoder und Decoder einen erhöhten Aufwand für die Speicherung der Referenzbilder mit sich bringt.

Statt einer DCT werden in H.264 verschiedene Integer-Transformationen eingesetzt, die auf Blöcke der Größe 4x4 Pixel angewendet werden. In bestimmten Fällen der Codierung von Chrominanzinformationen ist eine Transformation der Größe 2x2 vorgesehen. Die geringen und teilweise variablen Blockgrößen erschweren eine Reduktion von Blockartefakten mit reinen Postprocessingverfahren deutlich. Es ist jedoch in H.264 ein adaptives Deblocking-Filter zur Störungsreduktion spezifiziert.

Das Filter ist zur direkten Integration in den Encoder und Decoder vorgesehen. Es wird jeweils nach der inversen Transformation angewendet (In-Loop Filter), was sinnvoll ist, da prädizierte Bilder auf diese Weise von bereits artefaktreduzierten Referenzbildern profitieren können. Das Filter selbst bearbeitet Blockgrenzbereiche und unterscheidet dabei zwei Filtermodi für homogene und andere Bildbereiche. Die Filterung wird abhängig von Quantisierungsinformationen durchgeführt. Eine detaillierte Beschreibung des Filters findet sich in [LJL<sup>+</sup>03].

H.264 sieht gegenüber MPEG-2 eine flexiblere Handhabung von Makroblöcken vor. Makroblöcke können so genannten Slice Groups zugeordnet werden, die aus mehreren Slices bestehen können. Als Konsequenz daraus ist es möglich, dass Makroblöcke in einer Reihenfolge übertragen werden, die nicht der üblichen Raster-scan-Reihenfolge entsprechen muss. Durch die flexiblen Zuordnungsmöglichkeiten können z. B. auch Vordergrund- und Hintergrund-Slices definiert werden.

Das Ziel einer deutlich erhöhten Codiereffizienz konnte mit H.264 im Vergleich zu älteren Videocodierverfahren erreicht werden - allerdings auf Kosten eines erheblich gesteigerten Implementierungsaufwands. Das Verfahren eignet sich für einen Einsatz in Systemen, in denen sehr große Bilddatenmengen komprimiert werden sollen. Dazu zählen z. B. HDTV-Rundfunksysteme und die HD-Nachfolger der DVD.

## 2.3 Artefakte der digitalen Codierung und Übertragung

In einem Bildverarbeitungssystem können an vielen Stellen bzw. in vielen Teilkomponenten verschiedenste Störungen - auch Artefakte genannt - entstehen, die dem Nutzsignal überlagert werden und dieses degradieren. In Abbildung 2.9 ist veranschaulicht, welche typischen Störquellen in den Komponenten eines digitalen TV-Systems existieren. Im Folgenden werden einzelne Störquellen bzw. die von ihnen erzeugten Störungen näher betrachtet. Den Schwerpunkt bilden dabei die Codierartefakte.

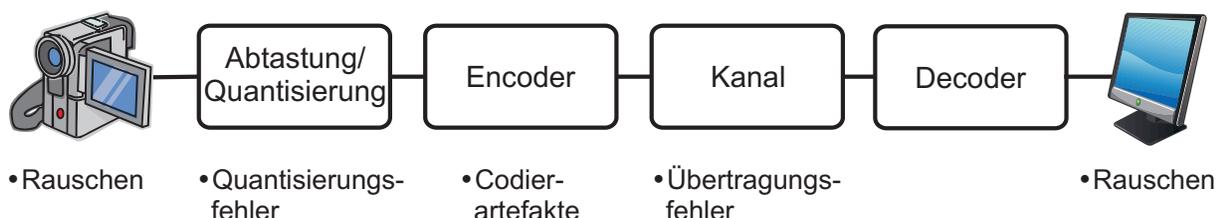


Abbildung 2.9: Störquellen in einem digitalen TV-System

Rauschquellen sind in Bildverarbeitungssystemen stets vorhanden. Von Bedeutung sind sie hauptsächlich im Umfeld der analogen Teilkomponenten, welche auch in digitalen TV-Systemen bei der Bildakquisition und Bildwiedergabe nahezu immer beteiligt sind. Insofern sollte Rauschen, das man als messbare Schwankung von physikalischen Größen beschreiben kann, grundsätzlich nicht vernachlässigt werden. Bei der Bildaufnahme mit CCD- oder CMOS-Kameras entstehen im Wesentlichen Schrotrauschen und Dunkelstromrauschen. Dies ist auf Zufallsprozesse bei der Generation von diskreten Ladungsträgern zurückzuführen [Jos01]. Daneben tritt in jedem Wirkwiderstand und somit in jedem realen elektronischen Bauteil thermisches Rauschen aufgrund der Brownschen Bewegung der

Ladungsträger auf. Bei der Abtastung von Filmmaterial entsteht Filmkornrauschen als Nebeneffekt der schwankenden Dichte von im Material vorhandenen Silberhalogeniden [SB00].

Rauschen und die Reduktion von Rauschstörungen (siehe hierzu z. B. [Jos01, PK04]) sind nicht Gegenstand dieser Arbeit. Es sei jedoch angemerkt, dass Rauschen im Zusammenhang mit digitaler Bildcodierung nicht grundsätzlich unbedeutend ist, da sich Rauschen negativ auf die Codiereffizienz von Videocodierverfahren auswirkt [SK00]. Insofern ist darauf zu achten, am Eingang von Encodern möglichst rauscharme Bildsignale zu verwenden, also gegebenenfalls eine Rauschreduktion vorzuschalten, um eine maximale Effizienz zu erreichen.

Die in Abbildung 2.9 aufgeführten Quantisierungsfehler entstehen als Nebeneffekt jeder skalaren Quantisierung, also der Überführung einer kontinuierlichen Signaldarstellung in eine diskrete Signaldarstellung (z. B. im Rahmen der Analog/Digital-Wandlung). Dabei wird ein Intervall an reellen Eingangswerten einem einzigen Ausgangswert zugewiesen, wodurch eine Quantisierungsoperation nicht eindeutig umkehrbar ist und Informationsverlust mit sich bringt. Typischerweise wird ein Eingangswert  $x$  im Intervall  $[(m - \frac{1}{2}) \cdot \Delta, (m + \frac{1}{2}) \cdot \Delta]$  auf den Ausgangswert  $x_q = m \cdot \Delta$  abgebildet.  $\Delta$  steht hier sowohl für die Quantisierungsstufenhöhe als auch für die Quantisierungsintervallbreite. Es ist allerdings nicht zwingend, dass diese beiden Größen identisch sind.  $m$  kann als Ordnungszahl des Intervalls bzw. der zugehörigen Quantisierungsstufe betrachtet werden und sowohl positiv als auch negativ ganzzahlig sein. Der Quantisierungsfehler  $e_q$  lässt sich als Differenz zwischen Originalwert und quantisiertem Wert darstellen:

$$e_q = x - x_q. \quad (2.6)$$

Er ist begrenzt auf den Wertebereich  $-\frac{\Delta}{2} \leq e_q \leq \frac{\Delta}{2}$ . Bei hinreichend feiner Quantisierung (kleine Quantisierungsstufenhöhe  $\Delta$ ) kann eine Gleichverteilung für den Quantisierungsfehler angenommen und die Fehlervarianz zu  $\sigma_q^2 = \frac{\Delta^2}{12}$  berechnet werden [Str05]. Somit kann die Quantisierungsstörwirkung bei einer hinreichend feinen Quantisierung und hinreichend komplexen Nutzsignalen durch Addition von Rauschen mit weißer spektraler Leistungsdichte mit der Leistung  $P_q = \frac{\Delta^2}{12}$  und gleichverteilter Amplitude (im Rahmen der möglichen Werte) zum unquantisierten Eingangssignal beschrieben werden [SB00, OSB04]. Bei grober Quantisierung oder einfachen Nutzsignalen (z. B. Sprung) kann der Quantisierungsfehler nicht als unkorreliert zum Nutzsignal betrachtet werden.

Zu den in Abbildung 2.9 genannten Übertragungsfehlern zählen alle Fehler, die während der Übertragung eines Signals über einen Kanal entstehen können. Ein Kanal kann in diesem Zusammenhang auch durch die Speicherung eines Signals auf einem Speichermedium und das spätere Auslesen gebildet werden. Zu unterscheiden sind grundsätzlich analoge und digitale Systeme. Übertragungsstörungen in analogen Rundfunksystemen lassen sich in guter Näherung durch weißes Rauschen mit gaußförmiger Amplitudenverteilung mo-

dellieren [Hen98]. Daneben können in Bildübertragungssystemen, die eine Frequenzmodulation verwenden, kurze, horizontal ausgedehnte helle oder dunkle Streifen, so genannte „Fische“, auftreten, die einen eher impulsförmigen Charakter aufweisen. Da analoge Übertragungstrecken in der Praxis zunehmend an Bedeutung verlieren, soll hierauf nicht näher eingegangen werden.

Auf digitalen Übertragungstrecken können Störungen Bit- oder Burstfehler verursachen, also die Verfälschung oder die Unterdrückung einzelner oder zusammenhängender Informationsträger. Die Folgen von Bit- und Burstfehlern können sehr unterschiedlich sein und hängen von vielen Faktoren ab, z. B. vom eingesetzten Kanalcodierer/-decoder und dem verwendeten Video-Codec. Die Kanalcodierung hat unter anderem die Aufgabe, einzelne Störungen auf dem Kanal verlustfrei auszugleichen. Dies ist jedoch nur bei einer begrenzten Anzahl und Konzentration von Störungen möglich. Können Störungen auf dem Übertragungskanal vom Kanaldecoder nicht mehr beseitigt werden, verhindern diese eine korrekte Verarbeitung der Nutzsignale im Video-Decoder. Im Falle einer MPEG-Codierung können sich Burst-Fehler beispielsweise durch verfälschte oder komplett fehlende Blöcke, Makroblöcke oder Slices im decodierten Bild auswirken (Beispiele in [PKL97, SL99]). Auch einzelne Bitfehler können an ungünstigen Stellen im Videostrom ähnliche Fehler verursachen [Sch04]. Für den Fall, dass die Decodierung durch starke Störungen unterbrochen werden sollte, sind in Videocodierverfahren typischerweise Mechanismen integriert, die einen neuen Beginn der Decodierung z. B. am Beginn von Slices oder I-Frames zulassen. Da sich diese Arbeit nicht mit den Auswirkungen und der Beseitigung von Übertragungsstörungen beschäftigt, sei für weitere Informationen auf die Literatur verwiesen.

Die in Abbildung 2.9 aufgeführten Codierartefakte treten beim Einsatz aller verlustbehafteten Bildkompressionsverfahren auf. Die Ursache von Codierartefakten liegt in der Quantisierung von Bilddaten bzw. transformierten Bilddaten. Im Folgenden soll eine Übersicht über die störendsten und häufigsten Codierartefakte gegeben werden. Dabei werden hauptsächlich solche Artefakte betrachtet, die im Rahmen einer blockbasierten Bildcodierung entstehen, da blockbasierte Codierverfahren bislang die mit Abstand größte Praxisrelevanz haben. Beispielsweise wird in den weit verbreiteten und seit Jahren eingesetzten digitalen TV-Rundfunksystemen DVB-S und DVB-C als Codierverfahren MPEG-2 eingesetzt. Um über einen Transponder mehrere TV-Kanäle übertragen zu können, werden für jeden einzelnen Kanal oft nur geringe Datenraten vorgesehen, was zwangsläufig zu Codierartefakten führt. Im terrestrischen DVB-T-System ist diese Tendenz besonders ausgeprägt, da ebenfalls eine blockbasierte Videocodierung nach MPEG-2 eingesetzt wird, im Vergleich zu DVB-S und DVB-C jedoch eine deutlich geringere Bandbreite zur Verfügung steht. Auf Video-CDs wird eine blockbasierte Videocodierung nach MPEG-1 in Verbindung mit recht geringen Datenraten eingesetzt, was in vielen Fällen ebenfalls zur Entstehung sichtbarer Störungen führt. Im Falle der DVD (MPEG-2) treten Codierartefakte aufgrund von hohen Datenraten nur in geringem Maße auf. Gleiches gilt momentan auch für die an Bedeutung gewinnenden HDTV-Rundfunksysteme und die HD-Nachfolger der DVD. Darüber hinaus

werden in diesen Systemen modernere Video-Codecs (z. B. H.264) eingesetzt, die integrierte Maßnahmen zur Verringerung von Codierartefakten besitzen. Dennoch arbeiten auch diese Codecs blockbasiert und verlustbehaftet und Störungen bleiben nicht in jedem Fall aus.

Als typische Arten von Codierartefakten sind Blockartefakte bzw. Blocking, DCT-Basisbilder, Staircase-Effekte, Ringing und Blurring zu nennen. Eine umfassende Übersicht liefert [YW98]. Dort werden auch weniger bedeutende Artefakte oder solche, die sich letztlich als spezielle Ausprägung einer der bereits genannten Arten von Artefakten interpretieren lassen, genannt. In Abbildung 2.10 werden in einem Beispielbild zunächst einige der im Folgenden betrachteten Codierartefakte gezeigt.

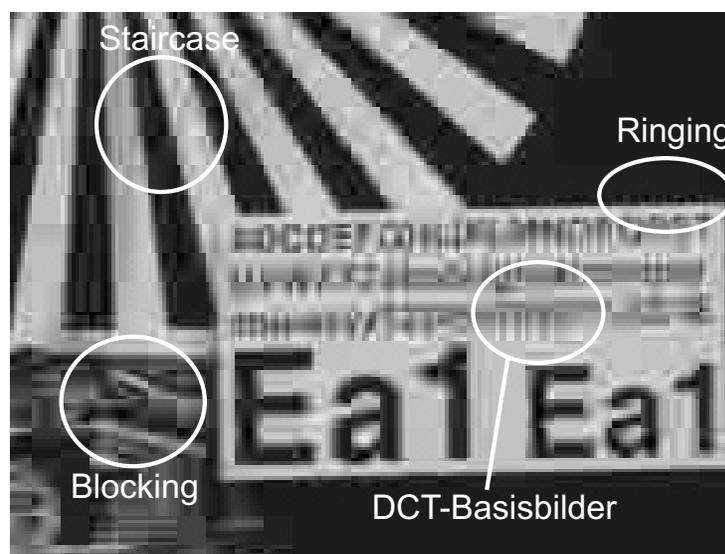


Abbildung 2.10: Beispiele für Codierartefakte

### Blocking

Das am häufigsten auftretende und oftmals störendste Codierartefakt ist das so genannte Blocking [Win05, Ric03, Tra01, RL83]. Es äußert sich durch Diskontinuitäten an bestimmten Bildpositionen, die durch die Lage der Blockgrenzen bzw. die Blockgröße des verwendeten Codiervorgangs vorgegeben sind. Blockartefakte treten sowohl in blockbasiert codierten Einzelbildern als auch in Bildsequenzen auf. Sie können auf mehrere Ursachen zurückgeführt werden.

In Einzelbildern bzw. in Intra-Frames besteht die Hauptursache in der isolierten Verarbeitung jedes einzelnen Bildblocks gleicher Größe. In jedem Block entstehen durch die

Quantisierung der Transformationskoeffizienten individuelle Quantisierungs- bzw. Codierfehler, die in ihrer Stärke und Charakteristik von den Fehlern in den benachbarten Blöcken abweichen können [YW98]. Intensitätssprünge oder ein unnatürlicher Wechsel von Mustern bzw. Texturen an Blockübergängen sind die Folge. Das Auftreten der Störungen in regelmäßigen Abständen sowie die horizontale und vertikale Ausrichtung der Diskontinuitäten sind dafür verantwortlich, dass das HVS besonders empfindlich für Blockartefakte ist. In hohem Maße für die beschriebenen Störungen verantwortlich ist eine grobe Quantisierung des DC-Koeffizienten [Ric03]. Da dieser Koeffizient die durchschnittliche Intensität bzw. Helligkeit eines Blocks bestimmt, werden Quantisierungsfehler durch abrupt wechselnde Grundintensitäten bzw. -helligkeiten von Blöcken sichtbar, was die Positionen der Blockgrenzen hervortreten lässt. Dies ist insbesondere in homogenen Bildbereichen sehr störend. Solche Artefakte findet man z. B. in Abbildung 2.13 bei 3,5 Mbit/s in der Rasenfläche, auch wenn hier die Homogenität selbst ein Resultat grober Quantisierung und nicht schon im Original vorhanden ist.

Aber auch die grobe Quantisierung signifikanter bzw. verhältnismäßig großer AC-Koeffizienten (dies sind meist die niederfrequenten Koeffizienten) ist für die Sichtbarkeit von Blockartefakten bedeutend. Denn hierdurch entstehen die bereits erwähnten Wechsel von Mustern bzw. Texturen an Blockgrenzen. Die typischerweise weichen Übergänge zwischen Blöcken werden beeinträchtigt und das „Zusammenpassen“ benachbarter Blöcke wird gestört [Ric03]. Solche Artefakte findet man in Abbildung 2.13 hauptsächlich bei 5 Mbit/s und bei 8 Mbit/s. Es sei erwähnt, dass je mehr Aktivität in einem Bildbereich vorhanden ist, Blockartefakte umso stärker maskiert werden können. Für die Sichtbarkeit ist demnach nicht nur die Stärke der Quantisierung, sondern auch der Bildinhalt verantwortlich [YW98].

In prädiktiv codierten Frames können ebenfalls Blockartefakte entstehen. Hierfür gibt es hauptsächlich zwei Ursachen. Zum einen werden die im Encoder ermittelten Prädiktionsfehlerbilder ebenso einer blockweisen Transformation und einer Quantisierung unterzogen wie Intra-codierte Bilder. Insofern können auch in Prädiktionsfehlerbildern Diskontinuitäten an Blockgrenzen entstehen, welche letztlich in den decodierten Inter-Frames sichtbar werden. Im Falle einer guten Bewegungskompensation oder in recht homogenen Bildbereichen ist der Prädiktionsfehler jedoch typischerweise klein und wird an vielen (auch benachbarten) Stellen zu Null quantisiert [YW98]. Daher sind in prädiktiv codierten Frames quantisierungsbedingte Blockartefakte in aller Regel weniger stark ausgeprägt als in Intra-Frames. Zum anderen können Probleme bei der Bewegungskompensation dazu führen, dass in prädiktiv codierte Frames unpassende Makroblöcke eingefügt werden, welche deutlich sichtbar sein können. Die hierdurch hervorgerufenen Blockartefakte haben allerdings nicht die Größe von Blöcken, sondern die von Makroblöcken. Probleme bei der Bewegungskompensation treten insbesondere dann auf, wenn verschiedene bewegte Objekte mit unterschiedlichen Bewegungsrichtungen aufeinander treffen. Die Veränderungen von Frame zu Frame sind dann so groß, dass zum Teil keine gut passenden Makroblöcke

als Grundlage für die Bewegungskompensation gefunden werden können. Der beschriebene Effekt wird jedoch durch die Berücksichtigung des Prädiktionsfehlers bei der Decodierung von prädiktiv codierten Frames verringert oder maskiert. Eine Maskierung tritt dann auf, wenn durch die Quantisierung des Prädiktionsfehlers neue rauschartige Artefakte entstehen, siehe hierzu [YW98].

Die Bewegungskompensation in der hybriden Videocodierung kann auch dazu führen, dass Blockartefakte, die in den als Referenz dienenden Frames vorhanden sind, in prädiktiv codierten Frames in das Innere von Blöcken verschoben werden. In diesem Zusammenhang wird gelegentlich von *false edges* gesprochen. Das beschriebene Problem wird durch die Subtraktion des Prädiktionsfehlers verringert. Je mehr Aktivität in einem Bildbereich vorhanden ist, umso mehr ist mit einer Beseitigung oder Maskierung von *false edges* zu rechnen [YW98].

Die bislang beschriebenen Effekte sind bei Verwendung von Farbsystemen, die Helligkeits- und Farbdifferenzkanäle verwenden (z. B. YCrCb), insbesondere im Helligkeitskanal kritisch, da der Helligkeitskanal in solchen Systemen der Hauptinformations-träger ist und das HVS für Informationen in den Chrominanzkanälen weniger empfindlich ist [CG02, Ric03]. Da die Chrominanzkanäle aber grundsätzlich in der gleichen Weise verarbeitet werden wie der Luminanzkanal, kann grobe Quantisierung der Chrominanz-Koeffizienten zu Artefakten führen, die denen im Luminanzkanal entsprechen. Sie machen sich durch Farbsprünge oder farblich unpassende blockartig ausgedehnte Bereiche in decodierten Bildern bemerkbar. Durch die Unterabtastung der Chrominanzkanäle sind solche Farbstörungen auf Bereiche in der Größe von Makroblöcken ausgedehnt. Auch durch die Bewegungskompensation können blockartig ausgedehnte Farbfehler entstehen, da die Bewegungsschätzung bei üblichen Codierverfahren nur in der Luminanz durchgeführt wird und die Bewegungsvektoren für die Chrominanz übernommen werden. Die Bewegungsschätzung ermittelt in ihrem Suchbereich einen Makroblock, der den geringsten Prädiktionsfehler in der Luminanz ergibt. In der Chrominanz kann der Prädiktionsfehler allerdings gleichzeitig groß sein, was zu den beschriebenen Störungen führt [YW98]. Insgesamt sind sprungartige und blockartig ausgedehnte Farbfehler aufgrund des im Vergleich zum Luminanzkanal geringeren Informationsgehalts in den Chrominanzkanälen und der geringeren Empfindlichkeit des HVS für Farbstörungen deutlich seltener zu beobachten als entsprechende Artefakte in der Luminanz.

In der Praxis treten insbesondere bei niedrigen Datenraten tendenziell mehr Blockartefakte in Intra-Frames als in Inter-Frames auf (siehe auch [LJL<sup>+</sup>03]). Dies deckt sich grundsätzlich mit den beschriebenen Ursachen und Mechanismen der Entstehung von Blocking-Störungen. Allerdings verhalten sich verschiedene Encoder diesbezüglich durchaus unterschiedlich. Entscheidend hierfür ist die im Encoder umgesetzte Strategie zum Erreichen einer vorgegeben Datenrate. Manche Encoder quantisieren die Prädiktionsfehlerbilder besonders grob, was eine bessere Qualität der Intra-Frames erlaubt, aber die Qualität der Inter-Frames verschlechtert, andere Encoder gehen entsprechend anders vor.

Der in dieser Arbeit eingesetzte MPEG-2 Encoder der Software Simulation Group [MSS] erzeugt besonders starke Blockartefakte in den Intra-Frames. In unbewegten Bildbereichen kann dies zu im Zeitverlauf pulsierenden Bildinhalten führen, da die Blockartefakte in den auf ein Intra-Frame folgenden Inter-Frames durch die Subtraktion des Codierfehlers fortlaufend verringert werden, in jedem Intra-Frame aber wieder neu entstehen [Tra01].

## DCT-Basisbilder

Ein weiteres typisches Codierartefakt bei der blockbasierten Bild- bzw. Videokompression äußert sich durch das Sichtbarwerden von bestimmten Strukturen, die sich auf die in den meisten Codierverfahren eingesetzte DCT zurückführen lassen. In Gleichung (2.1) wurde die DCT bereits eingeführt. Durch die Transformation werden Bildsignale bzw. Blöcke aus Bildsignalen in eine Darstellung überführt, in der die Bildinformationen durch Cosinusschwingungen verschiedener Frequenzen dargestellt werden. Jeder DCT-Koeffizient repräsentiert ein Produkt aus einer horizontalen und einer vertikalen Cosinusschwingung. Bei einer Transformation von  $M \cdot N$  Bildpunkten ergeben sich auch  $M \cdot N$  Koeffizienten und somit ebensoviele Produkte von Cosinusschwingungen. Diese werden oft *Basissschwingungen* oder *Basisfunktionen* der DCT genannt. Die diskreten Basisfunktionen  $\varphi_{k,l}$  sind für eine Transformationsgröße von  $M \cdot N$  Pixeln definiert durch:

$$\varphi_{k,l}(m, n) = \cos \left[ \frac{\pi (2m + 1) k}{2M} \right] \cdot \cos \left[ \frac{\pi (2n + 1) l}{2N} \right]. \quad (2.7)$$

Es gilt  $k, m \in 0 \dots M - 1$  und  $l, n \in 0 \dots N - 1$ . In Abbildung 2.11 sind für eine Blockgröße von 8x8 Bildpunkten die DCT-Basisfunktionen dargestellt. Wenn im Rahmen einer Bildkompression durch Quantisierung in einzelnen transformierten Bildblöcken neben dem DC-Koeffizienten hauptsächlich ein einziger Koeffizient einen deutlich von Null verschiedenen Wert behält, wird nach der Rücktransformation im decodierten Bild die mit dem entsprechenden Koeffizienten korrespondierende DCT-Basisfunktion sichtbar. Man spricht in diesem Zusammenhang auch von sichtbaren *DCT-Basisbildern*. Dieser Effekt ist hauptsächlich bei niedrigen Datenraten zu beobachten. Das Übrigbleiben von nur sehr wenigen DCT-Koeffizienten nach der Quantisierung, was auch bei höheren Datenraten vorkommt, führt zu vergleichbaren Effekten, die dann aus einer Überlagerung weniger DCT-Basisbilder bestehen.

Da DCT-Basisbilder in ihrer örtlichen Ausdehnung der im Codierverfahren verwendeten Blockgröße entsprechen, ist eine klare Abgrenzung von Blockartefakten nicht möglich. Es sei erwähnt, dass auch in prädiktiv codierten Bildern sichtbare DCT-Basisbilder entstehen können, wenn die Quantisierung des Prädiktionsfehlerbildes in einzelnen Blöcken zum Übrigbleiben nur eines (oder sehr weniger) AC-Koeffizienten führt.

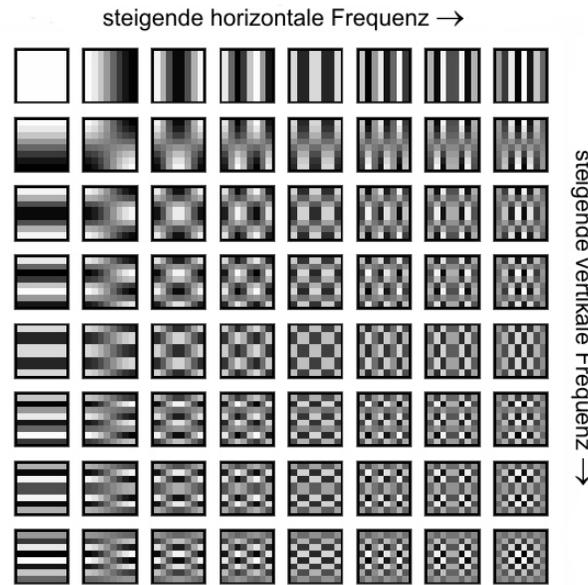


Abbildung 2.11: DCT-Basisbilder

## Staircase effect

Bei Betrachtung der in Abbildung 2.11 dargestellten DCT-Basisbilder bzw. DCT-Basisfunktionen fällt auf, dass diese im Wesentlichen horizontal oder vertikal ausgerichtete Strukturen enthalten, insbesondere die mit niederfrequenten Koeffizienten korrespondierenden Basisbilder. Somit ist zwar die Codierung horizontaler oder vertikaler Strukturen mit nur wenigen DCT-Koeffizienten möglich [MK03, Plo89], die Codierung diagonaler Strukturen ist jedoch problematisch und erfordert viele (vor allem hochfrequente) Koeffizienten [Ric03]. Da diese im Vergleich zu den niederfrequenten Koeffizienten aber in der Regel kleiner sind und dazu noch besonders grob quantisiert werden, werden sie bei niedrigen Datenraten oft ausgelöscht. Die Folge ist, dass diagonale Strukturen und Kanten hauptsächlich durch niederfrequente Basisfunktionen, die stark ausgeprägte horizontale und vertikale Komponenten enthalten, angenähert werden.

Je nach Winkel und Lage einer diagonalen Kante innerhalb eines Blocks wird diese entweder durch mehrheitlich horizontal oder mehrheitlich vertikal ausgerichtete Basisfunktionen angenähert. In ungünstigen Fällen kann dies von Block zu Block wechseln. Man beobachtet in der Praxis bei starker Kompression von Bildinhalten, dass angenäherte diagonal verlaufende Kantenabschnitte in benachbarten Blöcken oft nicht optimal zueinander passen. Es entstehen treppenartige Sprünge bzw. Diskontinuitäten an Blockgrenzen, was man als *staircase effect* bezeichnet. Dadurch, dass der staircase effect genauso wie Blockartefak-

te durch Diskontinuitäten an Blockgrenzen geprägt ist, ist eine Unterscheidung zwischen staircase effect und Blocking nur in begrenztem Maße möglich und sinnvoll.

### Ringing

Die mit der DCT verbundene Energiekonzentration auf wenige niederfrequente Koeffizienten und die üblicherweise besonders grobe Quantisierung von hochfrequenten Koeffizienten führen im Kontext der Bilddatenkompression oft zu einem Wegfall nahezu aller hochfrequenten Koeffizienten. Dadurch kommt es einerseits zu einer Auslöschung von Bilddetails, andererseits aber auch zu einer ungenauen Rekonstruktion von Kanten, die relevante Anteile in allen Frequenzbereichen enthalten.

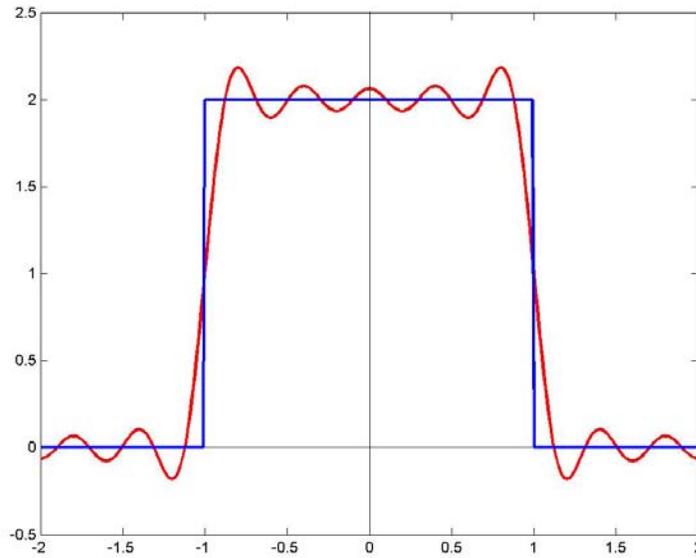
Die genannten Probleme äußern sich in decodierten Bildblöcken durch Überschwinger parallel zu Kanten, die an Blockgrenzen abrupt enden. Man bezeichnet derartige Störungen als *ringing-Artefakte*. Die Überschwinger lassen sich auf das so genannte *Gibbsche Phänomen* zurückführen, das besagt, dass bei der Rekonstruktion eines unstetigen periodischen Signals mit einer endlichen Anzahl an Basisfunktionen grundsätzlich Überschwinger mit einer Höhe von ca. 9 % der Sprunghöhe des unstetigen Signals entstehen [MH04]. Die Anzahl der zur Rekonstruktion verwendeten Basisfunktionen wirkt sich lediglich auf die Breite des ersten Überschwingers aus.

Um die Entstehung von Ringing-Artefakten zu veranschaulichen, ist in Abbildung 2.12 eine Periode der Approximation einer periodischen Rechteckfunktion, welche sich bekanntermaßen durch eine Fourierreihe darstellen lässt, mit fünf Cosinusschwingungen dargestellt. Es werden deutliche Überschwinger an den Sprungstellen des Rechtecks sichtbar. Eine vergleichbare Approximation einer Rechteckfunktion mitsamt der Überschwinger ist auch mit den Basisfunktionen der DCT realisierbar. Dies erklärt die Entstehung von Ringing-Artefakten im Rahmen einer DCT-basierten Bilddatenkompression bei der Darstellung von Kanten mit nur wenigen Koeffizienten.

In der Literatur findet man zusätzlich zum Ringing ein *Mosquito* genanntes Artefakt, das sich auf Ringing-Artefakte zurückführen lässt. Bei Mosquito handelt es sich um zeitlich variierende Ringing-Artefakte an gleichbleibender Stelle im Bild, wodurch insbesondere in Nachbarschaft zu Kanten eine rauschartige Unruhe entstehen kann, in der die typischen Charakteristika von Ringing-Artefakten (Überschwinger parallel zu Kanten) sichtbar sind.

### Blurring/Unschärfe

*Blurring* bzw. Unschärfe macht sich in Bildern durch Detailverluste und verschliffene Kanten bemerkbar. Solche Effekte sind zwar oftmals ein Resultat der begrenzten Bandbreite einzelner bildverarbeitender Komponenten, treten aber auch als Codierartefakte in der Praxis auf. Die im Rahmen von vielen Video- oder Bildkompressionsverfahren durchgeführte Quantisierung von Transformationskoeffizienten ist in aller Regel so gestaltet,



**Abbildung 2.12:** Approximation einer periodischen Rechteckfunktion mit Hilfe von fünf Cosinusschwingungen

dass insbesondere die mit hohen Ortsfrequenzanteilen korrespondierenden Koeffizienten sehr grob quantisiert (und daher oft ausgelöscht) werden. Genau diese Koeffizienten sind aber für die Darstellung von feinen Bilddetails verantwortlich [Tra01, YW98]. Bei mittleren und niedrigen Datenraten ist somit nach der Decodierung Unschärfe bzw. ein Fehlen von Details im Bildmaterial zu beobachten.

Blurring-Effekte als Resultat einer groben Quantisierung bzw. Auslöschung hochfrequenter Koeffizienten sind subjektiv einer Tiefpassfilterung von detaillierten Bildbereichen ähnlich, können jedoch örtlich durch die blockweise Verarbeitung im Rahmen der Videocodierung klar begrenzt sein. Kritisch für den beschriebenen Effekt sind besonders Bildbereiche mit hoher örtlicher Aktivität und feinen Texturen, also z. B. Rasenflächen oder hochfrequente Strukturen [YW98]. Blurring kann sich aber - wie bereits erwähnt - auch an Kanten durch deren Verschleifung bemerkbar machen.

In bidirektional prädiktiv codierten Frames (B-Frames) gibt es eine weitere Ursache für Blurring. Dort kann es im Rahmen der Bewegungskompensation zur Mittelung von Bildinhalten aus zwei Referenzframes kommen, was ebenfalls zu Detailverlusten und Verschleifungen führen kann. Durch die Subtraktion des Prädiktionsfehlers wird der Effekt jedoch verringert [YW98].

## **Abhängigkeit der Stärke von Codierartefakten von der Datenrate**

Aussagen zur Stärke von Codierartefakten bei bestimmten Datenraten sind nur mit Einschränkungen möglich, selbst wenn das im betrachteten Videoverarbeitungssystem eingesetzte Codierverfahren bekannt ist. Der Grund dafür ist, dass die Datenrate sich nicht allein auf die Stärke von Artefakten und somit auf die Bildqualität auswirkt, sondern in hohem Maße auch der Bildinhalt selbst [Tra01] und der verwendete Encoder. Je mehr (hochfrequente) Details in einem Bild vorhanden sind, umso größer ist die zur störungsfreien bzw. störungsarmen Codierung nötige Datenmenge. Andererseits können Bilddetails, die nicht oder nicht vollständig durch Quantisierungsverluste verloren gehen, auch die Sichtbarkeit von Störungen in unmittelbarer Nähe maskieren. Ein weiterer wichtiger Aspekt ist, dass sich die Charakteristik von Bildinhalten direkt auf die Art der sichtbar werdenden Störungen auswirkt. Bei blockbasierter Codierung treten Blockartefakte und sichtbare Basisbilder außer bei sehr hohen Datenraten nahezu immer auf. Blockartig ausgedehnte Farbstörungen sind hingegen hauptsächlich in Bildbereichen mit hoher Farbsättigung und nur bei geringen bis sehr geringen Datenraten zu beobachten. Ringing-Artefakte können - wie bereits erwähnt - nur dann entstehen, wenn kontrastreiche Kanten im Bild vorhanden sind. Staircase-Artefakte können ebenfalls nur im Bereich von Kanten entstehen, allerdings ist ein diagonaler Verlauf der Kanten eine weitere notwendige Bedingung. Blurring-Effekte treten - wie erwähnt - hauptsächlich dann auf, wenn im Bildmaterial Bereiche hoher örtlicher Aktivität, wie z. B. feine Texturen oder auch Kanten, vorhanden sind.

Da die praxisrelevanten Standards zur Videocodierung nur den Decoder bzw. den vom Decoder decodierbaren Datenstrom spezifizieren, kann sich auch der Encoder deutlich auf die Bildqualität bei einer vorgegebenen Datenrate auswirken. Im Encoder besteht z. B. die Möglichkeit, den I-Frames mehr Priorität bei der Verteilung der zur Verfügung stehenden Datenrate einzuräumen als den übrigen Frames. Auch eine entsprechend umgekehrte oder eine ausgewogene Priorisierung sind möglich. Die eingesetzte Priorisierung hat direkten Einfluss auf Art und Ausprägung der entstehenden Artefakte. Es existieren auch Encoder, die eine Optimierung der Verteilung der zur Verfügung stehenden Datenrate in mehreren Codierdurchläufen durchführen. Dadurch kann die Bildqualität deutlich gesteigert bzw. die Sichtbarkeit von Artefakten deutlich verringert werden. Für Echtzeitanwendungen sind solche Encoder allerdings kaum geeignet, so dass unter Annahme vergleichbarer Datenraten in Rundfunkanwendungen meist mit einer schlechteren Bildqualität gerechnet werden muss als beispielsweise in einem DVD-System. Von großem Einfluss auf die Bildqualität bzw. auf die Stärke von Störungen ist auch die Qualität des eingesetzten Bewegungsschätzers in Verbindung mit der Größe und Geschwindigkeit bewegter Bildanteile.

Trotz der Abhängigkeit der Bildqualität bzw. der Stärke von Artefakten von den genannten Faktoren ist es unstrittig, dass bei niedrigen Datenraten grundsätzlich mehr und deutlichere Artefakte auftreten als bei hohen Datenraten. Um einen Eindruck von der



**Abbildung 2.13:** Ausschnitt aus Frame 24 (Intra-Frame) der Sequenz *football*, codiert mit verschiedenen Datenraten

Abhängigkeit der Stärke von Codierartefakten von der Datenrate zu vermitteln, sind in den Abbildungen 2.13 und 2.14 exemplarisch Ausschnitte aus den Sequenzen *football* und *wheel* abgebildet, die mit verschiedenen Datenraten nach MPEG-2 komprimiert wurden.

Bei *football* ist eine klare Zunahme von Störungen mit sinkender Datenrate erkennbar. Durch die vielen Details und die schnellen Bewegungen in der Sequenz ist der Störgrad aber selbst bei 8 Mbit/s recht hoch. Man erkennt Blockartefakte, den DCT-Basisbildeffekt und Ringing-Störungen im Bereich der Werbetafeln. Bei 5 Mbit/s sind die Artefakte noch ausgeprägter und es werden deutliche Blurring-Störungen sichtbar. Bei noch niedrigerer Datenrate besteht die Rasenfläche schließlich nur noch aus homogenen Blöcken unterschiedlicher Helligkeitsstufen. Überall im Bildausschnitt finden sich Blockartefakte und im Bereich der Linien erkennt man Staircase-Artefakte.

Etwas unkritischer für eine MPEG-2 Codierung ist die Sequenz *wheel*. Sie enthält recht wenige Texturen, allerdings viele kontrastreiche Kanten. Bei 8 Mbit/s ist die Bildqualität im Vergleich zu *football* deutlich besser. Man erkennt nur einige Ansätze von Block- bzw.

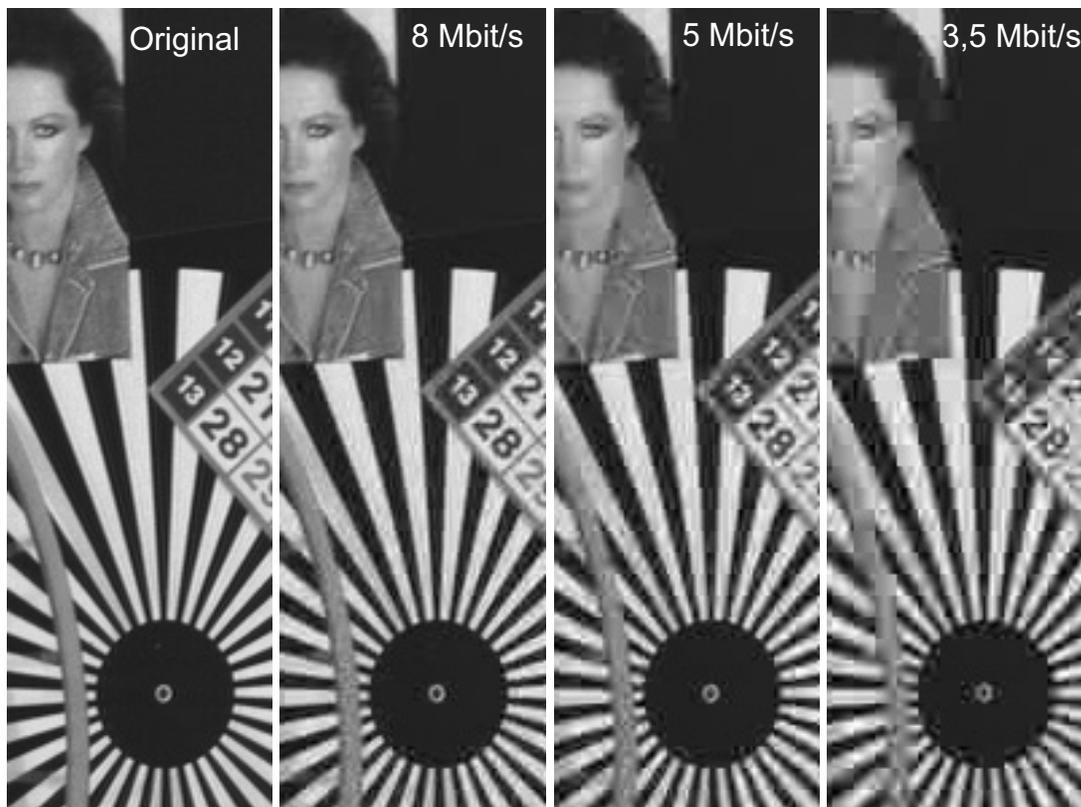


Abbildung 2.14: Ausschnitt aus Frame 24 (Intra-Frame) der Sequenz *wheel*, codiert mit verschiedenen Datenraten

Staircase-Artefakten im Bereich der Sektorscheibe und schwache Ringing-Störungen zwischen den Sektoren. Bei 5 Mbit/s sind die genannten Störungen stärker ausgeprägt und bei 3,5 Mbit/s ist der gesamte Bildausschnitt von Block- und Staircase-Artefakten gezeichnet. Außerdem sind starke Ringing-Artefakte sowie einige DCT-Basisbilder im oberen Bereich sichtbar.

Um zu demonstrieren, wie sehr sich Bewegung oder Stillstand in einer Sequenz auf die Bildqualität auswirken können, ist in Abbildung 2.15 das Ergebnis der Codierung einer Wiederholung eines einzelnen Bildes aus der Sequenz *football* - also einer Bildfolge ohne Bewegungen - bei zwei Datenraten im Vergleich zum Ergebnis der Codierung der ursprünglichen Sequenz *football* dargestellt. Man sieht, dass der Wegfall von Bewegungen zu einer deutlich besseren Bildqualität führt. Selbst bei 5 Mbit/s sind weniger Störungen sichtbar als in der ursprünglichen Sequenz bei 8 Mbit/s.

In Tabelle 2.1 sind objektive Messwerte zu den gezeigten Bildern abgedruckt. Es wurden Messungen des PSNR und eines Maßes für die Stärke von Blocking-Störungen nach



**Abbildung 2.15:** Ausschnitt aus Frame 24 der Sequenz *football24stat*, codiert mit verschiedenen Datenraten

[Pia05a] vorgenommen. Beide Maße werden in Kapitel 4 vorgestellt. Ein Absinken des PSNR deutet auf eine größer werdende Unähnlichkeit zwischen Referenz bzw. Original und Testbild hin. Es ist klar erkennbar, dass die Unähnlichkeit zwischen Original und codierten Bildern in allen Fällen mit sinkender Datenrate zunimmt. Gleichzeitig steigen die Blocking-Werte mit sinkender Datenrate an. Die objektiven Messwerte bestätigen also die bereits geschilderten subjektiven Eindrücke.

Um einen Eindruck zu gewinnen, mit welchen Datenraten im realen Fernschrundfunkbetrieb gerechnet werden kann, wurden im August und September 2007 25 Datenströme von jeweils ca. 10 - 15 Sekunden Länge im DVB-S Rundfunk und 20 vergleichbare Datenströme im DVB-T Rundfunk aufgezeichnet und analysiert. Es wurden jeweils mehrere öffentlich-rechtliche Sender, bekannte Privatsender, Lokalsender (nur DVB-S) sowie einige Spartenprogramme berücksichtigt. Datenraten wurden jeweils sowohl für gesamte Streams als auch für kurze Abschnitte in der Länge einer Group of Pictures (GOP) berechnet. Im DVB-S Rundfunk konnte eine maximale kurzfristige Datenrate von mehr als 8 Mbit/s und

**Tabelle 2.1:** Blocking- und PSNR-Messwerte bei unterschiedlichen Datenraten

Sequenz/Bild	Blocking	PSNR
football	1,02	-
football, 8 Mbit/s	1,49	33,05
football, 5 Mbit/s	1,80	30,83
football, 3,5 Mbit/s	2,41	28,01
wheel	1,01	-
wheel, 8 Mbit/s	1,36	33,00
wheel, 5 Mbit/s	1,69	29,35
wheel, 3,5 Mbit/s	2,35	26,27
football (statisch)	1,02	-
football (statisch), 8 Mbit/s	1,19	37,62
football (statisch), 5 Mbit/s	1,33	35,02

eine durchschnittliche Datenrate von mehr als 7 Mbit/s für einen ganzen Stream ermittelt werden. Im DVB-T Rundfunk waren die entsprechenden Datenraten mit ca. 6,5 Mbit/s und ca. 4,3 Mbit/s deutlich kleiner. Auch insgesamt waren die Datenraten im DVB-T Rundfunk erwartungsgemäß kleiner als im DVB-S Rundfunk. Datenraten von weniger als 2,5 Mbit/s waren in den aufgezeichneten DVB-T Streams außer bei den Hauptprogrammen des öffentlich-rechtlichen Rundfunks und bei den großen Privatsendern nicht unüblich und wurden bei kleinen und wenig bekannten Sendern auch im DVB-S Rundfunk beobachtet. Die kleinsten gemessenen kurzfristigen Datenraten waren mit ca. 1 Mbit/s bei DVB-S und DVB-T ähnlich niedrig.

Die angegebenen Zahlen können und sollen ausschließlich als Orientierungshilfe dienen, da sie stark vom Aufnahmezeitpunkt bzw. Bildinhalt und der Auswahl der Fernsehsender abhängen. Sie lassen jedoch keinen Zweifel daran, dass in DVB-Systemen zeitweise Datenraten eingesetzt werden, die insbesondere bei komplexem Bildinhalt und viel Bewegung deutliche Codierartefakte erwarten lassen, was durch Praxiserfahrungen bestätigt wird. Es sei angemerkt, dass die in dieser Arbeit z. B. in Form von Beispielen vermittelten Qualitätseindrücke bei bestimmten Datenraten nicht ohne Weiteres auf die in DVB-Systemen zu erwartenden Qualitäten bei gleichen Datenraten übertragen werden können, da das Testmaterial in dieser Arbeit grundsätzlich eine Auflösung von 720x576 Bildpunkten hatte und progressiv (zumeist bei einer Framerate von 60 Hz) codiert wurde, im SD DVB-Rundfunk aber Interlaced-Codierung und eine z. T. verringerte horizontale Auflösung eingesetzt werden. Wie bereits erwähnt, wirken sich auch die jeweils eingesetzten Encoder erheblich auf die Bildqualität aus.

# 3 Grundlagen der Regressionsanalyse

In Kapitel 4 sowie insbesondere in Kapitel 8 werden Qualitätsbewertungsverfahren vorgestellt, die die Prädiktion von absoluter Bildqualität oder einer Bildqualitätsveränderung durch Anwendung von Bildverarbeitungsverfahren auf Basis mehrerer Teilmaße, welche bestimmte spezielle Qualitätsmerkmale erfassen, ermöglichen. Mit Hilfe der Teilmaße sind direkt differenzierte Analysen der Qualitätsmerkmale bzw. der Auswirkungen von Bildverarbeitungsverfahren auf die Qualitätsmerkmale möglich. Im Hinblick auf eine einfache Durchführbarkeit und Vergleichbarkeit von Qualitätsbewertungen ist jedoch in der Praxis häufig die Abbildung mehrerer solcher Maße auf ein einziges, möglichst aussagekräftiges Gesamtmaß oder sehr wenige Maße, die bestimmte Merkmalsklassen zusammenfassen, wünschenswert. Eine Kombination von Maßen kann auch sinnvoll sein, um eine Qualitätsbewertung mit besonders hoher Prädiktionsgenauigkeit zu erhalten oder um ein Qualitätsbewertungsverfahren auf einem breiten Spektrum von Eingangsdaten anwendbar zu machen.

Die Statistik bietet als eine Lösung von derartigen Abbildungs- bzw. Kombinationsproblemen die Regressionsanalyse an, die sich mit der Modellierung des Zusammenhangs zwischen einer Untersuchungsvariable oder Zielgröße  $Y$  und einer oder mehreren erklärenden Variablen oder Einflussgrößen  $X$  bzw.  $X_j$ , mit  $j = 1 \dots M$ , beschäftigt [Tie87].  $Y$  wird auch Regressand und die  $X_j$  werden Regressoren genannt. Bevor eine Regressionsanalyse durchgeführt wird, sollten unbedingt umfangreiche theoretische Vorüberlegungen, z. B. über die Art des Zusammenhangs, angestellt werden [UM06]. Ziel ist die Aufstellung eines theoretischen Modells, das mit Hilfe der Regressionsanalyse getestet oder konkretisiert bzw. in ein statistisches Modell übersetzt werden kann. Zwar kann die Regressionsanalyse auch „blind“ durchgeführt werden, d. h. es kann ein mathematischer Zusammenhang zwischen beliebigen Größen hergestellt werden, jedoch ist es sehr fraglich, ob die dabei erzielten Ergebnisse sinnvoll sind und sich letztlich theoretisch erklären lassen [UM06]. Gerade die an späterer Stelle beschriebene schrittweise Regression ist diesbezüglich mit Gefahren verbunden. Im Folgenden werden für diese Arbeit relevante Aspekte der Regressionsanalyse betrachtet.

## 3.1 Einfache lineare Regression

Ausgegangen wird zunächst vom Fall des einfachen linearen Regressionsmodells. Es dient der Beschreibung einer linearen Abhängigkeit zwischen einer Variable und einer einzigen

anderen Variable. Das einfache lineare Regressionsmodell lässt sich formal durch folgende Modellgleichung beschreiben [Sch03]:

$$Y_i = \beta_0 + \beta_1 \cdot x_i + U_i, \quad \text{mit } i = 1 \dots N. \quad (3.1)$$

$x_i$  ist einer von  $N$  bestimmten Werten der Einflussgröße  $X$  und  $Y_i$  sind entsprechende potenzielle Beobachtungen der Zielgröße  $Y$ . Die Regressionskoeffizienten  $\beta_0$  und  $\beta_1$  beschreiben den deterministischen Teil des Zusammenhangs zwischen Einfluss- und Zielgröße.  $U_i$  sind Störvariablen, die den stochastischen Teil des Zusammenhangs darstellen. Dadurch wird der Gesamteffekt nicht betrachteter Einflussgrößen berücksichtigt. Der Wert von  $U_i$  lässt sich für ein vorgegebenes  $x_i$  nicht bestimmen. Daher können die  $U_i$  als Zufallsvariablen betrachtet werden, was folglich auch für die  $Y_i$  gilt.

Das einfache lineare Regressionsmodell soll nun im Zusammenhang mit einer realen Problemstellung konkretisiert werden. Beschrieben werden soll z. B. die Änderung des Stroms durch einen Widerstand bei Änderung der angelegten Spannung, ausgehend von einem Versuch, in dem der Strom bei verschiedenen angelegten Spannungen gemessen wird. Es wird von einem linearen Zusammenhang ausgegangen. Resultat des Versuchs sind Wertepaare  $(x_i, y_i)$ , wobei die  $x_i$  den angelegten Spannungen und die  $y_i$  den entsprechenden gemessenen Strömen entsprechen. Die  $y_i$  sind Stichprobenrealisationen der Variablen  $Y_i$  im linearen Regressionsmodell. Das Problem entspricht der Suche einer Ausgleichsgeraden durch eine Punktwolke aus Punkten, die durch die Wertepaare  $(x_i, y_i)$  bestimmt sind. Formal liegt folgende Situation vor:

$$y_i = \beta_0 + \beta_1 \cdot x_i + u_i, \quad \text{mit } i = 1 \dots N. \quad (3.2)$$

Hinter den  $u_i$  als Realisationen der Zufallsvariablen  $U_i$  verbergen sich in diesem Beispiel Messungenauigkeiten und möglicherweise nicht berücksichtigte Einflüsse, z. B. Temperaturänderungen am Widerstand während des Versuchs. Gesucht ist nun eine Schätzung für den deterministischen Teil des Zusammenhangs zwischen Zielgröße (Strom) und Einflussgröße (Spannung), also eine Schätzung  $b_0, b_1$  der Regressionskoeffizienten  $\beta_0, \beta_1$  der Grundgesamtheit.  $b_1$  stellt hier eine Schätzung des Kehrwertes des Widerstands, also des Leitwertes dar. Auf Basis der Parameter  $b_0, b_1$  lässt sich eine Schätzung  $\hat{y}_i$  des deterministischen Teils der Zielgröße in Abhängigkeit von der Einflussgröße angeben:

$$\hat{y}_i = b_0 + b_1 \cdot x_i. \quad (3.3)$$

Zur Bestimmung bzw. Schätzung der Parameter  $b_0, b_1$  wird üblicherweise die Methode der kleinsten Quadrate angewendet, die die Summe der quadrierten Abweichungen zwischen den Messwerten  $y_i$  und den Schätzwerten  $\hat{y}_i$  minimiert [Kle03, Tie87].

$$\sum_{i=1}^N (y_i - \hat{y}_i)^2 \rightarrow \text{Minimum}. \quad (3.4)$$

Der Schätzer besitzt die so genannte *BLUE*-Eigenschaft (Best Linear Unbiased Estimator, „bester linearer unverzerrter Schätzer“). Das bedeutet, dass der Schätzer erwartungstreu ist, d. h. es gilt  $E\{b_i\} = \beta_i$ , und dass er sich durch minimale Varianz unter den erwartungstreuen Schätzern auszeichnet [Tie87]. Durch partielle Differenziation der Summe (3.4) nach  $b_0$  und  $b_1$  und Nullsetzen erhält man zwei Gleichungen mit zwei Unbekannten. Die Lösung der Gleichungen mit  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$  und  $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$  lautet:

$$b_0 = \bar{y} - b \cdot \bar{x} \quad \text{und} \quad (3.5)$$

$$b_1 = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \frac{\text{cov}(x, y)}{\text{var}(x)}. \quad (3.6)$$

## 3.2 Multiple lineare Regression

Oft - so auch in den in Kapitel 8 dieser Arbeit betrachteten Aufgabenstellungen - reicht das einfache lineare Regressionsmodell nicht aus, da eine Abhängigkeit einer Zielgröße  $Y$  von mehreren Einflussgrößen  $X_j$ , mit  $j = 1 \dots M$  beschrieben werden muss. Es ist der Übergang vom Modell der einfachen linearen Regression auf das Modell der multiplen linearen Regression notwendig. Dabei muss auf die Erfüllung einer Grundvoraussetzung geachtet werden: die mindestens näherungsweise lineare Abhängigkeit der Zielgröße von den Einflussgrößen in dem relevanten bzw. betrachteten Wertebereich der Einflussgrößen. Dieser Wertebereich wird z. B. im Fall von Bildqualitätsmaßen, die Codierstörungen erfassen, durch die verwendeten Testbilder und Kompressionsraten bestimmt. Gleichung (3.1) wird zur Beschreibung einer linearen Abhängigkeit der Zielgröße von  $j$  Einflussgrößen folgendermaßen erweitert:

$$Y_i = \beta_0 + \sum_{j=1}^M \beta_j x_{ji} + U_i. \quad (3.7)$$

Eine Realisation des Modells lässt sich beschreiben durch

$$y_i = \beta_0 + \sum_{j=1}^M \beta_j x_{ji} + u_i. \quad (3.8)$$

Zur Schätzung der Modellparameter bzw. Regressionskoeffizienten  $b_0$  bis  $b_M$  kann wiederum die Methode der kleinsten Quadrate angewendet werden [Tie87]. Es gilt nun für die Schätzung des deterministischen Teils der Zielgröße

$$\hat{y}_i = b_0 + \sum_{j=1}^M b_j \cdot x_{ji} \quad (3.9)$$

oder nach Überführung in eine Matrix-Darstellung

$$\hat{\underline{y}} = \underline{X} \underline{b}. \quad (3.10)$$

Es sind

$$\hat{\underline{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{pmatrix}, \quad \underline{X} = \begin{pmatrix} x_{01} & x_{11} & \dots & x_{M1} \\ x_{02} & x_{12} & \dots & x_{M2} \\ \vdots & \vdots & & \vdots \\ x_{0N} & x_{1N} & \dots & x_{MN} \end{pmatrix}, \quad \underline{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_M \end{pmatrix}. \quad (3.11)$$

Für alle Werte in der ersten Spalte von  $\underline{X}$  gilt im Allgemeinen  $x_{0i} = 1$ . Wenn erzwungen werden soll, dass die Regression durch einen bestimmten Punkt geht, was der Einführung einer Nebenbedingung entspricht, müssen Anpassungen vorgenommen werden. Oft ist es bekannt oder sinnvoll, dass die Regression durch den Ursprung verlaufen soll. Das bedeutet, dass für den Fall, dass alle Einflussgrößen Null sind, auch die Zielergröße Null ist. Die erste Spalte von  $\underline{X}$  wird dann gestrichen und der Koeffizient  $b_0$  existiert nicht bzw. ist Null.

Eine Schätzung der Regressionskoeffizienten erhält man durch Lösung der Normalengleichung [Har02, GI94]

$$\underline{X}^T \underline{X} \underline{b} = \underline{X}^T \underline{y}, \quad \text{mit } \underline{y} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{pmatrix}. \quad (3.12)$$

$\underline{X}^T$  ist die Transponierte von  $\underline{X}$ . Wenn sichergestellt ist, dass die Matrix  $\underline{X}$  den Rang  $M+1$  hat, also die Anzahl der Wertekombinationen  $(x_{1i}, x_{2i}, \dots, x_{Mi}, y_i)$  größer oder gleich der Zahl der Einflussgrößen plus Eins ist [Tie87], existiert eine Lösung und lautet:

$$\underline{b} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y}. \quad (3.13)$$

Die Durchführung einer multiplen linearen Regression auf Basis von Wertekombinationen  $(x_{1i}, x_{2i}, \dots, x_{Mi}, y_i)$ , die durch Experimente - oder wie im Kontext dieser Arbeit durch Messungen und subjektive Tests - ermittelt wurden, lässt sich mit gängiger Statistik-Software durchführen, weshalb auf Details der Berechnung verzichtet wird. Die Zielgröße  $y_i$  entspricht in den ab Kapitel 7 beschriebenen Untersuchungen der gemittelten subjektiven Bewertung eines Bildes bzw. Bildpaares  $i$  durch die Probanden eines subjektiven Tests, während die  $x_{ji}$  die Werte von Qualitätsmaßen  $X_j$ , mit  $j = 1 \dots M$  sind, die für das  $i$ 'te Bildpaar gemessen werden müssen. Ziel ist es, durch Durchführung der Regression Koeffizienten  $b_0$  bis  $b_M$  zu erhalten, die eine möglichst genaue Schätzung  $\hat{y}_i$  der subjektiven Bewertung eines Bildpaares ermöglichen.

Nun sollen einige Konfidenzintervalle angegeben werden. Die  $1 - \alpha$  Konfidenzintervalle für die Regressionskoeffizienten  $\beta_j$  lauten nach [Har02, Tie87]:

$$\left[ b_j - \sqrt{\sigma^2 c_{jj}} \cdot t_{N-M-1;1-\alpha/2} \ ; \ b_j + \sqrt{\sigma^2 c_{jj}} \cdot t_{N-M-1;1-\alpha/2} \right]. \quad (3.14)$$

Sie geben an, in welchem Bereich die Koeffizienten  $\beta_j$  mit einem Vertrauen von  $1 - \alpha$  liegen. Die Fehlervarianz  $\sigma^2$  kann auf Basis der Streuung der Residuen geschätzt werden [Har02], welche folgendermaßen definiert sind:

$$d_i = y_i - \hat{y}_i. \quad (3.15)$$

Die Zahl der Freiheitsgrade bei der multiplen Regression mit  $M$  Einflussgrößen und  $N$  Stichproben beträgt  $N - M - 1$  und somit berechnet sich die Schätzung der Varianz zu

$$\hat{\sigma}^2 = \frac{1}{N - M - 1} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (3.16)$$

Wenn eine Regression durch den Ursprung erzwungen wird, muss beachtet werden, dass ein Freiheitsgrad mehr existiert, da kein Achsenabschnitt geschätzt wird. Die Summe muss dann also durch  $N - M$  dividiert werden. Die  $c_{jj}$  in Gleichung (3.14) sind die Hauptdiagonalelemente der Matrix  $\underline{C} = (\underline{X}^T \underline{X})^{-1}$  und die  $t_{N-M-1;1-\alpha/2}$  sind Quantile der Studentischen t-Verteilung (siehe z. B. [Har02]). In den meisten Fällen wird  $\alpha = 0,05$  gewählt. Das  $1 - \alpha$  Konfidenzintervall für  $E\{Y_i\}$ , also für die deterministische Komponente der Zielgröße in der Grundgesamtheit, lautet mit gegebenem  $\underline{x}_i^T = (1, x_{1i}, x_{2i}, \dots, x_{Mi})$  nach [Tie87]:

$$\left[ \hat{y}_i - \hat{\sigma} \sqrt{\underline{x}_i^T \underline{C} \underline{x}_i} \cdot t_{N-M-1;1-\alpha/2} \ ; \ \hat{y}_i + \hat{\sigma} \sqrt{\underline{x}_i^T \underline{C} \underline{x}_i} \cdot t_{N-M-1;1-\alpha/2} \right]. \quad (3.17)$$

Das Intervall gibt an, in welchem Bereich die Zielgröße in der Grundgesamtheit  $Y_i$  für bestimmte Einflussgrößen  $x_{ji}$  mit einem Vertrauen von  $1 - \alpha$  liegt. Das  $1 - \alpha$  Konfidenzintervall für einen individuellen Wert  $y_i$  bei gegebenen  $x_{ji}$  lautet nach [Tie87, EH04]:

$$\left[ \hat{y}_i - \hat{\sigma} \sqrt{1 + \underline{x}_i^T \underline{C} \underline{x}_i} \cdot t_{N-M-1;1-\alpha/2} \ ; \ \hat{y}_i + \hat{\sigma} \sqrt{1 + \underline{x}_i^T \underline{C} \underline{x}_i} \cdot t_{N-M-1;1-\alpha/2} \right]. \quad (3.18)$$

Das Intervall gibt für eine bestimmte Kombination der Einflussgrößen  $x_{ji}$  an, in welchem Bereich eine zukünftige Realisation der Zufallsvariablen bzw. Zielgröße  $y_i$  mit einem Vertrauen von  $1 - \alpha$  liegen wird. Hier wird also sowohl die deterministische als auch die stochastische Komponente einer Realisation  $y_i$  betrachtet.

Während einer Regressionsanalyse sollte geprüft werden, ob das Regressionsmodell konsistent mit den Daten ist. Bei der linearen Regression bedeutet dies speziell, dass geprüft werden muss, ob eine mindestens in guter Näherung lineare Abhängigkeit der Zielgröße von den Einflussgrößen in den betrachteten Wertebereichen der Einflussgrößen besteht.

Dazu können Diagramme, in denen die (gemessenen) Werte der Zielgröße  $y_i$  gegen die Werte der einzelnen Einflussgrößen  $x_{ji}$  aufgetragen werden, analysiert werden. Bei einem linearen Zusammenhang sollten die Punkte im Diagramm um eine gedachte Gerade angeordnet sein. Deutlicher sichtbar werden Verletzungen der Annahme von Linearität oft, indem für jede Einflussgröße  $x_{ji}$  ein Diagramm betrachtet wird, in dem die Residuen  $d_i$  über den Werten der entsprechenden Einflussgröße  $x_{ji}$  aufgetragen sind. Solange in den Diagrammen keine regelmäßigen Muster um den Mittelwert der Residuen erkennbar sind, kann von linearen Zusammenhängen ausgegangen werden [Kle03, GI94]. Die Verteilung der Werte in Richtung der Abszisse ist nicht von Bedeutung, da diese davon abhängt, welche Einflussgrößenwerte untersucht wurden oder in einem ungeplanten Versuch auftraten [Kle03]. Laut [GI94] deutet ein nicht erkennbares Muster in einem Residuendiagramm, in dem die Residuen gegen die geschätzte Zielgröße aufgetragen werden, ebenfalls auf die Konsistenz der Daten mit einem linearen Modell hin.

Bei nichtlinearen Zusammenhängen zwischen Zielgröße und Einflussgrößen können nichtlineare Regressionsmodelle verwendet werden [FKPT04, Har02, GI94]. Nichtlineare Regressionsanalysen sind mathematisch aufwändig und erfordern oft die Anwendung iterativer Methoden [FKPT04]. Des Weiteren sind die Ergebnisse sehr vom verwendeten Ansatz abhängig und daher schwer vergleichbar [Fec00]. Insgesamt ist die Theorie der nichtlinearen Regression im Vergleich zu der der linearen Regression wesentlich weniger weit entwickelt [GI94].

In vielen Fällen ist es möglich, auch wenn der Verdacht besteht, dass kein linearer Zusammenhang zwischen einer Einfluss- und der Zielgröße besteht, dennoch ein lineares Regressionsmodell anzuwenden. Dazu muss die entsprechende Einflussgröße geeignet transformiert werden [Kle03, Har02, Tie87]. Wenn beispielsweise der Verdacht einer quadratischen Abhängigkeit in der Form  $Y = \beta_0 + \beta_1 \cdot T^2$  besteht, so kann durch die Transformation  $X = T^2$  der lineare Zusammenhang  $Y = \beta_0 + \beta_1 \cdot X$  hergestellt werden. Auch wurzelförmige, exponentielle, logarithmische und weitere Zusammenhänge lassen sich geeignet transformieren. Es empfiehlt sich aus den oben genannten Gründen zunächst immer, geeignete Variablentransformationen und ein lineares Regressionsmodell in Betracht zu ziehen, bevor auf ein nichtlineares Regressionsmodell zurückgegriffen wird.

Die Güte der Anpassung eines linearen Regressionsmodells an die Versuchsdaten kann mit Hilfe des (multiplen) Bestimmtheitsmaßes  $B$  bewertet werden. Es gilt

$$B = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} = \frac{\text{var}(\hat{y})}{\text{var}(y)} = 1 - \frac{\text{var}(d)}{\text{var}(y)} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}. \quad (3.19)$$

Also ist das Bestimmtheitsmaß das Verhältnis der Varianz der geschätzten Zielgröße zur Varianz der beobachteten Zielgröße. Unter Beachtung von Gleichung (3.9) kann man auch sagen, dass das Bestimmtheitsmaß angibt, welcher Anteil der Varianz der Zielgröße durch die Varianz der Regressoren erklärt werden kann [Har02]. Das Bestimmtheitsmaß kann nur

Werte zwischen Null und Eins annehmen. Je größer es ist, umso besser ist die Anpassung des Regressionsmodells.

Im Falle der linearen Regression unter Verwendung der Methode der kleinsten Quadrate entspricht das Bestimmtheitsmaß  $B$  dem Quadrat des so genannten (multiplen) Korrelationskoeffizienten  $r$ , welcher die Stärke des linearen Zusammenhangs zwischen der Zielgröße  $Y$  und den Regressoren  $X_j$  angibt [EH04, Sch03]. Er ist genauso groß wie die Pearsonsche Korrelation zwischen geschätzter Zielgröße und beobachteter Zielgröße [Tie87]. Es gilt

$$B = r^2 = \frac{\text{cov}^2(y, \hat{y})}{\text{var}(y) \cdot \text{var}(\hat{y})}. \quad (3.20)$$

Erwähnung finden soll an dieser Stelle noch der Overall-F-Test. Es ist ein statistischer Test, der eine Aussage darüber macht, ob wenigstens eine Einflussgröße in einem multiplen linearen Regressionsmodell einen signifikanten Erklärungsbeitrag zur Zielgröße liefert [FKPT04]. Die Nullhypothese des Tests besagt, dass keine der Einflussgrößen einen Beitrag zur Erklärung der Zielgröße liefert. Die Alternativhypothese besagt, dass dies für mindestens eine Einflussgröße nicht gilt:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_M = 0, \quad H_A : \beta_j \neq 0 \text{ für mindestens ein } j. \quad (3.21)$$

Die Nullhypothese ist dann zugunsten der Alternativhypothese zu verwerfen, wenn bei gegebenem Vertrauensniveau  $1 - \alpha$  (typischerweise  $\alpha = 0,05$ ) gilt:

$$\frac{B}{1 - B} \cdot \frac{N - M - 1}{M} > F_{M, N - M - 1; 1 - \alpha} \quad (3.22)$$

Die  $F_{M, N - M - 1; 1 - \alpha}$  sind die Quantile der F-Verteilung. Siehe dazu z. B. [Har02].

Zur Beantwortung der Frage, ob alle Einflussgrößen einen signifikanten Beitrag zur Erklärung der Zielgröße liefern, existiert ebenfalls ein statistischer Test, der im Abschnitt zur schrittweisen multiplen Regression vorgestellt wird.

### 3.3 Voraussetzungen der linearen Regression

Damit ein Regressionsschätzer die BLUE-Eigenschaft (s. o.) besitzt und keine grundsätzlichen Schwierigkeiten bei der Analyse eines Regressionsmodells zu erwarten sind, müssen vier Voraussetzungen bezüglich der Störvariablen erfüllt sein [Tie87, Sch03, Kle03]. Die Voraussetzungen werden im folgenden Abschnitt vorgestellt und es wird darauf eingegangen, wie die Erfüllung einzelner Voraussetzungen geprüft werden kann. Darüber hinaus wird angegeben, welche Konsequenzen die Nichterfüllung einzelner Voraussetzungen hat.

1. Die erste Voraussetzung betrifft den Erwartungswert der Störvariablen, der gleich Null sein muss:

$$E\{U_i\} = 0, \quad \text{für alle } i = 1 \dots N. \quad (3.23)$$

Dieser Erwartungswert bezieht sich auf die Verteilung der Störvariablen bzw. Residuen in der Grundgesamtheit, welche im Gegensatz zu den Residuen der betrachteten Stichprobe nicht bekannt sind. Deshalb kann nicht geprüft werden, ob die Voraussetzung erfüllt ist oder nicht. Es sei erwähnt, dass der Mittelwert der Residuen der Stichprobe bei Anwendung der Methode der kleinsten Quadrate automatisch Null ist [UM06, Tie87].

Sollte die Voraussetzung nicht erfüllt sein, ergeben sich keinerlei kritische Probleme, da nur mit einer verzerrten Schätzung des Regressionskoeffizienten  $b_0$  (auch Intercept genannt) gerechnet werden muss. Dieser Koeffizient ist aber in aller Regel von geringem theoretischen oder analytischen Interesse [UM06], weshalb weitere Betrachtungen nicht notwendig sind.

2. Die zweite Voraussetzung fordert, dass die Varianz der Störvariablen konstant - das heißt unabhängig von den Einflussgrößen - sein muss:

$$\text{var}(U_i) = \sigma^2, \quad \text{für alle } i = 1 \dots N. \quad (3.24)$$

Man spricht auch von Homoskedastizität. Wird die Voraussetzung nicht erfüllt, können nach wie vor mit Hilfe der Methode der kleinsten Quadrate erwartungstreue Regressionskoeffizienten geschätzt werden, die Aussagekraft von Konfidenzintervallen und Signifikanztests wird jedoch deutlich eingeschränkt. Homoskedastizität kann am einfachsten durch visuelle Analyse der Residuen, aufgetragen über den Schätzungen der Zielgröße  $\hat{y}_i$ , überprüft werden. Es kann von Homoskedastizität ausgegangen werden, solange die einzelnen Punkte zufällig angeordnet sind und im weitesten Sinne ein waagrechtes Band bilden [FKPT04, Kle03, Har02]. Dagegen spricht man von Heteroskedastizität, wenn ein regelmäßiges Muster erkennbar wird. Ist Heteroskedastizität nicht zu tolerieren, so kann die Problematik mit Hilfe einer gewichteten kleinsten Quadrate Schätzung bzw. der Aitken-Schätzung angegangen werden [Har02, GI94, Tie87].

3. Die dritte Voraussetzung fordert, dass die Störvariablen unkorreliert sein müssen:

$$\text{cov}(U_i, U_j) = 0, \quad \text{für } i \neq j \quad \text{und } i, j = 1 \dots N. \quad (3.25)$$

Auch wenn diese Voraussetzung nicht erfüllt ist, lassen sich weiterhin erwartungstreue Regressionskoeffizienten mit Hilfe der Methode der kleinsten Quadrate

schätzen. Es besteht jedoch die Gefahr der Berechnung fehlerhafter Konfidenzintervalle und Signifikanztests. Eine übliche Methode zur Überprüfung der Unkorreliertheit der Störvariablen ist der Durbin-Watson-Test [Har02, SW77]. Dieser basiert auf der Analyse der Differenzen aufeinander folgender Residuen. Die auf den Differenzen aufbauende Teststatistik muss für Unkorreliertheit innerhalb eines Intervalls liegen, dessen Größe von der Anzahl der Regressoren und von der Anzahl der Stichproben abhängt. Bei Regressionen durch den Ursprung müssen beim Durbin-Watson-Test Besonderheiten beachtet werden, zu denen sich Details in [Kra71, Kin81] finden. Des Weiteren kann zur Überprüfung der Unkorreliertheit der Störvariablen eine grafische Analyse der Residuen, die dazu in der Reihenfolge der Versuchsdurchführung aufzutragen sind, durchgeführt werden [UM06, Kle03]. Es sollte ein waagerechtes Band ohne besondere Auffälligkeiten wie z. B. eine Zu- oder Abnahme der Größe der Residuen im Versuchsverlauf erkennbar sein. Eine weitere Möglichkeit zur Aufdeckung korrelierter Störvariablen besteht in der Durchführung einer Regression der in Versuchsreihenfolge angeordneten Residuen auf die um einen Wert verschobene Folge der Residuen [UM06]. Wenn sich ein signifikanter Regressionskoeffizient ergibt, muss von korrelierten Störvariablen ausgegangen werden.

4. Die vierte Voraussetzung besagt, dass die Störvariablen einer Normalverteilung folgen müssen. Sie steht nicht im Zusammenhang mit der BLUE-Eigenschaft der geschätzten Regressionskoeffizienten. Ist die Voraussetzung nicht erfüllt, ist jedoch von Einschränkungen bei der Interpretation von statistischen Tests und Intervallschätzungen auszugehen. Die Ergebnisse solcher Untersuchungen gelten insbesondere bei genügend großem Stichprobenumfang aber immer noch approximativ [Sch03, Tie87] bzw. können als hinreichend genau angesehen werden, wenn die Normalverteilungsannahme approximativ gilt [FKPT04]. In [Tie87] und in [Sch03] wird eine Verletzung der Normalverteilungsannahme sogar in der Praxis für wenig bedeutsam und damit für tolerabel erachtet. Soll dennoch eine Überprüfung vorgenommen werden, empfiehlt sich laut [Kle03] eine grafische Analyse der Residuen im Wahrscheinlichkeitsnetz. In einer solchen Darstellung liegen alle Punkte näherungsweise auf einer Geraden, wenn die Annahme einer Normalverteilung gerechtfertigt ist. Anderenfalls werden deutliche Krümmungen sichtbar. Sehr ähnlich läuft die Überprüfung mit Hilfe eines Q-Q-Plots ab [Har02]. Eine rechnerische Überprüfung kann mit Hilfe des  $\chi^2$ -Anpassungstests oder den auch für kleine Stichprobenumfänge geeigneten Kolmogoroff-Smirnov- oder Lilliefors-Tests durchgeführt werden [Har02]. Eine Möglichkeit, die Verletzung der Normalverteilungsannahme zu beseitigen, besteht darin, Werte aus dem Datensatz zu eliminieren, die besonders große Residuen hervorrufen, und die Regressionsanalyse auf dem bereinigten Datensatz durchzuführen [Sch03]. Selbstverständlich sollten nur Werte entfernt werden, wenn sich dies begründen lässt.

### 3.4 Auswahl von Regressoren und Multikollinearität

Beim Vorliegen mehrerer Regressoren stellt sich die Frage, ob alle Regressoren notwendig sind, um ein hinreichendes Modell zu erstellen. So ist es möglich, dass einzelne Regressoren zur Erklärung der Zielgröße überflüssig sind [Har02]. Dies trifft insbesondere dann zu, wenn Abhängigkeiten zwischen den Regressoren bestehen, was in der Praxis sehr oft vorkommt und auch nicht ohne Weiteres vermieden werden kann. Man spricht dann von stochastischer Multikollinearität. Zwar ist die BLUE-Eigenschaft der Regressionsschätzung auch bei Multikollinearität gegeben [UM06], jedoch können die Koeffizienten mit großen Streuungen behaftet sein. Die Konsequenzen daraus sind, dass einzelne Schätzwerte der Zielgröße recht weit vom entsprechenden Wert der Grundgesamtheit entfernt sein können, und dass die Aussagekraft von Konfidenzintervallen und statistischen Tests abnimmt [Tie87]. Darüber hinaus können bei starker Multikollinearität bereits kleine Änderungen an der Datenbasis zu deutlichen Änderungen der Regressionskoeffizienten führen. Somit besteht durch Weglassen einzelner Werte der Datenbasis und anschließende Berechnung neuer Koeffizienten auch eine Möglichkeit zu überprüfen, ob ein Verdacht auf Multikollinearität besteht (Stabilitätstest). Ein Beweis ist dies jedoch nicht. Ebenso belegen hohe Korrelationen zwischen einzelnen Einflussgrößen allein keine Multikollinearität. Eine recht gute Möglichkeit zur Aufdeckung von Multikollinearität besteht in der Durchführung von mehreren multiplen Regressionen unter ausschließlicher Verwendung der Einflussgrößen. Eine der Einflussgrößen wird jeweils als Zielgröße behandelt. Wenn die Differenz  $1 - B = 1 - r^2$  kleiner als 0,10 ist, kann von hoher Multikollinearität ausgegangen werden. Da der Nachweis von Multikollinearität grundsätzlich mit Schwierigkeiten verbunden ist, empfiehlt sich immer die Durchführung mehrerer Tests. Für weitere Details sei auf [UM06, Tie87] verwiesen.

Die einfachste Methode zur Beseitigung von Multikollinearität ist der Ausschluss von Einflussgrößen aus dem Regressionsmodell, die maßgeblich für die Multikollinearität verantwortlich sind. Dabei ergibt sich jedoch ein vollkommen neues Regressionsmodell, das mit dem zugehörigen theoretischen Modell oft nicht mehr in Einklang gebracht werden kann. Insofern scheidet diese Vorgehensweise oft aus. Des Weiteren kann zur Beseitigung von Multikollinearität die Faktorenanalyse angewendet werden [Tie87]. Dabei werden voneinander abhängige Variablen auf eine kleinere Zahl an unabhängigen Variablen abgebildet. Da die neuen Größen kaum noch sinnvoll interpretierbar sind und sich vollkommen anders verhalten können als die ursprünglichen Einflussgrößen, ist auch diese Vorgehensweise mit erheblichen Problemen verbunden. Eine Methode zur Beseitigung von Multikollinearität, die durch so genannte Wechselwirkungen (siehe weiter unten) verursacht wird, besteht aus der Zentrierung von Variablen bzw. Einflussgrößen [UM06]. Bei der Zentrierung wird von einer Einflussgröße ihr arithmetisches Mittel (bezogen auf die Stichprobe) subtrahiert. Dies kann unter Umständen die Interpretation der Werte einer Zielgröße erschweren. Des Weiteren hat eine Zentrierung erhebliche Auswirkungen auf den Regressionskoeffizienten  $\beta_0$ . Eine Regression durch den Ursprung verliert in aller Regel dadurch ihren Sinn.

Letztlich muss vor Anwendung von Techniken zur Beseitigung bzw. Reduzierung von Multikollinearität auch die Frage gestellt werden, ob nicht eine Korrelation zwischen Einflussgrößen besteht, die inhaltlicher Natur ist und ausgehend vom theoretischen Modell sogar erklärt werden kann. In dieser Arbeit wurden z. B. Qualitätsmaße entwickelt, die auf bestimmte Merkmale von Bildkompressionsartefakten reagieren. Es ist klar, dass solche Merkmale nicht unabhängig voneinander auftreten und somit zwangsläufig Abhängigkeiten zwischen den Qualitätsmaßen bestehen. In solchen Fällen muss kritisch geprüft werden, ob die Probleme, die durch Multikollinearität verursacht werden, tragbar sind. Soll die Multikollinearität unbedingt verringert werden, könnte sich als erster Schritt eine Zentrierung der Einflussgrößen anbieten.

Grundsätzlich sollte bei der Aufstellung eines Regressionsmodells darauf geachtet werden, möglichst keine Einflussgrößen aufzunehmen, die keinen nennenswerten Beitrag zur Erklärung der Zielgröße liefern. Jedoch muss dabei vermieden werden, dass bei diesem Vorgehen tatsächlich bedeutende Einflussgrößen ausgeschlossen werden. Es kommt oft vor, dass einzelne Einflussgrößen nur gemeinsam mit einer oder mehreren anderen Einflussgrößen einen bedeutenden Beitrag zur Erklärung der Zielgröße liefern. Man spricht dann von so genannten Suppressorvariablen [UM06]. Suppressorvariablen können für sich betrachtet nur sehr gering mit der Zielgröße korrelieren und ihre Aufnahme in ein Regressionsmodell kann das Bestimmtheitsmaß des Regressionsmodells nur wenig erhöhen, solange sich nicht bestimmte andere Größen ebenfalls im Modell befinden. Die Einschätzung, ob eine Einflussgröße unbedeutend ist, gestaltet sich daher problematisch und es gilt stets darauf zu achten, einzelne Einflussgrößen nicht vorschnell aus dem Modell auszuschließen. Dabei kann es hilfreich sein, Überlegungen zur Konsistenz eines Regressionsmodells mit dem der Problemstellung zugrunde liegenden theoretischen Modell anzustellen.

In der Praxis kommt es oft vor, dass sich nicht nur die einzelnen Einflussgrößen auf die Zielgröße auswirken, sondern dass auch ein bedeutender Einfluss auf die Zielgröße durch das direkte Zusammenwirken mindestens zweier Einflussgrößen besteht. Man spricht dann von Wechselwirkung oder Interaktion [UM06, Kle03, WJ99]. In einem Regressionsmodell wird Interaktivität durch die Aufnahme neuer Variablen, die Funktionen bereits im Modell befindlicher Einflussgrößen sind, berücksichtigt. Typischerweise werden in der Praxis Wechselwirkungen durch multiplikative Faktoren aus jeweils zwei Einflussgrößen modelliert. Bei der Berücksichtigung von Wechselwirkungen besteht grundsätzlich die Gefahr des Auftretens weiterer Multikollinearität. Die Problematik wird deutlich verringert, indem Wechselwirkungsfaktoren auf Basis zentrierter Einflussgrößen verwendet werden [UM06].

## 3.5 Schrittweise Regression

Eine formale iterative Methode zur Suche einer Linearkombination aus Einflussgrößen, die mit einem Minimum an Variablen eine maximale Erklärung der Zielgröße liefert, ist

die schrittweise Regression [Kle03, FH84]. Es besteht große Gefahr, dass dabei Modelle entstehen, die sich theoretisch kaum erklären lassen, oder die zu sehr auf den verwendeten Datensatz optimiert sind. Auch können Variablen aus dem Modell herausfallen, die aus theoretischer Sicht als wichtig einzustufen sind und durchaus einflussstark sind. Tatsächlich kann im Rahmen einer schrittweisen Regression weitgehend auf theoretische Vorüberlegungen verzichtet werden, denn die Auswahl von Variablen wird der formalen Methodik überlassen [UM06]. Ob die Aufstellung von Modellen ohne theoretische Grundlage in der Forschungspraxis sinnvoll ist, muss in vielen Fällen stark angezweifelt werden. Die schrittweise Regression soll dennoch beschrieben werden, da sie ausgehend von einem theoretischen Modell zur Suche von Modellvarianten oder zur Erweiterung des Modells angewendet werden kann.

Eine Möglichkeit der Durchführung einer schrittweisen Regression ist die iterative Durchführung der folgenden drei Schritte.

1. Zunächst wird mit Hilfe der einfachen linearen Regression ermittelt, welche Einflussgröße zum größten Bestimmtheitsmaß  $B$  der Zielgröße führt. Sie wird in das multiple Regressionsmodell übernommen.
2. Nun wird geprüft, welche weitere Einflussgröße die größte Erhöhung des Bestimmtheitsmaßes bei Aufnahme in das multiple Regressionsmodell ergibt. Wenn die Abhängigkeit zwischen Zielgröße und Einflussgröße signifikant ist, wird die Einflussgröße im Modell belassen.
3. Schritt 2 wird wiederholt, solange sich Einflussgrößen finden lassen, für die sich eine signifikante Abhängigkeit zeigen lässt.

Es kann auch der umgekehrte Weg gegangen werden, d. h. man geht von einem Regressionsmodell mit allen Einflussgrößen aus und beseitigt schrittweise diejenigen, die für die Erklärung der Zielgröße nicht signifikant sind. Beide Vorgehensweisen sind nicht äquivalent und können unterschiedliche Modelle ergeben.

Die Signifikanz der einzelnen Einflussgrößen wird bestimmt, indem für jede Einflussgröße ein statistischer Signifikanztest (t-Test) durchgeführt wird [Har02, GI94]. Es muss beachtet werden, dass ein solcher Test mit Unschärfen belegt ist, wenn die weiter oben genannten vier Voraussetzungen im Zusammenhang mit der linearen Regression nicht erfüllt sind [Kle03, Fec00]. Ausgegangen wird von der Hypothese, dass der Regressionskoeffizient  $\beta_j$  gleich einem konstanten Wert  $\beta_j^0$  ist und der Alternativhypothese, dass dies nicht der Fall ist:

$$H_0 : \beta_j = \beta_j^0, \quad H_A : \beta_j \neq \beta_j^0. \quad (3.26)$$

Die Nullhypothese ist dann zugunsten der Alternativhypothese zu verwerfen, wenn bei gegebenem Vertrauensniveau  $1 - \alpha$  (typischerweise  $\alpha = 0,05$ ) gilt:

$$\frac{|b_j - \beta_j^0|}{\sqrt{\sigma^2 c_{jj}}} > t_{N-M-1; 1-\alpha/2}. \quad (3.27)$$

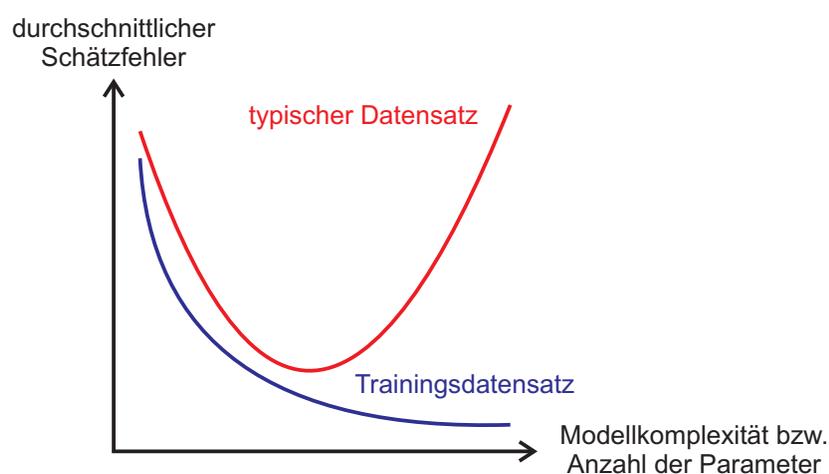
Zur Schätzung der Fehlervarianz  $\sigma^2$  siehe Gleichung (3.16). Konkret wird geprüft, ob eine Einflussgröße für die Erklärung der Zielgröße signifikant ist, indem  $\beta_j^0 = 0$  gesetzt wird und getestet wird, ob die Nullhypothese abgelehnt wird.

## 3.6 Abschließende Bemerkungen zur Regressionsanalyse

Einzelne Ausreißer in der Datenbasis haben insbesondere bei Verwendung der Methode der kleinsten Quadrate große Auswirkungen auf die geschätzten Regressionskoeffizienten. Daher sollte grundsätzlich über eine Elimination von Ausreißern nachgedacht werden. In [Sch03] wird vorgeschlagen, Ausreißer durch grafische Analyse des Residuendiagramms zu identifizieren, indem auf besonders auffällige Punkte geachtet wird. In [Har02] wird zu demselben Zweck die grobe Regel angegeben, dass die Wertekombination  $(x_{1i}, x_{2i}, \dots, x_{Mi}, y_i)$  einen Ausreißer darstellt, wenn das zugehörige auf  $\hat{\sigma}$  normierte Residuum  $d_i$  einen Betrag größer als Drei aufweist. Eine andere Möglichkeit stellt die Verwendung von Schätzern dar, die weniger kritisch auf Ausreißer reagieren. [Sch03] enthält eine Aufzählung von robusten Schätzern und entsprechende Literaturverweise. Die Grundidee besteht darin, dass die Schätzer stark abweichende Punkte weniger gewichten als wenig abweichende Punkte, so dass Ausreißer sich weniger stark auf die Schätzung auswirken. Ein Beispiel ist der gewichtete kleinste Quadrate Schätzer [GI94]. Robuste Schätzungen lassen sich oft nur durch iterative Algorithmen berechnen.

Es muss an dieser Stelle noch erwähnt werden, dass eine Aussage über die Bedeutung einzelner Einflussgrößen durch Vergleich der Regressionskoeffizienten nicht ohne Weiteres möglich ist. Gründe hierfür sind Multikollinearität, unterschiedliche Maßstäbe der Einflussgrößen und unterschiedliche Vorzeichen der Regressionskoeffizienten. Für weitere Informationen diesbezüglich siehe [Tie87]. Des Weiteren muss davor gewarnt werden, ein Regressionsmodell ohne Vorbehalte über den untersuchten Bereich der Einflussgrößen hinaus zu extrapolieren, da dabei erhebliche Fehler entstehen können [Kle03, Sto93]. Schließlich muss die Grundvoraussetzung einer linearen Regression - die mindestens näherungsweise lineare Abhängigkeit der Zielgröße von den Einflussgrößen - außerhalb der untersuchten Bereiche der Einflussgrößen nicht mehr gelten. Ob eine Extrapolation über den untersuchten Bereich der Einflussgrößen hinaus möglich ist, und wie groß die zu erwartenden Abweichungen sind, kann nur auf Basis eines sinnvollen theoretischen Modells des untersuchten Systems abgeschätzt werden.

Grundsätzlich gilt bei der Aufstellung von Regressionsmodellen, dass mit jeder Hinzunahme einer weiteren Einflussgröße bzw. eines weiteren Maßes die Korrelation zwischen modellbasierter Schätzung und den zu schätzenden Daten mindestens gleich bleibt, in der Regel aber erhöht wird. Gleichzeitig steigt aber die Gefahr von Unplausibilitäten und des Auftretens eines oder mehrerer der bereits angesprochenen statistischen Probleme im Zusammenhang mit der Regressionsanalyse. Es kann somit beobachtet werden, dass sehr komplexe Regressionsmodelle oft sehr gut an einen einzelnen Datensatz (Trainingsdatensatz) angepasst sind, aber keine akkurate Schätzung in anderen Datensätzen ermöglichen.



**Abbildung 3.1:** Veranschaulichung des Verhaltens des Schätzfehlers in Abhängigkeit von der Modellkomplexität

In Abbildung 3.1 ist qualitativ veranschaulicht, wie sich die Modellkomplexität von Regressionsmodellen auf den durchschnittlichen Schätzfehler im Trainingsdatensatz und in anderen Datensätzen auswirken kann. Wie bereits erwähnt, wird mit zunehmender Komplexität die Anpassung eines Modells an den Trainingsdatensatz immer besser. Für die Praxis viel entscheidender ist aber oft ein kleiner mittlerer Schätzfehler in anderen Datensätzen, was entsprechend der Abbildung nicht unbedingt bei Verwendung sehr komplexer Modelle gegeben ist. Diese Überlegung sollte daher stets bei der Aufstellung von Regressionsmodellen berücksichtigt werden, insbesondere dann, wenn der in der Praxis (und auch in dieser Arbeit) bedeutende Fall gegeben ist, dass auf Basis eines verhältnismäßig kleinen Trainingsdatensatzes ein Modell erzeugt werden soll, welches sich durch akkurate Schätzungen in einem breiten Spektrum an anderen Daten auszeichnet. Bei sehr umfangreichen Trainingsdatensätzen ist eine Verringerung der Probleme im Zusammenhang mit komplexen Modellen möglich, so dass komplexe Modelle dann insgesamt gegenüber einfacheren Modellen mehr im Vorteil sein können.

# 4 Objektive Verfahren zur Bildqualitätsbewertung

Das Problem der Qualitätsbewertung von Bildern und Bildsequenzen ist so alt wie die Bildverarbeitung selbst. Da im Bereich der Störungsreduktion meist Algorithmen entwickelt werden, die darauf abzielen, die vom Menschen wahrgenommenen Störungen zu reduzieren und auf diese Weise die empfundene Bildqualität zu verbessern, erscheint es sinnvoll, solche Algorithmen auch von menschlichen Betrachtern subjektiv bewerten zu lassen. Der Zeit- und Kostenaufwand für subjektive Tests ist jedoch sehr groß, wenn reproduzierbare Ergebnisse entstehen und erprobte Standards bzw. Empfehlungen (z. B. [ITU02]) eingehalten werden sollen.

Daher existieren diverse Ansätze zur objektiven Bewertung der Qualität von Bildern und Bildsequenzen, deren Zweck es ist, die subjektiven Eindrücke von menschlichen Betrachtern möglichst genau und reproduzierbar zu schätzen. Viele dieser Ansätze eignen sich nur für Qualitätsbewertungen in speziellen Anwendungsbereichen (z. B. in der Rauschreduktion oder der Bildcodierung). Nach einem universell einsetzbaren Werkzeug zur Bildqualitätsbewertung wird nach wie vor gesucht. In [Fec99] wird aufgrund der enormen Komplexität des HVS bezweifelt, dass jemals ein Messverfahren gefunden werden kann, das in allen denkbaren Anwendungsszenarien befriedigende Ergebnisse liefert. Es wird vorgeschlagen, für unterschiedliche Aufgaben auch unterschiedliche Messverfahren einzusetzen.

In diesem Kapitel soll eine Übersicht über objektive Verfahren zur Qualitätsbewertung von Bildern oder Bildsequenzen gegeben werden. Es sei vorweggenommen, dass sich die Untersuchungen an späterer Stelle in dieser Arbeit auf Qualitätsbewertungen in Einzelbildern konzentrieren. Im Folgenden wird zunächst dargestellt, in welche Klassen sich verschiedene Ansätze zur Bewertung bzw. Messung der Bildqualität nach unterschiedlichen Kriterien einteilen lassen. Anschließend werden mehrere für diese Arbeit relevante Ansätze zur objektiven Bildqualitätsbewertung aus verschiedenen Klassen vorgestellt. Eine Relevanz der vorgestellten Ansätze ist deshalb gegeben, weil diese entweder von anderen Autoren zur Bewertung der Wirksamkeit von Deblocking-Verfahren - der Hauptaufgabenstellung dieser Arbeit - eingesetzt wurden, oder zumindest zur Bewertung der Qualität komprimierten Bildmaterials - einem verwandten Thema - entwickelt oder optimiert wurden.

[Esk00] enthält eine Übersicht über eine Vielzahl an objektiven Qualitätsmaßen. Es wird zwischen Maßen, die nur ein einziges Eingangsbild und Maßen, die zwei Eingangsbilder benötigen und einen Vergleich zwischen den Bildern durchführen, unterschieden. Der

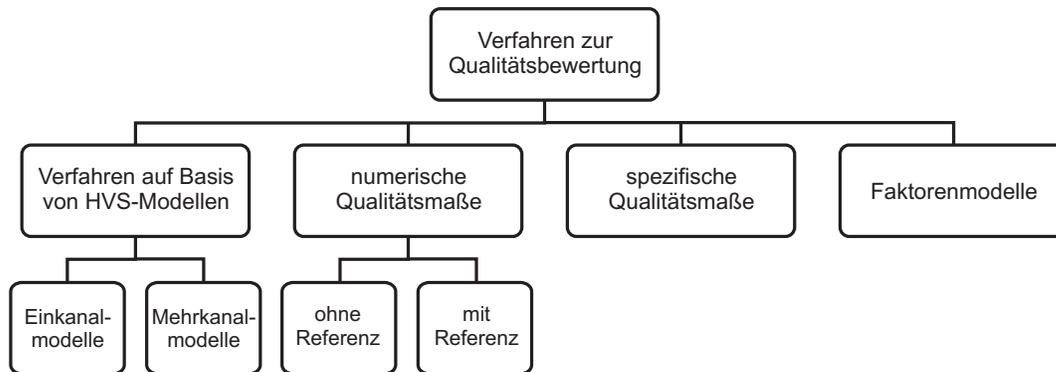
Autor spricht von univariaten und bivariaten Maßen. Es sei an dieser Stelle erwähnt, dass bivariate Maße, also solche, die ein Vergleichs- bzw. Referenzbild benötigen, prinzipbedingt nicht für eine Echtzeit-Kontrolle der Qualität von z. B. zu sendendem Bildmaterial geeignet sind, da eine Referenz in solchen Fällen oft nicht zur Verfügung steht. Für die Untersuchungen in dieser Arbeit ergeben sich durch den Einsatz von bivariaten Maßen jedoch keine Einschränkungen, da hier grundsätzlich Referenzen bzw. Vergleichsbilder verfügbar sind.

Des Weiteren wird in [Esk00] eine Unterscheidung von Qualitätsmaßen nach Art der Messung vorgeschlagen. Es werden die Klassen *Numerische Messverfahren*, *Modelle des menschlichen visuellen Systems* und *grafische Messverfahren* genannt. Zur ersten Klasse zählen alle Maße, die auf oft einfachen Rechenoperationen basieren und die Qualität über eine einzige Kennzahl beschreiben (z. B. MSE, durchschnittliche Intensitätsdifferenz oder verschiedene Korrelationsmaße). Zur zweiten Klasse zählen Maße, in die auf geeignete Art und Weise Modelle von Teilen des menschlichen visuellen Systems (HVS) eingebunden sind, um die von menschlichen Betrachtern wahrgenommene Bildqualität möglichst gut abschätzen zu können (Einkanalmodele, Mehrkanalmodele). Zur dritten Klasse zählen Maße, die qualitätsrelevante Merkmale von Bildern grafisch hervorheben oder aufbereiten und eine Qualitätseinschätzung auf Basis umfangreicher Daten ermöglichen (z. B. Histogramme oder Differenzbilder).

[ASS02] enthält ebenfalls eine umfangreiche Übersicht über verschiedene Qualitätsmaße. Es wird eine Einteilung in sechs Gruppen nach Art der in den Maßen verwendeten Informationen vorgenommen. Die Autoren nennen Intensitätsdifferenzmaße (z. B. MSE), Maße auf Basis von Korrelationen, Maße auf Basis der Analyse von Kanten (z. B. Kantenverschiebung), Maße im Spektralbereich, HVS-basierte Maße und kontext-basierte Maße (Analyse von Bildpunktnachbarschaften oder ganzen Bildbereichen).

In [Fec99] werden verschiedene Methoden zur Bildqualitätsmessung miteinander verglichen. Eine Unterscheidung erfolgt dort zwischen Modellen des menschlichen visuellen Systems und Maßen, die auf der Extraktion spezifischer Störsignale (z. B. Blocking und Mosquito-Noise) und deren Bewertung basieren. Darüber hinaus wird die Verknüpfung solcher Maße in Form von so genannten Faktorenmodellen behandelt. Eine weitere Unterteilung wird nach der Anzahl der Eingangsbilder bzw. nach Varianten mit oder ohne Referenz vorgenommen. In [WLYQ96] findet man eine vergleichbare Einteilung von Bildqualitätsmaßen.

Abbildung 4.1 zeigt eine grafische Übersicht über Ansätze zur Bildqualitätsbewertung. Viele der Kategorien aus der bereits zitierten Literatur finden sich dort wieder. Recht klar abgrenzbar ist die Klasse der Maße auf Basis von Modellen des menschlichen visuellen Systems. Eine weitere Einteilung in Ein- und Mehrkanalmodele ist möglich. Daneben stellen die numerischen Maße eine wichtige Klasse dar. Zu dieser Klasse sollen hier alle Maße gezählt werden, die z. B. auf Intensitätsdifferenz- oder Korrelationsmessungen basieren und einen einzigen numerischen Qualitätswert für ein Bild oder Bildpaar liefern (z. B. MSE



**Abbildung 4.1:** Übersicht über Verfahren zur Bildqualitätsbewertung

oder PSNR). Hier ist eine weitere Unterteilung in Maße mit und ohne Referenz möglich. Als dritte Klasse werden hier Maße, die spezifische Störungen (z. B. Blockartefakte oder Kantenverschleifungen) analysieren, betrachtet. Da solche Maße typischerweise sehr stark an bestimmte Störmerkmale angepasst sind, ist eine Einteilung in weitere Unterklassen kaum möglich. Die vierte Klasse bilden schließlich die Faktorenmodelle, welche mehrere Qualitätsmaße aus den anderen drei Klassen enthalten und geeignet miteinander kombinieren. Da in dieser Klasse eine sehr große Vielfalt besteht, ist eine weitere Unterteilung nicht sinnvoll. Grafische Messverfahren werden in der Abbildung nicht berücksichtigt, da sie eher eine Interpretationshilfe darstellen und die Bildqualität nicht in Form von einfachen Maßzahlen abschätzen können.

Weitere Literaturquellen, in denen Qualitätsmaße miteinander verglichen werden, sind [LR05, WSM01, EF95]. Sehr umfangreiche Darstellungen des Themas Bildqualität finden sich in [WR06, Win05].

## 4.1 MSE und PSNR

Der mittlere quadratische Fehler (MSE, mean squared error) ist ein Maß für die quadratische Differenz zwischen zwei Bildern oder Bildsequenzen und gehört zur Klasse der numerischen Maße mit Referenz. Je unähnlicher sich die miteinander zu vergleichenden Bilder sind, desto größer wird der MSE. Die Berechnungsvorschrift des MSE für ein Testbild  $s(x, y)$  und ein Referenzbild  $s_{ref}(x, y)$  ist in Gleichung 4.1 gegeben. Der MSE entspricht der Varianz des Differenzbildes  $\sigma_{diff}^2$ .  $N$  und  $M$  bezeichnen die Bildbreite und Bildhöhe in Pixeln.

$$MSE = \sigma_{diff}^2 = \frac{1}{N \cdot M} \sum_{x=1}^N \sum_{y=1}^M (s(x, y) - s_{ref}(x, y))^2. \quad (4.1)$$

Der MSE ist ein sehr oft eingesetztes Qualitätsmaß, da viele Verfahren und Anwendungen in der Bildsignalverarbeitung die Minimierung des mittleren quadratischen Fehlers - also des MSE - zum Ziel haben [Win05]. In [Esk00] findet man die Aussage, dass der MSE nach wie vor unabdingbar für die Bewertung von Kompressionsalgorithmen sei, da andere verlässliche Qualitätsmaße fehlten. Es sei jedoch auch erwähnt, dass der MSE in vielen Fällen nicht optimal mit den subjektiven Eindrücken von menschlichen Beobachtern korreliert. Auf Gründe wird im Folgenden noch eingegangen.

Der PSNR (peak signal to noise ratio) ist ein mit dem MSE verwandtes Fehler- bzw. Gütemaß, das in der Bildsignalverarbeitung sehr häufig eingesetzt wird. Der PSNR stellt ein Maß zur Beurteilung der Ähnlichkeit zweier Bilder bzw. Bildsequenzen dar. Zur Berechnung muss wie beim MSE neben dem Testbild  $s(x, y)$  grundsätzlich ein Referenzbild  $s_{ref}(x, y)$  zur Verfügung stehen. Je größer die Ähnlichkeit zwischen Referenz- und Testbild ist, desto höher fällt der PSNR-Wert aus. Die folgende Formel gibt die Berechnungsvorschrift für den PSNR eines Bildes an:

$$PSNR = 10 \cdot \log \left[ \frac{s_{max}^2}{\frac{1}{N \cdot M} \sum_{x=1}^N \sum_{y=1}^M (s(x, y) - s_{ref}(x, y))^2} \right] [dB] = 10 \cdot \log \left[ \frac{s_{max}^2}{MSE} \right] [dB]. \quad (4.2)$$

$N$  und  $M$  bezeichnen auch hier die Bildbreite und Bildhöhe in Pixeln und  $s_{max}$  den Signalspitzenwert, welcher in üblichen Bildverarbeitungssystemen mit 8 Bit Intensitätsauflösung (pro Kanal) 255 beträgt.

PSNR und MSE sind übliche bzw. häufig eingesetzte Qualitätsmaße, wenn es um die objektive Bewertung der Qualität von Bildmaterial mit überlagerten rauschartigen Störungen oder um die Bewertung von Rauschreduktionsverfahren geht. Im Rahmen einer in [ASS02] vorgestellten statistischen Untersuchung der Eignung verschiedener Qualitätsmaße zur Prädiktion von Bildqualität bei verrauschtem Bildmaterial schneidet der PSNR bzw. MSE auch vergleichsweise gut ab. Andererseits wird in [Jos01] gezeigt, dass die Prädiktionsgüte von PSNR und MSE bei einer Qualitätsbewertung von verrauschtem Bildmaterial begrenzt ist und besser geeignete Ansätze existieren. Davon abgesehen muss grundsätzlich mit einer eingeschränkten bzw. geringen Eignung von PSNR und MSE zur Qualitätsbewertung gerechnet werden, wenn andere Arten von Störungen im Bildmaterial vorhanden sind. Der PSNR, wie auch der MSE, berücksichtigen keinerlei Eigenschaften des menschlichen visuellen Systems und der im Bildmaterial vorhandenen Störungen. Dies ist insbesondere im Kontext von Codierartefakten problematisch, da solche Artefakte nur an bestimmten Stellen auftreten und sehr spezielle örtliche und zeitliche Eigenschaften aufweisen. Hinzu kommt, dass der PSNR alle Abweichungen zwischen zwei Testbildern gleich gewichtet. So haben beispielsweise Bilder, die großflächige Störungen mit geringer Intensität enthalten, den gleichen PSNR-Wert wie solche mit lokal begrenzten Störungen

hoher Intensität. Menschliche Betrachter würden beide Fälle hingegen mit hoher Wahrscheinlichkeit unterschiedlich bewerten. Auch die unterschiedliche Reaktion des HVS auf Reize unterschiedlicher Frequenz und Maskierungseffekte werden beim PSNR und MSE nicht berücksichtigt. Auf grundsätzliche Probleme im Zusammenhang mit dem PSNR wird in [Win05, Ric03, Ohm95] eingegangen.

Im Hinblick auf die Bewertung von Artefaktreduktionsverfahren sind PSNR und MSE besonders problematisch, da starke Bilddatenkompression zu einer Vielzahl an Codierartefakten in den codierten Bildern führt, was die Ähnlichkeit zum ungestörten Originalbild stark vermindert. Eine Reduktion von Codierartefakten ändert an dieser Tatsache in der Regel wenig, zumal das Ziel vieler Verfahren nicht die Wiederherstellung des ungestörten Originalbildes, sondern eine geschickte Kaschierung der Codierartefakte ist. In bestimmten Fällen kann eine Reduktion von Codierartefakten sogar zu einer weiteren Verringerung der Ähnlichkeit eines Bildes gegenüber dem Originalbild führen, was einer Verringerung des PSNR bzw. einer Vergrößerung des MSE entspricht. [BBK05, ASS02, YWWC02, Fec99, KK95] benennen die schlechte Eignung des PSNR bzw. MSE für eine Qualitätsbewertung in Anwendungsfällen, in denen Codierartefakte im Vordergrund stehen. Andererseits zeigen Untersuchungen in [VQE00, VQE03, Esk00, MRLP<sup>+</sup>06], dass der PSNR bzw. ein darauf basierendes Maß zumindest für eine absolute Qualitätsbewertung von komprimierten Bildsequenzen eingeschränkt geeignet ist. In der Praxis werden PSNR und MSE vielfach sowohl zur absoluten Qualitätsbewertung komprimierten Bildmaterials als auch zum Vergleich von Artefaktreduktionsverfahren eingesetzt.

Gelegentlich findet man in der Literatur den Begriff DPSNR (differential peak signal to noise ratio) [KYL<sup>+</sup>04, AP04]. Damit ist die Differenz des PSNR-Wertes eines gefilterten Bildes ( $PSNR_B$ ) und des PSNR-Wertes eines gestörten Bildes ( $PSNR_A$ ) gemeint. Auf diese Weise können z. B. Qualitätsgewinne durch die Anwendung von Störungsreduktionsverfahren oder durch Modifikationen an Video-Codecs angegeben werden. Gleichbedeutend mit dem DPSNR ist der ebenfalls gelegentlich verwendete Begriff PSNR<sub>i</sub> (peak signal to noise ratio improvement) [Jos01].

$$DPSNR = PSNR_B - PSNR_A = PSNR_i. \quad (4.3)$$

## 4.2 Gradientenbasierte Blocking-Messung

Die gradientenbasierte Blocking-Messung ist ein Ansatz aus der Klasse der spezifischen Qualitätsmaße, der spezielle Merkmale von Blockartefakten berücksichtigt. Er ist daher prinzipiell zur Qualitätsbewertung blockbasiert komprimierter Bilder oder auch zur Bewertung der Wirksamkeit von Artefaktreduktionsverfahren geeignet. Es existieren in der Literatur verschiedene Varianten gradientenbasierter Messverfahren, von denen im Folgenden einige vorgestellt werden. Eine sehr häufig zitierte Variante ist das gradientenbasierte

Blocking-Maß nach [WY97]. Die Funktionsweise des Maßes, das eine Blockgröße von 8x8 Pixeln voraussetzt, soll zunächst für horizontal verlaufende Blockgrenzen skizziert werden. Ausgegangen wird von einem Messwert, der die Stärke von Diskontinuitäten bzw. die Größe von Intensitätsdifferenzen (Gradienten) an Blockgrenzen zusammenfasst:

$$M_h = \left[ \sum_{k=1}^{N_c/8-1} \left\| \underline{W}_k \left[ \underline{f}_{c(8 \cdot k)} - \underline{f}_{c(8 \cdot k+1)} \right] \right\|_2^2 \right]^{1/2}. \quad (4.4)$$

Die  $\underline{f}_{c_j}$  mit  $j = 1 \dots N_c$  sind die  $N_c$  Spaltenvektoren, aus denen sich ein komplettes Bild zusammensetzen lässt.  $\underline{W}_k = \text{diag} [w_{1j}, w_{2j}, \dots, w_{N_r j}]$ , mit der Bildhöhe  $N_r$  und  $j = 8 \cdot k$  für  $k = 1, 2, \dots, N_c/8 - 1$ , dient der Gewichtung von Intensitätsdifferenzen mit dem Ziel, Nichtlinearitäten beim Helligkeitsempfinden des HVS und Maskierungseffekte zu berücksichtigen. Es werden Intensitätsdifferenzen in besonders dunklen und besonders hellen Bildbereichen, sowie in Bildbereichen mit hoher Aktivität geringer gewichtet als in anderen Bildbereichen, da das menschliche Auge dort weniger empfindlich für die Wahrnehmung von Diskontinuitäten ist. Die Gewichtungsfaktoren  $w_{ij}$  werden folgendermaßen berechnet:

$$w_{ij} = \begin{cases} \lambda \cdot \ln \left( 1 + \frac{\sqrt{\mu_{ij}}}{1 + \sigma_{ij}} \right) & \text{für } \mu_{ij} \leq \varsigma \\ \ln \left( 1 + \frac{\sqrt{255 - \mu_{ij}}}{1 + \sigma_{ij}} \right) & \text{sonst} \end{cases}. \quad (4.5)$$

$\mu_{ij}$  ist der Intensitätsmittelwert eines Vektors aus 16 waagrecht benachbarten Pixeln, die sich zu beiden Seiten der Position erstrecken, für die der Gewichtungsfaktor berechnet werden soll. Die Position des Gewichtungsfaktors entspricht der Position Acht des Vektors, wenn die Positionszählung mit Eins beginnt. Die Blockgrenze ist zwischen den Positionen Acht und Neun zu finden.  $\sigma_{ij}$  ist der Mittelwert der Standardabweichungen des linken und des rechten Teilvektors. Laut [WY97] gilt  $\varsigma = 81$  und  $\lambda = 1, 1520$ .

$M_h$  muss noch auf die durchschnittliche Intensitätsdifferenz benachbarter Pixel, die sich nicht auf Blockgrenzen befinden, normiert werden, so dass sich der Blocking-Messwert für horizontal verlaufende Blockgrenzen zu  $M_{hGBIM} = M_h/E$  ergibt. *GBIM* steht für *Generalized Block-Edge Impairment Metric*. Es gilt

$$E = \frac{1}{7} \sum_{n=1}^7 S_n \quad (4.6)$$

und

$$S_n = \left[ \sum_{k=1}^{N_c/8-1} \left\| \underline{W}_k \left[ \underline{f}_{c(8 \cdot k+n)} - \underline{f}_{c(8 \cdot k+n+1)} \right] \right\|_2^2 \right]^{1/2}. \quad (4.7)$$

Die Berechnung eines Blocking-Messwertes für vertikal verlaufende Grenzen erfolgt entsprechend. Der Blocking-Messwert für das gesamte Bild ist der Mittelwert der Messwerte für horizontal und vertikal verlaufende Blockgrenzen. Je größer der Blocking-Messwert wird, umso größer sind die prädierten Blocking-Störungen. Ungestörte Bilder haben einen Blocking-Messwert von ca. Eins. Einige Autoren verwenden das beschriebene Maß, um die Wirksamkeit einzelner Deblocking-Verfahren nachzuweisen. Hierzu wird die jeweilige Verringerung des durch dieses Bildqualitätsmaß ermittelten Blocking-Messwertes verwendet. Leicht veränderte Varianten des Blocking-Maßes finden sich in [Sut98, SW97].

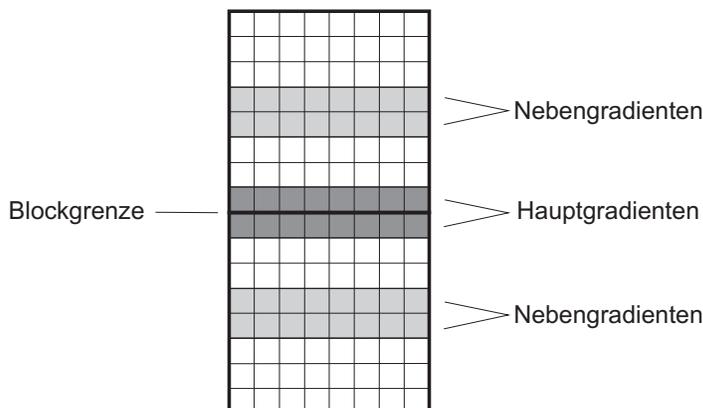
Ein etwas anderer Ansatz wird in [Tra98] gewählt. Zwar bilden auch hier Pixelintensitätsdifferenzen an unterschiedlichen Rasterlagen periodischer Raster die Basis, jedoch wird eine Einteilung des Bildes in Makroblöcke der Größe 16x16 Pixel vorgenommen, so dass 16 Rasterlagen unterschieden werden können. Zur Beurteilung eines Bildes wird ein Vektor mit 16 Einträgen erzeugt, der die durchschnittlichen Intensitätsunterschiede benachbarter Pixel an allen 16 Rasterlagen enthält. In unkomprimierten Bildern haben alle 16 Einträge etwa dieselbe Größe. In komprimierten Bildern sind zwei Einträge jedoch deutlich größer als die restlichen 14. Die Unterschiede zwischen den beiden größeren und dem Mittelwert der kleineren Vektoreinträge dienen der Berechnung eines Qualitätswertes. Das Verfahren wurde zur Echtzeit-Bildqualitätsüberwachung in MPEG-basierten Übertragungssystemen vorgeschlagen und erprobt. In [Tra01] findet sich eine sehr umfassende Darstellung und Untersuchung des Verfahrens. Es werden Varianten mit und ohne Zugriff auf Referenzinformationen vorgestellt und es wird auf die Berücksichtigung von Maskierungseffekten eingegangen. Als neuer Aspekt werden Übertragungsstörungen und ihre Auswirkungen auf das Maß betrachtet.

Abschließend soll noch ein sehr aufwandsgünstig implementierbares gradientenbasiertes Blocking-Maß, das in [Pia05a] vorgestellt wird, in seiner Funktionsweise erläutert werden. Das Maß ist an eine blockbasierte Bildkompression mit 8x8 Pixeln Blockgröße angepasst. Gemessen werden absolute Intensitätsdifferenzen benachbarter Bildpunkte an Stellen mit zwei unterschiedlichen Abständen von Blockgrenzen. Intensitätsdifferenzen, die direkt über Blockgrenzen hinweg gemessen werden, werden Hauptgradienten genannt, während Intensitätsdifferenzen abseits von Blockgrenzen (hier in Blockmitte) Nebengradienten genannt werden. Abbildung 4.2 veranschaulicht die Lage von Haupt- und Nebengradienten am Beispiel zweier benachbarter Blöcke, die eine gemeinsame horizontale Grenze haben. Für benachbarte Blöcke, die eine gemeinsame vertikale Grenze haben, gelten entsprechende Lagen der Gradienten.

Zunächst sind für ein zu bewertendes Bild alle Hauptgradienten und eine gleiche Anzahl an Nebengradienten aufzuaddieren. Die Summe der Hauptgradienten für horizontale Blockgrenzen berechnet sich zu

$$HG_1 = \sum_x \sum_y abs(s(x, y + 1) - s(x, y)), \quad (4.8)$$

mit  $x \in \{0, \dots, PW - 1\}$  und  $y = 8m - 1, m \in \{1, \dots, (PH/8) - 1\}$ .



**Abbildung 4.2:** Lage von Haupt- und Nebengradienten

$PH$  ist die Bildhöhe in Zeilen und  $PW$  ist die Bildbreite in Spalten. Entsprechend berechnen sich die Summe der Hauptgradienten für vertikale Blockgrenzen und die Nebengradientensummen zu

$$HG_2 = \sum_x \sum_y abs(s(x+1, y) - s(x, y)), \quad (4.9)$$

mit  $x = 8n - 1$ ,  $n \in \{1, \dots, PW/8 - 1\}$  und  $y \in \{0, \dots, PW - 1\}$ ,

$$NG_1 = \sum_x \sum_y abs(s(x, y+1) - s(x, y)), \quad (4.10)$$

mit  $x \in \{0, \dots, PW - 1\}$  und  $y = 8m - 5$ ,  $m \in \{1, \dots, (PH/8) - 1\}$ ,

$$NG_2 = \sum_x \sum_y abs(s(x+1, y) - s(x, y)), \quad (4.11)$$

mit  $x = 8n - 5$ ,  $n \in \{1, \dots, PW/8 - 1\}$  und  $y \in \{0, \dots, PW - 1\}$ .

Nun werden die Hauptgradientensummen zu den Nebengradientensummen ins Verhältnis gesetzt. Es ergibt sich der Blocking-Messwert  $BL$ :

$$BL = \frac{HG_1 + HG_2}{NG_1 + NG_2}. \quad (4.12)$$

In natürlichen unkomprimierten Bildern sind die Summen der Haupt- und der Nebengradienten durchschnittlich gleich groß. Somit ist das Verhältnis der Gradientensummen etwa Eins. Großflächige periodische Strukturen können zu Abweichungen führen. In solchen Fällen muss das Maß mit Vorsicht interpretiert werden. In blockbasiert komprimierten Bildern werden Blocking-Werte gemessen, die größer als Eins sind. Dies liegt daran, dass Blockartefakte, also Diskontinuitäten an Blockgrenzen, die Hauptgradientensummen

ansteigen lassen, während kompressionsbedingte Informationsverluste in Blockinnenbereichen die Nebengradientensummen verkleinern. Beide Effekte treten umso stärker auf, je höher der Kompressionsgrad eines Bildes ist.

Die Blocking-Messwerte eignen sich direkt zur Abschätzung der in einem Bild vorhandenen Anzahl und Stärke von Blockartefaken. Zur Beurteilung der Wirksamkeit von Deblocking-Verfahren können Blocking-Messwerte vor und nach Anwendung eines Artefaktreduktionsverfahrens miteinander verglichen werden. Es ist zu beachten, dass sich bei diesem Vorgehen Probleme ergeben können, was grundsätzlich auch auf die anderen bereits vorgestellten gradientenbasierten Blocking-Maße zutrifft. Es kann der Fall eintreten, dass bestimmte Deblocking-Verfahren zu Messwerten unterhalb des theoretischen Optimums führen, welches durch Gleichheit der Gradientensumme an den Blockgrenzen und der Gradientensummen an den anderen Rasterlagen gegeben ist. In diesem Fall können die Messwerte einen Grad an Qualitätsverbesserung signalisieren, der höher ausfällt als er von menschlichen Betrachtern wahrgenommen werden würde. Wenn Gradienten über Blockgrenzen hinweg durch eine Filterung bzw. Artefaktreduktion kleiner werden als die übrigen Gradienten, kann dies theoretisch nicht mehr zur Verbesserung des Bildeindrucks beitragen. Bei extremen Verkleinerungen kann sogar wieder eine Verschlechterung der Bildqualität durch eine zu starke Glättungswirkung der verwendeten Filter eintreten. Die beschriebene Problematik ist beim Einsatz der skizzierten Qualitätsmaße zu berücksichtigen. Bei relativ niedrigen Datenraten und somit vielen Blockartefakten im zu bewertenden Bildmaterial sind jedoch kaum Schwierigkeiten zu erwarten.

Weitere gradientenbasierte Verfahren zur Blocking-Messung finden sich in [Sut03, Plo89]. Die letztgenannte Quelle dürfte eine der ersten Beschreibungen eines gradientenbasierten Blocking-Maßes sein.

## 4.3 Ein- und Mehrkanalmodelle der visuellen Wahrnehmung

Wie bereits in der Übersicht zu diesem Kapitel dargelegt, bilden Qualitätsmaße auf Basis von HVS-Modellen eine spezielle Klasse von Qualitätsbewertungsverfahren. Da das menschliche visuelle System äußerst komplex ist, schlägt sich diese Komplexität auch in den zu der Klasse gehörenden Ein- und Mehrkanalmodellen nieder. Da HVS-Modelle nicht im Fokus dieser Arbeit sind, soll im Folgenden nur auf einige grundlegende Merkmale solcher Modelle eingegangen werden. Dazu eignet sich das klar gegliederte Einkanalmodell der visuellen Wahrnehmung nach [Fec00], welches auf Untersuchungen in [Bud72, Lim79] aufbaut. Es dient der Modellierung der menschlichen Verarbeitung visueller Informationen mit dem Ziel, einen objektiven Wert für die absolute Qualität eines komprimierten Bildes bzw. einer komprimierten Bildsequenz zu ermitteln, der möglichst exakt mit dem mittleren

subjektiven Eindruck von menschlichen Betrachtern übereinstimmt. Bei der Entwicklung wurde besonders berücksichtigt, dass sichtbare Codierartefakte Reize des HVS im überschweligen Bereich darstellen. Abbildung 4.3 zeigt den prinzipiellen Aufbau des Modells an einem Blockschaltbild.

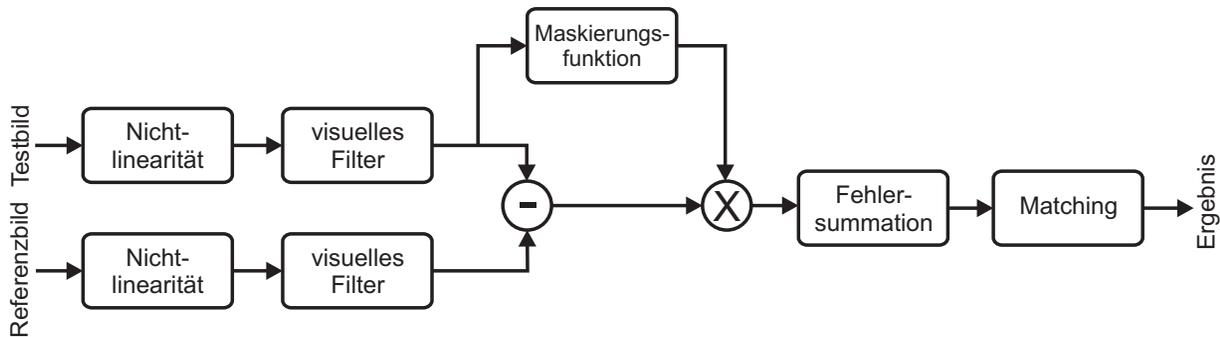


Abbildung 4.3: Einkanalmodell - Blockschaltbild

Das Einkanalmodell benötigt zur Ermittlung eines Qualitätswertes neben dem zu bewertenden Bild ein Referenzbild zum Vergleich. Ähnlich wie beim MSE und PSNR ist also grundsätzlich ein Zugriff auf Referenz-Bildmaterial notwendig. Die Funktionsblöcke auf der linken Seite der Abbildung, welche mit *Nichtlinearität* gekennzeichnet sind, tragen der Eigenschaft des menschlichen Auges Rechnung, Luminanzdifferenzen in unterschiedlichen Grundhelligkeitsbereichen verschieden wahrzunehmen. Das Verhalten des Auges ist also in Bezug auf die Helligkeitswahrnehmung nichtlinear, was im Einkanalmodell entsprechend berücksichtigt wird.

Die Funktionsblöcke, gekennzeichnet mit *visuelles Filter*, dienen der Berücksichtigung der frequenzabhängigen Störwahrnehmung des menschlichen Auges. Diese kann durch einen tiefpassartigen Frequenzgang des Filters hinreichend genau approximiert werden. Für nähere Informationen sei auf [Fec00] verwiesen.

Wie dem Blockschaltbild entnommen werden kann, wird nach dem visuellen Filter im Verarbeitungszweig des Testbildes eine Maskierungsfunktion ermittelt. Grundsätzlich ist auch die Ermittlung der Maskierungsfunktion im Verarbeitungszweig des Referenzbildes möglich. Welcher Zweig dafür am besten genutzt wird, hängt laut Untersuchungen in [Fec00] von der Anwendung ab (z. B. von dem zu bewertenden Videocodierverfahren). Die Maskierung stellt sicher, dass Störanteile, die im menschlichen visuellen System maskiert werden, keine Auswirkung auf den Qualitätswert des zu bewertenden Bildes haben. Maskierungseffekte werden hauptsächlich durch Signalsprünge ausgelöst und betreffen die umgebende Nachbarschaft, sind also in ihrer Ausdehnung lokal begrenzt. Die Detektion von Signalsprüngen erfolgt im hier betrachteten Einkanalmodell mit Hilfe einer Aktivitätsmessung auf Basis der Auswertung lokaler Pixeldifferenzen. Es wird eine Maskierungsfunktion

gebildet, die im Bereich von Signalsprüngen sehr kleine Werte und entfernt von Signalsprüngen den Wert Eins annimmt.

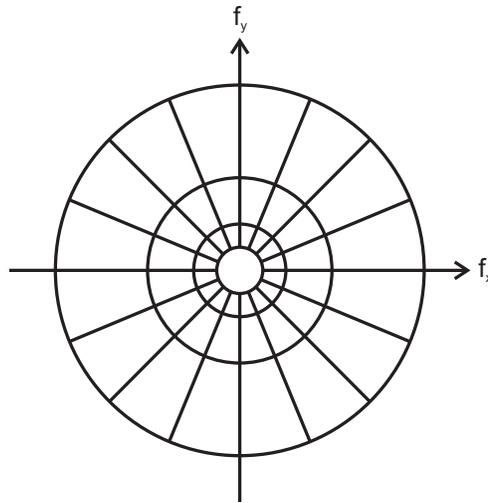
Im Zentrum des Blockschaltbildes erfolgt eine Differenzbildung zwischen Original- und Testbild, so dass nur noch Unterschiede zwischen den Bildern bzw. die Fehler im Testbild übrig bleiben, die den Störungen im zu beurteilenden Bild entsprechen. Maskierungseffekte werden durch Multiplikation des Differenzbildes mit der bereits erzeugten Maskierungsfunktion berücksichtigt.

Schließlich wird eine (gegebenenfalls nach Potenzierung) pixelweise Summation der Fehler und eine Normierung des Ergebnisses auf die Anzahl an untersuchten Pixelpositionen durchgeführt (anschließend gegebenenfalls inverse Potenzierung), wobei es vorgesehen ist, dies innerhalb von Fenstern begrenzter Größe zu tun und letztlich nur das Fenster für die weitere Verarbeitung auszuwählen, in dem die maximale Fehlersumme (normiert auf die Elementanzahl im Fenster) aufgetreten ist. Ganz rechts im Blockschaltbild befindet sich der Block *Matching*. Darunter ist eine Anpassung des ermittelten Fehlerwertes an die Eindrücke menschlicher Betrachter zu verstehen. Es wird eine Impairment-Skala verwendet, die eine Einordnung der Gesamtqualität eines Bildes oder einer Bildsequenz in Form eines MOS-Wertes (Mean Opinion Score) ermöglicht. Zur Ermittlung der Matching-Funktion ist die Durchführung von subjektiven Tests notwendig. Basis des Matchings ist eine lineare Funktion.

Für weitere Details zum vorgestellten Einkanalmodell der visuellen Wahrnehmung sei auf [Fec00] verwiesen. Auch für einige Untersuchungen in dieser Arbeit wurde ein derartiges Einkanalmodell eingesetzt. Anmerkungen dazu finden sich in Anhang C.

Ein weiteres Verfahren, das im weitesten Sinne als Einkanalmodell der visuellen Wahrnehmung betrachtet werden kann, wird in [KK95] vorgestellt. Es ist im Vergleich zu dem bereits beschriebenen Einkanalmodell stärker auf die Erfassung von Blocking-Störungen ausgelegt. Nahezu alle Komponenten des Modells nach [Fec00] sind auch in diesem Modell in ähnlicher Form enthalten. Die größten Unterschiede bestehen in der Konzentration auf Bildbereiche, die vermutlich Blockartefakte enthalten, wofür eine Kantendetektion zum Einsatz kommt, und in der Behandlung von Maskierungseffekten. Da eine getrennte Behandlung vertikal und horizontal ausgerichteter Störungen mit anschließender Kombination der entsprechenden Qualitätswerte erfolgt, könnte das Verfahren auch in die Klasse der Faktorenmodelle eingeordnet werden.

Mehrkanalmodelle der visuellen Wahrnehmung haben das Ziel, die Effekte der Störwahrnehmung des Menschen umfassender als Einkanalmodelle zu modellieren und dabei insbesondere Frequenz- und Orientierungsabhängigkeiten zu berücksichtigen. Dies geht mit einer äußerst hohen Komplexität und mit erheblichem Rechenaufwand einher. Die Komponenten der vorgestellten Einkanalmodelle finden sich typischerweise auch in Mehrkanalmodellen wieder. Der entscheidende Unterschied ist die Behandlung von Maskierungseffekten in mehreren Teilbändern. Dazu kann eine Bandaufspaltung eingesetzt werden, wie sie in Abbildung 4.4 veranschaulicht ist. Die Aufspaltung des Bildsignals er-



**Abbildung 4.4:** Bandaufspaltung in Frequenz- und Orientierungsbänder

folgt sowohl in Frequenz- als auch in Orientierungsbänder, deren Anzahlen je nach Verfahren unterschiedlich sein können. Grundlagen zur beschriebenen Art der Bandaufspaltung und zu Maskierungseffekten in den Teilbändern finden sich in [Dal93]. In [Fec00] wird ein Mehrkanalmodell bezüglich der Eignung zur Schätzung der von menschlichen Betrachtern wahrgenommenen absoluten Qualität komprimierter Bildsequenzen untersucht. Es erweist sich als problematisch, das Modell so anzupassen, dass eine akkurate Schätzung der Qualität von wenig gestörten bis hin zu stark gestörten Bildsequenzen möglich ist. Das untersuchte Modell wird schließlich als geeigneter zur Analyse schwelennahe Störungen (wie z. B. Rauschen niedriger Intensität) als zur Analyse überschwelliger Reize, zu denen Codierartefakte zählen, angesehen. In [YWWC02] wird ein Mehrkanalmodell vorgestellt, das nicht auf die Schätzung der absoluten Gesamtbildqualität, sondern auf die Schätzung der absoluten Stärke von Blockartefakten in unterschiedlich stark komprimiertem Bildmaterial optimiert ist. Mit diesem Modell können laut der veröffentlichten Untersuchungsergebnisse akkurate Schätzwerte ermittelt werden.

## 4.4 Faktorenmodelle

Einen für diese Arbeit besonders relevanten Ansatz zur Bewertung von Bildqualität stellen Verfahren dar, die aus einem Bild oder aus einer Bildsequenz verschiedene Arten oder Anteile von Merkmalen oder Störungen extrahieren und geeignet zu einem Qualitätsmaß verknüpfen. Bei solchen Ansätzen ist die Herangehensweise also nicht durch die Modellierung der Reaktion des HVS auf Störungen charakterisiert, sondern durch die gezielte Detektion oder Extraktion und Bewertung von Qualitätsmerkmalen. Ein typisches Verfah-

ren dieser Art wird in [HMM97] vorgestellt und ist mit seinen wesentlichen Komponenten in Abbildung 4.5 dargestellt.

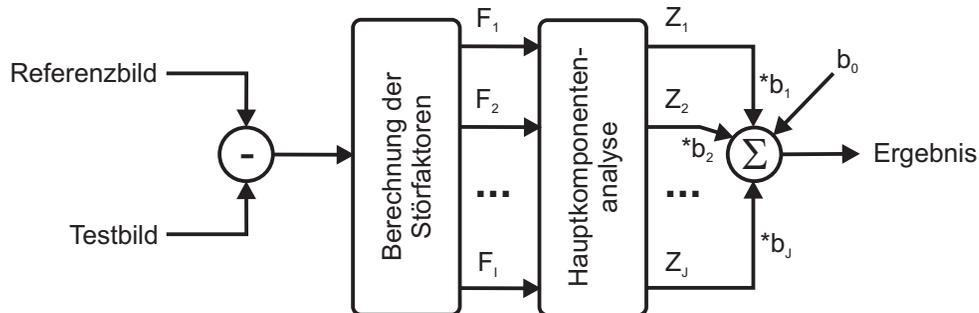


Abbildung 4.5: Faktorenmodell nach [HMM97]

Zentrale Komponente ist die Berechnung von Störfaktoren. Es kann eine beliebige Anzahl objektiver Maßzahlen verwendet werden, die z. B. so gestaltet sein können, dass sie mit der Stärke von bestimmten Arten von Störungen korrespondieren. In dem abgebildeten Verfahren werden die Störfaktoren auf Basis der Differenz aus Referenzbild und Testbild berechnet. Denkbar sind jedoch auch anders definierte Störfaktoren, die beispielsweise nur das Testbild oder beide Bilder als Eingangsdaten verwenden. Die Störfaktoren  $F_i$  werden mit Hilfe der Hauptkomponentenanalyse in die Faktoren bzw. Komponenten  $Z_i$  transformiert, welche orthogonal zueinander sind. Dadurch werden Abhängigkeiten zwischen den einzelnen Faktoren beseitigt und die Anzahl der relevanten Faktoren wird in vielen Fällen verringert. Problematisch ist, dass durch die Transformation die Interpretierbarkeit der neuen Faktoren  $Z_i$  gegenüber den ursprünglichen Faktoren deutlich erschwert werden kann. Die Hauptkomponentenanalyse muss nicht grundsätzlich durchgeführt werden. Andere Autoren (z. B. [Fec00, Frä98]) schlagen ähnliche Verfahren vor, die darauf verzichten.

Wichtig an dem Verfahren in Abbildung 4.5 ist, dass die einzelnen Störfaktoren oder die durch Transformation aus ihnen gebildeten Faktoren zu einem Gesamtqualitätsmaß kombiniert werden. Dies erfolgt durch Gewichtung und Addition der einzelnen Faktoren. Das Ergebnis bzw. die gemessene Bildqualität ist also eine Linearkombination aus einzelnen (Stör-)Faktoren. In [Fec00] wird für solche Verfahren zur Bildqualitätsbewertung der Begriff „Faktorenmodell“ benutzt, der auch in dieser Arbeit Verwendung findet. Das dort vorgestellte Faktorenmodell soll zunächst exemplarisch betrachtet werden, um einen detaillierten Einblick in die Funktionsweise eines Faktorenmodells zu liefern. In Abbildung 4.6 ist das Blockschaltbild des Verfahrens dargestellt.

Wie dem Schaltbild entnommen werden kann, benötigt auch dieser Ansatz neben einem zu bewertenden Bild den Zugriff auf ein Referenzbild. Das Referenzbild wird zunächst zur Durchführung einer Bildsegmentierung verwendet. Dabei erfolgt die Einteilung des Bildes in Kanten-, Textur- und Flächenbereiche. Jeder Bildpunkt wird nur einer dieser Klassen

oder gar nicht zugeordnet. Die Bildsegmentierung wird für die einzelnen Klassen sequenziell durchgeführt. Sie beinhaltet eine Verknüpfung von verschiedenen Operatoren zur Bildverarbeitung. Darunter befinden sich z. B. Sobel-Operatoren, Medianfilter, Erosionsoperatoren und Schwellwertentscheidungen. Für Details zur Segmentierung sei auf [Fec00] verwiesen.

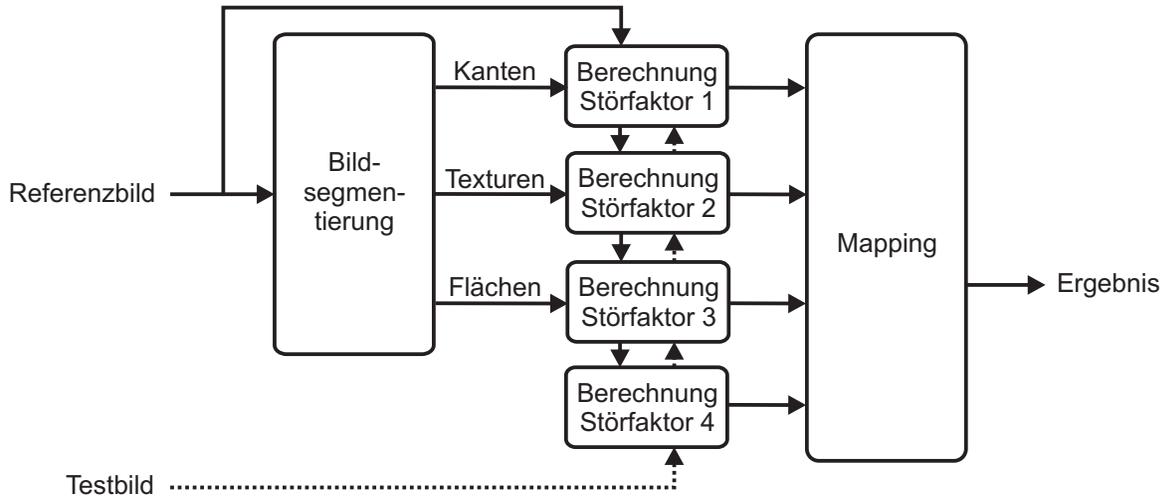


Abbildung 4.6: Faktorenmodell nach [Fec00]

Die Einteilung von Bildpunkten in die bereits genannten Klassen wird damit begründet, dass Fehler in den einzelnen Klassen vom menschlichen Auge unterschiedlich stark wahrgenommen werden. Es wird davon ausgegangen, dass gestörte Kanten bzw. Konturverläufe besonders auffällig sind. Ebenso sind Flächenbereiche für die Wahrnehmung von Störungen kritisch, da dort keinerlei Maskierungseffekte auftreten. In Texturbereichen kann hingegen von einer Maskierung von Störungen durch Texturen ausgegangen werden. Die Ermittlung einer Maskierungsfunktion, ähnlich der im bereits beschriebenen Einkanalmodell, ist in dem Faktorenmodell nach [Fec00] nicht vorgesehen.

In jedem Bildsegment wird als Störungsmaß  $w_i$  die Wurzel aus der Summe der quadratischen Abweichung zwischen Referenzbildpunkten  $s_{ref}(x, y)$  und Testbildpunkten  $s(x, y)$  berechnet und auf die Anzahl der in der Klasse  $i$  vorhandenen Bildpunkte  $N_i$  normiert, wobei  $B_i$  die Menge der zu einer Klasse gehörenden Bildpunkte bezeichnet:

$$w_i = \frac{1}{N_i} \sqrt{\sum_{(x,y) \in B_i} (s_{ref}(x, y) - s(x, y))^2}. \quad (4.13)$$

Als vierter Störfaktor  $w_4$  wird entsprechend die mittlere quadratische Abweichung zwischen dem gesamten Referenz- und dem Testbild hinzugezogen. Für jedes Bild werden also vier Störfaktoren berechnet, die jeweils eine Aussage über die Größe von Störungen

in bestimmten Bildbereichen machen. Diese vier Störmaße werden nun mit Hilfe einer Mapping-Funktion zu einem einzelnen Qualitätsmaß verrechnet und auf eine MOS-Skala abgebildet. Ziel ist eine möglichst gute Übereinstimmung des berechneten Qualitätsmaßes mit den Eindrücken menschlicher Betrachter, die in einem subjektiven Test ermittelt werden müssen. In [Fec00] wird eine lineare Verknüpfung zwischen den einzelnen Störmaßen und dem Gesamtqualitätsmaß  $\hat{q}$  vorgenommen:

$$\hat{q} = b_0 + b_1 \cdot w_1 + b_2 \cdot w_2 + b_3 \cdot w_3 + b_4 \cdot w_4. \quad (4.14)$$

Die Mapping-Funktion nimmt also eine Addition unterschiedlich gewichteter Stör- bzw. Qualitätsfaktoren vor. Die Gewichtungsfaktoren  $b_i$  werden mit Hilfe der multiplen linearen Regression ermittelt. Grundvoraussetzung für die Anwendbarkeit des vorgestellten Faktorenmodells auf ein Bild ist die Bestimmbarkeit der einzelnen Störfaktoren. Das bedeutet, dass in einem zu bewertenden Bild alle vorgesehenen Bildsegmente vorhanden sein sollten.

Das Prinzip des Faktorenmodells ermöglicht eine Anpassung der Qualitätsmessung an verschiedene Störungen oder unterschiedliche Ausprägungen von Störungen. Dazu kann es ausreichen, die einzelnen Gewichtungsfaktoren zu variieren. In [Fec00] wird das vorgestellte Faktorenmodell zur absoluten Qualitätsbewertung von unterschiedlich stark komprimierten Bildfolgen erfolgreich eingesetzt. Es kann eine hohe Korrelation der geschätzten Qualitätswerte mit den subjektiven Eindrücken von Betrachtern nachgewiesen werden. Dies gilt auch für verschiedene Kompressionsverfahren (MPEG-2, Wavelet-Codierung, fraktale Codierung), die grundsätzlich unterschiedliche Arten von Störungen hervorrufen. Für jedes Kompressionsverfahren ist nur ein angepasster Satz an Gewichtungsfaktoren bzw. Regressionskoeffizienten notwendig.

Es existieren weitere Ansätze zur Bildqualitätsbewertung auf Basis von Faktorenmodellen, die ebenfalls die Bewertung der absoluten Bildqualität von codierten Einzelbildern oder Bildsequenzen, zum Teil zum Vergleich oder zur Optimierung von Videokompressionsverfahren, anstreben. Diese Verfahren ermitteln - ähnlich wie das vorgestellte Verfahren - jeweils einen einzigen Qualitätswert für die Gesamtbildqualität auf einer geeigneten Skala. Die maßgeblichen Unterschiede der Verfahren bestehen in der Anzahl und der Art der einzelnen Stör- bzw. Qualitätsfaktoren. In [Frä98] werden z. B. drei Merkmale im Vergleich zwischen Referenz- und Testbild analysiert. Eines der Merkmale betrifft den mittleren lokalen Kontrastunterschied, ein weiteres Veränderungen im Bereich von Konturen und das dritte die Anzahl von Grauwerten in Analysefenstern der Größe 3x3 Pixel. In [HN94] werden verschiedene Faktorenmodelle mit zwei oder drei integrierten Qualitätsmaßen vorgestellt. Eines der Maße ist eine gradientenbasierte Blocking-Analyse und ein anderes ein Maß zur Bewertung von Mosquito- bzw. Ringing-Störungen. Daneben wird ein Störungsmaß verwendet, das Unterschiede zwischen Referenz- und Originalbild analysiert, wobei die lokale Bildaktivität und der Frequenzgang des HVS berücksichtigt werden. Das Faktorenmodell in [CG02] sieht die Verwendung von zwei Qualitätsmaßen vor. Das eine ist ein

Blocking-Maß im Frequenzbereich, während das andere die Ähnlichkeit von Konturen im Referenz- und im Testbild bewertet. Die Kombination der beiden Maße erfolgt über eine nichtlineare Funktion. Eine Besonderheit weist das in [TG00] vorgestellte Faktorenmodell zur Qualitätsschätzung von MPEG-Sequenzen auf. Es beinhaltet als einen Qualitätsfaktor ein Mehrkanalmodell der visuellen Wahrnehmung zur Unschärfemessung. Daneben führt es eine Blocking-Detektion im Frequenzbereich durch. Eines der ersten Faktorenmodelle zur Bildqualitätsbewertung dürfte der in [Miy88] präsentierte Ansatz sein.

## 4.5 Bewertung der Wirksamkeit von Deblocking-Verfahren

Die meisten der in diesem Kapitel vorgestellten Qualitätsbewertungsansätze wurden dazu entwickelt, die absolute Qualität von codierten - und somit mit Codierartefakten versehenen - Bildern oder Bildsequenzen über recht große Qualitätsbereiche zu schätzen, oder sie wurden zumindest erfolgreich zu diesem Zweck eingesetzt. Die Bewertungsaufgabe, die es in dieser Arbeit zu bewältigen gilt, erscheint grundsätzlich damit verwandt. Es geht um die Bewertung der Wirksamkeit von Deblocking-Verfahren in Bezug auf eine Bildqualitätsverbesserung in Bildmaterial mit (deutlich) sichtbaren Blockartefakten. Insofern ist es überlegenswert, ob und wie die vorgestellten Ansätze für die Untersuchungen in dieser Arbeit eingesetzt werden können. Naheliegend ist - wie bereits angedeutet - die Ermittlung der Qualität eines Bildes vor und nach Anwendung eines Deblocking-Verfahrens und ein Vergleich der entsprechenden Qualitätswerte z. B. über eine Differenzbildung. Tatsächlich wurden einige der vorgestellten Verfahren in ähnlicher Weise auch schon zur Performance-Bewertung von Deblocking-Verfahren eingesetzt. Dies betrifft im Wesentlichen die gradientenbasierten Ansätze und den PSNR [CWQ01, RK00, Sut03]. An späterer Stelle in dieser Arbeit werden einige der in diesem Kapitel vorgestellten Ansätze aufgegriffen und speziell zur Bewertung der Wirksamkeit von Deblocking-Verfahren bezogen auf eine Bildqualitätsverbesserung eingesetzt. Es sei vorweggenommen, dass die durchgeführten Untersuchungen ergaben, dass die untersuchten Ansätze bei dieser Aufgabe, bei der es auf die Erfassung von zum Teil recht kleinen Veränderungen bestimmter Störmerkmale ankommt, Schwächen bzw. Probleme aufweisen. Insgesamt konnte eine nur mäßige Übereinstimmung der entsprechenden Qualitätsschätzwerte mit den Eindrücken menschlicher Betrachter nachgewiesen werden.

Um zu besseren Ergebnissen bezüglich der Übereinstimmung mit menschlichen Eindrücken zu kommen, wurden im Rahmen der Arbeit neue Qualitätsbewertungsansätze speziell zur Bewertung der Wirksamkeit von Deblocking-Verfahren bei der Verbesserung der Qualität sichtbar gestörter Einzelbilder entwickelt. Diese in Kapitel 8 detailliert beschriebenen Ansätze basieren auf einem in demselben Kapitel vorgestellten allgemeinen Faktoren-

modell und kombinieren mehrere Einzelmaße, welche bestimmte Bild- oder Störmerkmale, die im Zusammenhang mit der Verringerung von Blockartefakten von Bedeutung sind, analysieren bzw. mit hoher Genauigkeit erfassen können. Bei der Entwicklung der neuen Qualitätsbewertungsansätze wurde beachtet, dass verschiedene Deblocking-Verfahren insgesamt oft nur verhältnismäßig kleine Qualitätsveränderungen in einem Bild hervorrufen oder das Bildmaterial nur sehr selektiv verändern, trotzdem für den menschlichen Betrachter aber unterschiedlich gut oder schlecht bei der Bewältigung ihrer Aufgabe erscheinen.

Die neuen Qualitätsbewertungsansätze zeigen auf einer Vergleichsskala (DMOS) direkt an, inwieweit ein Deblocking-Verfahren die Qualität eines Bildes verbessern kann. Dazu muss der Zugriff auf Referenz- und Testbildmaterial, in diesem Fall auf ungestörte, gestörte sowie auf gestörte und nachgefilterte Bilder möglich sein. Es werden nur rein örtliche Merkmale von Codierartefakten und die örtlichen Auswirkungen von Verfahren zu deren Beseitigung berücksichtigt, was zur Begrenzung des Untersuchungsumfangs notwendig war. Des Weiteren werden ausschließlich Luminanzinformationen ausgewertet, da diese - wie in Kapitel 2 bereits erwähnt - verglichen mit den Chrominanzinformationen weit bedeutender für das Auftreten und die Sichtbarkeit von Blockartefakten sind. Ergänzend sei erwähnt, dass Untersuchungen in [Fec00] gezeigt haben, dass eine Hinzunahme der Analyse von Chrominanzinformationen in die dort vorgestellten Verfahren zur Bewertung der absoluten Qualität MPEG-codierter Sequenzen kaum zu Verbesserungen führte, da aus den Chrominanzkanälen nahezu keine zusätzlichen relevanten Informationen zu gewinnen waren. Aufgrund der Verwandtschaft der Untersuchungen, musste auch hier mit ähnlichen Tendenzen gerechnet werden.

Es sei ebenfalls vorweggenommen, dass in Kapitel 9 gezeigt wird, dass mit den neuen Ansätzen zur Bewertung der Wirksamkeit von Deblocking-Verfahren Schätzungen möglich sind, die eine erheblich höhere Übereinstimmung mit den Eindrücken menschlicher Betrachter aufweisen, als Schätzungen von im Rahmen der Untersuchungen zu diesem Zweck eingesetzt (aber nicht dazu entwickelten) Vergleichsverfahren. Eine akkurate und stabile Wirksamkeitsbewertung von Deblocking-Verfahren in einem breiten Spektrum an Bildmaterial ist somit möglich.



# 5 Stand der Technik in der Artefaktreduktion

In diesem Kapitel werden verschiedene aus der Literatur bekannte Verfahren zur Reduktion von Blockartefakten, die für die in dieser Arbeit durchgeführten Untersuchungen relevant sind, vorgestellt. Dabei geht es ausschließlich um Blockartefakte als Resultat von Quantisierungsverlusten in blockbasierten Video-Codecs, also um eine Variante typischer Codierartefakte. Zwar können auch Übertragungsstörungen in digitalen Videoübertragungssystemen zu blockartigen Störungen führen, allerdings weisen diese eine andere Charakteristik auf als Codierartefakte (siehe Kapitel 2) und werden dementsprechend auch mit anderen bzw. speziellen Verfahren, so genannten Concealment-Verfahren, verringert bzw. beseitigt. Zum Thema Concealment-Verfahren sei auf die Literatur, z. B. [FK06, MK03, RB00, AB00], verwiesen. In Abbildung 5.1 ist ohne Anspruch auf Vollständigkeit ein Vorschlag zur Einteilung unterschiedlicher Ansätze zur Reduktion von Codierartefakten in Klassen gegeben, der auch von anderen Autoren in ähnlicher Form gewählt wird [YG97, SK98, SK00].

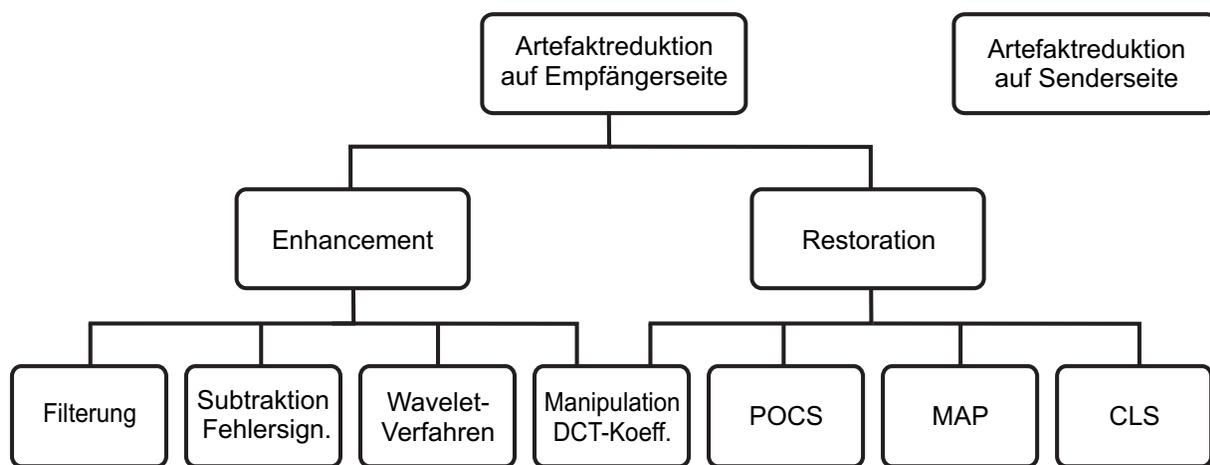


Abbildung 5.1: Überblick über Artefaktreduktionsverfahren

Grundsätzlich sind Maßnahmen zur Reduktion von Codierartefakten auf Sender- und Empfängerseite zu unterscheiden. Maßnahmen zur Artefaktreduktion auf Senderseite sind nicht Teil der Untersuchungen in dieser Arbeit. Dennoch soll erwähnt werden, dass beispielsweise eine senderseitige Rauschreduktion vor der Codierung dazu beiträgt, dass die

Bildqualität auf Empfängerseite bei gleicher Datenrate gesteigert werden kann, da das Rauschen nicht codiert und übertragen werden muss und somit mehr Kapazität für die relevanten Nutzdaten zur Verfügung steht. Neben dem Rauschen können auch bestimmte weniger relevante Bildinhalte vor der Codierung aus dem Nutzsignal entfernt werden, wodurch ebenfalls mehr Kapazität für relevantere Nutzdatenanteile zur Verfügung steht und die Bildqualität auf Empfängerseite insgesamt verbessert werden kann [SK00].

Die empfängerseitigen Artefaktreduktionsansätze lassen sich in Enhancement- und in Restoration-Techniken aufteilen, worauf in den folgenden Abschnitten näher eingegangen wird.

## 5.1 Restoration-Verfahren

Das Grundprinzip der Restoration-Verfahren (auch als Reconstruction- oder als Recovery-Verfahren bezeichnet) ist eine möglichst gute Rekonstruktion eines codierten Bildes ohne sichtbare Störungen und Artefakte bzw. eine möglichst genaue Schätzung des Originalbildes aus den codierten Daten. Über die am Decoder eintreffenden Daten (Koeffizienten, Quantisierungsinformationen etc.) hinaus, können vorab verfügbare Informationen über die im Bildmaterial vorhandenen Störungen oder die Eigenschaften des Kompressionsverfahrens zur Rekonstruktion genutzt werden.

Viele Restoration-Verfahren erfordern iterative oder rekursive Lösungen mit einem oft sehr hohen Rechenaufwand und sind insofern derzeit kaum für den Einsatz in Echtzeit-Systemen geeignet. Des Weiteren ist aufgrund der Art der benötigten Daten oft eine Implementierung in enger Verzahnung mit dem Decoder notwendig, was in vielen denkbaren Systemumgebungen eine Einschränkung darstellt oder den Einsatz ganz ausschließt (z. B. in vielen Endgeräten). Aus den genannten Gründen sind Restoration-Verfahren zur Verringerung von Blockartefakten vergleichsweise wenig praxisrelevant und liegen daher nicht im Fokus dieser Arbeit. Sie werden deshalb nicht im Detail betrachtet. Dennoch soll der Vollständigkeit halber ein kurzer Überblick über einige Restoration-Ansätze gegeben werden.

Eine Variante der Restoration-Verfahren sind die so genannten POCS-Verfahren (Projection Onto Convex Sets). Die Codierung und anschließende Decodierung eines Bildes bewirkt beim Einsatz von nicht verlustfreien Codecs, dass das decodierte Bild  $y$  nicht dem Originalbild  $x$  entspricht. Verantwortlich dafür sind Quantisierungsverluste. Es existiert eine Menge  $X$  von Bildern, die mögliche Originalbilder zum beobachteten bzw. decodierten Bild  $y$  enthält. Die Idee der POCS-Verfahren besteht in einer iterativen wechselseitigen Projektion des decodierten Bildsignals  $y$  auf mehrere konvexe Teilmengen von  $X$ . Auf diese Weise kann das Originalsignal  $x$  in mehreren Schritten bzw. Iterationen angenähert werden. Um mehrere konvexe Teilmengen von  $X$  zu definieren, sind Bedingungen (Constraints) notwendig, die die Teilmengen begrenzen. Mögliche Bedingungen betreffen beispielsweise den

Wertebereich der Transformationskoeffizienten, die in bestimmten bekannten Quantisierungsintervallen liegen müssen, den erlaubten Intensitätsbereich der einzelnen Bildpunkte oder bestimmte Kontrast- bzw. Schärfeeinschränkungen in Blockgrenzbereichen. In der Literatur findet man weitere Möglichkeiten und Vorschläge, a priori Wissen in geeignete Constraints umzuformulieren. Dabei ist es wichtig, dass die Constraints sinnvoll mit dem jeweiligen Anwendungszweck in Einklang gebracht werden. [Wen06, SK98] beschreiben das grundsätzliche Vorgehen bei POCS-Methoden. Details zu einzelnen Verfahren finden sich z. B. in [GAM02, YG97].

Eine weitere Variante der Restoration-Verfahren stellen die CLS-Verfahren (Constrained Least Squares) dar. Dabei wird typischerweise von der Modellierung eines decodierten Bildes  $y$  in der Form  $y = H \cdot x$  oder  $y = H \cdot x + n$  mit einem Verzerrungsoperator  $H$ , dem Originalbild  $x$  und einem Fehlersignal  $n$  ausgegangen [KP98, BK97]. Das Originalsignal soll möglichst exakt geschätzt werden, wofür ein Optimierungsansatz aufgestellt wird. Grundsätzlich gilt es, folgenden Ausdruck zu minimieren:

$$\|y - H\hat{x}\|^2. \quad (5.1)$$

$\hat{x}$  ist das geschätzte Originalbild. Da der Verzerrungsoperator  $H$  in der Praxis nichtlinear und ortsvariant ist [KP98], existiert in der Regel keine geschlossene Lösung für den Optimierungsansatz. Dies erfordert die Berücksichtigung von geeigneten Nebenbedingungen (Constraints) der Form  $\|A\hat{x}\|^2 \leq \varepsilon^2$ . Im Anwendungsfall der Reduktion von Codierartefakten wird für den Operator  $A$  oft ein blockadaptiver Hochpassoperator gewählt, wodurch zusammen mit der Schwelle  $\varepsilon^2$  eine Beschränkung hoher Frequenzanteile in Blockgrenzbereichen vorgenommen wird. In vielen Fällen (z. B. in [HKJP01, BK97]) wird der Optimierungsansatz und die Nebenbedingung in eine Funktion der Form

$$f(\hat{x}) = \|y - H\hat{x}\|^2 + \lambda \|A\hat{x}\|^2 \quad (5.2)$$

mit dem Anpassungsparameter  $\lambda$  überführt, welche minimiert werden muss. Weitere Informationen zu CLS-Verfahren finden sich auch in [SK98, Kau00].

Auch MAP-Verfahren (Maximum A Posteriori) zählen zu den Restoration-Verfahren. Die MAP-Schätzung  $\hat{x}$  des Originalbildes  $x$  ergibt sich bei vorliegendem gestörten Bild  $y$  zu

$$\hat{x} = \arg \max_x p(x|y). \quad (5.3)$$

$\hat{x}$  entspricht dem Bild aus dem Lösungsraum, das mit der größten a posteriori Wahrscheinlichkeit zum Bild  $y$  komprimiert wurde. Durch Anwendung des Bayes-Theorems und mit Hilfe der Log-Likelihood-Funktion lässt sich der Ausdruck in

$$\hat{x} = \arg \max_x \{ \log p(y|x) + \log p(x) - \log p(y) \} \quad (5.4)$$

überführen.  $p(y)$  kann vernachlässigt werden, da es bezüglich des Optimierungsparameters  $x$  konstant ist.  $p(y|x)$  kann ebenfalls vernachlässigt werden, da davon auszugehen ist, dass ein Codierverfahren bei konstanten Parametern aus dem Bild  $x$  immer dasselbe Bild  $y$  erzeugt. Dadurch ergibt sich folgende Vereinfachung:

$$\hat{x} = \arg \min_{x \in X} \{-\log p(x)\}. \quad (5.5)$$

$X$  ist die Menge aller Bilder, die durch das verwendete Codierverfahren zu  $y$  verarbeitet werden. Nun muss die a priori Wahrscheinlichkeit  $p(x)$  geeignet modelliert werden, wobei das spezielle Rekonstruktionsproblem, hier die Verringerung von Codierartefakten, zu berücksichtigen ist. Für Details und konkrete MAP-Verfahren sei auf [LCPH96, OS95] verwiesen. Die grundsätzliche Vorgehensweise bei der Anwendung von MAP-Verfahren wird in [SK98] beschrieben.

Eine weitere Möglichkeit, Codierartefakte zu verringern, die im weitesten Sinne ebenfalls zu den Restoration-Ansätzen gezählt werden kann, wird in [Nis88] vorgestellt. Allerdings ist das Ziel nicht eine Rekonstruktion von Bilddaten im Ortsbereich, sondern eine blockweise Schätzung bzw. Rekonstruktion von durch Quantisierungseffekte verloren gegangenen AC DCT-Koeffizienten. Die Schätzung erfolgt ausgehend von den DC-Koeffizienten umliegender Blöcke über die Lösung einfacher Gleichungen. Es können nur die niederfrequentesten AC-Koeffizienten geschätzt werden. Das Verfahren verspricht eine Verbesserung der Bildqualität somit nur in recht homogenen Bereichen. Eine weiterentwickelte Variante desselben Prinzips zur Reduktion von Codierartefakten findet sich in [SCA99].

## 5.2 Enhancement-Verfahren

Die Grundidee von Enhancement-Verfahren zur Artefaktreduktion ist die Anwendung geeigneter Maßnahmen, um die Sichtbarkeit von Codierartefakten im Bildmaterial zu verringern oder zu beseitigen. Dies muss nicht bedeuten, dass die uncodierten Originalbilder wiederhergestellt werden müssen, sondern dass die Bilddaten so zu verändern sind, dass die Ausprägung von Codierartefakten in die Nähe bzw. unter die Wahrnehmungsschwelle des HVS gebracht wird. Dazu können z. B. Filterungen oder die Subtraktion von Fehler-signalen vom gestörten Bild geeignet sein.

In den folgenden Abschnitten wird ein Überblick über in der Literatur vorgestellte Enhancement-Ansätze gegeben. Der Schwerpunkt liegt bei Verfahren zur Reduktion von Blockartefakten, welche für eine Implementierung als Postprocessing-Verfahren gut geeignet sind. Allerdings befinden sich in der Übersicht auch Verfahren, die eine Auskopplung von Quantisierungsinformationen aus dem Decoder erfordern und somit nicht ohne Weiteres in jeder Systemumgebung eingesetzt werden können. Auf so genannte Loop-Filter

[LJL<sup>+</sup>03, HCH<sup>+</sup>03] wird hier nicht explizit eingegangen, da sie eine Integration in die Prädiktionsschleife von Encoder und Decoder erfordern und in existierenden Standards, die keine Loop-Filter vorsehen, nicht ohne erhebliche Eingriffe in den bzw. Änderungen am Standard verwendbar sind. Grundsätzlich wäre es aber denkbar und möglich, die vorgestellten Verfahren selbst als Loop-Filter einzusetzen. Bei bereits vorhandenem Loop-Filter (z. B. in H.264-Systemen) ist ihr Einsatz zusätzlich möglich, um verbliebene Artefakte zu verringern. Ebenfalls nicht näher behandelt werden die in Abbildung 5.1 vermerkten Enhancement-Verfahren auf Basis der Manipulation von DCT-Koeffizienten, da auch sie für eine effiziente Umsetzung eine Integration in den bzw. erhebliche Änderungen am Decoder erfordern und somit für eine Verwendung in Systemumgebungen wie beispielsweise Endgeräten kaum geeignet sind. Es sei auf folgende Literaturquellen verwiesen. In [CWQ01] wird ein Verfahren vorgeschlagen, das eine adaptive Filterung von DCT-Koeffizienten zur Reduktion von Blockartefakten vornimmt. Die Größe der Filterfenster wird an die ebenfalls im DCT-Bereich messbare Aktivität von Bildbereichen angepasst. In [LB02] wird eine Blocking-Messung und Klassifizierung von Blöcken im DCT-Bereich durchgeführt. Durch eine an die Ergebnisse der Analysen angepasste Manipulation von DCT-Koeffizienten wird eine Verringerung von Blockartefakten bereits vor der Rücktransformation der Koeffizienten in den Ortsbereich sichergestellt. Besondere Beachtung finden Diskontinuitäten in homogenen Bereichen, welche ebenfalls bereits im DCT-Bereich erkannt und beseitigt werden, wobei auf den Erhalt echter Kanten geachtet wird. Eine Weiterentwicklung des Verfahrens findet sich in [WZF04].

### 5.2.1 Lineare Filterung von Blockgrenzen

Ein einfaches Verfahren zur Verringerung von Blockartefakten wird in [ANT92] vorgeschlagen, welches auf Untersuchungen in [RL83] zurückgeht. Blockartefakte lassen sich, wie in Kapitel 2 beschrieben, als Diskontinuitäten im Bereich der Blockgrenzen transformationscodierter Bilder beschreiben. Das Verfahren sieht vor, derartige Diskontinuitäten geeignet zu verschleifen. Zu diesem Zweck ist eine statische Tiefpassfilterung aller Blockgrenzen vorgesehen. Zur Vermeidung der Degradation unnötig vieler Bilddetails werden die inneren Bereiche der einzelnen Blöcke von der Filterung ausgenommen. Abbildung 5.2 veranschaulicht die prinzipielle Vorgehensweise.

Exemplarisch ist hier eine Filterung mit 2-Tap-Filtern dargestellt. Die Bearbeitung von horizontal und vertikal verlaufenden Blockgrenzen kann separiert durchgeführt werden, wobei die Reihenfolge nicht kritisch ist. Grundsätzlich muss die Lage der Blockgrenzen bekannt sein, was bei aktuellen Bildkompressionsverfahren in der Regel der Fall ist.

In [ANT92] wurden Tiefpassfilter mit unterschiedlichen Fensterformen und Fenstergrößen auf ihre Wirksamkeit hin untersucht. Darunter befanden sich quadratisch ausgehende Filter mit neun Koeffizienten, kreuzförmige Filter mit fünf Koeffizienten und 2-Tap Filter, deren Eingangspixel jeweils die direkt an einer Blockgrenze zwischen zwei Blöcken

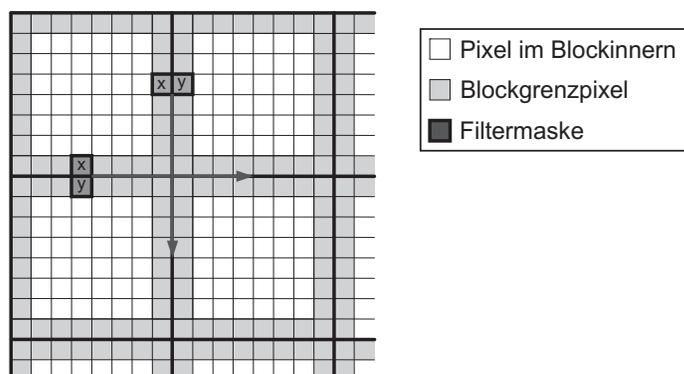


Abbildung 5.2: Filterung von Blockgrenzen zur Artefaktreduktion

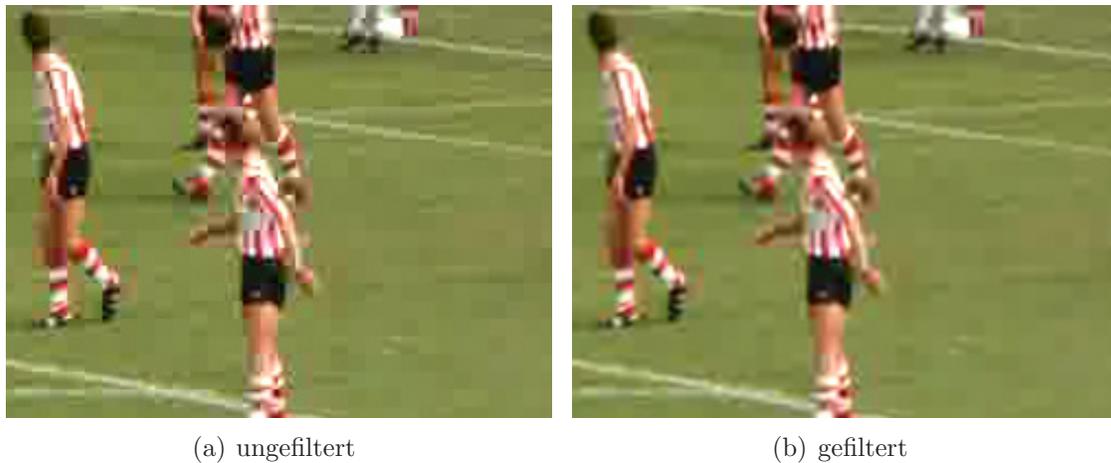
benachbarten Pixel darstellen. Für das zuletzt genannte Filter wurde bei Verwendung der Koeffizienten  $0,75$  und  $0,25$  in den meisten untersuchten gestörten Bildern die größte Verringerung des MSE nachgewiesen. Für die in Abbildung 5.2 dargestellten Pixel  $x$  und  $y$  lassen sich die gefilterten Pixel berechnen zu

$$\begin{aligned} x_f &= 0,75 \cdot x + 0,25 \cdot y, \\ y_f &= 0,75 \cdot y + 0,25 \cdot x. \end{aligned} \quad (5.6)$$

In Abbildung 5.3 wird an einem Ausschnitt aus Frame 24 der Sequenz *football*, codiert mit einer Datenrate von  $3,5$  Mbit/s (MPEG-2), die Wirksamkeit des Verfahrens gezeigt. Es sei darauf hingewiesen, dass hier und im Folgenden nur der Luminanzkanal gefiltert wurde, weshalb die stellenweise sichtbaren Farbfehler durch die Deblocking-Filterung nicht verändert werden. In Kapitel 7 wird auf die Entscheidung zur ausschließlichen Filterung der Luminanz näher eingegangen.

Im unbearbeiteten Bildausschnitt sind Blockartefakte sowohl in detaillierten Bereichen (Spieler), als auch im homogenen Hintergrund (Rasen) erkennbar. Teilweise treten auch DCT-Basisbilder hervor. Im gefilterten Bild ist eine deutliche Reduktion der Blockartefakte erkennbar, was insbesondere auf detaillierte Objekte (z. B. Fußballspieler) zutrifft. In den homogenen Hintergrundbereichen (Rasen) bleiben blockartige Artefakte weiterhin sichtbar, da an solchen Stellen die ausschließliche Filterung der Randpixel der zumeist nach der Codierung recht unstrukturierten Blöcke nicht ausreicht.

Die statische Filterung aller Blockgrenzen führt nicht zu einem generellen Unschärfedruck (siehe dazu auch die Testergebnisse in Kapitel 7). Allerdings können unter bestimmten Umständen sichtbare Blockraster im gefilterten Bild entstehen. Dies ist insbesondere dort der Fall, wo diagonal ausgerichtete bewegte Strukturen vorhanden sind. Das menschliche Auge nimmt die durch die statische Filterung hervorgerufene Veränderung von Bildinhalten in regelmäßigen Abständen als feststehendes Raster vor den bewegten

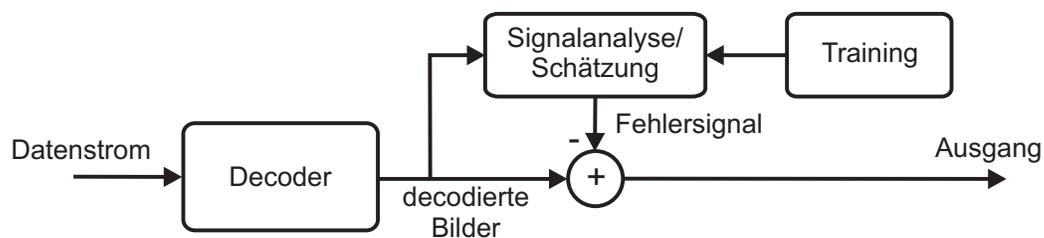


**Abbildung 5.3:** Bildbeispiel für die lineare Filterung von Blockgrenzen (Ausschnitt)

Strukturen wahr. Eine adaptive Filterung nur von Bereichen, die Blockartefakte aufweisen, ist zur Verringerung bzw. Beseitigung des Effektes und auch im Hinblick auf eine optimale Detailerhaltung grundsätzlich zu bevorzugen. Insgesamt ist das Verfahren bei der Reduktion von Blockartefakten wirksam. Es sind jedoch die genannten Schwächen und Einschränkungen zu beachten. Die Betrachtung weiterer Bilder und Sequenzen konnte die Erkenntnisse bestätigen.

### 5.2.2 Subtraktion von Fehlersignalen

Ein anderer Ansatz zur Reduktion von Codierartefakten wird in dem Verfahren nach [YCK00] gewählt. Dieses Verfahren diente auch als Grundlage für die Artefaktreduktion im SIPROS-Projekt, in dem es um die Signalverarbeitung für hochauflösende Matrixdisplays ging [BGK04]. Zunächst ist in Abbildung 5.6 die Funktionsweise an einem Blockschaltbild dargestellt.



**Abbildung 5.4:** Artefaktreduktion durch Subtraktion von Fehlersignalen

Die Grundidee des Verfahrens ist eine Schätzung eines durch Bildkompression entstandenen Fehlersignals und anschließende Subtraktion des Fehlers vom decodierten und fehlerhaften Bildsignal. Dazu wird senkrecht zu jeder Blockgrenzposition ein Vektor  $\underline{x}$  der Länge 16 Pixel (je acht Pixel aus den jeweils benachbarten zwei Blöcken) aus dem gestörten Bildsignal entnommen und analysiert. Die Verarbeitung horizontaler und vertikaler Blockgrenzen ist separiert durchzuführen. Zur Klassifizierung von Störungen werden zwei Maße, die *Boundary Discontinuity*  $D$  und die *Local Image Activity*  $P$  ermittelt. Die Boundary Discontinuity, die als eine Art Skalierungsfaktor dient, ist definiert durch

$$D = d_N - \frac{s_l - s_r}{2}, \quad (5.7)$$

mit  $d_N$  als Intensitätsdifferenz der beiden direkten Blockgrenzpixel und  $s_l$  und  $s_r$  als Summen aller Pixelintensitätsdifferenzen benachbarter Bildpunkte im linken (bzw. oberen) und rechten (bzw. unteren) Teilvektor normiert auf die Anzahl der addierten Pixeldifferenzen.

Die Local Image Activity  $P$  berechnet sich zu

$$P = \max(P_l, P_r). \quad (5.8)$$

$P_l$  gibt den Index des hochfrequentesten DCT-Koeffizienten, der nicht Null ist, einer 8-Punkt DCT des linken (bzw. unteren) Teilvektors an.  $P_r$  bezeichnet den entsprechenden Index für den rechten (bzw. oberen) Teilvektor. Durch  $P$  erfolgt auf diese Weise eine Einteilung der lokalen Bildaktivität in eine von acht Klassen. Für jede der acht Klassen muss in einer Lernphase, in der Originalbilder und entsprechende gestörte Bilder benötigt werden, ein Fehlervektor  $\hat{b}_p$  ermittelt und in einer Lookup-Tabelle gespeichert werden. Die Berechnungsvorschrift dafür lautet

$$\hat{b}_P = E \left( \frac{\underline{x} - x_0}{D} | P \right), \quad (5.9)$$

mit dem Originalpixelvektor  $x_0$  und dem Erwartungswert  $E$ , der hier dem Mittelwert entspricht. Das in der Lernphase verwendete Bildmaterial wirkt sich auf die Form der Fehlervektoren aus. Um möglichst universelle Vektoren zu erhalten, sollte ein breites Spektrum an Bildmaterial berücksichtigt werden. Die eigentliche Artefaktreduktion kann beschrieben werden durch

$$\hat{x} = \underline{x} - D \cdot \hat{b}_P. \quad (5.10)$$

Dabei ist  $\hat{x}$  der geschätzte Originalpixelvektor.

Die im Rahmen dieser Arbeit ermittelten Fehlervektoren sind in Abbildung 5.5 dargestellt. Sie zeigen die für Blockartefakte charakteristischen Signalsprünge an Blockgrenzen (hier in der Mitte der Vektoren) und wurden auf Basis von 18 unterschiedlichen Bildern,

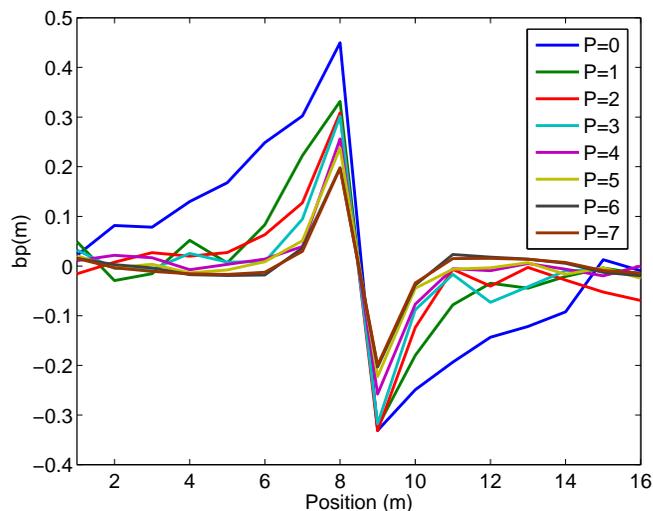


Abbildung 5.5: Ermittelte Fehlervektoren  $\hat{b}_p$

die einen Querschnitt typischen Bildmaterials und unterschiedlich ausgeprägter Artefakte darstellen, berechnet. Sehr ähnliche Vektoren sind auch in [YCK00] abgebildet.

In Abbildung 5.6 wird an einem Ausschnitt aus Frame 24 der Sequenz *football*, codiert mit einer Datenrate von 3,5 Mbit/s (MPEG-2), die Wirksamkeit des Verfahrens gezeigt. Es ist klar erkennbar, dass Blockartefakte reduziert werden, doch bleiben Ansätze von Artefakten in homogenen Bildbereichen (z. B. Rasen) sichtbar. Der subjektive Eindruck weist viele Gemeinsamkeiten mit der bereits beschriebenen linearen Filterung von Blockgrenzen auf. Auch hier gilt, dass die Wirksamkeit des Verfahrens nicht genügt, um Artefakte in allen Bildbereichen ausreichend zu reduzieren. Hervorzuheben ist die Unabhängigkeit von zusätzlichen Informationen aus dem codierten Datenstrom, so dass eine Anwendbarkeit bei ausschließlicher Verfügbarkeit bereits decodierter Bilder gegeben ist.

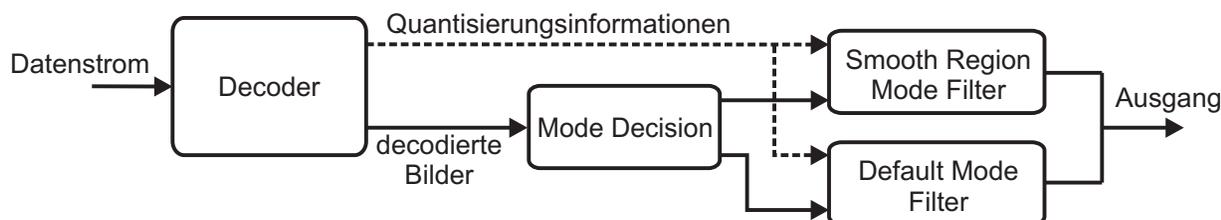
### 5.2.3 Adaptive Filterung von Blockgrenzen mit zwei Filtermodi

In [KYKR99] wurde ein adaptives Verfahren zur Verringerung von Blockartefakten auf Basis von zwei unterschiedlichen Filtermodi vorgestellt. Das Verfahren wurde als Option in den Anhang F des MPEG-4 Standards [ISO00, ISO04] aufgenommen. Zunächst ist in Abbildung 5.7 das Blockschaltbild des prinzipiellen Aufbaus dargestellt.

Bevor die Verringerung von Blockartefakten durchgeführt wird, werden die decodierten Bilddaten zunächst im Block *mode decision* analysiert. Dort wird für jede Blockgrenzposition die Homogenität (*flatness*) senkrecht zur Blockgrenze ermittelt. Dazu werden insgesamt neun Intensitätsdifferenzen benachbarter Pixel in einem Vektor senkrecht zur Blockgrenze



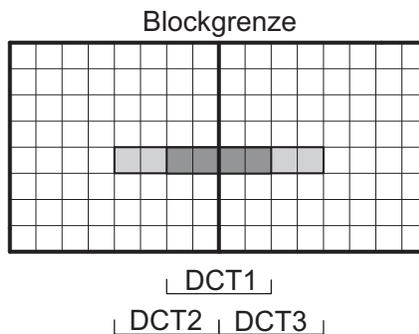
**Abbildung 5.6:** Bildbeispiel für die Artefaktreduktion durch Subtraktion von Fehlersignalen (Ausschnitt)



**Abbildung 5.7:** Reduktion von Blockartefakten mit zwei Filtermodi

ze berechnet und es wird gezählt, wie oft die Differenzen einen bestimmten Schwellwert unterschreiten. Liegt die Anzahl der Schwellwertunterschreitungen über einem weiteren Schwellwert, wird von einem homogenen Bildbereich (*smooth region mode* bzw. *DC offset mode*), sonst von einem detaillierten Bildbereich (*default mode*) ausgegangen. Für die mode decision muss die Lage der Blockgrenzen bekannt sein.

Im default mode werden zur Erhaltung möglichst vieler Details jeweils nur die beiden an einer Blockgrenze direkt benachbarten Pixel verändert. Basis der Filterung sind einzelne Koeffizienten dreier 4-Punkt-DCTs, deren Lage exemplarisch für eine Blockgrenzposition in Abbildung 5.8 veranschaulicht ist. Von besonderer Bedeutung ist der hochfrequenteste Koeffizient von DCT1, welcher maßgeblich für einen Intensitätssprung über eine Blockgrenze hinweg verantwortlich ist. Dieser vierte Koeffizient wird zur Verringerung von Blockartefakten verkleinert, bevor eine Rücktransformation durchgeführt wird. Zur Ermittlung des Skalierungsfaktors für den vierten Koeffizienten sind die Berechnung einzelner Koeffizienten von DCT2 und DCT3 sowie einige weitere einfache Operationen notwendig. Es muss sichergestellt werden, dass eine Veränderung der Intensitäten der Blockgrenzpixel nicht



**Abbildung 5.8:** Beispiel für die Lage der DCTs bei der Filterung im default mode

größer als die halbe unveränderte Intensitätssprunghöhe über die Blockgrenze ist, damit die Filterung eine glättende Wirkung hat. Da - wie bereits erwähnt - nur die beiden Pixel direkt an der Blockgrenze verändert werden, ergeben sich Vereinfachungsmöglichkeiten, die eine vollständige Hin- und Rücktransformationen unnötig und somit eine Verringerung des Rechenaufwands möglich machen (siehe [KYKR99]).

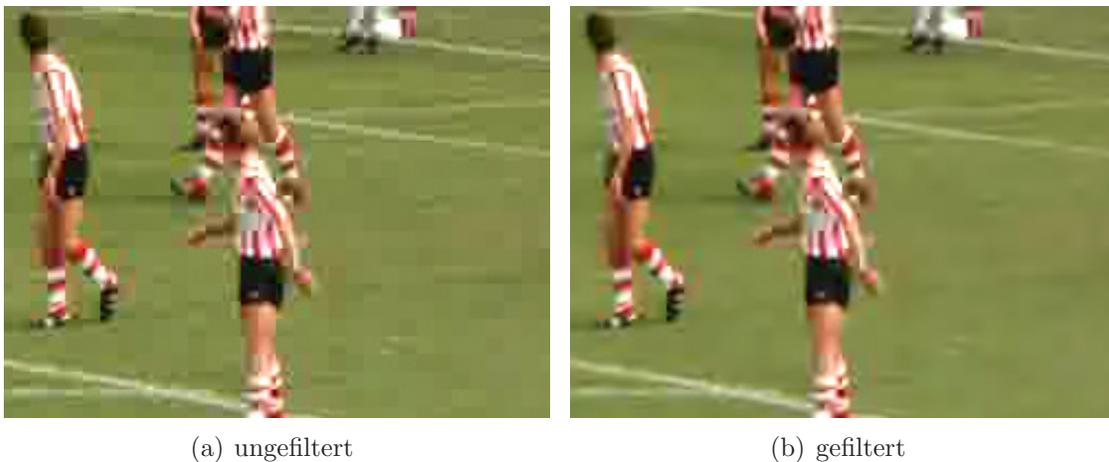
Im smooth region mode wird ein adaptives 9-Tap Glättungsfilter eingesetzt, das jeweils vier Pixel in den beiden an einer Blockgrenze benachbarten Blöcken modifiziert und somit weit in die Blöcke hineinreicht. Das Glättungsfilter stammt von einem Tiefpassfilter ab, enthält jedoch eine spezielle Randbehandlung, die abhängig von der Entfernung des zu berechnenden Ausgangspixels von der Blockgrenze die Eingangspixel (teils mehrfach) so auswählt, dass insgesamt nur auf ein Fenster von zehn Eingangspixeln (fünf auf jeder Seite der Blockgrenze) zugegriffen wird. Die Autoren sehen in dem beschriebenen Deblocking-Verfahren zuerst die Bearbeitung aller horizontalen und anschließend aller vertikalen Blockgrenzen vor.

Blockschaltbild 5.7 kann entnommen werden, dass die beiden Filter als zusätzliche Daten Quantisierungsparameter aus dem codierten Datenstrom benötigen. Die Quantisierungsparameter werden im Wesentlichen dazu eingesetzt, um zu entscheiden, ob bestimmte Charakteristika der Bilddaten auf Blockartefakte hindeuten oder nicht. So können beispielsweise Intensitätssprünge durch Kanten oder durch Artefakte hervorgerufen werden. Je gröber die Quantisierung ist, umso wahrscheinlicher ist jedoch ein Blockartefakt. Auf diese Weise helfen Quantisierungsparameter dabei, zu entscheiden, ob eine Filterung durchgeführt oder ausgesetzt wird. Auch wenn dieses Vorgehen sinnvoll ist, verhindert es den Einsatz des Artefaktreduktionsverfahrens in allen Umgebungen, in denen ein Zugriff auf Quantisierungsparameter nicht möglich ist (z. B. im Display).

In Abbildung 5.9 ist ein Beispiel für die Blocking-Reduktion mit dem beschriebenen Verfahren an einem Ausschnitt aus Frame 24 der Sequenz *football*, codiert mit 3,5 Mbit/s, dargestellt. Eine wirksame Reduktion von Blockartefakten ist sowohl in detaillierten als auch in homogenen Bildbereichen erkennbar. Die meisten Artefakte werden erkannt und

deutlich verringert, hauptsächlich in detaillierten Bildbereichen verbleiben jedoch einige Artefakte. Durch kleinere Änderungen im Algorithmus kann erreicht werden, dass nahezu alle Artefakte gefiltert werden, was jedoch auch mehr ungewollte Filterungen mit sich bringt, weshalb in der hier verwendeten Implementierung den Angaben aus der Literatur gefolgt wurde. In Hintergrundbereichen ist eine starke Glättungswirkung des Filters für den *smooth region mode* erkennbar. Je nach Bildmaterial kann dies dazu führen, dass Unschärfe bzw. eine Verschleifung von Details oder eine positive Auswirkung auf die Bildqualität wahrgenommen wird. Letzteres ist z. B. dann der Fall, wenn durch die stark glättende Wirkung der Filterung die Sichtbarkeit von DCT-Basisbildern deutlich verringert wird.

Insgesamt ist die Wirksamkeit des Verfahrens in den meisten Fällen hoch. Die Ergebnisse der Filterung sind im Schnitt über verschiedene Testbilder deutlich besser als bei der statischen Filterung von Blockgrenzen oder der Subtraktion von Fehlersignalen, da insbesondere stark sichtbare Artefakte in wenig detaillierten Bildbereichen stärker verringert werden.



**Abbildung 5.9:** Bildbeispiel für die Artefaktreduktion mit zwei Filtermodi nach [KYKR99] (Ausschnitt)

In [RK00] wird ein weiteres Deblocking-Verfahren mit zwei unterschiedlichen Filtermodi vorgestellt. Es hat einen mit dem Verfahren nach [KYKR99] vergleichbaren Aufbau, wurde jedoch mit dem Ziel entwickelt, mit weniger komplexen Komponenten und weniger Rechenaufwand auszukommen. Auch in diesem Verfahren werden zunächst an jeder Blockgrenzposition Bildpunkte analysiert, um einen Filtermodus auswählen zu können. Dazu werden die absoluten Differenzen der beiden Pixelintensitäten oberhalb und unterhalb (bzw. links und rechts) der Blockgrenze mit einem festen Schwellwert (5) verglichen. Sind beide Differenzen unterhalb der Schwelle, wird der analysierte Bereich als solcher mit *strong blocking effect*, sonst als solcher mit *weak blocking effect* klassifiziert. Es kann davon

ausgegangen werden, dass der strong blocking effect hauptsächlich in homogenen Bildbereichen auftritt, während der weak blocking effect eher in detaillierten Bereichen auftritt. Dementsprechend werden in beiden Fällen zur Glättung von Blockartefakten Filter mit unterschiedlicher örtlicher Wirkungsbreite eingesetzt.

Die beiden Filter sind so gestaltet, dass sie Artefakte bzw. Diskontinuitäten rampenförmig verschleifen. Sie sind mit dem Filter in Abbildung 6.5 vergleichbar. Allerdings wird hier den beiden Pixeln, die sich direkt an der Blockgrenze befinden, grundsätzlich derselbe Intensitätswert zugewiesen. Dadurch können zwar Intensitätssprünge an Blockgrenzen maximal geglättet werden, aber es zeigte sich während der Untersuchungen zu dieser Arbeit, dass dieses Vorgehen zur Entstehung horizontaler und vertikaler Strukturen in regelmäßigen Abständen führen kann, für die das HVS sehr empfindlich ist. Somit können - ähnlich wie bei der statischen Filterung von Blockgrenzen - insbesondere im Bereich von bewegten diagonal ausgerichteten Bildstrukturen Blockraster sichtbar werden. Da die Filterung hier adaptiv durchgeführt wird, ist der Effekt allerdings üblicherweise wenig ausgeprägt. Das Filter für den weak blocking effect verändert die Intensitäten von vier Bildpunkten, das für den strong blocking effect die Intensitäten von sechs Bildpunkten.

Ähnlich wie in dem Verfahren nach [KYKR99] kann die Filterung an jeder Blockgrenzposition ausgesetzt werden. Dazu wird der Zugriff auf Quantisierungsinformationen aus dem codierten Datenstrom benötigt. Gefiltert wird nur, wenn die absolute Intensitätsdifferenz von zwei Blockgrenzpixeln kleiner als das Produkt aus einem konstanten Faktor und dem zugehörigen Quantisierungsparameter aus dem Datenstrom ist. Anderenfalls wird von einer Kante ausgegangen, deren Steilheit nicht beeinträchtigt werden soll. Für die beiden Filtermodi sind unterschiedliche konstante Faktoren vorgesehen. Das beschriebene Verfahren wird separiert auf alle horizontalen und vertikalen Blockgrenzen angewendet. Die Lage der Blockgrenzen muss bekannt sein.

Die Wirksamkeit des Verfahrens ist etwas schlechter als die des Verfahrens nach [KYKR99]. In Abbildung 5.10 ist zu erkennen (siehe z. B. Fußballspieler links oder obere Bildmitte), dass mehr Artefakte unbearbeitet bleiben, was sich auch in anderen Bildern zeigte. Dafür ist die vereinfachte Prüfung, ob eine Filterung durchgeführt oder ausgesetzt werden soll, verantwortlich. Die vereinfachten Filter können sich je nach Bildmaterial im Vergleich zu [KYKR99] sowohl negativ als auch positiv bemerkbar machen. Grundsätzlich negativ zu bewerten ist die bereits erwähnte Gefahr der Entstehung von sichtbaren Blockrastern durch Anwendung des Verfahrens.

In [TCS05] und in [KIL04] finden sich Deblocking-Verfahren, die in ihrem grundsätzlichen Aufbau starke Ähnlichkeiten zu den beiden beschriebenen Verfahren mit zwei Filtermodi aufweisen. Sie wurden jedoch auf drei bzw. vier Filtermodi erweitert. Einige weitere Informationen zu diesen Verfahren sind in Kapitel 6 angegeben.



Abbildung 5.10: Bildbeispiel für die Artefaktreduktion mit zwei Filtermodi nach [RK00] (Ausschnitt)

### 5.2.4 Filterung von Wavelet-Koeffizienten

In der Literatur werden auch wavelet-basierte Verfahren zur Reduktion von Blockartefakten vorgestellt. In [Sie04] werden die Verfahren nach [XOZ97, KJKH98, WYT01] detailliert miteinander verglichen. Aus dem Vergleich geht hervor, dass das Verfahren nach [KJKH98] einen guten Kompromiss aus Wirksamkeit und Detailerhaltung darstellt, weshalb es im Folgenden beispielhaft für diese Gruppe von Verfahren näher vorgestellt wird. Abbildung 5.11 zeigt das Blockschaltbild.

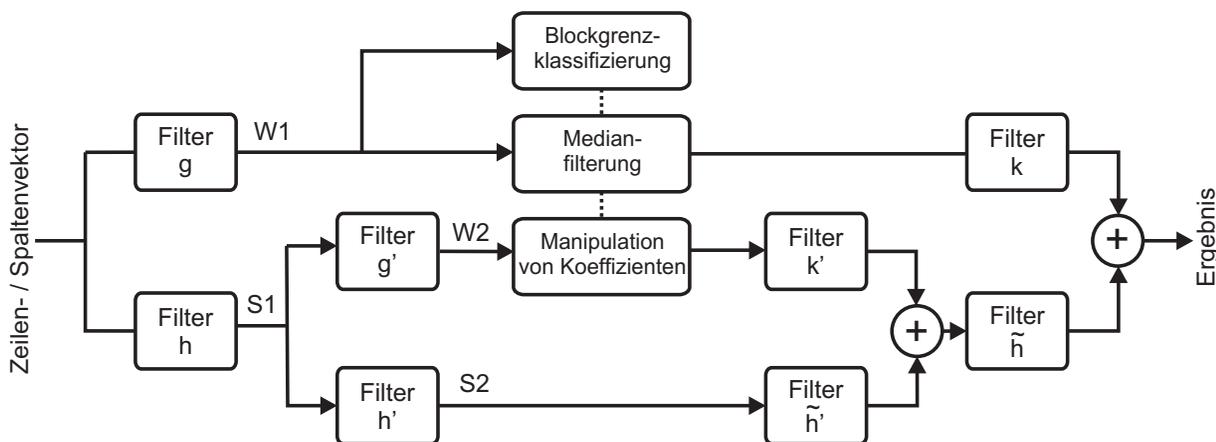


Abbildung 5.11: Artefaktreduktion auf Basis der Filterung von Wavelet-Koeffizienten, [KJKH98]

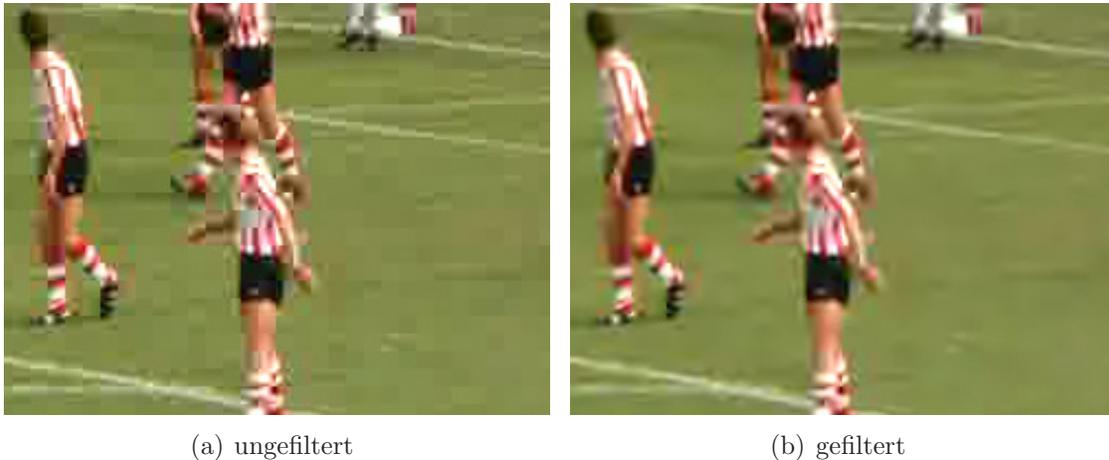
Das Verfahren verarbeitet einzelne Zeilen bzw. Spalten eines Bildes, welche als Vektoren betrachtet werden, getrennt voneinander. Die Vektoren werden mit Hilfe einer zwei-stufigen undezimierten eindimensionalen Wavelet-Transformation in den Wavelet-Bereich überführt. Eingesetzt wird ein Wavelet mit kurzer Ausdehnung und wenigen Oszillationen nach [MZ92]. Es eignet sich gut zur Lokalisierung von Signalsprüngen bzw. Diskontinuitäten und somit auch zur Lokalisierung von Blockartefakten. Eine für das Deblocking-Verfahren geeignete diskrete Realisierung der Wavelet-Transformation und Wavelet-Rücktransformation wird ebenfalls in [MZ92] vorgestellt, soll hier aber nicht im Detail betrachtet werden. Sie basiert auf Faltungen bzw. Filteroperationen mit kurzen Filtern. Die Filter in der zweiten Stufe (z. B. Filter  $g'$ ) entsprechen denen in der ersten Stufe (z. B. Filter  $g$ ), allerdings ist zwischen zwei Koeffizienten jeweils eine Null einzufügen.

Das Filter  $g$  (2 Taps) besitzt differenzierende und das Filter  $h$  (4 Taps) glättende Eigenschaften. Durch die Anordnung der Filter wie in Abbildung 5.11 werden Diskontinuitäten im Eingangssignal (Kanten und Blockartefakte) als Impulse in das W1-Band und als geglättete Impulse in das W2-Band extrahiert. Das W1-Band dient zunächst zur Analyse des Bildsignals. Es wird vektorweise eine Art von Blocking-Analyse, welche auf der Differenz der Varianzen des W1-Signals in Blockmitte und an Blockgrenzen basiert, durchgeführt. Hierauf aufbauend werden zwei Schwellwerte berechnet, die der Einteilung von Blockgrenzbereichen in drei Klassen dienen. Die drei Klassen korrespondieren mit bestimmten Signalverläufen im Ortsbereich: steile Kanten, weiche Kanten und flache bzw. homogene Bildbereiche. Es sei angemerkt, dass für dieses Verfahren die Position von Blockgrenzen im Ortsbereich bekannt sein muss. Wenn das der Fall ist, können die Blockgrenzen auch im Wavelet-Bereich lokalisiert werden.

Die Reduktion von Blockartefakten erfolgt nun durch Filterung von Wavelet-Koeffizienten, angepasst an die für eine Blockgrenze ermittelte Klasse. Steile Kanten sind mit hoher Wahrscheinlichkeit echte Kanten und somit relevante Bildinformationen, die nicht beeinträchtigt werden sollten. Daher erfolgt bei Identifizierung einer steilen Kante in diesem Verfahren keine Filterung. Weiche Kanten, so wie sie hier verstanden werden, können durchaus Intensitätssprünge kleiner Höhe enthalten. Diese sind mit hoher Wahrscheinlichkeit auf Blockartefakte zurückzuführen. Wie bereits erwähnt, werden Intensitätssprünge bzw. Diskontinuitäten als Impulse ins W1-Band extrahiert. Wird ein Blockgrenzbereich als weiche Kante identifiziert, erfolgt daher eine 3-Tap Medianfilterung im W1-Band, um ein mögliches Blockartefakt zu beseitigen. Da Blockartefakte in homogenen Bildbereichen besonders deutlich sichtbar sind, wird im Falle der Identifizierung eines Blockgrenzbereichs als homogener Bereich neben der Medianfilterung im W1-Band zusätzlich eine Verkleinerung von jeweils fünf Wavelet-Koeffizienten im W2-Band durchgeführt. Der Grad der Verkleinerung wird an die im W1-Band ermittelte Impulshöhe adaptiert.

Dem Blockschaltbild 5.11 kann entnommen werden, dass nach der Filterung der Wavelet-Koeffizienten eine Wavelet-Rücktransformation erfolgt, um die Signalvektoren in den Ortsbereich zurückzuführen. Abbildung 5.12 zeigt die Wirksamkeit der wavelet-

basierten Artefaktreduktion an einem Ausschnitt aus Frame 24 der Sequenz *football*, codiert mit 3,5 Mbit/s.



**Abbildung 5.12:** Bildbeispiel für die Artefaktreduktion auf Basis der Filterung von Wavelet-Koeffizienten (Ausschnitt)

Es ist offensichtlich, dass Blockartefakte gleichermaßen in detaillierten und in homogenen Bildbereichen deutlich verringert werden. Nahezu keine Artefakte bleiben un bearbeitet. Eine Tendenz zur Entstehung neuer störender Artefakte durch Anwendung des Verfahrens ist nicht sichtbar, jedoch ist eine Tendenz zu Unschärfbildung zu erkennen. Insgesamt kann von einem guten subjektiven Eindruck gesprochen werden.

Die Komplexität des Verfahrens ist unter den vorgestellten Verfahren als hoch einzuschätzen. Zwar kann die Wavelet-Hin- und Rücktransformation durch Filteroperationen realisiert werden, allerdings bedeutet die Bandaufspaltung, Verarbeitung und anschließende Zusammenführung der Bänder einigen Aufwand. Darüber hinaus benötigt die integrierte Blocking-Analyse den Zugriff auf ganze Zeilen bzw. ganze Spalten. Ein Zugriff auf ganze Spalten erfordert aber in der Praxis typischerweise die Speicherung eines ganzen Bildes, was je nach Anwendung einen zu großen Aufwand darstellen kann. Eine Verringerung des Speicherbedarfs ist ohne eine neue bzw. stark veränderte Blocking-Analyse nicht möglich. Es sei angemerkt, dass auch die anderen bereits vorgestellten Verfahren eine getrennte Verarbeitung aller horizontalen und vertikalen Blockgrenzen vorsehen, was die Speicherung eines ganzen Bildes erfordern kann. Durch ein geschicktes Verweben der Verarbeitung horizontaler und vertikaler Blockgrenzen sind jedoch Implementierungen denkbar, die nur einige Zeilenspeicher benötigen und im Vergleich mit den Originalvarianten sehr ähnliche Eigenschaften erwarten lassen.

# 6 Adaptive Reduktion von Blockartefakten mit lokaler Bildanalyse

Im Rahmen dieser Arbeit wurde ein neues adaptives Verfahren zur Reduktion von Blockartefakten in mehreren Varianten entwickelt. Dabei wurden Erkenntnisse berücksichtigt, die bei der Untersuchung existierender Verfahren gewonnen wurden. So sollte das Verfahren grundsätzlich unabhängig von Zusatzinformationen aus dem codierten Datenstrom (z. B. Quantisierungsparameter) sein und nur auf die bereits decodierten Daten angewiesen sein. Auf diese Weise kann ein besonders breites Spektrum an Einsatzmöglichkeiten erschlossen werden. Besonders wichtig war die Einsetzbarkeit des Verfahrens in Fernsehgeräten, da die Arbeiten im Rahmen eines Industrieprojekts mit der Micronas GmbH in München begonnen wurden. Außerdem sollte eine mit anderen bekannten Verfahren vergleichbare hohe Wirksamkeit bei der Reduktion von Blockartefakten erreicht werden und der Implementierungs- und Rechenaufwand sollte möglichst gering ausfallen.

Die gestellten Anforderungen ließen sich am besten mit einem klassischen Postprocessing-Ansatz verwirklichen, wodurch gleichzeitig sichergestellt werden konnte, dass ein Einsatz des Verfahrens ohne Eingriff in existierende Standards möglich [TCS05, SK98] und ein Maximum an Flexibilität gegeben ist. Es wurde zunächst der in Abbildung 6.1 dargestellte Grundaufbau entworfen, auf dessen Basis verschiedene Varianten des Verfahrens entwickelt und implementiert wurden. Der Grundaufbau lässt eine Trennung in zwei Hauptkomponenten erkennen: die lokale Bildanalyse und die adaptive Filterung des Bildmaterials, welche mit Hilfe von zwei unterschiedlichen Filtern durchgeführt werden kann. Diese Komponenten werden im Folgenden näher beschrieben (siehe auch [Pia05a, Pia05b]).

## 6.1 Lokale Bildanalyse

Die lokale Bildanalyse dient als Werkzeug, um die Filterung des Bildmaterials an lokale Bildeigenschaften anzupassen. Dabei werden im Wesentlichen zwei Ziele verfolgt. Erstens muss entschieden werden, ob das Bild in der Umgebung der Bildanalyse überhaupt gefiltert werden soll oder nicht. Eine unnötige Filterung würde die Bildqualität zwangsläufig verschlechtern. Zweitens muss die Filterung geeignet an die Bildumgebung angepasst werden.

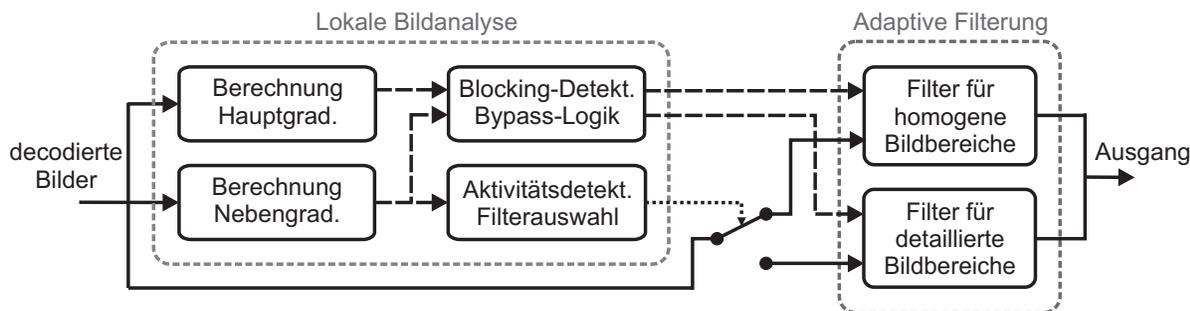
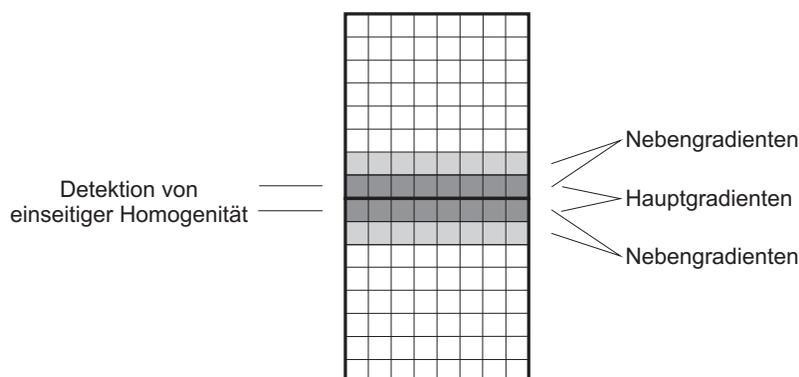


Abbildung 6.1: Reduktion von Blockartefakten mit lokaler Bildanalyse

Grundsätzlich ist es sinnvoll, nur dort zu filtern, wo mit hoher Wahrscheinlichkeit Blockartefakte sichtbar sind. Die Verringerung von Blockartefakten soll möglichst stark sein, aber gleichzeitig müssen möglichst viele Details erhalten werden. Zu diesem Zweck ist in homogenen Bildbereichen der Einsatz von breiten Filtern sinnvoll, da Blockartefakte dort besonders deutlich sichtbar sind und möglichst stark geglättet werden müssen. Eine Beeinträchtigung von Details ist in homogenen Bereichen nur in geringem Maße möglich. In detaillierten Bildbereichen ist der Detailerhalt hingegen wichtiger und die Sichtbarkeit von Blockartefakten durch Maskierungseffekte ohnehin geringer [KYKR99, Fec00, KIL04, TCS05]. Deshalb ist die Verwendung kurzer Filter anzustreben, um zwar eine Artefaktreduktion zu erreichen, aber Details möglichst wenig zu beeinträchtigen.

Die Untersuchung der in Kapitel 5 vorgestellten Algorithmen zur Reduktion von Blockartefakten ergab, dass mit nur zwei Filtermodi eine hohe Wirksamkeit bei der Artefaktreduktion sowie eine hinreichende Detailerhaltung erreicht werden kann. Daher wurde die lokale Bildanalyse zunächst so gestaltet, dass sie eine Auswahl zwischen zwei Filtermodi treffen kann und auch ohne Zugriff auf Quantisierungsinformationen entscheiden kann, ob eine Filterung notwendig ist oder nicht. Dazu wird zunächst die Berechnung von sogenannten Haupt- und Nebengradienten durchgeführt. Darunter sind Pixelintensitätsdifferenzen im Bereich von Blockgrenzen zu verstehen. Es wird wie in anderen Verfahren (siehe Kapitel 5) vorausgesetzt, dass die Lage des Blockrasters im Bild bekannt ist.

Die lokale Bildanalyse erstreckt sich maximal über den Bereich einer Blockgrenze und untersucht alle im Bild vorhandenen Blockgrenzen separat. Ob zunächst alle horizontalen oder vertikalen Blockgrenzen verarbeitet werden oder eine gemischte Verarbeitung horizontaler und vertikaler Grenzen erfolgt, spielt keine signifikante Rolle. Dadurch ist es möglich, die Reihenfolge so festzulegen, dass der Implementierungsaufwand des gesamten Verfahrens minimal wird. In Abbildung 6.2 ist am Beispiel einer horizontal verlaufenden Grenze zwischen zwei Blöcken veranschaulicht, auf welche Bildpunkte zur Berechnung der Gradienten zugegriffen werden muss.



**Abbildung 6.2:** Haupt- und Nebengradienten bei der lokalen Bildanalyse

Die Hauptgradienten sind die absoluten Pixelintensitätsdifferenzen der direkten Blockgrenzpixel zweier benachbarter Blöcke. Die für die weitere Verarbeitung relevante lokale Hauptgradientensumme ergibt sich für das obere linke Pixel des unteren betrachteten Blocks im Bild  $s$  an den Koordinaten  $(x_0, y_0)$  zu

$$HG_{x_0, y_0} = \sum_{n=0}^7 \text{abs}(s(x_0 + n, y_0 - 1) - s(x_0 + n, y_0)). \quad (6.1)$$

Die lokale Nebengradientensumme, die sich aus Pixelintensitätsdifferenzen innerhalb zweier durch eine gemeinsame Grenze verbundener Blöcke ergibt, lässt sich berechnen durch

$$NG_{x_0, y_0} = 0,5 \cdot \left( \sum_{n=0}^7 \text{abs}(s(x_0 + n, y_0 - 2) - s(x_0 + n, y_0 - 1)) \right) + 0,5 \cdot \left( \sum_{n=0}^7 \text{abs}(s(x_0 + n, y_0) - s(x_0 + n, y_0 + 1)) \right). \quad (6.2)$$

Die in Abbildung 6.1 dargestellte Aktivitätsdetektion kann direkt auf Basis der lokalen Nebengradientensumme erfolgen. Mit Hilfe eines Schwellwertvergleichs wird entschieden, ob der untersuchte Bildbereich homogen ist (niedrige Aktivität) oder nicht (höhere Aktivität). Je nach Ergebnis des Schwellwertvergleichs wird eines von zwei Filtern für die Reduktion von Artefakten ausgewählt. Die Ermittlung der Nebengradientensumme im direkten Blockgrenzbereich ist sinnvoll, da die nachfolgende Filterung in demselben Bildbereich stattfindet. Zusätzlich wird dadurch eine Implementierung des Verfahrens mit minimaler Anzahl an Zeilenspeichern ermöglicht.

Die im Blockschnittbild dargestellte Blocking-Detektion basiert auf dem Verhältnis von Hauptgradientensumme zu Nebengradientensumme. In unkomprimierten Bildern sind Haupt- und Nebengradienten im Schnitt etwa gleich groß. Das Verhältnis der Gradientensummen ist also ca. Eins. Konsequenzen einer blockbasierten Bildkompression sind

einerseits die Bildung von Diskontinuitäten im Bereich von Blockgrenzen und andererseits Informations- bzw. Detailverluste im Innern von Blöcken (siehe Kapitel 2). Dadurch werden die Hauptgradienten tendenziell größer, die Nebengradienten hingegen kleiner. Das Verhältnis der Gradientensummen wird folglich umso größer, je stärker sich Blockartefakte im Bild bemerkbar machen und eignet sich daher zur lokalen Blocking-Detektion [WY97, Tra98].

Um die Zuverlässigkeit des Blocking-Maßes zu erhöhen, empfiehlt es sich, einige Sonderfälle geeignet zu behandeln. So liegt in der Regel kein sichtbares Blocking vor, wenn sich Haupt- und Nebengradientensummen absolut nur wenig voneinander unterscheiden. Andererseits führt einseitige Homogenität entlang einer Blockgrenze oft zu einer besonders deutlichen Sichtbarkeit von Blockartefakten. Einseitige Homogenität kann durch einen Vergleich der Summen absoluter Pixelintensitätsdifferenzen entlang der Blockgrenze zweier benachbarter Blöcke erkannt werden (siehe Abbildung 6.2).

Eine erste Implementierung der lokalen Bildanalyse sah die Ermittlung eines einzigen Blocking- und Aktivitätsmaßes pro Bildblock vor. Dabei wurden die Hauptgradienten der oberen und linken Blockgrenze zusammengefasst. Die Nebengradienten, die auch der Aktivitätsbeurteilung dienen, wurden nahe der Blockmitte in horizontaler und vertikaler Richtung ermittelt. Die Vorgehensweise erwies sich insofern als problematisch, als dass kein sinnvoller Schwellwert für die Entscheidung, ob ein Block Artefakte aufweist oder nicht, gefunden werden konnte. So wurden bei diversen Schwellwerteinstellungen in manchen Bildern deutlich zu viele Blöcke als gestört erkannt, während in anderen Bildern viele Artefakte unerkant blieben. Der Grund dafür ist, dass manche Encoder offensichtlich dazu neigen, in bestimmten Bewegungssituationen hauptsächlich horizontal oder hauptsächlich vertikal verlaufende Blockartefakte zu erzeugen.



**Abbildung 6.3:** Mehrheitlich vertikal verlaufende Blockartefakte in der Sequenz *radsport*

Abbildung 6.3 zeigt exemplarisch an einem Bildausschnitt aus einer DVB-Übertragung das mehrheitliche Auftreten von Blockartefakten an vertikalen Blockgrenzen. Da ein gemeinsamer Blocking-Wert für horizontale und vertikale Blockgrenzen berechnet wird, ist es leicht nachvollziehbar, dass kein sinnvoller Schwellwert gefunden werden kann, der sowohl in der geschilderten Situation als auch beim gleichzeitigen Auftreten horizontal und vertikal verlaufender Artefakte eine zuverlässige Detektion von Blockartefakten ermöglicht.

Durch den Übergang auf eine separate Analyse und Filterung jeder Blockgrenze wurden die Probleme beseitigt. Es kann sogar gezeigt werden, dass es vorteilhaft sein kann, den Analysebereich noch weiter zu verkleinern und jede Hälfte einer Blockgrenze getrennt zu verarbeiten, da auf diese Weise bei bestimmten Konstellationen benachbarter DCT-Basisbilder Blocking-Störungen erkannt werden können, wo es sonst nicht der Fall ist. Auch die Filterung kann feiner an die lokale Bildsituation angepasst werden. Im Folgenden wird bei Analyse ganzer Blockgrenzen von *acht Pixel Analysebreite* und bei Analyse von halben Blockgrenzen von *vier Pixel Analysebreite* gesprochen. Abbildung 6.4 zeigt exemplarisch mehrere Stellen, an denen eine Analysebreite von vier Pixeln vorteilhaft ist. Bei einer Analysebreite von acht Pixeln müsste die Schwelle, die darüber entscheidet, ob Blocking-Störungen vorliegen oder nicht, so niedrig gewählt werden, dass an vielen anderen Stellen Artefakte erkannt werden würden, wo keine sind.



**Abbildung 6.4:** Stellen, an denen eine Analysebreite von vier Pixeln vorteilhaft ist

Eine weitere Verkleinerung des Analysebereichs ist nicht sinnvoll, da auf Basis von nur ein oder zwei Analysestellen keine zuverlässige Erkennung von Blockartefakten in 8x8 Pixel großen Blöcken gelingen kann. Die Auswirkungen der Verkleinerung des Analysebereichs von acht auf vier Pixel Breite ist mit einigen objektiven Qualitätsmetriken messbar, wirkt sich subjektiv aber kaum auf die Wirksamkeit aus, was an späterer Stelle noch gezeigt wird.

Der Übergang von blockweiser Analyse auf grenzweise Analyse vergrößert die Wirksamkeit hingegen deutlich.

## 6.2 Filter zur Störungsreduktion

In den meisten im Rahmen dieser Arbeit implementierten und untersuchten Varianten des neu vorgestellten Verfahrens zur Reduktion von Blockartefakten mit lokaler Bildanalyse kamen, wie in Abbildung 6.1 dargestellt, zwei grundsätzlich verschiedene Filter zum Einsatz. Für homogene Bildbereiche, also solche mit niedriger gemessener Aktivität, ist ein Filter der Wirkungsbreite sechs Pixel vorgesehen, das Diskontinuitäten an Blockgrenzen rampenförmig glättet. In homogenen Bildbereichen kann davon ausgegangen werden, dass keine oder nur minimale Detailinformationen vorhanden sind und durch die Filterung keine relevanten Informationen beeinträchtigt werden. Blockartefakte werden durch die breite Rampencharakteristik des Filters sehr wirkungsvoll verringert. Die Filterung wird nur dann durchgeführt, wenn bei der vorher durchgeführten lokalen Blocking-Messung ein einstellbarer Schwellwert überschritten wurde. In Abbildung 6.5 ist an einem Beispiel veranschaulicht, wie sich die Anwendung des Filters auf eine Diskontinuität bzw. ein Artefakt auswirkt.

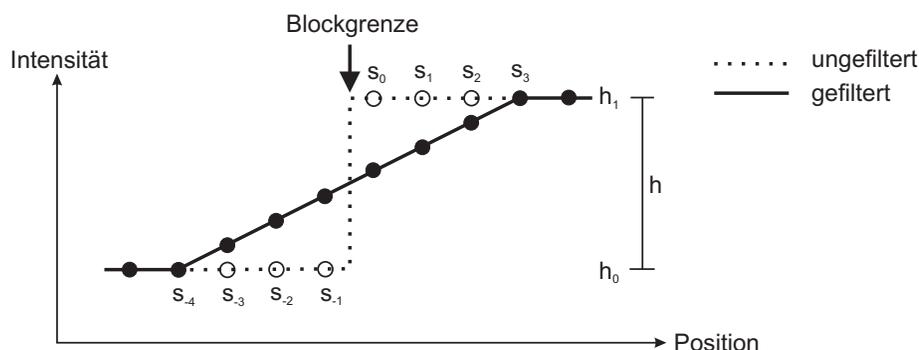


Abbildung 6.5: Filter für homogene Bildbereiche

Hat die Diskontinuität an der Blockgrenze die Höhe  $h = h_1 - h_0$ , so lassen sich die Ausgangspixel  $s'_i$  des Filters folgendermaßen berechnen:

$$\begin{aligned}
 s'_{-3} &= s_{-3} + h/7, \\
 s'_{-2} &= s_{-2} + 2h/7, \\
 s'_{-1} &= s_{-1} + 3h/7, \\
 s'_0 &= s_0 - 3h/7, \\
 s'_1 &= s_1 - 2h/7, \\
 s'_2 &= s_2 - h/7.
 \end{aligned}
 \tag{6.3}$$

$h_1$  entspricht der Intensität des ungefilterten Pixels  $s_0$ . Die Intensität des ungefilterten Pixels  $s_{-1}$  wird durch  $h_0$  beschrieben. Wenn die Diskontinuitätshöhe  $h$  über einem einstellbaren Schwellwert liegt, wird die Filterung zum Schutz von kontrastreichen Kanten ausgesetzt. Das Filter ist für eine Reduktion von Artefakten, die durch Bildkompressionsverfahren mit einer Blockgröße von 8x8 Pixeln entstehen können, ausgelegt. Eine Verkürzung des Filters für Blockgrößen von 4x4 Pixeln ist problemlos möglich.

In detaillierten Bildbereichen wird zum Schutz von Details ein Filter mit einer Wirkungsbreite von nur zwei Pixeln eingesetzt, das lesend allerdings auf vier Bildpunkte zugreift. Grundidee ist die Analyse jeder Blockgrenzposition mit Hilfe einer 4-Punkt-DCT senkrecht zur Blockgrenze, die anschließende Bearbeitung des für Blockartefakte verantwortlichen Koeffizienten und schließlich die Durchführung einer Rücktransformation (IDCT). In Abbildung 6.6 ist exemplarisch die Positionierung des DCT- bzw. IDCT-Fensters an einer Blockgrenzposition dargestellt.

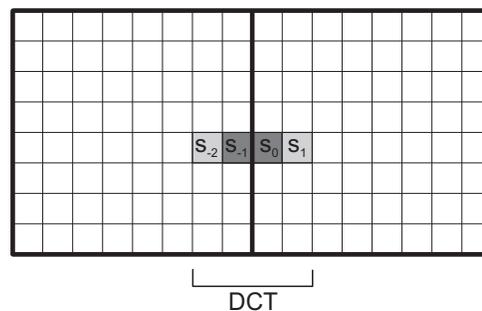


Abbildung 6.6: Filter für detaillierte Bildbereiche

Abbildung 6.7 zeigt die Basisvektoren einer 4-Punkt-DCT. Es ist erkennbar, dass der DCT-Basisvektor 3 maßgeblich mit dem Auftreten von Blockartefakten zusammenhängt, da dieser einen deutlichen Sprung in der Mitte - entspricht der Position der Blockgrenze im Analysefenster in Abbildung 6.6 - enthält. Mit diesem Basisvektor korrespondiert der hochfrequente antisymmetrische Koeffizient  $S_{DCT,3}$ , der als einziger DCT-Koeffizient für die weiteren Schritte von Interesse ist und daher wie folgt berechnet werden muss:

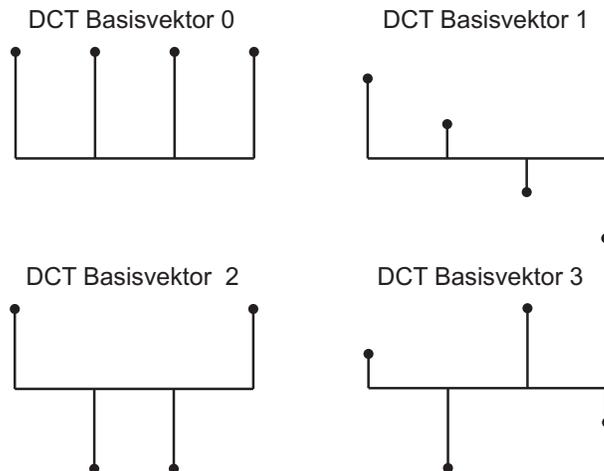
$$S_{DCT,3} = k_{3,0} \cdot s_{-2} + k_{3,1} \cdot s_{-1} + k_{3,2} \cdot s_0 + k_{3,3} \cdot s_1. \quad (6.4)$$

Die  $k_{i,j}$  sind nach [Sch98] für eine 4-Punkt-DCT definiert durch

$$k_{i,j} = \alpha_i \cdot \cos \frac{\pi(2j+1)i}{8}, \quad (6.5)$$

mit

$$\alpha_i = \begin{cases} \sqrt{\frac{1}{4}} & \text{für } i = 0 \\ \sqrt{\frac{2}{4}} & \text{für } 1 \leq i \leq 3 \end{cases}. \quad (6.6)$$



**Abbildung 6.7:** Veranschaulichung der Basisvektoren einer 4-Punkt-DCT

Heuristisch wurde ermittelt, dass eine Verminderung des Koeffizienten  $S_{DCT,3}$  um einen festen Faktor  $c = 0, 1$  wirksam und praxistauglich ist. Zur Verringerung von Blockartefakten in Detailbereichen reicht es aus, die beiden direkten Nachbarpixel an jeder Blockgrenzposition zu bearbeiten. Durch Rücktransformation ergeben sich die gefilterten Bildpunkte zu

$$\begin{aligned} s'_{-1} &= k_{0,1} \cdot S_{DCT,0} + k_{1,1} \cdot S_{DCT,1} + k_{2,1} \cdot S_{DCT,2} + k_{3,1} \cdot S'_{DCT,3}, \\ s'_0 &= k_{0,2} \cdot S_{DCT,0} + k_{1,2} \cdot S_{DCT,1} + k_{2,2} \cdot S_{DCT,2} + k_{3,2} \cdot S'_{DCT,3}. \end{aligned} \quad (6.7)$$

Die Rücktransformation der nicht gefilterten Bildpunkte lautet

$$\begin{aligned} s_{-1} &= k_{0,1} \cdot S_{DCT,0} + k_{1,1} \cdot S_{DCT,1} + k_{2,1} \cdot S_{DCT,2} + k_{3,1} \cdot S_{DCT,3}, \\ s_0 &= k_{0,2} \cdot S_{DCT,0} + k_{1,2} \cdot S_{DCT,1} + k_{2,2} \cdot S_{DCT,2} + k_{3,2} \cdot S_{DCT,3}. \end{aligned} \quad (6.8)$$

Durch Differenzbildung erhält man

$$\begin{aligned} s_{-1} - s'_{-1} &= k_{3,1} \cdot S_{DCT,3} \cdot (1 - c), \\ s_0 - s'_0 &= k_{3,2} \cdot S_{DCT,3} \cdot (1 - c), \end{aligned} \quad (6.9)$$

und somit schließlich eine einfache Möglichkeit zur Berechnung der gefilterten Bildpunkte:

$$\begin{aligned} s'_{-1} &= s_{-1} - k_{3,1} \cdot S_{DCT,3} \cdot (1 - c), \\ s'_0 &= s_0 - k_{3,2} \cdot S_{DCT,3} \cdot (1 - c). \end{aligned} \quad (6.10)$$

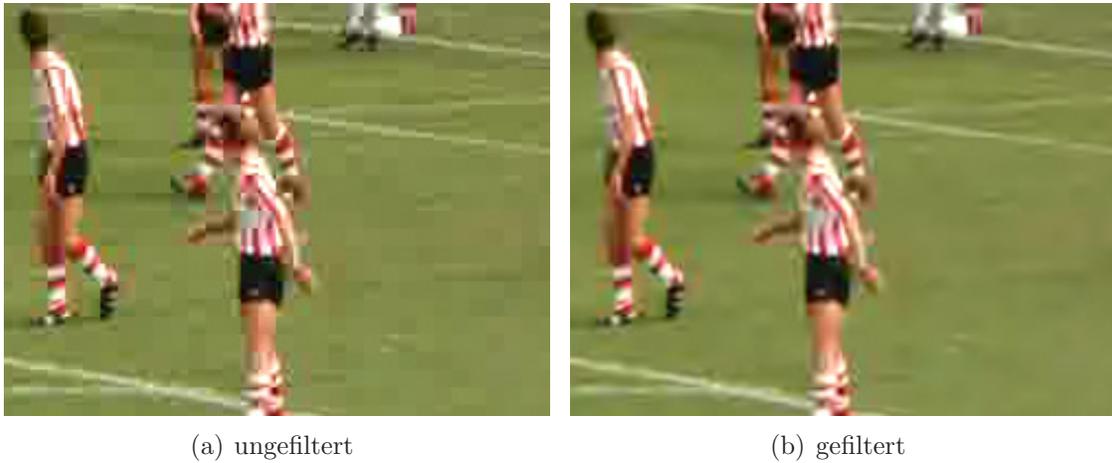
Es ist darauf zu achten, dass die Intensitätsänderung der Pixel nicht größer als die halbe Signalsprunghöhe an der Blockgrenze ausfallen darf, um eine Glättungswirkung zu erreichen und unerwünschte Effekte zu vermeiden. Dies kann durch eine Clipping-Operation realisiert werden.

In [KYKR99] wird zur Verringerung von Blockartefakten ebenfalls ein DCT-basiertes Filter eingesetzt. Bei diesem wird jedoch die Skalierung von  $S_{DCT,3}$  adaptiert, wozu zwei weitere 4-Punkt-DCTs notwendig sind. In Simulationen wurde überprüft, inwieweit sich der Einsatz des komplexeren Filters in dem hier vorgestellten Verfahren bezüglich einer Verbesserung der Wirksamkeit lohnt. Es konnten keine nennenswerten objektiven oder subjektiven Gewinne festgestellt werden, weshalb schließlich wieder eine feste Skalierung eingesetzt wurde. Eine Adaptivität des Filters ist dennoch gegeben, da die Filterung grundsätzlich nur dann durchgeführt wird, wenn die vorherige Blocking-Messung einen einstellbaren Schwellwert überschritten hat, und da sich letztlich die Intensitätswerte von vier Bildpunkten auf die Veränderung von  $s_{-1}$  und  $s_0$  auswirken.

## 6.3 Varianten des Verfahrens

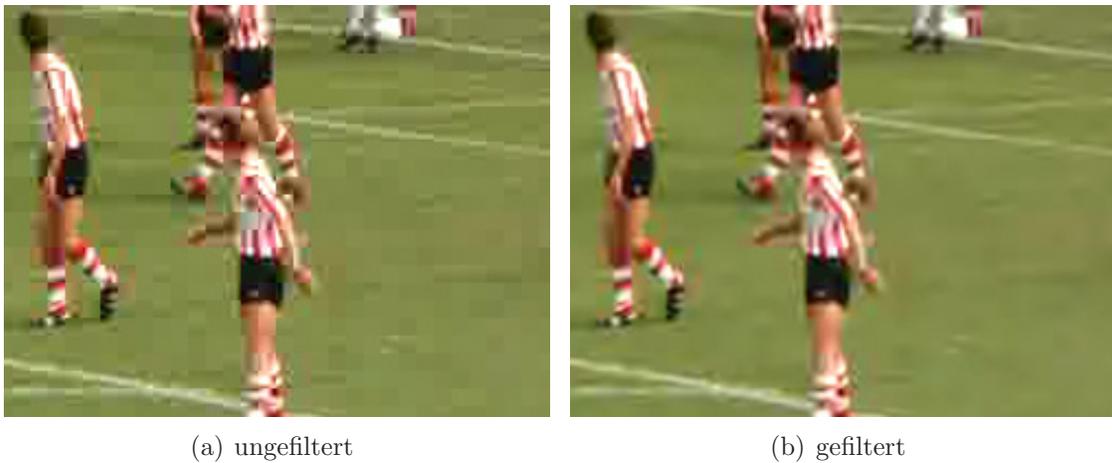
Wie bereits erwähnt, wurden auf Basis der beschriebenen Komponenten zur Bildanalyse und zur Filterung von Blockgrenzbereichen verschiedene Varianten des neuen Verfahrens zur Reduktion von Blockartefakten implementiert. In allen Varianten erfordern weder die lokale Bildanalyse noch die Filteroperationen einen Zugriff auf Zusatz- bzw. Quantisierungsinformationen aus dem codierten Datenstrom, wodurch ein flexibler Einsatz des Verfahrens auch in Endgeräten möglich ist. Unterschiede bestehen im Wesentlichen in der Analysebreite der lokalen Bildanalyse und in einigen Implementierungsdetails. Die verschiedenen Varianten stellen Zwischenschritte im Rahmen der Weiterentwicklung des Artefaktreduktionsverfahrens dar und sollen nicht im Einzelnen, sondern nur exemplarisch betrachtet werden. Die erste zu nennende Variante, auf die zum Teil schon eingegangen wurde, verwendet eine blockweise Bildanalyse und hat Probleme bei der sicheren Detektion von Artefakten. Dadurch ist die Wirksamkeit des Verfahrens eingeschränkt und eine weitere Untersuchung nicht lohnenswert. Relevant für die weiteren Untersuchungen sind hingegen die Varianten mit grenzweiser Bildanalyse, die vollständig auf den bereits vorgestellten Komponenten basieren und Analysebreiten von acht bzw. vier Bildpunkten sowohl in horizontaler als auch in vertikaler Richtung aufweisen. Diese werden im Folgenden *pi4b* (acht Pixel Analysebreite) und *pi5b* (vier Pixel Analysebreite) genannt. Es sei erwähnt, dass die Zahlen in den Namen den Entwicklungsschritt der jeweiligen Variante angeben, ansonsten aber keine Rückschlüsse auf Eigenschaften oder Implementierungsdetails erlauben.

Die Wirksamkeit der beiden letztgenannten Implementierungen ist hoch, was exemplarisch in den Abbildungen 6.8 (acht Pixel Analysebreite) und 6.9 (vier Pixel Analysebreite) nachvollzogen werden kann. Die Codierung der Sequenz *football* erfolgte mit 3,5 Mbit/s (MPEG-2). Blockartefakte werden sowohl in detaillierten als auch in homogenen Bildbereichen deutlich reduziert. Es bleiben nahezu keine Artefakte unbearbeitet. Auch die korrekte Auswahl der entsprechenden Filter gelingt, so dass nicht unnötig Details zerstört werden.



**Abbildung 6.8:** Bildbeispiel für die Artefaktreduktion mit lokaler Bildanalyse, acht Pixel Analysebreite (Ausschnitt)

Die Unterschiede zwischen beiden Varianten der Artefaktreduktion sind sehr gering. Es lassen sich theoretisch Situationen konstruieren, in denen eine feinere Analyse der zu filternden Bildbereiche ( $\pi_5b$ ) vorteilhaft ist. Andererseits kann es dann passieren, dass Blockgrenzen nur auf halber Breite gefiltert werden, was sichtbare Effekte nach sich ziehen könnte. In der Praxis konnten aber in verschiedensten Testbildern weder sichtbare Unterschiede zwischen den beiden Varianten noch Probleme durch die stellenweise Filtrierung von nur halben Blockgrenzen gefunden werden.



**Abbildung 6.9:** Bildbeispiel für die Artefaktreduktion mit lokaler Bildanalyse, vier Pixel Analysebreite (Ausschnitt)

In Tabelle 6.1 werden die beiden Varianten pi4b und pi5b, sowie die Variante pi7, auf die im folgenden Abschnitt eingegangen wird, anhand von PSNR-Messwerten verglichen. Auch wenn der PSNR generell zum Vergleich von unterschiedlichen Artefaktreduktionsverfahren wenig geeignet ist, wie in dieser Arbeit noch gezeigt wird, kann er im Rahmen der Analyse von solchen Verfahren dennoch mit Vorsicht zu interpretierende Zusatzinformationen liefern. Zum Beispiel kann der PSNR unnötige Filterungen in wenig gestörten Bildern, die mit hoher Datenrate codiert wurden, aufdecken. Diesbezüglich zeigen die Messwerte einen Vorteil der Variante pi4b gegenüber der Variante pi5b an. Offenbar kann sich die Variante pi4b etwas besser an ungestörte oder wenig gestörte Bilder adaptieren und vermeidet mehr unnötige Filterungen. Allerdings sei klargestellt, dass auch die Variante pi5b in uncodierten Sequenzen keine sichtbaren Störungen verursacht und somit gefahrlos in qualitativ guten Bildern eingesetzt werden kann. In stark gestörten Bildern fallen die Messwerte der Variante pi5b besser als die der Variante pi4b aus. Die Unterschiede sind jedoch vernachlässigbar klein, was den subjektiven Eindrücken entspricht.

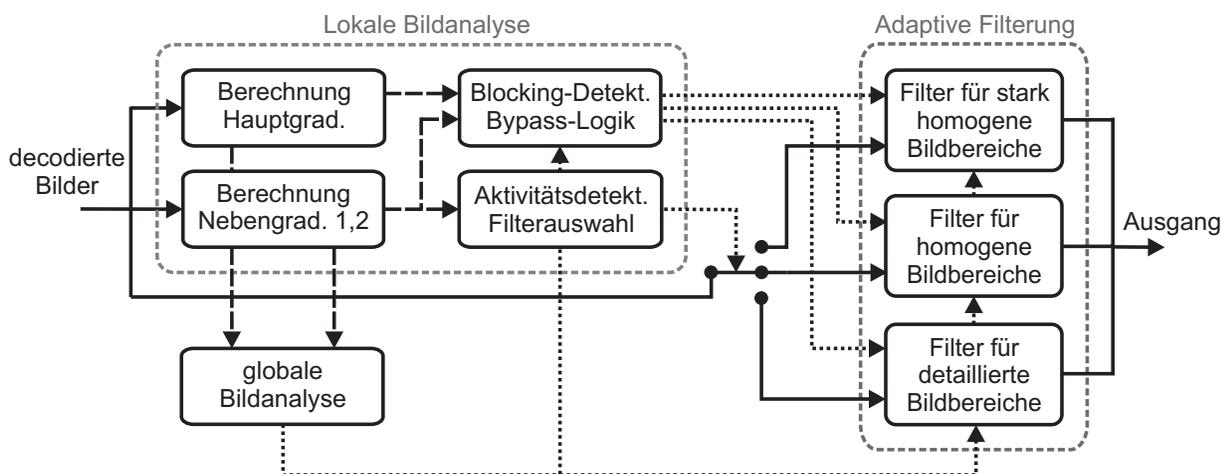
**Tabelle 6.1:** PSNR-Vergleich mehrerer Varianten der Artefaktreduktion mit lokaler (und globaler) Bildanalyse (in dB), Messungen im Luminanzkanal von Frame 24

Sequenz:	ungefiltert	pi4b	pi5b	pi7
football (uncodiert)	-	47.7379	45.4032	52.5216
football (3,5 MBit/s)	28.0127	28.4919	28.4973	28.4936
lena2_zoom2 (uncodiert)	-	52.4163	49.7464	54.1965
lena2_zoom2 (3,5 MBit/s)	31.7984	32.2414	32.262	32.2398
stockholm4 (uncodiert)	-	47.295	44.9843	52.4375
stockholm4 (3,5 MBit/s)	30.7311	31.0035	31.0163	30.9806
wheel (uncodiert)	-	47.331	45.2726	52.5306
wheel (3,5 MBit/s)	26.2736	26.9011	26.9277	26.9279

Zusammenfassend kann gesagt werden, dass eine generelle Empfehlung zur Analysebreite bei der lokalen Bildanalyse in dem hier betrachteten Deblocking-Verfahren nicht gegeben werden kann. Auf der Implementierungsseite spricht die etwas größere Regularität des gesamten Verfahrens für acht Pixel Analysebreite. Allerdings ist bei vier Pixeln Analysebreite eine Implementierung des Verfahrens mit nur fünf Zeilenspeichern möglich, was in Realisierungen, in denen die Artefaktreduktion nur eine von vielen Komponenten darstellt, von großem Vorteil sein kann. Grundsätzlich ist der Rechen- und Speicheraufwand aller Varianten als niedrig einzuschätzen.

## 6.4 Erweiterung des Verfahrens

Das in [Pia05a] vorgestellte und in den ersten Abschnitten dieses Kapitels beschriebene Verfahren zur Verringerung von Blockartefakten mit lokaler Bildanalyse ist adaptiv im Sinne der Filterauswahl, der Durchführung oder Aussetzung von Filterungen und der Anpassung der Filter an lokale Bildeigenschaften wie z. B. die Höhe von Intensitätssprüngen. Allerdings ist es notwendig, mehrere Parameter, insbesondere Schwellwerte, fest einzustellen. Im Rahmen dieser Arbeit wurden Erweiterungen an dem Verfahren vorgenommen, um die Adaptivität zu erhöhen und zu überprüfen, ob die Wirksamkeit weiter gesteigert werden kann. Abbildung 6.10 zeigt das Blockschaltbild des erweiterten Verfahrens.



**Abbildung 6.10:** Reduktion von Blockartefakten mit lokaler/globaler Bildanalyse und drei Filtermodi

Auffälligste Neuerung ist ein drittes Filter im Rahmen der adaptiven Filterung, wodurch eine feinere Anpassung der Filterung an verschiedene Bildsituationen ermöglicht werden soll. Es werden homogene und stark homogene Bereiche voneinander unterschieden und entsprechend angepasst gefiltert. Dies ist insofern sinnvoll, als dass homogene Bereiche für die Sichtbarkeit von Störungen von größerer Bedeutung sind als detaillierte Bildbereiche, in denen Maskierungseffekte die Sichtbarkeit von Artefakten verringern. In homogenen Bereichen kann einerseits eine Filterung mit besonders starker Glättungswirkung sinnvoll sein, andererseits gibt es aber auch Situationen, in denen ein Schutz letzter vorhandener Details wünschenswert ist. Eine feinere Anpassung der Filterung in homogenen Bereichen verspricht daher Gewinne. Zu diesem Zweck kommen nun zwei Filter zum Einsatz, die ähnlich wie das Filter des ursprünglichen Verfahrens in Abbildung 6.5 Diskontinuitäten an Blockgrenzen rampenförmig glätten. Die Filter unterscheiden sich in ihrer Wirkungsbreite bzw. in der Länge des rampenförmigen Verlaufs. Das Filter für stark

homogene Bildbereiche wurde auf acht Pixel Wirkungsbreite verlängert. Die Berechnung der Ausgangspixel wird wie folgt durchgeführt:

$$\begin{aligned}
s'_{-4} &= s_{-4} + h/9, \\
s'_{-3} &= s_{-3} + 2h/9, \\
s'_{-2} &= s_{-2} + 3h/9, \\
s'_{-1} &= s_{-1} + 4h/9, \\
s'_0 &= s_0 - 4h/9, \\
s'_1 &= s_1 - 3h/9, \\
s'_2 &= s_2 - 2h/9, \\
s'_3 &= s_3 - h/9.
\end{aligned} \tag{6.11}$$

Das Filter für weniger stark homogene Bereiche wurde auf vier Pixel Wirkungsbreite verkürzt und basiert auf Vorschrift (6.12). In Abbildung 6.11 sind beide Filter für homogene Bereiche dargestellt. Die Filter werden genauso wie im ursprünglichen Verfahren im Falle von großen Signalsprüngen an Blockgrenzen zum Schutz echter Kanten ausgeschaltet.

$$\begin{aligned}
s'_{-2} &= s_{-2} + h/5, \\
s'_{-1} &= s_{-1} + 2h/5, \\
s'_0 &= s_0 - 2h/5, \\
s'_1 &= s_1 - h/5.
\end{aligned} \tag{6.12}$$

Um eine adaptive Wahl zwischen drei Filtern zu ermöglichen, musste auch die lokale Bildanalyse erweitert werden. Dazu wurden weitere Nebengradienten, *Nebengradienten 2* genannt, eingeführt, die gegenüber den ursprünglichen Nebengradienten um ein Pixel weiter von der Blockgrenze entfernt sind (siehe Abbildung 6.12). Das Filter für stark homogene Bereiche wird nur dann ausgewählt, wenn sowohl die lokale Nebengradientensumme als auch die lokale Summe der Nebengradienten 2 unterhalb entsprechender Schwellwerte liegen. Deuten nur die Nebengradienten auf Homogenität hin, wird das Filter für weniger stark homogene Bereiche ausgewählt. Bei Betrachtung von Abbildung 6.12 fällt auf, dass die Messung aller Nebengradienten nicht den gesamten Bereich, der nach der lokalen Bildanalyse gegebenenfalls gefiltert wird, überdeckt. Dies ist jedoch nicht kritisch, da die Pixel, die nicht auf Homogenität bzw. Aktivität geprüft werden, durch das flache Filter für stark homogene Bereiche nur sehr wenig verändert werden und eine weitere Nebengradienten-Messung daher aus Aufwandsgründen nicht angeraten erscheint. Wichtig ist jedoch, dass sich die Messbereiche von Nebengradienten und Nebengradienten 2 überlappen, damit keine Intensitätssprünge zwischen den Messbereichen unentdeckt bleiben.

Eine weitere Veränderung des Artefaktreduktionsverfahrens gegenüber den ursprünglichen Varianten ist eine zusätzliche globale Bildanalyse. Diese besteht aus einer fortlaufenden Akkumulation der lokalen Haupt- und Nebengradienten und der Berechnung eines Verhältnisses aus akkumulierten Hauptgradienten zu akkumulierten Nebengradienten.

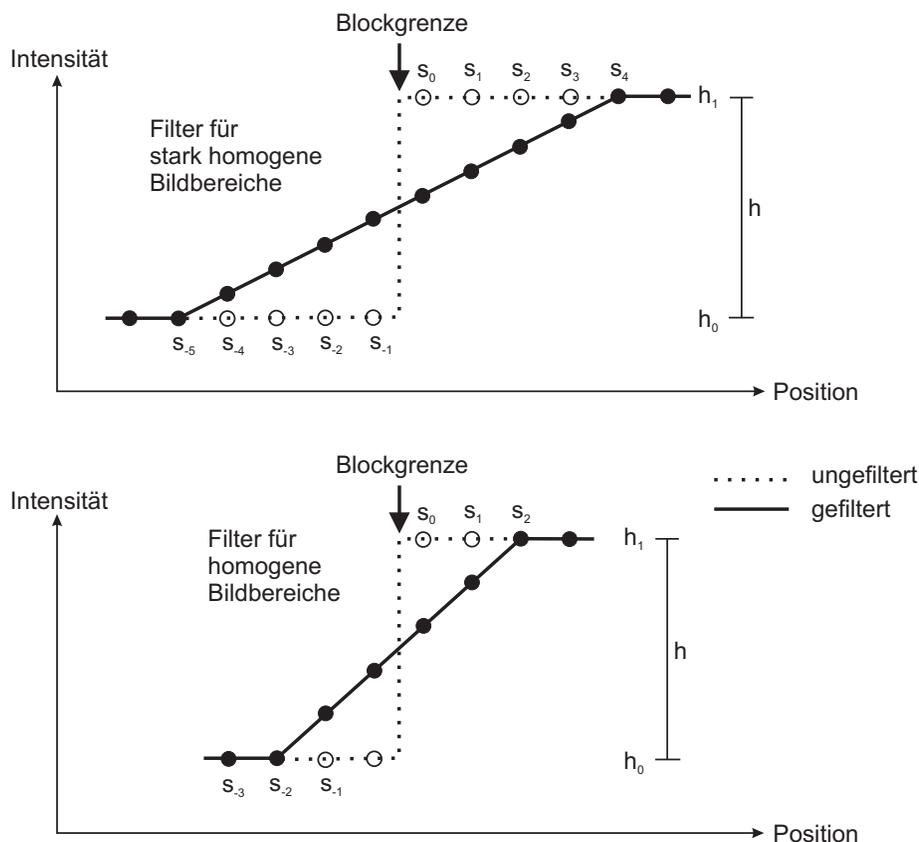
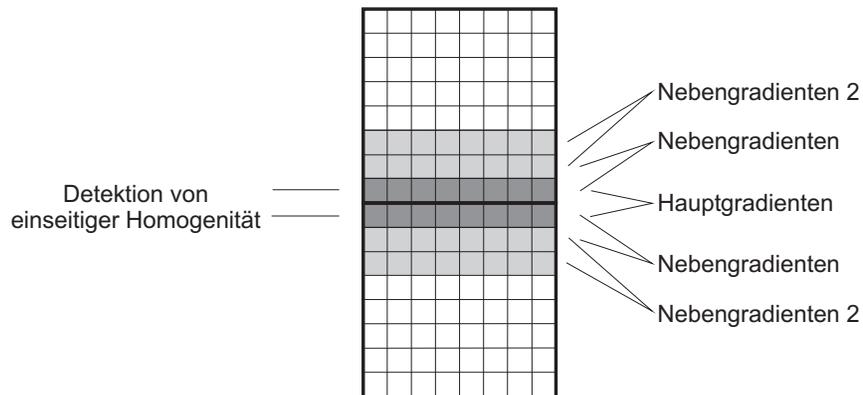


Abbildung 6.11: Filter für homogene und stark homogene Bildbereiche

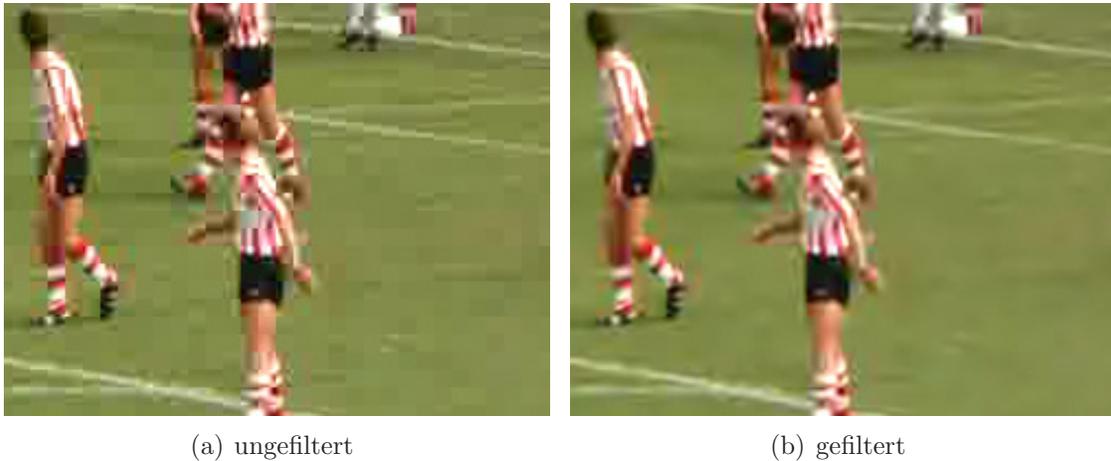
Auf diese Weise wird eine Maßzahl ermittelt, die eine grobe Abschätzung der Stärke von Blockartefakten im aktuellen Bild ermöglicht. Als Startwert wird jeweils die im vorherigen Bild ermittelte Maßzahl mit hohem Gewicht verwendet, um eine starke Schwankung des Maßes zu Beginn eines jeden neuen Bildes zu verhindern. Beim Starten des Algorithmus muss ein geeigneter Startwert vorgegeben werden. Die globale Bildanalyse wird dazu verwendet, eine häufige und starke Filterung zu erzwingen, wenn die ermittelte Maßzahl auf starke Blockartefakte hindeutet und eine seltene und schwache Filterung zu erzwingen, wenn die Maßzahl auf wenige bzw. schwache Blockartefakte hindeutet. In die Adaption werden Schwellwerte im Rahmen der lokalen Blocking-Detektion, die Schwellwerte zum Kantenschutz bei der Filterung homogener Bereiche und der Skalierungsfaktor  $c$  für den DCT-Koeffizienten  $S_{DCT,3}$  im Filter für detaillierte Bildbereiche einbezogen. Insgesamt kann so gleichermaßen ein guter Schutz von nahezu ungestörten Bildern und eine „aggressive“ Filterung von sehr stark gestörten Bildern erreicht werden. Das vorgestellte Verfahren wird im Folgenden *pi7* genannt.



**Abbildung 6.12:** Haupt- und Nebengradienten bei der lokalen Bildanalyse

In Abbildung 6.13 wird die Wirksamkeit des Verfahrens an einem Ausschnitt aus der Sequenz *football* gezeigt. Sowohl in detaillierten als auch in homogenen Bildbereichen werden Blockartefakte stark verringert. Relevante Details werden erhalten. Die Unterschiede zu den bereits vorgestellten einfacheren Varianten der Artefaktreduktion mit lokaler Bildanalyse sind jedoch gering, wie auch die PSNR-Messwerte in Tabelle 6.1 vermuten lassen. In uncodierten Bildern sind die PSNR-Messwerte deutlich besser als bei den anderen Varianten, was zeigt, dass die Adaption des Verfahrens an unterschiedliche Bildqualitätsstufen funktioniert und bei sehr guter Qualität nahezu keine Filterungen durchgeführt werden. Die Unterschiede in codierten Bildern sind gegenüber den einfacheren Varianten aber nicht signifikant. Letztlich ist dies nicht überraschend, da alle Varianten des Verfahrens viele Gemeinsamkeiten in der Funktionsweise aufweisen und sich nur in einigen Details unterscheiden. Aussagekräftiger als PSNR-Werte sind subjektive Eindrücke menschlicher Betrachter, die die Wirksamkeit des Verfahrens bei der Verbesserung der Qualität gestörter Bilder betreffen. In den im Rahmen dieser Arbeit durchgeführten subjektiven Tests (siehe Kapitel 7) konnte eine hohe Wirksamkeit des Verfahrens *pi7* nachgewiesen werden, die tendenziell etwas größer als die der einfacheren Varianten mit nur zwei Filtermodi und ohne globale Bildanalyse ist.

In Abbildung 6.14 ist am Beispiel eines I-Frames aus der Sequenz *football* dargestellt, an welcher Stelle im Bild welches der drei zur Verfügung stehenden Filter vom Verfahren *pi7* angewendet wird. Es soll veranschaulicht werden, wie sich das Verfahren an die Stärke von Blockartefakten in einem Bild und an die Aktivität von Bildbereichen adaptiert. Die zu Teilbild (a) gehörige Sequenz wurde mit 3,5 Mbit/s codiert und die zu Teilbild (b) gehörige Sequenz mit 8 Mbit/s. Stellen, die mit dem Filter für stark homogene Bildbereiche gefiltert werden, sind mit Weiß gekennzeichnet, solche, die mit dem Filter für weniger stark homogene Bildbereiche gefiltert werden, mit Mittelgrau und solche, die mit dem Filter für detaillierte Bildbereiche gefiltert werden, mit Dunkelgrau. Mit Schwarz sind Block-



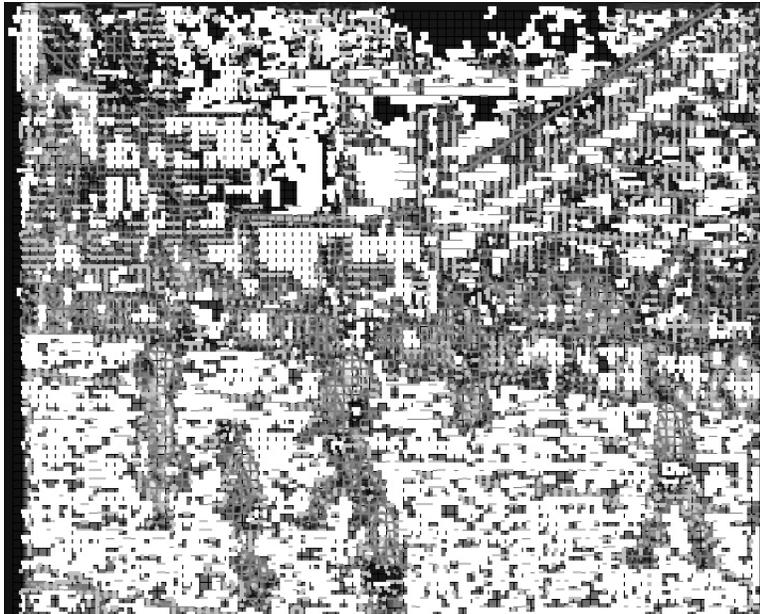
**Abbildung 6.13:** Bildbeispiel für die Artefaktreduktion mit lokaler/globaler Bildanalyse und drei Filtermodi (Ausschnitt)

grenzbereiche gekennzeichnet, an denen die Filterung ausgesetzt wird. Man erkennt, dass bei starker Kompression sehr viele Stellen als stark homogen eingestuft und entsprechend breit gefiltert werden. Bei niedriger Kompression sind solche Stellen hingegen in deutlich geringerer Anzahl erkennbar, während die Zahl der Stellen, an denen nicht gefiltert wird, zunimmt. Detaillierte Bildbereiche werden in beiden Fällen größtenteils richtig erkannt und entsprechend mit dem passenden Filter bearbeitet. Auch das mittellange Filter wird offenbar korrekt eingesetzt.

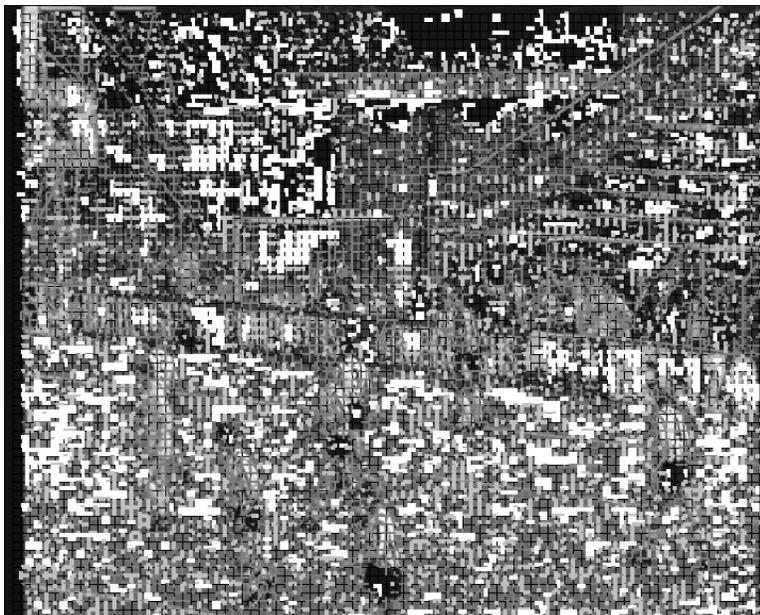
Es kann zusammenfassend festgehalten werden, dass durch die Erweiterung des in [Pia05a] vorgestellten Verfahrens zur Reduktion von Blockartefakten um einen dritten Filtermodus und eine globale Bildanalyse ein Verfahren entwickelt werden konnte, das eine hohe Wirksamkeit bei der Reduktion von Blockartefakten und bei der Verbesserung der Qualität gestörter Bilder besitzt. Die Gewinne gegenüber den ursprünglichen Varianten sind zwar nur moderat, das gilt gleichermaßen aber auch für den zusätzlichen Aufwand, da die globale Bildanalyse auf den Komponenten der lokalen Bildanalyse aufsetzt. Insbesondere wenn noch berücksichtigt wird, dass die erweiterte Variante offensichtlich einen besonders guten Schutz wenig oder gar nicht gestörter Bilder gewährleistet, erscheint das Verfahren für ein Postprocessing in Endgeräten wie z. B. LC-Displays gut geeignet und vorteilhaft gegenüber den ursprünglichen Varianten. Es sei erwähnt, dass in [TCS05] ein adaptives Verfahren zur Reduktion von Blockartefakten vorgestellt wird, das ebenfalls über drei Filtermodi verfügt. Es kombiniert hauptsächlich Komponenten aus den in Kapitel 5 vorgestellten Verfahren nach [KYKR99] und [RK00] und führt eine 3x3 Tiefpassfilterung von Bildbereichen durch, die nicht eindeutig als homogen oder detailliert klassifiziert werden können oder in denen keine eindeutigen Hinweise auf das Vorhandensein von Arte-

fakten bestehen. Die Wirksamkeit des Verfahrens wurde im Rahmen der in dieser Arbeit durchgeführten Tests analysiert (siehe Kapitel 7).

In einer weiteren Variante der im Rahmen dieser Arbeit vorgestellten Blocking-Reduktion wird der Fall von einseitiger Homogenität an Blockgrenzen berücksichtigt. Dazu sind ein Filter für detaillierte Bildbereiche und drei rampenförmige Filter für homogene Bildbereiche vorgesehen, wobei zwei der rampenförmigen Filter nicht symmetrisch zur Blockgrenze gestaltet sind, sondern auf die eine bzw. andere Seite der Blockgrenze verschoben sind. Die Homogenität von zwei benachbarten Blöcken wird im Blockgrenzbereich getrennt analysiert und bei nur einseitiger Homogenität wird jeweils ein Filter ausgewählt, das hauptsächlich auf die homogene Seite wirkt. Für die Artefaktreduktion mit Berücksichtigung einseitiger Homogenität konnten gegenüber den einfachen Varianten mit nur zwei Filtern keine subjektiven oder objektiven Gewinne in den durchgeführten Untersuchungen nachgewiesen werden. Daher wurde der Ansatz verworfen und nicht in die weiteren Untersuchungen einbezogen. Ein Verfahren zur Artefaktreduktion mit vergleichbarer Grundidee wird in [KIL04] propagiert. Da dieses in der Realisierung jedoch Unterschiede zu dem hier vorgestellten Ansatz aufweist, ist eine Übertragung der gewonnenen Erkenntnisse auf das Verfahren nach [KIL04] nicht sinnvoll.



(a) Datenrate 3,5 MBit/s



(b) Datenrate 8 MBit/s

**Abbildung 6.14:** Veranschaulichung der Adaptivität der Blocking-Reduktion mit drei Filtermodi; weiß/mittelgrau/dunkelgrau: Anwendung des Filters für stark homogene/weniger stark homogene/detaillierte Bildbereiche; schwarz: keine Filterung

## 7 Subjektive Tests

Qualitätswerte, die durch objektive Messverfahren ermittelt werden, sollen im Umfeld der Bildsignalverarbeitung in den meisten Fällen möglichst gut mit den subjektiven Eindrücken menschlicher Betrachter übereinstimmen. Diese Forderung war auch im Rahmen der vorliegenden Arbeit bei der Entwicklung eines neuen Verfahrens zur Bewertung der Wirksamkeit von Artefaktreduktionsverfahren zu erfüllen. Insofern war die Durchführung umfangreicher subjektiver Tests notwendig, um eine optimale Abbildung der einzelnen Maßzahlen bzw. Qualitätsfaktoren des auf einem Faktorenmodell basierenden neuen Verfahrens auf einen einzelnen Qualitätswert vornehmen zu können und um Daten zu gewinnen, an denen die Güten der Schätzungen der in dieser Arbeit untersuchten Qualitätsbewertungsverfahren beurteilt werden können.

Zweck des neuen Verfahrens, das in Kapitel 8 detailliert vorgestellt wird, ist die Ermittlung eines Qualitätswertes, der eine Aussage darüber macht, ob und um welchen Grad ein Deblocking-Verfahren die Gesamtqualität eines mit sichtbaren Blockartefakten versehenen - also im überschwelligem Bereich gestörten - Einzelbildes verbessern kann. Es ist klar, dass durch eine Qualitätsbewertung auf Basis von Einzelbildern zeitliche Effekte, die bei der Qualitätsanalyse von Videosequenzen eine nicht zu vernachlässigende Rolle spielen können, nicht erfasst werden. Um den Untersuchungsumfang zu begrenzen, konzentrierten sich die Untersuchungen in dieser Arbeit jedoch größtenteils auf örtliche Aspekte von Codierartefakten, was bei der Gestaltung der subjektiven Tests entsprechend zu berücksichtigen war. Ohnehin arbeiten die in die Untersuchungen einbezogenen Deblocking-Verfahren rein örtlich, so dass eine Konzentration auf die Bewertung örtlicher Effekte und Artefakteigenschaften auch aus diesem Grund sinnvoll war. Dennoch wurden einige Sequenzen in einen der subjektiven Tests aufgenommen, um erste Eindrücke über die Wirksamkeit der eingesetzten Deblocking-Verfahren in Bildsequenzen zu erhalten und um eine eventuelle Erzeugung ungewünschter neuer zeitlicher Artefakte durch einzelne Verfahren erkennen zu können. Aufgrund der größten Praxisrelevanz wurde farbiges Bildmaterial im YCrCb-Format für die subjektiven Tests verwendet. Auf eine Artefaktreduktion in den Farbdifferenzkanälen wurde jedoch verzichtet. Die Gründe dafür werden in diesem Kapitel noch dargelegt.

Grundsätzlich mussten zur Gewinnung von für die weiteren Untersuchungen geeigneten Daten jeweils zwei Bilder miteinander verglichen werden: ein gestörtes (im Folgenden *codiertes* Bild genannt) und ein gestörtes und nachgefiltertes Bild. In der ITU-R Recommendation BT.500 [ITU02] findet sich im Kapitel *Alternative methods of assessment* die Beschreibung eines für diese Problemstellung geeigneten subjektiven Testablaufs unter

den so genannten *stimulus-comparison-Methoden*. Die drei subjektiven Tests in dieser Arbeit wurden weitgehend an die beschriebenen Randbedingungen angelehnt und werden im Folgenden dargestellt und ausgewertet. Die ersten beiden Tests waren im Rahmen der Entwicklung des neuen Qualitätsbewertungsverfahrens von besonderer Wichtigkeit und werden daher detaillierter behandelt als der weniger umfangreiche dritte Test, der im Wesentlichen der Validierung des Verfahrens diente.

### 7.1 Randbedingungen

In einem abgedunkelten Raum mit geringer Resthelligkeit wurde ein High Definition LC-Display vom Typ *Sharp LC-37HV4E* mit 37 Zoll Bildschirmdiagonale ( $H=46$  cm Bildhöhe) aufgestellt und mit seiner nativen Auflösung von  $1366 \times 768$  Bildpunkten von einem Personal Computer mit der Sequenzanlagen-Software *Seqplay* [Dol03] angesteuert. Schwarzwert, Helligkeit und Kontrast wurden mit Hilfe eines Pluge Test Patterns (Pluge  $\hat{=}$  Picture Line-Up Generation Equipment) eingestellt. Mit einem Colorimeter vom Typ *Photo Research SpectraScan PR650* wurden in Display-Mitte folgende Leuchtdichten ermittelt:

- weißer Bildschirm:  $199,5 \text{ cd/m}^2$ ,
- schwarzer Bildschirm:  $1,05 \text{ cd/m}^2$ ,
- inaktiver Bildschirm: unterhalb der Messgrenze.

Um das Display herum wurden graue Vorhänge aufgehängt, um eine Ablenkung der Teilnehmer an den subjektiven Tests durch andere im Raum befindliche optische Reize zu vermeiden. Ein Foto des Aufbaus ist in Abbildung 7.1 dargestellt. Für ein Display der verwendeten Größe wird in [ITU02] ein Betrachtungsabstand im Bereich der sechsfachen Bildhöhe, also ca. 2,80 m, vorgeschlagen. Bei diesem Betrachtungsabstand konnten selbst von Experten auch in stark komprimierten Bildern nur vereinzelt und unter größter Anstrengung Blockartefakte wahrgenommen werden.

Grundsätzlich sollten die allgemeinen Empfehlungen zum Betrachtungsabstand aus der ITU-R Recommendation BT.500 heutzutage fallbezogen kritisch hinterfragt werden. Einerseits sorgen in typischen Wohnsituationen immer größer werdende Displays in Räumlichkeiten gleichbleibender Ausdehnung gegenüber früheren Zeiten für verringerte relative Betrachtungsabstände. Andererseits ist die Verfügbarkeit neuer mobiler Endgeräte (z. B. videofähige Mobiltelefone) zu nennen, deren Einsatz in der Praxis mit sehr unterschiedlichen (z. T. auch sehr großen) relativen Betrachtungsabständen einhergeht. Daher ist die Bindung des empfohlenen Betrachtungsabstands einzig an die Bildhöhe aktuell sehr fragwürdig, was auch in [Win05] ähnlich gesehen wird. Ein weiterer zu berücksichtigender Aspekt ist die Durchdringung des Marktes mit unterschiedlichen Display-Technologien.



**Abbildung 7.1:** Aufbau des subjektiven Tests

Um dies zu verdeutlichen, wurde für verschiedene am Arbeitsgebiet Schaltungen der Informationsverarbeitung der Technischen Universität Dortmund verfügbare Displays die Auflösung der Displays in  $cy/^\circ$  bei Einhaltung des empfohlenen relativen Betrachtungsabstands berechnet. Darüber hinaus wurde berechnet, welcher Betrachtungsabstand für eine Auflösung von  $30\text{ }cy/^\circ$  zu wählen wäre. Diese Auflösung liegt im Bereich des maximalen Auflösungsvermögens des menschlichen Auges unter optimalen Bedingungen, für das gemeinhin eine Bogenminute angenommen wird. Die Ergebnisse der Berechnungen finden sich in Tabelle 7.1.

**Tabelle 7.1:** Kennzahlen verschiedener Displays, Längenangaben in cm

Display	Bildhöhe	Zeilen	Abstand	Aufl. in $cy/^\circ$	Abst. für $30\text{ }cy/^\circ$
Dell 2405 PFW	32,5	1200	217,75	70,29	92,16
Sharp LC-37HV4E	46	768	276	40,31	205,05
Sharp LC-45GD1E	56	1080	313,6	52,92	176,79
NEC PDP 42"	52	480	301,6	24,36	371,82
Nokia 459A	23	768	179,4	52,35	102,52
Philips 200P3	31	1200	210,8	71,34	87,90

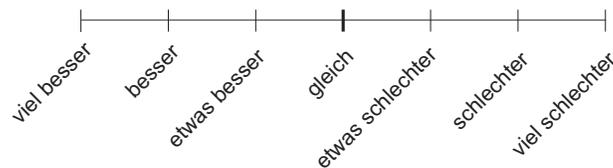
Vergleicht man beispielsweise das 42 Zoll Plasmapanel von NEC mit dem 45 Zoll großen LC-Display Sharp LC-45GD1E, so fällt eine deutlich unterschiedliche Auflösung bei vergleichbarer Bildschirmdiagonale auf. Dies wirkt sich auch auf den berechneten Betrachtungsabstand für eine Auflösung von  $30\text{ }cy/^\circ$  aus, der bei dem Plasmadisplay etwa doppelt so groß ist wie bei dem LC-Display. Diese Zahlen bestätigen die Zweifel an den Empfeh-

lungen zum Betrachtungsabstand in [ITU02] und legen insbesondere bei hochauflösenden Displays verkürzte Betrachtungsabstände nahe. Dies deckt sich auch mit Forderungen an HDTV-Systeme, die bereits in [WS91] formuliert wurden. Dort wurde als Betrachtungsabstand konkret die zweifache Bildhöhe genannt. In der Literatur finden sich verschiedene Fälle, in denen Autoren die Betrachtungsabstände in subjektiven Tests an ihre Anforderungen angepasst bzw. verkürzt haben und von den Empfehlungen der ITU-R abgewichen sind [VQE03, ASS02, KK95, HN94]. Es wurden in den genannten Quellen Betrachtungsabstände zwischen der dreifachen und fünffachen Bildhöhe verwendet. Entscheidend dürfte letztlich die Beantwortung der Frage sein, ob in einer bestimmten Anwendung, die unter anderem auch durch einen bestimmten Betrachtungsabstand gekennzeichnet ist, auf einem zu benutzenden Display sichtbare Darstellungsprobleme oder Störungen auftreten oder nicht. Bezüglich der für diese Arbeit relevanten Codierartefakte lässt sich die Frage auf Basis von Praxiserfahrung guten Gewissens mit ja beantworten, wenn an Systeme wie DVB-T, Video-CD oder an niederratige Videoübertragungen auf mobile Geräte bei jeweils praxistypischen Betrachtungsabständen gedacht wird.

Für die durchgeführten Tests wurde basierend auf den beschriebenen Überlegungen ein Betrachtungsabstand von 1,40 m gewählt, was bei dem verwendeten Display etwa der dreifachen Bildhöhe entspricht. Für die Probanden waren Blockartefakte dadurch ohne größere Schwierigkeiten, aber nicht überdeutlich sichtbar. Die Probanden bestätigten, dass die Tests ein hohes Maß an Konzentration erforderten. Bei dem gewählten Betrachtungsabstand in Kombination mit dem verwendeten Display ergibt sich eine Auflösung von ca. 21 *cy*/°. Höhere Auflösungen können vom menschlichen Auge ohnehin nur unter optimalen Bedingungen insbesondere bezüglich des Kontrasts von zu betrachtenden Details wahrgenommen werden. Von optimalen Bedingungen kann bei Betrachtung unterschiedlichsten Bildmaterials aber nicht ausgegangen werden. Gerade die in dieser Arbeit relevanten Blockartefakte zeichnen sich durch Diskontinuitäten mit eher kleinem Kontrastumfang aus. Es sei angemerkt, dass davon ausgegangen werden kann, dass sich die in einem subjektiven Test bei einem bestimmten Betrachtungsabstand ermittelten Ergebnisse durch Skalierung in gewissen Grenzen auch auf andere Betrachtungsabstände übertragen lassen. Die entsprechende Skalierungsfunktion müsste dazu allerdings in einem weiteren subjektiven Test, der verschiedene Betrachtungsabstände berücksichtigt, ermittelt werden. Ähnliches gilt für andere Display-Auflösungen.

In den im Rahmen dieser Arbeit durchgeführten subjektiven Tests, an denen jeder Proband einzeln teilnahm, wurden auf dem Display nebeneinander immer zwei progressive Bilder gezeigt: ein gestörtes und ein gestörtes und nachgefiltertes Bild. Es wurde bewusst rein progressives Bildmaterial verwendet, um Störungen durch das zur Darstellung auf progressiven Displays notwendige Deinterlacing zu vermeiden und somit einen Einfluss von potenziellen Deinterlacing-Störungen auf die Ergebnisse der subjektiven Tests auszuschließen. Auf welcher Seite welches Bild gezeigt wurde, war den Teilnehmern nicht bekannt und variierte von Bild zu Bild. Die Auflösung des Displays ermöglichte die nahezu

komplette Darstellung von zwei SD-Bildern der Auflösung 720x576 Bildpunkte. Lediglich am rechten und linken Rand der Bilder mußten jeweils ca. 20 Pixel weggeschnitten werden. Der Ablauf der Tests war vollautomatisch. Es wurde jeweils ein Paar von Bildern 20 Sekunden (beim dritten Test 23 Sekunden) lang nebeneinander gezeigt. Danach folgte die Einblendung von mittlerem Grau für eine Dauer von 8 Sekunden (beim dritten Test 7 Sekunden) zur Abgabe der Bewertung, bevor das nächste Paar von Bildern dargestellt wurde. Die Probanden erhielten Bögen mit 6 cm langen Vergleichsskalen wie in Abbildung 7.2, auf denen eine Qualitätsbewertung durch das Setzen von Strichen oder Kreuzen erfolgen sollte. Die Benutzung der Skalen sollte kontinuierlich erfolgen, das heißt, dass beliebige Zwischenwerte zwischen zwei Skalenstrichen erlaubt waren. Dadurch sollte ein mögliches Zögern der Probanden bei einem wahrgenommenen Qualitätsunterschied zwischen zwei Skalenstufen verringert werden. Zu bewerten war grundsätzlich die Qualität des linken Bildes bezogen auf die Qualität des rechten Bildes. Für die Auswertung der Testbögen wurden die Positionen der Kreuze gemessen und entsprechend der Comparison-Skala aus [ITU02] auf Werte im Bereich von -3 (viel schlechter) bis +3 (viel besser) abgebildet, so dass DMOS-Werte (Differential Mean Opinion Score) berechnet werden konnten.



**Abbildung 7.2:** In den subjektiven Tests eingesetzte Vergleichsskala

Eine ausreichende Sehschärfe der Probanden wurde mit Hilfe eines Snellen-Charts sichergestellt. Außerdem wurden die Probanden einem Ishihara-Test auf Farbschwächen unterzogen. Letztendlich wurden jedoch auch einige wenige Personen mit Farbschwächen zu den subjektiven Tests zugelassen, da es in den Tests grundsätzlich nicht um die Identifizierung oder Beurteilung von Farbfehlern ging und die eingesetzten Artefaktreduktionsverfahren ohnehin nur auf den Luminanzkanal angewendet wurden.

## 7.2 Bildmaterial

Als Ausgangsmaterial für die beiden ersten Tests dienten die YCrCb-Sequenzen *football*, *lena2\_zoom2*, *stockholm4* und *wheel* (siehe Anhang E). Damit sollten unterschiedliche Klassen von Bildmaterial abgedeckt werden: Sportübertragung, menschliches Porträt, „Landschaftsaufnahme“ und synthetisches Material. Es wurde darauf geachtet, dass ein ausgewogenes Verhältnis aus Konturen, Texturen und homogenen Bereichen im Bildmaterial vorhanden war. Die Sequenzen wurden mehrfach abwechselnd in richtiger und

in umgekehrter Abspielreihenfolge hintereinanderkopiert (Pingpong-Loop) und mit Hilfe des MPEG2-Encoders der MPEG Software Simulation Group [MSS], Testmodel 5, mit den Datenraten 3,5 Mbit/s und 5 Mbit/s codiert. Auf diese Weise ergaben sich bei dem gewählten Betrachtungsabstand sowohl Bilder bzw. Bildsequenzen mit deutlichen Artefakten, als auch solche mit vergleichsweise wenigen und schwach ausgeprägten Artefakten. Bei höheren Datenraten wären nicht mehr genügend Artefakte für einen sinnvollen Vergleich von Deblocking-Verfahren sichtbar gewesen. Zusätzlich wurden von den Sequenzen *football* und *wheel* Varianten erzeugt, die ausschließlich I-Frames bei einer Datenrate von 12 Mbit/s enthielten. Da sich mit einer reinen Intra-Codierung nur geringere Kompressionsgewinne erzielen lassen als mit einer hybriden Codierung, musste die Datenrate deutlich erhöht werden, um die Bildqualität nicht zu schlecht werden zu lassen. Auf die gestörten Bildsequenzen wurden insgesamt sieben Deblocking-Verfahren angewendet:

- statische Filterung von Blockgrenzen nach [ANT92],
- adaptive Filterung von Blockgrenzen nach [KYKR99],
- wavelet-basierte Blocking-Reduktion nach [KJKH98],
- Blocking-Reduktion mit lokaler Bildanalyse und zwei Filtermodi (siehe Kapitel 6),
- Blocking-Reduktion mit lokaler/globaler Bildanalyse und drei Filtermodi (siehe Kapitel 6),
- adaptive Filterung von Blockgrenzen nach [RK00],
- Subtraktion von Fehlersignalen nach [YCK00].

Grundsätzlich fand eine Artefaktreduktion nur in den Luminanzkanälen der Sequenzen statt. In YCrCb-Systemen ist der Luminanzkanal Träger der wesentlichen Bildinformationen und auch maßgeblich für die Entstehung bzw. Sichtbarkeit von Blockartefakten verantwortlich. Die in Kapitel 2 beschriebenen sprungartigen oder blockartig ausgedehnten Farbstörungen treten in der Praxis im Vergleich zu den klassischen Blockartefakten in der Luminanz deutlich seltener und vorzugsweise bei geringen bis sehr geringen Datenraten auf. Bevor die Entscheidung getroffen wurde, nur die Luminanzkanäle zu bearbeiten, wurde exemplarisch an mehreren stark gestörten Einzelbildern geprüft, wie sich eine zusätzliche Artefaktreduktion in den Chrominanzkanälen auf die Bildqualität auswirkt. Dazu wurde je eine Variante eines gestörten Bildes mit und ohne Deblocking-Filterung in den Chrominanzkanälen direkt miteinander verglichen. Bei einem geringen Betrachtungsabstand von ein bis zwei Bildhöhen konnten zwar stellenweise Unterschiede gefunden werden, allerdings führte das Deblocking der Chrominanzkanäle nicht in jedem Fall zu einer deutlichen Verringerung der Farbfehler, sondern oft nur zu einer Verschmierung

bzw. Aufweichung der Fehler. Die Verbesserung der Bildqualität durch das Deblocking der Chrominanzkanäle wurde als wesentlich geringer empfunden als die Verbesserung der Bildqualität durch das Deblocking des Luminanzkanals. Bei dem vorgesehenen Betrachtungsabstand von drei Bildhöhen waren schließlich nur sehr vereinzelt Gewinne durch die Bearbeitung der Farbdifferenzkanäle sichtbar. Insgesamt wurde der zusätzliche Aufwand für eine Deblocking-Filterung in den Chrominanzkanälen als nicht hinreichend lohnenswert erachtet und die Blocking-Reduktion daher auf die Luminanz beschränkt.

Durch die Codierung der vier erwähnten Sequenzen mit zwei Datenraten und die Anwendung von sieben Artefaktreduktionsverfahren ergab sich eine umfangreiche Auswahl an Bildmaterial für die beiden ersten subjektiven Tests. Zusätzlich wurden codierte und gefilterte Varianten der Sequenzen *matphone\_lauf0* und *coralreef* erzeugt, aus denen Einzelbilder im Rahmen der Einweisung der Probanden in den Ablauf der subjektiven Tests verwendet wurden. Dabei wurden den Probanden Blockartefakte gezeigt und die Auswirkungen von wirksamen und weniger wirksamen Artefaktreduktionsverfahren auf gestörte Bilder vorgeführt. Es wurde des Weiteren gezeigt, dass Verfahren zur Reduktion von Blockartefakten stellenweise Unschärfe erzeugen können, was die Probanden gegebenenfalls abwertend berücksichtigen sollten.

Für den dritten subjektiven Test, der zu Validierungszwecken diente, war es entscheidend, Bildmaterial zu verwenden, das keine großen Ähnlichkeiten zu dem Bildmaterial in den anderen Tests aufweist. Deshalb wurden als Basis die YCrCb-Sequenzen *fries2*, *pedestrian3* und *shields3* (siehe Anhang E) ausgewählt. Auch hier wurde auf ein ausgewogenes Verhältnis aus Konturen, Texturen und homogenen Bereichen geachtet. Da nur Einzelbilder aus den Sequenzen verwendet wurden, war keine Erzeugung von Pingpong-Loops notwendig. Die Sequenzen *fries2* und *pedestrian3* wurden mit einer Datenrate von 1,8 Mbit/s MPEG-2-codiert, die Sequenz *shields3* mit 3 Mbit/s. Die sehr niedrig erscheinende Datenrate von 1,8 Mbit/s wurde deshalb gewählt, weil die damit codierten Sequenzen nur 25 Frames/s im Gegensatz zu den anderen Sequenzen mit 60 Frames/s enthielten. Insgesamt ergaben sich Sequenzen mit deutlich erkennbaren Codierartefakten. Auf die gestörten Bildsequenzen (bzw. auf deren Luminanzkanal) wurden insgesamt acht Deblocking-Verfahren angewendet, darunter zwei, die in den ersten beiden Tests nicht eingesetzt wurden:

- Mittelwertfilterung des gesamten Bildes mit einem Filter der Größe 3x3 Pixel,
- statische Filterung von Blockgrenzen nach [ANT92],
- adaptive Filterung von Blockgrenzen nach [KYKR99],
- wavelet-basierte Blocking-Reduktion nach [KJKH98],
- Blocking-Reduktion mit lokaler/globaler Bildanalyse und drei Filtermodi (siehe Kapitel 6),

- adaptive Filterung von Blockgrenzen nach [RK00],
- adaptive Filterung von Blockgrenzen mit drei Filtermodi nach [TCS05],
- Subtraktion von Fehlersignalen nach [YCK00].

Somit stand auch für den dritten Test ein umfangreicher Pool an Bildmaterial zur Verfügung. Keine der Sequenzen aus den ersten Tests wurde erneut verwendet. Die Einführung der Probanden in den dritten Test erfolgte an Einzelbildern aus den Sequenzen *matphone\_lauf0* und *football*.

### 7.3 Test 1

Für den ersten subjektiven Test wurden 47 Einzelbildpaare und acht Bildsequenzpaare ausgewählt. Die Einzelbildpaare, welche am Anfang des Tests gezeigt wurden, enthielten ausschließlich I-Frames der codierten Sequenzen und deren nachgefilterte Entsprechungen. Die Sequenzen bestanden, wie bereits erwähnt, ebenfalls ausschließlich aus I-Frames. Der Grund hierfür ist im Zusammenhang mit dem Ziel des Tests - der Bewertung der Wirksamkeit von Deblocking-Verfahren - zu sehen. Voraussetzungen für eine sinnvolle Wirksamkeitsbewertung sind in diesem Fall eine deutliche Sichtbarkeit und ein häufiges Auftreten von Artefakten. Der eingesetzte MPEG-Encoder der Software Simulation Group [MSS] neigt (wie offenbar auch andere Encoder, siehe [LJL<sup>+</sup>03]) dazu, in I-Frames erheblich mehr Artefakte zu erzeugen als in anderen Frame-Typen und stellt dort die Voraussetzungen sicher. Hinzu kommt, dass in P-Frames und in B-Frames auch andere Arten von Artefakten entstehen, die z. B. auf Bewegungsschätzfehler zurückzuführen sind. Des Weiteren können Diskontinuitäten an Stellen auftreten, die aufgrund fester Verarbeitungsraster von nahezu keinem Deblocking-Verfahren verringert werden können. Es ist nicht sinnvoll, Artefakte in den zu bewertenden Bildern zu zeigen, gegen deren Verringerung die eingesetzten Verfahren machtlos sind. I-Frames enthalten solche Artefakte in der Regel nicht.

Die hauptsächliche Verwendung von Einzel- bzw. Standbildern hatte zweierlei Gründe. Zum einen war die Verwendung von Standbildern sinnvoll, weil das in dieser Arbeit entwickelte Qualitätsbewertungsverfahren (siehe Kapitel 8) genauso wie die implementierten Deblocking-Verfahren rein örtlich arbeitet und keine zeitlichen Effekte berücksichtigt. Zum anderen sollte den Probanden durch die Anzeige von Standbildern ausreichend Zeit und Ruhe zum detaillierten Vergleich von ungefiltertem und gefiltertem Bildmaterial verschafft werden. Als Kriterien sind in diesem Zusammenhang z. B. die Erhaltung von Details oder das Entstehen von Unschärfe zu nennen. Die Sequenzen am Ende des Tests wurden gezeigt, um mögliche Probleme einzelner Algorithmen, die nur in der Bewegung sichtbar sind, aufzudecken und um erste Erkenntnisse darüber zu gewinnen, inwieweit eine Übertragung

der Wirksamkeitsbewertung der untersuchten Verfahren von Standbildern auf Sequenzen möglich oder problematisch ist.

Da ein subjektiver Test keinesfalls länger als 30 Minuten dauern sollte [ITU02], wurde eine Auswahl des verfügbaren Bildmaterials gezeigt. Es wurden nahezu alle Kombinationen aus Sequenz bzw. Bild und Deblocking-Verfahren bei 3,5 Mbit/s aufgenommen, um umfangreiche Informationen zum Vergleich der Deblocking-Verfahren zu gewinnen. Bei 5 Mbit/s wurden nur *football* und *wheel* verwendet und es wurden Deblocking-Verfahren, die anderen sehr ähnlich sind, nicht in den Test einbezogen. Der Test wurde zur Plausibilitätskontrolle mit mehrfach gezeigten Bildpaaren und mit identischen Bildpaaren angereichert. Die ersten fünf gezeigten Bildpaare dienten ausschließlich der Gewöhnung der Probanden an den Test und wurden nicht ausgewertet. Die Reihenfolge der gezeigten Bildpaare folgte keinem System und kann als zufällig betrachtet werden. Als Sequenzen wurden ebenfalls *football* und *wheel* in jeweils vier unterschiedlichen nachgefilterten Varianten ausgewählt. Insgesamt ergab sich eine Länge des Tests von knapp 26 Minuten. Die Playlist des Tests findet sich in Anhang A.

An dem Test nahmen insgesamt 29 Personen teil, darunter 21 Nicht-Experten und acht Experten. Von den 29 Personen waren fünf Frauen und 24 Männer. Es wurden 21 Testbögen verteilt, auf denen für jedes Bildpaar bzw. Sequenzpaar ausschließlich eine Bewertung der Gesamtbildqualität des linken Bildes im Vergleich zum rechten Bild vorzunehmen war (siehe Abbildung 7.2). Darüber hinaus wurde in acht Fällen ein erweiterter Fragebogen ausgeteilt, auf dem neben einem Gesamtbildqualitätsvergleich auch ein Vergleich speziell bezogen auf die Stärke von Blockartefakten und bezogen auf die Bildschärfe durchgeführt werden sollte. Sinn des erweiterten Fragebogens war die Gewinnung zusätzlicher Informationen über die einzelnen Artefaktreduktionsverfahren. Insbesondere sollte ermittelt werden, wie sich die Reduktion speziell von Blockartefakten einerseits und Unschärfenbildung andererseits auf die von den Testteilnehmern wahrgenommene Gesamtbildqualität auswirken. Es zeigte sich, dass der erweiterte Fragebogen die Probanden oft überforderte, was diese auch äußerten. Beim Setzen von Kreuzen auf den jeweils drei Skalen kam es oft zu offensichtlichen Vertauschungen und die Analyse der Ergebnisse legte den Verdacht nahe, dass die Probanden den Gesamtqualitätsvergleich zu stark und sehr statisch von den anderen beiden Vergleichen abhängig machten. Als Konsequenz wurden die erweiterten Bögen zwar ausgewertet, die Ergebnisse wurden jedoch nicht für nachfolgende Untersuchungen verwendet.

## 7.4 Test 2

Nachdem sich die erweiterten Fragebögen in dem ersten subjektiven Test als problematisch erwiesen hatten, wurde ein zweiter subjektiver Test mit 16 Teilnehmern durchgeführt, darunter 11 Nicht-Experten und 5 Experten. Eine Frau nahm an dem Test teil, ansonsten nur

Männer. Ziel des Tests war die Gewinnung weiterer Informationen zu den untersuchten Artefaktreduktionsverfahren, was ursprünglich durch die erweiterten Fragebögen im Rahmen des ersten Tests abgedeckt werden sollte. Für 45 Bildpaare war die Schärfe des linken Bildes verglichen mit der des rechten Bildes zu bewerten und es musste angegeben werden, inwieweit sich die Qualität der Bilder speziell bezogen auf Blockartefakte unterschied. Auf einen Gesamtqualitätsvergleich wurde verzichtet. Das ausgewählte Testmaterial entsprach dem des ersten subjektiven Tests, allerdings wurden die doppelt gezeigten Bildpaare, die der Plausibilitätskontrolle dienten, gestrichen, und es wurden keine Bildsequenzen gezeigt. Dadurch konnte der Test um fünf Minuten auf 21 Minuten Dauer verkürzt werden, was aufgrund von Anmerkungen seitens der Teilnehmer des ersten Tests angeraten erschien. Da die Ergebnisse ohnehin einem Screening nach [ITU02] unterzogen wurden und die Ergebnisse als weniger kritisch für die weiteren Untersuchungen einzustufen waren, konnte die Plausibilitätskontrolle gegenüber dem ersten Test vereinfacht werden. Auch in diesem Test dienten die ersten fünf gezeigten Bildpaare ausschließlich der Gewöhnung der Probanden an den Testablauf und wurden nicht ausgewertet. Selbstverständlich wurden die Probanden in den Testablauf eingewiesen. Die Playlist für Test 2 findet sich in Anhang A.

### 7.5 Test 3

Für den dritten subjektiven Test, der - wie bereits erwähnt - der Validierung des im Rahmen dieser Arbeit entwickelten Qualitätsbewertungsverfahrens diente, wurden 31 Einzelbildpaare ausgewählt. Die Einzelbildpaare bestanden aus I-Frames der codierten Sequenzen und deren nachgefilterten Entsprechungen. Auch hier kam es wie in den ersten beiden Tests auf die deutliche Sichtbarkeit von Blockartefakten an und es galten grundsätzlich die bereits für den ersten Test beschriebenen Überlegungen. Es wurden alle für den dritten Test vorbereiteten Kombinationen aus Sequenzen bzw. Bildern und Deblocking-Verfahren aufgenommen, um einen möglichst umfangreichen Datensatz zu gewinnen. Auch zwei Bildpaare, die aus auf beiden Seiten ungefilterten und somit identischen Bildern bestanden, wurden in den Test eingefügt. Die ersten fünf gezeigten Bildpaare waren Duplikate von später nochmals gezeigten Paaren und dienten ausschließlich der Gewöhnung der Probanden an den Testablauf. Sie wurden nicht ausgewertet. Die Reihenfolge der gezeigten Bildpaare folgte keinem System und kann als zufällig betrachtet werden. Der Test dauerte ca. 16 Minuten. Es nahmen insgesamt 18 Personen teil, darunter zwölf Nicht-Experten und sechs Experten. Unter den 18 Personen waren zwei Frauen und 16 Männer. Es musste ausschließlich eine Bewertung für die Gesamtbildqualität des linken Bildes im Vergleich zum rechten Bild entsprechend Abbildung 7.2 abgegeben werden. Die Playlist für Test 3 findet sich in Anhang A.

## 7.6 Auswertung der subjektiven Testdaten

Da in den gezeigten Bild- oder Sequenzpaaren die mit Artefaktreduktionsverfahren bearbeiteten Bilder sowohl auf der rechten als auch auf der linken Bildschirmseite gezeigt wurden, die Probanden aber immer die Qualität des linken Bildes bezogen auf die des rechten Bildes bewerten sollten, mussten für die Auswertung systematische Vertauschungen der Vorzeichen der Bewertungen vorgenommen werden. Dies war immer dann der Fall, wenn das gefilterte Bild auf der rechten Seite gezeigt wurde, so dass eine Qualitätsverbesserung durch ein Verfahren nach Durchführung der Vertauschungen stets an einem positiven Bewertungsvorzeichen abgelesen werden konnte.

### 7.6.1 Ausschluss unzuverlässiger Daten

Wie bereits erwähnt, dienten in allen drei Tests die ersten fünf gezeigten Bildpaare zur Gewöhnung der Probanden an die Testbedingungen und wurden daher grundsätzlich nicht ausgewertet. Die zur Auswertung vorgesehenen restlichen Daten wurden zunächst auf versehentlich vertauschte Vorzeichen untersucht. Deutete die Bewertung eines Bildpaares beim Vergleich mit den Bewertungen der anderen Probanden eindeutig auf eine Vertauschung hin und war der Betrag der abgegebenen Bewertung mindestens 1,4 - also relativ groß - so wurde ein Rücktausch des Vorzeichens vorgenommen. Letztlich waren im Rahmen des ersten subjektiven Tests in den zur Auswertung verwendeten Testbögen nur bei fünf von ca. 800 Werten Änderungen der Vorzeichen notwendig. Im zweiten subjektiven Test war die Quote der vertauschten Vorzeichen höher, was damit zu begründen ist, dass die Abfrage zweier Bewertungen, die sich typischerweise gegenläufig verhalten, leichter zu Vertauschungen führen kann. Die sichere Erkennung von Vertauschungen war einfacher, da in der Regel beide Bewertungen eines Bildpaares betroffen waren. Der Schwellwert von 1,4 musste zum Auslösen eines Vorzeichenrücktauschs nur bei einer der paarweisen Bewertungen erreicht werden. Insgesamt wurden 14 Vorzeichen von ca. 1300 Werten geändert. Im dritten subjektiven Test wurde nur ein einziger Vorzeichenrücktausch vorgenommen, obwohl der Schwellwert auf 1,0 verringert wurde.

Die Testbögen des ersten subjektiven Tests wurden einer sehr restriktiven Plausibilitätskontrolle und anschließend einem Screening-Verfahren unterzogen, um unzuverlässig bewertende Probanden aus der weiteren Auswertung auszuschließen. Innerhalb des relevanten Testumfangs wurden entsprechend einem Vorschlag aus [ITU02] zwei Bildpaare gezeigt, die aus identischen Bildern bestanden. Darüber hinaus wurden drei Bildpaare zweimal gezeigt. Gab ein Proband bei mindestens einem der Bildpaare, die aus identischen Bildern bestanden, eine Bewertung ab, die um mehr als eine Skalenstufe von Null abwich, oder wurde bei mindestens einem der doppelt gezeigten Bildpaare eine um mehr als 1,5 Skalenstufen abweichende Bewertung abgegeben, wurde der Proband als unzuverlässig

eingestuft und nicht weiter berücksichtigt. Von den 21 Testbögen wurden auf diese Weise vier aussortiert.

Nach dem Rücktausch versehentlich vertauschter Vorzeichen und dem Ausschluss unzuverlässiger Probanden (entsprechend den beschriebenen Kriterien) wurden für jedes Bild- bzw. Sequenzpaar  $j$  aus den übrig gebliebenen  $N$  Testbögen die Stichprobenmittelwerte  $\bar{u}_{Qj}$  der abgegebenen Bewertungen der Gesamtqualität  $u_{Qij}$  berechnet:

$$\bar{u}_{Qj} = \frac{1}{N} \sum_{i=1}^N u_{Qij}. \quad (7.1)$$

$i$  kennzeichnet dabei einen bestimmten Probanden. Ebenfalls von Interesse waren die Standardabweichungen der Bewertungen

$$S_{Qj} = \sqrt{\sum_{i=1}^N \frac{(\bar{u}_{Qj} - u_{Qij})^2}{N-1}} \quad (7.2)$$

und die 95%-Vertrauensintervalle

$$[\bar{u}_{Qj} - \delta_{Qj} ; \bar{u}_{Qj} + \delta_{Qj}], \text{ mit } \delta_{Qj} = t_{N-1;1-\alpha/2} \cdot \frac{S_{Qj}}{\sqrt{N}}. \quad (7.3)$$

Die  $t_{N-1;1-\alpha/2}$  sind Quantile der Studentschen t-Verteilung (siehe [Kle03]). Es muss  $\alpha = 0,05$  gewählt werden. Die 95%-Vertrauensintervalle geben jeweils Bereiche an, in denen die absoluten Differenzen zwischen den Stichprobenmittelwerten und den „wahren“ Mittelwerten für sehr große Probandenanzahlen mit einer Wahrscheinlichkeit von 95 % liegen.

Im Rahmen des eingesetzten Screening-Verfahrens, das in [ITU02] vorgestellt wird, wird nach einer Häufung von auffälligen Bewertungen einzelner Probanden gesucht, was auf Unzuverlässigkeit der Probanden hindeutet. Unter auffälligen Bewertungen sind solche zu verstehen, die deutlich von den Stichprobenmittelwerten abweichen. Zunächst muss mit Hilfe eines so genannten  $\beta_2$ -Tests für jedes Bildpaar geprüft werden, ob die abgegebenen Bewertungen einer Normalverteilung unterliegen oder nicht. Wenn keine Normalverteilung vorliegt, wird eine größere Abweichung zugelassen. Damit von einer Normalverteilung ausgegangen werden kann, muss  $\beta_{2j}$  im Intervall  $[2 \leq \beta_{2j} \leq 4]$  liegen.  $\beta_{2j}$  wird folgendermaßen berechnet:

$$\beta_{2j} = \frac{m_4}{(m_2)^2}, \text{ mit } m_x = \frac{\sum_{i=1}^N (u_{Qij} - \bar{u}_{Qj})^x}{N}. \quad (7.4)$$

Für jeden Probanden  $i$  muss nun gezählt werden, wie oft seine Bewertungen nicht in folgenden Intervallen liegen:

$$\begin{aligned} [\bar{u}_{Q_j} - 2S_{Q_j} \leq u_{Q_{ij}} \leq \bar{u}_{Q_j} + 2S_{Q_j}] & \text{ bei Normalverteilung,} \\ [\bar{u}_{Q_j} - \sqrt{20}S_{Q_j} \leq u_{Q_{ij}} \leq \bar{u}_{Q_j} + \sqrt{20}S_{Q_j}] & \text{ bei anderer Verteilung.} \end{aligned} \quad (7.5)$$

Bei Abweichungen nach oben wird ein Zähler  $P_i$  inkrementiert, bei Abweichungen nach unten entsprechend ein Zähler  $Q_i$ . Laut [ITU02] ist ein Proband dann als unzuverlässig zu betrachten, wenn mit  $M$  gleich der Anzahl der Bildpaare gilt:

$$\frac{P_i + Q_i}{M} > 0,05 \quad \text{und} \quad \left| \frac{P_i - Q_i}{P_i + Q_i} \right| < 0,3. \quad (7.6)$$

Im ersten subjektiven Test wurde nur ein Proband durch den Screening-Algorithmus als unzuverlässig eingestuft. Es kann davon ausgegangen werden, dass der vorherige Plausibilitätstest bereits für eine Identifizierung und Aussonderung der unzuverlässigsten Probanden sorgte. Insgesamt blieben 16 Fragebögen zur Auswertung übrig. Nach Aussonderung unzuverlässiger Probanden wurden für die weiteren Auswertungen die Stichprobenmittelwerte, Standardabweichungen und Vertrauensintervalle neu berechnet. Erstaunlicherweise wurden im zweiten und dritten subjektiven Test keine Probanden durch den Screening-Algorithmus als unzuverlässig eingestuft, obwohl keine vorherigen Plausibilitätskontrollen durchgeführt wurden und das Screening im zweiten Test auf beide abgefragten Qualitätsbewertungen angewendet wurde. Dies kann damit erklärt werden, dass am zweiten und dritten Test mehrheitlich Probanden teilnahmen, die bereits am ersten Test teilgenommen hatten und mit der Testumgebung daher erheblich besser zurechtkamen. Es erscheint daher für zukünftige Untersuchungen empfehlenswert, vor der Durchführung eines subjektiven Tests einen Vortest mit denselben Probanden durchzuführen, der ausschließlich der Gewöhnung der Teilnehmer an die Testumgebung dient. Im zweiten subjektiven Test konnten alle 16 Bögen, im dritten Test alle 18 Bögen weiterverwendet werden.

## 7.6.2 Datenanalyse

Nun sollen die abgegebenen Qualitätsbewertungen für alle Standbildpaare betrachtet werden. In Abbildung 7.3 sind die DMOS-Werte (Differential Mean Opinion Score, hier den Stichprobenmittelwerten entsprechend) aus dem ersten und zweiten subjektiven Test gemeinsam in ein Diagramm eingetragen. Ebenso sind die entsprechenden 95%-Vertrauensintervalle angegeben. Zunächst wird auf einige allgemeine Tendenzen eingegangen.

Es fällt die Bewertung der Bildpaare 11 und 21 auf, die bei allen Bewertungskriterien nahe Null liegt. Es handelt sich dabei um Bildpaare, die aus identischen Bildern bestanden, was von den Probanden größtenteils richtig erkannt wurde. Ein weiterer wichtiger Punkt

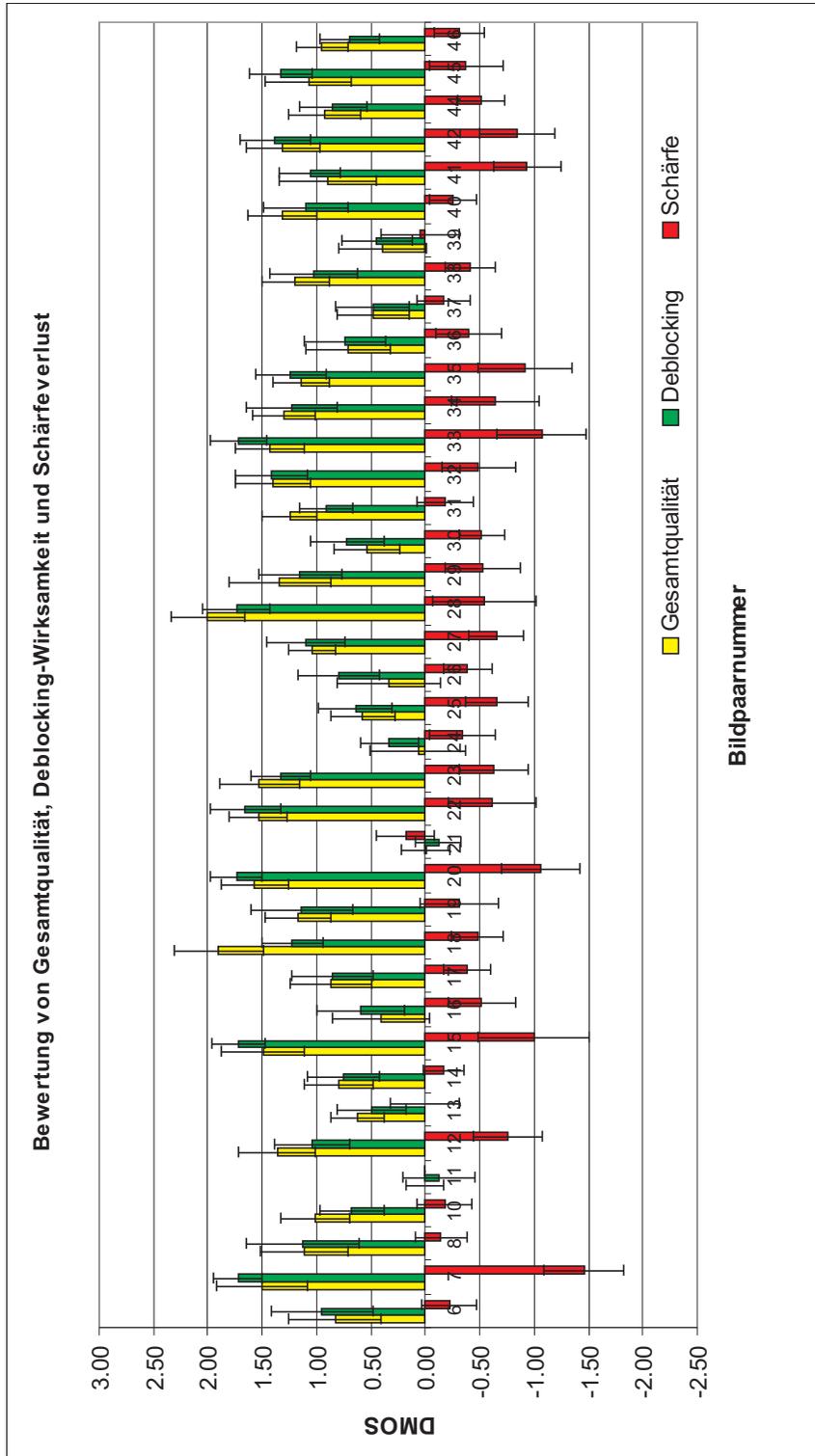


Abbildung 7.3: Ergebnisse der subjektiven Tests 1 und 2, Reihenfolge siehe Tabelle A.2

ist, dass die Veränderung der Gesamtqualität und die Veränderung der Sichtbarkeit speziell von Blockartefakten bei den einzelnen Bildpaaren sehr ähnlich beurteilt wurde. In den meisten Fällen wird die Verringerung von Blockartefakten etwas höher bewertet als die Verbesserung der Gesamtqualität. Es gibt jedoch auch einige Bildpaare, bei denen der umgekehrte Fall eintritt. Grundsätzlich ist zu beachten, dass hier (zufällige) Stichprobenmittelwerte miteinander verglichen werden, die sich nicht exakt reproduzieren lassen. Eine Wiederholung des Tests würde daher eine etwas andere Lage der einzelnen Bewertungen ergeben. Wichtig ist, dass in Abbildung 7.3 in nahezu allen Fällen eine deutliche Überschneidung der Vertrauensintervalle der Bewertungen, die die Gesamtqualität und die Reduktion von Blockartefakten betreffen, zu erkennen ist. Deshalb kann die Tendenz, dass die Probanden diese beiden Kriterien sehr ähnlich bewerten, als gesichert gelten. Auffällig ist die Bewertung von Bildpaar 18, da hier die Verbesserung der Gesamtqualität sehr viel höher eingestuft wird als die Qualitätsverbesserung speziell bezogen auf die Verringerung von Blockartefakten. Hier wirkt sich offenbar die starke Glättungswirkung des eingesetzten Deblocking-Verfahrens (*kim*) besonders positiv auf die Qualitätsverbesserung des Bildes *stockholm4* aus. Eine ähnliche Tendenz gibt es in Bildpaar 28 (*lena2\_zoom2*), wo ebenfalls das Verfahren *kim* eingesetzt wurde. Obwohl die Bildpaare 18 und 28 Auffälligkeiten aufweisen, besteht kein Grund, sie als Ausreißer zu betrachten und aus den weiteren Untersuchungen auszuschließen. Wird Abbildung 7.3 nochmals insgesamt betrachtet, kann eine Ausnutzung der zur Verfügung stehenden Skala zu ca. zwei Dritteln abgelesen werden. Es zeigt sich, dass die eingesetzten Artefaktreduktionsverfahren zum Teil eine deutliche Verbesserung der Bildqualität von gestörten Bildern bewirken können. Dass keine Bewertung im oberen Drittel der Skala liegt, ist nicht erstaunlich, da Postprocessing-Verfahren zur Verringerung von Blockartefakten letztlich nur Störungen kaschieren, nicht jedoch die Qualität eines ungestörten Bildes wiederherstellen können. Dennoch haben einzelne Probanden auch Bewertungen im oberen Drittel der Skala abgegeben.

Interessant sind auch die Bewertungen, die sich auf den Vergleich der Schärfe von Bildpaaren beziehen. Es zeigt sich, dass eine Reduktion von Blockartefakten durch die eingesetzten Verfahren grundsätzlich mit der Entstehung von Unschärfe einhergeht. Bildpaar 39 stellt die einzige Ausnahme dar, jedoch ist die durchschnittliche Bewertung des Schärfeunterschieds so klein, dass davon auszugehen ist, dass die Probanden keinen Schärfeunterschied feststellen konnten. Insgesamt ist die Veränderung der Schärfe durch die untersuchten Verfahren weit geringer als die Veränderung der anderen Kriterien und mit im Schnitt weniger als 0,5 Skaleneinheiten auch absolut betrachtet sehr klein. Es gibt einige wenige Bildpaare, in denen ein Unterschied der Schärfe im Bereich zwischen einer und 1,5 Skaleneinheiten angezeigt wird. Deutliche Auswirkungen auf die Bewertung des Gesamtqualitätsunterschieds sind aber auch in diesen Fällen nicht nachweisbar, d. h. die Gesamtqualität eines Bildes wird beim Auftreten von Unschärfe nicht wesentlich schlechter beurteilt als die Qualität speziell bezogen auf die Sichtbarkeit von Blockartefakten. Zusammenfassend kann gesagt werden, dass eine Verbesserung der Gesamtbildqualität in

den durchgeführten subjektiven Tests maßgeblich durch die Fähigkeit der eingesetzten Verfahren, Blockartefakte zu verringern, bestimmt wird und sich die gleichzeitige Entstehung von Unschärfe in der Regel nur begrenzt qualitätsmindernd auf die Eindrücke der Probanden auswirkt. Wie bereits erwähnt, ließen die erweiterten Fragebögen des ersten subjektiven Tests diesen Schluss nicht zu, was jedoch frühzeitig bezweifelt wurde.

Im Folgenden soll die Wirksamkeit der im Rahmen der subjektiven Tests eingesetzten Verfahren zur Reduktion von Codierartefakten verglichen und bewertet werden. Es sei bereits hier auf Abbildung 7.4 hingewiesen, die einige Daten aus den folgenden Tabellen grafisch darstellt. Zunächst ist in Tabelle 7.2 angegeben, wie die Gesamtqualitätsunterschiede zwischen den Bildern der gezeigten Bildpaare von den Probanden eingeschätzt wurden. Eine genaue Aufschlüsselung der Bildnamen findet sich in Anhang A. Die Tabelle ist so sortiert, so dass die größten Qualitätsverbesserungen am unteren Tabellenende zu finden sind. Der einzige in Klammern dargestellte Eintrag stammt aus der Trainingsphase des ersten subjektiven Tests und ist daher mit Vorsicht zu interpretieren. Da das entsprechende Bildpaar versehentlich im übrigen Testverlauf nicht erneut gezeigt wurde, erschien eine Aufnahme in die Tabelle aus Gründen der Vollständigkeit sinnvoll. In Anhang B befinden sich dieselben Testergebnisse nochmals getrennt nach Datenrate der Testbilder.

Auffällig sind die ersten beiden Tabellenzeilen, in denen keine Qualitätsverbesserung angezeigt wird. Da es sich hier um die beiden Bildpaare aus identischen Bildern handelt, ist dies vollkommen plausibel. Weiterhin fällt auf, dass bei niedriger Datenrate (3,5 Mbit/s), absolut betrachtet, stärkere Verbesserungen der Gesamtbildqualität erzielt werden als bei höherer Datenrate (5 Mbit/s). Auch das ist plausibel, da bei weniger vorhandenen Störungen auch weniger Verbesserungen durch die Artefaktreduktionsverfahren erzielt werden können. Die wirksamsten Verfahren sind die mit mehreren Filtermodi. Sie finden sich ausnahmslos in der Spitzengruppe, also am unteren Tabellenende, wieder, während die statische Filterung von Blockgrenzen (*avril*) und die Subtraktion von Fehlervektoren (*yang*) insgesamt als wenig wirksam einzuschätzen sind. Diese Aussage gilt mit Ausnahme von *wheel* unabhängig vom betrachteten Bild(-typ). Das Verfahren aus dem MPEG-4-Standard (*kim*) und die in dieser Arbeit vorgestellten Deblocking-Verfahren (*pi7*, *pi4b*) bieten die höchste Wirksamkeit. Die komplexere Variante *pi7* (drei Filtermodi, globale Bildanalyse) erzielt etwas größere Verbesserungen der Gesamtbildqualität als die einfachere Variante. Das auf geringen Rechenaufwand optimierte adaptive Verfahren *ramkishor* ist insgesamt weniger wirksam als andere adaptive Verfahren mit mehreren Filtermodi. Das waveletbasierte Verfahren *kim98* erzielt zum Teil deutliche, zum Teil aber auch nur mittelmäßige Qualitätsverbesserungen und verhält sich insofern etwas uneinheitlich. Bei niedrigen und hohen Datenraten zeigen die Deblocking-Verfahren untereinander verglichen jeweils ein ähnliches Verhalten. Das in dieser Arbeit vorgestellte Verfahren mit lokaler und globaler Bildanalyse *pi7* schneidet bei 5 Mbit/s jedoch sowohl im Bild *football* als auch im Bild *wheel* mit deutlichem Abstand am besten ab, was darauf hindeutet, dass durch die globale Bildanalyse eine besonders gute Anpassung der Filterung an die Bildqualität bzw.

**Tabelle 7.2:** Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Gesamtqualitätsunterschiede

Nr.	Bild	Qualität
21	wheel5m_L	-0.01
11	football3.5m_L	0.01
24	stockholm4_3.5m_avril_R	0.07
26	football5m_yang_R	0.34
39	wheel5m_ramkishor_L	0.39
16	wheel5m_kim98_R	0.41
37	football5m_avril_R	0.48
30	wheel5m_yang_R	0.53
25	wheel5m_avril_R	0.58
13	wheel5m_kim_L	0.63
36	wheel5m_pi7_R	0.71
(2)	(wheel3.5m_kim_R)	(0.77)
14	lena2_zoom2_3.5m_avril_L	0.79
6	football5m_ramkishor_L	0.83
17	football3.5m_avril_L	0.87
41	football5m_kim98_R	0.89
44	wheel3.5m_ramkishor_R	0.93
46	lena2_zoom2_3.5m_yang_L	0.95
10	stockholm4_3.5m_yang_L	1.02
27	wheel3.5m_yang_R	1.04
45	football5m_kim_L	1.08
8	wheel3.5m_avril_L	1.11
35	football3.5m_yang_L	1.14
19	lena2_zoom2_3.5m_kim98_R	1.17
38	stockholm4_3.5m_ramkishor_R	1.19
31	lena2_zoom2_3.5m_pi7_L	1.25
34	stockholm4_3.5m_pi7_R	1.30
42	stockholm4_3.5m_kim98_L	1.31
40	lena2_zoom2_3.5m_ramkishor_R	1.32
29	football5m_pi7_L	1.34
12	wheel3.5m_kim98_R	1.36
32	wheel3.5m_pi4b_L	1.40
33	football3.5m_ramkishor_R	1.43
15	football3.5m_pi4b_R	1.49
7	football3.5m_kim98_R	1.51
23	wheel3.5m_pi7_L	1.53
22	football3.5m_kim_R	1.54
20	football3.5m_pi7_L	1.57
18	stockholm4_3.5m_kim_L	1.90
28	lena2_zoom2_3.5m_kim_L	2.00

Datenrate gelingt. Auffällig sind noch die beiden unteren Tabellenzeilen. Die subjektiv empfundenen Qualitätsverbesserungen sind hier besonders hoch. In beiden Fällen kam das Verfahren *kim* zum Einsatz, welches sich durch eine besonders starke tiefpassartige Filterung in homogenen Bereichen auszeichnet. Wie bereits im Zuge der Analyse von Abbildung 7.3 angedeutet, wirkt sich diese Filterung durch die starke Glättung bestimmter Störkomponenten in den beiden Testbildern offenbar besonders günstig und somit qualitätssteigernd aus. Die Wirksamkeit des Verfahrens speziell in Bezug auf die Verringerung von Blockartefakten wurde von den Probanden geringer eingeschätzt (vergleiche Abbildung 7.3).

Tabelle 7.3 zeigt, wie die Teilnehmer des zweiten subjektiven Tests Bildqualitätsunterschiede in den paarweise gezeigten Bildern speziell bezogen auf Blockartefakte empfanden. Es sei darauf hingewiesen, dass die Nummerierung der Bildpaare an die des ersten subjektiven Tests angepasst wurde und nicht der tatsächlichen Nummerierung der Bildpaare im zweiten Test entspricht. Die Tabelle zeigt, dass auch hier offensichtlich die identischen Bildpaare von den meisten Probanden erkannt wurden. Sie finden sich in den oberen Tabellenzeilen wieder. Genauso wie bei der Betrachtung der Gesamtbildqualität erzielen sämtliche Verfahren bei niedriger Datenrate durchschnittlich größere Gewinne als bei höherer Datenrate. Eine ausgeprägte Abhängigkeit der Qualitätsverbesserung vom Bild(-typ) ist nicht erkennbar. Grundsätzlich schneiden die adaptiven Verfahren mit mehreren Filtermodi besser als die anderen Verfahren (*avril*, *yang*) ab, wenn nur die reine Reduktion von Blockartefakten betrachtet wird. Diesbezüglich gibt es zwei potenzielle Ausreißer (Nr. 8 und Nr. 35), die aber keine Systematik erkennbar werden lassen. Insgesamt zeigen die abgegebenen Bewertungen ähnliche Tendenzen wie die Bewertungen der Verbesserung der Gesamtqualität. Jedoch sind sie weniger klar ausgeprägt und wirken im Vergleich etwas durchmischer. Es fällt auf, dass sich die komplexere Variante des in dieser Arbeit vorgestellten Artefaktreduktionsverfahrens *pi7* in Bezug auf die Wirksamkeit bei der Verringerung von Blockartefakten nicht von der einfacheren Variante *pi4* absetzen kann.

Tabelle 7.4 zeigt die durch die Teilnehmer des zweiten subjektiven Tests abgegebenen Bewertungen bzgl. der wahrgenommenen Schärfeunterschiede in den paarweise gezeigten Bildern. Zunächst kann festgehalten werden, dass die beiden Bildpaare aus identischen Bildern (Nr. 11 und Nr. 21) ohne Zweifel erkannt wurden. Die Bewertungen für die übrigen Bildpaare zeigen, dass sämtliche Verfahren zur Reduktion von Blockartefakten die Bildschärfe verringern, was im Hinblick auf die innerhalb der Verfahren eingesetzten Filterungen den Erwartungen entspricht. Allerdings kann in vielen Fällen (ca. das obere Drittel der Tabelle) nicht von einem signifikanten Schärfeunterschied zwischen den Bildern eines Paares gesprochen werden. Insgesamt spielen sich die Verminderungen der Bildschärfe durch den Einsatz von Artefaktreduktionsverfahren fast ausnahmslos im Bereich von nur einer Skaleneinheit ab und sind somit gering. Das soll aber nicht darüber hinwegtäuschen, dass Verminderungen der Schärfe existieren. Auffällig ist eine Konzentration von einzelnen Bildern bzw. Bildtypen in bestimmten Abschnitten der Tabelle. Dies deutet darauf

**Tabelle 7.3:** Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Qualitätsunterschiede in Bezug auf Blockartefakte

Nr.	Bild	Qualität
11	football3.5m_L	-0.13
21	wheel5m_L	-0.12
24	stockholm4_3.5m_avril_R	0.33
39	wheel5m_ramkishor_L	0.45
37	football5m_avril_R	0.49
13	wheel5m_kim_L	0.50
16	wheel5m_kim98_R	0.59
25	wheel5m_avril_R	0.64
10	stockholm4_3.5m_yang_L	0.68
46	lena2_zoom2_3.5m_yang_L	0.70
30	wheel5m_yang_R	0.72
36	wheel5m_pi7_R	0.74
14	lena2_zoom2_3.5m_avril_L	0.76
26	football5m_yang_R	0.80
44	wheel3.5m_ramkishor_R	0.85
17	football3.5m_avril_L	0.86
31	lena2_zoom2_3.5m_pi7_L	0.91
6	football5m_ramkishor_L	0.95
(2)	wheel3.5m_kim_R	0.97
38	stockholm4_3.5m_ramkishor_R	1.03
12	wheel3.5m_kim98_R	1.04
41	football5m_kim98_R	1.06
27	wheel3.5m_yang_R	1.09
40	lena2_zoom2_3.5m_ramkishor_R	1.10
8	wheel3.5m_avril_L	1.13
19	lena2_zoom2_3.5m_kim98_R	1.14
29	football5m_pi7_L	1.16
18	stockholm4_3.5m_kim_L	1.23
34	stockholm4_3.5m_pi7_R	1.23
35	football3.5m_yang_L	1.24
23	wheel3.5m_pi7_L	1.33
45	football5m_kim_L	1.33
42	stockholm4_3.5m_kim98_L	1.38
32	wheel3.5m_pi4b_L	1.41
22	football3.5m_kim_R	1.66
15	football3.5m_pi4b_R	1.72
33	football3.5m_ramkishor_R	1.72
7	football3.5m_kim98_R	1.73
20	football3.5m_pi7_L	1.74
28	lena2_zoom2_3.5m_kim_L	1.74

**Tabelle 7.4:** Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Schärfeunterschiede

Nr.	Bild	Qualität
21	wheel5m_L	0.18
39	wheel5m_ramkishor_L	0.05
13	wheel5m_kim_L	0.01
11	football3.5m_L	0.00
8	wheel3.5m_avril_L	-0.14
14	lena2_zoom2_3.5m_avril_L	-0.17
37	football5m_avril_R	-0.17
10	stockholm4_3.5m_yang_L	-0.18
31	lena2_zoom2_3.5m_pi7_L	-0.18
6	football5m_ramkishor_L	-0.22
40	lena2_zoom2_3.5m_ramkishor_R	-0.26
19	lena2_zoom2_3.5m_kim98_R	-0.31
46	lena2_zoom2_3.5m_yang_L	-0.31
24	stockholm4_3.5m_avril_R	-0.34
45	football5m_kim_L	-0.38
17	football3.5m_avril_L	-0.39
26	football5m_yang_R	-0.39
36	wheel5m_pi7_R	-0.39
38	stockholm4_3.5m_ramkishor_R	-0.41
(2)	wheel3.5m_kim_R	-0.43
18	stockholm4_3.5m_kim_L	-0.48
32	wheel3.5m_pi4b_L	-0.49
44	wheel3.5m_ramkishor_R	-0.51
16	wheel5m_kim98_R	-0.52
30	wheel5m_yang_R	-0.52
29	football5m_pi7_L	-0.53
28	lena2_zoom2_3.5m_kim_L	-0.54
22	football3.5m_kim_R	-0.61
23	wheel3.5m_pi7_L	-0.63
34	stockholm4_3.5m_pi7_R	-0.64
27	wheel3.5m_yang_R	-0.65
25	wheel5m_avril_R	-0.66
12	wheel3.5m_kim98_R	-0.76
42	stockholm4_3.5m_kim98_L	-0.84
35	football3.5m_yang_L	-0.91
41	football5m_kim98_R	-0.94
15	football3.5m_pi4b_R	-1.00
20	football3.5m_pi7_L	-1.06
33	football3.5m_ramkishor_R	-1.07
7	football3.5m_kim98_R	-1.46

hin, dass der Bildinhalt für die Wahrnehmung von Schärfeveränderungen offenbar bedeutender ist als das eingesetzte Deblocking-Verfahren. Bezüglich der Abhängigkeit der Schärfeveränderungen von der Datenrate lässt sich keine belastbare Aussage machen.

Bei Betrachtung der einzelnen Deblocking-Verfahren fällt auf, dass die statische Filterung von Blockgrenzen (*avril*) mit ihren sehr kurzen Filtern nur zu geringer Unschärfbildung führt. Einzige Ausnahme ist Bildpaar Nr. 25, wo sich die fehlende Adaptivität offenbar negativ auswirkt. Das wavelet-basierte Deblocking-Verfahren *kim98* neigt in dem durchgeführten Test am meisten zu Unschärfe-Bildung. Die übrigen Verfahren verhalten sich diesbezüglich sehr uneinheitlich. Für die in dieser Arbeit vorgestellten Verfahren *pi4b* und *pi7* ist grundsätzlich eine Tendenz zu leichter Unschärfbildung erkennbar, es gibt jedoch auch Ausnahmen (Nr. 31, Nr. 36). Bei der Variante mit breiterem Filter für homogene Bereiche (*pi7*) wird minimal mehr Unschärfe von den Probanden empfunden als bei der anderen Variante.

Ergänzend zu den bereits vorgestellten Tabellen mit Ergebnissen der subjektiven Tests sind in Abbildung 7.4 einige der Ergebnisse in Form von Diagrammen veranschaulicht. In den einzelnen Diagrammen sind die abgegebenen Bewertungen jeweils für die verschiedenen Deblocking-Verfahren zusammengefasst, wobei nur Bilder, die mit 3,5 Mbit/s codiert wurden, berücksichtigt sind. Die insgesamt geringe Wirksamkeit von *avril* beim Deblocking und bei der Verbesserung der Gesamtbildqualität kann ebenso deutlich abgelesen werden, wie die nur sehr geringe Tendenz des Verfahrens, Unschärfe zu erzeugen. Die Wirksamkeit von *yang* wurde etwas höher als die von *avril* bewertet, allerdings ist auch die Tendenz zur Erzeugung von Unschärfe höher. Dies verwundert nicht, da im Gegensatz zu *avril* nicht nur die direkten Blockgrenzpixel bearbeitet werden, sondern auch Pixel, die weiter im Blockinnern liegen. Die bezüglich Blocking-Reduktion und Verbesserung der Gesamtbildqualität wirksamsten Verfahren sind offensichtlich *kim*, *pi7* und *kim98*. Das Verfahren *ramkishor* schneidet etwas schlechter ab. Die bereits erwähnte besonders hohe Wirksamkeit von *kim* im Falle der Bilder *lena2\_zoom2* und *stockholm4* ist schließlich in Abbildung 7.4 ebenso erkennbar wie die unter allen untersuchten Verfahren größte Tendenz zu Unschärfbildung von *kim98*.

Einige der in den subjektiven Tests abgegebenen Bewertungen sollen nun anhand ausgewählter Bildausschnitte aus dem Testmaterial veranschaulicht werden. In Abbildung 7.5 kann erkannt werden, dass die statische Filterung von Blockgrenzen mit kurzen Filtern (*avril*) Blockartefakte sichtbar reduzieren und die Bildqualität verbessern kann, wenn viele Kanten und wenige fein texturierte Bereiche im Bildmaterial vorhanden sind. Allerdings ist auch erkennbar, dass im oberen Bereich des Bildausschnitts Reste von Blockartefakten und im Bereich der Sektorscheibe leichte Staircase-Artefakte sichtbar bleiben. Stellenweise wäre der Einsatz von längeren Filtern hier vorteilhaft. Solche Filter werden von *pi7* adaptiv angewendet und so ist in Abbildung 7.5 erkennbar, dass die Wirksamkeit von *pi7* sowohl in Bezug auf die Wirksamkeit beim Deblocking als auch in Bezug auf die Wirksamkeit bei der Verbesserung der Gesamtbildqualität höher ist. Die beschriebenen Eindrücke

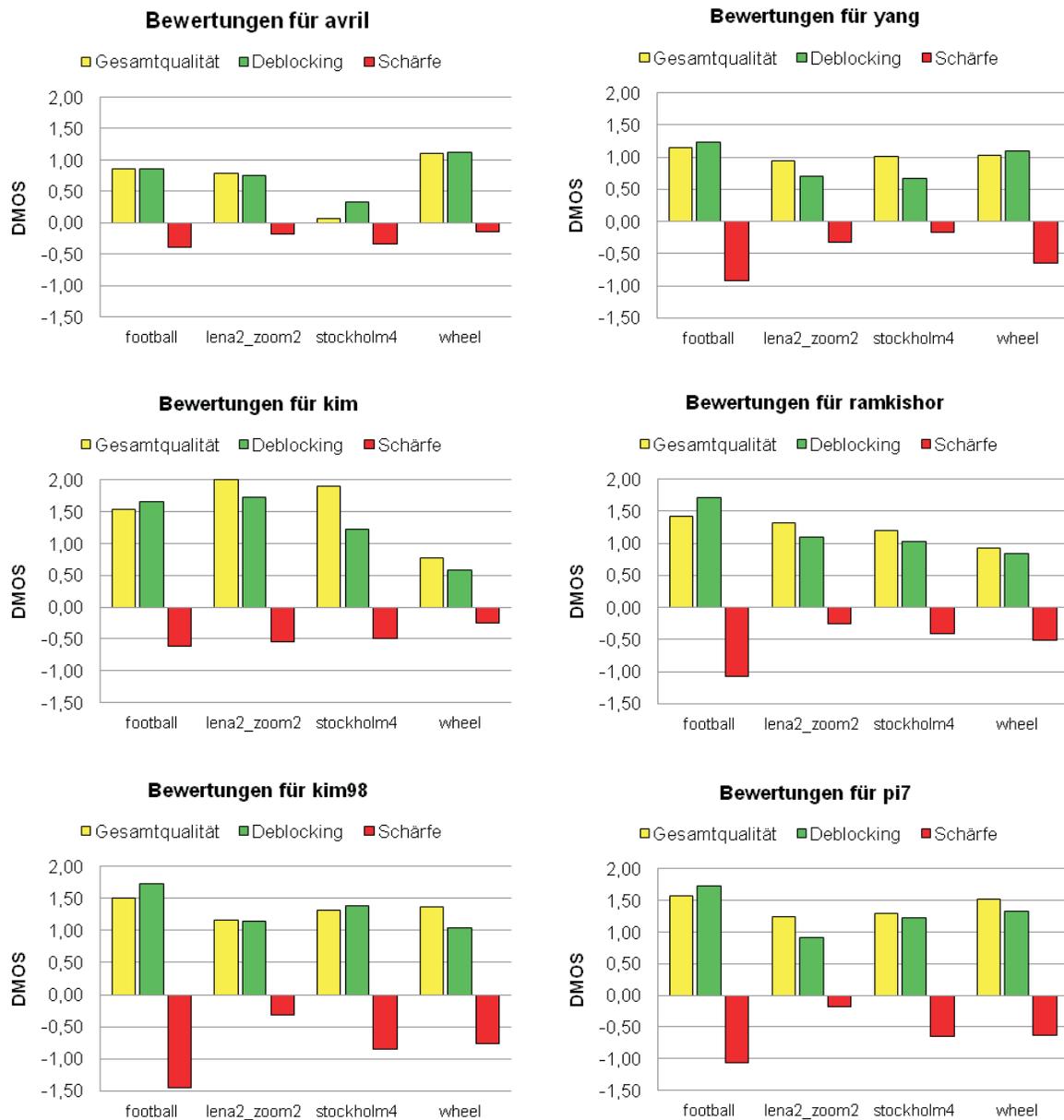
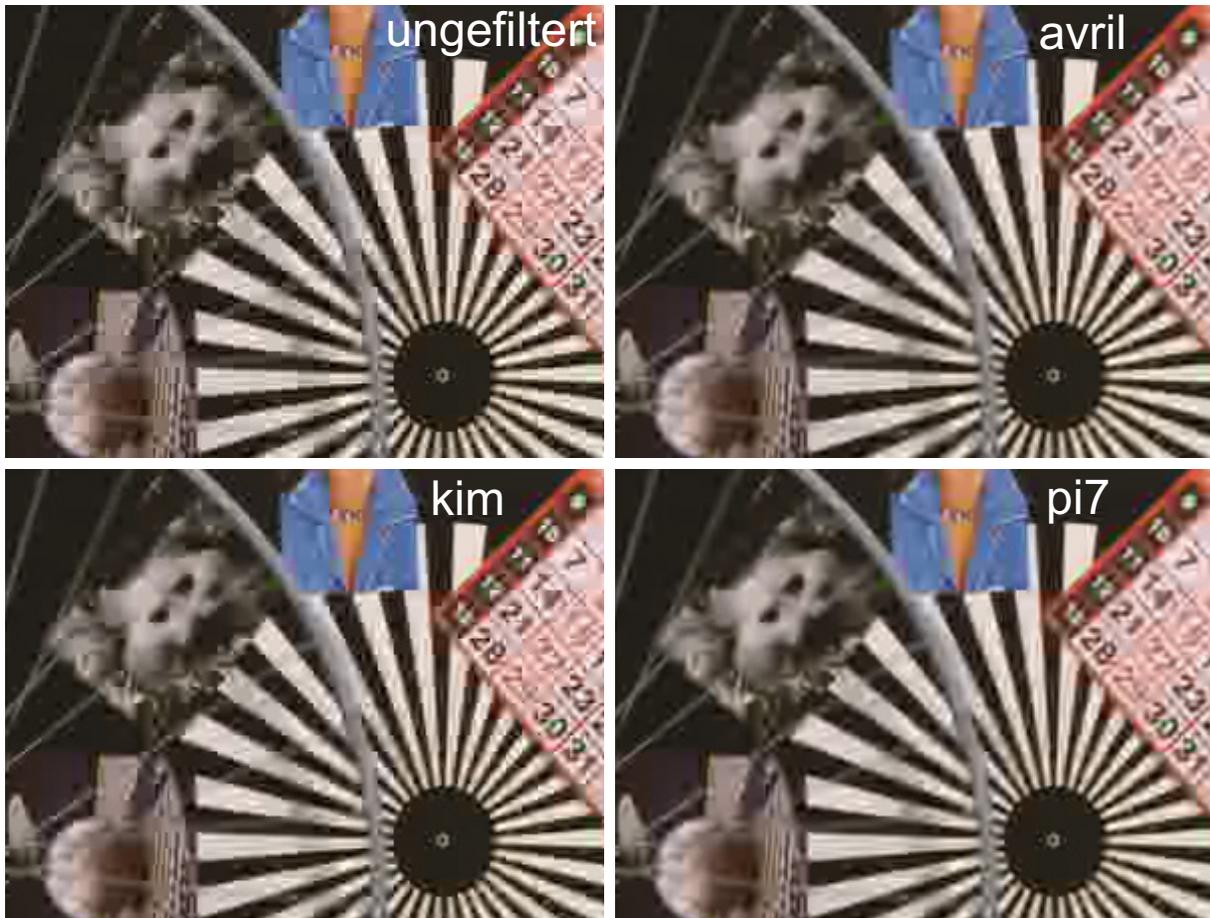


Abbildung 7.4: Bewertungen der einzelnen Deblocking-Verfahren bei 3,5 Mbit/s

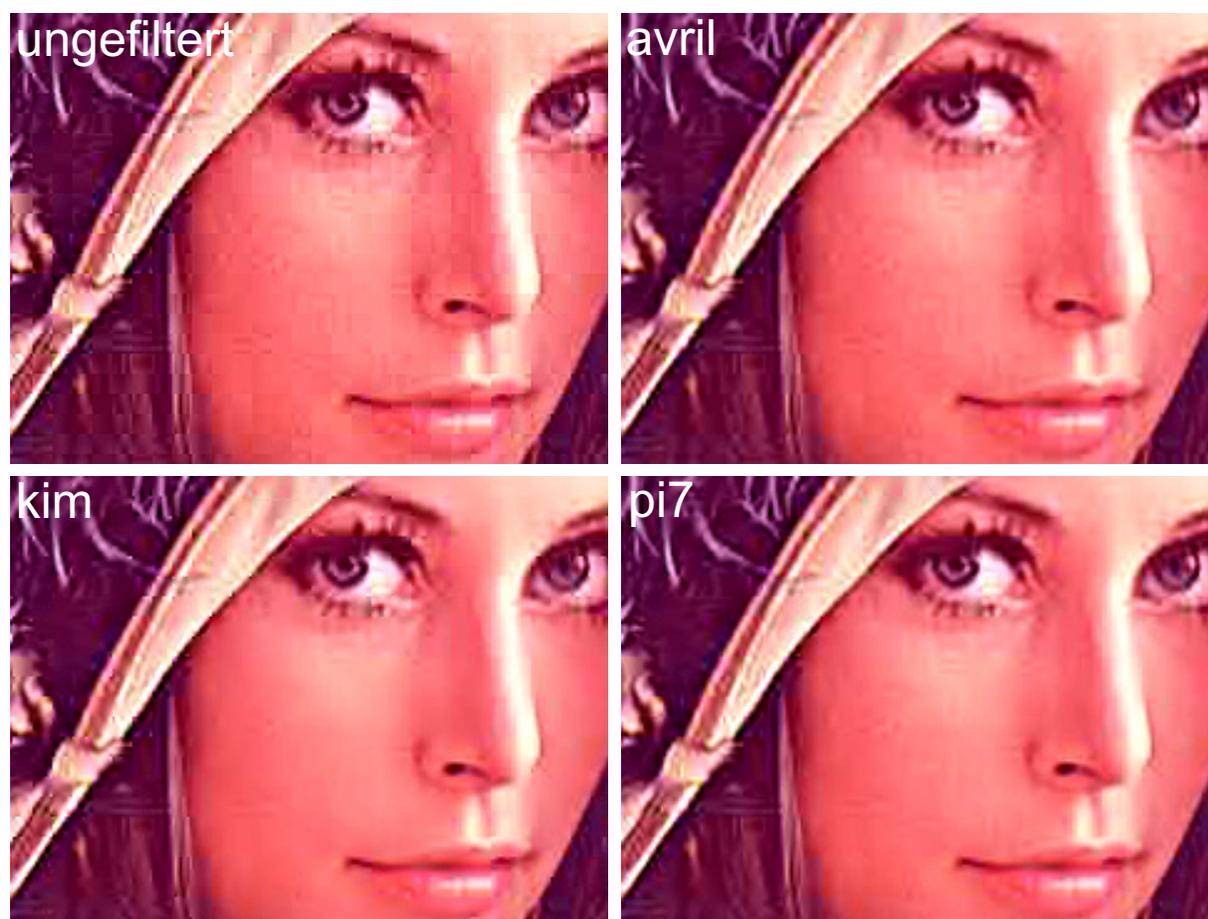
decken sich mit den Ergebnissen der subjektiven Tests, in denen *pi7* entsprechend besser bewertet wurde als *avril*. Recht schlecht bewertet wurde im Testbild *wheel* das Verfahren *kim*. Auch dies kann in Abbildung 7.5 nachvollzogen werden, da erkennbar ist, dass zahlreiche Artefakte nicht erkannt und somit nicht gefiltert wurden.



**Abbildung 7.5:** Bildausschnitte aus Frame 24 der Sequenz *wheel*, bearbeitet mit verschiedenen Deblocking-Verfahren

In Abbildung 7.6 kann insbesondere nachvollzogen werden, warum die Probanden der subjektiven Tests das Verfahren *kim* im Falle des Bildes *lena2\_zoom2* besonders gut bewertet haben. Zwar bleiben vereinzelt Artefakte ungefiltert, auffällig ist aber das wirksame Deblocking und die erhebliche Verbesserung der Bildqualität im Bereich des Gesichts. Hier wirkt sich die breite tiefpassartige Filterung von *kim* in homogenen Bereichen und die daraus resultierende Glättung besonders günstig aus. Die glättende Wirkung des Filters für homogene Bereiche im Verfahren *pi7* ist speziell um die Mitte des Bildausschnitts herum

geringer, was sich entsprechend negativ in den abgegebenen Bewertungen niederschlägt. Das kurze Filter im Verfahren *avril* ist am wenigsten dazu in der Lage, die Störungen im Gesicht hinreichend zu verringern, was sich ebenfalls mit den abgegebenen Bewertungen deckt.



**Abbildung 7.6:** Bildausschnitte aus Frame 48 der Sequenz *lena2\_zoom2*, bearbeitet mit verschiedenen Deblocking-Verfahren

Schließlich ist in Abbildung 7.7 sichtbar, warum im Falle des Bildes *stockholm4* das Deblocking-Verfahren *avril* als besonders wenig wirksam bewertet wurde. Das Verfahren ist an vielen Stellen im Bildausschnitt nicht dazu in der Lage, Blockartefakte deutlich zu verringern. Ansätze der Artefakte bleiben sichtbar und im Bereich des Wassers ist nur eine sehr geringe Wirkung erkennbar. Es sei allerdings erwähnt, dass bei diesem Testbild einige (ungewollte) Vertauschungen der durch die Probanden abgegebenen Bewertungen registriert wurden, welche den Mittelwert der Bewertungen verringerten (siehe auch Kapitel 8). Es kann daher davon ausgegangen werden, dass die tatsächliche Wirksamkeit

des Verfahrens etwas (aber nicht wesentlich) größer ist als die abgedruckten Ergebnisse des subjektiven Tests suggerieren. Das Verfahren *kim* ist aber ohne Zweifel im Vergleich zum Verfahren *avril* erheblich wirkungsvoller und verringert die vorhandenen Blockartefakte stark, was entsprechend in einer auch von den Probanden der subjektiven Tests wahrgenommenen verbesserten Bildqualität resultiert. Für weitere Bildbeispiele sei auf die Kapitel 5 und 6 verwiesen.



**Abbildung 7.7:** Bildausschnitte aus Frame 12 der Sequenz *stockholm4*, bearbeitet mit verschiedenen Deblocking-Verfahren

Wie bereits erwähnt, wurden im ersten subjektiven Test auch einige Bildsequenzpaare gezeigt, um mögliche Probleme einzelner Algorithmen, die nur in der Bewegung sichtbar sind, aufzudecken und um erste Erkenntnisse darüber zu gewinnen, inwieweit eine Übertragung der Wirksamkeitsbewertung der untersuchten Verfahren von Standbildern auf Sequenzen möglich oder problematisch ist. Die von den Betrachtern abgegebenen Bewertungen der wahrgenommenen Unterschiede der Gesamtqualitäten zwischen den Sequenzen eines jeden gezeigten Paares sind in Tabelle 7.5 abgedruckt. Ebenfalls in die Tabelle aufgenommen wurden die Bewertungen der Qualitätsunterschiede zwischen den Bildern der entsprechenden Einzelbildpaare. Im Fall von *football* wurden von den Probanden deutliche Qualitätsunterschiede wahrgenommen und es kann eine fast optimale Übereinstimmung der abgegebenen Bewertungen für Sequenzpaare und für Einzelbildpaare bei 3,5 Mbit/s festgestellt werden. Eine gute Übereinstimmung der Bewertungen für Sequenzpaare und für Einzelbildpaare bei 5 Mbit/s ist daher zwangsläufig nicht gegeben. Anders sind die Verhältnisse im Fall von *wheel*, obwohl dieselben Datenraten verwendet wurden wie

im Fall von *football*. Es fällt auf, dass die Probanden keine großen Qualitätsunterschiede zwischen den Sequenzen der gezeigten Sequenzpaare wahrgenommen haben und dass die Übereinstimmung der Bewertungen für Sequenzpaare und für Standbildpaare insgesamt vergleichsweise schlecht ist. Die bessere Übereinstimmung ist hier zwischen den Bewertungen für Sequenzpaare und für Einzelbildpaare bei 5 Mbit/s gegeben. Ein Grund für die nur schwach wahrgenommenen Qualitätsunterschiede bei der Sequenz *wheel* könnte sein, dass durch die kaum vorhandenen Texturen und große unbewegte Bereiche eine bessere Komprimierbarkeit im Vergleich zu *football* gegeben ist und somit insgesamt vergleichsweise weniger Artefakte, an denen Bildqualitätsverbesserungen ausgemacht werden können, vorhanden sind. Entscheidender dürften aber durch die Videokompression erzeugte zeitliche Artefakte sein, die in der Sequenz *wheel* in Form von rausch- bzw. mosquitoartigen Störungen im Bereich des Kalenders und der Buchstabentafel deutlich sichtbar werden. Diese werden durch die eingesetzten Deblocking-Verfahren kaum verringert, was die Empfindungen der Probanden vermutlich stark geprägt haben dürfte. Des Weiteren sind Detailverbesserungen durch die Deblocking-Verfahren denkbar, die in Bildfolgen untergehen und nur in Einzelbildern sichtbar sind.

**Tabelle 7.5:** Vergleich der wahrgenommenen Qualitätsunterschiede in Einzelbildpaaren und in Sequenzpaaren

Sequenz	12m_intra	3.5m	5m
wheel12m_intra_pi7_R	0.46	1.53	0.71
wheel12m_intra_yang_L	0.50	1.04	0.53
wheel12m_intra_kim_R	0.57	0.77	0.63
wheel12m_intra_avril_L	0.57	1.11	0.58
football12m_intra_avril_R	0.84	0.87	0.48
football12m_intra_yang_L	1.19	1.14	0.34
football12m_intra_kim_R	1.50	1.54	1.08
football12m_intra_pi7_L	1.53	1.57	1.34

Letztlich sind die vorliegenden Daten zur Bewertung der Wirksamkeit von Deblocking-Verfahren in Bildsequenzen nicht ausreichend, um daraus belastbare Aussagen ableiten zu können, weshalb auf solche rein spekulativen Aussagen verzichtet wird. Da der Fokus der weiteren Untersuchungen in dieser Arbeit ohnehin auf örtlichen Störmerkmalen liegt, werden zeitliche Störmerkmale im Folgenden nicht weiter betrachtet. Die Einbeziehung zeitlicher Einflüsse in die Bewertung der Wirksamkeit und in die Realisierung von Deblocking-Verfahren stellt jedoch ohne Zweifel einen interessanten Aspekt für zukünftige weitere Untersuchungen dar.

Nun soll noch in verkürzter Form auf die Ergebnisse des dritten subjektiven Tests eingegangen werden. Dazu ist in Tabelle 7.6 angegeben, wie die Unterschiede zwischen den Gesamtqualitäten der Bilder jedes gezeigten Bildpaares von den Probanden eingeschätzt

wurden. Fälle, in denen besonders große Qualitätsunterschiede von den Probanden empfunden wurden, finden sich unten in der Tabelle.

**Tabelle 7.6:** Subjektives Ranking der Bilder bzw. Verfahren im dritten subjektiven Test, Bewertung der Gesamtqualitätsunterschiede

Nr.	Bild	Qualität
15	pedestrian3_1.8m_L	0.06
24	fries2_1.8m_L	0.14
30	pedestrian3_1.8m_3x3_L	0.46
29	shields3_3m_3x3_R	0.63
22	fries2_1.8m_3x3_L	0.67
8	shields3_3m_avril_R	0.71
18	pedestrian3_1.8m_avril_R	0.73
23	shields3_3m_yang_L	0.83
25	pedestrian3_1.8m_tai_L	0.92
9	pedestrian3_1.8m_yang_R	0.95
13	shields3_3m_ramkishor_L	1.00
20	shields3_3m_tai_L	1.02
26	shields3_3m_pi7_R	1.04
10	shields3_3m_kim98_L	1.07
7	pedestrian3_1.8m_ramkishor_R	1.08
6	fries2_1.8m_yang_R	1.15
28	pedestrian3_1.8m_kim_L	1.24
11	fries2_1.8m_tai_R	1.25
27	fries2_1.8m_avril_L	1.26
31	fries2_1.8m_ramkishor_L	1.39
12	pedestrian3_1.8m_kim98_L	1.42
21	pedestrian3_1.8m_pi7_R	1.44
16	shields3_3m_kim_R	1.46
14	fries2_1.8m_kim98_R	1.65
17	fries2_1.8m_pi7_L	1.68
19	fries2_1.8m_kim_L	1.72

Zunächst muss festgehalten werden, dass auch in diesem Testdurchlauf die Bildpaare aus identischen Bildern von den Probanden eindeutig erkannt wurden. Sie finden sich an den ersten beiden Positionen. Eine durchschnittlich besonders deutliche Artefaktreduktion gelingt offensichtlich im Bild *fries2*, eine nicht so deutliche Reduktion im Bild *shields3*. Im Bild *pedestrian3* sind die Ergebnisse sehr gemischt. Des Weiteren ist interessant, aber auch plausibel, dass eine einfache statische 3x3 Mittelwertfilterung nur eine geringe Verbesserung von mit Codierartefakten versehenen Bildern ermöglicht. Alle drei Fälle, in denen eine Mittelwertfilterung durchgeführt wurde, finden sich ganz oben in der Tabelle. Bezüglich der Wirksamkeit der anderen eingesetzten Deblocking-Verfahren sind insgesamt Tendenzen ablesbar, die größtenteils denen aus den vorherigen Tests entsprechen. Die größte

Bildqualitätsverbesserung ist in allen Bildern mit den adaptiven Verfahren *kim* und *pi7* sowie mit dem wavelet-basierten Verfahren *kim98* möglich. Die Verfahren *avril* und *yang* weisen - genauso wie in den ersten beiden Tests - eine recht schwache Wirksamkeit auf. Die übrigen Verfahren *ramkishor* und *tai* sind im Mittelfeld einzuordnen. Bemerkenswert ist, dass sich das Deblocking-Verfahren *tai* bezüglich seiner Wirksamkeit nicht von dem sehr ähnlichen aber einfacheren Verfahren *ramkishor* absetzen kann. Tendenziell ist sogar das Gegenteil der Fall.

# 8 Neuer Ansatz zur Bewertung von Deblocking-Verfahren

Die Vielzahl an verfügbaren Verfahren zur Reduktion von Codierartefakten (siehe Kapitel 5 und 6) wirft das Problem der Auswahl eines geeigneten Verfahrens für einen bestimmten Anwendungszweck auf. Von besonderer Bedeutung ist ohne Zweifel die Frage nach der Wirksamkeit einzelner Verfahren bei der Verbesserung der Bildqualität von gestörtem Bildmaterial. Denkbar ist die Suche nach einem insgesamt besonders wirksamen Verfahren, oder nach einem Verfahren, das in bestimmtem Bildmaterial eine möglichst deutliche Verbesserung der Gesamtbildqualität ermöglicht. Ebenso interessant können Aussagen sein, die darauf abzielen, abzuschätzen, wie groß der Verlust an Wirksamkeit hinsichtlich der Bildqualitätsverbesserung durch Einsatz eines einfachen Verfahrens im Vergleich zu einem Referenzverfahren ist. Auch im Rahmen der Weiterentwicklung, Optimierung oder Neuentwicklung von Artefaktreduktionsverfahren für spezielle Anwendungen - beispielsweise Video-Streaming auf mobile Endgeräte - ist es von großer Bedeutung, bewerten zu können, wie stark die Verfahren die Bildqualität von gestörten Bildern verbessern können.

Grundsätzlich lassen sich solche Bewertungen mit Hilfe von subjektiven Tests (z. B. nach [ITU02]) ermitteln. Allerdings können oder sollen subjektive Tests aufgrund ihres hohen Aufwands an Vorbereitungen, Zeit, Material und Kosten oft nicht durchgeführt werden [ASS02, YWWC02]. Wünschenswert sind deshalb objektive Maße, die mit subjektiven Tests möglichst gut übereinstimmende Ergebnisse liefern. Dies gilt für viele Bewertungsaufgaben, darunter auch die objektive Bewertung der Wirksamkeit von Artefaktreduktionsverfahren bezogen auf eine Bildqualitätsverbesserung - zentrales Thema dieser Arbeit. Da es sich hierbei um ein sehr spezielles Thema bzw. Problem handelt, wird es in der Literatur bislang allenfalls am Rande behandelt. Lediglich Metriken zur Messung von Blockartefakten wurden in diesem Zusammenhang schon verwendet, ohne den Nachweis zu erbringen, dass damit eine Bewertung der Wirksamkeit von Deblocking-Verfahren möglich ist, die den Eindrücken menschlicher Betrachter weitgehend entspricht.

Wie an späterer Stelle in dieser Arbeit gezeigt wird, sind bekannte Metriken zur Messung von Blockartefakten tatsächlich nur sehr begrenzt zur Wirksamkeitsschätzung von Deblocking-Verfahren bezüglich einer Bildqualitätsverbesserung geeignet. Gleiches gilt für andere bereits in Kapitel 4 vorgestellte Qualitätsbewertungsansätze, insbesondere auch für solche, die speziell dazu entwickelt wurden, die absolute Qualität komprimierter und somit durch Codierartefakte gestörter Bilder bzw. Bildfolgen zu bewerten. Dies ist insofern nicht verwunderlich, als dass bei der Entwicklung der Qualitätsmaße meist spezielle Probleme

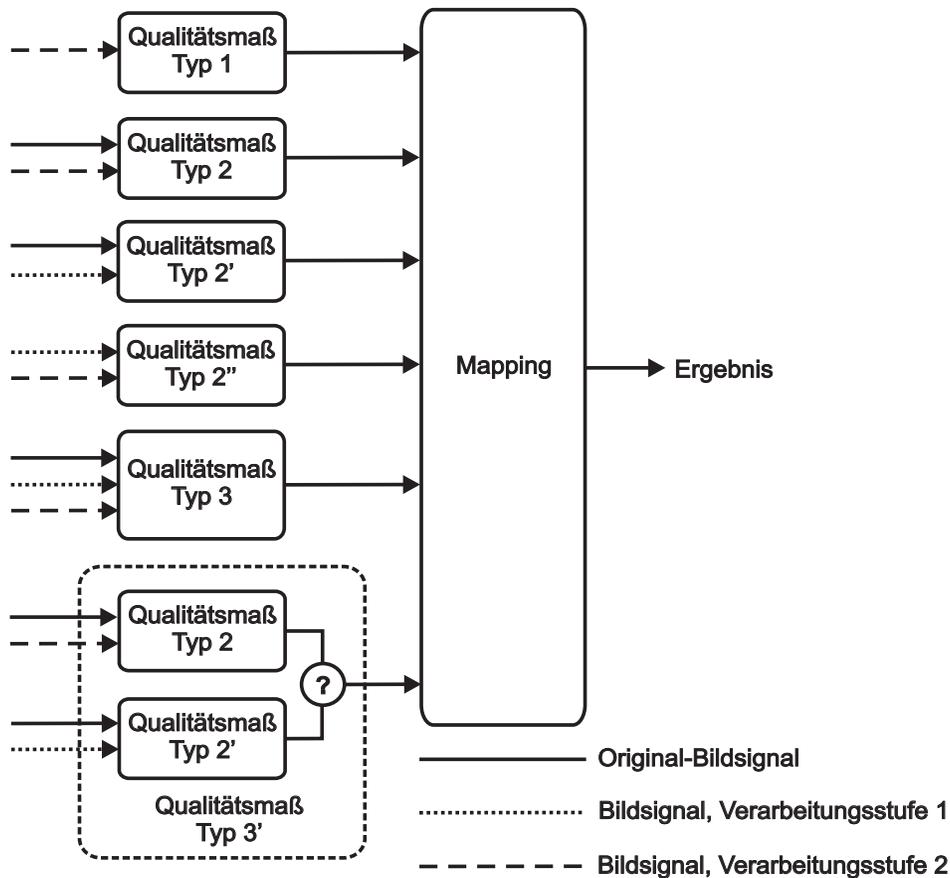
im Fokus standen und die Qualitätsmaße entsprechend stark auf die Problemstellung hin optimiert wurden. Somit kann für viele Verfahren eine gute Eignung zu ihrem primären Zweck nachgewiesen werden (z. B. Bewertung der absoluten Qualität komprimierter Bilder), eine Eignung für abweichende - zum Teil sogar verwandte - Aufgabenstellungen (z. B. Bewertung der Wirksamkeit von Deblocking-Verfahren) muss aufgrund der Spezialisierung jedoch nicht gegeben sein.

Auf eine spezielle Anpassung von Qualitätsbewertungsverfahren zu verzichten, um ein breiteres Anwendungsfeld zu erschließen, ist zwar wünschenswert, allerdings mit dem Problem verbunden, dass die Schätzgenauigkeit der Verfahren im Vergleich zu spezialisierten Metriken darunter leidet. Bis heute wurde kein Verfahren vorgestellt, das in einem breiten Anwendungsfeld eine akkurate Schätzung der von menschlichen Betrachtern wahrgenommenen Qualitätseindrücke ermöglicht. In [Fec99] wird aufgrund der hohen Komplexität des HVS sogar bezweifelt, dass ein solches Verfahren überhaupt entwickelt werden kann. Wenn eine hohe Schätzgenauigkeit gefordert ist, ist daher der Einsatz von speziell an den Anwendungszweck angepassten Qualitätsmaßen empfehlenswert [Fec99, BBK05]. Allerdings bedeutet dies einen hohen Aufwand, da für neue Anwendungen auch stets neue Qualitätsmaße entwickelt oder zumindest vorhandene Maße entsprechend modifiziert werden müssen. Um eine optimale Schätzgenauigkeit bei der Bewertung der Wirksamkeit von Deblocking-Verfahren in Bezug auf eine Verbesserung der Bildqualität von sichtbar gestörten Bildern (auf einer vergleichenden Skala) zu erhalten, wurde der Neuentwicklung eines geeigneten Qualitätsbewertungsverfahrens gegenüber der Anpassung existierender Verfahren in dieser Arbeit der Vorzug gegeben. Gleichzeitig sollte die Neuentwicklung ihrerseits allerdings die nötige Flexibilität bieten, um an neue oder veränderte Aufgabenstellungen angepasst werden zu können, ohne dabei viele Kompromisse eingehen zu müssen. Das neu entwickelte Verfahren wird in diesem Kapitel vorgestellt. Dabei wird zunächst von einem allgemeinen Faktorenmodell ausgegangen, bevor mehrere zum Einsatz in dem Modell entwickelte Teilmaße und mögliche Kombinationen in Form von linearen Regressionsmodellen vorgestellt werden.

## 8.1 Allgemeines Faktorenmodell

Die in Kapitel 4 vorgestellten Faktorenmodelle, welche auf der Struktur in Abbildung 4.5 basieren, ermöglichen grundsätzlich durch den Austausch der integrierten Maße eine Anpassung an neue bzw. gegenüber dem ursprünglichen Einsatzzweck veränderte Aufgabenstellungen. Sie stellen eine geeignete Basis zur Entwicklung von neuen Qualitätsbewertungsverfahren dar und erschienen auch als Grundlage zur Bewertung der Wirksamkeit von Deblocking-Verfahren, dem Thema dieser Arbeit, geeignet. Um allerdings im Vergleich zu der Struktur in Abbildung 4.5 noch mehr Flexibilität zu gewinnen und ein möglichst breites Spektrum an denkbaren Teilmaßen einsetzen zu können, wurde die ursprüngliche

Struktur erweitert und ein verallgemeinertes Faktorenmodell zur Bildqualitätsbewertung und zum Vergleich der Wirksamkeit von Bildverarbeitungsalgorithmen entworfen. In Abbildung 8.1 ist dieses allgemeine Faktorenmodell skizziert.



**Abbildung 8.1:** Allgemeines Faktorenmodell zur Bildqualitätsbewertung und zum Vergleich von Bildverarbeitungsalgorithmen

Das Modell unterscheidet drei verschiedene Arten von Eingangsdaten. Dies sind Referenz- bzw. Originalbilder, Bildsignale auf Verarbeitungsstufe 1 und Bildsignale auf Verarbeitungsstufe 2. Die beiden Verarbeitungsstufen bezeichnen verschiedene Positionen in einem Bildverarbeitungssystem, wobei das Modell nicht vorschreibt, wo genau sich diese Positionen befinden müssen. Bildsignale bzw. Bilder auf Verarbeitungsstufe 1 können z. B. codierte Bilder sein und Bildsignale auf Verarbeitungsstufe 2 können codierte und anschließend gefilterte bzw. nachverarbeitete Bilder sein. Unter codierten Bildern sind in diesem Zusammenhang codierte und bereits wieder decodierte Bilder zu verstehen, also Bilder, die Codierartefakte enthalten. Die verschiedenen Typen von Qualitätsmaßen unterscheiden sich nach Anzahl und Art der Eingangsdaten. Ein Qualitätsmaß vom Typ 1

ist z. B. einzig auf Basis eines bereits nachverarbeiteten bzw. gefilterten Bildes (Verarbeitungsstufe 2) berechenbar. Ein Qualitätsmaß vom Typ 2 verwendet als Eingangsdaten zwei Bilder, beispielsweise ein Originalbild und ein nachverarbeitetes Bild oder ein codiertes und ein nachverarbeitetes Bild. Als Beispiel für Qualitätsmaße vom Typ 2 seien der MSE und der PSNR genannt. Ein Qualitätsmaß vom Typ 3 benötigt hingegen sowohl ein Originalbild, als auch je ein Bild auf beiden Verarbeitungsstufen. Typ 3' als Spezialfall benötigt ebenfalls alle drei Eingangsbilder, jedoch lässt sich das Maß auf zwei Qualitätsmaße des Typs 2 zurückführen, die geeignet miteinander kombiniert werden müssen, was in der Abbildung durch ein Fragezeichen symbolisiert ist. Als Beispiel sei der DPSNR genannt. Die Verknüpfung ist in diesem Fall eine Differenzbildung. Grundsätzlich ließe sich das Modell durch die Berücksichtigung zusätzlicher Verarbeitungsstufen noch erweitern, was für diese Arbeit aber keine Relevanz hat und somit auch nicht weiter verfolgt wird.

Das allgemeine Faktorenmodell bietet die Flexibilität, um eine Vielzahl bekannter Qualitätsmetriken als Teilmaße integrieren zu können. Gleichzeitig bietet es viele Freiheiten bei der Entwicklung neuer Maße, die im Modell verwendet werden sollen. Bezüglich der Verwendung einer geeigneten Mapping-Funktion bestehen keine Vorgaben. Somit können auf Basis des allgemeinen Faktorenmodells Qualitätsbewertungsverfahren realisiert werden, die an sehr unterschiedliche - auch sehr spezielle - Aufgabenstellungen in der (Bild-)Signalverarbeitung angepasst sind. Das Spektrum der Möglichkeiten reicht von Aufgaben im Umfeld der absoluten Bildqualitätsbewertung über Bildqualitätsvergleiche bis hin zur objektiven Bewertung bestimmter Performance-Kriterien von Bildverarbeitungsalgorithmen. Auch zur Bewertung der Wirksamkeit von Deblocking-Verfahren stellt das allgemeine Faktorenmodell eine gut geeignete Basis dar, wie noch gezeigt wird.

## 8.2 Randbedingungen des neuen Ansatzes zur Bewertung von Deblocking-Verfahren

Der im Folgenden im Detail beschriebene neue Qualitätsbewertungsansatz wurde von Anfang an und ausschließlich auf die Bewertung der Wirksamkeit von Deblocking-Verfahren bezüglich einer Bildqualitätsverbesserung von sichtbar gestörten Bildern hin optimiert. Unter sichtbar gestörten Bildern sind solche zu verstehen, die bei gegebenem Betrachtungsabstand (siehe dazu Kapitel 7) einen Störgrad aufweisen, der von menschlichen Betrachtern ohne besondere Anstrengungen wahrzunehmen ist. Allerdings war es durchaus das Ziel, unter den genannten Bedingungen ein Spektrum von Bildern mit wenigen und vergleichsweise schwachen Störungen bis hin zu Bildern mit sehr vielen und starken Störungen abzudecken. Um eine akkurate und stabile Schätzung der Wirksamkeit von Deblocking-Verfahren in sehr unterschiedlichen Bildern zu erzielen, erschien es grundsätzlich sinnvoll, nicht nur ein einziges Merkmal im Kontext von Codierartefakten und deren Beseitigung

zu berücksichtigen, sondern mehrere Merkmale zu analysieren und die Analyseergebnisse zu kombinieren. Als Grundlage für den neuen Qualitätsbewertungsansatz erschien somit das allgemeine Faktorenmodell in Abbildung 8.1 gut geeignet. Es wurde durch die Neuentwicklung und Integration von mehreren Qualitätsmaßen, die bestimmte Bild- bzw. Störmerkmale analysieren, sowie durch die Wahl einer geeigneten Mapping-Funktion an die spezielle Aufgabenstellung angepasst. Bei der Entwicklung der einzelnen Qualitätsmaße wurde berücksichtigt, dass Deblocking-Verfahren gestörte Bilder oft nur verhältnismäßig wenig und mitunter sehr gezielt verändern und somit zur Bewältigung der gestellten Bewertungsaufgabe eine besonders genaue Erfassung von zum Teil nur kleinen Änderungen bestimmter Merkmale notwendig ist. Im Vordergrund standen Störmerkmale, die üblicherweise in intra-codierten Bildern auftreten. Der neue Qualitätsbewertungsansatz sollte objektive Bewertungen liefern, die eine hohe Übereinstimmung mit den subjektiven Eindrücken menschlicher Betrachter aufweisen und es sollte eine flexible Anpassbarkeit an veränderte Aufgabenstellungen gegeben sein. Nicht zuletzt wurde auch einer moderaten Komplexität bzw. einem akzeptablen Rechenaufwand einige Bedeutung beigemessen.

Das neue Verfahren analysiert rein örtliche Bild- bzw. Störmerkmale auf Frame-Basis und berücksichtigt aktuell keine zeitlichen Effekte. Da nahezu sämtliche in der Literatur veröffentlichten Deblocking-Verfahren ebenfalls rein örtlich auf Frame- oder Field-Basis arbeiten, ist die Analyse von örtlichen Störmerkmalen und den örtlichen Auswirkungen eines Deblockings offensichtlich im Hinblick auf die Bewertung der Wirksamkeit von Deblocking-Verfahren besonders bedeutend. Dennoch sei darauf hingewiesen, dass in diesem Kontext auch zeitliche Effekte eine nicht zu vernachlässigende Wirkung auf die Störwahrnehmung menschlicher Betrachter haben können, nicht zuletzt weil verschiedene Frame-Typen in Videosequenzen auch sehr unterschiedliche örtliche Störmerkmale im Zeitverlauf aufweisen können. Es war in dieser Arbeit jedoch nicht möglich, eine hinreichend umfassende Untersuchung zeitlicher Effekte einzubeziehen, da nahezu keine relevanten Voruntersuchungen zur Bewertung der Wirksamkeit von Artefaktreduktionsverfahren vorlagen und allein die Untersuchung örtlicher Merkmale von erheblichem Umfang war. Die Einbeziehung zeitlicher Effekte könnte ein interessanter Aspekt für erweiternde zukünftige Untersuchungen sein.

Die entwickelte Bildqualitätsanalyse ist auf die Luminanz beschränkt. Wie bereits dargestellt, ist der Luminanzkanal als Hauptinformationsträger maßgeblich für das Auftreten und die Sichtbarkeit von Blockartefakten verantwortlich. Eine deutliche Bildqualitätsverbesserung wird daher im Wesentlichen durch die Verringerung von Blockartefakten im Luminanzkanal erzielt. Ein zusätzliches Deblocking der Chrominanzkanäle führt nur sehr vereinzelt zu leichten Bildqualitätsverbesserungen und ist daher kaum lohnenswert, worauf in Kapitel 7 bereits eingegangen wurde. Da aus den genannten Gründen im Rahmen der durchgeführten subjektiven Tests beschlossen wurde, sämtliche eingesetzten Deblocking-Verfahren nur auf den Luminanzkanal anzuwenden, erübrigten sich hier Analysen der Chrominanzkanäle. Ergänzend sei angemerkt, dass Untersuchungen in [Fec00] ergeben ha-

ben, dass sich durch die Einbeziehung der Analyse von Chrominanzinformationen in die dort vorgestellten Verfahren zur Bewertung der absoluten Qualität MPEG-codierter Sequenzen kaum Verbesserungen erzielen ließen, da aus den Chrominanzkanälen nahezu keine zusätzlichen Informationen zu gewinnen waren. Aufgrund der Verwandtschaft der Untersuchungen bzw. Aufgabenstellungen musste auch im Kontext dieser Arbeit mit ähnlichen Tendenzen gerechnet werden, was die getroffene Entscheidung untermauert.

Als Eingangsmaterial verwendet das Verfahren jeweils ein Referenzbild, ein codiertes und decodiertes - also mit Artefakten versehenes - Bild und ein nachgefiltertes Bild. Auf diese Weise steht ein Maximum an verwertbaren Daten zur Qualitätsbewertung zur Verfügung. Da das Verfahren nicht der permanenten Kontrolle der Wirksamkeit von Artefaktreduktionsverfahren dient, sondern eine Analyse- und Entscheidungshilfe im Umfeld der Neu- und Weiterentwicklung sowie der Optimierung von Deblocking-Verfahren darstellen soll, bedeutet die Notwendigkeit des Zugriffs auf Referenzbildmaterial keine Einschränkung. Das neue Verfahren ermittelt einen Qualitätswert, der eine Aussage darüber machen soll, inwieweit ein Artefaktreduktionsverfahren ein gestörtes Bild bezüglich seiner Qualität verbessern kann. Der Qualitätswert wird auf einer Comparison-Skala (DMOS - Differential Mean Opinion Score) nach [ITU02], die den direkten Vergleich zweier Bilder (gestörtes und nachgefiltertes Bild) bezüglich ihrer Qualität ermöglicht, angegeben. Zur Optimierung des Verfahrens wurden die Daten aus den ersten beiden subjektiven Tests, welche in Kapitel 7 beschrieben wurden, verwendet.

### 8.3 Teilkomponenten/Qualitätsmaße

Ein bedeutender Schritt bei der Entwicklung des neuen Ansatzes zur Bewertung der Wirksamkeit von Deblocking-Verfahren bezogen auf eine Bildqualitätsverbesserung war die Identifizierung von Merkmalen, die im Zusammenhang mit der Kompression von Bildern stehen und an denen sich eine Qualitätsverbesserung oder Qualitätsverschlechterung ausmachen lässt. Dabei musste auch berücksichtigt werden, inwieweit die Anwendung von Deblocking-Verfahren solche Merkmale beeinflussen kann. Interessant erschienen in diesem Zusammenhang besonders folgende Merkmale:

- Diskontinuitäten an Blockgrenzen,
- „Glattheit“ im Bereich von Blockgrenzen,
- Grauwertanzahlen in örtlich begrenzten Bildbereichen, da diese z. B. durch Quantisierungseffekte vermindert werden können,
- Veränderung oder Degradation von Kanten bzw. Konturen,
- Ähnlichkeit oder Unähnlichkeit von Bildern.

Im Folgenden werden im Rahmen der Arbeit entwickelte Qualitätsmaße vorgestellt, die auf die Analyse einzelner der genannten Merkmale abgestimmt sind. Wie bereits erwähnt, war eine Kombination mehrerer dieser Maße auf Basis des allgemeinen Faktorenmodells in Abbildung 8.1 vorgesehen, was stellenweise zu berücksichtigen war. Wie an späterer Stelle in diesem Kapitel noch gezeigt wird, eignet sich eine Linearkombination zum Mapping der Qualitätswerte der einzelnen Maße auf ein Gesamtmaß.

### 8.3.1 Diskontinuitätsmaß

Eines der markantesten Merkmale einer blockbasierten Bildcodierung ist das Auftreten von Blockartefakten in decodierten Bildern. Diese äußern sich durch Diskontinuitäten an Blockgrenzen und sind für menschliche Betrachter besonders störend wahrnehmbar (siehe Kapitel 2). Es erschien daher geboten, ein Qualitätsmaß zu entwickeln, das auf der Analyse von Diskontinuitäten an Blockgrenzen basiert. Abbildung 8.2 veranschaulicht die Funktionsweise des *Diskontinuitätsmaßes*, das einem Maß des Typs 2' nach der in Abbildung 8.1 eingeführten Unterscheidung entspricht.

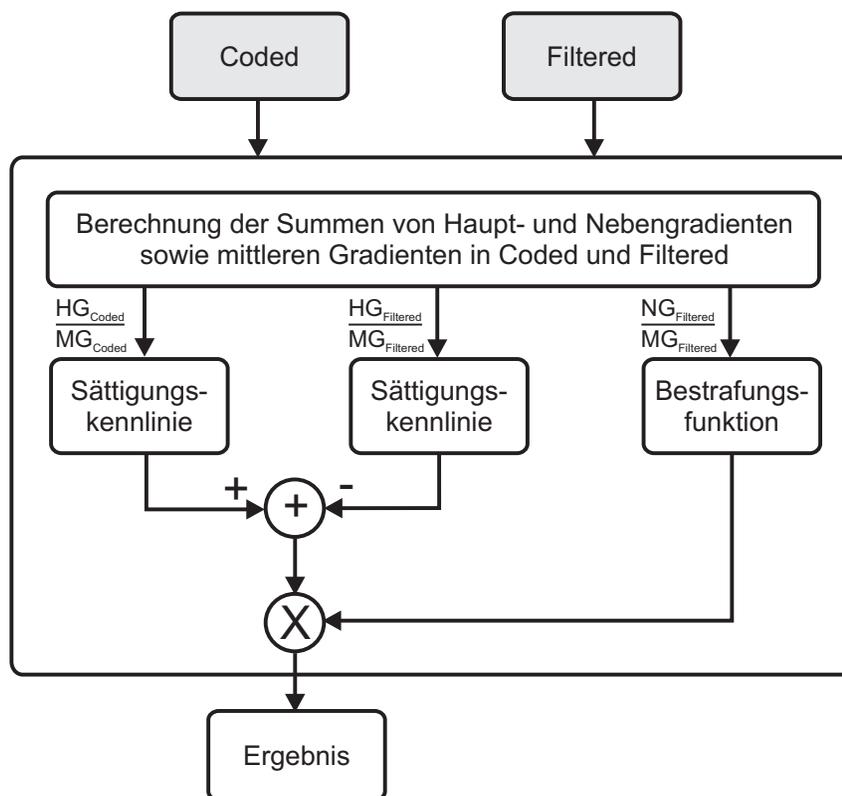
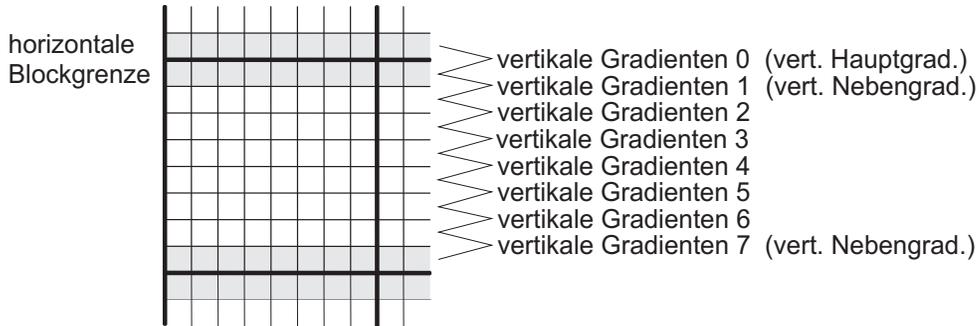


Abbildung 8.2: Funktionsweise des Diskontinuitätsmaßes

Es basiert auf der Berechnung und dem Vergleich der Summen absoluter Pixelintensitätsdifferenzen an verschiedenen Rasterlagen im codierten und im nachgefilterten Bild. Entsprechende Originalbilder werden nicht benötigt. Blockgrößen von 8x8 Pixeln werden vorausgesetzt. Mit Hauptgradienten werden Intensitätsdifferenzen über Blockgrenzen hinweg bezeichnet, während mit Nebengradienten Intensitätsdifferenzen in direkter Nachbarschaft zu Blockgrenzen bezeichnet werden. Abbildung 8.3 veranschaulicht am Beispiel einer horizontal verlaufenden Blockgrenze die Lage der vertikalen Gradienten. Für die Lage der horizontalen Gradienten und ihre Beziehung zu vertikal verlaufenden Blockgrenzen gilt Entsprechendes.



**Abbildung 8.3:** Veranschaulichung der Gradientenlagen

Die Berechnungsvorschrift für die Summe der Hauptgradienten  $HG$  in einem Bild  $s$  mit der Bildhöhe in Zeilen  $PH$  und der Bildbreite in Spalten  $PW$  setzt sich aus Teilsummen für vertikale und horizontale Gradienten zusammen und lautet:

$$HG = HG_V + HG_H, \quad (8.1)$$

$$HG_V = \sum_x \sum_y abs(s(x, y + 1) - s(x, y)), \quad (8.2)$$

mit  $x \in \{0, \dots, PW - 1\}$  und  $y = 8m - 1, m \in \{1, \dots, PH/8 - 1\}$ ,

$$HG_H = \sum_x \sum_y abs(s(x + 1, y) - s(x, y)), \quad (8.3)$$

mit  $x = 8n - 1, n \in \{1, \dots, PW/8 - 1\}$  und  $y \in \{0, \dots, PW - 1\}$ .

Entsprechend ist die Summe der Nebengradienten  $NG$  zu berechnen. Es ist zu berücksichtigen, dass doppelt so viele Nebengradienten (vertikale und horizontale Gradienten 1 sowie vertikale und horizontale Gradienten 7) wie Hauptgradienten (vertikale und horizontale Gradienten 0) existieren, weshalb die Nebengradienten mit  $\frac{1}{2}$  zu gewichtet sind. Des Weiteren wird der Mittelwert der Summen der Gradienten 2 bis 6,  $MG$ , benötigt.

$$MG = MG_V + MG_H, \quad (8.4)$$

$$MG_V = \frac{1}{5} \sum_i \sum_x \sum_y abs(s(x, y + i + 1) - s(x, y + i)), \quad (8.5)$$

mit  $x \in \{0, \dots, PW - 1\}$ ,  $y = 8m - 1$ ,  $m \in \{1, \dots, PH/8 - 1\}$ ,  $i \in \{2, \dots, 6\}$ ,

$$MG_H = \frac{1}{5} \sum_i \sum_x \sum_y abs(s(x + i + 1, y) - s(x + i, y)), \quad (8.6)$$

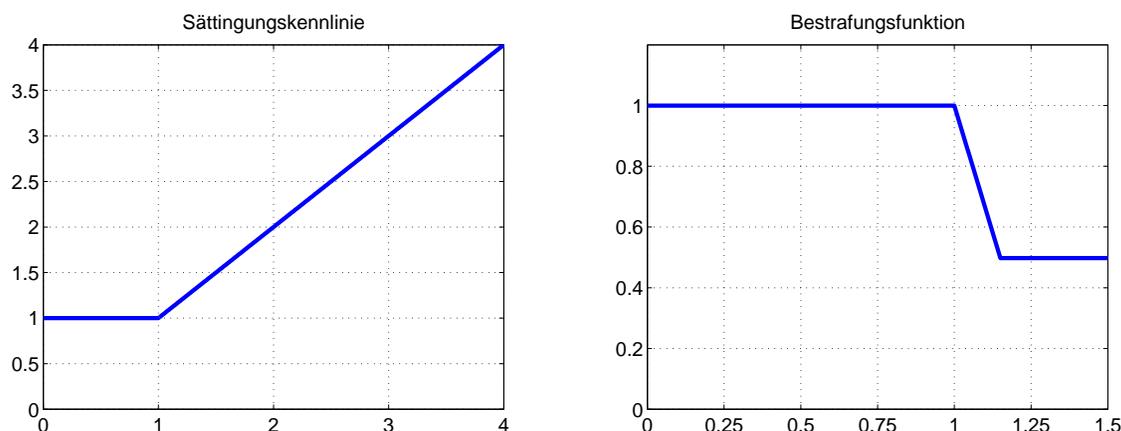
mit  $x = 8n - 1$ ,  $n \in \{1, \dots, PW/8 - 1\}$ ,  $y \in \{0, \dots, PW - 1\}$ ,  $i \in \{2, \dots, 6\}$ .

Die Gradienten an den übrigen Rasterlagen werden deshalb nicht bei der Berechnung von  $MG$  berücksichtigt, da sie durch eine Bildkompression besonders stark beeinflusst werden, was hier unerwünscht ist, da  $MG$  der Normierung dient.

Wie bereits in Kapitel 4 bei der Vorstellung gradientenbasierter Blocking-Maße dargestellt, vergrößert eine Bildkompression die Hauptgradienten im Vergleich zu denen in unkomprimierten Bildern. Die übrigen Gradienten werden durch Detailverluste, die mit einer Bildkompression einhergehen, tendenziell kleiner im Vergleich zu den entsprechenden Gradienten in unkomprimierten Bildern. Somit liefert das Verhältnis  $\frac{HG}{MG}$  ein Maß für das Vorhandensein und die Stärke von Blockartefakten bzw. Diskontinuitäten an Blockgrenzen. In natürlichen unkomprimierten Bildern ist das Verhältnis ca. Eins, in komprimierten Bildern ist es größer als Eins. Da Verfahren zur Reduktion von Blockartefakten Diskontinuitäten an Blockgrenzen verkleinern, kann ihre Wirksamkeit diesbezüglich durch Vergleich des Verhältnisses  $\frac{HG}{MG}$  vor und nach Anwendung eines Verfahrens beurteilt werden. Auf diesem Prinzip beruht das Diskontinuitätsmaß.

Es muss jedoch ein Problem berücksichtigt werden, das bereits in Kapitel 4 skizziert wurde. Eine zu starke Glättung von Blockgrenzbereichen kann dazu führen, dass die Hauptgradienten durchschnittlich kleiner werden als die übrigen Gradienten. Es sollte sichergestellt werden, dass in diesem Fall keine zu gute Bewertung der Wirksamkeit eines Deblocking-Verfahrens vorgenommen wird. Ein theoretisch optimales Deblocking-Verfahren beseitigt alle durch eine Bildkompression verursachten Diskontinuitäten vollständig, ohne dabei eine zu starke Glättung vorzunehmen. Sämtliche Kanten - insbesondere auf Blockgrenzen - werden gleichzeitig geschützt und nicht beeinträchtigt. Ein optimales Deblocking-Verfahren stellt das Verhältnis  $\frac{HG}{MG}$  auf einen Wert von Eins ein. Kein Verfahren kann Diskontinuitäten besser beseitigen als das angenommene theoretisch optimale Verfahren. Deshalb wird hier eine Verkleinerung des Verhältnisses  $\frac{HG}{MG}$  im Diskontinuitätsmaß nur bis zu einem Wert von Eins als Verbesserung gewertet. Somit existiert eine vom Diskontinuitätsmaß anzeigbare bildabhängige maximale Verringerung von Diskontinuitäten an Blockgrenzen. Wie aus Abbildung 8.2 abgelesen werden kann, wird die Begrenzung der anzeigbaren Verringerung von Diskontinuitäten durch eine (einheitenlose) Sättigungskennlinie realisiert. Diese ist in Abbildung 8.4 dargestellt.

Eine Bestrafung von Deblocking-Verfahren, die eine zu starke Glättung von Blockgrenzbereichen vornehmen, wurde in diesem Maß über die beschriebene Sättigung hinaus nicht implementiert, da in solchen Fällen keine deutlich negativen Auswirkungen auf die Bildqualität festgestellt wurden. Es wurde jedoch eine im Folgenden beschriebene Bestra-



**Abbildung 8.4:** Sättigungskennlinie und Bestrafungsfunktion im Diskontinuitätsmaß, Abszisse: Eingangswerte, Ordinate: Ausgangswerte

fung für den Fall vorgesehen, dass ein Deblocking-Verfahren Blockgrenzen mit sehr kurzen bzw. zu kurzen Filtern bearbeitet. Durch zu kurze Filter werden Diskontinuitäten von den Blockgrenzen in die direkten Nachbarbereiche der Blockgrenzen verschoben und bleiben insbesondere in homogenen Bildbereichen weiterhin sichtbar. Die Hauptgradienten liefern keine Hinweise auf ein solches Problem, aber die Nebengradienten sind zur Erkennung des beschriebenen Effektes geeignet. Das Verhältnis  $\frac{NG}{MG}$  ist nach einer Bildkompression typischerweise etwas kleiner als Eins. Eine Deblocking-Filterung mit sehr kurzen Filtern ist durch eine starke Verkleinerung der Hauptgradienten und eine damit verbundene Vergrößerung der Nebengradienten geprägt. Dies führt dazu, dass das Verhältnis  $\frac{NG}{MG}$  Werte größer als Eins annimmt, welche daher als Hinweis auf sehr kurze Filter gelten können.

Die größten Werte des Verhältnisses  $\frac{NG}{MG}$ , die im Rahmen der durchgeführten Untersuchungen gemessen wurden, lagen bei etwa 1,25. Dies musste beim Entwurf einer geeigneten Bestrafungskennlinie, die - wie in Abbildung 8.2 erkennbar - eine vom Verhältnis  $\frac{NG}{MG}$  des gefilterten Bildes abhängige Gewichtung des Diskontinuitätsmaßes vornimmt, berücksichtigt werden. Ebenso musste berücksichtigt werden, dass kurze Filter insbesondere in homogenen Bildbereichen zwar nur eine begrenzte Verringerung der Wahrnehmbarkeit von Diskontinuitäten an Blockgrenzen ermöglichen, ein gewisses Maß an Verringerung jedoch nicht abgestritten werden kann. Die Bestrafungskennlinie durfte daher nicht bis auf Null abfallen. Die Forderungen an die Kennlinie waren insgesamt ein flacher Verlauf für Eingangswerte bis Eins, ein steiler Abfall für Eingangswerte zwischen Eins und ca. 1,1 ... 1,25, sowie eine Sättigung auf einem Niveau von etwa  $\frac{2}{3}$  ...  $\frac{1}{3}$ . Die für die weiteren Untersuchungen ausgewählte Bestrafungskennlinie ist in Abbildung 8.4 dargestellt. Sie wurde heuristisch so gewählt, dass sie zu einer guten Übereinstimmung des Verhaltens des Diskontinuitätsmaßes mit den Eindrücken menschlicher Betrachter positiv beiträgt.

Insgesamt weist das Diskontinuitätsmaß eine hohe Korrelation mit den Bewertungen auf, die im Rahmen der in dieser Arbeit durchgeführten subjektiven Tests ermittelt wurden. Dies gilt erwartungsgemäß für die Eindrücke der Probanden speziell bezogen auf die Verringerung von Blockartefakten, aber auch für die Eindrücke bzgl. einer Gesamtbildqualitätsverbesserung.

### 8.3.2 Glättungsmaß

Um der besonderen Bedeutung von Diskontinuitäten und deren Glättung im Zusammenhang mit der Bewertung der Wirksamkeit von Deblocking-Verfahren Rechnung zu tragen, wurde ein weiteres Maß entwickelt, das auf die Analyse dieser Merkmale abzielt. Das *Glättungsmaß* unterscheidet sich vom bereits vorgestellten Diskontinuitätsmaß in der Funktionsweise erheblich, wie Abbildung 8.5 entnommen werden kann. Es wird nicht die Veränderung von Gradienten - also Intensitätsdifferenzen benachbarter Pixel - bewertet, sondern die Veränderung der Standardabweichungen  $\sigma = \sqrt{\frac{1}{7} \sum_{i=1}^8 (s_i - \bar{s})^2}$  von Bildpunktvektoren der Länge acht Pixel, die über Blockgrenzen hinweg verlaufen.  $\bar{s}$  ist der Mittelwert eines solchen Bildpunktvektors. Abbildung 8.6 veranschaulicht exemplarisch die Lage eines horizontalen und eines vertikalen Analysevektors. Intensitätssprünge an Blockgrenzen führen tendenziell zu großen Standardabweichungen der Analysevektoren. Glättungsfiler mit großer Wirkungsbreite wirken sich hingegen stark verkleinernd auf die Standardabweichungen der Vektoren aus. Daher sind die Standardabweichungen zur Bewertung der Wirksamkeit von Deblocking-Verfahren geeignet.

Grundsätzlich ist zunächst vorgesehen, die Standardabweichungen der Bildpunktvektoren an allen Blockgrenzpositionen im codierten und im durch ein Artefaktreduktionsverfahren bearbeiteten Bild miteinander zu vergleichen. Ist die Standardabweichung im gefilterten Bild kleiner als im ungefilterten Bild, kann von einer Glättung ausgegangen werden. Die Differenz zwischen den Standardabweichungen der beiden miteinander verglichenen Vektoren wird dann akkumuliert. Es sollten jedoch nur Differenzen akkumuliert werden, wenn die entsprechende Glättung mit hoher Wahrscheinlichkeit zu einer Verbesserung der Bildqualität beiträgt. Ansonsten würden Deblocking-Verfahren, die unnötige Filterungen durchführen, durch das Glättungsmaß erheblich zu gut bewertet werden.

Um das zu verhindern, werden Blockgrenzbereiche aus der Messung ausgeschlossen, an denen keine Hinweise auf das Vorhandensein deutlicher Blockartefakte bestehen. Ob vom Vorhandensein von Artefakten ausgegangen werden muss oder nicht, wird mit Hilfe einer lokalen Diskontinuitätsmessung entschieden. Dazu wird ähnlich vorgegangen, wie im Rahmen der in Kapitel 6 vorgestellten adaptiven Reduktion von Blockartefakten mit lokaler Bildanalyse. An jeder horizontalen und vertikalen Blockgrenze werden Summen benachbarter Pixelintensitätsdifferenzen berechnet. Die Intensitätsdifferenzen direkt über die Blockgrenze hinweg werden als Hauptgradienten, die Intensitätsdifferenzen in Nach-

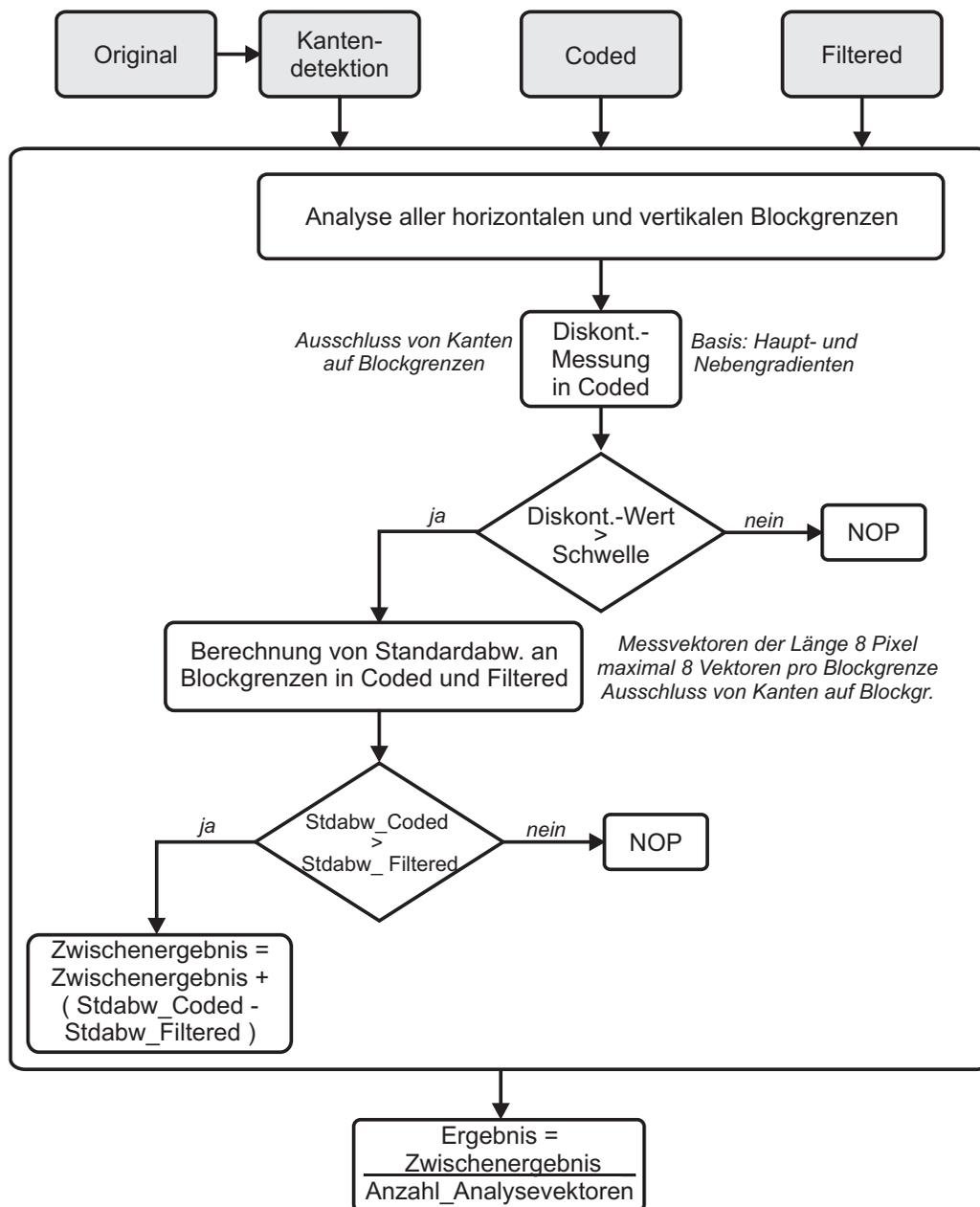
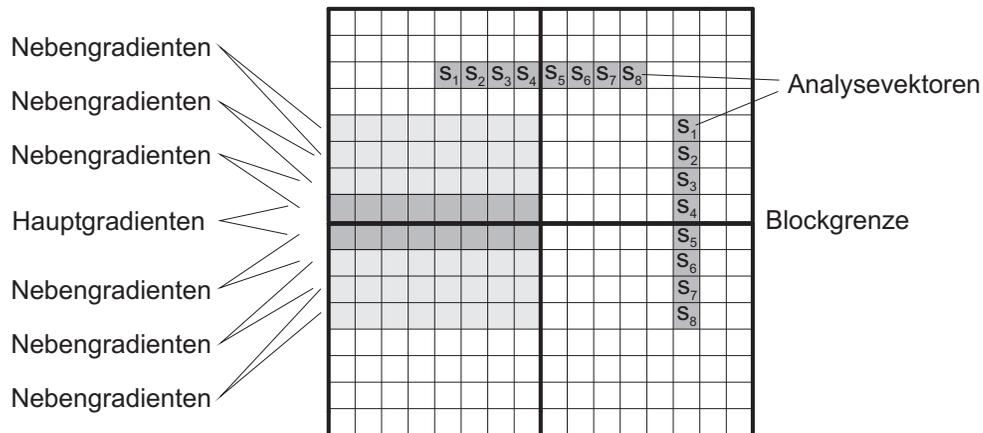


Abbildung 8.5: Funktionsweise des Glättungsmaßes, NOP  $\hat{=}$  no operation



**Abbildung 8.6:** Beispiel für die Lage der Analysevektoren und der Gradienten

barschaft zu den Blockgrenzen werden als Nebengradienten bezeichnet. Abbildung 8.6 veranschaulicht exemplarisch die Lage der Gradienten an einer horizontal verlaufenden Blockgrenze. Stellen, an denen Kanten bzw. Konturen direkt auf Blockgrenzpositionen liegen, werden aus der lokalen Diskontinuitätsmessung ausgeschlossen, um die Messung nicht zu verfälschen. Dies betrifft sowohl die Haupt- als auch die zugehörigen Nebengradienten. Zur Identifizierung von Kanten wird ein kombinierter Gradienten-Laplace-Kantendetektor (Schwellen: 90, -400), dessen Funktionsweise in Anhang C beschrieben wird, eingesetzt. Als Grundlage für die Kantendetektion dient das Originalbild, da dort Kanten mit der größten Genauigkeit gefunden werden können. Es hat sich als vorteilhaft erwiesen, die Kantendetektion nicht zu empfindlich einzustellen, um nicht unnötig viele Stellen aus der Analyse auszuschließen.

Aus dem Verhältnis der berechneten Haupt- zu den berechneten Nebengradienten (gewichtet mit  $\frac{1}{6}$ , da sechsmal mehr Neben- als Hauptgradienten aufsummiert werden) ergibt sich ein lokaler Diskontinuitäts-Messwert. Nur wenn dieser größer als ein Schwellwert ist, wird die Blockgrenze im Glättungsmaß weiter berücksichtigt. Allerdings bleiben Stellen, an denen Kanten auf Blockgrenzen identifiziert wurden, grundsätzlich aus den weiteren Analysen ausgeschlossen, da Deblocking-Filter dort in aller Regel nur ungewollte Kantenschleifungen hervorrufen können, die nicht positiv bewertet werden dürfen. Als Schwellwerte im Rahmen der lokalen Diskontinuitätsmessung wurden Werte zwischen Zwei und Drei untersucht. Bei einem Wert von Zwei war in bestimmten Testbildern keine deutliche Abnahme der Zahl an Analysestellen mit steigender Datenrate zu erkennen, während bei einem Wert von Drei bereits eine sehr starke Ausdünnung der Analysestellen bei mittleren Datenraten festzustellen war. Der Schwellwert 2,3 erwies sich als guter Kompromiss und auch im Hinblick auf die Kombination des Glättungsmaßes mit anderen Qualitätsmaßen als geeignete Lösung. Es sei angemerkt, dass als Nebeneffekt der Funktionsweise

der lokalen Diskontinuitätsmessung indirekt auch eine ungewollte positive Bewertung der Verschleifung von größeren Konturbereichen in direkter Nachbarschaft von Blockgrenzen verhindert wird. Konturen in der Nähe von Blockgrenzen vergrößern die Nebengradienten, wodurch die Wahrscheinlichkeit, dass der lokale Blocking-Wert über der Schwelle 2,3 liegt, verringert und eine ungewollte positive Bewertung von Glättungen unwahrscheinlicher wird.

Wie Abbildung 8.5 entnommen werden kann, werden die tatsächlich akkumulierten Differenzen der Standardabweichungen auf die Zahl der möglichen Analysevektoren normiert. Dadurch wird sichergestellt, dass in Bildern, in denen nur wenige Stellen existieren, an denen eine Glättung zu Qualitätsverbesserungen führen kann, das Maß einen kleineren Qualitätsgewinn anzeigt als in Bildern, in denen viele solcher Stellen existieren. Dies entspricht subjektiven Wahrnehmungen, denn Bildqualitätsverbesserungen durch Glättungen an wenigen Stellen wird ein menschlicher Betrachter weniger stark empfinden als Qualitätsverbesserungen an vielen Stellen. Insgesamt wird das Glättungsmaß also durch die Zahl der Stellen, an denen eine Glättung mit hoher Wahrscheinlichkeit zu einer Verbesserung der Bildqualität führt und durch das Ausmaß der Glättungen beeinflusst.

Das Glättungsmaß weist eine sehr hohe Korrelation mit den im Rahmen dieser Arbeit durchgeführten subjektiven Tests auf. Dies gilt sowohl für die Bewertung der Wirksamkeit von Artefaktreduktionsverfahren bezogen auf eine Verbesserung der Gesamtbildqualität als auch speziell bezogen auf die Verringerung von Blockartefakten. Nach der in Abbildung 8.1 eingeführten Unterscheidung entspricht das Glättungsmaß einem Qualitätsmaß des Typs 3, auch wenn der Zugriff auf das Originalbild nur der Durchführung einer zuverlässigen Kantendetektion dient.

Statt des Vergleichs von Standardabweichungen wurde auch ein Vergleich von Varianzen untersucht. Die Korrelationen des Glättungsmaßes mit subjektiven Eindrücken wurden dadurch jedoch erheblich vermindert. Dies liegt vermutlich daran, dass die Varianzen durch ihre quadratische Charakteristik in ungefilterten und gefilterten Bildern punktuell sehr unterschiedliche Werte haben können. Unter bestimmten Umständen können einige wenige Varianzdifferenzen daher einen erheblichen Einfluss auf das Glättungsmaß haben, was menschliche Betrachter nicht in ähnlicher Weise empfinden.

### 8.3.3 Konturdegradationsmaß

Die Anwendung von Verfahren zur Reduktion von Codierartefakten hat in gestörten Bildern nicht nur positive Auswirkungen. Da solche Verfahren, insbesondere wenn sie als reine Postprocessing-Verfahren realisiert sind, in den meisten Fällen auf einer Filterung von Bildbereichen oder einer damit vergleichbaren Operation basieren, besteht grundsätzlich die Gefahr einer ungewollten Verarbeitung von Stellen, in denen keine Störungen sichtbar sind. Dort kann eine Verschlechterung der Bildqualität durch Anwendung eines Artefaktreduktionsverfahrens eintreten. Um die Auswirkungen einer Artefaktreduktion umfassend

beurteilen oder bewerten zu können, kann es daher sinnvoll sein, auch negative Aspekte zu berücksichtigen. Ein negativer Aspekt kann z. B. die Verschleifung oder Beeinträchtigung von Kanten bzw. Konturverläufen sein, eine für menschliche Betrachter gut wahrnehmbare Störung. Das im Folgenden beschriebene Qualitätsmaß ist im Hinblick auf derartige Probleme entwickelt worden. Abbildung 8.7 zeigt die Funktionsweise des *Konturdegradationsmaßes*, das einem Maß des Typs 3 nach Abbildung 8.1 entspricht.

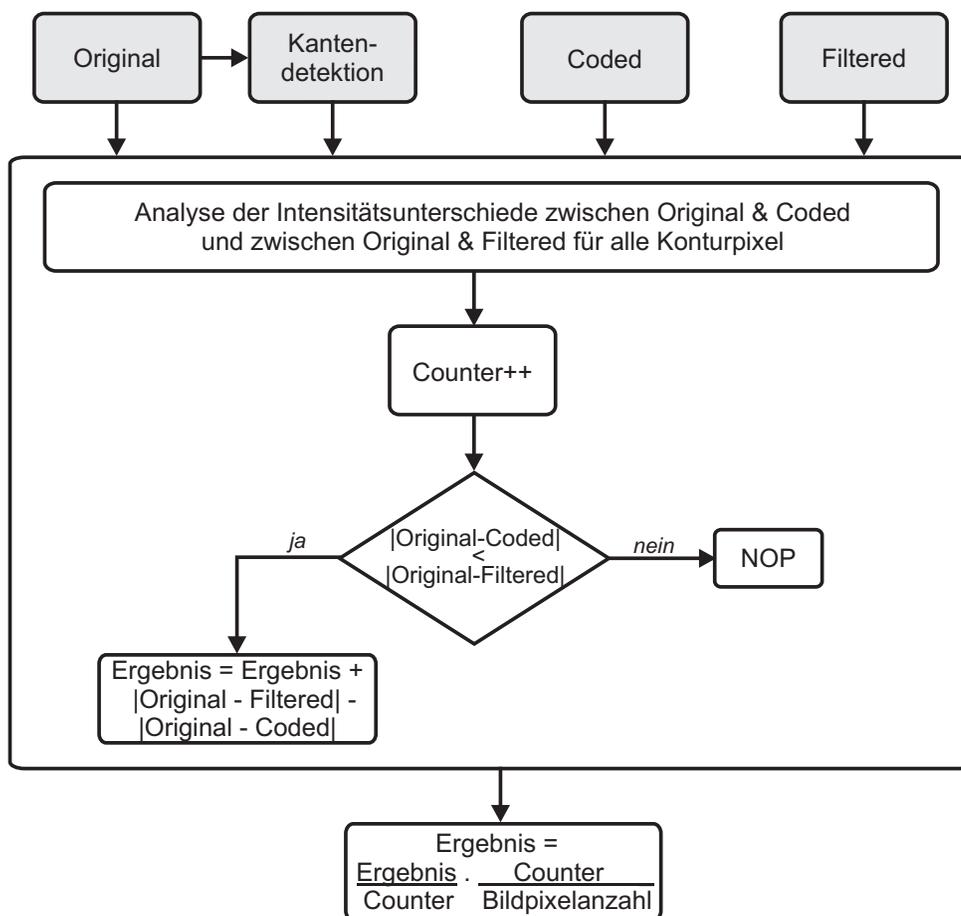


Abbildung 8.7: Funktionsweise des Konturdegradationsmaßes

Zunächst müssen Konturverläufe im Bildmaterial identifiziert werden. Dazu wird eine kombinierte Gradienten-Laplace-Kantendetektion (siehe Anhang C) mit den Schwellwerten 40 und -80 im Originalbild durchgeführt. Die gewählten Schwellen sorgen für eine empfindliche Kantendetektion, die eine verhältnismäßig niedrige Rauschempfindlichkeit besitzt. Es kann davon ausgegangen werden, dass sehr viele Kanten, kantenähnliche Signalverläufe und kontrastreiche Details für die weitere Verarbeitung identifiziert werden.

An den identifizierten Bildstellen wird bildpunktweise geprüft, ob die absolute Intensitätsdifferenz zwischen Original und codiertem Bild kleiner als die zwischen Original und gefiltertem Bild ist. Ist dies der Fall, muss damit gerechnet werden, dass die Anwendung eines Artefaktreduktionsverfahrens an der entsprechenden Stelle zu einer Degradation der im codierten Bild (noch) vorhandenen Informationen geführt hat. Die Verschleifung einer Kante oder eines Details, das durch die Bildcodierung nur wenig verändert wurde, wird z. B. zu einer messbaren Intensitätsdifferenzvergrößerung führen. Gemessene Intensitätsdifferenzvergrößerungen werden akkumuliert.

Nachdem das gesamte Bild bzw. die darin enthaltenen Konturpunkte analysiert wurden, werden die akkumulierten Intensitätsdifferenzvergrößerungen normiert und skaliert. Die Normierung erfolgt auf die Zahl der Konturpunkte. Dadurch wird sichergestellt, dass das Maß umso größer wird, je häufiger und stärker Degradationen unter den Konturpixeln auftreten. Jedoch wird einem menschlichen Betrachter die Degradation sehr weniger vorhandener Konturbereiche in einem Bild weniger unangenehm auffallen als die vergleichbar starke Degradation sehr vieler im Bild vorhandener Konturbereiche. Daher wird noch eine Skalierung des Maßes mit dem Verhältnis der Zahl an Konturpunkten zur Gesamtzahl der Bildpunkte vorgenommen. Selbstverständlich können Normierung und Skalierung zusammengefasst werden, was der Normierung des Maßes auf die Gesamtzahl der Pixel in einem Bild entspricht. Das Konturdegradationsmaß korreliert stark mit den im Rahmen dieser Arbeit durchgeführten subjektiven Tests bezogen auf die Unschärfewahrnehmung der Probanden. Eine hohe Korrelation mit den weiteren abgefragten Qualitätskriterien besteht erwartungsgemäß nicht.

### 8.3.4 Degradationsmaß

Als Alternative oder Ergänzung zum bereits vorgestellten Konturdegradationsmaß wurde ein Maß entwickelt, das ebenfalls auf die Erfassung ungewünschter Auswirkungen einer Artefaktreduktion abzielt. Es basiert auf dem gleichen Prinzip wie das Konturdegradationsmaß, wurde allerdings auf die Analyse aller Punkte eines Bildes, unabhängig davon, ob diese in Konturbereichen liegen oder nicht, erweitert. Abbildung 8.8 zeigt die Funktionsweise des *Degradationsmaßes*.

Es wird bildpunktweise geprüft, ob die absolute Intensitätsdifferenz zwischen Original und codiertem Bild kleiner als die zwischen Original und gefiltertem Bild ist. Wenn dies der Fall ist, wird von einer Degradationswirkung der Artefaktreduktion ausgegangen und der Intensitätsdifferenzunterschied akkumuliert. Das Maß erfasst somit die Tendenz von Artefaktreduktionsverfahren, bildpunktweise die Unähnlichkeit gestörter Bilder im Vergleich zu den Originalbildern weiter zu vergrößern. Die Normierung des Maßes kann einfacher gestaltet werden als beim Konturdegradationsmaß, da alle Bildpunkte unabhängig vom Vorhandensein bestimmter Bildmerkmale analysiert werden. Es wird der Wert der akkumulierten Intensitätsdifferenzunterschiede auf die Anzahl der Gesamtbildpunkte normiert.

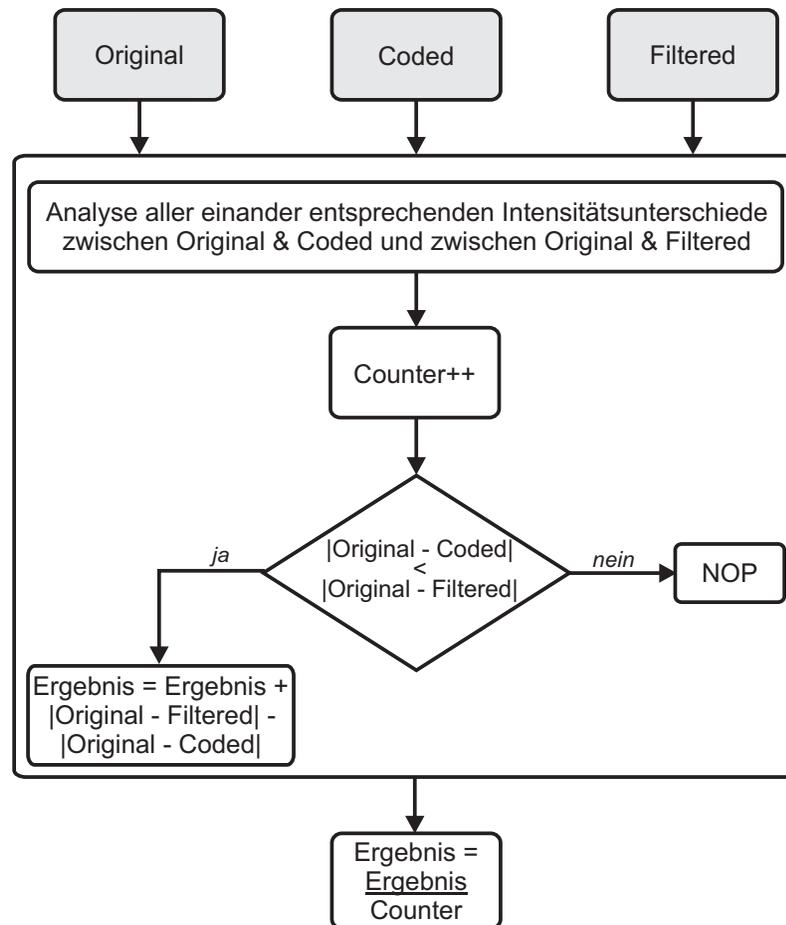


Abbildung 8.8: Funktionsweise des Degradationsmaßes

Auf diese Weise wirken sich sowohl die Stärke von Degradationen als auch deren Häufigkeit auf das Maß aus. Das Degradationsmaß korreliert stark mit den Unschärfe-Eindrücken der Teilnehmer an den im Rahmen dieser Arbeit durchgeführten subjektiven Tests. Die Korrelationen mit den anderen abgefragten Kriterien fallen geringer aus. Das Degradationsmaß ist ein Maß des Typs 3 im Sinne von Abbildung 8.1.

### 8.3.5 Wechselwirkung aus Diskontinuitäts- und Glättungsmaß

In Kapitel 3 wurde bereits das Thema Wechselwirkung zwischen Einflussgrößen im Rahmen der Regressionsanalyse eingeführt. Die Wechselwirkung zwischen Diskontinuitäts- und Glättungsmaß wurde als weiteres Maß in die Untersuchungen bzw. die an späterer Stelle in diesem Kapitel beschriebenen Regressionsanalysen mit einbezogen. Damit können Effekte

berücksichtigt werden, die sich durch das Zusammenwirken der beiden Maße ergeben. Das kann z. B. eine Kompensationswirkung sowohl in positiver als auch in negativer Richtung sein, wenn beide Maße gleichzeitig besonders große Werte annehmen. In der Praxis der Versuchsauswertung ist die Berücksichtigung von Wechselwirkungen üblich [Kle03]. Im Regelfall wird unter Wechselwirkung das Produkt zweier Variablen bzw. Einflussgrößen verstanden, was auch hier der Fall ist (Produkt aus Diskontinuitätsmaß und Glättungsmaß). Wird eine Wechselwirkung in ein Regressionsmodell eingefügt, müssen unbedingt auch die Variablen, von denen die Wechselwirkung abhängt, in das Modell aufgenommen werden.

### 8.3.6 Weitere Maße

Es wurden weitere Maße entwickelt, die im Folgenden nur in verkürzter Form vorgestellt werden sollen, da sie für die empfehlenswerten Varianten des in dieser Arbeit entworfenen kombinierten Ansatzes zur Wirksamkeitsbewertung von Deblocking-Verfahren eine weniger wichtige Rolle spielen.

#### Intensitätsannäherungsmaß

Grundlage des Intensitätsannäherungsmaßes ist die Überprüfung, inwieweit eine Artefaktreduktion zu einer Annäherung von Pixelintensitätsdifferenzen zwischen Original und gefiltertem Bild im Vergleich zu den Pixelintensitätsdifferenzen zwischen Original und codiertem Bild führt. Das Maß stellt im Grunde die Umkehrung des Degradationsmaßes dar und soll Bildqualitätsverbesserungen erfassen. Das Funktionsprinzip entspricht dem in Abbildung 8.8 dargestellten Prinzip. Jedoch wird nicht geprüft, ob  $|Original - Coded| < |Original - Filtered|$  gilt, sondern ob  $|Original - Coded| > |Original - Filtered|$  gilt. Wenn die Bedingung erfüllt ist, wird die Differenz der Wurzeln aus den absoluten Intensitätsdifferenzen akkumuliert, also  $\sqrt{|Original - Coded|} - \sqrt{|Original - Filtered|}$ . Die Korrelation des Maßes mit den Ergebnissen der subjektiven Tests wird auf diese Weise erheblich gegenüber der Akkumulation von Differenzen aus absoluten Intensitätsdifferenzen  $|Original - Coded| - |Original - Filtered|$  erhöht. Offenbar nehmen menschliche Betrachter moderate Bildqualitätsverbesserungen (hier Intensitätsannäherungen) an vielen Stellen im Bild deutlicher wahr als einzelne große Verbesserungen an sehr wenigen Stellen. Allerdings muss beim potenziellen Einsatz des Verfahrens im Rahmen eines Regressionsmodells überprüft werden, ob sich möglicherweise durch die Wurzel-Operationen nicht vernachlässigbare Nichtlinearitätsprobleme ergeben. Insgesamt korreliert das Maß sehr gut mit den subjektiven Eindrücken bzgl. der Gesamtqualitätsunterschiede und der Reduktion von Blockartefakten. Die Korrelation mit den Unschärfe-Eindrücken ist deutlich geringer.

### Grauwertmaß

Das Grauwertmaß ist ein Maß für Bildqualitätsverbesserungen. Es basiert auf der Überlegung, dass Quantisierungsverluste im Rahmen einer Bildkompression im Regelfall zur Verringerung der Anzahl an Grauwerten, die in örtlich begrenzten Bildbereichen auftreten, führen. Von besonderer Bedeutung sind bezüglich dieses Effekts rampenförmige Intensitätsübergänge und kontrastschwache Texturen. Artefaktreduktionsverfahren können verlorene Grauwerte unter günstigen Umständen teilweise wiederherstellen. Insbesondere trifft dies auf rampenförmige Intensitätsübergänge zu. Das Grauwertmaß sucht nach Bildbereichen, in denen die Grauwertanzahl durch Bildkompression verringert und durch Anwendung einer Artefaktreduktion wieder erhöht wurde. Die Größe der einzelnen sich nicht überlappenden Bildbereiche beträgt hier eine Blockgröße. Variationen sind diesbezüglich prinzipiell möglich. Wird ein Bereich gefunden, der die Kriterien erfüllt, wird die Anzahl der neu erzeugten Grauwerte akkumuliert. Das Grauwertmaß wird schließlich auf die Zahl der untersuchten Bildbereiche - also auf die Blockanzahl - normiert. Dadurch wird das Maß sowohl durch die Anzahl der neu erzeugten Grauwerte als auch durch die Anzahl der Bildbereiche mit neu erzeugten Grauwerten beeinflusst, wodurch eine gute Anpassung an subjektive Eindrücke gelingt. Es ist wichtig zu erwähnen, dass Konturen und ihre direkte Nachbarschaft aus der Analyse ausgeschlossen werden, damit Konturverschleifungen nicht positiv durch das Maß ausgelegt werden. Zu diesem Zweck wird eine Kantendetektion und eine anschließende Dilatation des Kantenbildes eingesetzt. Auch das Grauwertmaß weist hohe Korrelationen mit den subjektiven Eindrücken bzgl. der Gesamtqualitätsunterschiede und der Reduktion von Blockartefakten auf. Die Korrelation mit den Unschärfe-Eindrücken ist etwas geringer.

### Differenzvergrößerungsmaß

Das Differenzvergrößerungsmaß ist ebenfalls ein Maß für Bildqualitätsverbesserungen. Es analysiert, inwieweit eine Deblocking-Filterung die Intensitätsdifferenzen horizontal und vertikal benachbarter Bildpunkte vergrößert. Es werden nur Bildbereiche analysiert, die im codierten Bild homogen sind. Auf diese Weise soll festgestellt werden, ob eine Deblocking-Filterung verlorengegangene Aktivität, insbesondere z. B. rampenförmige Grauwertübergänge, wiederherstellen kann. Bezüglich der zu erfassenden Merkmale bestehen also einige Parallelen zum Grauwertmaß. Blockgrenzen werden grundsätzlich nicht betrachtet, da eine Deblocking-Filterung dort in aller Regel keine Intensitätsdifferenzvergrößerung hervorrufen wird. Wird eine Intensitätsdifferenzvergrößerung im gefilterten Bild verglichen mit dem codierten Bild festgestellt, wird die Vergrößerung akkumuliert, wenn die Intensitätsdifferenz nicht größer als die entsprechende Intensitätsdifferenz im Originalbild ist. Die Erzeugung ungewünschter Aktivität in ursprünglich homogenen Bildbereichen wird somit nicht positiv bewertet. Das Differenzvergrößerungsmaß wird so normiert, dass

sich sowohl die Größe als auch die Häufigkeit von Intensitätsdifferenzvergrößerungen auf den Wert des Maßes auswirken. Es korreliert stark mit den im Rahmen der durchgeführten subjektiven Tests ermittelten Eindrücken der Probanden bzgl. der Gesamtqualitätsunterschiede und der Reduktion von Blockartefakten und schwächer mit den Unschärfe-Eindrücken.

### Differenzverkleinerungsmaß

Das Differenzverkleinerungsmaß basiert auf dem Vergleich der Intensitäten horizontal und vertikal benachbarter Pixel im codierten und im gefilterten Bild und hat das Ziel, durch die Deblocking-Filterung verursachte Degradationen zu erfassen. Zunächst werden im codierten Bild Stellen gesucht, an denen die Intensitätsdifferenz benachbarter Pixel über einem kleinen Schwellwert liegt. Es wird davon ausgegangen, dass bei sehr kleinen Intensitätsdifferenzen keine Degradation sichtbar ist. Dann wird geprüft, ob die entsprechende Intensitätsdifferenz im gefilterten Bild kleiner ist. Ist dies der Fall, liegt ein Hinweis auf eine Degradationswirkung der Deblocking-Filterung bzw. auf Informationsverlust vor und die Differenzverkleinerung wird akkumuliert. Blockgrenzen werden grundsätzlich aus der Analyse ausgeschlossen, da dort durch die gewollte Glättung von Blockartefakten mit Differenzverkleinerungen gerechnet werden muss, die zur Verbesserung der Bildqualität beitragen und deshalb nicht negativ bewertet werden dürfen. Sowohl die Häufigkeit von Degradationen als auch deren Stärke wirken sich auf den Wert des Differenzverkleinerungsmaßes aus. Es korreliert hoch mit den in subjektiven Tests ermittelten Unschärfe-Eindrücken der Probanden und deutlich niedriger mit den übrigen abgefragten Kriterien.

## 8.4 Kombination von Qualitätsmaßen

Auf Basis theoretischer Überlegungen und der Durchführung umfangreicher Regressionsanalysen wurde nach geeigneten Kombinationsmöglichkeiten der bereits vorgestellten Maße zu einem Gesamtqualitätsmaß gesucht. Ziel war eine möglichst gute und genaue Schätzung der Wirksamkeit von Deblocking-Verfahren bezogen auf eine Verbesserung der Bildqualität in Kombination mit einer hohen Stabilität der Schätzung in einem breiten Spektrum an Bildmaterial. Das heißt, dass auch eine gute Schätzung in Bildern und bei Verwendung von Verfahren, die nicht in den ersten beiden subjektiven Tests dieser Arbeit eingesetzt wurden und somit nicht zur Berechnung von Regressionskoeffizienten beitragen, angestrebt wurde. Zu diesem Zweck wurde während der Entwicklungsphase fortlaufend die Plausibilität der Schätzungen potenzieller Modellvarianten an verschiedensten Bildern überprüft. Dazu gehörten auch besonders kritische Bilder, z. B. solche, die mit starken statischen Tiefpassfiltern bearbeitet wurden und somit eine deutliche Unschärfe aufwiesen, oder mit sehr hohen Datenraten codierte Bilder, die nur sehr wenige Artefakte aufwiesen.

Solche Bilder waren deshalb als kritisch zu betrachten, da bei einzelnen Qualitätsmaßen mit außerordentlich hohen oder niedrigen Werten gerechnet werden musste.

Es erschien grundsätzlich sinnvoll, mindestens ein Maß, das eine Bildqualitätsverbesserung erfassen kann (z. B. Diskontinuitätsmaß) mit mindestens einem Maß, das eine Bildqualitätsverschlechterung erfassen kann (z. B. Degradationsmaß), zu kombinieren. Dadurch sollte auch eine korrekte Beurteilung von Deblocking-Verfahren, die zwar einerseits störende Merkmale von Blockartefakten beseitigen oder verringern, andererseits aber gleichzeitig andere Störungen hervorrufen, sichergestellt werden. Die Erkenntnis aus den durchgeführten subjektiven Tests, dass die Wirksamkeit von Verfahren, die besonders starke Unschärfe verursachen, bezogen auf eine Bildqualitätsverbesserung von menschlichen Betrachtern stets geringer eingeschätzt wird als die Wirksamkeit der Verfahren speziell bezogen auf eine Verringerung von Blockartefakten, stützt diese Forderung, auch wenn sich die Unschärfeigung einzelner Verfahren insgesamt nicht stark negativ auf die Bewertungen durch die Betrachter ausgewirkt hat.

Während der Untersuchungen zeigte sich, dass viele Kombinationen von Maßen, die theoretisch sinnvoll erscheinen, in einem Regressionsmodell zu Problemen führen. Ein bedeutender Grund dafür sind Korrelationen zwischen den einzelnen Maßen. Gerade wenn - wie hier - Maße verwendet werden sollen, die zwar einzelne Merkmale von Störungen untersuchen, diese Merkmale aber fast ausschließlich gemeinsam auftreten und von Artefaktreduktionsverfahren auch gemeinsam beeinflusst werden, sind solche Korrelationen unvermeidbar. Korrelationen bzw. Multikollinearitäten zwischen Maßen können einerseits dazu führen, dass die Kombination der Maße eine Zielgröße - in diesem Falle die Wirksamkeit von Deblocking-Verfahren - kaum besser erklärt als die einzelnen Maße für sich betrachtet. In solchen Fällen kann für einzelne Maße in der Kombination keine statistische Signifikanz nachgewiesen werden. Andererseits können besonders starke Multikollinearitäten auch die Ergebnisse von Regressionsanalysen verzerren. Weitere Probleme können bei der Kombination von Maßen entstehen, wenn sich herausstellt, dass in der Kombination einzelne Voraussetzungen der linearen Regression verletzt werden (z. B. Unkorreliertheit der Störvariablen). Dann ist entweder die Durchführung einer linearen Regression nicht mehr sinnvoll oder es muss zumindest mit einer Unschärfe der Ergebnisse gerechnet werden.

Leider lässt sich während der Entwicklung von Qualitätsmaßen nicht vorhersagen, ob diese gut mit anderen Maßen kombinierbar sind oder ob erhebliche statistische Probleme dabei auftreten werden. Dies kann erst am Ende der Entwicklungsphase überprüft werden und die Notwendigkeit von erheblichen Veränderungen oder Überarbeitungen ergeben. Im Folgenden werden mehrere Kombinationen der bereits vorgestellten Qualitätsmaße beschrieben, die eine gute Schätzung der Wirksamkeit von Deblocking-Verfahren ermöglichen und keine kritischen Probleme im Rahmen der linearen Regression verursachen.

Es sei darauf hingewiesen, dass in den folgenden Untersuchungen und in den Auswertungen ein einzelnes Bildpaar des ersten subjektiven Tests nicht berücksichtigt wurde. Es ist das Bildpaar Nummer 24 (siehe Tabelle 7.2) und wurde deshalb ausgeschlossen, weil es

während der durchgeführten Regressionsanalysen immer wieder als Ausreisser verdächtigt wurde. Bei Überprüfung der von den Probanden des subjektiven Tests abgegebenen Bewertungen bzgl. des Gesamtqualitätsunterschieds zwischen den Bildern des Paares fiel auf, dass einige negative Bewertungen abgegeben wurden, was auf eine Verwechslung des gefilterten und ungefilterten Bildes hindeutet. Die Kriterien für eine relativ sichere Identifizierung von einzelnen Vertauschungen waren jedoch nicht erfüllt und so wurde nicht zurückgetauscht, was insgesamt zu einem sehr niedrigen Mittelwert der abgegebenen Bewertungen führte. Die Wirksamkeit des entsprechenden Deblocking-Verfahrens wird also mit hoher Wahrscheinlichkeit nicht korrekt beschrieben. Ein Ausschluss des Datensatzes aus den weiteren Untersuchungen erschien daher sinnvoll.

### 8.4.1 Variante 1

Zunächst soll eine Modellvariante aus der Kombination dreier Maße vorgestellt werden, welche auf theoretischen Überlegungen zum Aufbau eines Gesamtqualitätsmaßes basiert und sich mit Hilfe der Regressionsanalyse gut mit einem statistischen Modell in Einklang bringen lässt. Es erscheint sinnvoll, ein oder mehrere Maße, die Bildqualitätsverbesserungen durch Deblocking-Algorithmen erfassen können, mit einem oder mehreren Degradationsmaßen, die auf unerwünschte Nebeneffekte einer Blocking-Reduktion reagieren, zu kombinieren. Eine sinnvolle Modellvariante, die den Vorüberlegungen genügt, ist die Kombination des Diskontinuitätsmaßes mit dem Glättungsmaß und dem Konturdegradationsmaß. Tabelle 8.1 enthält die mit der Methode der kleinsten Quadrate für das Modell ermittelten Regressionskoeffizienten. Zur Berechnung der Koeffizienten wurden - wie auch bei den im Folgenden beschriebenen Modellvarianten - ausschließlich die Bewertungen der Wirksamkeit von Artefaktreduktionsverfahren bezogen auf eine Gesamtbildqualitätsverbesserung aus dem ersten subjektiven Test herangezogen. Es wurde eine Regression durch den Ursprung erzwungen ( $b_0 = 0$ ), da es sinnvoll erschien, dass das Modell beim Vergleich zweier identischer Bilder keinerlei Qualitätsveränderung anzeigt. Ohnehin ist der Koeffizient  $b_0$  in dieser Modellvariante im Gegensatz zu den in Tabelle 8.1 aufgeführten Koeffizienten nicht signifikant.

**Tabelle 8.1:** Regressionskoeffizienten der Modellvariante 1

Maß	Koeffizient
Diskontinuitätsmaß (Einflussgröße 1)	0,6272
Glättungsmaß (Einflussgröße 2)	2.3319
Konturdegradationsmaß (Einflussgröße 3)	-0.8854

Dem primären Zweck - die Messung von Bildqualitätsverbesserungen - entsprechend sind die Regressionskoeffizienten der ersten beiden Maße positiv. Das Konturdegradati-

onsmaß hingegen hat einen negativen Koeffizienten. Die Vorzeichen der Koeffizienten verhalten sich also plausibel. Degradationen von Konturen wirken sich verringernd auf das Gesamtmaß aus.

Die statistische Analyse (siehe dazu auch Kapitel 3) der vorgestellten Modellvariante ergab keine Auffälligkeiten, die auf eine Verletzung der Voraussetzungen der linearen Regression oder einen generell ungeeigneten Modellansatz hindeuten. Die Unkorreliertheit der Störvariablen konnte mit dem Durbin-Watson-Test nachgewiesen werden. Es ist auch kein Trend der Residuen im Versuchsverlauf erkennbar (siehe Abbildung 8.9 oben links). Die Homoskedastizität wurde mit Hilfe des Streudiagramms in Abbildung 8.9 oben rechts überprüft. Da kein regelmäßiges Muster erkennbar ist, kann von Homoskedastizität ausgegangen werden. Abbildung 8.9 enthält auch Streudiagramme der Residuen über den einzelnen Einflussgrößen. Diese lassen ebenfalls keine regelmäßigen Muster erkennen. Daraus lässt sich schließen, dass die Wahl eines linearen Modells gerechtfertigt ist und keine Fehlspezifikation vorliegt.

Die Modellvariante wurde mit Hilfe der Durchführung von mehreren multiplen Regressionen jeweils einer der Einflussgrößen auf die übrigen Einflussgrößen auf Multikollinearität geprüft. Zwar besteht hier schon allein durch das ähnliche Verhalten des Diskontinuitäts- und des Glättungsmaßes ein gewisser Grad an Kollinearität, ein problematisches Ausmaß konnte jedoch nicht nachgewiesen werden. Schließlich wurde geprüft, ob die Residuen normalverteilt sind. Im Lilliefors-Test wurde die Hypothese, dass eine Normalverteilung vorliegt, nicht abgelehnt und auch der QQ-Plot in Abbildung 8.10 lässt trotz zweier auffälliger Werte darauf schließen, dass die Annahme einer Normalverteilung gerechtfertigt ist. Aus statistischer Sicht bestehen also keine Vorbehalte gegenüber der Modellvariante 1.

Abbildung 8.9 unten rechts zeigt das Streudiagramm der Zielgröße  $y$  über der geschätzten Zielgröße  $y_e = \hat{y}$ . Es ist erkennbar, dass die Punkte um eine gedachte Gerade herum angeordnet sind und insgesamt nur recht kleine Abweichungen auftreten. Die beiden auffälligen Punkte am oberen Ende des Diagramms entsprechen den Bildpaaren 18 und 28 im ersten subjektiven Test. Bei der Auswertung der Ergebnisse des subjektiven Tests wurde bereits darauf hingewiesen, dass die Bewertung der Qualitätsverbesserung in diesen beiden Fällen auffällig hoch ausfällt. Belastbare Anhaltspunkte für eine Behandlung der Bewertungen als Ausreißer liegen jedoch nicht vor. Die in Modellvariante 1 betrachteten Störmerkmale ermöglichen offensichtlich keine optimale Erfassung der überdurchschnittlich hohen Qualitätsverbesserungen in den beiden Fällen. Insofern fallen die Schätzungen der Wirksamkeit der Artefaktreduktion geringer aus als die subjektiven Eindrücke der Probanden. Die entsprechenden Residuen sind daher vergleichsweise groß. Da auffällige Abweichungen zwischen den subjektiven Eindrücken und den Schätzwerten offenbar recht selten auftreten und die Abweichungen absolut betrachtet nicht zu groß erscheinen, können sie jedoch akzeptiert werden.

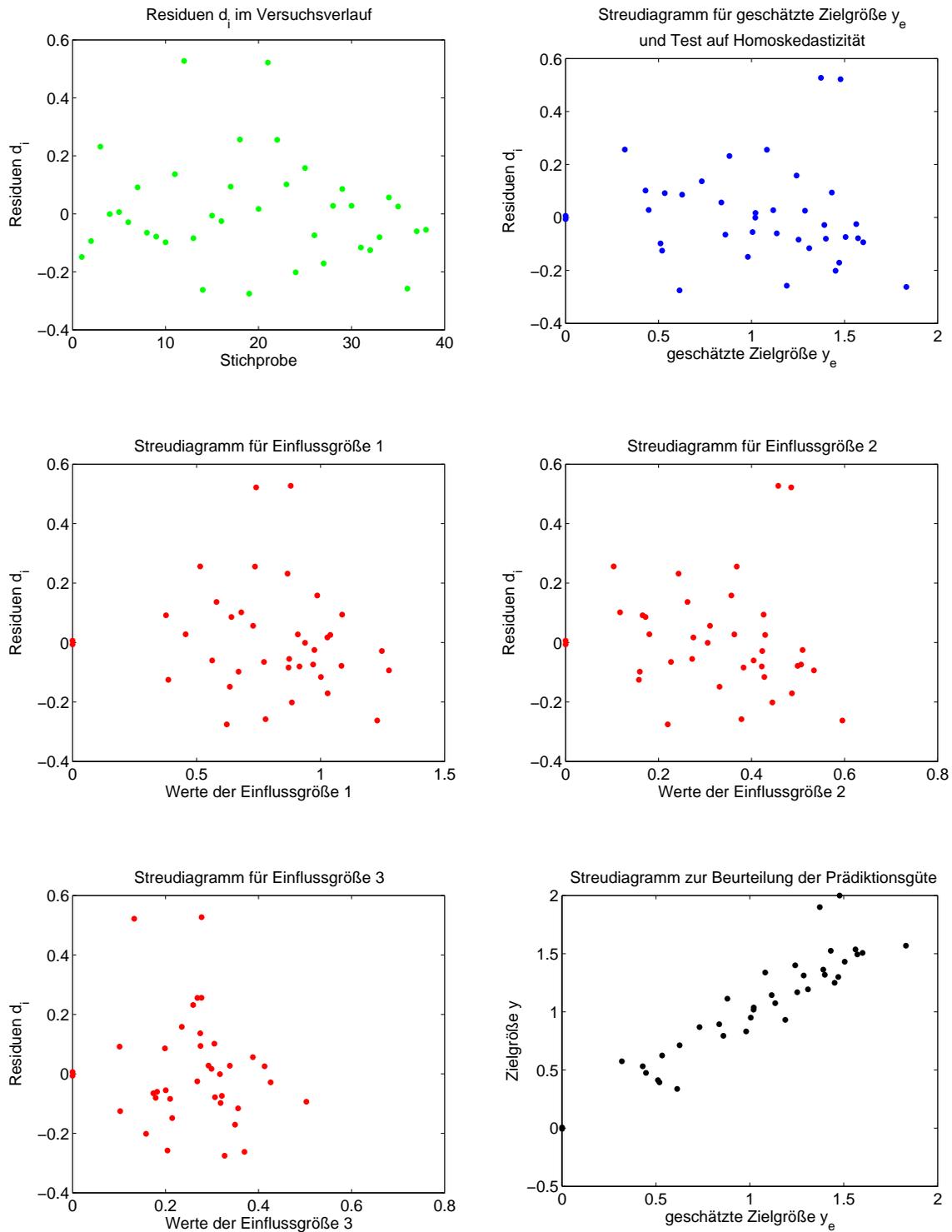


Abbildung 8.9: Streudiagramme zur statistischen Analyse der Modellvariante 1

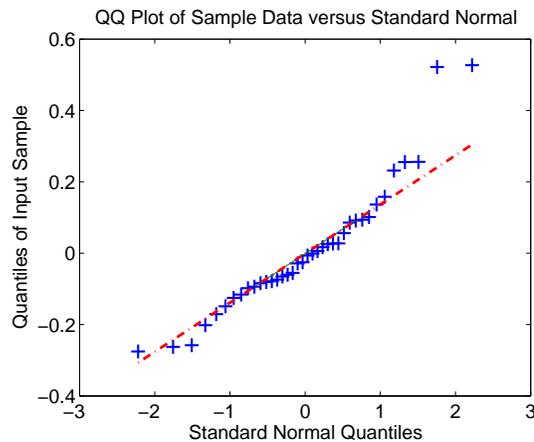


Abbildung 8.10: QQ-Plot zur statistischen Analyse der Modellvariante 1

### 8.4.2 Variante 2

Die im Folgenden vorgestellte Modellvariante 2 basiert ebenfalls auf der Überlegung, dass es sinnvoll ist, eine Kombination von Maßen für eine Qualitätsverbesserung mit Degradationsmaßen vorzunehmen. Auch hier zeigt die Regressionsanalyse, dass sich die Variante gut mit einem statistischen Modell in Einklang bringen lässt. Die eingesetzten Maße sind das Diskontinuitätsmaß, das Glättungsmaß, das Degradationsmaß und die Wechselwirkung aus Diskontinuitäts- und Glättungsmaß. Tabelle 8.2 enthält die mit der Methode der kleinsten Quadrate ermittelten Regressionskoeffizienten für die Modellvariante 2. Auch hier wurde aus dem bereits bei Variante 1 erwähnten Grund eine Regression durch den Ursprung ( $b_0 = 0$ ) erzwungen. Es sei angemerkt, dass auch hier der Koeffizient  $b_0$  im Gegensatz zu den in der Tabelle aufgeführten Koeffizienten insignifikant ist.

Tabelle 8.2: Regressionskoeffizienten der Modellvariante 2

Maß	Koeffizient
Diskontinuitätsmaß (Einflussgröße 1)	0,8298
Glättungsmaß (Einflussgröße 2)	3,4238
Degradationsmaß (Einflussgröße 3)	-0,8807
Wechselwirkung (Einflussgröße 4)	-0.9313

Die Koeffizienten für das Diskontinuitäts- und für das Glättungsmaß sind positiv, was plausibel ist, da diese Maße der Bewertung von Merkmalen, die mit einer Bildqualitätsverbesserung einhergehen, dienen. Die Koeffizienten für die beiden anderen Maße sind negativ. Bezüglich des Degradationsmaßes ist die Plausibilität offensichtlich, aber auch bezüglich

der Wechselwirkung erscheint ein negativer Koeffizient nicht unsinnig, da die Wechselwirkung auf diese Weise kompensierend zum Gesamtqualitätsmaß beitragen kann. Die Kompensation kommt besonders dann zum Tragen, wenn beide Qualitätsverbesserungsmaße gleichzeitig hohe Werte annehmen, was sich offenbar in vielen Fällen zu stark vergrößernd auf das Gesamtqualitätsmaß auswirkt. Die Wechselwirkung, versehen mit einem negativen Regressionskoeffizienten, verringert das Gesamtqualitätsmaß in solchen Fällen wieder.

Die statistische Analyse der Modellvariante 2 ergab zunächst keine kritischen Auffälligkeiten, die auf eine Verletzung der Voraussetzungen der linearen Regression hindeuten oder den Modellansatz als ungeeignet erscheinen lassen. Der Durbin-Watson-Test lieferte keinen Verdacht auf Korreliertheit der Störvariablen. Auch ein Trend der Residuen im Versuchsverlauf ist nicht erkennbar (siehe Abbildung 8.11 oben links). Das Streudiagramm in Abbildung 8.11 oben rechts liefert keinen Hinweis auf Homoskedastizität. Da auch in den übrigen Streudiagrammen keine auffälligen Muster erkennbar sind, kann zunächst davon ausgegangen werden, dass die Annahme eines linearen Modells zulässig ist und keine Fehlspezifikation vorliegt. Es wurden jedoch geringe Auffälligkeiten in dem in Abbildung 8.12 links dargestellten Diagramm gefunden, in dem die Zielgröße über den Werten der Wechselwirkung (Einflussgröße 4) aufgetragen ist. Es besteht der Verdacht, dass die Punkte um eine gekrümmte Linie herum angeordnet sind, so dass die Beschreibung des Zusammenhangs der dargestellten Größen durch eine nichtlineare Funktion angeraten erscheint. Es wurde somit die Verwendung der Wurzel aus der Wechselwirkung als vierte Einflussgröße untersucht. Dadurch konnte zwar eine etwas bessere Anpassung zwischen Zielgröße und Einflussgröße 4 erreicht werden, es entstanden aber so große Multikollinearitätsprobleme, dass diese Variante nicht eingesetzt werden konnte. Der Einsatz einer Zentrierung zur Verringerung der Multikollinearitätsprobleme war aus arithmetischen Gründen (Wurzeln aus negativen Zahlen innerhalb eines reellen Modells) nicht möglich. Ein weiteres Problem bei der Verwendung der Wurzel aus der Wechselwirkung bestand darin, dass für einzelne Maße sehr große Koeffizienten berechnet wurden, die zu einer Unausgewogenheit der Gewichtung der Maße führte. Es bestand die Gefahr des Auftretens von erheblichen Instabilitätsproblemen in einzelnen Bildern. Letztlich wurde an der ursprünglichen Variante, also der Verwendung der Wechselwirkung zwischen Diskontinuitäts- und Glättungsmaß, festgehalten. Da die in Abbildung 8.12 links zu vermutende Krümmung insgesamt nicht besonders ausgeprägt ist, kann ohnehin durchaus ein approximativ linearer Zusammenhang zwischen Einflussgröße 4 und Zielgröße angenommen werden, was durch eine hohe ermittelte Korrelation zwischen den Größen gestützt wird.

Die Durchführung von mehreren multiplen Regressionen jeweils einer der Einflussgrößen auf die übrigen Einflussgrößen deckte Multikollinearitätsprobleme auf. Verantwortlich dafür ist die Wechselwirkung zwischen Diskontinuitäts- und Glättungsmaß, da ein Herauslassen der Wechselwirkung die Multikollinearitätsanzeichen deutlich verkleinert. Es wurde eine Zentrierung von Diskontinuitäts- und Glättungsmaß in Erwägung gezogen, jedoch verlor dabei die Wechselwirkung ihre Signifikanz, während der Koeffizient  $b_0$  hoch

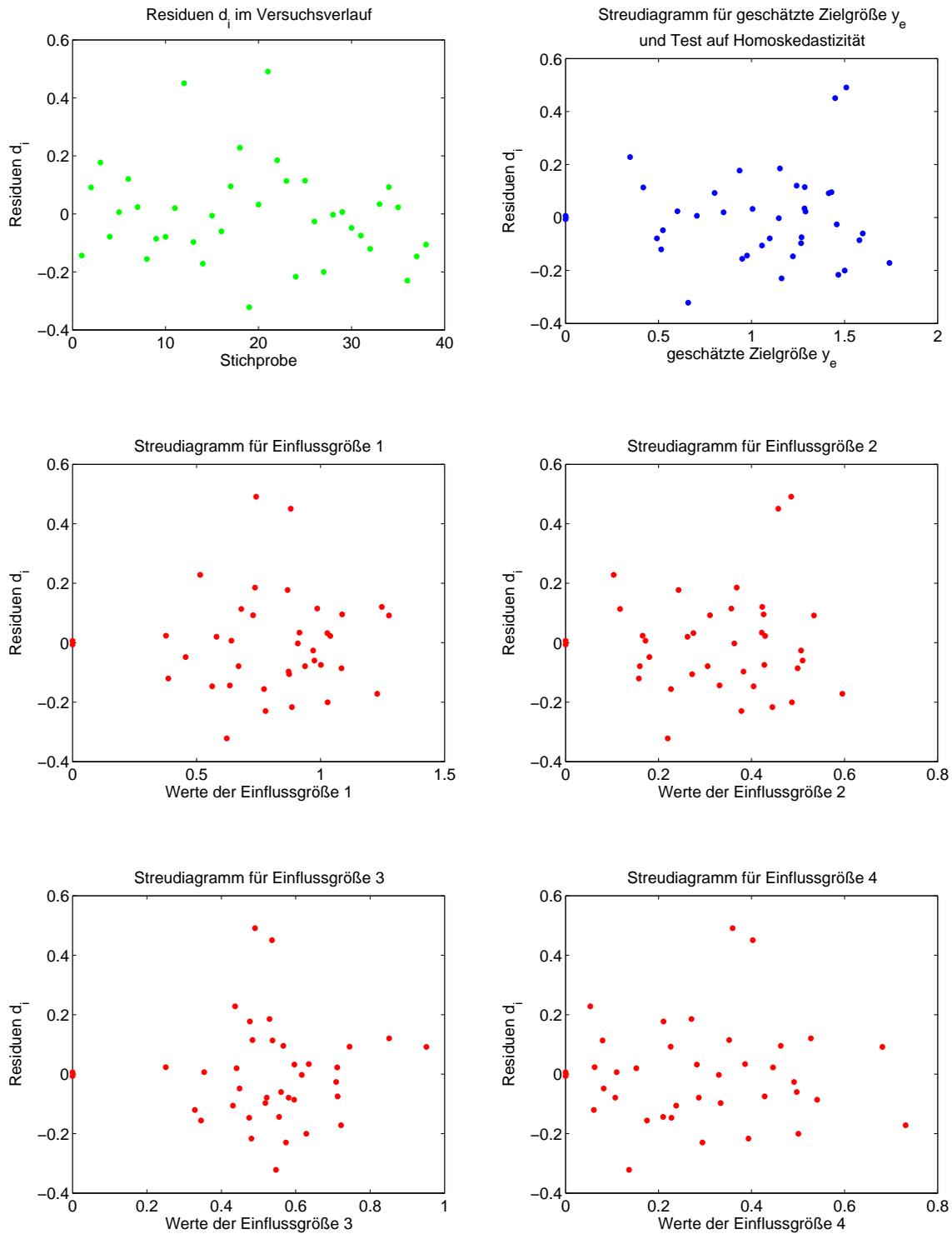
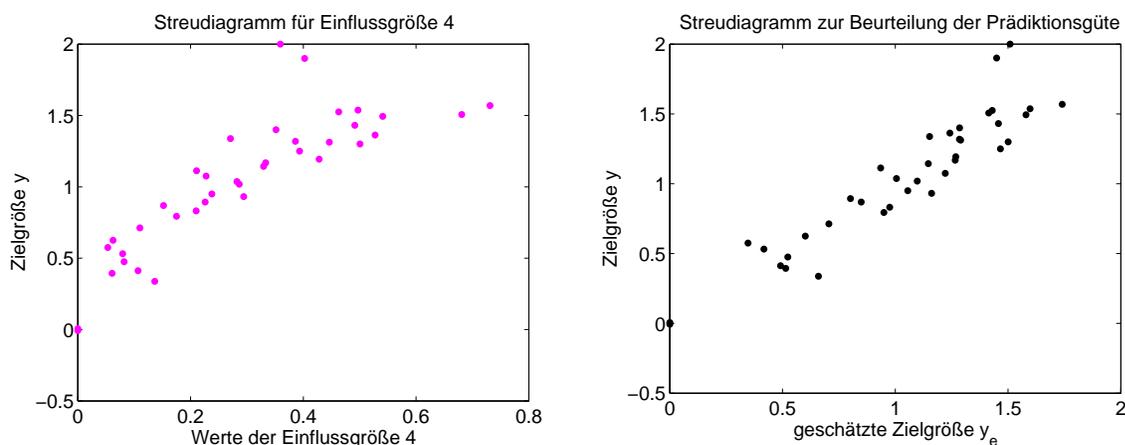


Abbildung 8.11: Streudiagramme zur statistischen Analyse der Modellvariante 2



**Abbildung 8.12:** Streudiagramm der Zielgröße über Einflussgröße 4 und Streudiagramm der Zielgröße über der geschätzten Zielgröße, Modellvariante 2

signifikant wurde. Eine Regression durch den Ursprung verlor somit jeglichen Sinn. Die entstehenden Probleme sprachen letztlich gegen eine Zentrierung. Die Multikollinearität muss daher in Kauf genommen werden. Die in Tabelle 8.2 angegebenen Koeffizienten sind somit wahrscheinlich mit Unschärfen versehen. Beim Herauslassen einzelner Bildpaare aus dem Datensatz und dem testweisen Berechnen neuer Koeffizienten traten tolerable Schwankungen auf, so dass von moderaten Unschärfen ausgegangen werden kann. Schließlich wurde geprüft, ob die Residuen in Modellvariante 2 normalverteilt sind. Im Lilliefors-Test wurde die Hypothese, dass eine Normalverteilung vorliegt, nicht abgelehnt und auch der QQ-Plot in Abbildung 8.13 lässt nicht darauf schließen, dass die Annahme einer Normalverteilung ungerechtfertigt ist.

Zusammenfassend kann zur Modellvariante 2 gesagt werden, dass aus statistischer Sicht zwei Auffälligkeiten bestehen. Es gibt einen Hinweis auf einen nichtlinearen Zusammenhang zwischen der Zielgröße und einer der Einflussgrößen. Wie bereits erwähnt, kann jedoch approximativ von einem linearen Zusammenhang ausgegangen werden, so dass sich hieraus kein Problem ergibt. Nicht zu leugnen ist die Existenz von Multikollinearität. Für die Regressionsschätzung gilt jedoch weiterhin die BLUE-Eigenschaft. Es muss nur mit Unschärfen bezüglich der Werte der ermittelten Koeffizienten gerechnet werden. Abbildung 8.12 rechts zeigt das Streudiagramm der Zielgröße  $y$  über der geschätzten Zielgröße  $y_e$ . Es ist erkennbar, dass die Punkte um eine gedachte Gerade herum angeordnet sind und insgesamt nur recht kleine Abweichungen auftreten, die im Durchschnitt minimal geringer ausfallen als in Modellvariante 1. Ein Nichtlinearitätsproblem ist nicht erkennbar. Die beiden auffälligen Punkte am oberen Ende des Diagramms entsprechen auch hier den Bildpaaren 18 und 28 im subjektiven Test.

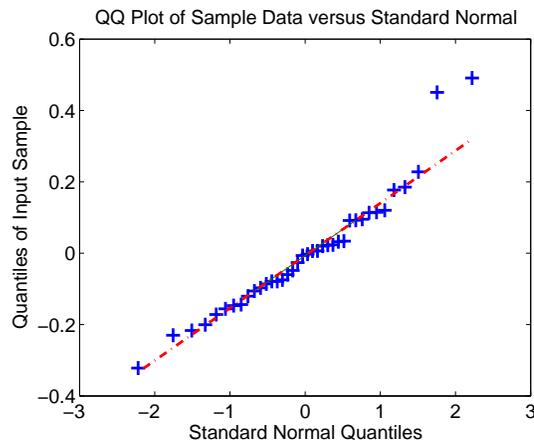


Abbildung 8.13: QQ-Plot zur statistischen Analyse der Modellvariante 2

### 8.4.3 Variante 3

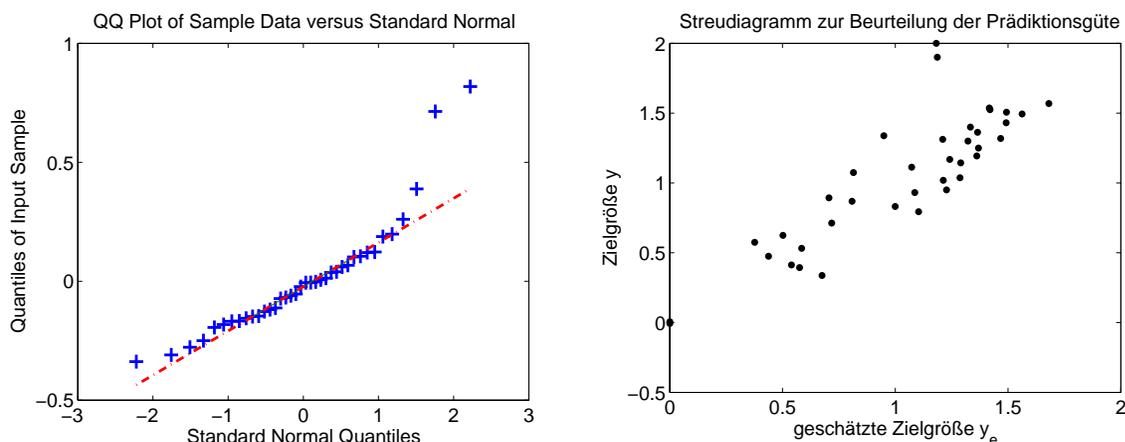
Die Modellvariante 3 ist nicht als erste Wahl zu betrachten und soll hauptsächlich zur Demonstration der möglichen Gewinne durch eine Kombination mehrerer Maße über einen Regressionsansatz dienen, wenn die einzelnen Maße nicht über einen derart hohen Korrelationsgrad verfügen wie das vorgestellte Glättungsmaß (siehe dazu auch Kapitel 9). Dazu wurde das Glättungsmaß ausgeschlossen und eine Kombination aus Diskontinuitätsmaß, Differenzvergrößerungsmaß und Konturdegradationsmaß untersucht. Tabelle 8.3 enthält die mit der Methode der kleinsten Quadrate ermittelten signifikanten Regressionskoeffizienten für die Modellvariante 3. Es wurde wieder eine Regression durch den Ursprung erzwungen und somit der auch hier ohnehin insignifikante Koeffizient  $b_0$  ausgeschlossen.

Tabelle 8.3: Regressionskoeffizienten der Modellvariante 3

Maß	Koeffizient
Diskontinuitätsmaß (Einflussgröße 1)	1,2872
Differenzvergrößerungsmaß (Einflussgröße 2)	2,4232
Konturdegradationsmaß (Einflussgröße 3)	-2,0131

Die Vorzeichen der Regressionskoeffizienten ergeben ein plausibles Modell, da nur der Koeffizient des Konturdegradationsmaßes ein negatives Vorzeichen hat. Die statistische Analyse des Modells wurde mit der Suche nach Trends im Versuchsverlauf und dem Durbin-Watson-Test begonnen. Es ergaben sich dabei keine Auffälligkeiten, so dass von einer Unkorreliertheit der Störvariablen ausgegangen werden kann. Das Streudiagramm der Residuen über der geschätzten Zielgröße lieferte keine Hinweise, die gegen Homoske-

distizität sprechen. Die Analyse von Streudiagrammen der Residuen über den einzelnen Einflussgrößen ergab schließlich ebenfalls keine Auffälligkeiten, so dass davon ausgegangen werden kann, dass die Annahme eines linearen Modells gerechtfertigt ist und keine Fehlspezifikation vorliegt. Auch Modellvariante 3 wurde mit Hilfe der Durchführung von mehreren multiplen Regressionen jeweils einer der Einflussgrößen auf die übrigen Einflussgrößen auf Multikollinearität geprüft. Ein problematisches Ausmaß an Multikollinearität konnte nicht gefunden werden, auch wenn eine verhältnismäßig hohe Korrelation zwischen Diskontinuitäts- und Differenzvergrößerungsmaß besteht. Als problembehaftet erwies sich schließlich die Analyse der Verteilung der Residuen. Im Lilliefors-Test wurde die Hypothese, dass die Residuen normalverteilt sind, abgelehnt. Auch der QQ-Plot in Abbildung 8.14 deutet auf Abweichungen von einer Normalverteilung hin. Es konnte nachgewiesen werden, dass die Probleme durch die bereits mehrfach auffällig gewordenen Bildpaare 18 und 28 verursacht werden. Der Ausschluss der entsprechenden Testdaten beseitigt jeden Verdacht auf nicht normalverteilte Residuen. Modellvariante 3 beschreibt offenbar die beiden kritischen Datensätze besonders ungenau, so dass sie als Ausreißer erscheinen. Darauf deutet auch das Streudiagramm der Zielgröße  $y$  über der geschätzten Zielgröße  $y_e$  in Abbildung 8.14 hin, wo die beiden kritischen Punkte am oberen Rand des Diagramms zu erkennen sind. Die übrigen Punkte sind um eine gedachte Gerade herum angeordnet. Die Abweichungen sind insgesamt etwas größer als bei den bereits vorgestellten Modellvarianten. Ein Nichtlinearitätsproblem ist nicht erkennbar.



**Abbildung 8.14:** QQ-Plot und Streudiagramm von Zielgröße über geschätzter Zielgröße, Modellvariante 3

Da, wie bereits erwähnt, keine Grundlage gegeben ist, die Bewertungen der Bildpaare 18 und 28 als Ausreißer zu behandeln, können diese aus den Analysen nicht ausgeschlossen werden. Es besteht dann jedoch der Verdacht, dass die Residuen nicht gleichverteilt sind, was zu Unschärfen bei den ermittelten Regressionskoeffizienten führen kann. Der Nach-

weis, dass die Normalverteilungsannahme nur durch zwei kritische Werte verletzt wird, berechtigt zu der Annahme, dass das Unschärfeproblem nicht überbewertet werden sollte. Aus demselben Grund ist auch von gültigen Ergebnissen der durchgeführten statistischen Analysen auszugehen.

#### 8.4.4 Variante 4

Abschließend soll der Vollständigkeit halber eine Kombination von Maßen vorgestellt werden, die das Ergebnis einer schrittweisen Regression ausgehend von allen Maßen darstellt. Es bleiben fünf Qualitätsmaße übrig, deren zugehörige Regressionskoeffizienten signifikant sind. Es sind das Glättungsmaß, das Grauwertmaß, das Differenzvergrößerungsmaß, das Konturdegradationsmaß und das Differenzverkleinerungsmaß. Die Modellvariante lässt sich jedoch nicht plausibel mit einem theoretischen Modell in Einklang bringen, da die Regressionskoeffizienten teilweise Vorzeichen aufweisen, die nicht dem primären Zweck des zugehörigen Maßes entsprechen. So hat das Grauwertmaß einen negativen Regressionskoeffizienten und das Differenzverkleinerungsmaß einen positiven Regressionskoeffizienten, obwohl das erstgenannte Maß primär Bildqualitätsverbesserungen und das letztgenannte Maß primär Bildqualitätsverschlechterungen erfassen soll. Hinzu kommt, dass die statistische Analyse der Modellvariante erhebliche Zweifel an einer Normalverteilung der Störgrößen aufkommen lässt.

Die genannten Auffälligkeiten und Probleme bestätigen, dass die Anwendung einer schrittweisen Regression allenfalls unterstützend in explorativen Phasen von Forschungsaufgaben eingesetzt werden sollte (siehe auch [UM06]). Das automatische Finden eines sinnvollen Regressionsmodells ohne theoretische Vorüberlegungen ermöglicht die schrittweise Regression nicht oder nur in seltenen Fällen. Sie ermöglicht jedoch eine sehr gute Anpassung von Schätzern an einzelne Datensätze, was sich hier durch eine sehr hohe Korrelation der geschätzten Wirksamkeit von Deblocking-Verfahren bezogen auf eine Bildqualitätsverbesserung mit den Ergebnissen des ersten subjektiven Tests äußert (siehe Kapitel 9). Bei Verwendung anderen Bildmaterials muss mit deutlichen Fehlschätzungen gerechnet werden, was an späterer Stelle noch gezeigt wird. Insgesamt sind die Auffälligkeiten bei der Analyse von Modellvariante 4 so groß, dass keine Empfehlung zur Nutzung dieser Kombination von Maßen ausgesprochen werden kann. Daher wird die Modellvariante auch nicht im Detail vorgestellt. Sie soll nur stellenweise zu Vergleichszwecken in den weiteren Untersuchungen dienen.

#### 8.4.5 Anmerkungen zur Entwicklung und Kombination von Qualitätsmaßen

Die Kombination mehrerer Qualitätsmaße zu einem Gesamtqualitätsmaß mit dem Ziel, die Schätzgüte oder andere Eigenschaften zu verbessern, ist ohne Zweifel sinnvoll und führt

in vielen Fällen zum Erfolg, wie in dieser Arbeit und auch in den Untersuchungen anderer Autoren gezeigt wird (siehe Kapitel 4). Die lineare Regressionsanalyse ist in diesem Zusammenhang ein wichtiges Werkzeug. Allerdings lassen sich nicht beliebige Maße sinnvoll miteinander kombinieren, was sich in den durchgeführten Untersuchungen daran zeigte, dass sich aus der Vielzahl an implementierten Maßen nur wenige Kombinationen finden ließen, die keine nicht zu tolerierenden statistischen Probleme aufwiesen. Letztlich ist es bei der Entwicklung von Maßen nicht abschätzbar, inwieweit diese mit anderen Maßen kombinierbar sind. Das Optimum - sich zueinander orthogonal verhaltende Maße, d. h. Maße, die sich völlig unabhängig voneinander verhalten - kann in der Praxis nicht erzwungen werden [UM06, Tie87]. Dennoch sollen einige Erkenntnisse bzw. Beobachtungen zum Entwurf geeigneter Maße wiedergegeben werden.

Grundsätzlich bereitet die Kombination von Maßen, die sehr miteinander verwandte Merkmale analysieren, Schwierigkeiten, welche sich in Form von starken Multikollinearitäten äußern. Leider lässt sich dieses Problem bei einer Aufgabenstellung, wie sie in dieser Arbeit vorliegt, kaum wirkungsvoll verhindern. Die Merkmale spezifischer Störungen wie z. B. Blockartefakte treten in aller Regel gemeinsam auf und werden von Störungsreduktionsverfahren auch gemeinsam beeinflusst. Es kann nur versucht werden, trotzdem Merkmale zu finden, die sich nicht zu gleichartig bzw. möglichst unterschiedlich verhalten. Dazu kann die Aufnahme und Analyse von zusätzlichen Daten, die nicht nur die Gesamtwirkung von Störungen oder Störungsreduktionsverfahren betreffen, hilfreich sein. In dieser Arbeit wurde zu dem genannten Zweck der zweite subjektive Test durchgeführt. Allerdings besteht gleichzeitig die Gefahr, einzelne Maße auf Basis der zusätzlichen Daten zu sehr in eine bestimmte Richtung zu optimieren, wodurch letztlich Multikollinearitätsprobleme sogar zunehmen können.

Während der Entwicklung der in dieser Arbeit vorgestellten Maße wurde die Erfahrung gemacht, dass sich stets Schwierigkeiten ergaben, wenn versucht wurde, sowohl Qualitätsverbesserungen als auch Qualitätsverschlechterungen in einem einzigen Maß zu erfassen und gegeneinander zu verrechnen. Als Beispiel sei ein Glättungsmaß genannt, das nicht nur gewollte, sondern auch ungewollte Glättungen zu erfassen versucht und entsprechend gegeneinander abwägt. Derartige Ansätze wurden verfolgt, führten aber in der Regel zu starken Instabilitäten in bestimmten Bildern oder zu Problemen bei der Kombination mit anderen Maßen. Die Erkenntnisse reichen nicht aus, um diese Probleme zu verallgemeinern, jedoch waren die Probleme in den hier durchgeführten Untersuchungen stark ausgeprägt. Als unproblematischer hat sich der Ansatz erwiesen, Maße zu entwickeln, die entweder eine Qualitätsverbesserung oder eine Qualitätsverschlechterung erfassen, und diese in Form von geeigneten Regressionsmodellen miteinander zu kombinieren. Die Daten aus dem zweiten subjektiven Test waren hilfreich bei der Entwicklung und Optimierung der einzelnen Qualitätsmaße unter dem genannten Gesichtspunkt.

Ein Kriterium, das ebenfalls frühzeitig bei der Entwicklung von Maßen berücksichtigt werden sollte, ist die korrekte Reaktion der Maße auf unterschiedliche Ausprägungen der

zu erfassenden Merkmale bzw. ein (möglichst) monotonen Verhalten der Maße. Das bedeutet, dass z. B. die Verringerung von Störungen in wenig gestörten Bildern bzw. an wenigen Stellen zu kleineren Messwerten eines entsprechenden Maßes führen sollte als eine vergleichbare Verringerung von Störungen in deutlich gestörten Bildern bzw. an vielen Stellen. Es sollte diesbezüglich ein Einklang mit den subjektiven Eindrücken menschlicher Betrachter bestehen. Darüber hinaus sollte bei der Entwicklung von Maßen frühzeitig abgeschätzt werden, ob diese in bestimmten Situationen - hier Bildern - zu Instabilitäten neigen könnten, um entsprechend darauf reagieren zu können.

Besonders wichtig bei der Kombination von Maßen in Form von Regressionsmodellen ist die Plausibilität der Modelle. Es ist nicht sinnvoll, Maße in irgendeiner kaum zu erklärenden Form zu kombinieren, nur um eine besonders gute Anpassung der Kombination an einen speziellen Datensatz zu erreichen. Das Ziel der Kombination ist eine stabile und akkurate Qualitätsbewertung in einer breiten Palette an Bildmaterial und nicht die optimale Korrelation mit einem einzigen Datensatz. Die Plausibilität eines Regressionsmodells äußert sich insbesondere durch die Art und Anzahl der integrierten Maße unter gleichzeitiger Betrachtung der Vorzeichen und Größen der Regressionskoeffizienten. Es ist zu beachten, dass die Größen der Regressionskoeffizienten grundsätzlich im Zusammenhang mit den Wertebereichen der entsprechenden Maße zu interpretieren sind. Die schrittweise Regression dürfte in den seltensten Fällen zu plausiblen Modellen führen, was die hier durchgeführten Untersuchungen bestätigen. Auf das Thema der optimalen Modellkomplexität wird in Kapitel 9 noch eingegangen.



# 9 Schätzgüte des neuen Ansatzes und Vergleich mit anderen Verfahren

Die in Kapitel 8 vorgestellten Qualitätsmaße und Kombinationen aus Maßen sollen nun bezüglich Ihrer Schätzgenauigkeit und Stabilität bei der Bewertung der Wirksamkeit von Deblocking-Verfahren analysiert werden. Dazu dienen im Wesentlichen die Daten, die im Rahmen der in dieser Arbeit durchgeführten subjektiven Tests ermittelt wurden. In die Untersuchungen werden auch andere bekannte Qualitätsbewertungsansätze einbezogen. Darunter sind gradientenbasierte Blocking-Metriken, die in der Literatur vereinzelt zum Vergleich von Artefaktreduktionsverfahren herangezogen werden [CWQ01, RK00, Sut03], Einkanal- und Faktorenmodelle, welche primär zur Bewertung der absoluten Qualität komprimierter Bildfolgen entwickelt wurden, sowie PSNR und MSE. Die Schätzgüten werden anhand von Korrelationsgraden und mittleren Schätzfehlern beurteilt. Im Vordergrund steht die Eignung der genannten Qualitätsbewertungsansätze zur Schätzung der Wirksamkeit von Deblocking-Verfahren bzgl. der Verbesserung der Gesamtbildqualität. Es werden jedoch auch die Korrelationen der neu vorgestellten Qualitätsmaße mit den Kriterien, die im zweiten subjektiven Test dieser Arbeit bewertet wurden, betrachtet.

Insbesondere für eine Beurteilung der im Rahmen dieser Arbeit entwickelten Qualitätsbewertungsverfahren wichtig ist die Untersuchung der Schätzgüte in *unbekannten* Bildern, also in Bildern, die nicht während der Entwicklung der Verfahren verwendet wurden. Zu diesem Zweck werden einerseits die Daten aus dem dritten subjektiven Test herangezogen. Andererseits wird die Stabilität und Plausibilität der Qualitätsbewertungsverfahren in besonders kritischen Bildern untersucht. Dazu zählen z. B. Bilder, die mit starken Tiefpassfiltern bearbeitet wurden, oder Bilder, die nur sehr wenige Störungen enthalten, weil in solchen Bildern mit außerordentlich hohen oder niedrigen Werten einzelner Maße gerechnet werden muss.

## 9.1 Schätzgüte in den ersten beiden subjektiven Tests

In Tabelle 9.1 sind die Korrelationen der in Kapitel 8 vorgestellten Qualitätsmaße mit den in den ersten beiden subjektiven Tests abgegebenen Bewertungen angegeben. Es kann festgehalten werden, dass die Maße grundsätzlich ihren primären Zweck erfüllen. Die Maße, die eine Qualitätsverbesserung durch die Anwendung von Deblocking-Verfahren erfassen

und bewerten sollen, korrelieren mit Ausnahme des Glättungsmaßes und der Wechselwirkung aus Diskontinuitäts- und Glättungsmaß am besten mit den Deblocking-Bewertungen (BL) und etwas schlechter mit den Bewertungen, die sich auf Unterschiede der Gesamtbildqualität beziehen (Q). Dies ist plausibel, da die Maße stark an Merkmale angepasst sind, die im Zusammenhang mit Blockartefakten stehen und durch Deblocking-Verfahren maßgeblich beeinflusst werden. Die subjektive Gesamtqualitätsverbesserung muss nicht in dem Maße empfunden werden wie die charakteristischen störenden Merkmale von Blockartefakten verringert werden. Die Korrelationen der Maße für Qualitätsverbesserungen mit den Unschärfe-Bewertungen (SC) sind in allen Fällen am geringsten. Umgekehrt ist das Verhalten bei den Maßen, die eine Degradationswirkung der Deblocking-Verfahren erfassen sollen. Diese korrelieren erheblich besser mit den Unschärfe-Bewertungen als mit den anderen Bewertungen, auch wenn die Maße nicht unbedingt so gestaltet sind, dass sie ausschließlich von Unschärfenbildung beeinflusst werden.

**Tabelle 9.1:** Korrelationen der Qualitätsmaße

Maß	Korrel. Q	Korrel. BL	Korrel. SC
Diskontinuitätsmaß	0,798	0,804	-0,704
Glättungsmaß	0,913	0,904	-0,643
Grauwertmaß	0,730	0,801	-0,675
Intensitätsannäherungsmaß	0,798	0,842	-0,656
Differenzvergrößerungsmaß	0,757	0,829	-0,706
Konturdegradationsmaß	0,430	0,569	-0,795
Degradationsmaß	0,633	0,736	-0,801
Differenzverkleinerungsmaß	0,561	0,648	-0,721
Wechselwirkung	0,847	0,844	-0,719

Die Korrelationen der Maße liegen insgesamt auf einem hohen Niveau. Besonders auffällig ist das Glättungsmaß. Es weist eine so hohe Korrelation mit den Bewertungen, die sich auf Gesamtqualitätsunterschiede und speziell auf die Reduktion von Blockartefakten beziehen, auf, dass eine Kombination mit weiteren Maßen zunächst kaum notwendig erscheint. Wie an späterer Stelle noch gezeigt wird, gibt es jedoch entscheidende Gründe, die für eine Kombination sprechen.

In Tabelle 9.2 sind die Korrelationen der Qualitätsschätzwerte der vorgestellten kombinierten Qualitätsbewertungsverfahren und weiterer Qualitätsbewertungsansätze mit den subjektiven Eindrücken zur Gesamtqualität, die im ersten subjektiven Test ermittelt wurden, angegeben. Darüber hinaus sind die mittleren Abweichungen zwischen Qualitätsschätzwerten und subjektiven Eindrücken abgedruckt. Um den PSNR und MSE zur Bewertung der Wirksamkeit von Deblocking-Verfahren verwenden zu können, wurden die Differenzen aus den PSNR- bzw. MSE-Werten der codierten und der entspre-

chenden nachgefilterten Bilder betrachtet. Diese Differenzen wurden DPSNR bzw. DMSE genannt (siehe auch Kapitel 4). Für weitere Untersuchungen und die Berechnung der mittleren Abweichungen mussten DPSNR und DMSE geeignet gewichtet und auf die in den subjektiven Tests verwendete DMOS-Skala abgebildet werden. Dazu wurden lineare Regressionen der subjektiven Eindrücke auf DPSNR bzw. DMSE durch den Ursprung durchgeführt. Entsprechend wurde mit einem Einkanalmodell der visuellen Wahrnehmung, wie bereits in Kapitel 4 vorgestellt, sowie mit den ebenfalls in Kapitel 4 vorgestellten Blocking-Messverfahren nach [Pia05a] und [WY97] umgegangen. Auch hier wurden die Differenzen der Qualitätsschätzungen für codierte und gefilterte Bilder betrachtet.

**Tabelle 9.2:** Güte der Wirksamkeitsschätzung im ersten subjektiven Test

Maß/Verfahren	Korrel. mit Q	mittlere Abw.
Glättungsmaß	0,913	0,154
Modellvariante 1	0,926	0,130
Modellvariante 2	0,936	0,123
Modellvariante 3	0,863	0,164
Modellvariante 4	0,948	0,108
GBIM [WY97]	0,560	0,345
Blocking [Pia05a]	0,689	0,245
DMSE	0,269	0,657
DPSNR	0,240	0,551
Einkanalmodell	0,288	0,666
Vergleichs-Faktorenmodell 1	0,904	0,159
Vergleichs-Faktorenmodell 2	0,283	0,570

Ebenfalls in den Vergleich aufgenommen wurde ein Faktorenmodell nach Vorbild des in Kapitel 4 vorgestellten Modells nach [Fec00]. Es wurde an die Aufgabenstellung in dieser Arbeit angepasst, indem zwar auch hier separate Fehlermaße für Kontur-, Textur- und Flächenbereiche sowie für das Gesamtbild verwendet wurden, allerdings die Differenzen dieser Maße zwischen codierten und gefilterten Bildern berechnet und als Einflussgrößen eines neuen Regressionsmodells verwendet wurden. Es wurde eine Regression durch den Ursprung erzwungen, damit beim Vergleich identischer Bilder der Schätzwert Null ausgegeben werden konnte. Die beschriebenen Anpassungen ließen eine Erhöhung der Schätzgenauigkeit bei der hier gegebenen Aufgabenstellung erwarten. Auf eine umfangreiche statistische Analyse des angepassten Faktorenmodells wurde verzichtet, da die Anpassung bekannter Qualitätsbewertungsansätze an neue Aufgabenstellungen nicht das zentrale Thema dieser Arbeit ist. Es sollte lediglich die Frage beantwortet werden, ob das als Basis dienende Verfahren grundsätzlich durch moderate Änderungen an die Aufgabe

der Wirksamkeitsbewertung von Deblocking-Verfahren angepasst werden kann und ob dabei eine gute Schätzungsgüte erzielt werden kann. Im Folgenden wird das beschriebene Modell *Vergleichs-Faktorenmodell 1* genannt.

Eine nicht speziell an die in dieser Arbeit gegebene Aufgabe angepasste Variante des Faktorenmodells wurde ebenfalls untersucht. Hier wurden einzelne Fehlermaße für Kontur-, Textur- und Flächenbereiche sowie für das Gesamtbild, wie in Kapitel 4 skizziert, direkt als Einflussgrößen eines Regressionsmodells benutzt, welches auf Basis von subjektiven Testdaten aus [Bre06] für die absolute Qualitätsbewertung von MPEG-2-komprimierten Sequenzen ausgelegt wurde. Das Modell erreichte dabei eine hohe Korrelation von 0,95. Die Qualität von codierten und entsprechenden nachgefilterten Bildern wurde mit Hilfe des Modells geschätzt, um anschließend auf Basis der Differenzen der Schätzwerte für die absoluten Bildqualitäten eine Bewertung der Wirksamkeit von Deblocking-Verfahren bezogen auf eine Bildqualitätsverbesserung vornehmen zu können. Die Differenzen der absoluten Bildqualitäten wurden auf die hier verwendete DMOS-Skala abgebildet. Diese Variante des Faktorenmodells wird im Folgenden *Vergleichs-Faktorenmodell 2* genannt.

Bei Betrachtung der Ergebnisse in Tabelle 9.2 fällt zunächst die sehr geringe Korrelation des DPSNR und des DMSE mit den subjektiven Qualitätseindrücken auf. Auch die mittleren Abweichungen zwischen Schätzwerten und subjektiven Eindrücken liegen auf einem hohen Niveau und erreichen im Falle des DMSE fast zwei Drittel einer Skaleneinheit. Die Werte zeigen eindeutig, dass DPSNR und DMSE nicht dazu geeignet sind, die Wirksamkeit von Deblocking-Verfahren miteinander zu vergleichen, auch wenn dies in vielen Fällen praktiziert wird [TCS05, KIL04, RK00, YCK00, KYKR99, KJKH98]. Das Ergebnis verwundert jedoch nicht, da viele Autoren die Eignung von PSNR und MSE zur Bewertung der absoluten Qualität komprimierter Bilder bereits angezweifelt haben (siehe Kapitel 4) und hier eine verwandte Aufgabenstellung vorliegt, die aber tendenziell eine noch höhere Genauigkeit bei der Erfassung von Störmerkmalen, die im Zusammenhang mit einer Bildkompression stehen, erfordert.

Nicht besser als mit DPSNR und DMSE gelingt die Bewertung der Wirksamkeit von Deblocking-Verfahren auf Basis des untersuchten Einkanalmodells, obwohl ein solches Modell ohne Zweifel sehr gut zur Bewertung der absoluten Qualität komprimierten Bildmaterials geeignet ist (siehe [Fec00]). Sowohl Korrelationsgrad als auch mittlere Abweichung liegen hier auf dem Niveau von DPSNR und DMSE. Eine Betrachtung der Fehlerwerte, die das Einkanalmodell berechnet, ergab, dass diese stark durch die Datenrate beeinflusst werden, jedoch oft nur wenig durch die Anwendung verschiedener Deblocking-Verfahren bei gleicher Datenrate. Damit ist das Verhalten der Fehlerwerte im Hinblick auf den Zweck, zu dem das den Untersuchungen zugrunde liegende Einkanalmodell ursprünglich entwickelt wurde - die Erfassung der absoluten Qualität von komprimiertem Bildmaterial über einen weiten Datenratenbereich - günstig. Hingegen ist das Verhalten der Werte im Hinblick auf die hier gegebene Aufgabenstellung - die Wirksamkeitsbewertung von Deblocking-Verfahren - ungünstig. Hinzu kommt, dass die Tendenz der Fehlerwerte nach Anwendung

verschiedener Deblocking-Verfahren sehr häufig nicht subjektiven Eindrücken entspricht. Zusammenfassend kann das Fazit gezogen werden, dass das untersuchte Einkanalmodell zur Bewältigung der Aufgaben in dieser Arbeit kaum geeignet ist.

Ebenfalls zur Bewertung der Wirksamkeit von Deblocking-Verfahren bezogen auf eine Verbesserung der Bildqualität kaum geeignet ist das Vergleichs-Faktorenmodell 2. Es korreliert mit den subjektiven Eindrücken nur wenig besser als DPSNR und DMSE. Die mittlere Abweichung der Schätzwerte von den Qualitätswerten aus dem ersten subjektiven Test liegt ebenfalls auf dem Niveau von DPSNR und DMSE. Es sei allerdings klar betont, dass das Modell zu einem anderen Zweck - der Erfassung der absoluten Qualität von komprimierten Bildfolgen über einen weiten Datenratenbereich - entwickelt bzw. optimiert wurde, der nicht der in dieser Arbeit gegebenen Aufgabenstellung entspricht. Daher verwundern die Ergebnisse nicht. Das an die Aufgabenstellung dieser Arbeit angepasste Vergleichs-Faktorenmodell 1 erzielt allerdings erheblich bessere Ergebnisse bei der Wirksamkeitsschätzung von Deblocking-Verfahren. Es wird ein hoher Korrelationsgrad von etwa 0,9 und eine mittlere Abweichung zwischen Schätzwerten und subjektiven Eindrücken von weniger als einem Fünftel einer Skaleneinheit erzielt. Dennoch erreicht die Schätzgüte nicht ganz die Güte, die sich mit den besten Faktorenmodellen, die speziell zur Wirksamkeitsbewertung von Deblocking-Verfahren entwickelt wurden, erreichen lässt. Darauf wird noch eingegangen.

Deutlich besser als DPSNR und DMSE und insgesamt mittelmäßig gut zur Bewertung der Wirksamkeit von Deblocking-Verfahren geeignet sind erwartungsgemäß gradientenbasierte Blocking-Maße. Der exemplarisch untersuchte Ansatz nach [Pia05a] korreliert recht gut mit den subjektiven Eindrücken menschlicher Betrachter. Die mittlere Abweichung beträgt 0,245 Skaleneinheiten und damit weniger als 40 % der mittleren Abweichung des DMSE. Im direkten Vergleich schlechter ist die Güte der Wirksamkeitsschätzung, die mit der Blocking-Metrik nach [WY97] (GBIM) erreicht werden konnte. Dies zeigt sich sowohl am Korrelationsgrad als auch an der mittleren Abweichung. Bezogen auf die in dieser Arbeit relevante Aufgabenstellung wirkt sich die Integration von Komponenten des HVS in das gradientenbasierte Verfahren im Vergleich zum einfacheren Verfahren nach [Pia05a] eher negativ aus.

Das im Rahmen dieser Arbeit entwickelte Glättungsmaß eignet sich sehr gut zur Bewertung der Qualität von Deblocking-Verfahren, was an dem hohen Korrelationsgrad von über 0,91 und einer mittleren Abweichung von nur 0,15 Skaleneinheiten abgelesen werden kann. Weitere Verbesserungen der Schätzgüte können mit den kombinierten Modellen in unterschiedlichem Ausmaß erreicht werden. Über die Verbesserung des Korrelationsgrades und die Verringerung der mittleren Abweichung hinaus bringen die Kombinationen noch weitere Vorteile mit sich, die z. B. akkurate Schätzwerte in bestimmten kritischen Bildern betreffen. Darauf wird an späterer Stelle eingegangen. In Tabelle 9.2 fällt auf, dass Modellvariante 4 eine herausragend hohe Schätzgüte, repräsentiert durch einen Korrelationsgrad von fast 0,95 und eine mittlere Abweichung von wenig mehr als einer zehntel

Skaleneinheit, aufweist. Allerdings sei an die bereits geäußerten Plausibilitätsprobleme der Variante erinnert, die starke Zweifel an der Übertragbarkeit der Ergebnisse auf andere Bilddaten aufkommen lassen. Insofern dient Modellvariante 4 hier im Wesentlichen als Benchmark, der anzeigt, welche Schätzgüte erreichbar ist, wenn das Regressionsmodell optimal an einen bestimmten Datensatz angepasst wird, ohne dabei auf Plausibilität, Stabilität sowie einen moderaten Aufwand zu achten. Insbesondere Modellvariante 2 ist von der Performance des Benchmark-Modells sowohl in Bezug auf den Korrelationsgrad als auch in Bezug auf die mittlere Abweichung nicht weit entfernt. Modellvariante 1 weist ebenfalls eine hohe Schätzgüte auf. Modellvariante 3 darf nicht mit den anderen Varianten oder dem stärker korrelierenden Glättungsmaß verglichen werden, da das Glättungsmaß bewusst ausgeschlossen wurde. Es zeigt sich aber, dass die lineare Kombination der drei integrierten Maße eine deutliche Verbesserung der Schätzgüte verglichen mit den Einzelmaßen ermöglicht. Diese ist größer als bei den Kombinationen, die das Glättungsmaß enthalten, und zeigt die Möglichkeiten und Chancen auf, die ein lineares Regressionsmodell bieten kann.

Um abschätzen bzw. prognostizieren zu können, ob eine zuverlässige Bewertung der Wirksamkeit von Deblocking-Verfahren bezogen auf Gesamtbildqualitätsverbesserungen mit den in dieser Arbeit entwickelten Qualitätsbewertungsansätzen auch bei der Analyse von Bildern oder Deblocking-Verfahren zu erwarten ist, die nicht im Rahmen der Berechnung der Regressionskoeffizienten eingesetzt wurden, fanden weitere Untersuchungen statt. Der im ersten subjektiven Test ermittelte Datensatz wurde zufällig in zwei Teile aufgespalten. Ein Teil, der etwa zwei Drittel der Daten umfasste, wurde zur Berechnung neuer Regressionskoeffizienten benutzt (Lerndatensatz). Der andere Teil, bestehend aus den restlichen Daten (Testdatensatz), wurde zur Ermittlung neuer Korrelationsgrade und mittlerer Abweichungen zwischen Schätzwerten und subjektiven Eindrücken verwendet. Die Gütewerte wurden also ausschließlich auf Basis von Daten ermittelt, die nicht zur Berechnung der Regressionskoeffizienten herangezogen wurden. Um stabile Zahlen zu erhalten, wurde die zufällige Aufspaltung der Daten, die anschließende Koeffizientenberechnung und die Ermittlung der Gütewerte jeweils 10.000 Mal durchgeführt. Die mittleren Ergebnisse finden sich in Tabelle 9.3. Das Einkanalmodell und das Vergleichs-Faktorenmodell 2 wurden in dieser und in den folgenden Untersuchungen aufgrund des in den bereits beschriebenen Untersuchungen vergleichsweise schwachen Abschneidens unter den komplexen Verfahren nicht weiter berücksichtigt. Es sei darauf hingewiesen, dass in Anhang B eine Tabelle abgedruckt ist, die ausgewählte Werte der Tabellen 9.2 bis 9.4 umfasst und somit den Vergleich einzelner Werte vereinfacht.

Zunächst fallen in Tabelle 9.3 die verhältnismäßig kleinen Unterschiede der Gütewerte für DMSE und DPSNR im Vergleich zu den im gesamten ersten Test ermittelten Werten (siehe Tabelle 9.2) auf. Dies verwundert nicht, da bei diesen einfachen Maßen weder deutliche Verbesserungen noch deutliche Verschlechterungen zu erwarten waren. Ähnliches gilt für die beiden gradientenbasierten Blocking-Maße, auch wenn die Abweichungen hier im

**Tabelle 9.3:** Güte der Wirksamkeitsschätzung bei Trennung der Datensätze zur Berechnung der Regressionskoeffizienten und der Gütemaße

Maß/Verfahren	Korrel. mit Q	mittlere Abw.
Glättungsmaß	0,912	0,160
Modellvariante 1	0,916	0,145
Modellvariante 2	0,920	0,145
Modellvariante 3	0,845	0,182
Modellvariante 4	0,928	0,133
GBIM [WY97]	0,548	0,357
Blocking [Pia05a]	0,662	0,253
DMSE	0,267	0,677
DPSNR	0,248	0,574
Vergleichs-Faktorenmodell 1	0,867	0,188

Vergleich etwas höher ausfallen. Des Weiteren fallen die sehr geringen Verschlechterungen bzgl. Korrelation und mittlerer Abweichung beim Glättungsmaß auf. Dadurch wird die allgemein gute Eignung des Maßes zur Bewertung der Wirksamkeit von Deblocking-Verfahren bestätigt. Damit verglichen etwas größer sind die Abweichungen der Gütewerte für die in dieser Arbeit entwickelten kombinierten Wirksamkeitsbewertungsverfahren gegenüber den Angaben in Tabelle 9.2. Der Grund liegt darin, dass mit zunehmender Anzahl an Teilmaßen letztlich eine immer bessere Anpassung eines Modells an einen bestimmten Datensatz erfolgen kann. Gleichzeitig werden Schätzungen in anderen Datensätzen in den meisten Fällen nicht so gut sein wie in dem Datensatz, der bei der Entwicklung des Modells herangezogen wurde, was die größeren Abweichungen erklärt. Entscheidend ist jedoch, dass die Abweichungen trotzdem insgesamt klein bleiben sollten. Große Abweichungen würden gegen die Stabilität des Qualitätsbewertungsverfahrens oder für Ungereimtheiten im Rahmen der Regressionsanalyse sprechen. Es läge dann die Vermutung nahe, dass zwar ein funktionaler Zusammenhang zwischen den verwendeten Qualitätsmaßen und dem Lern Datensatz gefunden wurde, das Modell aber davon abgesehen falsch spezifiziert ist und seine Aufgabe nicht erfüllen kann. Darauf deuten die angegebenen Zahlen nicht hin, da die Unterschiede der Gütewerte in den Tabellen 9.2 und 9.3 für die kombinierten Modellvarianten verhältnismäßig klein sind. Die größten Abweichungen treten erwartungsgemäß bei Modellvariante 4 auf, allerdings sind sie nicht so groß wie aufgrund der Plausibilitätsprobleme erwartet werden konnte. Die insgesamt kleinsten Abweichungen sind für Modellvariante 1, der bzgl. des Aufbaus klarsten und plausibelsten Modellvariante, ablesbar. Absolut betrachtet liegen die in dieser Untersuchung ermittelten durchschnittlichen Abweichungen zwischen Qualitätsschätzwerten und subjektiven Eindrücken für die kombinierten Modellvarianten laut Tabelle 9.3 bei maximal 0,145 Skaleneinheiten (Variante

3 ausgenommen) im Vergleich zu fast 0,7 Skaleneinheiten beim DMSE und 0,57 Skaleneinheiten beim DPSNR. Die Korrelationsgrade der drei besten Modellvarianten zwischen 0,916 und 0,928 zeigen zusammen mit den kleinen mittleren Abweichungen eine sehr hohe Übereinstimmung der geschätzten Qualitätsbewertungen mit den subjektiven Eindrücken menschlicher Betrachter auch in Fällen an, die keine Berücksichtigung beim Berechnen der Regressionskoeffizienten fanden. Beim Vergleichs-Faktorenmodell 1 sind die Unterschiede von Korrelation und mittlerer Abweichung zu den Werten in Tabelle 9.2 verhältnismäßig groß. Es besteht ein erster Verdacht, dass das Modell zu einer Überanpassung an den Lerndatensatz im hier gegebenen Aufgabenumfeld neigt. Dies hängt vermutlich damit zusammen, dass das Modell trotz der vorgenommenen Anpassungen letztlich nicht für die hier vorliegende Aufgabenstellung entwickelt wurde und keine daraufhin optimierten Maße enthält. Grundsätzlich musste aus diesem Grund mit einer Verschlechterung der Gütewerte gerechnet werden.

## 9.2 Schätzgüte im dritten subjektiven Test

Um weitere Erkenntnisse zum Verhalten der untersuchten Qualitätsbewertungsverfahren in Fällen, die nicht bei der Berechnung der Regressionskoeffizienten bzw. Mapping-Funktionen berücksichtigt wurden, zu gewinnen, wurde der in Kapitel 7 beschriebene dritte subjektive Test durchgeführt. Er enthielt kein einziges Testbild, das in den ersten beiden subjektiven Tests verwendet wurde. Auch eine große Ähnlichkeit zwischen den Bildern der ersten Tests und des dritten Tests war nicht gegeben. Des Weiteren wurden im dritten subjektiven Test zwei neue Deblocking-Verfahren berücksichtigt. Dies waren eine statische 3x3 Mittelwertfilterung gesamter Bilder und das adaptive Verfahren mit drei Filtermodi nach [TCS05]. In Tabelle 9.4 sind die Korrelationsgrade der Schätzwerte verschiedener Ansätze zur Bewertung der Wirksamkeit von Deblocking-Verfahren mit den Eindrücken der Teilnehmer des dritten subjektiven Tests abgedruckt. Des Weiteren sind die mittleren Abweichungen zwischen geschätzten Qualitätsunterschieden und subjektiv empfundenen Qualitätsunterschieden angegeben. Es sei klargestellt, dass sämtliche innerhalb der einzelnen Qualitätsbewertungsverfahren verwendeten Mapping-Funktionen zur Berechnung der DMOS-Werte auf Basis des ersten subjektiven Tests ermittelt wurden.

Die Modellvarianten 1 und 2 des in dieser Arbeit entwickelten Wirksamkeitsbewertungsansatzes korrelieren auch in diesem Test mit Werten über 0,92 sehr gut mit den subjektiven Eindrücken der Probanden und damit kaum schlechter als mit den Eindrücken im ersten subjektiven Test. Dies verdeutlicht die Stabilität der Modellvarianten und zeigt, dass diese offensichtlich auch in Bildmaterial, das nicht zur Berechnung der Regressionskoeffizienten verwendet wurde, eine akkurate Schätzung der Wirksamkeit von Artefaktreduktionsverfahren bezogen auf eine Bildqualitätsverbesserung ermöglichen. Die geringen Erhöhungen der mittleren Abweichungen zwischen den Schätzwerten und den

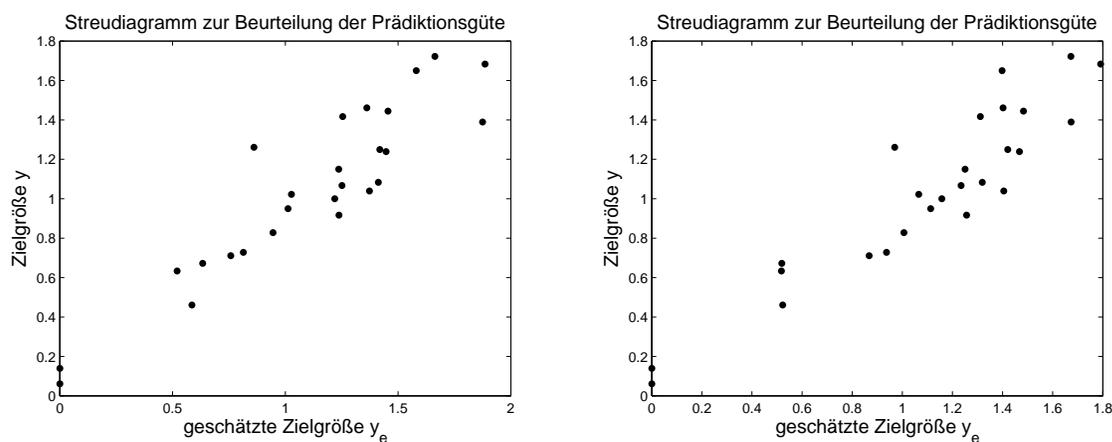
**Tabelle 9.4:** Zahlen zur Güte der Wirksamkeitsschätzung im dritten subjektiven Test

Maß/Verfahren	Korrel. mit Q	mittlere Abw.
Glättungsmaß	0,878	0,169
Modellvariante 1	0,921	0,159
Modellvariante 2	0,926	0,163
Modellvariante 3	0,858	0,274
Modellvariante 4	0,571	0,346
GBIM [WY97]	0,666	0,326
Blocking [Pia05a]	0,778	0,217
DMSE	0,349	0,907
DPSNR	0,401	0,731
Vergleichs-Faktorenmodell 1	0,400	0,803

subjektiven Eindrücken im Vergleich zum ersten subjektiven Test stützen die Erkenntnisse. Ähnliches gilt für Modellvariante 3, allerdings auf einem niedrigeren Güteniveau der Wirksamkeitsschätzung. Es fällt auf, dass die bis hierher genannten Ergebnisse sehr große Ähnlichkeiten mit denen in Tabelle 9.3 aufweisen. Wie bereits beschrieben, sind dort die Schätzgüten von Qualitätsbewertungsverfahren in während der Koeffizientenberechnung nicht berücksichtigten Fällen, entstanden durch Aufteilung der Daten des ersten subjektiven Tests in Lern- und Testdaten, enthalten.

Interessant an den Gütewerten in Tabelle 9.4 ist das verhältnismäßig schlechte Abschneiden der Modellvariante 4. Hier fällt die Schätzgüte im Vergleich zum ersten subjektiven Test deutlich schlechter aus. Dies war jedoch zu erwarten, da Modellvariante 4 von vornherein als unplausibles Verfahren eingeschätzt wurde. Es ermöglicht eine gute Beschreibung der subjektiven Eindrücke der Teilnehmer des ersten subjektiven Tests, an den es sehr gut angepasst ist. Eine universelle Einsetzbarkeit in beliebigem Bildmaterial ist jedoch nicht gegeben. Gute Ergebnisse liefert hingegen das Glättungsmaß. Zwar ist der Korrelationsgrad im Vergleich zu den Modellvarianten 1 und 2 deutlich kleiner, die durchschnittliche Abweichung zwischen Schätzungen und subjektiven Eindrücken ist jedoch auf einem ähnlichen Niveau. Die Bewertung der Wirksamkeit von Deblocking-Verfahren auf Basis von DMSE und DPSNR erweist sich auch im dritten subjektiven Test als stark problembehaftet. Zwar sind die Korrelationsgrade sogar etwas größer als im ersten Test, allerdings sind sie nach wie vor - absolut betrachtet - klein, und die durchschnittlichen Abweichungen zwischen Schätzungen und subjektiven Eindrücken sind mit 0,7 bzw. 0,9 Skaleneinheiten sehr groß. Der Grund für das etwas bessere Abschneiden von DPSNR und DMSE im Vergleich zum ersten subjektiven Test dürften die beiden bereits mehrfach erwähnten auffälligen Bewertungen mit dem Charakter von Ausreißern (Bildpaare 18 und 28) im ersten Test sein, die sich negativ auf die Gütewerte der meisten Verfahren ausge-

wirkt haben. Im dritten subjektiven Test gab es keine besonders auffälligen Fälle, was sich günstig auf die Schätzgüte auswirken kann. Dies zeigt sich auch an den Schätzgüten der gradientenbasierten Qualitätsbewertungsverfahren. Die Korrelationsgrade mit den Daten des dritten subjektiven Tests sind größer als die Korrelationsgrade mit den Daten des ersten subjektiven Tests. Die mittleren Abweichungen zwischen den Schätzwerten und den subjektiven Eindrücken sind minimal kleiner. Das Vergleichs-Faktorenmodell 1 weist in dieser Untersuchung nur einen Korrelationsgrad auf, der auf dem Niveau der Korrelationsgrade von DMSE und DPSNR und somit erheblich niedriger als im ersten Test liegt. Die mittlere Abweichung zwischen geschätzten Qualitätsunterschieden und subjektiv empfundenen Qualitätsunterschieden ist ebenfalls mit der von DMSE und DPSNR vergleichbar. Auch dies entspricht einer deutlichen Verschlechterung gegenüber dem ersten subjektiven Test. Der Grund dürfte, wie bereits im Abschnitt zuvor angegeben, sein, dass das Modell nicht für die hier vorliegende Aufgabenstellung entwickelt wurde und keine daraufhin optimierten Maße enthält. Die vorgenommenen Anpassungen an die Aufgabenstellung ermöglichen zwar offenkundig eine gute Anpassung des Modells an die Daten des ersten subjektiven Tests bzw. an den Lerndatensatz, ergeben jedoch kein stabiles Verfahren zur Bewertung der Wirksamkeit von Deblocking-Verfahren in einem breiten Spektrum an Bildmaterial, wie die Zahlen zeigen.



**Abbildung 9.1:** Streudiagramme zur Beurteilung der Schätzgüte der Modellvarianten 1 und 2 im dritten subjektiven Test

In Abbildung 9.1 sind ergänzend zu den Angaben in Tabelle 9.4 die Streudiagramme der Zielgröße  $y$  über der geschätzten Zielgröße  $y_e$  für die Modellvarianten 1 (links) und 2 (rechts) des im Rahmen dieser Arbeit entwickelten Wirksamkeitsbewertungsansatzes abgedruckt. Auch hier wurden die auf Basis des ersten subjektiven Tests ermittelten Mapping-Funktionen bzw. Regressionskoeffizienten verwendet. In beiden Fällen ist erkennbar, dass die Punkte um eine gedachte Gerade herum angeordnet sind. Hinweise auf nichtlineare

Abhängigkeiten bestehen nicht, so dass sich auch hier zeigt, dass die Aufstellung linearer Modelle gerechtfertigt ist. In den Diagrammen ist ebenfalls sichtbar, dass in beiden Fällen die Abweichungen zwischen geschätzter Zielgröße und Zielgröße recht klein sind. Die Streudiagramme weisen letztlich keine besonderen Auffälligkeiten auf und sind mit denen für den ersten subjektiven Test (siehe Kapitel 8) vergleichbar.

### 9.3 Plausibilität und Stabilität

Um Eindrücke bezüglich der Plausibilität und Stabilität der neu vorgestellten Regressionsmodelle zur Bewertung der Wirksamkeit von Deblocking-Verfahren zu gewinnen, wurden weitere Untersuchungen durchgeführt. Zunächst wurden auf Basis des dritten subjektiven Tests für die Modellvarianten 1 bis 4 und für ein Modell, das ausschließlich das Glättungsmaß enthält, neue Regressionskoeffizienten berechnet. Bei alleiniger Verwendung des Glättungsmaßes sind die Regressionskoeffizienten, die auf Basis des ersten und des dritten subjektiven Tests berechnet wurden, nahezu identisch. Das Glättungsmaß alleine scheint also Bildqualitätsverbesserungen akkurat zu erfassen und ermöglicht eine stabile Schätzung dieses Kriteriums. Ähnliches gilt für Modellvariante 1.

In Tabelle 9.5 sind die neu berechneten Regressionskoeffizienten für Variante 1 im Vergleich zu den im ersten Test ermittelten Koeffizienten angegeben. Es fallen keine großen Unterschiede bzw. Abweichungen auf, was auf eine hohe Stabilität der Qualitätsschätzung auf Basis des Modells hindeutet. Gleiches gilt für Modellvariante 2, da die neu berechneten Koeffizienten ebenfalls keine großen Abweichungen aufweisen (siehe Tabelle 9.6).

**Tabelle 9.5:** Regressionskoeffizienten der Modellvariante 1, ermittelt auf Basis der Daten aus dem ersten und dem dritten subjektiven Test im Vergleich

Maß	Koeffizient aus Test 1	Koeffizient aus Test 3
Diskontinuitätsmaß	0,6272	0,4977
Glättungsmaß	2,3319	2,2476
Konturdegradationsmaß	-0,8854	-0.6462

Bei Modellvariante 3 treten deutlichere Abweichungen auf (siehe Tabelle 9.7), weshalb bei Verwendung dieser Variante mit größeren Unschärfen bei der Qualitätsschätzung gerechnet werden muss. Davon musste allerdings auch schon aufgrund der Ergebnisse der bisherigen Untersuchungen (größere Schätzfehler, geringere Korrelationen) ausgegangen werden. Der Vollständigkeit halber sei erwähnt, dass sich die neuen Regressionskoeffizienten für Modellvariante 4 schließlich sehr stark von dem ursprünglichen Satz an Koeffizienten unterscheiden. Es tritt sogar ein Vorzeichenwechsel auf. Die bisherigen Zweifel an der Stabilität und Plausibilität der Variante 4 werden dadurch bestätigt.

**Tabelle 9.6:** Regressionskoeffizienten der Modellvariante 2, ermittelt auf Basis der Daten aus dem ersten und dem dritten subjektiven Test im Vergleich

Maß	Koeffizient aus Test 1	Koeffizient aus Test 3
Diskontinuitätsmaß	0,8298	0,5167
Glättungsmaß	3,4238	3,9727
Degradationsmaß	-0,8807	-0,7810
Wechselwirkung	-0,9313	-1,0847

**Tabelle 9.7:** Regressionskoeffizienten der Modellvariante 3, ermittelt auf Basis der Daten aus dem ersten und dem dritten subjektiven Test im Vergleich

Maß	Koeffizient aus Test 1	Koeffizient aus Test 3
Diskontinuitätsmaß	1,2872	0,9955
Differenzvergrößerungsmaß	2,4232	1,4386
Konturdegradationsmaß	-2,0131	-0,4520

In weiteren Untersuchungen wurde überprüft, ob sich die vorgestellten Modelle und das Glättungsmaß in besonderen Fällen nachvollziehbar verhalten. Dazu können z. B. hohe Datenraten, extreme Schärfe- und Detailverluste durch die Anwendung von Artefaktreduktionsverfahren, sowie P- und B-Frames gezählt werden. Es wurden mit hoher Datenrate codierte Bilder, mit einem 5x5 Mittelwertfilter nachgefilterte gestörte Bilder, sowie gefilterte P- bzw. B-Frames in die Untersuchungen einbezogen.

In einem ersten Schritt wurde die Bewertung der Wirksamkeit von Deblocking-Verfahren in einzelnen I-Frames der mit einer Datenrate von 8 Mbit/s codierten Sequenzen *football* und *wheel* betrachtet. Rein subjektiv sind in den Bildern nur wenige schwach ausgeprägte Artefakte erkennbar, so dass keine deutliche Bildqualitätsverbesserung möglich ist. Da die Bewertung der Wirksamkeit von Artefaktreduktionsverfahren in Bildern, die kaum Artefakte enthalten, nur bedingt sinnvoll ist, ging es hier darum, zu überprüfen, ob sich die Wirksamkeitsschätzung grundsätzlich erwartungsgemäß bzw. unkritisch verhält. Dies ist offenbar der Fall, denn die Wirksamkeit der Artefaktreduktionsverfahren wird in allen Fällen geringer geschätzt als in Bildern, die mit niedrigerer Datenrate codiert wurden, und die Schätzwerte fallen insgesamt klein aus. Die Unterschiede zwischen den geschätzten Wirksamkeiten durch die einzelnen Verfahren sind im Vergleich zu höheren Datenraten absolut betrachtet klein. Die größten und allenfalls minimal zu großen Schätzwerte (Maximum 0,83) werden bei alleiniger Verwendung des Glättungsmaßes berechnet. Es ist möglich, dass hier vereinzelte ungewollte Glättungen gemessen und positiv ausgelegt werden. Bei den kombinierten Messverfahren (mit Ausnahme der ohnehin problema-

tischen Variante 4) machen sich die integrierten kompensierenden Komponenten wie z. B. das Degradationsmaß bemerkbar und führen zu etwas geringeren Schätzwerten (Maximum 0,77). Insgesamt gibt es keine kritischen Auffälligkeiten. Die Anwendung der verschiedenen Varianten zur Wirksamkeitsschätzung ist also offenbar auch bei hohen Datenraten bzw. geringer Anzahl oder Ausprägung von Artefakten grundsätzlich ohne besondere Probleme möglich, auch wenn der Sinn zweifelhaft ist.

In einem zweiten Schritt wurde überprüft, wie sich die Schätzung der Wirksamkeit von Deblocking-Verfahren in P- bzw. B-Frames verhält. Der verwendete MPEG-Encoder erzeugt in solchen Frames (insbesondere in B-Frames) tendenziell weniger bzw. weniger deutlich ausgeprägte Artefakte als in I-Frames (siehe Abbildung 9.2). Hinzu kommt, dass blockartige Artefakte durch die Bewegungskompensation an Stellen auftreten können, die nicht dem üblichen Blockraster entsprechen. Darauf sind die untersuchten Deblocking-Verfahren und auch die Verfahren zur Wirksamkeitsschätzung nicht spezialisiert. Insgesamt ist in prädiktiv codierten Frames mit geringeren geschätzten Wirksamkeiten als in I-Frames zu rechnen.

(a) *football*, Frame 24 (I-Frame)(b) *football*, Frame 25 (B-Frame)

**Abbildung 9.2:** Bildbeispiel zur Veranschaulichung der unterschiedlichen Ausprägung von Artefakten in unterschiedlichen Frame-Typen

In Tabelle 9.8 sind einige mit der Variante 1 der Wirksamkeitsschätzung ermittelte Schätzwerte für verschiedene Frames aus den Sequenzen *football* und *wheel* angegeben. Die Schätzung der Wirksamkeit der eingesetzten Deblocking-Verfahren fällt insbesondere in dem B-Frame deutlich niedriger aus als in dem I-Frame. Für die statischen Filterverfahren wird sogar vereinzelt eine Bildqualitätsverschlechterung prognostiziert. Absolut betrachtet sind die Unterschiede zwischen den Bewertungen der verschiedenen Deblocking-Verfahren in den prädiktiv codierten Frames klein. Die Schätzwerte für die Wirksamkeit in dem P-Frame sind insgesamt höher als die für die Wirksamkeit in dem B-Frame, im

Vergleich zu den Werten für das I-Frame aber trotzdem auf einem niedrigen Niveau. Die in Tabelle 9.8 angegebenen Werte konnten bei subjektiver Betrachtung der analysierten Frames grundsätzlich nachvollzogen werden. Es ist nicht zu bestreiten, dass die Deblocking-Verfahren in P-Frames eine geringere Wirksamkeit bzgl. einer Bildqualitätsverbesserung aufweisen als in I-Frames und dass die Wirksamkeit in B-Frames tendenziell noch geringer ist. Gleichzeitig sind in den betrachteten P- und B-Frames insgesamt weniger bzw. weniger deutlich ausgeprägte Artefakte vorhanden. Die Schätzwerte der Wirksamkeiten sind insofern plausibel. Ob sie insgesamt auch auf dem korrekten Niveau sind, kann nicht abschließend beurteilt werden, da hierzu weitere subjektive Testdaten notwendig wären. Es kann aber festgehalten werden, dass sich die Wirksamkeitsschätzung auch in P- und B-Frames nicht auffällig verhält. Dennoch sollten die Schätzwerte grundsätzlich wegen der erwähnten Spezialisierung der beteiligten Verfahren und Komponenten auf I-Frames mit Vorsicht interpretiert werden. Die anderen Varianten der Wirksamkeitsschätzung sowie das Glättungsmaß verhielten sich sehr ähnlich und werden daher nicht gesondert betrachtet.

**Tabelle 9.8:** Bewertung der Wirksamkeit von Deblocking-Verfahren in verschiedenen Frame-Typen durch Modellvariante 1, Angabe des DMOS

Sequenz, Verfahren	I-Frame (24)	B-Frame (25)	P-Frame (27)
football, avril	0.73	-0.05	0.12
football, kim	1.56	0.21	0.52
football, ramkishor	1.51	0.02	0.35
football, pi7	1.83	0.16	0.50
football, 3x3 MW	0.48	-0.35	-0.15
wheel, avril	0.88	0.02	0.28
wheel, kim	1.15	0.32	0.66
wheel, ramkishor	1.19	0.15	0.60
wheel, pi7	1.43	0.25	0.66
wheel, 3x3 MW	0.58	-0.17	0.11

In einem dritten Schritt wurde überprüft, wie sich das Glättungsmaß und die kombinierten Modellvarianten verhalten, wenn im Zuge einer Artefaktreduktion starke Schärfe- bzw. Detailverluste auftreten. Zu diesem Zweck wurden gestörte Bilder mit einer statischen 5x5 Mittelwertfilterung bearbeitet. Subjektiv werden dadurch nicht alle Artefakte ausreichend verringert, insgesamt werden aber deutlich zu viele relevante Bilddetails degradiert, was insbesondere in wenig gestörten Bildern unangenehm auffällt. Tabelle 9.9 zeigt Ergebnisse der durchgeführten Untersuchungen. Bei alleiniger Verwendung des Glättungsmaßes treten erhebliche Probleme auf. Die Bewertung der Wirksamkeit der 5x5 Mittelwertfilterung fällt in allen Fällen zu hoch aus. Bei stark gestörten Bildern liegen die Bewertungen

auf dem Niveau der subjektiv wirksamsten anderen Deblocking-Verfahren und bezogen auf die beschriebenen subjektiven Eindrücke eindeutig zu hoch. In weniger stark gestörten Bildern sind die Schätzwerte der Wirksamkeit zwar kleiner, liegen aber über dem Niveau der subjektiv wirksamsten Verfahren und somit ebenfalls deutlich zu hoch. Ein ähnliches Verhalten zeigt Variante 4 der kombinierten Wirksamkeitsschätzung. Die Varianten 1 und 2 verhalten sich hingegen plausibler. Im Fall von stark gestörten Bildern prognostizieren sie für die 5x5 Mittelwertfilterung in der Regel noch eine minimale Bildqualitätsverbesserung, während bei weniger stark gestörten Bildern eine Bildqualitätsverschlechterung prognostiziert wird. Variante 2 reagiert stärker auf die durch die Tiefpassfilterung verursachten Degradationen und liefert Schätzwerte, die vermutlich besser mit den subjektiven Eindrücken menschlicher Betrachter übereinstimmen dürften. Variante 3 reagiert zu stark auf die Tiefpassfilterung und prognostiziert unrealistisch große Bildqualitätsverschlechterungen, die bei weniger stark gestörten Bildern im Bereich von drei Skaleneinheiten liegen. Es kann festgehalten werden, dass Deblocking-Verfahren, die mit starken Schärfe- oder Detaildegradationen einhergehen, für die Wirksamkeitsschätzung kritisch sind. Dies zeigt sich besonders durch die erheblich zu guten Bewertungen durch das Glättungsmaß. Es ist daher sinnvoll, Maße in die Wirksamkeitsschätzung zu integrieren, die Degradationen erfassen und berücksichtigen. Eine recht ausgewogene Integration weisen die Variante 1 und besonders die Variante 2 auf.

**Tabelle 9.9:** Bewertung der Wirksamkeit einer 5x5 Mittelwertfilterung bei der Blocking-Reduktion durch verschiedene Verfahren, Angabe des DMOS

Bild (jeweils Frame 24)	Glättungsm.	Var. 1	Var. 2	Var. 3	Var. 4
football, 3,5 Mbit/s	1,75	0,62	0,21	-1,14	1,60
football, 5 Mbit/s	1,47	-0,19	-0,77	-2,52	0,86
football, 8 Mbit/s	0,99	-0,90	-1,74	-3,34	0,04
wheel, 3,5 Mbit/s	1,40	0,59	-0,01	-0,71	1,70
wheel, 5 Mbit/s	0,69	-0,50	-1,33	-2,06	0,90
wheel, 8 Mbit/s	0,30	-1,27	-2,35	-3,20	0,28

An dieser Stelle sollen noch einige Anmerkungen zum Bildmaterial, das zur Bewertung von Deblocking-Verfahren verwendet werden sollte, gemacht werden. Wie an anderer Stelle bereits erwähnt, wurden für die Untersuchungen in dieser Arbeit Bilder bzw. Sequenzen verwendet, die ein ausgewogenes Verhältnis aus Konturen, Texturen und homogenen Bereichen enthalten. Für Beispiele sei auf Anhang E verwiesen. Die vorgestellten Verfahren zur Wirksamkeitsschätzung sind daher primär auf derartige Bilder optimiert, was keine große Einschränkung darstellt, da typisches Bildmaterial sowohl Konturen, Texturen als auch homogene Bereiche enthält. Ein genau definiertes Verhältnis der einzelnen Anteile ist nicht notwendig. Dennoch muss mit Unschärfen bei der Wirksamkeitsschätzung gerechnet

werden, wenn Bilder verwendet werden, die Missverhältnisse bezüglich des Vorhandenseins und der Größe der genannten Bildbereiche aufweisen. Es konnte die Tendenz beobachtet werden, dass die Wirksamkeitsschätzung in Bildern, die extrem viele Details bzw. sehr viele Konturen und Texturen enthalten, zu niedrig ausfallen kann. Als Beispiele seien bildfüllende Blumenwiesen oder eine Hafenan­sicht mit vielen Segelbooten samt zugehöriger Takelage genannt. Der Grund dafür ist, dass in solchen Fällen nach Anwendung von Deblocking-Verfahren oft sehr große Degradationswerte gemessen werden, was zu niedrigen Wirksamkeitsschätzwerten führt. Die Verhältnisse zwischen Degradationswert und anderen Messwerten sind im Vergleich zu den Verhältnissen bei ausgewogeneren Bildern zu unterschiedlich, so dass Regressionsmodelle, deren Koeffizienten auf Basis von eher ausgewogenen Bildern berechnet wurden, zu Fehlschätzungen neigen. Im Falle von Bildern mit sehr wenig Aktivität, also sehr großen unstrukturierten Bereichen wie z. B. Wasserflächen (ohne viele Wellen), sind umgekehrte Tendenzen zu beobachten. Hier können insbesondere das Diskontinuitätsmaß und teilweise auch das Glättungsmaß untypisch große Werte annehmen, während das Konturdegradationsmaß aufgrund fehlender Konturen untypisch klein werden kann. Dadurch entstehen ebenfalls Verhältnisse zwischen den Einzelmaßen, die sich so stark von denen in ausgewogeneren Bildern unterscheiden können, dass es zu verzerrten und in diesem Fall zu etwas zu großen Schätzungen kommen kann. Die beschriebene Problematik könnte verringert werden, indem für kombinierte Verfahren zur Bewertung der Wirksamkeit von Artefaktreduktionsverfahren unterschiedliche Koeffizientensätze für sehr detailreiche, sehr homogene und ausgewogene Bilder ermittelt und bereitgehalten werden. Eine primitive globale Bildaktivitätsmessung zur Einteilung eines Testbildes in eine der drei Klassen bzw. zur Auswahl des am besten geeigneten Koeffizientensatzes dürfte ausreichen, um eine akkurate Wirksamkeitsschätzung in allen genannten Bildtypen zu ermöglichen.

## 9.4 Zusammenfassung und Interpretation der Ergebnisse

Zusammenfassend kann auf Basis der durchgeführten Untersuchungen festgestellt werden, dass die im Rahmen dieser Arbeit entwickelten Qualitätsmaße und insbesondere deren Kombinationen über einen linearen Regressionsansatz eine akkurate und zuverlässige Bewertung der Wirksamkeit von Deblocking-Verfahren bezüglich einer Verbesserung der Bildqualität ermöglichen. Dies gilt nicht nur auf einem einzelnen Qualitätsniveau, sondern in unterschiedlich stark gestörten Bildern bzw. bei unterschiedlichen Datenraten. Die Korrelationen mit den Eindrücken menschlicher Betrachter können je nach Maß bzw. Kombination sehr hoch sein, was nicht nur im ersten subjektiven Test gezeigt wurde, sondern auch im aussagekräftigeren dritten subjektiven Test, in dem kein einziges Test-

bild verwendet wurde, das auch zum Berechnen der Regressionskoeffizienten diente. Auch eine akkurate Schätzung der Wirksamkeit der beiden in diesem Test neu hinzugefügten Deblocking-Verfahren gelang. Mit den Modellvarianten 1 und 2 konnten Korrelationsgrade von mehr als 0,92 erreicht werden (wobei der Testdatensatz nicht dem Lerndatensatz entsprach). Damit sind die Schätzgüten ganz erheblich besser als die auf Basis des MSE oder PSNR erreichbaren Schätzgüten, welche in den durchgeführten Untersuchungen bestenfalls durch Korrelationen im Bereich von 0,35 bzw. 0,4, in anderen Fällen aber noch darunter gekennzeichnet waren. Die weiteren untersuchten Ansätze zur Wirksamkeitsbewertung weisen insgesamt ebenfalls eine deutlich schlechtere Schätzgüte auf, was auch für die vereinzelt zur Bewertung von Deblocking-Verfahren herangezogenen gradientenbasierten Blocking-Metriken gilt.

Die kleinsten für den neuen Bildqualitätsbewertungsansatz ermittelten durchschnittlichen Abweichungen zwischen den geschätzten Wirksamkeiten der untersuchten Artefaktreduktionsverfahren und den von menschlichen Betrachtern empfundenen Wirksamkeiten liegen bei etwa 0,16 Skaleneinheiten auf der DMOS-Skala (im aussagekräftigeren dritten subjektiven Test). Diese Abweichungen gelten für die Modellvarianten 1 und 2. Sie sind erheblich kleiner als die bei Verwendung von MSE oder PSNR erreichbaren Abweichungen von 0,73 bzw. 0,91 Skaleneinheiten, die anzeigen, wie problematisch ein Wirksamkeitsvergleich von Deblocking-Verfahren auf Basis von MSE und PSNR ist. Mittlere Abweichungen im Bereich von 0,16 Skaleneinheiten auf der DMOS-Skala sind hingegen für eine Einschätzung der Wirksamkeit von Deblocking-Verfahren klein genug. Die neuen kombinierten Qualitätsmaße können somit eine besonders wichtige Eigenschaft von Deblocking-Verfahren, deren Leistungsfähigkeit bei der Qualitätsverbesserung gestörter Bilder, objektiv bewerten. Allerdings sollte - wie fast immer beim Einsatz objektiver Maße im Kontext der Bildqualitätsbewertung - berücksichtigt werden, dass auch kleine Abweichungen zwischen geschätzten und „wahren“ Qualitätsgrößen in bestimmten Fällen zu Fehlinterpretationen führen können, zumal einzelne Abweichungen größer ausfallen als die angegebenen mittleren Abweichungen. Wenn die neuen Qualitätsbewertungsverfahren im Rahmen der Auswahl eines Deblocking-Verfahrens für einen gegebenen Anwendungszweck zwei Verfahren bei der Verbesserung der Bildqualität gestörter Bilder als sehr ähnlich wirksam einschätzen, sollten weitere Qualitätsmaße, z. B. die Werte der in die neuen Verfahren integrierten Maße oder auch die subjektiven Einschätzungen eines oder mehrerer Experten betrachtet werden, um einen sicheren Wirksamkeitsvergleich zu gewährleisten. Allerdings können moderate Schätzfehler in einem solchen Fall auch tolerabel sein, wenn weitere Eigenschaften von Artefaktreduktionsverfahren wie z. B. Rechenaufwand, Speicheraufwand, Unabhängigkeit von Seiteninformationen etc. berücksichtigt werden.

Bezüglich der Eigenschaften der verschiedenen in dieser Arbeit entwickelten Modellvarianten zur Bewertung der Wirksamkeit von Deblocking-Verfahren enthüllten die durchgeführten Untersuchungen interessante Unterschiede. Eine sehr hohe Korrelation mit den Daten des ersten subjektiven Tests konnte mit den Modellvarianten 1, 2 und 4 sowie mit

dem Glättungsmaß alleine erreicht werden. Die Modellvariante 3 und die anderen neu vorgestellten Maße verhielten sich diesbezüglich schlechter. Die Korrelationen der Modelle mit den Daten des dritten subjektiven Tests waren allerdings in einem Fall entscheidend anders: die Wirksamkeitsschätzungen auf Basis von Modellvariante 4 wiesen erhebliche Schätzfehler auf, während mit den Modellvarianten 1 und 2 sehr genaue Schätzungen möglich waren. Dies ist durch eine Überanpassung von Modellvariante 4 an den Lern- bzw. Trainingsdatensatz - also die Daten aus dem ersten subjektiven Test - zu erklären. Der gleiche Effekt tritt bei dem an die Aufgabenstellung dieser Arbeit angepassten Vergleichs-Faktorenmodell 1 auf.

Wie bereits in Kapitel 3 angemerkt, lassen sich komplexe Regressionsmodelle oft sehr gut an einen einzelnen Trainingsdatensatz anpassen, ermöglichen aber keine akkurate Schätzung in anderen Datensätzen, wie diese beiden Beispiele zeigen. Die Gefahr einer Überanpassung sollte daher bei der Entwicklung und Verwendung von Regressionsmodellen stets im Auge behalten werden. Sie ist eng mit der Plausibilität der Modelle verbunden. Gerade komplexe Regressionsmodelle sind bei genauer Betrachtung (z. B. was die Vorzeichen der Koeffizienten betrifft) oft unplausibel oder zumindest bezogen auf den betrachteten Anwendungszweck unplausibel. Modellvariante 4 des neu vorgestellten Qualitätsbewertungsansatzes ist unter den untersuchten Regressionsmodellen das komplexeste und gleichzeitig das Modell, das schon bei der statistischen Analyse große Plausibilitätszweifel aufgeworfen hatte. Insofern sind die Ergebnisse und Auffälligkeiten nicht überraschend. Auch beim Vergleichs-Faktorenmodell 1 musste letztlich mit Problemen gerechnet werden, da es keine speziell für die Aufgabenstellung entwickelten bzw. optimierten Maße enthält.

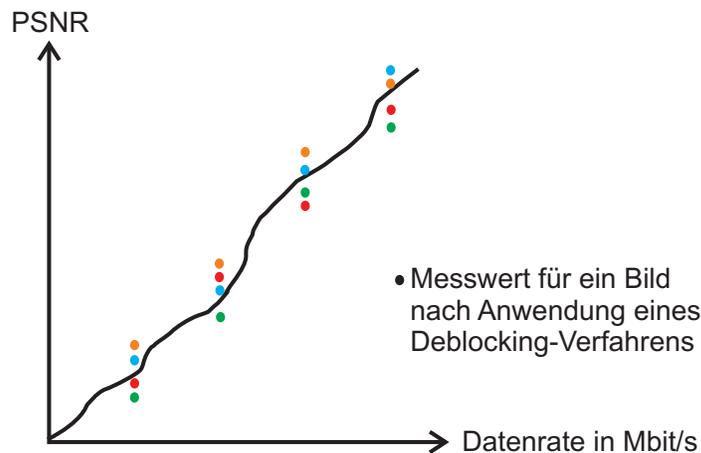
Wenn die Komplexität der Varianten des neu vorgestellten Qualitätsbewertungsansatzes im Vergleich betrachtet wird, hat sich in dieser Arbeit eine mittlere Modellkomplexität als optimal erwiesen, was mit der Darstellung in Abbildung 3.1 konform ist. Die größte Schätzgüte wurde insgesamt mit den Modellvarianten, die drei oder vier Qualitätsmaße enthalten, erzielt (Varianten 1 und 2), während sich Variante 4 - wie dargestellt - als überangepasst und suboptimal erwies. Die ausschließliche Verwendung eines einzelnen Qualitätsmaßes erwies sich im Vergleich mit den Varianten 1 und 2 ebenfalls als suboptimal, wenn auch in einigen Punkten als vorteilhaft gegenüber der Verwendung von Modellvariante 4.

Die speziell zur Abschätzung von Plausibilität und Stabilität der neu vorgestellten Qualitätsbewertungsmodelle durchgeführten Untersuchungen ergaben insbesondere bei den Modellvarianten 1 und 2 keine kritischen Probleme. Die Ähnlichkeit der auf Basis des ersten und des dritten subjektiven Tests berechneten Regressionskoeffizienten spricht für eine hohe Stabilität der beiden Varianten. Dies gilt nicht für Modellvariante 4 und nur eingeschränkt für Variante 3. Das Glättungsmaß alleine liefert in den meisten Fällen stabile Schätzwerte. Es konnte jedoch gezeigt werden, dass es sich problematisch verhält und falsche Schätzwerte liefert, wenn die Anwendung von Deblocking-Verfahren mit sehr starken Glättungen einhergeht. Die Modellvarianten 1 und 2 verhalten sich in solchen Fällen

deutlich unkritischer, während die Modellvarianten 3 und 4 diesbezüglich ebenfalls als problematisch einzuschätzen sind.

Die Schätzung der Wirksamkeit von Deblocking-Verfahren mit den vorgestellten Modellvarianten in P- und in B-Frames sowie in sehr wenig gestörten Bildern wurde betrachtet und ergab keine besonderen Unplausibilitäten. Es sei aber daran erinnert, dass die Modelle nicht auf diese Fälle hin optimiert sind und ein Einsatz in solchen Fällen generell fragwürdig ist. Die bestmögliche bzw. genaueste Wirksamkeitsschätzung wird gewährleistet, wenn Bildmaterial verwendet wird, das ein ausgewogenes Verhältnis von Konturen, Texturen und homogenen Bildbereichen aufweist und wenn die Schätzungen in I-Frames durchgeführt werden. Ein Maß an Artefakten, das für menschliche Betrachter ohne größere Anstrengungen wahrnehmbar ist, sollte vorhanden sein. Sind die Voraussetzungen näherungsweise erfüllt, ist insbesondere mit den Modellvarianten 1 und 2 eine akkurate und stabile Schätzung der Wirksamkeit von Deblocking-Verfahren bezogen auf eine Bildqualitätsverbesserung in einer breiten inhaltlichen Palette an Bildmaterial zu erwarten.

Um zu veranschaulichen, warum PSNR und MSE durchaus mit Einschränkungen geeignet sein können, um die absolute Qualität codierter Bilder oder Bildsequenzen zu schätzen (siehe z. B. [VQE00, VQE03, Esk00, MRLP<sup>+</sup>06]), jedoch kaum zur Bewertung der Wirksamkeit von Deblocking-Verfahren geeignet sind, ist in Abbildung 9.3 ein Diagramm abgedruckt, in dem qualitativ gezeigt wird, wie sich der PSNR eines Testbildes bei der Codierung mit verschiedenen Datenraten und bei Anwendung verschiedener Deblocking-Verfahren verhalten könnte.



**Abbildung 9.3:** Veranschaulichung des Verhaltens des PSNR bei Anwendung verschiedener Deblocking-Verfahren bei unterschiedlichen Datenraten

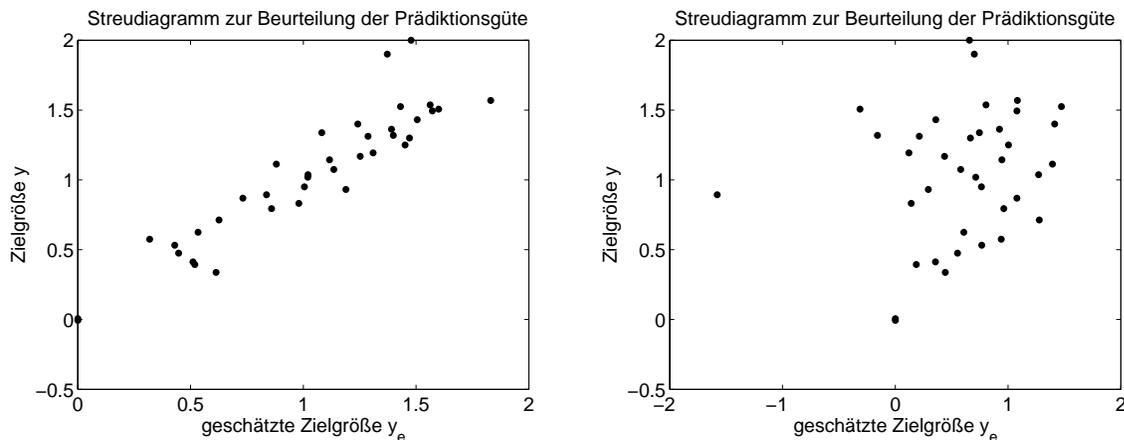
An der durchgezogenen Linie ist ablesbar, dass der PSNR tendenziell mit steigender Datenrate zunimmt. In der Praxis könnte diese Linie durch deutlichere Schwankungen

oder auch durch einen insgesamt gekrümmten Verlauf gekennzeichnet sein. Auch bei unterschiedlichen Bildern muss jeweils mit einer etwas unterschiedlichen Reaktion des PSNR auf eine Codierung mit verschiedenen Datenraten gerechnet werden (siehe Tabelle 2.1). Dies ändert jedoch nichts an der grundsätzlichen Tendenz des PSNR, mit steigender Datenrate größer zu werden. Die von menschlichen Betrachtern wahrgenommene Bildqualität wird mit steigender Datenrate ebenfalls zunehmen. Selbst wenn dies nicht in gleichem Maße der Fall ist wie die Zunahme des PSNR, wird jedoch klar, dass der PSNR die steigende Bildqualität mit zunehmender Datenrate mit nicht allzu großen Ungenauigkeiten beschreiben kann, weshalb er zu diesem Zweck nach wie vor oft eingesetzt wird.

Anders ist die Situation bei der Messung des PSNR nach Anwendung verschiedener Deblocking-Verfahren bei konstanter Datenrate. Die Messwerte liegen alle in einem kleinen Intervall, d. h. die Veränderungen des PSNR sind klein gegenüber den Veränderungen, die durch signifikant unterschiedliche Datenraten hervorgerufen werden. Die Anwendung von Deblocking-Verfahren führt zu einer Anordnung der Messwerte um die Messwerte der ungefilterten Bilder herum, welche auf der durchgezogenen Linie liegen. Viele Verfahren verbessern den PSNR, andere verschlechtern ihn jedoch. Das Ranking der Verfahren untereinander kann bei unterschiedlichen Datenraten schwanken. Hinzu kommt, dass die von Menschen empfundenen Bildqualitätsveränderungen nicht mit diesem Ranking übereinstimmen müssen, was schon allein daran abgelesen werden kann, dass Deblocking-Verfahren in der Regel die Bildqualität verbessern, selbst wenn der PSNR leicht abnimmt. Diese Überlegungen zeigen, warum der PSNR und auch der MSE zwar eine ungefähre Schätzung der absoluten Qualität codierten Bildmaterials über weite Datenratenbereiche ermöglichen, jedoch kaum dazu geeignet sind, die durch Deblocking-Verfahren verursachten Bildqualitätsveränderungen akkurat, streuungsarm und den Empfindungen menschlicher Betrachter hinreichend genau entsprechend zu erfassen.

Ähnliches gilt auch für das in die Untersuchungen einbezogene Einkanalmodell und das Vergleichs-Faktorenmodell 2. Diese Qualitätsbewertungsverfahren wurden speziell zur korrekten und mit der menschlichen Wahrnehmung möglichst gut übereinstimmenden Erfassung von Qualitätsunterschieden bei der Codierung von Bildsequenzen mit unterschiedlichen Datenraten und somit zur Erfassung verhältnismäßig großer Bildqualitätsunterschiede entwickelt bzw. daraufhin optimiert. Dementsprechend ermöglichen sie in diesem Kontext auch eine akkurate Qualitätsschätzung. Die zum Teil sehr geringen und nur im Detail unterschiedlichen Veränderungen, die durch die Anwendung von verschiedenen Deblocking-Verfahren an Bildern entstehen, werden durch die Verfahren hingegen nicht geeignet erfasst. Die Reaktionen der Maße auf die durch die Deblocking-Verfahren hervorgerufenen Veränderungen sind ähnlich wie beim PSNR oft klein. Sie weisen vergleichsweise große Streuungen auf oder entsprechen grundsätzlich nicht subjektiven Eindrücken, was letztlich zu schlechten Korrelationen und häufig zu großen Abweichungen zwischen „wahren“ Qualitätswerten und Schätzwerten bei der Wirksamkeitsbewertung von Deblocking-Verfahren führt. Die gradientenbasierten Qualitätsmetriken wurden zwar auch

zur Erfassung recht großer Bildqualitätsunterschiede, wie sie durch eine Codierung mit unterschiedlichen Datenraten verursacht werden, entwickelt, verhalten sich bei der Wirksamkeitsbewertung von Deblocking-Verfahren jedoch günstiger als z. B. MSE und PSNR, weil sie Merkmale analysieren, die im Kontext der Anwendung von Deblocking-Verfahren von Belang sind. Letztlich sind aber auch die untersuchten gradientenbasierten Metriken nicht auf die Aufgabenstellung in dieser Arbeit optimiert und ihre Schätzwerte weisen daher in dem genannten Zusammenhang mittelgroße Streuungen und somit Schätzgüten auf, die nur im Mittelfeld liegen. Die in dieser Arbeit entwickelten Qualitätsbewertungsansätze erfassen die verschiedenen Auswirkungen, die die Anwendung von Deblocking-Verfahren auf gestörte Bilder hat, hingegen sehr deutlich und genau, da sie Maße enthalten, die speziell daraufhin optimiert sind. Auch die neuen Verfahren weisen Streuungen auf, allerdings ist das Ausmaß der Streuungen im Vergleich zu den (gewünschten) Reaktionen auf die zu erfassenden Merkmalsveränderungen klein. Dadurch sind erheblich größere Korrelationen und kleinere Abweichungen zwischen „wahren“ Qualitätswerten und Schätzwerten als mit den untersuchten Vergleichsverfahren möglich, wie in diesem Kapitel gezeigt wurde. Abbildung 9.4 veranschaulicht dies exemplarisch anhand des Vergleichs der Streudiagramme der Zielgröße über der geschätzten Zielgröße für Modellvariante 1 und DPSNR (Datenbasis erster subjektiver Test). Während die einzelnen Punkte im Falle von Modellvariante 1 nur wenig um eine gedachte Gerade herum streuen, sind die Streuungen im Falle des DPSNR erheblich.



**Abbildung 9.4:** Streudiagramme zur Beurteilung der Schätzgüte der Modellvariante 1 (links) im Vergleich zum DPSNR (rechts)

Zusammenfassend kann gesagt werden, dass die Empfehlung aus [Fec99], spezielle Qualitätsmaße für spezielle Aufgabenstellungen zu verwenden, in dieser Arbeit bestätigt wird. Das den Untersuchungen zugrunde liegende Einkanalmodell und das Vergleichs-Faktorenmodell 2 sind z. B. gut oder sehr gut dazu geeignet (und auch dazu entwickelt bzw.

optimiert worden), unterschiedliche Qualitätsniveaus von durch Codierartefakte gestörten Bildsequenzen zu erfassen. Die Qualitätsniveaus werden durch verschiedene Datenraten in einem verhältnismäßig großen Bereich bestimmt. Mit Einschränkungen sind auch MSE und PSNR zur Erfassung großer Qualitätsunterschiede geeignet. All diese Ansätze zeigen jedoch Schwächen bei der Erfassung feiner Bildveränderungen bzw. Qualitätsunterschiede, wie sie bei der Wirksamkeitsbewertung von Deblocking-Verfahren von Belang sind. Die untersuchten gradientenbasierten Qualitätsmetriken können solche Bildveränderungen bzw. Qualitätsunterschiede zwar besser erfassen, allerdings zeigt sich, dass auch sie nicht daraufhin optimiert sind. Das an die Aufgabenstellung der Arbeit angepasste Vergleichs-Faktorenmodell 1 machte im ersten subjektiven Test zwar einen viel versprechenden Eindruck, im dritten Test zeigte sich jedoch, dass auch dieses Modell offenbar keine akkurate, streuungsarme und subjektiven Eindrücken entsprechende Erfassung relevanter Unterschiede zwischen codierten und nachgefilterten Bildern in einem breiten inhaltlichen Spektrum an Bildern ermöglicht. Die im Rahmen dieser Arbeit entwickelten Qualitätsbewertungsansätze sind hingegen speziell daraufhin optimiert, auf einem bestimmten (aber durchaus variablen) Qualitätsniveau kleinere bis moderate Bildmerkmalsveränderungen, die im Zusammenhang mit der Verringerung von Blockartefakten stehen, zuverlässig und genau zu erfassen. Resultierend daraus sind sie sehr gut zu dem bei ihrer Entwicklung zugrunde liegenden Zweck - der Bewertung der Wirksamkeit von Deblocking-Verfahren bezüglich einer Bildqualitätsverbesserung - geeignet. Zur absoluten Qualitätsbewertung codierter Bilder sind sie hingegen vollkommen ungeeignet bzw. lassen sich dazu gar nicht anwenden, da sie grundsätzlich nur vergleichende Qualitätswerte ermitteln können. Letztlich ist es nicht schwer nachvollziehbar, dass Verfahren, die zur Bewältigung einer bestimmten Aufgabe ausgelegt sind, sich für andere Aufgaben als wenig geeignet erweisen können. Dies gilt insbesondere dann, wenn die Aufgaben einen oder mehrere entscheidende Unterschiede aufweisen. Ein solcher Unterschied ist z. B. die Notwendigkeit der Erfassung recht großer Merkmals- bzw. Bildqualitätsunterschiede bei der absoluten Qualitätsbewertung codierter Bildmaterials im Gegensatz zur Notwendigkeit der Erfassung sehr feiner Merkmals- bzw. Bildqualitätsunterschiede bei der Wirksamkeitsbewertung von Deblocking-Verfahren.

# 10 Zusammenfassung und Ausblick

Die in der Praxis sehr bedeutsame Kompression von Bildern und Bildsequenzen geht insbesondere bei hohen Kompressionsgraden mit Störungen bei der Bildwiedergabe einher. Die wohl bedeutendsten und am häufigsten auftretenden Störungen in diesem Zusammenhang sind Blockartefakte, welche z. B. auch in DVB- und DVD-Systemen deutlich sichtbar werden können. Zur Beseitigung bzw. Reduktion von Blockartefakten existiert eine Vielzahl an Verfahren, die sehr unterschiedliche Eigenschaften haben. Mehrere dieser Verfahren wurden im Rahmen dieser Arbeit vorgestellt und untersucht. Darüber hinaus wurde ein neues adaptives Verfahren zur Reduktion von Blockartefakten entwickelt, welches sich insbesondere dadurch auszeichnet, dass es aufwandsgünstig implementierbar ist und im Gegensatz zu vielen anderen Verfahren keinerlei Zugriff auf Quantisierungsinformationen benötigt. Es ist somit als reines Postprocessing-Verfahren für Endgeräte wie z. B. LC-Displays geeignet. Verschiedene Varianten des Verfahrens, die alle eine deutliche Reduktion von Blockartefakten ermöglichen, wurden vorgestellt. In einer speziellen Variante gelingt durch eine Kombination aus globaler und lokaler Bildanalyse eine besonders gute Anpassung des Verfahrens an die Stärke der im Bild vorhandenen Artefakte, so dass eine deutliche Artefaktreduktion bei gleichzeitig guter Detailerhaltung insbesondere auch in wenig gestörten Bildern möglich ist.

Von besonderem Interesse im Zusammenhang mit Deblocking-Verfahren generell bzw. bei der Auswahl eines geeigneten Verfahrens für einen bestimmten Anwendungszweck ist die Wirksamkeit der Verfahren bezogen auf eine Verbesserung der Gesamtbildqualität. Eine Abschätzung kann durch äußerst aufwändige subjektive Tests oder mit Hilfe von objektiven Qualitätsmaßen erfolgen. Als objektive Qualitätsmaße werden in der Literatur zu diesem Zweck in den meisten Fällen der PSNR und vereinzelt gradientenbasierte Ansätze verwendet. In dieser Arbeit wurde untersucht, inwieweit mit unterschiedlichen bekannten Verfahren zur Bildqualitätsbewertung eine Schätzung der Wirksamkeit von Deblocking-Verfahren in Einzelbildern möglich ist, die in guter Übereinstimmung mit den Ergebnissen subjektiver Tests ist. Es zeigte sich, dass teilweise enorme Abweichungen zwischen geschätzten und tatsächlich von menschlichen Betrachtern wahrgenommenen Wirksamkeiten bezogen auf Gesamtbildqualitätsverbesserungen bestehen. Dies gilt insbesondere für PSNR und MSE, aber auch für weitere in die Untersuchungen einbezogene Bildqualitätsbewertungsansätze, die in der Literatur zuvor erfolgreich zur Bewertung der absoluten Qualität komprimierter Bilder oder Bildfolgen eingesetzt wurden.

Ein wesentlicher Grund, warum sich manche Qualitätsbewertungsansätze zur Erfassung der absoluten Qualität von komprimiertem Bildmaterial gut eignen, bei der Be-

wertung der Wirksamkeit von Verfahren zur Verringerung von Codierartefakten - einer nicht unverwandten Aufgabe - jedoch erhebliche Schwächen aufweisen, liegt in entscheidenden Unterschieden bei den zu erfassenden Qualitäts- bzw. Störmerkmalen. Verfahren zur absoluten Qualitätsschätzung von komprimiertem Bildmaterial sind in der Regel auf eine akkurate Qualitätsschätzung über weite Qualitäts- bzw. Datenratenbereiche ausgelegt. Eine signifikante Qualitäts- bzw. Datenratenveränderung geht in Bildern auch mit deutlichen Veränderungen verschiedenster Merkmale einher. Auf die Erfassung solcher Merkmale sind die Qualitätsbewertungsverfahren entsprechend optimiert. Die Anwendung von Deblocking-Verfahren führt hingegen an Bildern insgesamt oft nur zu sehr geringen Veränderungen, die aber sehr selektiv und auf bestimmte (Stör-)Merkmale konzentriert sind. Qualitätsbewertungsverfahren, welche zur absoluten Qualitätsbewertung komprimierter Bilder oder Bildfolgen entwickelt wurden, sowie PSNR und MSE erfassen die relevanten Merkmale im Kontext einer Bildqualitätsveränderung durch Deblocking oft nicht oder zu ungenau (bzw. mit zu starken Streuungen versehen). Um die von menschlichen Betrachtern wahrnehmbaren Unterschiede der Wirksamkeit von Deblocking-Verfahren akkurat schätzen zu können, sind Ansätze notwendig, die stark auf die Erfassung der für diesen Zweck relevanten Bild- bzw. Störmerkmale in ihrer entsprechenden Ausprägung optimiert sind. Es wird klar, dass für spezielle Qualitätsbewertungsaufgaben auch speziell angepasste Qualitätsmaße verwendet werden sollten, wenn sehr hohe Schätzgüten gefordert sind.

Aus diesem Grund wurde im Rahmen der vorliegenden Arbeit ein neuer Ansatz zur Bewertung der Wirksamkeit von Deblocking-Verfahren in Einzelbildern entwickelt und untersucht. Es wurden zunächst mehrere Qualitätsmaße entwickelt, die speziell an die Eigenschaften von Störungen, welche im Zusammenhang mit blockbasierter hybrider Transformationscodierung und der Beseitigung von Codierartefakten auftreten, angepasst sind. Von großer Bedeutung sind diesbezüglich Diskontinuitäten an Blockgrenzen, glatte Helligkeitsverläufe in Bildbereichen mit wenig Aktivität und die Degradation von Konturen oder Bilddetails. Darüber hinaus wurde untersucht, inwieweit Kombinationen mehrerer Qualitätsmaße in Form von Faktorenmodellen die Genauigkeit bei der Schätzung der Wirksamkeit von Deblocking-Verfahren sowie weitere Eigenschaften der Qualitätsschätzung verbessern können. Es konnten auf Basis umfangreicher linearer Regressionsanalysen mehrere Modellvarianten entwickelt werden, die zur Bewertung der Auswirkungen von Deblocking-Verfahren auf die Gesamtbildqualität gut geeignet sind. Diese beinhalten sowohl Maße, welche auf die Erfassung von Bildqualitätsverbesserungen durch die Anwendung von Artefaktreduktionsverfahren abzielen, als auch Maße, die auf die Erfassung von Bildqualitätsverschlechterungen abzielen. Andere Kombinationen, die z. B. ausschließlich Maße beinhalten, die Bildqualitätsverbesserungen erfassen, erwiesen sich aus statistischen oder aus Stabilitätsgründen als problematisch.

Die Entwicklung und Auswahl geeigneter Qualitätsmaße für kombinierte Qualitätsbewertungsmodelle stellte sich insgesamt als ein sehr aufwändiger Prozess dar, da die Eigenschaften einer Kombination während der Entwicklungsphase kaum abgeschätzt und

---

erst im komplett implementierten Modell untersucht werden können. Allerdings bieten kombinierte Qualitätsbewertungsansätze in Form von Faktorenmodellen eine sehr große Flexibilität, die eine Anpassung an diverse Qualitätsbewertungsaufgaben erlaubt, wenn ein recht hoher Entwicklungsaufwand in Kauf genommen wird. Vorteilhaft ist es, wenn in einem Faktorenmodell Maße für sehr unterschiedliche relevante Merkmale kombiniert werden können und nicht mehrere Maße für recht ähnliche Merkmale verwendet werden müssen. Im letztgenannten Fall, der in dieser Arbeit gegeben war, tritt der Effekt auf, dass sich meist nur Faktorenmodelle mit einer recht kleinen Zahl an signifikanten Teilmaßen finden lassen. Das bedeutet, dass z. B. durch Kombination von drei bis vier Qualitätsmaßen die Schätzgenauigkeit eines Qualitätsbewertungsverfahrens gegenüber der ausschließlichen Verwendung eines einzigen Maßes zwar erhöht werden kann, die Aufnahme weiterer Maße jedoch zu keiner signifikanten Erhöhung mehr führt. Schuld daran sind Abhängigkeiten zwischen den Qualitätsmaßen, so genannte Multikollinearitäten.

Grundsätzlich muss bei der Kombination von Maßen stets auf Plausibilität der entstehenden Modelle geachtet werden. Es kann sonst passieren, dass Modelle zu sehr an den Trainingsdatensatz angepasst werden, sich bei Anwendung der Modelle zur Qualitätsschätzung in anderen Daten aber erhebliche Ungenauigkeiten oder Instabilitäten zeigen. Insbesondere die Anwendung der schrittweisen Regression birgt diesbezüglich Gefahren, was in den durchgeführten Untersuchungen bestätigt werden konnte. Die geschilderte Problematik kann bei im Umfang begrenzten Trainingsdatensätzen besonders deutlich zum Tragen kommen.

Die Korrelationen der Schätzwerte der in dieser Arbeit entwickelten und für detaillierte Untersuchungen ausgewählten Qualitätsbewertungsmodelle mit den subjektiven Eindrücken menschlicher Betrachter liegen bei ca. 0,92. Sie sind damit erheblich größer als die auf Basis des MSE und PSNR erzielbaren Korrelationen von günstigstenfalls 0,35 bzw. 0,4. Auch in die Untersuchungen einbezogene gradientenbasierte Blocking-Maße, welche vereinzelt in der Literatur zur Bewertung von Deblocking-Verfahren eingesetzt wurden, erreichen diese Schätzgüte bei Weitem nicht. Die mittleren Abweichungen zwischen geschätzten und wahrgenommenen Wirksamkeiten bzgl. der Verbesserung der Gesamtbildqualität betragen bei den neuen Modellen nur ca. 0,16 Skaleneinheiten auf einer DMOS-Skala von 0...3. Demgegenüber betragen die mittleren Abweichungen 0,73 Skaleneinheiten bei Verwendung des PSNR und sogar 0,91 Skaleneinheiten bei Verwendung des MSE. Die Datenbasis für die angegebenen Zahlen wurde im Rahmen von drei umfangreichen subjektiven Tests ermittelt.

Die neuen Ansätze zur Qualitätsbewertung eignen sich insgesamt zur Erfüllung ihres Zwecks erheblich besser als bislang verwendete Qualitätsmaße. Sie können eine besonders wichtige Eigenschaft von Deblocking-Verfahren, deren Wirksamkeit bei der Qualitätsverbesserung unterschiedlich stark gestörter Bilder, objektiv und akkurat schätzen. Wenn allerdings durch die neuen Qualitätsbewertungsansätze die Wirksamkeit von zwei Verfahren sehr ähnlich eingeschätzt wird, sollten weitere Qualitätsmaße oder die sub-

jektiven Eindrücke eines oder mehrerer Experten hinzugezogen werden, um eine sichere Wirksamkeitseinschätzung zu gewährleisten. In einem solchen Fall bringt auch ein kleiner Schätzfehler die Gefahr von Fehlinterpretationen mit sich. Werden bei der Auswahl eines Deblocking-Verfahrens für einen bestimmten Anwendungszweck neben der Wirksamkeit auch Kriterien wie Rechenaufwand, Speicheraufwand, Unabhängigkeit der Verfahren von Seiteninformationen etc. berücksichtigt, dürften mittlere Abweichungen im Bereich von 0,16 Skaleneinheiten aber auch in einem solchen Fall hinreichend klein sein.

Um die Schätzgüten der neuen Ansätze zur Bewertung der Wirksamkeit von Artefaktreduktionsverfahren weiter zu verbessern, sind Erweiterungen denkbar. Bislang ist es empfehlenswert, Testbilder zu verwenden, die ein ausgewogenes Maß an Texturen, Konturen und homogenen Bereichen enthalten. In Testbildern, die diesbezüglich Missverhältnisse aufweisen, ist mit Unschärfen bei der Qualitätsschätzung zu rechnen. Eine Verbesserung könnte dadurch erzielt werden, dass in den Faktorenmodellen mehrere Koeffizientensätze zur Gewichtung der integrierten Qualitätsmaße für bestimmte Bildklassen vorgehalten werden, welche über eine verhältnismäßig primitive globale Bildklassifizierung ausgewählt werden könnten. Damit wäre es möglich, eine optimale Anpassung der Modelle an ein besonders breites Spektrum an Bildmaterial zu gewährleisten. Dies erscheint in Bezug auf eine weitere Verbesserung der Schätzgüte vielversprechender als der Versuch, ein sehr breites Spektrum an Bildmaterial bei der Berechnung eines einzigen Koeffizientensatzes zu berücksichtigen.

Ebenfalls über unterschiedliche Koeffizientensätze oder möglicherweise über eine einfache Skalierung von Koeffizienten könnte eine Anpassung der Faktorenmodelle an leicht veränderte Aufgabenstellungen erfolgen. Darunter sind z. B. veränderte Betrachtungsabstände - hier wurde die dreifache Bildhöhe verwendet - oder Unterschiede bei der wahrgenommenen Wirksamkeit von Deblocking-Verfahren auf unterschiedlichen Displaytypen - hier wurde ein LCD verwendet - zu verstehen. Letzlich wäre für solche Anpassungen im Wesentlichen die Durchführung weiterer subjektiver Tests notwendig.

Eine weitere Verbesserung der Schätzgenauigkeit der entwickelten Faktorenmodelle könnte durch die Berücksichtigung von Eigenschaften des menschlichen visuellen Systems in den verwendeten Qualitätsmaßen erzielt werden. Von Bedeutung könnten der Frequenzgang des HVS, Maskierungseffekte in Bildbereichen hoher Aktivität und Nichtlinearitäten beim Kontrastempfinden sein. Es wäre denkbar, solche Effekte beispielsweise bei der Bewertung der Verringerung bzw. Glättung von Diskontinuitäten an Blockgrenzen sowie bei der Bewertung der Beeinträchtigung von Konturen zu berücksichtigen. Ein in dieser Arbeit zu Vergleichszwecken untersuchtes Blocking-Maß, welches Eigenschaften des HVS berücksichtigt, wies allerdings keine bessere Schätzgüte als ein einfacheres Blocking-Maß ohne Berücksichtigung solcher Eigenschaften auf. Klarheit über die Vor- und Nachteile der Berücksichtigung von Eigenschaften des HVS können allerdings nur weitere Untersuchungen bringen.

---

Besonders interessante Möglichkeiten zur Erweiterung der neuen Faktorenmodelle bieten sich durch Anpassungen bzw. Erweiterungen mit dem Ziel der expliziten Bewertung der Wirksamkeit von Deblocking-Verfahren in Bildsequenzen. Dazu müssten einerseits die unterschiedliche Ausprägung von Artefakten und die unterschiedliche Wirksamkeit von Artefaktreduktionsverfahren in verschiedenen Frame-Typen berücksichtigt werden, andererseits auch zeitliche Maskierungseffekte, die im menschlichen visuellen System wirksam werden oder durch die Darstellung auf einem Display bedingt sind. Des Weiteren müsste beachtet werden, dass menschliche Betrachter in ihrer Wahrnehmung oft besonders stark und nachhaltig auf schlechte Bildqualität, die nur in verhältnismäßig kurzen Zeiträumen - also z. B. in einzelnen Bildern einer Bildsequenz - gegeben ist, reagieren. In diesem Zusammenhang sind weitere Untersuchungen möglich und wünschenswert.

Die in dieser Arbeit gewählten Ansätze zur Bildqualitätsbewertung bieten die notwendige Flexibilität, um auch für andere Aufgaben eingesetzt werden zu können. Auf der Hand liegt z. B. die Bewertung der Wirksamkeit von Deringing-Verfahren. Dazu müssten ein oder mehrere geeignete Ringing-Maße entwickelt werden. Ideen dazu werden in Anhang D vorgestellt. Es wäre auch denkbar, Blocking-, Ringing- und Degradationsmaße in ein einziges Faktorenmodell zu integrieren, um die Wirksamkeit von Artefaktreduktionsverfahren bei der Verringerung unterschiedlicher Artefakte sowie die ungewünschten Nebenwirkungen erfassen und bewerten zu können.

Auch über die Bewertung von Verfahren zur Reduktion von Codierartefakten hinaus sind Einsatzmöglichkeiten der hier vorgestellten Qualitätsbewertungsansätze denkbar. Entscheidend ist, dass sich im Kontext des Einsatzes mehrere Merkmale finden lassen, die mit geeigneten Qualitätsmaßen analysiert werden können. Dabei gibt es grundsätzlich keine Vorgaben bezüglich der von den Qualitätsmaßen benötigten Daten (Referenzbilder, gestörte Bilder, nachgefilterte Bilder etc.). Vorstellbar wäre z. B. die Kombination mehrerer Rausch- und Degradationsmaße zu einem Faktorenmodell, das die Schätzung der Wirksamkeit von Rauschreduktionsverfahren zum Ziel hat. In diesem Zusammenhang könnten auch die Auswirkungen von Rauschen und Rauschreduktionsverfahren auf verschiedene Bildbereiche (homogen, detailliert usw.) berücksichtigt werden. Ebenfalls denkbar wären Faktorenmodelle zur Bewertung von Deinterlacing-Verfahren. Bedeutende Merkmale in diesem Zusammenhang könnten Zähnchenstrukturen, Unschärfbildung und auch zeitliche Merkmale wie z. B. Flackerstörungen sein.



# Literaturverzeichnis

- [AB00] ALKACHOUH, Z. und M.G. BELLANGER: *Fast DCT-Based Spatial Domain Interpolation of Blocks in Images*. IEEE Transactions on Image Processing, Vol. 9, No. 4, S. 729–732, April 2000.
- [ANT92] AVRIL, C. und T. NGUYEN-TRONG: *Linear filtering for reducing blocking effects in orthogonal transform image coding*. IEEE Signal Processing Magazine, Vol. 14, Issue 2, März 1992.
- [AP04] ASCENSO, J. und F. PEREIRA: *H.264/AVC Fine Grain Scalability using Bit-plane Coding*. 5th International Workshop on Image Analysis for Multimedia Interactive Services, Lissabon, April 2004.
- [ASS02] AVCIBAS, I., B. SANKUR und K. SAYOOD: *Statistical evaluation of image quality measures*. Journal of Electronic Imaging, Vol. 11(2), April 2002.
- [BBK05] BARKOWSKY, M., J. BIALKOWSKI und A. KAUP: *Subjektiver Videobetrachtungstest für niederratige Multimedia-Szenarien*. ITG-Fachbericht Elektronische Medien 188 (11. Dortmunder Fernsehseminar), S. 169–174. VDE-Verlag, September 2005.
- [BGK04] BLUME, V., K. GRÜNEBERG und M. KRÜTZMANN: *Videosignalverarbeitung für hochauflösende Großflächen-Matrixdisplays*. Fernseh- und Kinotechnik, 57. Jahrgang, Nr. 4, 2004.
- [BK97] BANHAM, M.R. und A.K. KATSAGGELOS: *Digital image restoration*. Journal of Electronic Imaging, Vol. 1(2), S. 24–41, April 1997.
- [Bre06] BREMENKAMP, D.: *Objektive Bewertung der Qualität von gestörten Bildsequenzen mit Hilfe des Faktorenmodells*. Studienarbeit S04-2005 am AG Schaltungen der Informationsverarbeitung der Technischen Universität Dortmund, 2006.
- [BS95] BRONSTEIN, I.N. und K.A. SEMENDJAJEW: *Taschenbuch der Mathematik*. Verlag Harri Deutsch, 1995.
- [Bud72] BUDRIKIS, Z.L.: *Visual fidelity Criterium and Modeling*. Proceedings of IEEE, Vol. 60, S. 771–779, 1972.

- [CG02] CHAN, R.W.-S. und P. GOLDSMITH: *Modeling and validation of a psychovisually based image quality evaluator for DCT-based compression*. Signal Processing: Image Communication, Vol. 17, No. 6, S. 485–495, Juli 2002.
- [CWQ01] CHEN, T., H.R. WU und B. QIU: *Adaptive Postfiltering of Transform Coefficients for the Reduction of Blocking Artifacts*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 5, S. 594–602, Mai 2001.
- [Dal93] DALY, S.: *The visible differences predictor: An algorithm for the assessment of image fidelity*. In: Watson, A.: Digital Images and Human Vision, S. 179–206. MIT Press, 1993.
- [dH00] HAAN, G. DE: *Video Processing for Multimedia Systems*. University Press Eindhoven, 2000.
- [DKL<sup>+</sup>05] DOLAR, C., G. KOHLMAYER, S. LENKE, P. PIASTOWSKI und H. SCHRÖDER: *Entwicklungstrends der digitalen Videosignalverarbeitung*. 11. Dortmunder Fernsehseminar, S. 23–36, September 2005.
- [Dol03] DOLAR, C.: *Entwicklung einer Sequenzanlage auf Basis eines Standard-PC*. Studienarbeit S04-2003 am AG Schaltungen der Informationsverarbeitung der Technischen Universität Dortmund, 2003.
- [EBU75] *E.B.U. Standard for Chromaticity Tolerances for Studio Monitors*. Tech. 3213-E, 1975.
- [EF95] ESKICIOGLU, A.M. und P.S. FISHER: *Image Quality Measures and their Performance*. IEEE Transactions on Communications, Vol. 43, No. 12, S. 2959–2965, Dezember 1995.
- [EH04] ELPELT, B. und J. HARTUNG: *Grundkurs Statistik*. Oldenbourg, 2004.
- [Esk00] ESKICIOGLU, A.M.: *Quality measurement for monochrome compressed images in the past 25 years*. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Istanbul, S. 1907–1910, Juni 2000.
- [Fec99] FECHTER, F.: *Technische Bildqualitätsmessung - Ein Vergleich alternativer Methoden*. Tagungsband des 8. Dortmunder Fernsehseminars, September 1999.
- [Fec00] FECHTER, F.: *Zur Beurteilung der Qualität komprimierter Bildfolgen*. VDI-Verlag, 2000.

- 
- [FH84] FAHRMEIR, L. und A. HAMERLE: *Multivariate statistische Verfahren*. Walter de Gruyter, 1984.
- [FK06] FRIEBE, M. und A. KAUP: *3D-Deblocking for Error Concealment in Block-Based Video Decoding Systems*. Picture Coding Symposium, Peking, April 2006.
- [FKPT04] FAHRMEIR, L., R. KÜNSTLER, I. PIGEOT und G. TUTZ: *Statistik - Der Weg zur Datenanalyse*. Springer, 2004.
- [Frä98] FRÄNTI, P.: *Blockwise distortion measure for statistical and structural errors in digital images*. Signal Processing: Image Communication, Vol. 13, S. 89–98, 1998.
- [GAM02] GUNTURK, B.K., Y. ALTUNBASAK und R.M. MERSEREAU: *Multiframe Blocking-Artifact Reduction for Transform-Coded Video*. IEEE Transactions On Circuits and Systems for Video Technology, Vol. 12, No. 4, April 2002.
- [GI94] GRAYBILL, F.A. und H.K. IYER: *Regression Analysis*. Duxbury Press, 1994.
- [Har02] HARTUNG, J.: *Statistik*. Oldenbourg, 2002.
- [HCH<sup>+</sup>03] HUANG, Y.-W., T.-W. CHEN, B.-Y. HSIEH, T.-C. WANG, T.-H. CHANG und L.-G. CHEN: *Architecture Design for Deblocking Filter in H.264/JVT/AVC*. IEEE International Conference on Multimedia and Expo, S. 693–696, Juli 2003.
- [Hei99] HEIKEL, S.: *Bildqualitätsbewertung auf Basis eines Einkanalmodells der visuellen Wahrnehmung*. Studienarbeit S14-99 am AG Schaltungen der Informationsverarbeitung der Technischen Universität Dortmund, 1999.
- [Hen98] HENTSCHEL, CH.: *Video-Signalverarbeitung*. B. G. Teubner, 1998.
- [HKJP01] HWANG, I.K., S.J. KIM, W.C. JOUNG und J.K. PAIK: *Constraint-Reduced Regularization for Reducing Blocking Artifacts in Compressed Video*. Proceedings of the 11th International Packet Video Workshop, Kyungju, Korea, S. 309–313, April 2001.
- [HMM97] HORITA, Y., M. MIYAHARA und T. MURAI: *Objective Picture Quality Scale for Monochrome Video*. Proceedings of the Picture Coding Symposium, Berlin, S. 123–128, September 1997.

- [HN94] HAMADA, H. und S. NAMBA: *A Study on Objective Picture Quality Scales for Pictures Digitally Encoded for Broadcast*. IEICE Transactions on Communications, Vol. E77-B, No. 12, S. 1480–1488, Dezember 1994.
- [ISO92] *Digital compression and coding of continuous-tone still images*. ISO/IEC 10918-1/ITU-T Rec. T.81, 1992.
- [ISO93] *Information technology - coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 2: Video*. ISO/IEC 11172-2, 1993.
- [ISO96] *Information technology: generic coding of moving pictures and associated audio information: Video*. ISO/IEC 13818-2 / ITU-T Rec. H.262, 1996.
- [ISO98] *Information technology: coding of audio-visual objects - Part 2: visual*. ISO/IEC 14996-2, 1998.
- [ISO00] *Information technology: coding of audio-visual objects - Part 2: visual*. ISO/IEC 14496-2, Amendment 1: Visual Extensions, 15.7.2000.
- [ISO02] *Overview of the MPEG-4 Standard*. ISO/IEC JTC1/SC29/WG11 N4668, März 2002.
- [ISO03] *Coding of audio-visual objects - Part 10: Advanced Video Coding*. ISO/IEC 14496-10:2003 / ITU-T Recommendation H.264, 2003.
- [ISO04] *Information technology: coding of audio-visual objects - Part 2: visual*. ISO/IEC 14996-2, third edition, 1.6.2004.
- [ITU02] *Methodology for the subjective assessment of the quality of television pictures*. ITU-T Recommendation BT.500-11, 2002.
- [Jos01] JOSTSCHULTE, K.: *Rauschreduktion in Bildsignalen mit teilbandzerlegenden Medianfiltern*. Shaker, 2001.
- [Kau00] KAUP, A.: *Image restoration for frame- and object-based video coding using an adaptive constrained least-squares approach*. Signal Processing, Vol. 80, No. 11, S. 2337–2345, 2000.
- [KIL04] KWON, K.-K., S.-H. IM und D.-S. LIM: *Picture quality improvement in MPEG-4 video coding using simple adaptive filter*. Proceedings of the 12th annual ACM international conference on Multimedia, New York, S. 284–287, Oktober 2004.

- 
- [Kin81] KING, M.L.: *The Durbin-Watson Bounds Test and Regressions without an Intercept*. Australian Economic Papers, Vol. 20, No. 36, S. 161–170, Juni 1981.
- [KJKH98] KIM, N.C., I.H. JANG, D.H. KIM und W.H. HONG: *Reduction of Blocking Artifact in Block-Coded Images Using Wavelet Transform*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, No. 3, S. 253–257, Juni 1998.
- [KK95] KARUNASEKERA, S.A. und N.G. KINGSBURY: *A Distortion Measure for Blocking Artifacts in Images based on Human Visual Sensitivity*. IEEE Transactions on Image Processing, Vol. 4, No. 6, S. 713–724, Juni 1995.
- [Kle03] KLEPPMANN, W.: *Taschenbuch Versuchsplanung - Produkte und Prozesse optimieren*. Hanser, 2003.
- [KP98] KIM, T.K. und J.K. PAIK: *Fast Image Restoration for Reducing Block Artifacts based on Adaptive Constrained Optimization*. Journal of Visual Communication and Image Representation, Vol. 9, No. 3, S. 234–242, September 1998.
- [Kra71] KRAMER, G.: *On the Durbin-Watson Bounds Test in the Case of Regression Through the Origin*. Jahrbücher für Nationalökonomie und Statistik, Band 185, S. 345–358, 1971.
- [KYKR99] KIM, S.D., J. YI, H.M. KIM und J.B. RA: *A Deblocking Filter with Two Separate Modes in Block-Based Video Coding*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 1, Februar 1999.
- [KYL<sup>+</sup>04] KIM, Y.-H., J.W. YOO, S.-W. LEE, J. SHIN, J. PAIK und H.K. JUNG: *Adaptive mode decision for H.264 encoder*. Electronic Letters, Vol. 40, No. 19, September 2004.
- [LB02] LIU, S. und A.C. BOVIK: *Efficient DCT-Domain Blind Measurement and Reduction of Blocking Artifacts*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12, No. 12, S. 1139–1149, Dezember 2002.
- [LCPH96] LUO, J., C.W. CHEN, K.J. PARKER und T.S. HUANG: *Artifact Reduction in Low Bit Rate DCT-Based Image Compression*. IEEE Transactions on Image Processing, Vol. 5, No. 9, S. 1363–1368, September 1996.
- [Lim79] LIMB, J.O.: *Distortion criteria of the human viewer*. IEEE Transactions on Systems Man and Cybernetics, Vol. 9, S. 778–793, 1979.

- [LJL<sup>+</sup>03] LIST, P., A. JOCH, J. LAINEMA, G. BJONTEGAARD und M. KARCZEWICZ: *Adaptive Deblocking Filter*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, Juli 2003.
- [LR05] LEONTARIS, A. und A.R. REIBMAN: *Comparison of blocking and blurring metrics for video compression*. IEEE International Conference on Acoustics, Speech and Signal Processing, Philadelphia, USA, Vol. 2, S. 585–585, März 2005.
- [Mal99] MALLAT, S.: *A wavelet tour of signal processing - second edition*. Academic Press, 1999.
- [MH04] MEFFERT, B. und O. HOCHMUTH: *Werkzeuge der Signalverarbeitung*. Pearson Studium, 2004.
- [Miy88] MIYAHARA, M.: *Quality Assessments for Visual Service*. IEEE Communications Magazine, S. 51–60, Oktober 1988.
- [MK03] MEISINGER, K. und A. KAUP: *Örtliche Fehlerverschleierung von gestört empfangenen Bilddaten durch frequenzselektive Extrapolation*. Tagungsband des 10. Dortmunder Fernsehseminars, S. 189–194, September 2003.
- [MRLP<sup>+</sup>06] MARTINEZ-RACH, M., O. LOPEZ, P. PINOL, , M.P. MALUMBRES und J. OLIVER: *A Study of Objective Quality Assessment Metrics for Video Codec Design and Evaluation*. Proceedings of the 8th IEEE International Symposium on Multimedia, S. 517–524, Dezember 2006.
- [MSS] *MPEG Software Simulation Group (MSSG)*.  
<http://www.mpeg.org/MPEG/MSSG/>.
- [MZ92] MALLAT, S. und S. ZHONG: *Characterization of signals from multiscale edges*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, S. 710–732, Juli 1992.
- [NH95] NETRAVALI, A.N. und B.G. HASKELL: *Digital Pictures - Representation, Compression and Standards, Second Edition*. Plenum Press, 1995.
- [Nis88] NISS, B.: *Prediction of AC coefficients from the DC values*. ISO/IEC JTC1/SC2/WG8 N745, Mai 1988.
- [OBL<sup>+</sup>04] OSTERMANN, J., J. BORMANS, P. LIST, D. MARPE, M. NARROSCHKE, F. PEREIRA, T. STOCKHAMMER und T. WEDI: *Video Coding with H.264/AVC: Tools, Performance and Complexity*. IEEE Circuits and Systems Magazine, First Quarter, 2004.

- 
- [Ohm95] OHM, J.-R.: *Digitale Bildcodierung - Repräsentation, Kompression und Übertragung von Bildsignalen*. Springer, 1995.
- [OS95] O’ROURKE, T.P. und R.L. STEVENSON: *Improved Image Decompression for Reduced Transform Coding Artifacts*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 5, No. 6, Dezember 1995.
- [OSB04] OPPENHEIM, A.V., R.W. SCHAFER und J.R. BUCK: *Zeitdiskrete Signalverarbeitung*. Pearson Studium, 2004.
- [Pia05a] PIASTOWSKI, P.: *Adaptive Reduktion von Blockartefakten mit lokaler Bildanalyse*. 3. Thüringer Medienseminar, Erfurt, Juni 2005.
- [Pia05b] PIASTOWSKI, P.: *System zur Decoder-unabhängigen Reduktion von Blockartefakten*. 11. Dortmunder Fernsehseminar, S. 215–218, September 2005.
- [PK04] PIASTOWSKI, P. und G. KOHLMAYER: *Neues Verfahren zur kombinierten örtlich/zeitlichen Rauschreduktion und verbesserte Bewegungsschätzung für neue Display-Technologien*. FKTG Jahrestagung, Koblenz, Mai 2004.
- [PKL97] PARK, J.W., J.W. KIM und S.U. LEE: *DCT Coefficients Recovery-Based Error Concealment Technique and its Application to the MPEG-2 Bit Stream Error*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No. 6, S. 845–854, Dezember 1997.
- [Plo89] PLOMPEN, R.: *Motion Video Coding for Visual Telephony*. CIP - Gegevens Koninklijke Bibliotheek, 1989.
- [PS04] PROAKIS, J.G. und M. SALEHI: *Grundlagen der Kommunikationstechnik*. Pearson Studium, 2004.
- [RB00] RAMAN, A. und M. BABU: *A Low Complexity Error Concealment Scheme for MPEG-4 Coded Video Sequences*. 10. Annual Symposium on Multimedia Communications and Signal Processing, S. 729–732, November 2000.
- [RH96] RAO, K.R. und J.J. HWANG: *Techniques and Standards for Image, Video and Audio Coding*. Prentice Hall, 1996.
- [Ric03] RICHARDSON, I.E.G.: *Video Codec Design - Developing Image and Video Compression Systems*. Wiley, 2003.
- [Ric04] RICHARDSON, I.E.G.: *H.264 and MPEG-4 Video Compression - Video Coding for Next-generation Multimedia*. Wiley, 2004.

- [RK00] RAMKISHOR, K. und P. KARANDIKAR: *A Simple and Efficient Deblocking Algorithm for Low Bit-Rate Video Coding*. IEEE International Symposium on Consumer Electronics, Hongkong, China, Dezember 2000.
- [RL83] REEVE, H.C. und J.S. LIM: *Reduction of blocking effects in image coding*. Proceedings of ICASSP, Boston, S. 1212–1215, April 1983.
- [SB00] SCHRÖDER, H. und H. BLUME: *Mehrdimensionale Signalverarbeitung - Band 2: Architekturen und Anwendungen für Bilder und Bildsequenzen*. B.G. Teubner, 2000.
- [SCA99] SHIN, T., K. CHO und B.-H. AHN: *Block effect reduction with content-based AC prediction in an MPEG-2 compressed video*. International Conference on Consumer Electronics, S. 232–233, Juni 1999.
- [Sch95] SCHMITTE, A.: *Konzeption und Implementierung einer objektorientierten Bibliothek zur digitalen Bildverarbeitung*. Studienarbeit S16-1994 am Lehrstuhl Nachrichtentechnik der Technischen Universität Dortmund, 1995.
- [Sch98] SCHRÖDER, H.: *Mehrdimensionale Signalverarbeitung - Band 1: Algorithmische Grundlagen für Bilder und Bildsequenzen*. B.G. Teubner, 1998.
- [Sch03] SCHLITTEGEN, R.: *Einführung in die Statistik*. Oldenbourg, 2003.
- [Sch04] SCHUCHERT, T.: *Concealmentverfahren zur Störungsreduktion in digital übertragenen/gespeicherten Videosequenzen*. Studienarbeit S4-2004 am AG Schaltungen der Informationsverarbeitung der Technischen Universität Dortmund, 2004.
- [Sha48] SHANNON, C.E.: *A mathematical theory of communication*. Bell System Technical Journal, Vol. 27, 1948.
- [Sie04] SIEBERT, A.: *Wavelet-basierte Störungsreduktion in Videosequenzen*. Diplomarbeit D3-2004 am AG Schaltungen der Informationsverarbeitung der Technischen Universität Dortmund, 2004.
- [SK98] SHEN, M.-Y. und C.-C.J. KUO: *Review of Postprocessing Techniques for Compression Artifact Removal*. Journal of Visual Communication and Image Representation, Vol. 9, No. 1, S. 2–14, März 1998.
- [SK00] SEGALL, C. und A. KATSAGGELOS: *Pre- and post-processing algorithms for compressed video enhancement*. Proceedings of the Asilomar Conference on Signals and Systems, S. 1369–1373, Oktober 2000.

- 
- [SL99] SHYU, H. und J. LEOU: *Detection and Concealment of Transmission Errors in MPEG-2 Images - A Genetic Algorithm Approach*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 6, S. 937–948, September 1999.
- [Sto93] STOYAN, D.: *Stochastik für Ingenieure und Naturwissenschaftler*. Akademie Verlag, 1993.
- [Str05] STRUTZ, T.: *Bilddatenkompression*. Vieweg, 2005.
- [Sut98] SUTHAHARAN, S.:  *$L_\infty$ -based Distortion Measure for Blocking Artifacts in Digital Video*. IEEE International Caracas Conference on Devices, Circuits and Systems, Venezuela, S. 235–238, 1998.
- [Sut03] SUTHAHARAN, S.: *A perceptually significant block-edge impairment metric for digital video coding*. Proceedings of the 2003 International Conference on Multimedia and Expo, Vol. 1, S. 585–588, 2003.
- [SW77] SAVIN, N.E. und K.J. WHITE: *The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors*. Econometrica, Vol. 45, No. 8, S. 1989–1996, November 1977.
- [SW97] SUTHAHARAN, S. und H.R. WU: *A Distortion Measure for Blocking Artifacts in Digital Video*. International Conference on Information, Communications and Signal Processing, Singapore, S. 1566–1569, September 1997.
- [SWLG02] SUN, X., F. WU, S. LI und W. GAOS: *In-loop deblocking filter for block-based video coding*. International Conference on Signal Processing (ICSP), Peking, Vol. 1, S. 33–36, August 2002.
- [TCS05] TAI, S.-C., Y.-Y. CHEN und S.-F. SHEU: *Deblocking Filter for Low Bit Rate MPEG-4 Video*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 6, S. 733–741, Juni 2005.
- [TG00] TAN, K.T. und M. GHANBARI: *A Multi-Metric Objective Picture-Quality Measurement Model for MPEG Video*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 7, S. 1208–1213, Oktober 2000.
- [Tie87] TIEDE, M.: *Statistik - Regressions- und Korrelationsanalyse*. Oldenbourg, 1987.
- [Tra98] TRAUBERG, M.: *Ein neues Verfahren für die Bildqualitätsüberwachung*. FKT, S. 411–416, 1998.

- [Tra01] TRAUBERG, M.: *Ein Verfahren zur Qualitätsbewertung datenreduzierter Bildfolgen*. Shaker Verlag, 2001.
- [UM06] URBAN, D. und J. MAYERL: *Regressionsanalyse: Theorie, Technik und Anwendung*. VS Verlag für Sozialwissenschaften, 2006.
- [VQE00] *Validation of Objective Models of Video Quality Assessment*. Final Report from the Video Quality Experts Group, Juni 2000.
- [VQE03] *Validation of Objective Models of Video Quality Assessment, Phase II*. Final Report from the Video Quality Experts Group, August 2003.
- [Wen88] WENDLAND, B.: *Fernsehtechnik - Band 1: Grundlagen*. Hüthig, 1988.
- [Wen06] WENTKER, M.: *Bildverbesserung mit Restoration-Verfahren für neue Displays*. Diplomarbeit D18-2005 am AG Schaltungen der Informationsverarbeitung der Technischen Universität Dortmund, 2006.
- [Win05] WINKLER, S.: *Digital video quality - vision models and metrics*. Wiley, 2005.
- [WJ99] WEIHS, C. und J. JESSENBERGER: *Statistische Methoden zur Qualitätssicherung und -optimierung*. Wiley-VCH, 1999.
- [WLYQ96] WU, H., C. LAMBRECHT, M. YUEN und B. QIU: *Quantitative Quality and Impairment Metrics for Digitally Coded Images and Image Sequences*. Proceedings of Australian Telecommunication Networks and Applications Conference, Dezember 1996.
- [WR06] WU, H.R. und K.R. RAO: *Digital video image quality and perceptual coding*. CRC Press, 2006.
- [WS91] WENDLAND, B. und H. SCHRÖDER: *Fernsehtechnik - Band 2: Systeme und Komponenten zur Farbbildübertragung*. Hüthig, 1991.
- [WSM01] WINKLER, S., A. SHARMA und D. MCNALLY: *Perceptual video quality and blockiness metrics for multimedia streaming applications*. Proceedings of the International Symposium on Wireless Personal Multimedia Communications, Aalborg, Denmark, S. 547–552, September 2001.
- [Wu94] WU, X.: *Synthetische Kantenversteilerung zur Verbesserung der Bildschärfe*. VDI-Verlag, 1994.
- [WY97] WU, H.R. und M. YUEN: *A Generalized Block-Edge Impairment Metric for Video Coding*. IEEE Signal Processing Letters, Vol. 4, No. 11, 1997.

- 
- [WYT01] WU, S., H. YAN und Z. TAN: *An efficient wavelet-based deblocking algorithm for highly compressed images*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, Issue 11, S. 1193–1198, November 2001.
- [WZF04] WANG, C., W.-J. ZHANG und X.-Z. FANG: *Adaptive reduction of blocking artifacts in DCT domain for highly compressed images*. IEEE Transactions on Consumer Electronics, Vol. 50, Issue 2, S. 647–654, Mai 2004.
- [XOZ97] XIONG, Z., M.T. ORCHARD und Y.-Q. ZHANG: *A deblocking algorithm for JPEG compressed images using overcomplete wavelet representations*. IEEE Transactions on Circuits and Systems for Video Technology, 7, Issue 2, S. 433–437, April 1997.
- [YCK00] YANG, J., H. CHOI und T. KIM: *Noise Estimation for Blocking Artifacts Reduction in DCT Coded Images*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 7, Oktober 2000.
- [YG97] YANG, Y. und N.P. GALATSANOS: *Removal of Compression Artifacts Using Projections onto Convex Sets and Line Process Modeling*. IEEE Transactions on Image Processing, Vol. 6, No. 10, Oktober 1997.
- [YW98] YUEN, M. und H.R. WU: *A survey of hybrid MC/DPCM/DCT video coding distortions*. Signal Processing, 70, S. 247–278, 1998.
- [YWWC02] YU, Z., H.R. WU, S. WINKLER und T. CHEN: *Vision-model-based impairment metric to evaluate blocking artifacts in digital video*. Proceedings of the IEEE, Vol. 90, Issue 1, S. 154–169, Januar 2002.



# Abbildungsverzeichnis

2.1	Typisches Bildverarbeitungssystem . . . . .	5
2.2	Blockschaltbild eines JPEG-Encoders . . . . .	8
2.3	Aufbau eines hybriden Video-Encoders, Inter-Codierung, angelehnt an [Str05]	11
2.4	Aufbau eines hybriden Video-Decoders . . . . .	12
2.5	Struktur eines MPEG-Streams . . . . .	12
2.6	Bildtypen und Übertragungsreihenfolge . . . . .	13
2.7	Frame DCT Coding und Field DCT Coding . . . . .	14
2.8	Beispiel für die Definition von Video-Objekten . . . . .	16
2.9	Störungsquellen in einem digitalen TV-System . . . . .	18
2.10	Beispiele für Codierartefakte . . . . .	21
2.11	DCT-Basisbilder . . . . .	25
2.12	Approximation einer periodischen Rechteckfunktion mit Hilfe von fünf Co-sinusschwingungen . . . . .	27
2.13	Ausschnitt aus Frame 24 (Intra-Frame) der Sequenz <i>football</i> , codiert mit verschiedenen Datenraten . . . . .	29
2.14	Ausschnitt aus Frame 24 (Intra-Frame) der Sequenz <i>wheel</i> , codiert mit verschiedenen Datenraten . . . . .	30
2.15	Ausschnitt aus Frame 24 der Sequenz <i>football24stat</i> , codiert mit verschiedenen Datenraten . . . . .	31
3.1	Veranschaulichung des Verhaltens des Schätzfehlers in Abhängigkeit von der Modellkomplexität . . . . .	46
4.1	Übersicht über Verfahren zur Bildqualitätsbewertung . . . . .	49
4.2	Lage von Haupt- und Nebengradienten . . . . .	54
4.3	Einkanalmodell - Blockschaltbild . . . . .	56
4.4	Bandaufspaltung in Frequenz- und Orientierungsbänder . . . . .	58
4.5	Faktorenmodell nach [HMM97] . . . . .	59
4.6	Faktorenmodell nach [Fec00] . . . . .	60
5.1	Überblick über Artefaktreduktionsverfahren . . . . .	65
5.2	Filterung von Blockgrenzen zur Artefaktreduktion . . . . .	70
5.3	Bildbeispiel für die lineare Filterung von Blockgrenzen (Ausschnitt) . . . . .	71
5.4	Artefaktreduktion durch Subtraktion von Fehlersignalen . . . . .	71

---

5.5	Ermittelte Fehlervektoren $\hat{b}_p$ . . . . .	73
5.6	Bildbeispiel für die Artefaktreduktion durch Subtraktion von Fehlersignalen (Ausschnitt) . . . . .	74
5.7	Reduktion von Blockartefakten mit zwei Filtermodi . . . . .	74
5.8	Beispiel für die Lage der DCTs bei der Filterung im default mode . . . . .	75
5.9	Bildbeispiel für die Artefaktreduktion mit zwei Filtermodi nach [KYKR99] (Ausschnitt) . . . . .	76
5.10	Bildbeispiel für die Artefaktreduktion mit zwei Filtermodi nach [RK00] (Ausschnitt) . . . . .	78
5.11	Artefaktreduktion auf Basis der Filterung von Wavelet-Koeffizienten, [KJKH98] . . . . .	78
5.12	Bildbeispiel für die Artefaktreduktion auf Basis der Filterung von Wavelet-Koeffizienten (Ausschnitt) . . . . .	80
6.1	Reduktion von Blockartefakten mit lokaler Bildanalyse . . . . .	82
6.2	Haupt- und Nebengradienten bei der lokalen Bildanalyse . . . . .	83
6.3	Mehrheitlich vertikal verlaufende Blockartefakte in der Sequenz <i>radsport</i> . . . . .	84
6.4	Stellen, an denen eine Analysebreite von vier Pixeln vorteilhaft ist . . . . .	85
6.5	Filter für homogene Bildbereiche . . . . .	86
6.6	Filter für detaillierte Bildbereiche . . . . .	87
6.7	Veranschaulichung der Basisvektoren einer 4-Punkt-DCT . . . . .	88
6.8	Bildbeispiel für die Artefaktreduktion mit lokaler Bildanalyse, acht Pixel Analysebreite (Ausschnitt) . . . . .	90
6.9	Bildbeispiel für die Artefaktreduktion mit lokaler Bildanalyse, vier Pixel Analysebreite (Ausschnitt) . . . . .	90
6.10	Reduktion von Blockartefakten mit lokaler/globaler Bildanalyse und drei Filtermodi . . . . .	92
6.11	Filter für homogene und stark homogene Bildbereiche . . . . .	94
6.12	Haupt- und Nebengradienten bei der lokalen Bildanalyse . . . . .	95
6.13	Bildbeispiel für die Artefaktreduktion mit lokaler/globaler Bildanalyse und drei Filtermodi (Ausschnitt) . . . . .	96
6.14	Veranschaulichung der Adaptivität der Blocking-Reduktion mit drei Filtermodi; weiß/mittelgrau/dunkelgrau: Anwendung des Filters für stark homogene/weniger stark homogene/detaillierte Bildbereiche; schwarz: keine Filterung . . . . .	98
7.1	Aufbau des subjektiven Tests . . . . .	101
7.2	In den subjektiven Tests eingesetzte Vergleichsskala . . . . .	103
7.3	Ergebnisse der subjektiven Tests 1 und 2, Reihenfolge siehe Tabelle A.2 . . . . .	112
7.4	Bewertungen der einzelnen Deblocking-Verfahren bei 3,5 Mbit/s . . . . .	120

---

7.5	Bildausschnitte aus Frame 24 der Sequenz <i>wheel</i> , bearbeitet mit verschiedenen Deblocking-Verfahren . . . . .	121
7.6	Bildausschnitte aus Frame 48 der Sequenz <i>lena2_zoom2</i> , bearbeitet mit verschiedenen Deblocking-Verfahren . . . . .	122
7.7	Bildausschnitte aus Frame 12 der Sequenz <i>stockholm4</i> , bearbeitet mit verschiedenen Deblocking-Verfahren . . . . .	123
8.1	Allgemeines Faktorenmodell zur Bildqualitätsbewertung und zum Vergleich von Bildverarbeitungsalgorithmen . . . . .	129
8.2	Funktionsweise des Diskontinuitätsmaßes . . . . .	133
8.3	Veranschaulichung der Gradientenlagen . . . . .	134
8.4	Sättigungskennlinie und Bestrafungsfunktion im Diskontinuitätsmaß, Abszisse: Eingangswerte, Ordinate: Ausgangswerte . . . . .	136
8.5	Funktionsweise des Glättungsmaßes, NOP $\hat{=}$ no operation . . . . .	138
8.6	Beispiel für die Lage der Analysevektoren und der Gradienten . . . . .	139
8.7	Funktionsweise des Konturdegradationsmaßes . . . . .	141
8.8	Funktionsweise des Degradationsmaßes . . . . .	143
8.9	Streudiagramme zur statistischen Analyse der Modellvariante 1 . . . . .	150
8.10	QQ-Plot zur statistischen Analyse der Modellvariante 1 . . . . .	151
8.11	Streudiagramme zur statistischen Analyse der Modellvariante 2 . . . . .	153
8.12	Streudiagramm der Zielgröße über Einflussgröße 4 und Streudiagramm der Zielgröße über der geschätzten Zielgröße, Modellvariante 2 . . . . .	154
8.13	QQ-Plot zur statistischen Analyse der Modellvariante 2 . . . . .	155
8.14	QQ-Plot und Streudiagramm von Zielgröße über geschätzter Zielgröße, Modellvariante 3 . . . . .	156
9.1	Streudiagramme zur Beurteilung der Schätzgüte der Modellvarianten 1 und 2 im dritten subjektiven Test . . . . .	170
9.2	Bildbeispiel zur Veranschaulichung der unterschiedlichen Ausprägung von Artefakten in unterschiedlichen Frame-Typen . . . . .	173
9.3	Veranschaulichung des Verhaltens des PSNR bei Anwendung verschiedener Deblocking-Verfahren bei unterschiedlichen Datenraten . . . . .	179
9.4	Streudiagramme zur Beurteilung der Schätzgüte der Modellvariante 1 (links) im Vergleich zum DPSNR (rechts) . . . . .	181
C.1	Kombinierter Gradienten-Laplace-Kantendetektor . . . . .	216
E.1	Im Rahmen der Arbeit verwendete Bilder bzw. Sequenzen . . . . .	222
E.2	Im Rahmen der Arbeit verwendete Bilder bzw. Sequenzen . . . . .	225



# Tabellenverzeichnis

2.1	Blocking- und PSNR-Messwerte bei unterschiedlichen Datenraten . . . . .	32
6.1	PSNR-Vergleich mehrerer Varianten der Artefaktreduktion mit lokaler (und globaler) Bildanalyse (in dB), Messungen im Luminanzkanal von Frame 24 . . . . .	91
7.1	Kennzahlen verschiedener Displays, Längenangaben in cm . . . . .	101
7.2	Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Gesamtqualitätsunterschiede . . . . .	115
7.3	Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Qualitätsunterschiede in Bezug auf Blockartefakte . . . . .	117
7.4	Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Schärfeunterschiede . . . . .	118
7.5	Vergleich der wahrgenommenen Qualitätsunterschiede in Einzelbildpaaren und in Sequenzpaaren . . . . .	124
7.6	Subjektives Ranking der Bilder bzw. Verfahren im dritten subjektiven Test, Bewertung der Gesamtqualitätsunterschiede . . . . .	125
8.1	Regressionskoeffizienten der Modellvariante 1 . . . . .	148
8.2	Regressionskoeffizienten der Modellvariante 2 . . . . .	151
8.3	Regressionskoeffizienten der Modellvariante 3 . . . . .	155
9.1	Korrelationen der Qualitätsmaße . . . . .	162
9.2	Güte der Wirksamkeitsschätzung im ersten subjektiven Test . . . . .	163
9.3	Güte der Wirksamkeitsschätzung bei Trennung der Datensätze zur Berechnung der Regressionskoeffizienten und der Gütemaße . . . . .	167
9.4	Zahlen zur Güte der Wirksamkeitsschätzung im dritten subjektiven Test . . . . .	169
9.5	Regressionskoeffizienten der Modellvariante 1, ermittelt auf Basis der Daten aus dem ersten und dem dritten subjektiven Test im Vergleich . . . . .	171
9.6	Regressionskoeffizienten der Modellvariante 2, ermittelt auf Basis der Daten aus dem ersten und dem dritten subjektiven Test im Vergleich . . . . .	172
9.7	Regressionskoeffizienten der Modellvariante 3, ermittelt auf Basis der Daten aus dem ersten und dem dritten subjektiven Test im Vergleich . . . . .	172
9.8	Bewertung der Wirksamkeit von Deblocking-Verfahren in verschiedenen Frame-Typen durch Modellvariante 1, Angabe des DMOS . . . . .	174

9.9	Bewertung der Wirksamkeit einer 5x5 Mittelwertfilterung bei der Blocking-Reduktion durch verschiedene Verfahren, Angabe des DMOS . . . . .	175
A.1	Untersuchte Deblocking-Verfahren . . . . .	207
A.2	Playlist des ersten subjektiven Tests . . . . .	208
A.3	Playlist des zweiten subjektiven Tests . . . . .	209
A.4	Playlist des dritten subjektiven Tests . . . . .	210
B.1	Gütwerte für die in Kapitel 9 untersuchten Wirksamkeitsbewertungsverfahren, Zusammenfassung der Tabellen 9.2 bis 9.4 . . . . .	211
B.2	Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Gesamtbildqualitätsunterschiede, 3,5 Mbit/s . . . . .	212
B.3	Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Gesamtbildqualitätsunterschiede, 5 Mbit/s . . . . .	212
B.4	Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Qualitätsunterschiede in Bezug auf Blockartefakte, 3,5 Mbit/s . . . . .	213
B.5	Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Qualitätsunterschiede in Bezug auf Blockartefakte, 5 Mbit/s . . . . .	213
B.6	Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Schärfeunterschiede, 3,5 Mbit/s . . . . .	214
B.7	Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Schärfeunterschiede, 5 Mbit/s . . . . .	214

# A Playlists

Im Folgenden sind die Playlists der drei durchgeführten subjektiven Tests abgedruckt. Die Namen der Sequenz- bzw. Bildpaare sind folgendermaßen zu interpretieren:

Zunächst ist der Name der ursprünglichen Sequenz (z. B. *wheel*) angegeben, gefolgt von der Datenrate, mit der die Sequenz komprimiert wurde (z. B. 3.5m für 3,5 Mbit/s), und dem angewendeten Deblocking-Algorithmus (z. B. *avril*). *L* oder *R* kennzeichnen, auf welcher Seite eines Paares die nachgefilterte Version des Bildes bzw. der Sequenz in den subjektiven Tests zu sehen war (links/rechts). Die Deblocking-Algorithmen wurden nach dem erstgenannten Autor in der zugehörigen Veröffentlichung benannt. In Tabelle A.1 sind Einzelheiten zu den verschiedenen Verfahren aufgeführt und es ist angegeben, wo die Algorithmen in dieser Arbeit vorgestellt werden und welche Literaturquelle eine ausführliche Beschreibung enthält. Nur im ersten subjektiven Test wurden einige Sequenzen gezeigt (an den letzten acht Positionen der Playlist für den ersten Test), ansonsten ausschließlich Einzelbilder. Als Einzelbilder wurden stets dieselben Frames aus den entsprechenden Sequenzen verwendet. Im Falle von *stockholm4* war dies Frame 12, im Falle von *football*, *pedestrian3* und *wheel* war es Frame 24, und im Falle von *fries2*, *lena2\_zoom2* und *shields3* war es Frame 48. Die Tabellen A.2, A.3 und A.4 enthalten die Playlists der durchgeführten subjektiven Tests. Da die gezeigten Bildpaare in den ersten beiden Tests in Reihenfolge und Position nicht identisch waren, jedoch bis auf eine Ausnahme in beiden Tests vorkamen, ist in Tabelle A.3 auch angegeben, an welcher Position das entsprechende Bildpaar im ersten Test gezeigt wurde. Zu Bildpaar 45 gab es keine Entsprechung im ersten Test. Wann immer in dieser Arbeit Bildpaarnummern, die sich auf die ersten beiden Tests beziehen, genannt werden, gelten die Positionsnummern des ersten subjektiven Tests.

**Tabelle A.1:** Untersuchte Deblocking-Verfahren

Name	Funktionsweise	Kapitel	Quelle
3x3	statische 3x3 Mittelwertfilterung von Bildern	-	-
avril	statische Filterung von Blockgrenzen	5	[ANT92]
kim	adaptive Filterung von Blockgrenzen, zwei Filtermodi	5	[KYKR99]
kim98	wavelet-basierte Blocking-Reduktion	5	[KJKH98]
pi4b	Blocking-Reduktion mit lokaler Bildanalyse und zwei Filtermodi	6	[Pia05a]
pi7	Blocking-Reduktion mit lokaler/globaler Bildanalyse und drei Filtermodi	6	-
ramkishor	adaptive Filterung von Blockgrenzen, zwei Filtermodi	5	[RK00]
tai	adaptive Filterung von Blockgrenzen, drei Filtermodi	6	[TCS05]
yang	Subtraktion von Fehlersignalen	5	[YCK00]

Tabelle A.2: Playlist des ersten subjektiven Tests

Nr.	Sequenz
1	football3.5m_avril_L
2	wheel3.5m_kim_R
3	stockholm4.3.5m_yang_L
4	football3.5m_kim_R
5	wheel3.5m_avril_L
6	football5m_ramkishor_L
7	football3.5m_kim98_R
8	wheel3.5m_avril_L
10	stockholm4.3.5m_yang_L
11	football3.5m_L
12	wheel3.5m_kim98_R
13	wheel5m_kim_L
14	lena2_zoom2.3.5m_avril_L
15	football3.5m_pi4b_R
16	wheel5m_kim98_R
17	football3.5m_avril_L
18	stockholm4.3.5m_kim_L
19	lena2_zoom2.3.5m_kim98_R
20	football3.5m_pi7_L
21	wheel5m_L
22	football3.5m_kim_R
23	wheel3.5m_pi7_L
24	stockholm4.3.5m_avril_R
25	wheel5m_avril_R
26	football5m_yang_R
27	wheel3.5m_yang_R
28	lena2_zoom2.3.5m_kim_L
29	football5m_pi7_L
30	wheel5m_yang_R
31	lena2_zoom2.3.5m_pi7_L
32	wheel3.5m_pi4b_L
33	football3.5m_ramkishor_R
34	stockholm4.3.5m_pi7_R
35	football3.5m_yang_L
36	wheel5m_pi7_R
37	football5m_avril_R
38	stockholm4.3.5m_ramkishor_R
39	wheel5m_ramkishor_L
40	lena2_zoom2.3.5m_ramkishor_R
41	football5m_kim98_R
42	stockholm4.3.5m_kim98_L
44	wheel3.5m_ramkishor_R
45	football5m_kim_L
46	lena2_zoom2.3.5m_yang_L
48	football12m_intra_avril_R
49	wheel12m_intra_yang_L
50	wheel12m_intra_pi7_R
51	football12m_intra_pi7_L
52	wheel12m_intra_kim_R
53	football12m_intra_kim_R
54	wheel12m_intra_avril_L
55	football12m_intra_yang_L

Tabelle A.3: Playlist des zweiten subjektiven Tests

Nr.	Nr. T1	Sequenz
1	1	football3.5m_avril_L
2	2	wheel3.5m_kim_R
3	3	stockholm4.3.5m_yang_L
4	4	football3.5m_kim_R
5	5	wheel3.5m_avril_L
6	6	football5m_ramkishor_L
7	7	football3.5m_kim98_R
8	8	wheel3.5m_avril_L
9	10	stockholm4.3.5m_yang_L
10	11	football3.5m_L
11	12	wheel3.5m_kim98_R
12	13	wheel5m_kim_L
13	14	lena2_zoom2_3.5m_avril_L
14	15	football3.5m_pi4b_R
15	16	wheel5m_kim98_R
16	17	football3.5m_avril_L
17	18	stockholm4.3.5m_kim_L
18	19	lena2_zoom2_3.5m_kim98_R
19	20	football3.5m_pi7_L
20	21	wheel5m_L
21	22	football3.5m_kim_R
22	23	wheel3.5m_pi7_L
23	24	stockholm4.3.5m_avril_R
24	25	wheel5m_avril_R
25	26	football5m_yang_R
26	27	wheel3.5m_yang_R
27	28	lena2_zoom2_3.5m_kim_L
28	29	football5m_pi7_L
29	30	wheel5m_yang_R
30	31	lena2_zoom2_3.5m_pi7_L
31	32	wheel3.5m_pi4b_L
32	33	football3.5m_ramkishor_R
33	34	stockholm4.3.5m_pi7_R
34	35	football3.5m_yang_L
35	36	wheel5m_pi7_R
36	37	football5m_avril_R
37	38	stockholm4.3.5m_ramkishor_R
38	39	wheel5m_ramkishor_L
39	40	lena2_zoom2_3.5m_ramkishor_R
40	41	football5m_kim98_R
41	42	stockholm4.3.5m_kim98_L
42	44	wheel3.5m_ramkishor_R
43	45	football5m_kim_L
44	46	lena2_zoom2_3.5m_yang_L
45	-	wheel3.5m_kim_R

Tabelle A.4: Playlist des dritten subjektiven Tests

Nr.	Sequenz
1	fries2_1.8m_avril_L
2	pedestrian3_1.8m_avril_R
3	pedestrian3_1.8m_pi7_R
4	fries2_1.8m_ramkishor_L
5	shields3_3m_kim_R
6	fries2_1.8m_yang_R
7	pedestrian3_1.8m_ramkishor_R
8	shields3_3m_avril_R
9	pedestrian3_1.8m_yang_R
10	shields3_3m_kim98_L
11	fries2_1.8m_tai_R
12	pedestrian3_1.8m_kim98_L
13	shields3_3m_ramkishor_L
14	fries2_1.8m_kim98_R
15	pedestrian3_1.8m_L
16	shields3_3m_kim_R
17	fries2_1.8m_pi7_L
18	pedestrian3_1.8m_avril_R
19	fries2_1.8m_kim_L
20	shields3_3m_tai_L
21	pedestrian3_1.8m_pi7_R
22	fries2_1.8m_3x3_L
23	shields3_3m_yang_L
24	fries2_1.8m_L
25	pedestrian3_1.8m_tai_L
26	shields3_3m_pi7_R
27	fries2_1.8m_avril_L
28	pedestrian3_1.8m_kim_L
29	shields3_3m_3x3_R
30	pedestrian3_1.8m_3x3_L
31	fries2_1.8m_ramkishor_L

# B Ergänzende Ergebnisse

Ergänzend zu den in Kapitel 9 abgedruckten Gütewerten für die dort untersuchten Verfahren zur Wirksamkeitsbewertung von Deblocking-Algorithmen fasst Tabelle B.1 die Tabellen 9.2 bis 9.4 zusammen. Für Details zu den angegebenen Werten sei entsprechend auf Kapitel 9 verwiesen.

**Tabelle B.1:** Gütewerte für die in Kapitel 9 untersuchten Wirksamkeitsbewertungsverfahren, Zusammenfassung der Tabellen 9.2 bis 9.4

Maß/System	Kor. 9.2	Kor. 9.3	Kor. 9.4	Abw. 9.2	Abw. 9.3	Abw. 9.4
Glättungsmaß	0,913	0,912	0,878	0,154	0,160	0,169
Modellvariante 1	0,926	0,916	0,921	0,130	0,145	0,159
Modellvariante 2	0,936	0,920	0,926	0,123	0,145	0,163
Modellvariante 3	0,863	0,845	0,858	0,164	0,182	0,274
Modellvariante 4	0,948	0,928	0,571	0,108	0,133	0,346
GBIM [WY97]	0,560	0,548	0,666	0,345	0,357	0,326
Blocking [Pia05a]	0,689	0,662	0,778	0,245	0,253	0,217
DMSE	0,269	0,267	0,349	0,657	0,677	0,907
DPSNR	0,240	0,248	0,401	0,551	0,574	0,731
Vergleichs-Faktorenmodell 1	0,904	0,867	0,400	0,159	0,188	0,803

Die folgenden Tabellen zeigen als Ergänzung zu den Tabellen in Kapitel 7 nach Datenrate getrennt, wie die Teilnehmer der ersten beiden subjektiven Tests die Qualitätsunterschiede zwischen den Bildern der gezeigten Bildpaare bewertet haben. Es ist der DMOS angegeben. Tabelle B.2 enthält die Bewertungen der Gesamtbildqualitätsunterschiede, wobei nur Bilder berücksichtigt sind, welche aus Sequenzen stammen, die mit einer Datenrate von 3,5 Mbit/s codiert wurden. Die Tabelle ist so sortiert, dass sich die größte wahrgenommene Qualitätsverbesserung unten befindet. Tabelle B.3 enthält entsprechend Bewertungen für Bilder aus Sequenzen, die mit 5 Mbit/s codiert wurden. Die Tabellen B.4 und B.5 enthalten die Bewertungen der Qualitätsunterschiede speziell bezogen auf die von den Probanden wahrgenommenen Blocking-Störungen. Entsprechend enthalten die Tabellen B.6 und B.7 die Bewertungen der von den Probanden wahrgenommenen Schärfeunterschiede zwischen den Bildern der gezeigten Bildpaare. Die Bildpaare mit den größten wahrgenommenen Schärfeunterschieden befinden sich in den Tabellen ganz unten.

**Tabelle B.2:** Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Gesamtbildqualitätsunterschiede, 3,5 Mbit/s

Nr.	Bild	Qualität
11	football3.5m_L	0.01
24	stockholm4_3.5m_avril_R	0.07
(2)	(wheel3.5m_kim_R)	(0.77)
14	lena2_zoom2_3.5m_avril_L	0.79
17	football3.5m_avril_L	0.87
44	wheel3.5m_ramkishor_R	0.93
46	lena2_zoom2_3.5m_yang_L	0.95
10	stockholm4_3.5m_yang_L	1.02
27	wheel3.5m_yang_R	1.04
8	wheel3.5m_avril_L	1.11
35	football3.5m_yang_L	1.14
19	lena2_zoom2_3.5m_kim98_R	1.17
38	stockholm4_3.5m_ramkishor_R	1.19
31	lena2_zoom2_3.5m_pi7_L	1.25
34	stockholm4_3.5m_pi7_R	1.30
42	stockholm4_3.5m_kim98_L	1.31
40	lena2_zoom2_3.5m_ramkishor_R	1.32
12	wheel3.5m_kim98_R	1.36
32	wheel3.5m_pi4b_L	1.40
33	football3.5m_ramkishor_R	1.43
15	football3.5m_pi4b_R	1.49
7	football3.5m_kim98_R	1.51
23	wheel3.5m_pi7_L	1.53
22	football3.5m_kim_R	1.54
20	football3.5m_pi7_L	1.57
18	stockholm4_3.5m_kim_L	1.90
28	lena2_zoom2_3.5m_kim_L	2.00

**Tabelle B.3:** Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Gesamtbildqualitätsunterschiede, 5 Mbit/s

Nr.	Bild	Qualität
21	wheel5m_L	-0.01
26	football5m_yang_R	0.34
39	wheel5m_ramkishor_L	0.39
16	wheel5m_kim98_R	0.41
37	football5m_avril_R	0.48
30	wheel5m_yang_R	0.53
25	wheel5m_avril_R	0.58
13	wheel5m_kim_L	0.63
36	wheel5m_pi7_R	0.71
6	football5m_ramkishor_L	0.83
41	football5m_kim98_R	0.89
45	football5m_kim_L	1.08
29	football5m_pi7_L	1.34

**Tabelle B.4:** Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Qualitätsunterschiede in Bezug auf Blockartefakte, 3,5 Mbit/s

Nr.	Bild	Qualität
11	football3.5m_L	-0.13
24	stockholm4_3.5m_avril_R	0.33
10	stockholm4_3.5m_yang_L	0.68
46	lena2_zoom2_3.5m_yang_L	0.70
14	lena2_zoom2_3.5m_avril_L	0.76
44	wheel3.5m_ramkishor_R	0.85
17	football3.5m_avril_L	0.86
31	lena2_zoom2_3.5m_pi7_L	0.91
-	wheel3.5m_kim_R	0.97
38	stockholm4_3.5m_ramkishor_R	1.03
12	wheel3.5m_kim98_R	1.04
27	wheel3.5m_yang_R	1.09
40	lena2_zoom2_3.5m_ramkishor_R	1.10
8	wheel3.5m_avril_L	1.13
19	lena2_zoom2_3.5m_kim98_R	1.14
18	stockholm4_3.5m_kim_L	1.23
34	stockholm4_3.5m_pi7_R	1.23
35	football3.5m_yang_L	1.24
23	wheel3.5m_pi7_L	1.33
42	stockholm4_3.5m_kim98_L	1.38
32	wheel3.5m_pi4b_L	1.41
22	football3.5m_kim_R	1.66
15	football3.5m_pi4b_R	1.72
33	football3.5m_ramkishor_R	1.72
7	football3.5m_kim98_R	1.73
20	football3.5m_pi7_L	1.74
28	lena2_zoom2_3.5m_kim_L	1.74

**Tabelle B.5:** Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Qualitätsunterschiede in Bezug auf Blockartefakte, 5 Mbit/s

Nr.	Bild	Qualität
21	wheel5m_L	-0.12
39	wheel5m_ramkishor_L	0.45
37	football5m_avril_R	0.49
13	wheel5m_kim_L	0.50
16	wheel5m_kim98_R	0.59
25	wheel5m_avril_R	0.64
30	wheel5m_yang_R	0.72
36	wheel5m_pi7_R	0.74
26	football5m_yang_R	0.80
6	football5m_ramkishor_L	0.95
41	football5m_kim98_R	1.06
29	football5m_pi7_L	1.16
45	football5m_kim_L	1.33

**Tabelle B.6:** Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Schärfeunterschiede, 3,5 Mbit/s

Nr.	Bild	Qualität
11	football3.5m_L	0.00
8	wheel3.5m_avril_L	-0.14
14	lena2_zoom2_3.5m_avril_L	-0.17
10	stockholm4_3.5m_yang_L	-0.18
31	lena2_zoom2_3.5m_pi7_L	-0.18
40	lena2_zoom2_3.5m_ramkishor_R	-0.26
19	lena2_zoom2_3.5m_kim98_R	-0.31
46	lena2_zoom2_3.5m_yang_L	-0.31
24	stockholm4_3.5m_avril_R	-0.34
17	football3.5m_avril_L	-0.39
38	stockholm4_3.5m_ramkishor_R	-0.41
-	wheel3.5m_kim_R	-0.43
18	stockholm4_3.5m_kim_L	-0.48
32	wheel3.5m_pi4b_L	-0.49
44	wheel3.5m_ramkishor_R	-0.51
28	lena2_zoom2_3.5m_kim_L	-0.54
22	football3.5m_kim_R	-0.61
23	wheel3.5m_pi7_L	-0.63
34	stockholm4_3.5m_pi7_R	-0.64
27	wheel3.5m_yang_R	-0.65
12	wheel3.5m_kim98_R	-0.76
42	stockholm4_3.5m_kim98_L	-0.84
35	football3.5m_yang_L	-0.91
15	football3.5m_pi4b_R	-1.00
20	football3.5m_pi7_L	-1.06
33	football3.5m_ramkishor_R	-1.07
7	football3.5m_kim98_R	-1.46

**Tabelle B.7:** Subjektives Ranking der Bilder bzw. Verfahren, Bewertung der Schärfeunterschiede, 5 Mbit/s

Nr.	Bild	Qualität
21	wheel5m_L	0.18
39	wheel5m_ramkishor_L	0.05
13	wheel5m_kim_L	0.01
37	football5m_avril_R	-0.17
6	football5m_ramkishor_L	-0.22
45	football5m_kim_L	-0.38
26	football5m_yang_R	-0.39
36	wheel5m_pi7_R	-0.39
16	wheel5m_kim98_R	-0.52
30	wheel5m_yang_R	-0.52
29	football5m_pi7_L	-0.53
25	wheel5m_avril_R	-0.66
41	football5m_kim98_R	-0.94

# C Anmerkungen zur Software-Implementierung

## Allgemeine Hinweise zur implementierten Software

In der im Rahmen dieser Arbeit implementierten Software zur Artefaktreduktion und Qualitätsbewertung wurde eine am Lehrstuhl Kommunikationstechnik der Universität Dortmund entwickelte C++ Bibliothek zum Zugriff auf Bilddaten verwendet, deren erste Version auf [Sch95] zurückgeht. Die Bibliothek ermöglicht die Verwaltung von und den Zugriff auf Bildpuffer innerhalb von C++ Programmen. Jedes einzelne Pixel eines Bildpuffers kann komfortabel angesprochen werden. Darüber hinaus können Bilddaten von Datenträgern gelesen und auf Datenträgern gespeichert werden. Es wird ein proprietäres Datenformat eingesetzt, das Bilddaten getrennt nach Farbkanälen unkomprimiert in einzelnen Dateien abspeichert. Bildfolgen werden in Form von aneinandergehängten Einzelbildern in den Farbkanal-Dateien abgespeichert. Informationen über die Beschaffenheit von Bildern oder Bildfolgen werden in zusätzlichen Textdateien festgehalten. Dazu gehören z. B. Zeilenzahl, Spaltenzahl, Bildanzahl und Framerate.

Die einzelnen Algorithmen zur Artefaktreduktion wurden unter Verwendung der genannten Bibliothek als Kommandozeilentools in C++ implementiert. Quell- und Zieldateien werden ebenso wie zugehörige Konfigurationsdateien, die Parameter wie Koeffizienten oder Schwellwerte enthalten, als Aufrufparameter übergeben. Dadurch ist ein Batch-Betrieb der Algorithmen möglich. Vergleichbares gilt für die untersuchten bzw. entwickelten Qualitätsbewertungstools. Auch diese wurden als Kommandozeilentools implementiert, denen Bilddateien und Parameter in Form von Aufrufparametern übergeben werden.

Sämtliche Tools wurden so implementiert, dass sie sowohl monochrome Bilder und Bildsequenzen als auch Bilder und Bildsequenzen im YUV- bzw. YCrCb-Format verarbeiten können. Sowohl die Artefaktreduktion als auch die Qualitätsbewertung erfolgt jedoch ausschließlich im Luminanz-Kanal. Wie in Kapitel 7 dargelegt, konnten keine nennenswerten Bildqualitätsgewinne durch eine zusätzliche Artefaktreduktion in den Chrominanzkanälen wahrgenommen werden, weshalb die Chrominanzkanäle aus den Untersuchungen ausgeschlossen wurden. Daher war auch keine Verarbeitung der Chrominanzkanäle in der Software notwendig.

## Zugriff auf Quantisierungsparameter

Da einige der untersuchten Deblocking-Verfahren eine Adaption von Filtern auf Basis von Quantisierungsparametern vorsehen, musste eine Möglichkeit geschaffen werden, auf Quantisierungsparameter zuzugreifen. Dazu wurde der MPEG2-Decoder der MPEG Software Simulation Group [MSS], Testmodel 5, welcher im Quelltext bezogen werden kann, in der Art modifiziert, dass er die Quantisierungsparameter eines MPEG-Streams in eine Datei schreiben kann. Solche Dateien können - sofern benötigt - von den implementierten Deblocking-Verfahren eingelesen und entsprechend verarbeitet werden. Dabei ist zu berücksichtigen, dass die Decodierreihenfolge von Bildern in einem MPEG-Stream nicht der Anzeigereihenfolge entsprechen muss, was sich auch auf die Positionen von Quantisierungsparametern in den genannten Dateien auswirkt. Um das Problem zu lösen, wurden auch Informationen zur Decodier- und Anzeigereihenfolge von Bildern aus dem MPEG-Decoder exportiert und im Rahmen der Artefaktreduktion berücksichtigt.

## Kantendetektion

Einige Komponenten der in Kapitel 8 vorgestellten Ansätze zur Bewertung der Wirksamkeit von Deblocking-Verfahren benötigen eine Kantendetektion. Diese wird hier in Grundzügen vorgestellt. Die Grundlage bildet ein in [Wu94] vorgestellter kombinierter Gradienten-Laplace-Operator. Abbildung C.1 zeigt die Funktionsweise der Kantendetektion.

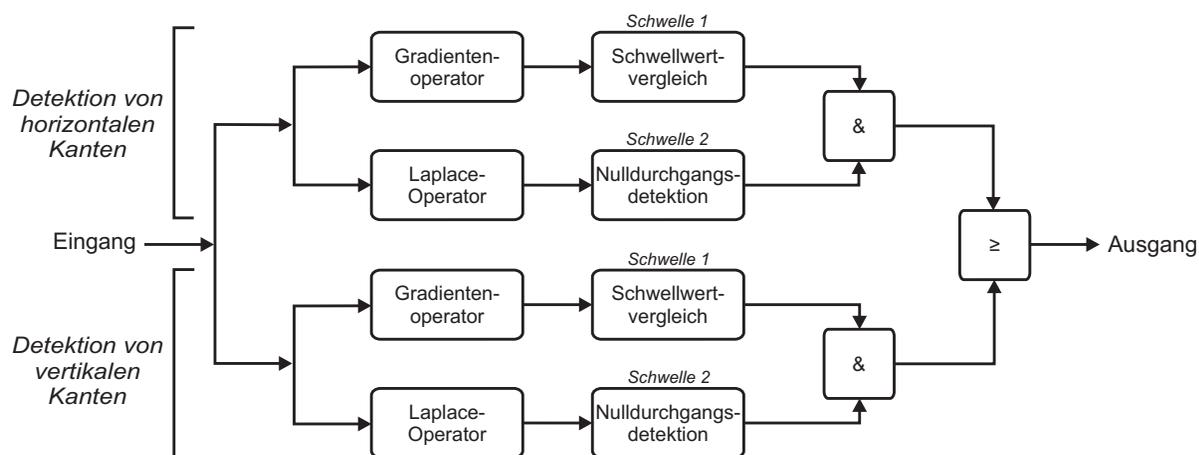


Abbildung C.1: Kombiniertes Gradienten-Laplace-Kantendetektor

Zunächst sind zwei Zweige zu erkennen, in denen eine getrennte Detektion horizontal und vertikal verlaufender Kanten vorgenommen wird. Innerhalb der beiden Zweige

---

wird parallel eine Kantendetektion mit einem Gradientenoperator und eine Kantendetektion mit einem Laplace-Operator durchgeführt. Gradientenbasierte Kantendetektionen ermitteln Kantenverläufe durch die Suche nach großen Steigungen im Bildsignal. Die dazu notwendige Berechnung der ersten Ableitung des Bildsignals kann über Differenzen angenähert werden [Sch98]. Kantendetektionen auf Basis von Laplace-Operatoren ermitteln Kantenverläufe durch die Suche nach Nulldurchgängen der zweiten Ableitung des Bildsignals. Auch dabei sind Näherungen über die Bildung zweiter Differenzen möglich [Sch98]. Um die Rauschempfindlichkeit der Operatoren zu verringern, wurden Tiefpass-Filterungen parallel zum Kantenverlauf integriert. Als Gradientenoperator für vertikal verlaufende Kanten wurde letztlich folgendes Filter verwendet:

$$g = \begin{bmatrix} -1 & 1 \\ -2 & 2 \\ -1 & 1 \end{bmatrix}. \quad (\text{C.1})$$

Als Laplace-Operator für vertikal verlaufende Kanten wurde folgendes Filter eingesetzt:

$$l = \begin{bmatrix} 1 & -2 & 1 \\ 2 & -4 & 2 \\ 1 & -2 & 1 \end{bmatrix}. \quad (\text{C.2})$$

Wenn der absolute Ausgangswert des Gradientenoperators oberhalb einer Detektionsschwelle liegt, werden beide Pixel in der Mitte des Filters (den Positionen der Koeffizienten -2 und 2 entsprechend) als Kantenpixel markiert. Dies geschieht aber nur dann, wenn gleichzeitig mit Hilfe des Laplace-Operators ein Nulldurchgang der zweiten Ableitung des Bildsignals zwischen den beiden Pixeln ermittelt wird. Dazu muss der Laplace-Operator auf beide Pixel angewendet werden. Wenn das Produkt der Ausgangswerte negativ ist, liegt ein Nulldurchgang vor. Um die Rauschempfindlichkeit des Laplace-Operators weiter zu verringern, wurde in dieser Arbeit stets gefordert, dass das Produkt kleiner als eine (negative) Rauschschwelle sein muss. Im Detektionszweig für horizontal verlaufende Kanten werden entsprechende Operationen ausgeführt.

Die Detektionsschwelle ergibt bei einem Wert von ca. 30 eine sehr empfindliche Kantendetektion, die auch einige Texturanteile als Kanten erkennt, und bei einem Wert von ca. 100 eine Kantendetektion, die nahezu keine Texturanteile als Kanten erkennt und bei der erste Lücken in Kantenverläufen auftreten. Die Einstellung der Rauschschwelle ist sehr unkritisch. Bei Werten bis etwa -1000 kann von einer Verringerung der Rauschempfindlichkeit ausgegangen werden, ohne dass ein maßgeblicher Einfluss auf die Eigenschaften der gesamten Kantendetektion erkennbar ist. Bei erheblich niedrigeren Schwellen kommt es zur Entstehung von Lücken in Kantenverläufen. Zur Beeinflussung der grundsätzlichen Eigenschaften der Kantendetektion sollte jedoch im Wesentlichen die Detektionsschwelle verändert werden, während die Rauschschwelle tatsächlich nur der moderaten Verringerung der Rauschempfindlichkeit dienen sollte. In dieser Arbeit wurden gute Erfahrungen

gemacht, wenn die Rauschschwelle auf einen Wert im Bereich der zweifachen bis fünffachen Detektionsschwelle (multipliziert mit -1) eingestellt wurde. Die Angaben zu den verwendeten Schwellen im Rahmen der Kantendetektion in dieser Arbeit bestehen stets aus Wertepaaren. Der erste Wert ist die Detektionsschwelle, der zweite Wert ist die Rauschschwelle. Für den implementierten Kantendetektor lassen sich folgende Eigenschaften angeben:

- separierte Detektion horizontal und vertikal verlaufender Kanten und Verknüpfung über eine Oder-Operation,
- gleichzeitige Detektion von Kanten mit Hilfe von Gradienten- und Laplace-Operatoren und Verknüpfung über eine Und-Operation,
- integrierte Tiefpass-Filterungen zur Verringerung der Rauschempfindlichkeit,
- grundsätzliche Kennzeichnung von Pixelpaaren, zwischen denen ein Intensitätssprung bzw. eine Kante verläuft.

## Einkanalmodell

Für die Untersuchungen in dieser Arbeit wurde ein Einkanalmodell der visuellen Wahrnehmung nach Vorbild des in Kapitel 4 vorgestellten Modells nach [Fec00] eingesetzt. Das nichtlineare Helligkeitsempfinden des Auges und Chrominanzinformationen wurden nicht berücksichtigt, da sich dadurch laut [Fec00] nahezu keine negativen Auswirkungen auf die Modellgüte ergeben. Maskierungseinflüsse wurden im Gegensatz zur Darstellung in Abbildung 4.3 nicht auf Basis des Testbildmaterials, sondern auf Basis des Referenzbildmaterials ermittelt. Dadurch sollte vermieden werden, dass sich Artefakte selbst maskieren können und es sollte sichergestellt werden, dass beim Vergleich gefilterter und nicht gefilterter Bilder jeweils die gleichen Maskierungseffekte zum Tragen kommen. Es sei angemerkt, dass sich in [Fec00] die Ermittlung von Maskierungseinflüssen in der Referenz speziell im Falle von MPEG-2 als leicht vorteilhaft herausgestellt hat. Das Modell war bestimmungsgemäß auf die absolute Qualitätsbewertung (MPEG-2-)komprimierter Sequenzen ausgelegt und die Parametrierung orientierte sich entsprechend an Werten, die nach Angaben in [Fec00] dazu geeignet sind. Die Breite des Fensters für die lokale Fehlersummutation wurde auf 35 Pixel eingestellt, um wenigstens einige komplette Bildblöcke mit dem Fenster zu überdecken. Das Modell erreichte bei der absoluten Qualitätsbewertung von MPEG-2-Sequenzen in einem entsprechenden Testdatensatz eine hohe Korrelation von 0,914. Details zur Implementierung des Modells finden sich in [Hei99]. Um eine Bewertung der Wirksamkeit von Artefaktreduktionsverfahren vornehmen zu können, wurden die Differenzen der für codierte und gefilterte Bilder gemessenen Fehlerwerte verarbeitet. Das Matching auf eine Vergleichsskala (DMOS) erfolgte über einen linearen Regressionsansatz durch den Ursprung. Weitere spezielle Anpassungen an die Aufgabenstellung erfolgten nicht.

# D Messung der Wirksamkeit von Deringing-Verfahren

Um eine besonders umfassende Bewertung der Wirksamkeit von Artefaktreduktionsverfahren zu ermöglichen oder die in dieser Arbeit vorgestellten Qualitätsbewertungsansätze speziell an Ringing-Artefakte anzupassen, sind ein oder mehrere Ringing-Maße zur Integration in die entsprechenden Faktorenmodelle notwendig. Es wurden einige Voruntersuchungen zur Entwicklung von Ringing-Maßen durchgeführt, die vielversprechend verliefen. Im Folgenden sollen einige Ideen skizziert werden, die die mögliche Funktionsweise eines geeigneten und verhältnismäßig einfachen Ringing-Maßes betreffen.

Ringing-Artefakte äußern sich durch Oszillationen der Pixelintensität parallel zu Kantenverläufen. Sie sind örtlich auf die Ausdehnung von Blöcken begrenzt und werden insbesondere in homogenen Bildbereichen in Nachbarschaft zu Kanten sichtbar. Eine wirksame Ringing-Filterung sollte die Oszillationen verringern oder beseitigen. Es zeigte sich während der Voruntersuchungen, dass Oszillationen, die durch Ringing-Artefakte hervorgerufen werden, mit Hilfe von Varianzanalysen innerhalb von kleinen Analysefenstern (z. B. 3x3 Pixel) detektiert bzw. gemessen werden können.

Ein Ringing-Maß könnte zunächst eine Kantendetektion im Originalbild durchführen, um Stellen zu identifizieren, an denen die Möglichkeit des Auftretens von Ringing-Artefakten besteht. Anschließend könnten Varianz-Messungen in der Nachbarschaft der Kantenverläufe sowohl im Originalbild als auch im gestörten und im gefilterten Bild durchgeführt werden. Eine wirksame Ringing-Reduktion würde sich dadurch äußern, dass einer relativ niedrigen Varianz (homogener Bereich) im Originalbild eine höhere Varianz im gestörten Bild gegenübersteht, welche im gefilterten Bild wieder verringert ist. Der Grad der Verringerung könnte zur Bewertung der Wirksamkeit des Artefaktreduktionsverfahrens akkumuliert werden. Es sollte darauf geachtet werden, Varianzanalysen nur in Blöcken durchzuführen, in denen auch Kanten vorhanden sind, da sich Ringing-Artefakte nicht über Blockgrenzen hinaus ausdehnen können. Aus demselben Grund sollten die Analysefenster grundsätzlich so platziert werden, dass sie keine Blockgrenzen überlappen. Schließlich sollte noch eine geeignete Normierung der akkumulierten Verringerungen der Varianz durchgeführt werden, so dass das Qualitätsmaß sowohl auf die Häufigkeit von Verringerungen als auch auf deren Grad reagieren kann.

Es konnte gezeigt werden, dass ein Maß, das auf den geschilderten Ideen basiert, deutlich auf die Anwendung von Deringing-Verfahren reagiert und dass wenig wirksame Verfahren offenbar auch zu kleineren Messwerten führen als stark wirksame Verfahren. Auch

mit einem ähnlichen Ansatz, der statt auf Varianzmessungen auf einem Vergleich von Pixelintensitätsdifferenzen im Originalbild, im gestörten Bild und im gefilterten Bild nahe an Kanten basiert, wurden erste vielversprechende Messergebnisse erzielt.

# E Verwendetes Bildmaterial

Die wichtigsten für die Untersuchungen in dieser Arbeit verwendeten Sequenzen bzw. Bilder sollen im Folgenden in alphabetischer Reihenfolge vorgestellt werden. Dazu ist jeweils ein Einzelbild abgebildet und es werden die bedeutendsten Charakteristika beschrieben. Wenn nicht anders bezeichnet, haben die ausschließlich progressiven Sequenzen bzw. die Einzelbilder eine Auflösung von 720x576 Pixeln.

## Sequenz *coralreef*

Die Sequenz *coralreef* ist ein Ausschnitt aus der HD-Produktion *Coral Reef Adventure*. Aus der Originalsequenz (720p) wurde ein Ausschnitt der Länge 60 Frames in der Größe 720x576 Pixel entnommen. Abbildung E.1 oben links zeigt Frame 36 der Sequenz. Die Bildfolge zeigt einen Kameraflug über eine Wasseroberfläche. Zunächst ist im Vordergrund ein schnell fahrendes Boot zu sehen, der Hintergrund enthält feine Details in Form von Wolken. Es folgt ein Schnitt, der mit einem starken Farbwechsel und dem Auftauchen von Bergen im Hintergrund einhergeht. Durch nochmaliges Anhängen des Sequenzausschnitts in umgekehrter Abspielreihenfolge wurde die Gesamtlänge auf 120 Frames verlängert. Beim Abspielen ergibt sich eine Pingpong-Bewegung. Die Sequenz diente hauptsächlich zur Einweisung der Probanden in die ersten beiden durchgeführten subjektiven Tests.

## Sequenz *football*

Die Sequenz *football*, dargestellt in Abbildung E.1 oben rechts, zeigt einige Fußballspieler, die sich auf einem Fußballplatz unterschiedlich schnell bewegen. Im Hintergrund sind Bäume, ein Strommast, ein Gebäude und ein Teil einer nicht besetzten Zuschauertribüne sichtbar. Der Spielfeldrand wird durch Reklametafeln mit unterschiedlich großen Werbeschriften abgegrenzt. Die Kamera zoomt im Laufe der Sequenz in Richtung des Balls, der sich vor einer der Reklametafeln befindet. Durch diesen Zoom entstehen im Außenbereich der Sequenz, beispielsweise an der Tafel links oben im Bild, Bewegungsgeschwindigkeiten von etwa zwei bis drei Pixel/Bild. Die schnellsten Bewegungen mit acht bis zehn Pixel/Bild sind im Bereich des Fußballspielers, ganz links im Bild, zu finden. Die Sequenz enthält insgesamt viele Kanten sowie horizontale und vertikale Strukturen. Sie ist aufgrund des Kamerazooms und der damit verbundenen globalen Bewegung für typische Videokompressionsverfahren kritisch und somit für Untersuchungen interessant. Sie stellt ein praxisnahes Beispiel für eine Sportübertragung dar. Die Original-Framerate der Sequenz beträgt 50



Abbildung E.1: Im Rahmen der Arbeit verwendete Bilder bzw. Sequenzen

---

Frames/Sekunde. In den Untersuchungen im Rahmen dieser Arbeit wurde *football* jedoch als Sequenz mit 60 Frames/Sekunde behandelt, um eine Frameraten-Konversion bei der Darstellung auf LC-Panels zu vermeiden. Dadurch werden die ursprünglichen Bewegungen etwas beschleunigt. Die Sequenz hat eine Länge von 120 Frames, wobei die letzten 60 Frames den ersten 60 Frames in umgekehrter Reihenfolge entsprechen, so dass sich eine Pingpong-Bewegung ergibt.

### Sequenz *fries2*

*fries2* - Frame 48 dargestellt in Abbildung E.1 Mitte links - ist eine Sequenz, die eine Szene in einem Fastfood-Restaurant zeigt. Das Bedienpersonal bringt Speisen und Getränke aus dem Küchenbereich zum Verkaufstresen, an dem mehrere Kunden warten. Dabei schwenkt die Kamera schnell von rechts nach links, wodurch im Hintergrund ständig neue Ausschnitte aus den Räumlichkeiten des Restaurants sichtbar werden. Die Sequenz enthält viele schnelle Bewegungen und deutliche Konturen. Dies führt bei einer MPEG-2-Codierung zu starken Codierartefakten bei niedrigen Datenraten. Die Framerate beträgt 25 Frames/s bei einer Gesamtlänge von 60 Frames. Ein Pingpong-Loop wurde nicht integriert. Die Framerate von 25 Frames/s sowie der fehlende Pingpong-Loop bereiteten in den Untersuchungen dieser Arbeit trotz Verwendung eines Displays mit einer Wiedergaberate von 60 Frames/s keine Schwierigkeiten, da nur Standbilder aus *fries2* gezeigt wurden. Gleiches gilt auch für die weiteren im Folgenden vorgestellten Sequenzen mit Frameraten von 24 oder 25 Hz.

### Sequenz *lena2\_zoom2*

Zur Erzeugung der Sequenz *lena2\_zoom2* wurde ein Ausschnitt im Seitenverhältnis 4:3 aus dem bekannten Testbild *Lena*, das das Gesicht einer Frau zeigt, entnommen. Zur Erzeugung von Bewegung wurde über eine Länge von 12 Frames auf den Bildmittelpunkt hin gezoomt. Anschließend wurde ebenfalls über eine Länge von 12 Frames wieder zurück gezoomt. Die Sequenz besteht aus fünf Wiederholungen der beschriebenen Bewegung, so dass sich eine Gesamtlänge von 120 Frames ergibt. In Abbildung E.1 Mitte rechts ist Bild 24 der Sequenz dargestellt. Die Sequenz wurde ausgewählt, da das menschliche visuelle System für Fehler in Bildern, die Gesichter zeigen, recht empfindlich ist. Die eingefügte Zoom-Bewegung ist für Video-Encoder kritisch und führt bei geringen Datenraten zu deutlichen Artefakten.

### Sequenz *mat\_phone*

Die Sequenz *mat\_phone* zeigt einen Ausschnitt aus der Kinoproduktion *Die Matrix*. Es wird, wie in Abbildung E.1 unten links anhand von Frame 12 dargestellt, ein Innenstadtbereich mit im Hintergrund sichtbaren Hochhäusern, Bäumen und Fahrzeugen gezeigt. Im

Vordergrund durchqueren Fußgänger mit unterschiedlichen Bewegungsgeschwindigkeiten bis zu etwa acht Pixel/Bild (horizontal) die Szene und links verlässt eine Person mit etwas geringerer Geschwindigkeit eine Telefonzelle. Zusätzlich erhält die Szene durch einen Kameraschwenk weitere Bewegung in horizontaler Richtung. Die Sequenz zeichnet sich durch eine Vielzahl unterschiedlich schneller Bewegungen aus, was sie zusammen mit den enthaltenen Konturen und Details für Videokompressionsverfahren kritisch macht. Sie stellt ein typisches Beispiel für Actionfilm-Material dar. Die Original-Framerate der Sequenz beträgt 24 Bilder/Sekunde, es erfolgte aber eine Behandlung als Sequenz mit 25 Bildern/s. *mat.phone* diente hauptsächlich zur Einweisung der Probanden in die durchgeführten subjektiven Tests.

### Sequenz *pedestrian3*

Die Sequenz *pedestrian3* ist ein Ausschnitt aus der HD-Testsequenz *Pedestrian Area* im Bildformat 16:9, die sich schnell bewegende Menschen in einer Fußgängerzone zeigt. Im Hintergrund sind Gebäudefassaden und Schaufenster zu sehen. Um eine SD-Sequenz im Format 4:3 zu erzeugen, wurde die HD-Sequenz (1080p) zunächst auf die korrekte Zeilenzahl und somit auf die Auflösung 1024x576 skaliert. Dann wurde ein Ausschnitt der Größe 720x576 Pixel entnommen, der pro Frame um zwei Pixel von links nach rechts wandert. Abbildung E.1 unten links zeigt Frame 24 aus der Sequenz der Länge 60 Frames. Durch den wandernden Bildausschnitt und die sich bewegenden Fußgänger in einem recht kontur- und detailreichen Umfeld ist die Sequenz kritisch für die Videocodierung, was bei niedrigen Datenraten zu deutlichen Artefakten führt. Die Framerate beträgt 25 Frames/s.

### Sequenz *shields3*

In Abbildung E.2 oben links ist Frame 48 der Sequenz *shields3* dargestellt. Die Sequenz zeigt eine Szene, in der ein Mensch vor einer Wand aus diversen Wappen steht und auf eines der Wappen zeigt. Es sind viele Details und Konturen sowie Texturen um die Wappen herum vorhanden. Die Kamera schwenkt zunächst schnell von rechts nach links und zoomt anschließend sehr schnell und ungleichmäßig auf den Zeigefinger des Menschen. Die Sequenz ist ein Ausschnitt der Größe 720x576 Pixel und der Länge 60 Frames aus der HD-Testsequenz *Knightshields* (720p). Die Framerate beträgt 60 Frames/s. Durch die schnellen Kamerabewegungen und den Kontur- und Detailreichtum der Szene entstehen bei einer Codierung mit niedriger Datenrate klar sichtbare Codierartefakte, weshalb die Sequenz für die Untersuchungen in dieser Arbeit gut geeignet ist.



Abbildung E.2: Im Rahmen der Arbeit verwendete Bilder bzw. Sequenzen

### Sequenz *stockholm4*

Die Sequenz *stockholm4* ist ein Ausschnitt der Größe 720x576 Pixel aus der HD-Testsequenz *Stockholm Pan* (Originalauflösung 1280x720 Pixel). Die Framerate der Sequenz beträgt 60 Bilder/Sekunde. In Abbildung E.2 oben links ist Bild 12 der Sequenz dargestellt. Die Sequenz zeigt einen Kameraschwenk über die Altstadt von Stockholm. Es sind viele Details, reguläre Strukturen und eine Wasserfläche sichtbar. Im Vordergrund bewegen sich einige Fahrzeuge. Die Sequenz ist 120 Frames lang, wobei die letzten 60 Frames den ersten 60 Frames in umgekehrter Reihenfolge entsprechen, so dass sich eine Pingpong-Bewegung ergibt. Bei einer Codierung mit geringen Datenraten entstehen deutliche Artefakte.

## Sequenz *wheel*

Bei der Sequenz *wheel*, aus der Frame 12 in Abbildung E.2 unten dargestellt ist, handelt es sich um eine unter Laborbedingungen aufgezeichnete Testsequenz. Es wird ein Aufbau gezeigt, der aus einem rotierenden Speichenrad besteht, welches sich vor einer ebenfalls rotierenden Scheibe befindet und diese teilweise überdeckt. Die Speichen bewegen sich im äußeren Bereich des Rades mit einer recht hohen Geschwindigkeit von neun bis zehn Pixeln pro Bild, wenn Bereiche betrachtet werden, in denen diese Bewegung rein horizontal oder rein vertikal ist. An der Scheibe, die in schwarze und weiße Sektoren unterteilt ist, sind mehrere Objekte befestigt, welche der Sequenz unterschiedlich große Details hinzufügen. Die Objekte bewegen sich mit einer Geschwindigkeit von maximal vier bis fünf Pixeln/Bild. Der Hintergrund des gesamten Aufbaus ist größtenteils schwarz.

Die Sequenz ist aufgrund der beiden Rotationsbewegungen und der bewegten unterschiedlich detailreichen Objekte interessant. Die Speichen stellen hochfrequente Strukturen verschiedener Ausrichtung und Bewegungsgeschwindigkeit dar, an denen nach Anwendung von Videokompressionsverfahren starke Artefakte auftreten können. Die Original-Framerate dieser synthetischen Sequenz beträgt 50 Frames/Sekunde. In den Untersuchungen im Rahmen dieser Arbeit wurde *wheel* jedoch als Sequenz mit 60 Frames/Sekunde behandelt, was durch den synthetischen Charakter keinerlei Probleme oder Auffälligkeiten verursacht. Die Sequenz ist 100 Frames lang, wobei die letzten 50 Frames den ersten 50 Frames in umgekehrter Reihenfolge entsprechen, so dass sich eine Pingpong-Bewegung ergibt.

# Lebenslauf

## Persönliche Daten

Name	Piastowski
Vorname	Patrick
Geburtsdatum	19.05.1976
Geburtsort	Herne

## Ausbildung

1995	Abitur am Otto-Hahn-Gymnasium in Herne
1995-1996	Zivildienst in der Christopherus-Schule in Bochum
1996-2001	Studium der Elektrotechnik an der Technischen Universität Dortmund
08/2001	Abschluss des Studiums mit Diplom
16.09.2008	Promotion an der Technischen Universität Dortmund

## Beruflicher Werdegang

10/2001 - 09/2006	Wissenschaftlicher Mitarbeiter am Arbeitsgebiet Schaltungen der Informationsverarbeitung an der Technischen Universität Dortmund
seit 05/2007	Consultant bei der ASKON ConsultingGroup GmbH, Düsseldorf, Schwerpunkt: Automotive Software Engineering