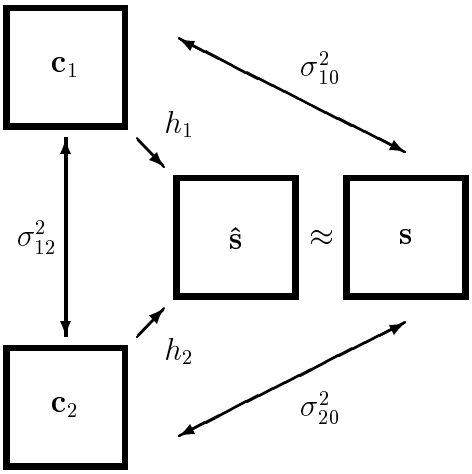
Lehrstuhl für Nachrichtentechnik I
Universität Erlangen-Nürnberg
Prof. Dr.-Ing. B. Girod



Bewegungskompensierte Multihypothesen-Langzeitprädiktion

Diplomarbeit im Studiengang Elektrotechnik

vorgelegt von

Markus Flierl

betreut von

Dipl.-Ing. Thomas Wiegand

19. September 1997

Erklärung

„Ich versichere, daß ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und daß die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.“

Erlangen, 19. September 1997

An dieser Stelle möchte ich all denen danken, die mir bei der Anfertigung dieser Diplomarbeit zur Seite standen. Vor allem möchte ich mich bei meinen Eltern bedanken, die mir ihre Unterstützung jederzeit zugesichert haben. Bei der Durchführung der umfangreichen Simulationen wurde ich von Bernd Westrich unterstützt. Im besonderen bin ich Thomas Wiegand für seine fachliche Betreuung und stete Diskussionsbereitschaft zu Dank verpflichtet.

Zeit und Zeitlosigkeit hängen zusammen. Augenblick und Ewigkeit kämpfen in uns. Daraus entstehen all unsere Widersprüche, unser Trotz, unsere Engstirnigkeit, unser Glaube und unser Kummer.
Wir müssen damit rechnen, daß unsere Lieder eines Tages ein Ende nehmen. Vielleicht kommt auch für den größten Künstler ein Moment, in dem er nicht mehr Kunst machen will oder muß. Und vielleicht schätzen wir sein Schaffen gerade dann noch höher ein; weil es diesen Augenblick gegeben hat, in dem er über sein Werk hinausgelangt ist.

— *Arvo Pärt*

Inhaltsverzeichnis

Zusammenfassung **xi**

Liste der verwendeten Formelzeichen und Abkürzungen **xiii**

Abbildungsverzeichnis **xv**

1 Einleitung **1**

2 Multihypothesen-Prädiktion **3**
2.1 Multihypothesen-Modell 3
2.2 Schätzung der Hypothesen 5
2.2.1 Schätzkriterium für die Bewegungsparameter 5
2.2.2 Algorithmus zur Auswahl der Hypothesen 7
2.2.3 Optimale Anzahl der Hypothesen 14
2.3 Entwurf des Multihypothesen-Prädiktors 20
2.3.1 Iterativer Entwurfsalgorithmus 21
2.3.2 Optimale Interpolationskoeffizienten 22
2.3.3 Interpretation des Entwurfs 29

3 Variationen zur Multihypothesen-Prädiktion **31**
3.1 Bidirektionale Multihypothesen-Prädiktion 31
3.1.1 Optimale Wahl des Blocks für eine Hypothese 31
3.1.2 Optimale Wahl der Blöcke für mehrere Hypothesen 32
3.2 Langzeitprädiktion mit überlappenden Blöcken 35
3.2.1 Modell für die Blocküberlappung 35
3.2.2 Überlappende Bewegungsschätzung 36
3.2.3 Optimale Gewichtskoeffizienten 37
3.2.4 Entwurf des Prädiktors 37
3.3 Interblock-Prädiktion der Translation 41
3.3.1 Translationsmodell 41
3.3.2 Translationsgeschwindigkeit des aktuellen Blocks 41
3.3.3 Translationsgeschwindigkeit der Nachbarblöcke 42

4 Schlußbemerkungen **45**

A Wiener-Lösung mit Nebenbedingung **47**

B Huffman-Codierung **49**

C Simulationsbedingungen **53**

D Methode der n besten Matches **55**

E Signaldarstellung und Systemkomponenten **57**

Literaturverzeichnis **59**

Zusammenfassung

In der vorliegenden Arbeit wird ein Verfahren zur bewegungskompensierten Langzeitprädiktion vorgestellt, das n Hypothesen aus mehreren Bildern linear überlagert. Es wird ein lokal optimaler Algorithmus zur Auswahl der Hypothesen präsentiert. Im Rahmen eines Rate-Distortion-Modells wird blockweise die optimale Anzahl der Hypothesen bestimmt. Wir interpretieren die bewegungskompensierte Multihypothesen-Langzeitpädiktion als Vektorquantisierung und lösen das Problem des Prädiktorentwurfs mit Algorithmen zum Entwurf eines Vektorquantisierers. Weiterhin untersuchen wir die bidirektionale Multihypothesen-Prädiktion und die blocküberlappende Langzeitprädiktion. Zur Verbesserung der Rate-Distortion-Effizienz schlagen wir ein Modell zur Interblock-Prädiktion der Multihypothesen-Translation vor.

Liste der verwendeten Formelzeichen und Abkürzungen

Signale

$\mathbf{S}$	Zufallsvariable der Originalblöcke
$\hat{\mathbf{S}}$	Zufallsvariable der geschätzten Blöcke
$\mathbf{C}$	(Vektorielle) Zufallsvariable der Hypothesen
$\mathbf{B}$	Zufallsvariable des Multihypothesencodes
$\mathbf{\Delta}$	(Vektorielle) Zufallsvariable der Multihypothesen-Translation
$\mathbf{D}$	(Vektorielle) Zufallsvariable der differentiellen Multihypothesen-Translation
$\hat{\mathbf{V}}$	(Vektorielle) Zufallsvariable der geschätzten Translationsgeschwindigkeit

Funktionen

$d(\mathbf{s}, \hat{\mathbf{s}})$	Distortion zwischen den Blöcken $\mathbf{s}$ und $\hat{\mathbf{s}}$
$f_{\mathbf{S}}(\mathbf{s})$	Wahrscheinlichkeitsfunktion der Zufallsvariablen $\mathbf{S}$
$f_{\hat{\mathbf{S}}\|\mathbf{S}}(\hat{\mathbf{s}}\|\mathbf{s})$	Bedingte Wahrscheinlichkeitsfunktion der Zufallsvariablen $\hat{\mathbf{S}}$
$r(\mathbf{b})$	Rate des Codewortes $\mathbf{b}$
$\|\cdot\|_2^2$	Quadratische l_2-Norm
$\|\mathbf{b}\|$	Länge des Codewortes $\mathbf{b}$
$\alpha(\cdot)$	Bewegungskompensierte Multihypothesen-Prädiktion
$\beta(\cdot)$	Multihypothesen-Bewegungskompensation
$\gamma(\cdot)$	Entropie-Codierung der Bewegungsparameter

Funktionale

D	Distortion-Funktional
H	Entropie-Funktional
J	Lagrange-Funktional
P	Wahrscheinlichkeits-Funktional
R	Raten-Funktional
$\mu_{\mathbf{S}}$	Mittelwert-Funktional der Zufallsvariablen $\mathbf{S}$
$\varphi_{\mathbf{CS}}$	Korrelation der Zufallsvariablen $\mathbf{C}$ und $\mathbf{S}$
$\sigma^2_{\mathbf{CS}}$	Kovarianz der Zufallsvariablen $\mathbf{C}$ und $\mathbf{S}$

Operatoren

$E\{\cdot\}$	Erwartungswert-Operator
$\cdot \circ \cdot$	Verknüpfungsoperator zweier Funktionen

Mengen

$\mathcal{B}$	Ereignisraum des Translationscodes
$\mathcal{L}$	Blockgitter
$\mathcal{V}$	Ereignisraum der Blöcke

Symbole

C	Komplexität des Optimierungsproblems
I	Anzahl der Iterationen
N	Maximale Anzahl der Hypothesen
a	Betrag der maximalen örtlichen Verschiebung
b	Betrag der maximalen bedingten Verschiebung
h	Vektor der Interpolationskoeffizienten
m	Anzahl der Bilder, die für die Schätzung zur Verfügung stehen
n	Anzahl der Hypothesen
s	Kantenlänge des Blocks
u	Eins-Vektor
λ	Lagrange-Parameter

Abkürzungen

CLGA	Chou Lookabaugh Gray Algorithm
ECVQ	Entropy Constrained Vector Quantization
GLA	Generalized Lloyd Algorithm
HP	Half-Pel
ICM	Iterated Conditional Modes
IP	Integer-Pel
MAP	Maximum a posteriori
MHC	Multi-Hypothesis Compensation, Multi-Hypothesis Code
MHE	Multi-Hypothesis Estimation
OHSA	Optimal Hypothesis Selecting Algorithm
PD	Peak Distortion
QCIF	Quarter Common Intermediate Format
VQ	Vector Quantization

Abbildungsverzeichnis

2.1 *Modell für die bewegungskompensierte Multihypothesen-Prädiktion* 4
2.2 *Veranschaulichung einer Hypothese am 2-D Block* 5
2.3 *Veranschaulichung zweier Hypothesen am 2-D Block* 6
2.4 *Teilung einer Hypothese in zwei Hypothesen* 8
2.5 *Algorithmus zur Auswahl der Hypothesen* 9
2.6 *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke,* $b = 4$ *und* $\lambda = 0$ 11
2.7 *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke,* $b = 4$ *und* $\lambda = 0$ 11
2.8 *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke und* $\lambda = 0$ *in Abhängigkeit der Größe des bedingten Suchbereichs* 12
2.9 *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke und* $\lambda = 0$ *in Abhängigkeit der Größe des bedingten Suchbereichs* 12
2.10 *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* $b = 4$ *und* $\lambda = 0$ 13
2.11 *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* $b = 4$ *und* $\lambda = 0$ 13
2.12 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke* 15
2.13 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke* . 15
2.14 *Partielle Raten über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke, Integer-Pel-Genauigkeit,* $N = 4$ *und* $m = 10$ 16
2.15 *Partielle Raten über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke, Integer-Pel-Genauigkeit,* $N = 4$ *und* $m = 10$ 16
2.16 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke* 17
2.17 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke* . 17
2.18 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke* 18
2.19 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke* . 18

2.20 *Partielle Raten über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 4×4 *Blöcke, Integer-Pel-Genauigkeit,* $N = 4$ *und* $m = 10$ 19
2.21 *Partielle Raten über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 4×4 *Blöcke, Integer-Pel-Genauigkeit,* $N = 4$ *und* $m = 10$ 19
2.22 *Quantisierungsmodell* 20
2.23 *Iterativer Entwurfsalgorithmus* 22
2.24 *Konvergenz der Lagrange-Kosten für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $m = 10$ 24
2.25 *Koeffizienten für 2 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $m = 10$ 25
2.26 *Koeffizienten für 3 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $m = 10$ 25
2.27 *Koeffizienten für 4 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $m = 10$ 26
2.28 *Konvergenz der Lagrange-Kosten für die Trainingssequenzen (QCIF, 7.5 fps),* $\lambda = 100$, $m = 1$ *und* 4×4 *Blöcke* 27
2.29 *Koeffizienten für 2 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* $\lambda = 100$, $m = 1$ *und* 4×4 *Blöcke* 27
2.30 *Koeffizienten für 3 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* $\lambda = 100$, $m = 1$ *und* 4×4 *Blöcke* 28
2.31 *Koeffizienten für 4 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* $\lambda = 100$, $m = 1$ *und* 4×4 *Blöcke* 28
2.32 *Interpretation des Entwurfs* 29

3.1 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s), Integer-Pel-Genauigkeit und* 16×16 *Blöcke zum Verhalten der bidirektionalen Prädiktion.* 32
3.2 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s), Integer-Pel-Genauigkeit und* 16×16 *Blöcke zum Verhalten der bidirektionalen Prädiktion.* . 32
3.3 *Konvergenz der Lagrange-Kosten für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $N = 4$ 33
3.4 *Koeffizienten für 2 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $N = 4$ 33
3.5 *Koeffizienten für 3 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $N = 4$ 34
3.6 *Koeffizienten für 4 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $N = 4$ 34
3.7 *Direkte Nachbarn des aktuellen Blocks* 35
3.8 *Aufteilung der Blöcke in 4 Teilgitter bei der Prädiktion mit überlappenden Blöcken.* 36
3.9 *Konvergenz der Lagrange-Kosten für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$, $m = 10$ *und* $n = 1$ 38
3.10 *Initiale Interpolationskoeffizienten für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$, $m = 10$ *und* $n = 1$ 38

3.11 *Interpolationskoeffizienten bei der 1. Iteration für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$, $m = 10$ *und* $n = 1$ 39
3.12 *Interpolationskoeffizienten bei der 5. Iteration für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$, $m = 10$ *und* $n = 1$ 39
3.13 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke,* $n = 1$ *und Integer-Pel-Genauigkeit zum Verhalten der Blocküberlappung* 40
3.14 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke,* $n = 1$ *und Integer-Pel-Genauigkeit zum Verhalten der Blocküberlappung* 40
3.15 *Interblock-Prädiktion der Translationsgeschwindigkeit* 42
3.16 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke, Integer-Pel-Genauigkeit,* $m = 10$ *und* $N = 4$ *zum Verhalten der Interblock-Prädiktion.* 43
3.17 *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke, Integer-Pel-Genauigkeit,* $m = 10$ *und* $N = 4$ *zum Verhalten der Interblock-Prädiktion.* 43

B.1 *Huffman-Algorithmus zur Konstruktion eines binären präfixfreien Codes* . . 49
B.2 *Konstruktion eines Huffman-Baums* 50

C.1 *Veranschaulichung des arithmetischen Mittels* $D_a = \frac{1}{2}(D[1]+D[2])$ *und des geometrischen Mittels* $D_g = (D[1]D[2])^{\frac{1}{2}}$ *am Halbkreis mit dem Durchmesser* $D[1]+D[2]$. *Der geometrische Mittelwert ist immer kleiner oder gleich dem arithmetischen Mittelwert.* 54

D.1 *Prädiktionsfehler über die Anzahl der Hypothesen für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcken* 56
D.2 *Prädiktionsfehler über die Anzahl der Hypothesen für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcken* 56

E.1 *Modell des äquivalenten Testkanals* 57

Kapitel 1

Einleitung

Die *bewegungskompensierte Prädiktion* stellt derzeit das effizienteste Verfahren zur Reduktion zeitlicher Korrelation in Videosequenzen dar. Algorithmen zur Videocodierung, wie z.B. die ITU-T Empfehlung H.263 [1], integrieren diese Verfahren und sind damit in der Lage, Videosequenzen sowohl bei sehr niedrigen als auch bei hohen Datenraten zu verarbeiten. Theoretische Untersuchungen der bewegungskompensierten Prädiktion in Hinblick auf den Einsatz in einem hybriden Codierschema werden in [2] erstmals angestellt. Eine Verbesserung der Prädiktion durch eine Erhöhung der Genauigkeit bei der bewegungskompensierten Prädiktion wird in [3] dokumentiert.

Algorithmen zur Videocodierung verwenden das zuletzt decodierte Bild zur zeitlichen Dekorrelation des aktuellen Bildes. In [4] wird erstmals ein *Verfahren zur bewegungskompensierten Langzeitprädiktion* vorgestellt, das neben dem letzten auch früher decodierte Bilder zur zeitlichen Dekorrelation des aktuellen Bildes nutzt.

Der Begriff „Multihypothesen-Prädiktion“ wurde durch die Tatsache geprägt, daß Videocoder zur Prädiktion des aktuellen Bildes mehr als ein bewegungskompensiertes Prädiktionssignal verwenden. Beispiele hierfür sind die blocküberlappende Prädiktion [5] und die bidirektionale Prädiktion [6]. Eine umfassende *Effizienzanalyse der bewegungskompensierten Multihypothesen-Prädiktion* für die Videocodierung wird erstmals in [7] angestellt.

Wir interpretieren die *bewegungskompensierte Multihypothesen-Langzeitprädiktion* als ein Kommunikationssystem, das im Rahmen der Rate-Distortion-Theorie [8, 9, 10] optimiert wird. Die Minimierung des Verzerrungsmaßes (Distortion) für gegebene Raten-Nebenbedingung lösen wir mit Hilfe eines Lagrange-Ansatzes [11].

Die n lokal optimalen Hypothesen für die bewegungskompensierte Multihypothesen-Prädiktion bestimmen wir mit einem iterativen Algorithmus. Dieser vermeidet den Test aller möglichen Hypothesenkombinationen, indem er n bedingt optimale Lösungen sukzessiv verbessert und im Konvergenzfall eine lokal optimale Lösung liefert. Der Algorithmus orientiert sich dabei an den *Iterated Conditional Modes* nach Besag [12]. Die optimale Anzahl der Hypothesen eines Blocks determinieren wir mit Hilfe des Rate-Distortion-Kriteriums.

Für die bewegungskompensierte Multihypothesen-Langzeitprädiktion benötigen wir neben den optimalen Hypothesen auch die optimalen Interpolationskoeffizienten für eine

gewichtete Überlagerung der Hypothesen. Zur Lösung dieses nichtlinearen Optimierungsproblems wird ein iterativer Algorithmus verwendet, der für die Vektorquantisierung mit Entropie-Nebenbedingung [13] bekannt ist. Dieser ist eine erweiterte Variante des *Generalized Lloyd Algorithm* [14].

In **Kapitel 2** stellen wir ein Multihypothesen-Modell vor, das die bewegungskompensierte Multihypothesen-Langzeitprädiktion faßt. Nach der Definition des Schätzkriteriums für die Multihypothesen-Translation präsentieren wir Algorithmen zur Auswahl der optimalen Hypothesen und zur optimalen Anzahl der Hypothesen. Zur Bestimmung der optimalen Interpolationskoeffizienten erläutern wir einen iterativen Entwurfsalgorithmus. Das Kapitel beschließen wir mit einer Interpretation des Entwurfs.

Kapitel 3 behandelt die bidirektionale Multihypothesen-Prädiktion, die Langzeitprädiktion mit überlappenden Blöcken und die Interblock-Prädiktion der Multihypothesen-Translation. Diese Abschnitte variieren das Thema der Multihypothesen-Prädiktion und sollen die Vielschichtigkeit dieser Grundidee andeuten.

Im **Anhang** wird die Wiener-Lösung mit Nebenbedingung im Detail behandelt, ein Überblick über die Huffman-Codierung gegeben, die Simulationsbedingungen festgehalten und die verwendete Signaldarstellung skizziert. Eine nicht optimale Methode zum Auffinden der n Hypothesen ist zum Vergleich angegeben.

Kapitel 2

Multihypothesen-Prädiktion

2.1 Multihypothesen-Modell

Der blockbasierten bewegungskompensierten Prädiktion [2] liegt folgendes Modell zugrunde: Das aktuelle Bild wird in sich nicht überlappende Blöcke eingeteilt. Für jeden Block des aktuellen Bildes ist im vorherigen Bild ein Block identischer Größe zu finden, der durch Translation in örtlicher und zeitlicher Richtung die Position des Blocks im aktuellen Bild einnimmt. Die optimalen Translationsparameter sind dabei zu bestimmen.

Die Rate-Distortion-Theorie [8] stellt ein Optimierungskriterium zur Verfügung. Man veranschauliche sich ein Kommunikationssystem, das ein Originalsignal $\mathbf{S}$ in einen Code $\mathbf{B}$ transformiert und daraus wiederum die Rekonstruktion $\hat{\mathbf{S}}$ gewinnt. Das Kommunikationssystem wird dabei durch die mittlere Distortion zwischen Original und Rekonstruktion

$$D = E\left\{d(\mathbf{S}, \hat{\mathbf{S}})\right\} \tag{2.1}$$

und durch die mittlere Rate des Codes

$$R = E\left\{r(\mathbf{B})\right\} \tag{2.2}$$

charakterisiert. Das Kommunikationssystem bezeichnet man dann als optimal, wenn bei gegebener mittlerer Rate die mittlere Distortion zwischen Original und Rekonstruktion minimiert ist. Diese Formulierung mit Hilfe der Raten-Nebenbedingung kann mit einem Lagrange-Ansatz [11] in ein Optimierungsproblem ohne Nebenbedingung transformiert werden. Dabei wird das Distortion-Funktional um die mit einem Lagrange-Parameter gewichteten mittleren Rate additiv ergänzt.

$$J(\lambda) = E\left\{d(\mathbf{S}, \hat{\mathbf{S}})\right\} + \lambda E\left\{r(\mathbf{B})\right\} \tag{2.3}$$

Das optimale Kommunikationssystem für den gewählten Parameter λ erhält man durch Minimierung der Lagrange-Kosten $J(\lambda)$.

Für die blockbasierte bewegungskompensierte Multihypothesen-Prädiktion legen wir in dieser Arbeit folgendes Modell zugrunde: Wiederum wird das aktuelle Bild in sich nicht überlappende Blöcke eingeteilt. Für jeden Block des aktuellen Bildes sind nun aber im

vorherigen Bild n Blöcke identischer Größe zu finden, die durch Translation in örtlicher und zeitlicher Richtung die Position des Blocks im aktuellen Bild einnehmen. Die n Blöcke sind dabei durch skalare Koeffizienten zu gewichten. Die Translationsparameter der n Blöcke sind auch nach dem Rate-Distortion-Kriterium zu optimieren.

Bisher wurden die für die Schätzung zur Verfügung stehenden Blöcke aus dem vorherigen Bild herangezogen. Trifft diese Restriktion zu, so sprechen wir von *Kurzzeitprädiktion*. Ziehen wir hingegen Blöcke aus den m vorherigen Bildern zur Schätzung heran, so sprechen wir für $m > 1$ von *Langzeitprädiktion*. Für die Langzeitprädiktion ist somit eine explizite zeitliche Translation zu codieren.

Wir konkretisieren nun die oben angeführten Modellvorstellungen. Bezeichnet $\mathbf{s}$ einen *Originalblock* im aktuellen Bild k, so versuchen wir durch Prädiktion den *geschätzten Block* $\hat{\mathbf{s}}$ zu bestimmen. $\mathbf{c}_\nu$ beschreibt den Block, der zur Schätzung herangezogen wird und sich in einem beliebigen vorherigen Bild $k - \kappa$ mit $\kappa > 0$ befindet. Die Suchraumgröße in zeitlicher Richtung m stellt die Bilder $\{k-m, k-m+1, \ldots, k-1\}$ für die Schätzung zur Verfügung.

Der geschätzte Block $\hat{\mathbf{s}}$ ergibt sich aus einer Linearkombination der zur Verfügung stehenden Blöcke $\mathbf{c}_\nu$ mit den Gewichten h_ν.

$$\hat{\mathbf{s}} = \sum_{\nu=1}^{n} \mathbf{c}_\nu h_\nu = \begin{pmatrix} \mathbf{c}_1 & \ldots & \mathbf{c}_n \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \mathbf{c}h \tag{2.4}$$

Für die Kurzzeitprädiktion befinden sich ein $\mathbf{c}$-Block im vorherigen Bild $k-1$; Langzeitprädiktion erlaubt einem $\mathbf{c}$-Block eine Position in einem beliebigen vorherigen Bild $k - \kappa$, wobei $1 \leq \kappa \leq m$ gilt.

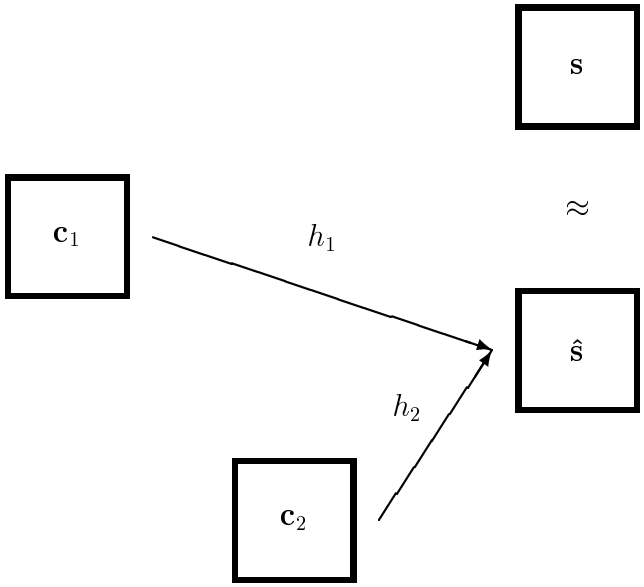


Abbildung 2.1: *Modell für die bewegungskompensierte Multihypothesen-Prädiktion*

Abbildung 2.1 veranschaulicht das Modell für die bewegungskompensierte Multihypothesen-Prädiktion für den Fall zweier Hypothesen $n = 2$. Um die Darstellung im folgenden zu vereinfachen, wählen wir an geeigneter Stelle zwei Hypothesen. Die allgemeine Darstellung mit n Hypothesen dient nur der Vollständigkeit.

2.2 Schätzung der Hypothesen

2.2.1 Schätzkriterium für die Bewegungsparameter

Für die Optimierung interpretieren wir die bewegungskompensierte Multihypothesen-Prädiktion als ein Kommunikationssystem. Dem Originalblock $\mathbf{s}$ wird ein Bewegungscode $\mathbf{b}$ zugeordnet, der dem Empfänger des Codes erlaubt, die Rekonstruktion (den geschätzten Block) $\hat{\mathbf{s}}$ zu erzeugen.

Wir erläutern zuerst das Vorgehen zur Schätzung der Bewegungsparameter für die bewegungskompensierte Prädiktion für eine Hypothese $n = 1$. Dafür haben wir im Suchraum einen Block $\mathbf{c}_1$ zu finden, der in diesem Fall identisch mit dem geschätzten Block $\hat{\mathbf{s}}$ ist.

Die Distortion zwischen Original und Rekonstruktion berechnen wir aus der quadratischen l_2-Norm der Blöcke. Für diesen Fall interpretieren wir die $s \times s$-Blöcke als Vektoren der Dimension s^2.

$$d(\mathbf{s}, \hat{\mathbf{s}}) = \|\mathbf{s} - \hat{\mathbf{s}}\|_2^2 \tag{2.5}$$

Den Bewegungscode $\mathbf{b}$ erhalten wir durch eine Codierung γ der Bewegungsparameter $\mathbf{\Delta}_{x_1}$, $\mathbf{\Delta}_{y_1}$ und $\mathbf{\Delta}_{t_1}$ des Blocks $\mathbf{c}_1$. Die Rate des Bewegungscodes ist dabei die Länge des Bewegungscodewortes.

$$r(\mathbf{b}) = |\gamma(\mathbf{c}_1)| \tag{2.6}$$

Wir geben uns die Länge $|\gamma(\cdot)|$ der Bewegungscodewörter vor und finden den optimalen Block $\mathbf{c}_1$ durch Minimierung des Lagrange-Funktionals.

$$\min_{\mathbf{c}_1} \left\{ \|\mathbf{s} - \mathbf{c}_1\|_2^2 + \lambda |\gamma(\mathbf{c}_1)| \right\} \tag{2.7}$$

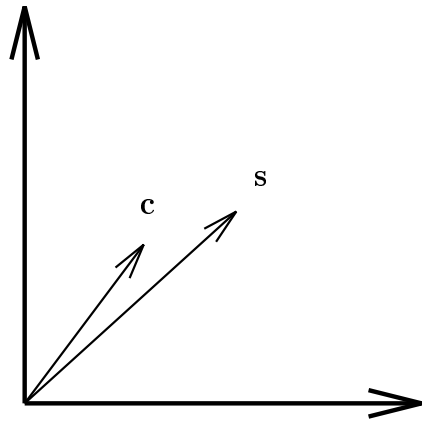


Abbildung 2.2: *Veranschaulichung einer Hypothese am 2-D Block*

Abbildung 2.2 veranschaulicht die Schätzung einer Hypothese am zwei-dimensionalen Block, d.h. an einem Block mit zwei Bildelementen. In dieser Darstellung werden der Originalblock $\mathbf{s}$ und der Block $\mathbf{c}_1$ als Vektoren abgebildet. Der Vektor $\mathbf{c}_1$ ist so zu wählen, daß der Lagrange-Abstand

$$\|\mathbf{s} - \mathbf{c}_1\|_2^2 + \lambda |\gamma(\mathbf{c}_1)| \tag{2.8}$$

minimiert wird.

Im weiteren erläutern wir das Vorgehen zur Schätzung der Bewegungsparameter für die bewegungskompensierte Prädiktion für zwei Hypothesen $n = 2$. Dafür haben wir im Suchraum die Blöcke $\mathbf{c}_1$ und $\mathbf{c}_2$ zu finden. Der geschätzte Block $\hat{\mathbf{s}}$ ergibt sich aus der Linearkombination der Blöcke $\mathbf{c}_1$ und $\mathbf{c}_2$.

$$\hat{\mathbf{s}} = \mathbf{c}_1 h_1 + \mathbf{c}_2 h_2. \tag{2.9}$$

Die Distortion berechnet sich wie im Fall einer Hypothese nach **Gleichung 2.5**. Den Bewegungscode $\mathbf{b}$ erhalten wir durch eine Codierung γ der Bewegungsparameter $\mathbf{\Delta}_{x_\nu}$, $\mathbf{\Delta}_{y_\nu}$ und $\mathbf{\Delta}_{t_\nu}$ für jeden Block $\mathbf{c}_\nu$. Die Rate des Bewegungscodes ist dabei die Summe der Längen der Bewegungscodewörter.

$$r(\mathbf{b}) = |\gamma(\mathbf{c}_1)| + |\gamma(\mathbf{c}_2)| \tag{2.10}$$

Wir geben uns die Länge $|\gamma(\cdot)|$ der Bewegungscodewörter und die Gewichte h_1, h_2 vor und finden das optimale Blockpaar $(\mathbf{c}_1, \mathbf{c}_2)$ durch Minimierung des Lagrange-Funktionals.

$$\min_{(\mathbf{c}_1,\mathbf{c}_2)} \left\{ \|\mathbf{s} - \mathbf{c}_1 h_1 - \mathbf{c}_2 h_2\|_2^2 + \lambda \left(|\gamma(\mathbf{c}_1) + |\gamma(\mathbf{c}_2)| \right) \right\} \tag{2.11}$$

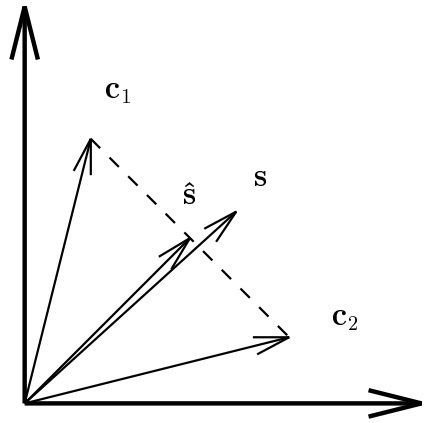


Abbildung 2.3: *Veranschaulichung zweier Hypothesen am 2-D Block*

Abbildung 2.3 veranschaulicht die Schätzung zweier Hypothesen am zwei-dimensionalen Block, d.h. an einem Block mit zwei Bildelementen. In dieser Darstellung werden die Blöcke $\mathbf{c}_1$, $\mathbf{c}_2$ und der Originalblock $\mathbf{s}$ als Vektoren abgebildet. Das Vektorpaar $(\mathbf{c}_1, \mathbf{c}_2)$ ist so zu wählen, daß der Lagrange-Abstand

$$\|\mathbf{s} - \mathbf{c}_1 h_1 - \mathbf{c}_2 h_2\|_2^2 + \lambda \left(|\gamma(\mathbf{c}_1)| + |\gamma(\mathbf{c}_2)| \right) \tag{2.12}$$

minimiert wird.

Schließlich fassen wir das Vorgehen bei der Schätzung der Bewegungsparameter bei der bewegungskompensierten Multihypothesen-Prädiktion mit n Hypothesen zusammen. Die Distortion zwischen Originalblock und geschätztem Block berechnen wir nach **Gleichung 2.5**. Den Bewegungscode $\mathbf{b}$ erhalten wir durch eine Codierung γ der Bewegungsparameter

$\mathbf{\Delta}_{x_\nu}$, $\mathbf{\Delta}_{y_\nu}$ und $\mathbf{\Delta}_{t_\nu}$ für jeden Block $\mathbf{c}_\nu$. Die Rate des Multibewegungscodes ist dabei die Summe der Längen der Bewegungscodewörter.

$$r(\mathbf{b}) = |\gamma(\mathbf{c})| = \sum_{\nu=1}^{n} |\gamma(\mathbf{c}_\nu)| \tag{2.13}$$

Sei die Codiervorschrift γ und die Koeffizienten

$$h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \tag{2.14}$$

der Linearkombination gegeben. Dann erhalten wir die optimale bewegungskompensierte Multihypothesen-Prädiktion für den Originalblock $\mathbf{s}$, wenn der Vektor der Blöcke

$$\mathbf{c} = \begin{pmatrix} \mathbf{c}_1 & \dots & \mathbf{c}_n \end{pmatrix} \tag{2.15}$$

das Lagrange-Funktional minimiert.

$$\min_{\mathbf{c}} \left\{ \|\mathbf{s} - \mathbf{c}h\|_2^2 + \lambda\, |\gamma(\mathbf{c})| \right\} \tag{2.16}$$

Wir haben ein Rate-Distortion-Kriterium angegeben, daß den optimalen Multibewegungscode für jeden Originalblock bestimmt. Ein Algorithmus, der eine lokal optimale Lösung zu diesem Minimierungsproblem liefert, wird im nächsten Abschnitt beschrieben.

2.2.2 Algorithmus zur Auswahl der Hypothesen

Bevor wir den Algorithmus zur Auswahl der Hypothesen angeben, beginnen wir mit einer Komplexitätsbetrachtung für die vollständige Suche von n Hypothesen. Lassen wir für jede Translation $(\mathbf{\Delta}_{x_\nu}, \mathbf{\Delta}_{y_\nu}, \mathbf{\Delta}_{t_\nu})$ einen Suchbereich der Größe $[-a, a] \times [-a, a] \times [-m, -1]$ zu, so ergibt sich für n Hypothesen eine Komplexität von

$$C_v = \left[m(2a+1)^2 \right]^n \tag{2.17}$$

Suchpositionen. Nehmen wir zum Beispiel für die örtliche Verschiebung ein Maximum von $a = 15$ bzw. für die zeitliche Verschiebung ein Maximum von $m = 10$ Positionen an. Für die Bestimmung von $n = 4$ Hypothesen bei vollständiger Suche ergibt sich eine Komplexität von $C_v = 8.5 \cdot 10^{15}$ Suchpositionen. Der hier angegebene lokal optimale Algorithmus wählt iterativ die Hypothesen bei einer deutlich geringeren Komplexität.

Der Algorithmus zur Auswahl von n Hypothesen setzt voraus, daß die Länge der Bewegungscodewörter $|\gamma(\cdot)|$ und die Koeffizienten h_ν gegeben sind.

Zur anschaulichen Erklärung wählen wir zwei Hypothesen. Der iterative Algorithmus gliedert sich für diesen Fall in vier Schritte.

0: Für die Initialisierung bestimmen wir durch vollständige Suche die beste Hypothese für den Fall $n = 1$. Die Position dieser besten Hypothese wird beiden Initialisierungshypothesen zugeordnet, wobei diese mit den Koeffizienten h_1 und h_2 assoziiert werden.

1: Hypothese $\mathbf{c}_2$ wird nun als konstant betrachtet. Für die bedingte Suche werden alle Positionen des bedingten Suchbereichs der Hypothese $\mathbf{c}_1$ geprüft und diejenige gewählt, die das Lagrage-Funktional am stärksten minimiert.

2: Die Hypothese $\mathbf{c}_1$ wird nun an der zuvor gefundenen Position fixiert und als konstant betrachtet. Eine zu Schritt 1 analoge bedingte Suche für die Hypothese $\mathbf{c}_2$ minimiert das Lagrange-Funktional maximal.

3: Solange das Lagrange-Funktional weiterhin minimiert werden kann, ist mit Schritt 1 fortzufahren. Andernfalls ist eine lokal optimale Lösung für die Auswahl von zwei Hypothesen gefunden.

Dem interativen Algorithmus ist zu entnehmen, daß in keinem Schritt eine Zunahme des Lagrange-Funktionals zugelassen wird; das Lagrange-Funktional ist somit über die Iterationen monoton abnehmend.

Betrachten wir den Algorithmus im Detail und lassen dabei n Hypothesen zu. Nach **Gleichung 2.16** ist das Lagrange-Funktional

$$j(\mathbf{c}) = \|\mathbf{s} - \mathbf{c}h\|_2^2 + \lambda\,|\gamma(\mathbf{c})| \tag{2.18}$$

über alle möglichen Hypothesenkombinationen zu minimieren. Der iterative Algorithmus vermeidet den Test aller möglichen Hypothesenkombinationen, indem er n bedingt optimale Lösungen sukzessiv verbessert und im Konvergenzfall eine lokal optimale Lösung liefert. Dazu sind in jeder Iteration jeweils n bedingt optimale Hypothesen zu bestimmen.

Wir betrachten für unsere Überlegungen die μ-te Hypothese ($1 \leq \mu \leq n$) von n Hypothesen gesondert. Wir benötigen für die bedingte Optimierung der μ-ten Hypothese einen bedingten Suchbereich, den wir für die bedingte Lösung $\mathbf{c}_\mu^{(i+1)}$ im $i+1$-ten Schritt vollständig absuchen. Dazu wählen wir eine lokale Umgebung um die Position der Hypothese $\mathbf{c}_\mu^{(i)}$, die wir im Schritt i bestimmt haben. Die Form des bedingten Suchbereichs ist ein 3-D-Quader $[-b, b] \times [-b, b] \times [-b, b]$, der durch den bedingten Suchbereichsparameter b festgelegt wird.

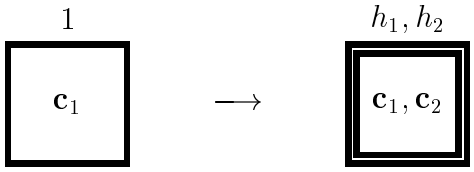


Abbildung 2.4: *Teilung einer Hypothese in zwei Hypothesen*

Die Initialisierung für den iterativen Algorithmus erfolgt dabei durch *Teilung einer Hypothese*. **Abbildung 2.4** veranschaulicht das Prinzip. Die optimale Hypothese für $n = 1$ wird wie üblich durch vollständige Suche mit Gewicht 1 gestimmt. Diese so gefundene Hypothese wird nun geteilt, d.h. die Position der geteilten Hypothesen ist identisch mit der der 1-Hypothese und die Gewichte der geteilten Hypothesen summieren sich zu 1. Die beiden geteilten Hypothesen stehen für die Initialisierung zur Verfügung. Sind n Hypothesen für die Initialisierung notwendig, so wird die 1-Hypothese n-fach geteilt.

Diese Art der Teilung aus einer Hypothese stellt bei stark bewegtem Sequenzinhalt eine gute Initialisierung sicher, da der iterative Algorithmus bei kleinem bedingten Suchbereich sehr schnell in ein lokales Optimum konvergiert. Bei Sequenzen mit geringer Bewegungsdynamik genügt für die Initialisierung die Null-Verschiebung in das vorherige Bild.

Für eine Realisierung ist der Konvergenzfall $i \to \infty$ nicht praktikabel. Da das Lagrange-Funktional über die Iterationen monoton abnimmt, wird mit dem Abbruchkriterium

$$\frac{j(\mathbf{c}^{(i)}) - j(\mathbf{c}^{(i+1)})}{j(\mathbf{c}^{(i)})} < 0.005 \tag{2.19}$$

die Quasi-Konvergenz geprüft. **Abbildung 2.5** faßt den Algorithmus zur Auswahl der Hypothesen zusammen.

0: Für das Optimierungsproblem ist bei n Hypothesen das Zielfunktional für jeden Originalblock $\mathbf{s}$ durch

$$j(\mathbf{c}_1, \ldots, \mathbf{c}_\mu, \ldots, \mathbf{c}_n) = \left\| \mathbf{s} - \sum_{\substack{\nu=1 \\ \nu \neq \mu}}^{n} \mathbf{c}_\nu h_\nu - \mathbf{c}_\mu h_\mu \right\|_2^2 + \lambda \left(\sum_{\substack{\nu=1 \\ \nu \neq \mu}}^{n} |\gamma(\mathbf{c}_\nu)| + |\gamma(\mathbf{c}_\mu)| \right)$$

definiert. Gegeben seien die Koeffizienten h_ν, die Längen der Verschiebungscodewörter $|\gamma(\cdot)|$ und der Lagrange-Parameter λ. Initialisiere den Algorithmus mit n Hypothesen $(\mathbf{c}_1^{(0)}, \ldots, \mathbf{c}_n^{(0)})$ und setze $i := 0$.

1: Wähle jeweils aus den n Hypothesen die μ-te aus. Beginne dabei mit der ersten Hypothese und ende mit der n-ten.

a: Betrachte alle außer der μ-ten Hypothese als fixiert. Wähle für den bedingten Suchbereich der Hypothese $\mathbf{c}_\mu^{(i+1)}$ eine lokale Umgebung der Hypothese $\mathbf{c}_\mu^{(i)}$.

b: Minimiere das Lagrange-Funktional durch vollständige Suche im bedingten Suchbereich der Hypothese $\mathbf{c}_\mu^{(i+1)}$.

$$\min_{\mathbf{c}_\mu^{(i+1)}} j(\mathbf{c}_1^{(i+1)}, \ldots, \mathbf{c}_{\mu-1}^{(i+1)}, \mathbf{c}_\mu^{(i+1)}, \mathbf{c}_{\mu+1}^{(i)}, \ldots, \mathbf{c}_n^{(i)})$$

2: Solange das Lagrange-Funktional minimiert werden kann, fahre mit Schritt 1 fort und setzt $i := i + 1$.

Abbildung 2.5: *Algorithmus zur Auswahl der Hypothesen*

Theoretisch kann man das Schätzproblem zum Auffinden der n Hypothesen als eine *maximum a posteriori (MAP)* Schätzung auffassen. Nach der Bayes'schen Vorstellung entspricht dem beobachteten Ereignis der Originalblock $\mathbf{s}$ und dem Modell-Ereignis der Vektor der Hypothesen $\mathbf{c}$. Für die MAP Schätzung ist die bedingte Wahrscheinlichkeit des Modell-Ereignises $\mathbf{c}$ bei beobachtetem Ereignis $\mathbf{s}$ zu maximieren.

$$\max_{\mathbf{c}} P(\mathbf{c}|\mathbf{s}) \tag{2.20}$$

Die MAP-Wahrscheinlichkeit läßt sich nach dem Bayes'schen Satz [15] derart faktorisieren, daß die μ-te Hypothese explizit für eine bedingte Optimierung zur Verfügung steht.

$$P(\mathbf{c}|\mathbf{s}) = P(\mathbf{c}_\mu|\mathbf{c}_1, \ldots, \mathbf{c}_{\mu-1}, \mathbf{c}_{\mu+1}, \ldots, \mathbf{c}_n, \mathbf{s}) P(\mathbf{c}_1, \ldots, \mathbf{c}_{\mu-1}, \mathbf{c}_{\mu+1}, \ldots, \mathbf{c}_n|\mathbf{s}) \tag{2.21}$$

Die daraus resultierende iterative Lösung des MAP Schätzproblems wird in [12] als *Iterated Conditional Modes (ICM)* bezeichnet. Diese Methode verhindert ein Abnehmen der MAP-Wahrscheinlichkeit während der Iteration und stellt somit die Konvergenz sicher.

Die Komplexitätsbetrachtung für den iterativen Algorithmus führt zu folgenden Ergebnissen: Für jede bedingte Suche wird ein Bereich der Größe $[-b, b] \times [-b, b] \times [-b, b]$ untersucht. Die Initialisierung erfolgt mit der Lösung für eine Hypothese. Für n Hypothesen und I Iterationen ergibt sich eine Komplexität von

$$C_i = m(2a+1)^2 + In(2b+1)^3 \tag{2.22}$$

Suchpositionen. Ergänzen wir das Beispiel zu Beginn des Abschnitts, indem wir für den bedingten Suchbereich ein Maximum von $b = 4$ Positionen und $I = 3$ Iterationen zulassen, so erhalten wir für die Komplexität des iterativen Algorithmus $C_i = 1.8 \cdot 10^4$ Suchpositionen und reduzieren somit den Aufwand im Vergleich zur vollständigen Suche um den Faktor $4.6 \cdot 10^{11}$.

Abbildungen 2.6 und 2.7 zeigen Simulationsergebnisse für die Testsequenzen. Dabei gelten die Simulationsbedingungen nach **Anhang C**. Die Raten-Nebenbedingung hat auf die Ergebnisse keinen Einfluß ($\lambda = 0$). Wir lassen eine maximale örtliche Verschiebung von $a = 15$ zu. Für den bedingten Suchbereich setzen wir $b = 4$. Die Abbildungen zeigen Ergebnisse für eine Suchraumgröße von einem Bild ($m = 1$) und von zehn Bildern ($m = 10$). Für die Half-Pel-Suchpositionen sind die Originalbilder bilinear interpoliert. Die Gewichtskoeffizienten sind für alle Hypothesen zu $\frac{1}{n}$ gesetzt.

Die Ergebnisse lassen dabei folgende Beobachtungen zu.

1. Zwei Hypothesen erzielen einen größeren Prädiktionsgewinn als eine Prädiktionsgenauigkeit von einem halben Bildelement. Dies steht im Einklang zu den theoretischen Aussagen in [7].

2. Eine Verdoppelung der Anzahl der Hypothesen verursacht keine konstante Zunahme des Prädiktionsgewinns. Dies steht im Widerspruch zu der theoretischen Aussage in [7], die statistische Unabhängigkeit zwischen den Hypothesen annimmt.

3. Die Effekte der Multihypothesen-Prädiktion und der Langzeitprädiktion addieren sich. Darüber hinaus ist festzustellen, daß sich durch die Vergrößerung des Suchbereichs bei der Langzeitprädiktion ein zusätzlicher Gewinn bei der Multihypothesen-Prädiktion erzielen läßt. Dieser beträgt bei $n = 8$ Hypothesen und $m = 10$ Bildern ca. 0.4 dB.

4. Der Gewinn durch Half-Pel-Prädiktion nimmt im Vergleich zur Integer-Pel-Prädiktion für zunehmende Hypothesenzahl ab. Interpretiert man Half-Pel-Prädiktion als Multihypothesen-Prädiktion [7], so ist dieser Effekt in Beobachtung 1 bereits festgestellt. Half-Pel-Prädiktion erhöht im Prinzip die effektive Anzahl an Hypothesen.

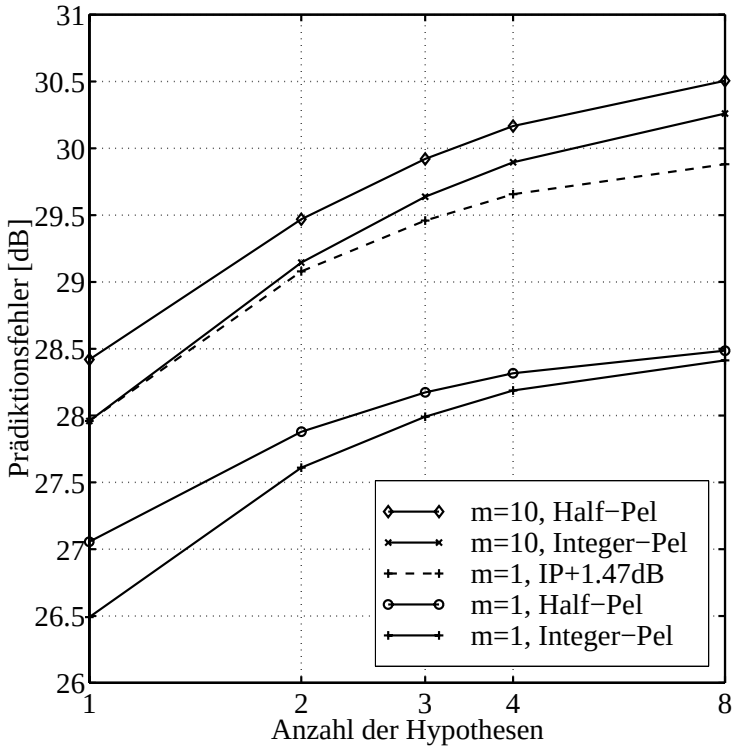


Abbildung 2.6: *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke,* $b = 4$ *und* $\lambda = 0$

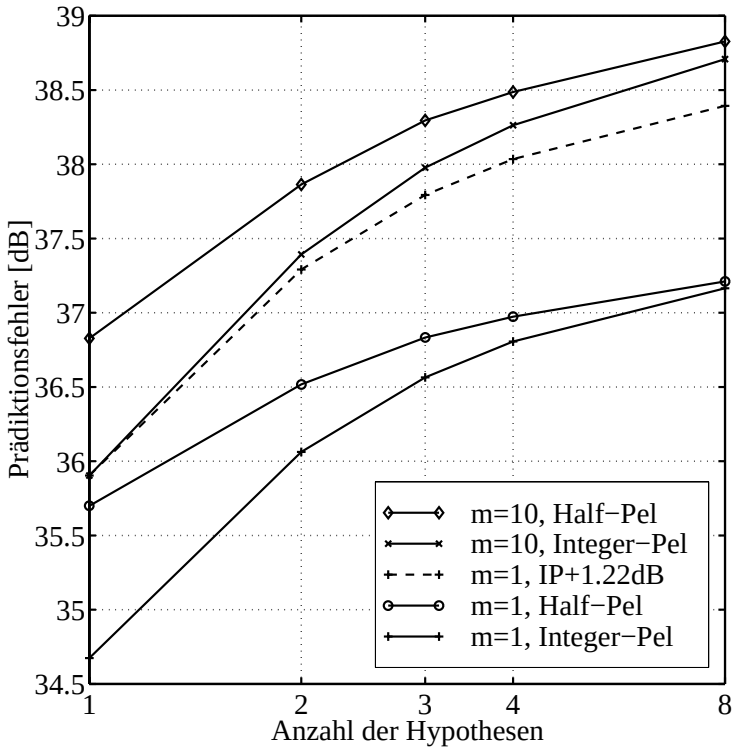


Abbildung 2.7: *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke,* $b = 4$ *und* $\lambda = 0$

Abbildungen 2.8 und 2.9 zeigen Simulationsergebnisse für variierende Größe des bedingten Suchbereichs. Es wird ein Maximum von $b = 2$, $b = 4$ und $b = 8$ zugelassen. Bei Langzeitprädiktion ist eine größere Sensitivität bezüglich der Größe des bedingten Suchbereichs festzustellen, da diese auch in Zeitrichtung variiert wird.

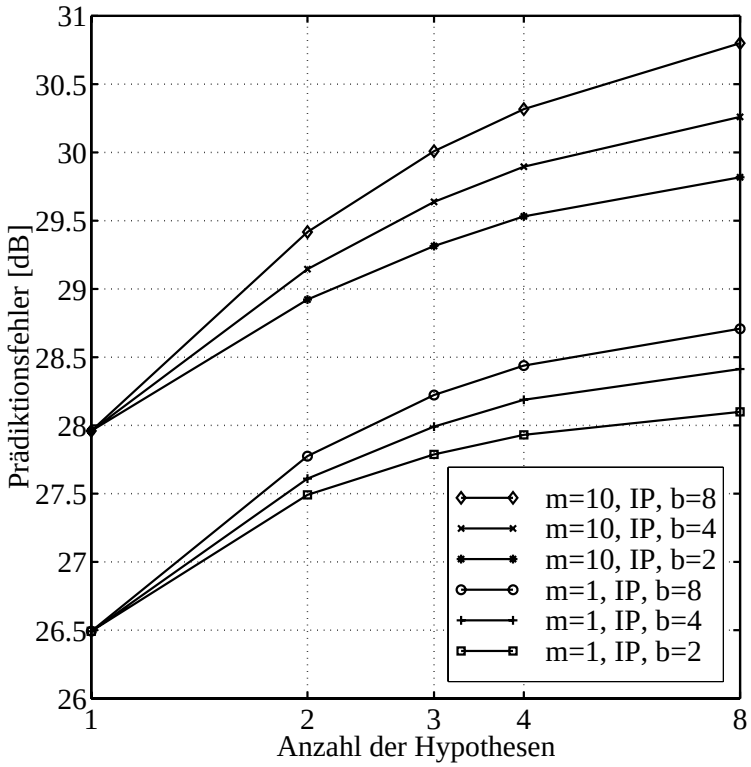


Abbildung 2.8: *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke und* $\lambda = 0$ *in Abhängigkeit der Größe des bedingten Suchbereichs*

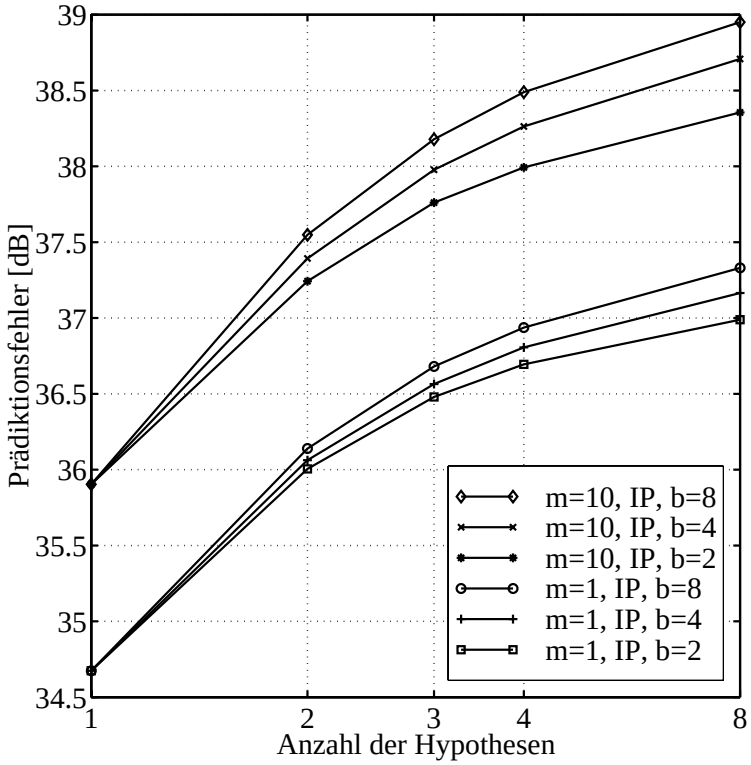


Abbildung 2.9: *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke und* $\lambda = 0$ *in Abhängigkeit der Größe des bedingten Suchbereichs*

Abbildungen 2.10 und 2.11 zeigen Simulationsergebnisse für variierende Blockgrößen. Die Testsequenzen werden in 4×4, 8×8 bzw. 16×16 Blöcke eingeteilt. Die Prädiktionsfehler können verglichen werden, da die Distortion-Werte auf Bildelemente bezogen

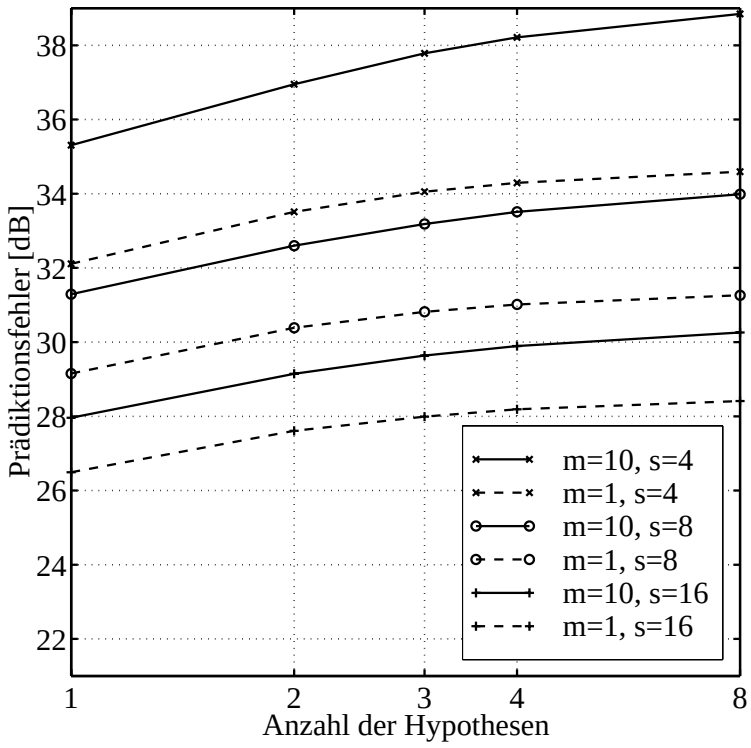


Abbildung 2.10: *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s), $b = 4$ und $\lambda = 0$*

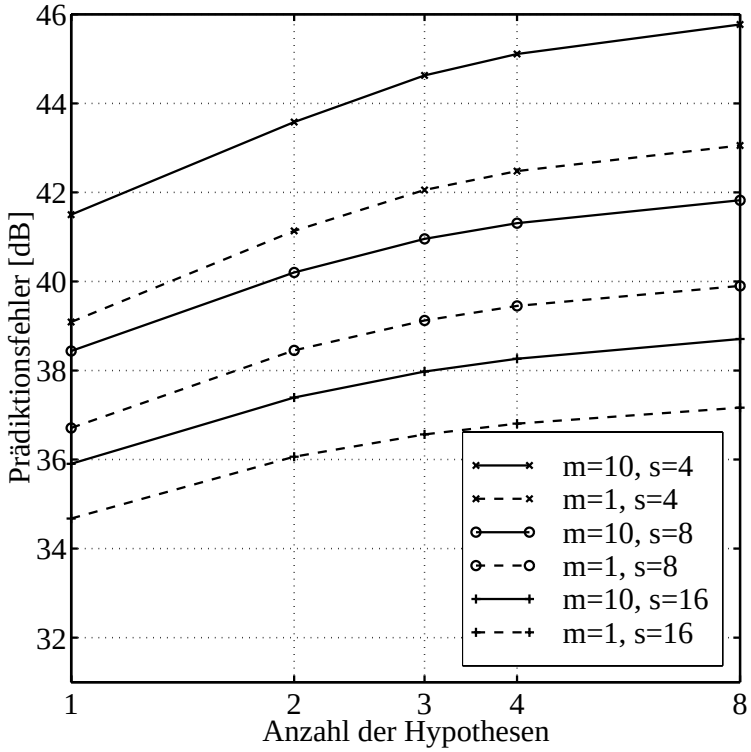


Abbildung 2.11: *Prädiktionsfehler über der Anzahl der Hypothesen für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s), $b = 4$ und $\lambda = 0$*

werden. Genaueres dazu findet man im **Anhang C**. Für die Simulationen bleiben die obigen Randbedingungen erhalten, wobei wir im folgenden $b = 4$ wählen.

1. Grundsätzlich ist festzustellen, daß ein Gewinn durch Codierung mehrerer Hypothe-

sen unabhängig von der Blockgröße zu beobachten ist. Dieser Effekt ist unabhängig von der Größe des Suchraumes.

2. Dieser Gewinn vergrößert sich leicht für abnehmende Blockgrößen. Durch Verringerung der Dimensionalität der zu schätzenden Vektoren (Blöcke) bei gegebener Hypothesenzahl ist es möglich, exakter zu prädizieren.

Prinzipiell scheint es sinnvoll, einen mit 4 Hypothesen codierten 16×16 Block mit 4 8×8 Blöcken, die jeweils mit einer Hypothese spezifiziert sind, zu vergleichen. Beide Codiermethoden führen für die Testsequenz „Mother and Daughter“ zu vergleichbaren Prädiktionsergebnissen, wohingegen für die Testsequenz „Foreman“ im Mittel eine Zerlegung in kleinere Blöcke effizienter ist. Da aber bei der vorliegenden Simulation alle Blöcke jeweils nach der gleichen Methode codiert sind, ist eine Aussage über die Effizienz der beiden Methoden für einzelne Blöcke nicht möglich. Um eine Entscheidung darüber treffen zu können, ist es notwendig, beide Codiermethoden für jeden Block zuzulassen.

2.2.3 Optimale Anzahl der Hypothesen

Bisher wurden alle Blöcke mit konstanter Anzahl an Hypothesen codiert. Für ein Rate-Distortion-optimiertes Prädiktionsverfahren ist es aber notwendig, die optimale Hypothesenzahl für jeden Block zu bestimmen. Diese Notwendigkeit untersuchen wir in diesem Abschnitt.

Um die Rate-Distortion-optimale Anzahl an Hypothesen $\hat{n}$ bei gegebener maximaler Hypothesenzahl N für jeden Originalblock zu bestimmen, wird diejenige Codierung gewählt, die die geringsten Lagrange-Kosten für jeden Block besitzt.

$$\min_{n:1 \leq n \leq N} \left\{ \left\| \mathbf{s} - \mathbf{c}^{(n)} h^{(n)} \right\|_2^2 + \lambda |\gamma(\mathbf{c}^{(n)})| \right\} \tag{2.23}$$

Für N optimale Multihypothesen-Codierungen ist die Entscheidung nach **Gleichung 2.23** optimal. Die n optimalen Hypothesen werden unabhängig von den $n+1$ optimalen Hypothesen gefunden.

Abbildungen 2.12 und 2.13 zeigen Simulationsergebnisse zur optimalen Anzahl der Hypothesen. **Anhang C** erläutert die generellen Randbedingungen. Die Prädiktion wird aus den 10 vorherigen Bildern bei Integer-Pel-Genauigkeit bestimmt. Weiterhin verwenden wir den optimalen Huffman-Code für die Bewegungsparameter und die optimalen Interpolationskoeffizienten, deren Berechnung wir im nächsten Abschnitt untersuchen. Wir vergleichen den Prädiktionsfehler über der Rate des Multihypothesencodes für konstante und optimale Hypothesenzahlen. Der Lagrange-Parameter wird dabei exponentiell variiert $(25, 50, 100, \ldots, 1600)$.

Der Algorithmus zur Bestimmung der optimalen Hypothesenzahl stellt sicher, daß sich das Rate-Distortion-Verhalten für abnehmende Rate des Multihypothesencodes dem Verhalten einer Hypothese anpaßt. Gleichzeitig bleibt für maximale Rate der Gewinn durch die Multihypothesen-Codierung von bis zu $N = 4$ Hypothesen erhalten. Die Ergebnisse lassen darauf schließen, daß durch die Raten-Nebenbedingung die Originalblöcke mit unterschiedlichen Hypothesenzahlen codiert werden und somit ein verbessertes Rate-Distortion-Verhalten festzustellen ist. Diese Verbesserung ist durch zwei Modelle zu erklären.

1. Die Zerlegung eines Originalblocks in n_1 Hypothesen bzw. in n_2 Hypothesen hat keinen Einfluß auf den Prädiktionsfehler.

2. Eine Erhöhung des Prädiktionsfehlers durch eine Codierung mit weniger Hypothesen wird durch die Ratenreduktion mehr als ausgeglichen.

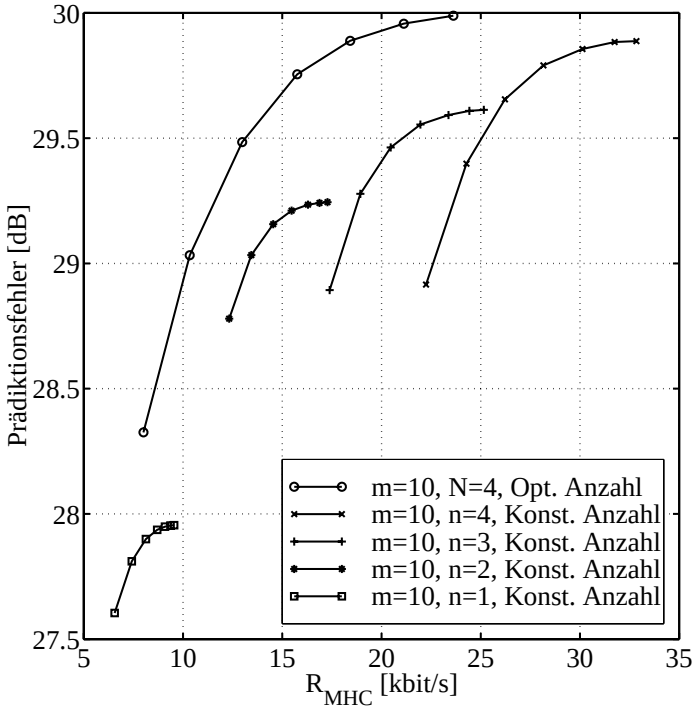


Abbildung 2.12: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke*

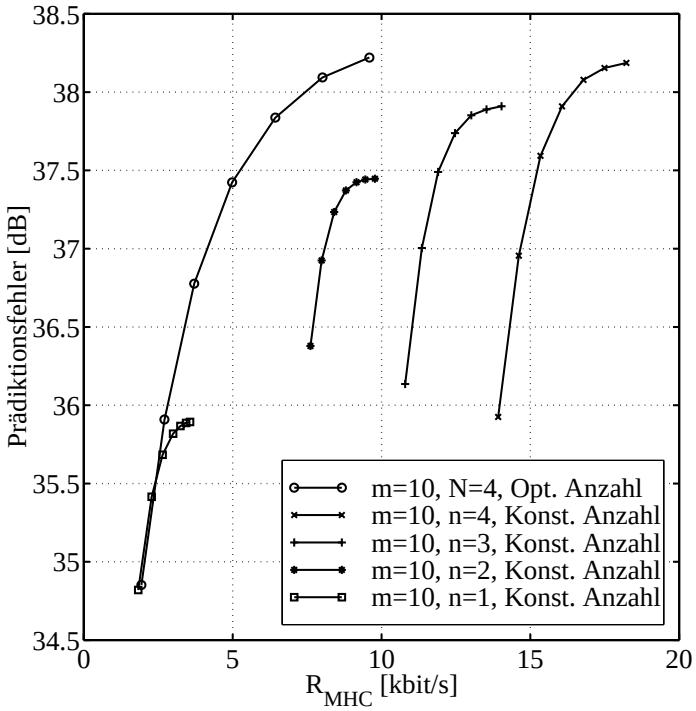


Abbildung 2.13: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke*

Abbildungen2.14 und 2.15 visualisiert die Aufteilung der Rate des Multihypothesencodes; je größer die Rate, desto dominanter wird die maximal zulässige Hypothesenzahl. Hypothese Null kennzeichnet einen uncodierten Block (Sonderfall einer 1-Hypothese).

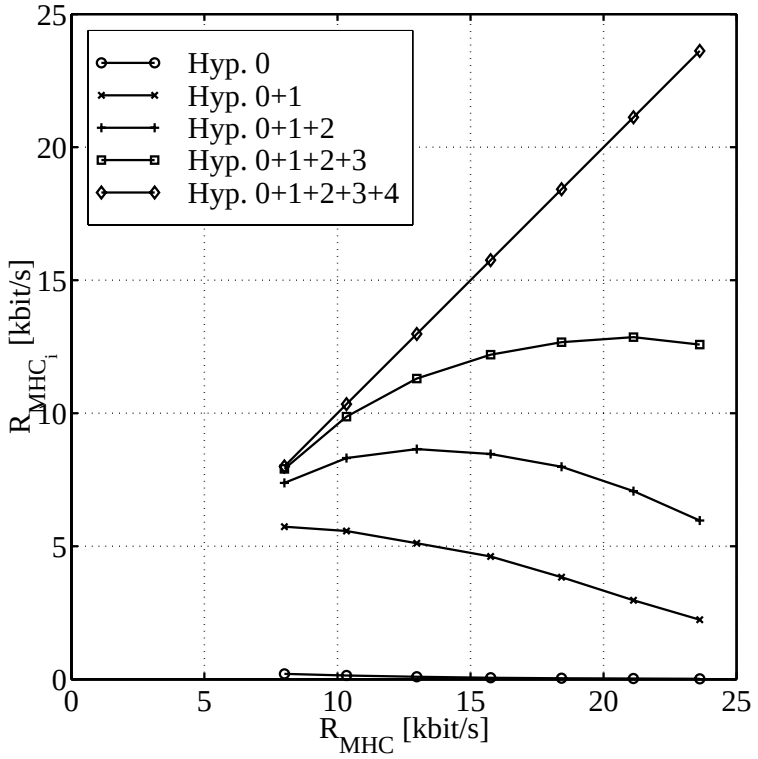


Abbildung 2.14: *Partielle Raten über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke, Integer-Pel-Genauigkeit,* $N = 4$ *und* $m = 10$

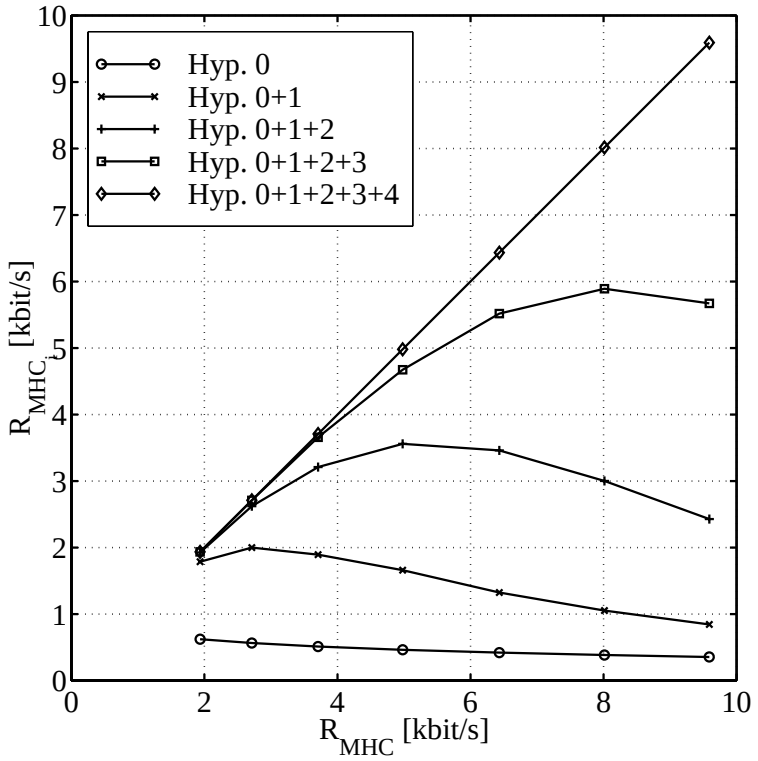


Abbildung 2.15: *Partielle Raten über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke, Integer-Pel-Genauigkeit,* $N = 4$ *und* $m = 10$

Abbildungen 2.16 und 2.17 zeigen Ergebnisse zur Multihypothesen-Prädiktion bei Half-Pel-Genauigkeit. Die obigen Simulationsrandbedingungen gelten weiterhin. Es ist festzustellen, daß sich der Half-Pel-Gewinn für zunehmende Hypothesenzahl reduziert und nur noch für hohe Bewegungsraten zu beobachten ist. Dies ist damit zu erklären, daß ein Teil der Half-Pel-Positionen bei einer Hypothese bereits durch die Multihypothesen-Codierung mit Integer-Pel-Genauigkeit berücksichtigt werden.

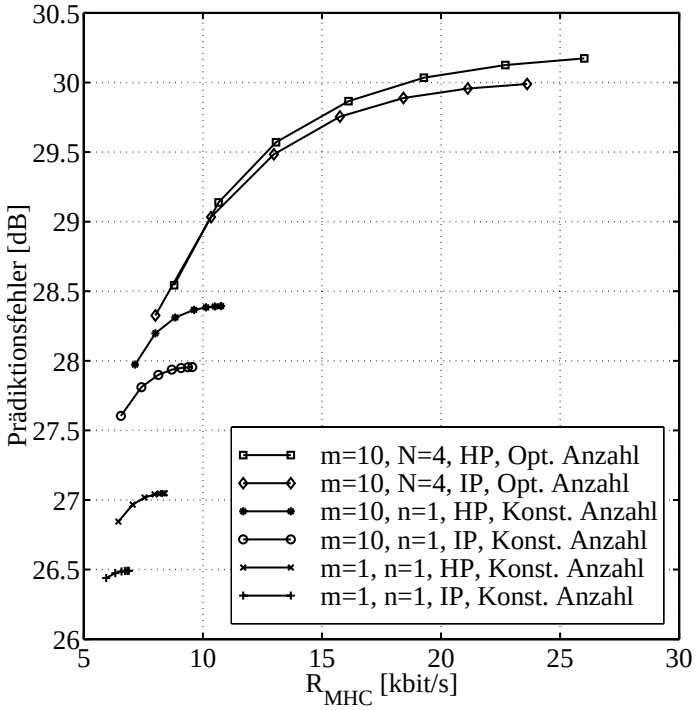


Abbildung 2.16: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke*

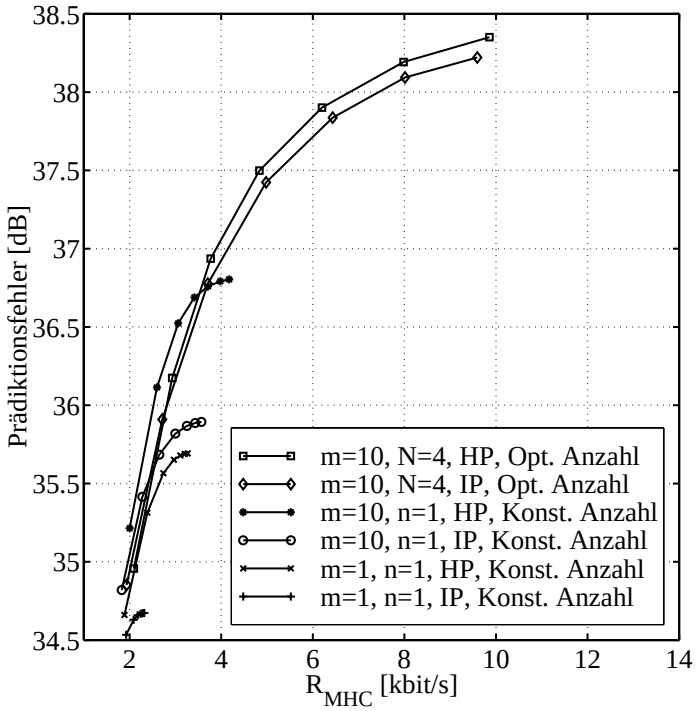


Abbildung 2.17: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke*

Abbildungen 2.18 und 2.19 zeigen Ergebnisse zur Multihypothesen-Prädiktion von Originalblöcken mit unterschiedlicher Größe für verschiedene Multihypothesen-Raten.

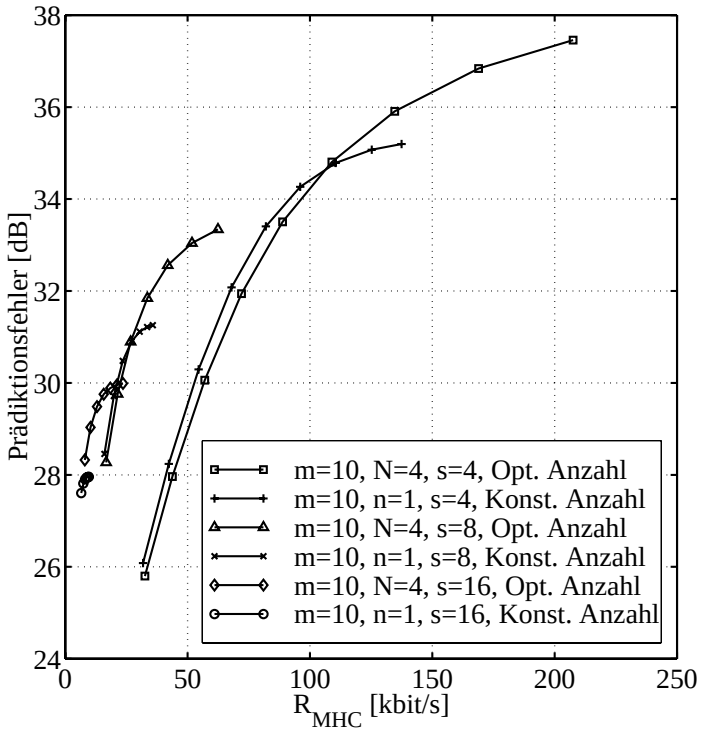


Abbildung 2.18: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke*

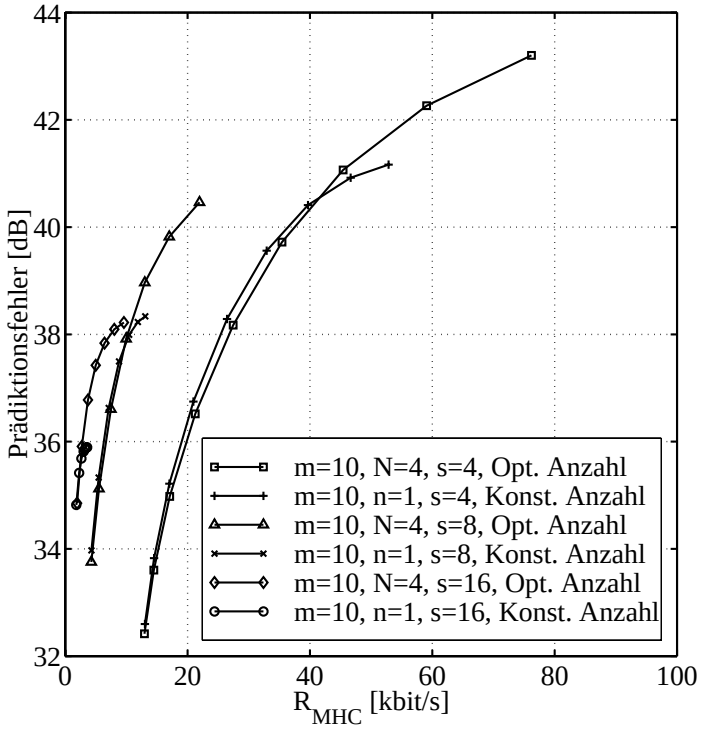


Abbildung 2.19: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcke*

Wie wir bereits aus den Simulationen zur Blockgröße bei $\lambda = 0$ (**Abbildungen 2.10 und 2.11**) gesehen haben, tritt der Multihypothesen-Effekt unhabhängig von der Blockgröße

auf. Diese Simulation gibt Auskunft, bei welcher Rate der Effekt eintritt. Der Wert für λ, bei dem der Effekt relevant wird, verschiebt sich für abnehmende Blockgrößen zu kleineren Werten, d.h. der Gewinn stellt sich nur bei zunehmend höheren Raten ein.

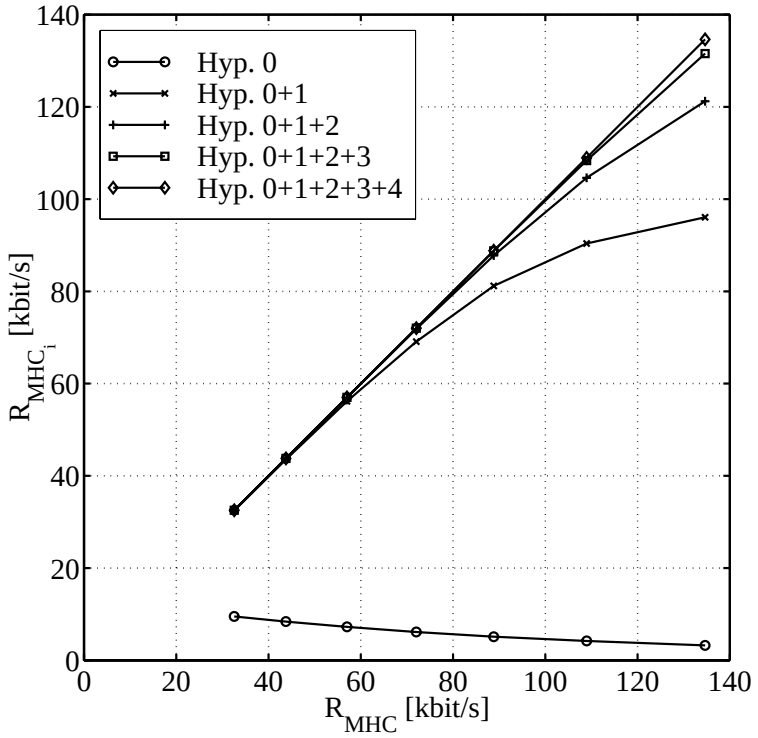


Abbildung 2.20: *Partielle Raten über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s), 4×4 Blöcke, Integer-Pel-Genauigkeit, $N = 4$ und $m = 10$*

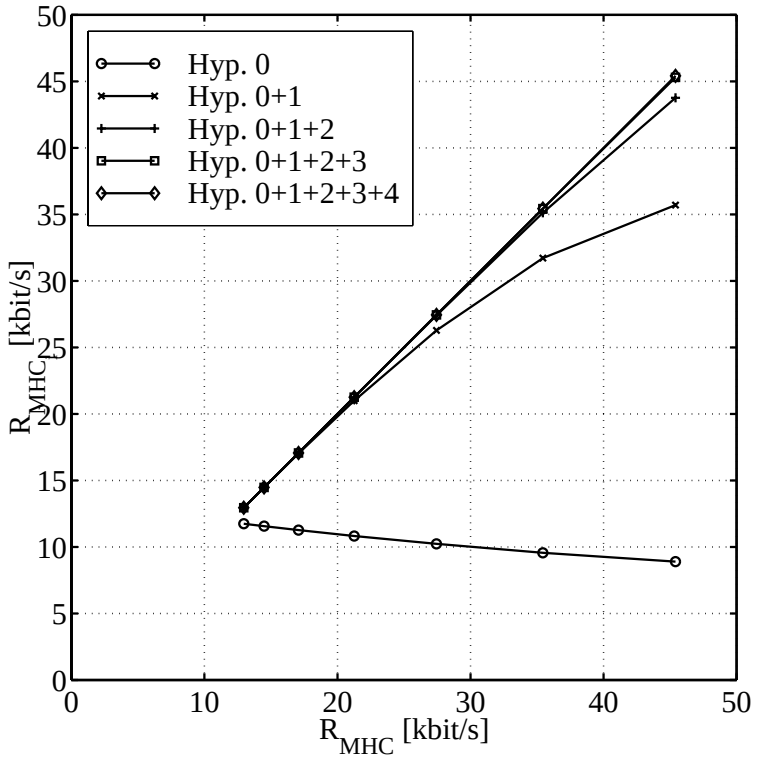


Abbildung 2.21: *Partielle Raten über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s), 4×4 Blöcke, Integer-Pel-Genauigkeit, $N = 4$ und $m = 10$*

Abbildungen2.20 und 2.21 visualisiert die Aufteilung der Rate des Multihypothesencodes für 4×4 Blöcke durch eine Darstellung der kumulierten partiellen Raten. Hypothese Null kennzeichnet einen uncodierten Block, der ein Sonderfall einer 1-Hypothese ist. In dieser Darstellung wird der Multihypothesen-Effekt durch einen Abfall der partiellen Rate der 1-Hypothese unter die Winkelhalbierende sichtbar.

2.3 Entwurf des Multihypothesen-Prädiktors

In den vorangehenden Abschnitten wurden bei gegebener Codiervorschrift $\gamma(\cdot)$ und gegebenen Interpolationskoeffizienten h die lokal optimalen Hypothesen bestimmt. Diese bedingte Optimierung werden wir in diesem Abschnitt verallgemeinern und im Rahmen eines Rate-Distortion-Modells sowohl die optimale Codiervorschrift als auch die optimalen Interpolationskoeffizienten in einem iterativen Entwurfsalgorithmus determinieren. Dazu betrachten wir das Problem der bewegungskompensierten Prädiktion als Vektorquantisierung und den Entwurf des Multihypothesen-Prädiktors als Quantisierer-Entwurf. Der *Generalized Lloyd Algorithm* (GLA) [14] bzw. die um die Entropie-Nebenbedingung erweiterte Variante von *Chou, Lookabaugh and Gray* (CLGA) [13] lösen dieses Entwurfsproblem.

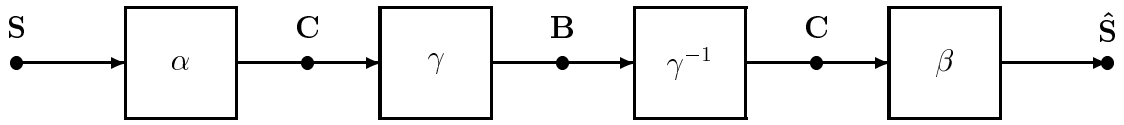


Abbildung 2.22: *Quantisierungsmodell*

Abbildung 2.22 zeigt ein abstraktes Modell für die bewegungskompensierte Prädiktion. Die Originalblöcke $\mathbf{S}$ werden durch die Multihypothesen-Codierung α in ihre Hypothesen $\mathbf{C}$ zerlegt. Die örtlichen und zeitlichen Positionen der Hypothesen werden durch eine Entropiecodierung γ auf einen Multihypothesen-Bewegungscode $\mathbf{B}$ abgebildet. Bei der Bewegungskompensation wird die Entropiecodierung invertiert und die daraus resultierenden Hypothesen in der Abbildung β durch eine Linearkombination auf die geschätzten Blöcke $\hat{\mathbf{S}}$ transformiert.

Die Optimierung des Quantisierungsmodells mit Raten-Nebenbedingung lösen wir durch Minimierung des Lagrange-Funktionals.

$$J(\alpha, \beta, \gamma, \lambda, \mathbf{S}) = E\left\{\left\|\mathbf{S} - \hat{\mathbf{S}}\right\|_2^2 + \lambda|\mathbf{B}|\right\} \tag{2.24}$$

Mit Hilfe des Quantisierungsmodells stellen wir die geschätzten Blöcke durch die Abbildungen α und β,

$$\hat{\mathbf{S}} = \beta \circ \alpha(\mathbf{S}) \tag{2.25}$$

bzw. den Bewegungscode $\mathbf{B}$ durch die Abbildung α und γ

$$\mathbf{B} = \gamma \circ \alpha(\mathbf{S}) \tag{2.26}$$

dar. Somit erhalten wir für das Lagrange-Funktional

$$J(\alpha, \beta, \gamma, \lambda, \mathbf{S}) = E\left\{\|\mathbf{S} - \beta \circ \alpha(\mathbf{S})\|_2^2 + \lambda|\gamma \circ \alpha(\mathbf{S})|\right\} \tag{2.27}$$

einen Ausdruck, der nur von den drei Abbildungen, λ und der Zufallsvariablen $\mathbf{S}$ der Originalblöcke abhängt. Die Minimierung des Lagrange-Funktionals

$$\min_{\alpha,\beta,\gamma} J(\alpha, \beta, \gamma, \lambda_c, \mathbf{S}) \tag{2.28}$$

für die drei Abbildungen bei konstantem $\lambda = \lambda_c$ und konstanter Wahrscheinlichkeitsfunktion der Originalblöcke $f_{\mathbf{S}}(\mathbf{s})$ erfolgt in Anlehnung an den CLGA durch einen iterativen Entwurfsalgorithmus.

2.3.1 Iterativer Entwurfsalgorithmus

Im folgenden wählen wir $\lambda = \lambda_c$ als konstant und geben uns eine Wahrscheinlichkeitsfunktion der Originalblöcke $f_{\mathbf{S}}(\mathbf{s})$ vor.

1. Im ersten Teilschritt bestimmen wir die optimalen Hypothesen $\mathbf{c} = \alpha(\mathbf{s})$. Dazu nehmen wir die Abbildungen β und γ als konstant an.

$$\begin{aligned} & \min_{\alpha} E\left\{\|\mathbf{S} - \beta_c \circ \alpha(\mathbf{S})\|_2^2 + \lambda_c |\gamma_c \circ \alpha(\mathbf{S})|\right\} \\ \Longrightarrow \quad & \alpha(\mathbf{s}) = \operatorname*{argmin}_{\mathbf{c}} \left\{\|\mathbf{s} - \mathbf{c}h_c\|_2^2 + \lambda_c |\gamma_c(\mathbf{c})|\right\} \end{aligned} \tag{2.29}$$

   Gleichung 2.29 beschreibt das Schätzkriterium zum Auffinden der optimalen Hypothesen nach **Abschnitt 2.2.1**

2. Im zweiten Teilschritt bestimmen wir die optimale Abbildung γ. Dazu nehmen wir die Abbildung α und β als konstant an und erhalten somit auch eine konstante Wahrscheinlichkeitsfunktion für die Hypothesen $f_{\mathbf{C}}(\mathbf{c})$.

$$\begin{aligned} & \min_{\gamma} E\left\{\|\mathbf{S} - \beta_c \circ \alpha_c(\mathbf{S})\|_2^2 + \lambda_c |\gamma \circ \alpha_c(\mathbf{S})|\right\} \\ \Longrightarrow \quad & \min_{\gamma} E\left\{|\gamma \circ \alpha_c(\mathbf{S})|\right\} \\ \Longrightarrow \quad & \min_{\gamma} E\left\{|\gamma(\mathbf{C})|\right\} \end{aligned} \tag{2.30}$$

   Gleichung 2.30 besagt, daß der optimale Code eine mittlere Codewortlänge von $\sum_{\mathbf{c}} |\gamma(\mathbf{c})| f_{\mathbf{C}}(\mathbf{c})$ aufzuweisen hat. Für eine endliche Anzahl von Indizes wird das Minimierungsproblem mit dem Huffman-Algorithmus (**Anhang B**) gelöst.

3. Im dritten Teilschritt bestimmen wir die optimale Abbildung β. Dazu nehmen wir die Abbildungen α und γ als konstant an.

$$\begin{aligned} & \min_{\beta} E\left\{\|\mathbf{S} - \beta \circ \alpha_c(\mathbf{S})\|_2^2 + \lambda_c |\gamma_c \circ \alpha_c(\mathbf{S})|\right\} \\ \Longrightarrow \quad & \min_{h} E\left\{\|\mathbf{S} - \mathbf{C}_c h\|_2^2\right\} \end{aligned} \tag{2.31}$$

   Gleichung 2.31 wird als *Wiener-Problem* (**Anhang A**) bezeichnet und bestimmt die optimalen Interpolationskoeffizienten. Die Terminologie der Vektorquantisierung bezeichnet $\mathbf{c}h$ als den besten Zentroiden der Voronoi-Region $\{\mathbf{S}|\alpha_c(\mathbf{S}) = \mathbf{c}\}$ [16].

Analog zum iterativen Algorithmus zur Auswahl der Hypothesen verursacht jeder Schritt keine Verschlechterung, d.h. das Lagrange-Funktional ist auch für diesen Entwurfsalgorithmus monoton abnehmend. Die Quasi-Konvergenz kann ebenfalls durch ein Abbruchkriterium bestimmt werden. Praktisch ist der Algorithmus konvergiert, wenn der relative Fehler aufeinanderfolgender Interationsschritte kleiner als 0.5% ist.

0: Initialisiere den Algorithmus mit einer Entropiecodierung γ und mit Interpolationskoeffizienten h. Stelle sicher, daß die Wahrscheinlichkeitsfunktion der Originalblöcke $f_{\mathbf{S}}(\mathbf{s})$ über die Iterationen konstant bleibt. Wähle einen konstanten Lagrange-Parameter λ.

1: Bei gegebener Entropiecodierung der Bewegungsparameter γ und Interpolationskoeffizienten h gilt für die *optimale Multihypothesen-Codierung* des Originalblocks $\mathbf{s}$:

$$\min_{\mathbf{c}} \left\{ \|\mathbf{s} - \mathbf{c}h\|_2^2 + \lambda|\gamma(\mathbf{c})| \right\}$$

2: Bei gegebener Multihypothesen-Codierung $\mathbf{C}$ und Interpolationskoeffizienten h erhält man für die *optimale Entropiecodierung* γ eine Huffmancodierung.

$$\min_{\gamma} E\left\{|\gamma(\mathbf{C})|\right\}$$

3: Bei gegebener Multihypothesen-Codierung $\mathbf{C}$ und Entropiecodierung der Bewegungsparameter γ gilt für die *optimalen Interpolationskoeffizienten* h:

$$\min_{h} E\left\{ \|\mathbf{S} - \mathbf{C}h\|_2^2 \right\}$$

4: Solange eine weitere Minimierung möglich ist, fahre mit Schritt 1 fort.

Abbildung 2.23: *Iterativer Entwurfsalgorithmus*

Abbildung 2.23 faßt den iterativen Entwurfsalgorithmus zusammen. Die Wahrscheinlichkeitsfunktion der Originalblöcke $f_{\mathbf{S}}(\mathbf{s})$ ist unbekannt und somit durch eine Trainingsmenge zu approximieren. Der Einfluß der Trainingsmenge auf die Konvergenz wird hier nicht weiter untersucht. An dieser Stelle sei auf die Untersuchungen von M.J. Sabin und R.M. Gray zum *Generalized Lloyd Algorithm* [17] hingewiesen.

2.3.2 Optimale Interpolationskoeffizienten

Im dritten Schritt des iterativen Entwurfsalgorithmus werden die bedingt optimalen Interpolationskoeffizienten durch die Wiener-Lösung bestimmt. In diesem Abschnitt geben wir eine Lösung des Wiener-Problems an, die zusätzlich eine weitere Nebenbedingung erfüllt.

Dazu formulieren wir die Optimalitätsbedingung nach **Gleichung 2.31** mit Hilfe der Autokorrelation der Originalblöcke $\varphi_{\mathbf{SS}}$, der Autokorrelation der Hypothesen $\varphi_{\mathbf{CC}}$ und der Kreuzkorrelation $\varphi_{\mathbf{CS}}$.

$$\min_{h} \left\{ \varphi_{\mathbf{SS}} - 2h^T \varphi_{\mathbf{CS}} + h^T \varphi_{\mathbf{CC}} h \right\} \tag{2.32}$$

Die Mittelwerte der Blöcke sind relativ groß und um numerische Probleme bei der Lösung zu vermeiden, formulieren wir das bedingte Optimierungsproblem mit Kovarianzen. Dies ist möglich, da wir für das Schätzproblem grundsätzlich eine Identität zwischen mittlerem Originalblock und mittlerem geschätzten Block fordern.

$$E\left\{\mathbf{S}\right\} = E\left\{\hat{\mathbf{S}}\right\} \tag{2.33}$$

In der Kovarianz-Notation des bedingten Optimierungsproblems

$$\min_h \left\{\sigma^2_{\mathbf{SS}} - 2h^T\sigma^2_{\mathbf{CS}} + h^T\sigma^2_{\mathbf{CC}}h\right\} \tag{2.34}$$

bezeichnet $\sigma^2_{\mathbf{SS}}$ die Varianz des Originalblocks, $\sigma^2_{\mathbf{CC}}$ die Kovarianz der Hypothesen und $\sigma^2_{\mathbf{CS}}$ den Vektor der Kreuzterme. Die Kovarianzen sind dabei aus den Daten der Trainingsmenge zu schätzen.

Neben den Wiener-Bedingungen (**Gleichungen 2.33 und 2.34**) ist für die Bestimmung der bedingten optimalen Interpolationskoeffizienten eine weitere Nebenbedingung notwendig. Dazu betrachten wir den mittleren Block $\mu_{\mathbf{C}_\nu}$ der ν-ten Hypothese $\mathbf{C}_\nu$.

$$\mu_{\mathbf{C}_\nu} = E\left\{\mathbf{C}_\nu\right\} \tag{2.35}$$

Der Algorithmus zur Auswahl der Hypothesen stellt sicher, daß alle n Hypothesen $\mathbf{c}_1, \ldots,$ $\mathbf{c}_n$ aus dem gleichen Suchraum stammen. Der Suchraum der ν-ten Hypothese ist identisch mit dem der $\nu+1$-ten Hypothese. Somit ist der mittlere Block der ν-ten Hypothese $\mu_{\mathbf{C}_\nu}$ identisch mit dem mittleren Block der $\nu+1$-ten Hypothese $\mu_{\mathbf{C}_{\nu+1}}$. Die Wahrscheinlichkeit, mit der eine spezifische Hypothese aus ihrem Suchraum gewählt wird, spielt für die Überlegungen primär keine Rolle. Die Identität der Suchräume ist dagegen sehr wichtig. Vollständige Induktion führt zu dem Ergebnis, daß die mittleren Blöcke aller n Hypothesen identisch sind.

$$\mu_{\mathbf{C}_1} = \mu_{\mathbf{C}_2} = \cdots = \mu_{\mathbf{C}_n} \tag{2.36}$$

Für die bewegungskompensierte Prädiktion eines Originalblocks zur Zeit k stehen Blöcke aus vorherigen Bildern zur Verfügung. Ein Teil dieser Blöcke waren zu einem früheren Zeitpunkt wiederum Originalblöcke. Analog zu den obigen Überlegungen kommen wir zu dem Schluß, daß der mittlere Originalblock $\mu_{\mathbf{S}}$ identisch zum mittleren Block der ν-ten Hypothese $\mu_{\mathbf{C}_\nu}$ ist.

$$\mu_{\mathbf{C}} = \mu_{\mathbf{S}}u^T \tag{2.37}$$

$u^T = (1, 1, \cdots, 1)$ kennzeichnet den Eins-Vektor der Dimension n. Mit dieser Aussage und **Gleichung 2.33** können wir die Nebenbedingung ableiten.

$$E\left\{\hat{\mathbf{S}}\right\} = E\left\{\mathbf{C}h\right\} = \mu_{\mathbf{C}}h = \mu_{\mathbf{S}}u^Th = E\left\{\mathbf{S}\right\}u^Th$$

Die somit erhaltene Nebenbedingung

$$u^Th = 1 \tag{2.38}$$

besagt, daß sich die Interpolationskoeffizienten zu Eins summieren. Die mittlere Helligkeit der prädizierten Videosequenz entspricht also der mittleren Helligkeit der Originalsequenz.

Durch Lösung des Wiener-Problems mit der Nebenbedingung $u^T h = 1$ erhalten wir nun die bedingt optimalen Interpolationskoeffizienten.

$$h = \sigma^2_{\mathbf{CC}}{}^{-1} \left(\sigma^2_{\mathbf{CS}} - \frac{u^T \sigma^2_{\mathbf{CC}}{}^{-1} \sigma^2_{\mathbf{CS}} - 1}{u^T \sigma^2_{\mathbf{CC}}{}^{-1} u} u \right) \tag{2.39}$$

In **Anhang A** wird das Problem im Detail behandelt. Ein Lagrange-Ansatz führt auch in diesem Fall zum Ziel.

In diesem Abschnitt haben wir bisher die bedingt optimalen Interpolationskoeffizienten bestimmt. Mit dem iterativen Entwurfsalgorithmus gelingt es uns, die lokal optimalen Interpolationskoeffizienten zu determinieren. Dazu haben wir für den Entwurf eine Trainingsmenge festzulegen. In dieser Arbeit wählen wir 18 Videosequenzen von je zehn Sekunten: *Akiyo, Bream, Car Phone, Children, Coastguard, Container Ship, Fun Fair, Hall Monitor, Mobile, News, Salesman, Sean, Silent, Stefan, Table Tennis, Total Destruction, Tunnel* und *Weather*. Für die örtliche Auflösung wählen wir das QCIF-Format und für die zeitliche Auflösung 7.5 Vollbilder pro Sekunde.

Abbildung 2.24 zeigt die Konvergenz der Lagrange-Kosten für die Trainingssequenzen, 16×16 Blöcke, $\lambda = 100$ und $m = 10$. Der Entwurfsalgorithmus ist dabei mit einem gleich verteilten Codebuch und gleich großen Gewichten $\frac{1}{n}$ initialisiert. Das lokale Minimum ist praktisch schon nach dem ersten Schritt erreicht.

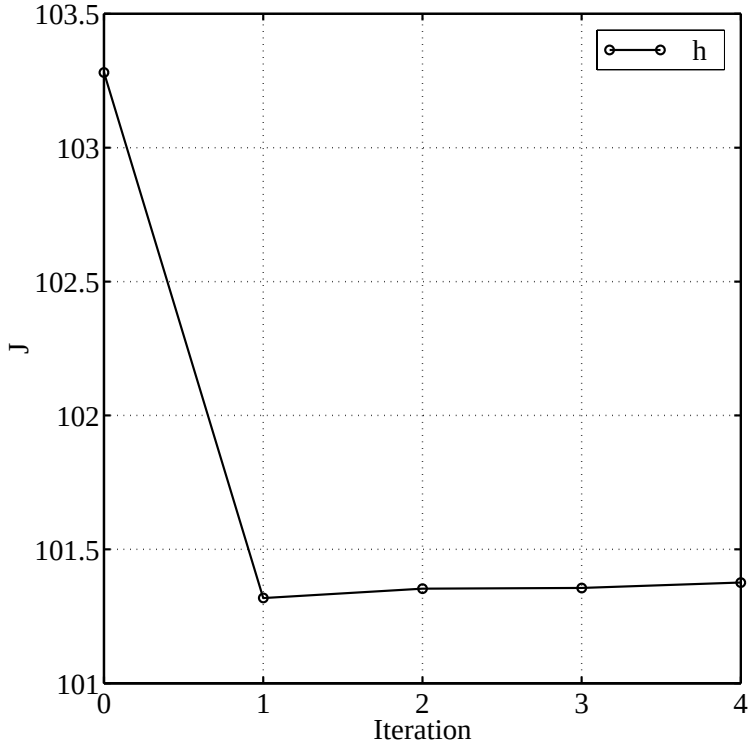


Abbildung 2.24: *Konvergenz der Lagrange-Kosten für die Trainingssequenzen (QCIF, 7.5 fps), 16×16 Blöcke, $\lambda = 100$ und $m = 10$*

Abbildungen 2.25, 2.26 und 2.27 zeigen die Konvergenz der Interpolationskoeffizienten für 2, 3 und 4 Hypothesen. Zur Verifikation der Konvergenz ist ebenfalls ein Entwurf mit einer beliebig anderen Initialisierung der Interpolationskoeffizienten protokolliert. Bereits nach fünf Schritten konvergieren unabhängig von der Initialisierung die Interpola-

tionskoeffizienten zu ihren Fixpunkten. Es ist bemerkenswert, daß diese sehr nahe bei $\frac{1}{n}$ liegen und die ihnen assoziierten Hypothesen somit nahezu gleich gewichten.

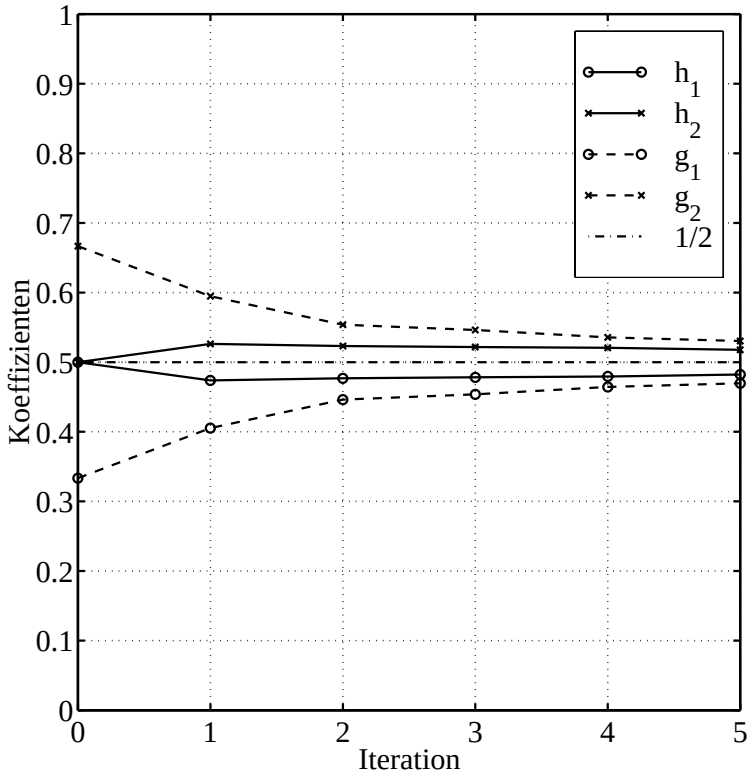


Abbildung 2.25: *Koeffizienten für 2 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $m = 10$

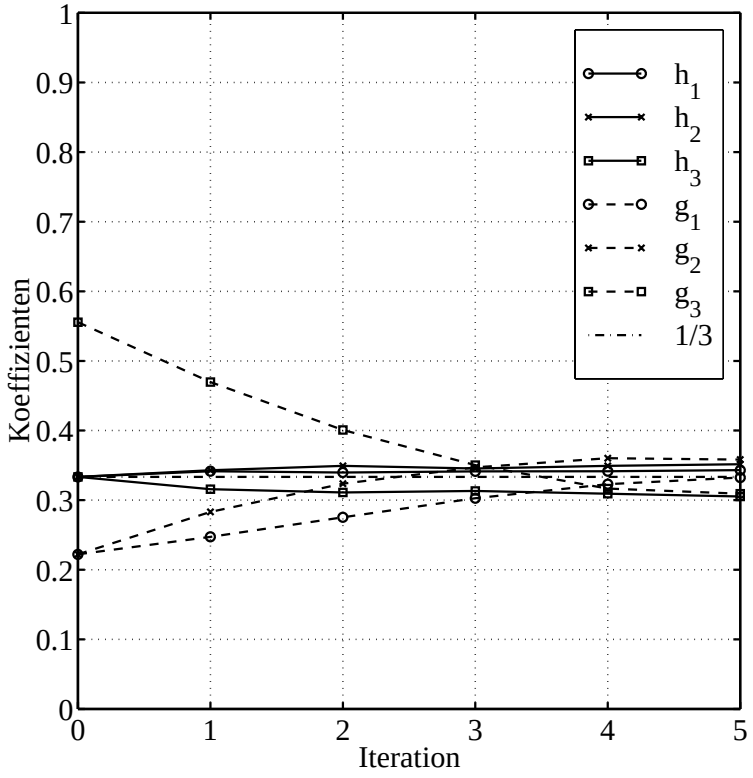


Abbildung 2.26: *Koeffizienten für 3 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $m = 10$

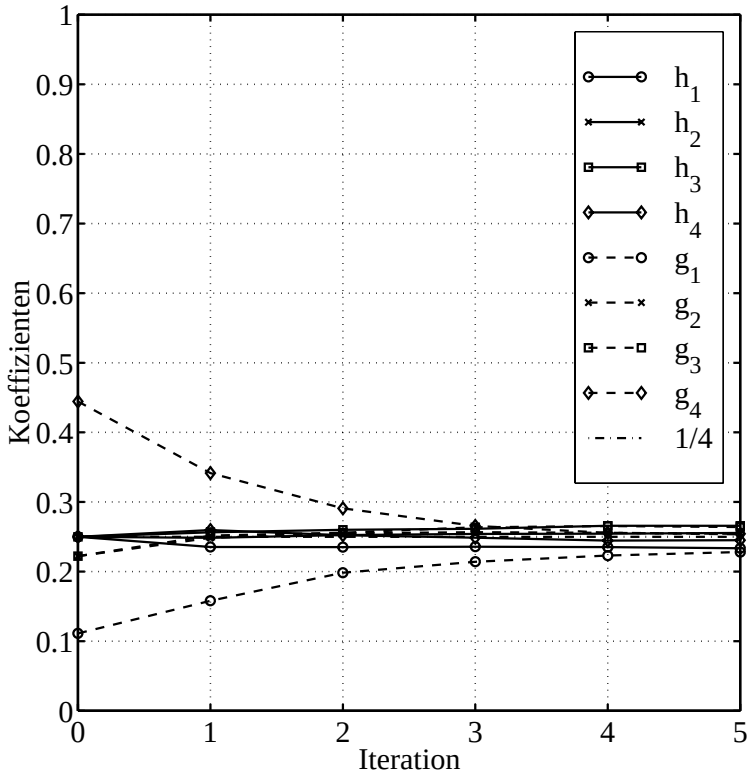


Abbildung 2.27: *Koeffizienten für 4 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $m = 10$

Der vorangehende Entwurf wurde mit Hypothesen aus den jeweils $m = 10$ vorherigen Bildern realisiert. Wir sind nun daran interessiert, ob die zeitliche Suchraumgröße einen Einfluß auf die lokal optimalen Interpolationskoeffizienten hat. Für einen Vergleich wählen wir $m = 1$.

Eine Ad-hoc-Untersuchung zur Konvergenz der Interpolationskoeffizienten führt zu folgendem Satz: Für zunehmende Blockgröße ist die Größe des Suchraumes zu vergrößern. Wir vermuten, daß bei gegebener Suchraumgröße die Zuverlässigkeit der Schätzung der Kovarianz $\sigma^2_{\mathbf{CC}}$ für zunehmende Blockdimension abnimmt. Die Degradation der Zuverlässigkeit, so vermuten wir weiter, verhindert eine Konvergenz der Koeffizienten. Aufgrund dieser Vermutungen wählen wir eine Blockgröße von $s = 4$.

Abbildung 2.28 zeigt die Konvergenz der Lagrange-Kosten für die Trainingssequenzen, $\lambda = 100$, $m = 1$ und 4×4 Blöcke. Wiederum ist der Entwurfsalgorithmus mit einem gleich verteilten Codebuch und mit gleich großen Gewichten $\frac{1}{n}$ initialisiert.

Abbildungen 2.29, 2.30 und 2.31 zeigen die Konvergenz der Interpolationskoeffizienten für 2, 3 und 4 Hypothesen. Zur Verifikation der Konvergenz ist ebenfalls ein Entwurf mit einer beliebig anderen Initialisierung der Interpolationskoeffizienten protokolliert. Für diese Konstellation ($m = 1$, $s = 4$) konvergieren unabhängig von der Initialisierung die Interpolationskoeffizienten zu ihren Fixpunkten.

Es ist bemerkenswert, daß im Vergleich zum letzten Entwurf der Abstand der Fixpunkte zu den Werten $\frac{1}{n}$ größer ist. Wir vermuten, daß eine Hypothese für die Schätzung relevanter ist als die restlichen $n - 1$ Hypothesen. Wir vermuten weiterhin, daß der Suchraum für $m = 1$ zu klein ist, um mehrere gleichgewichtige Hypothesen zu finden.

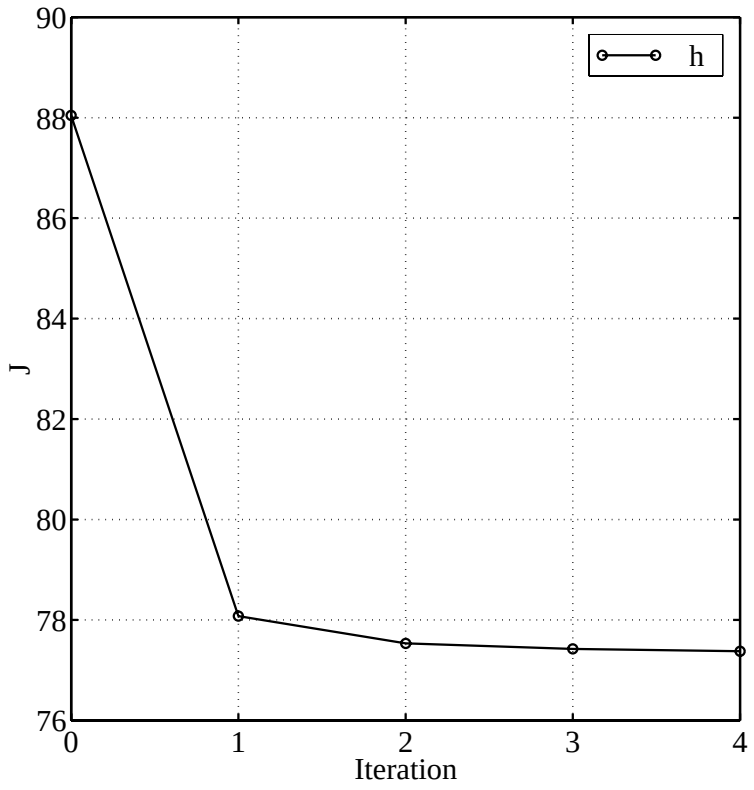


Abbildung 2.28: *Konvergenz der Lagrange-Kosten für die Trainingssequenzen (QCIF, 7.5 fps),* $\lambda = 100$, $m = 1$ *und* 4×4 *Blöcke*

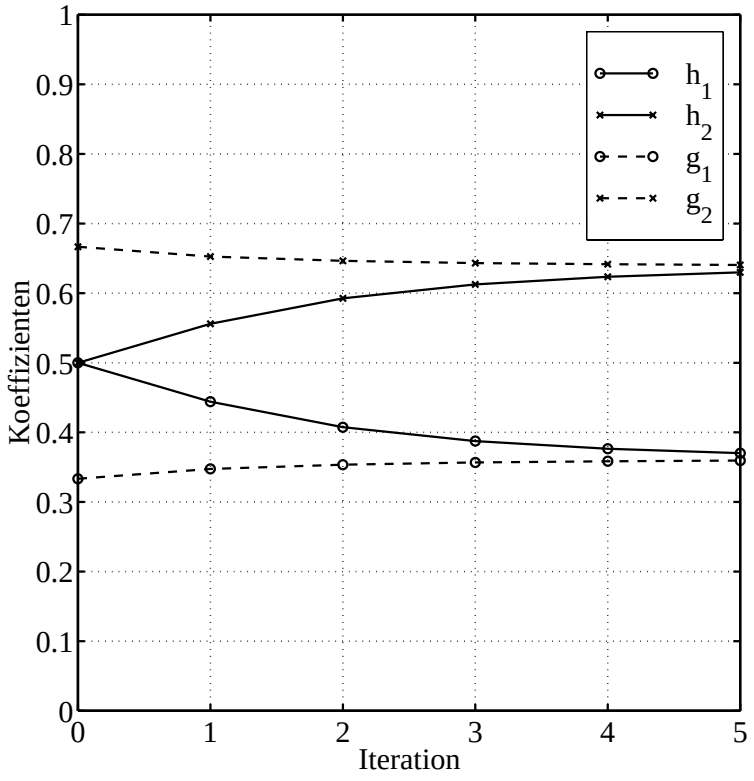


Abbildung 2.29: *Koeffizienten für 2 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* $\lambda = 100$, $m = 1$ *und* 4×4 *Blöcke*

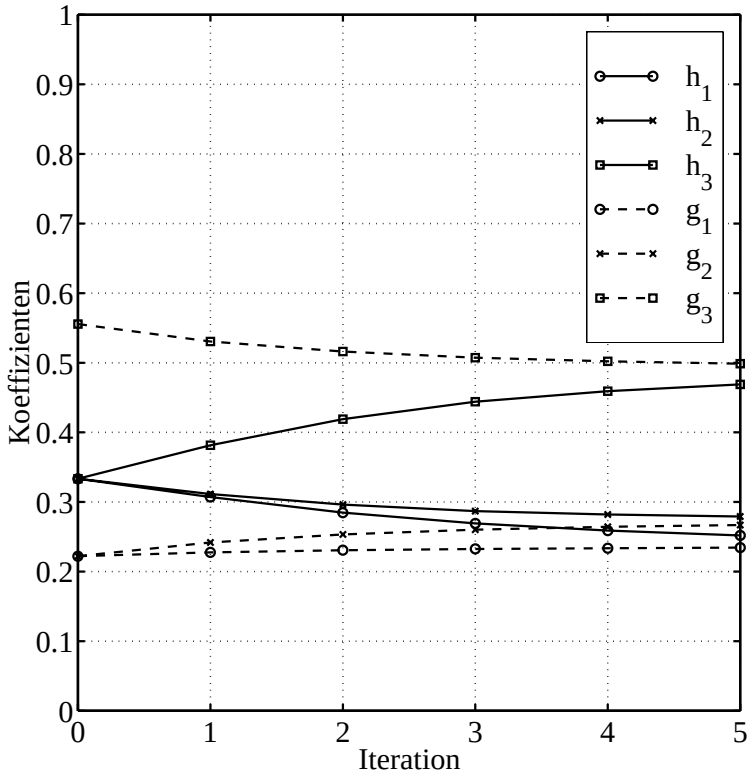


Abbildung 2.30: *Koeffizienten für 3 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* $\lambda = 100$, $m = 1$ *und* 4×4 *Blöcke*

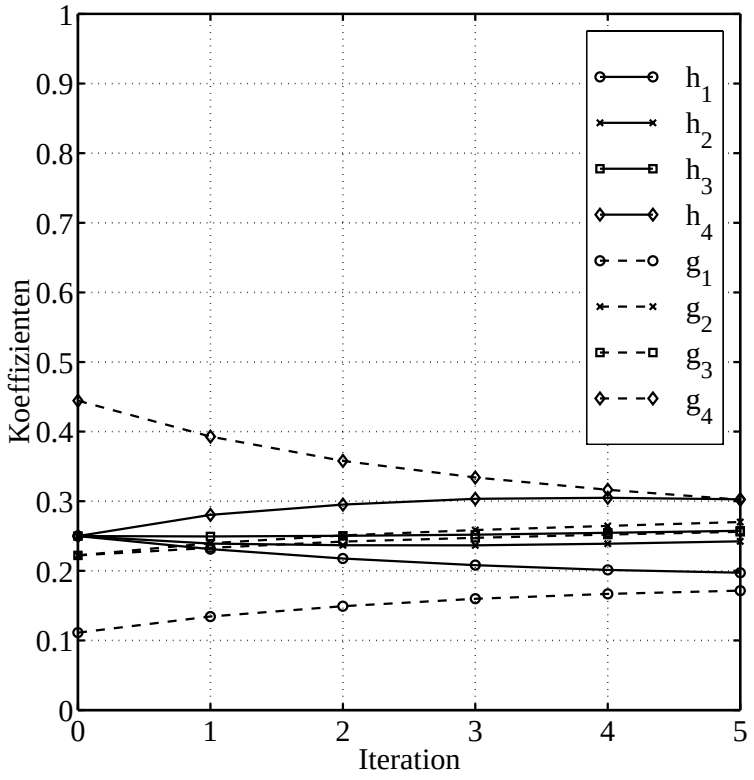


Abbildung 2.31: *Koeffizienten für 4 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* $\lambda = 100$, $m = 1$ *und* 4×4 *Blöcke*

2.3.3 Interpretation des Entwurfs

Der Entwurfsalgorithmus hat zum Ziel, das Lagrange-Funktional nach **Gleichung 2.24** zu minimieren. In diesem Abschnitt interpretieren wir das abstrakte Optimierungskriterium anhand des Multihypothesen-Modells und treffen Aussagen über die Beziehungen zwischen den Hypothesen und über die Beziehungen der Hypothesen zum Originalblock.

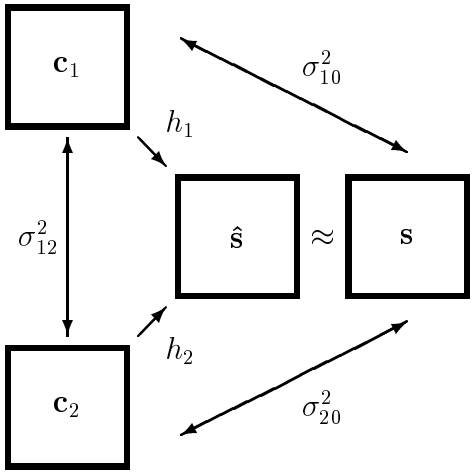


Abbildung 2.32: *Interpretation des Entwurfs*

Wir wählen nach **Abbildung 2.32** ein Modell mit $n = 2$ Hypothesen und nehmen an, daß wir die optimale Hypothesen-Codierung, das optimale Codebuch und die optimalen Interpolationskoeffizienten vorliegen haben.

$$\sigma_{\mathbf{SS}}^2 = \sigma_{00}^2 \quad \sigma_{\mathbf{CS}}^2 = \begin{pmatrix} \sigma_{10}^2 \\ \sigma_{20}^2 \end{pmatrix} \quad \sigma_{\mathbf{CC}}^2 = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 \end{pmatrix} \quad h = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \tag{2.40}$$

Für das Modell geben wir uns explizit die Varianz $\sigma_{\mathbf{SS}}^2$, den Kovarianzvektor $\sigma_{\mathbf{CS}}^2$ und die Kovarianzmatrix $\sigma_{\mathbf{CC}}^2$ vor. Damit erhalten wir für $\lambda = 0$ aus dem allgemeinen Optimierungskriterium (**Gleichung 2.28**) das spezielle Optimierungskriterium.

$$\min \left\{ \sigma_{00}^2 - \sigma_{10}^2 - \sigma_{20}^2 + \frac{1}{4}\sigma_{11}^2 + \frac{1}{2}\sigma_{12}^2 + \frac{1}{4}\sigma_{22}^2 \right\} \tag{2.41}$$

Die Varianzen der beiden Hypothesen σ_{11}^2 und σ_{22}^2 sind ebenso gegeben wie die Varianz der Originalblöcke σ_{00}^2. Wir haben an anderer Stelle festgestellt, daß Originalblöcke und Hypothesen zum Teil identische statistische Eigenschaften aufweisen. Um die positiv definite Summe zu minimieren, ist

1. eine minimale Korrelation zwischen den Hypothesen ($\min \sigma_{12}^2$) und
2. eine maximale Korrelation der Hypothesen zum Originalblock ($\max \sigma_{10}^2, \max \sigma_{20}^2$)

anzustreben. Für statistisch unabhängige Hypothesen ist die Kovarianz zwischen den Hypothesen null ($\sigma_{12}^2 = 0$).

Kapitel 3

Variationen zur Multihypothesen-Prädiktion

3.1 Bidirektionale Multihypothesen-Prädiktion

In **Kapitel 2** haben wir für die Multihypothesen-Prädiktion eine Langzeitprädiktion bzw. eine Kurzzeitprädiktion der Hypothesen zugelassen. Der Suchraum der Hypothesen bestand aus den m vorherigen Bildern $\mathbf{s}[t-1]$, $\mathbf{s}[t-2]$, ..., $\mathbf{s}[t-m]$. Bei der bidirektionalen Prädiktion setzt sich der Suchraum aus dem vorherigen Bild $\mathbf{s}[t-1]$ und dem nächsten Bild $\mathbf{s}[t+1]$ zusammen. In diesem Suchraum sind n Hypothesen zu finden. Dabei werden die möglichen Positionen der Hypothesen innerhalb des Suchraumes nicht eingeschränkt.

MPEG's interpolative Prädiktion schränkt die möglichen Positionen der zwei Hypothesen ein. Die erste Hypothese befindet sich im vorherigen Bild, die zweite im nächsten. S. Wu und A. Gersho schlagen in [6] einen iterativen Algorithmus für die gemeinsame Schätzung der beiden Hypothesen vor. Jedoch erhalten sie die hypothesenbezogene Einschränkung des Suchraumes aufrecht. Der Suchraum-Effekt der bidirektionalen Prädiktion wird vom Multihypothesen-Effekt der bidirektionalen Prädiktion nicht getrennt. In diesem Abschnitt versuchen wir diese beiden Phänomene zu differenzieren.

3.1.1 Optimale Wahl des Blocks für eine Hypothese

Das Auffinden einer Hypothese erfolgt mit dem Algorithmus des Block-Matchings. Es wird diejenige Hypothese gewählt, die im Suchraum, bestehend aus dem vorherigen und dem nächsten Bild, die Lagrange-Kosten minimiert.

$$\|\mathbf{s} - \mathbf{c}_1\|_2^2 + \lambda|\gamma(\mathbf{c}_1)|$$

Natürlich muß bei der Codierung der Verschiebung die zeitliche Komponente ebenfalls explizit berücksichtigt werden. Der Entwurf des Prädiktors beschränkt sich dabei auf die Entropie-Codierung der Verschiebungsparameter.

Abbildungen 3.1 und 3.2 zeigen das Verhalten der bidirektionalen Prädiktion mit einer Hypothese bei Integer-Pel-Genauigkeit. Zum Vergleich ist die kausale bewegungskompensierte Prädiktion aus dem vorherigen Bild ($m = 1$) eingetragen.

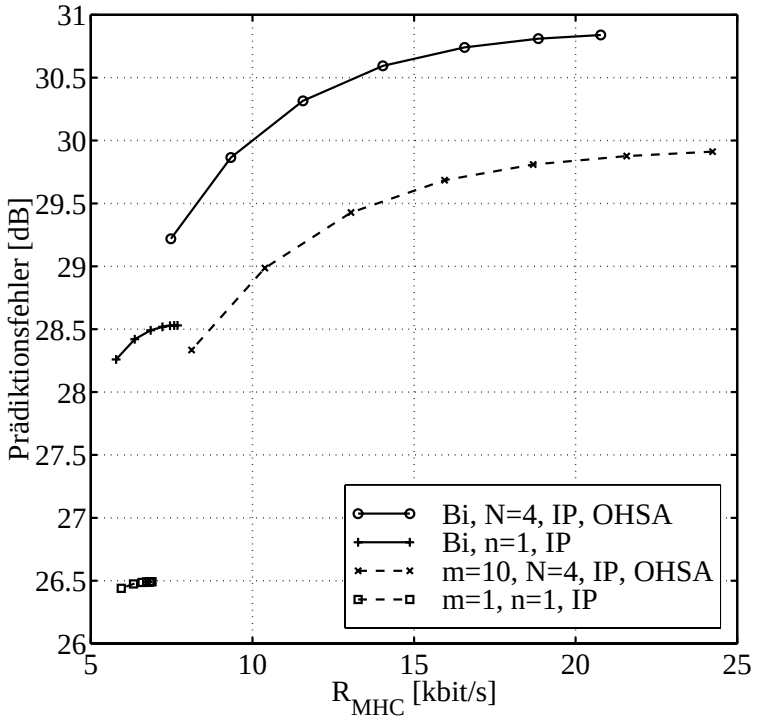


Abbildung 3.1: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s), Integer-Pel-Genauigkeit und* 16×16 *Blöcke zum Verhalten der bidirektionalen Prädiktion.*

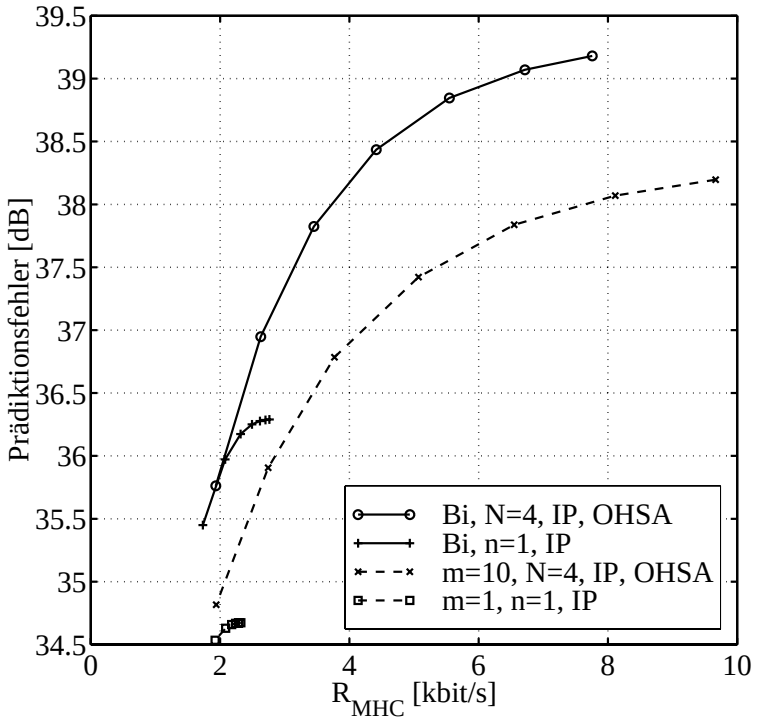


Abbildung 3.2: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s), Integer-Pel-Genauigkeit und* 16×16 *Blöcke zum Verhalten der bidirektionalen Prädiktion.*

3.1.2 Optimale Wahl der Blöcke für mehrere Hypothesen

An dieser Stelle verallgemeinern wir die bekannte bidirektionale Prädiktion und lassen mehr als zwei Hypothesen für jeden zu schätzenden Originalblock zu. Den Suchraum erstrecken wir jedoch ohne Einschränkung über das vorherige und das nächste Bild.

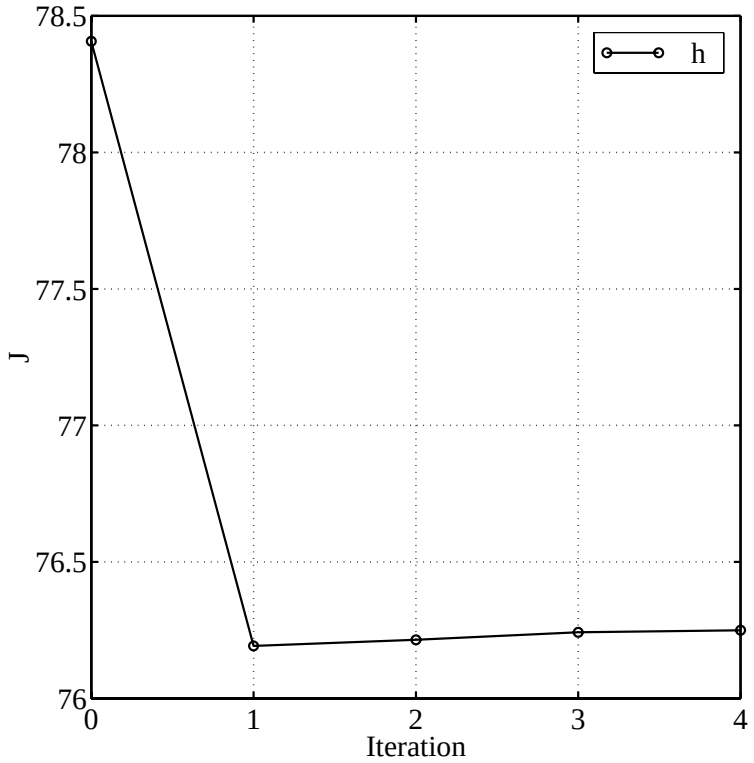


Abbildung 3.3: *Konvergenz der Lagrange-Kosten für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $N = 4$

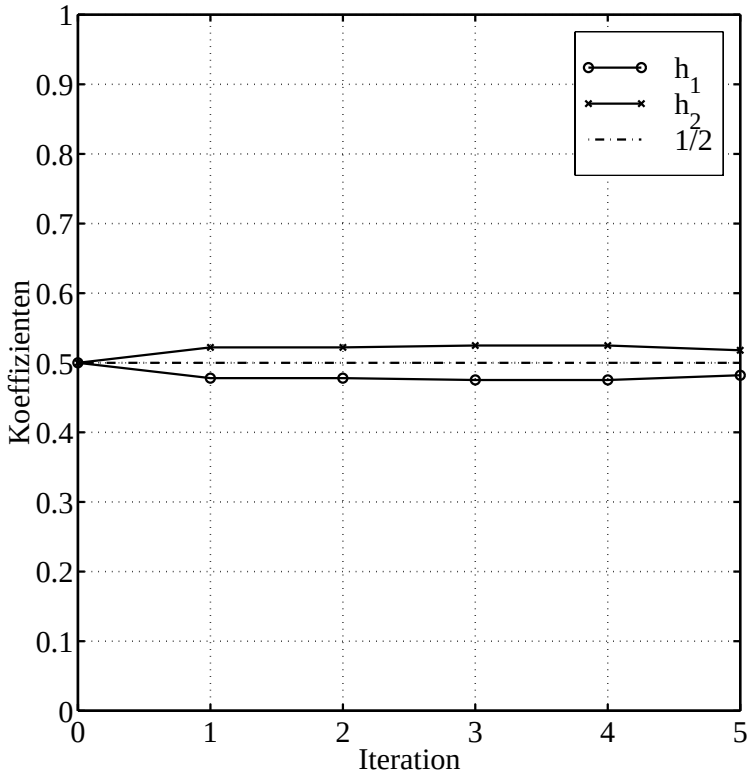


Abbildung 3.4: *Koeffizienten für 2 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $N = 4$

Die optimalen Hypothesen bestimmen wir mit dem Algorithmus zur Auswahl der Hypothesen nach **Abschnitt 2.2.2**. Darüber hinaus wählen wir für die maximale Anzahl der Hypothesen $N = 4$. Die Vorschrift zur Bestimmung der optimalen Anzahl nach **Ab-**

schnitt 2.2.3 determiniert für jeden Block die optimale Hypothesenzahl.

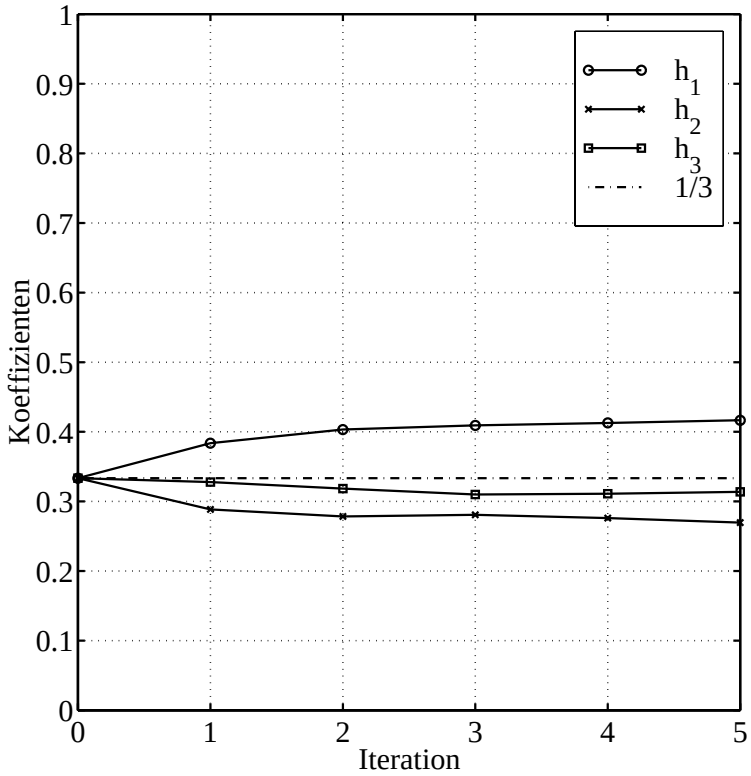


Abbildung 3.5: *Koeffizienten für 3 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $N = 4$

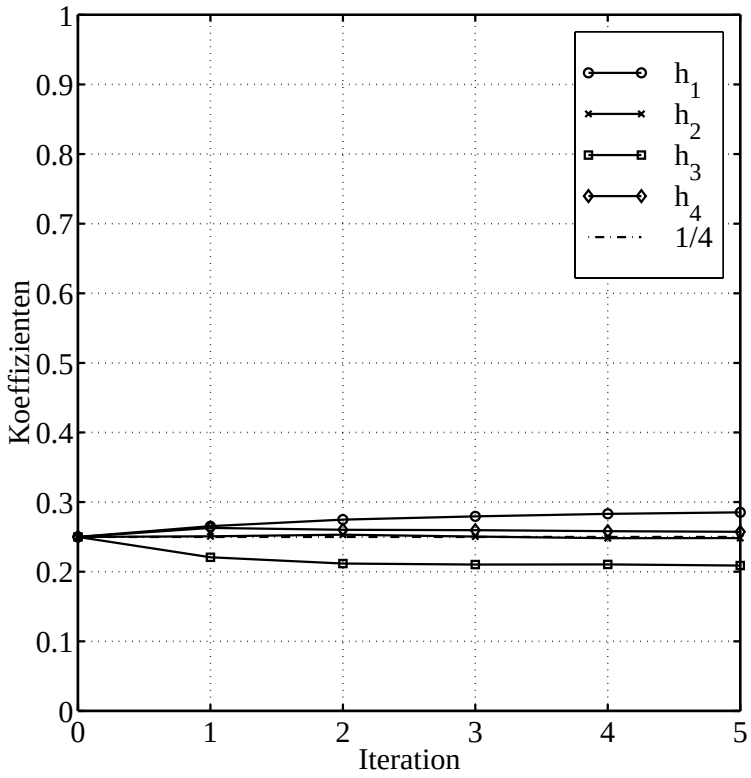


Abbildung 3.6: *Koeffizienten für 4 Hypothesen über den Iterationen für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$ *und* $N = 4$

Abbildung 3.3 zeigt die Konvergenz der Lagrange-Kosten für die 16×16 Blöcke der

Trainingssequenzen bei $\lambda = 100$. Der Entwurf des Multihypothesen-Prädiktors für den vorliegenden bidirektionalen Suchraum liefert die optimale Entropie-Codierung und die optimalen Interpolationskoeffizienten. Die Quasi-Fixpunkt-Koeffizienten für 2, 3 und 4 Hypothesen kann man den **Abbildungen 3.4, 3.5 und 3.6** entnehmen.

Wir beziehen uns jetzt noch einmal auf die **Abbildungen 3.1 und 3.2**. Diese geben zusätzlich Auskunft über das Rate-Distortion-Verhalten der bidirektionalen Prädiktion mit bis zu $N = 4$ Hypothesen. Es ist bemerkenswert, daß der Multihypothesen-Gewinn im Vergleich zur kausalen Prädiktion annähernd in der gleichen Größenordnung auftritt. Der Multihypothesen-Gewinn, um dies noch einmal deutlich zu unterstreichen, resultiert ausschließlich aus der Erhöhung der maximal zulässigen Hypothesenzahl bei konstanter Suchraumgröße. Bidirektionale Prädiktion erzielt a priori einen größeren Suchraum-Gewinn durch ihren antikausalen Charakter. Der Multihypothesen-Effekt addiert sich aber weiterhin.

3.2 Langzeitprädiktion mit überlappenden Blöcken

Kurzzeitprädiktion mit überlappenden Blöcken wird in [5] ausführlich untersucht. Wir behandeln die überlappende Prädiktion in diesem Abschnitt um die Analogien zu der in **Kapitel 2** vorgestellten Multihypothesen-Prädiktion darzulegen.

3.2.1 Modell für die Blocküberlappung

Für die Blocküberlappung wählen wir ein elementares Modell. Bei der überlappenden Prädiktion haben die direkten Nachbarn des aktuellen Blocks einen Einfluß auf die zu schätzenden Bewegungsparameter. Wir ordnen deshalb die direkten Nachbarn um den aktuellen Block an und assoziieren jeden Block mit der Translation $\mathbf{\Delta}_\nu$. **Abbildung 3.7** skizziert diese Zuordnung.

$\mathbf{\Delta}_2$	$\mathbf{\Delta}_3$	$\mathbf{\Delta}_4$
$\mathbf{\Delta}_5$	$\mathbf{\Delta}_1$	$\mathbf{\Delta}_6$
$\mathbf{\Delta}_7$	$\mathbf{\Delta}_8$	$\mathbf{\Delta}_9$

Abbildung 3.7: *Direkte Nachbarn des aktuellen Blocks*

Jeder Block bei der überlappenden Prädiktion wird nur mit einer Translation $\mathbf{\Delta}_\nu$ codiert. Bei der Multihypothesen-Prädiktion nach **Kapitel 2** werden dagegen n Translationen codiert, um n Hypothesen aufzufinden. Die überlappende Prädiktion nutzt aber die Translationen der örtlichen Nachbarn, um ihre Hypothesen zu adressieren.

Die Adressierung der Hypothesen mittels der örtlichen Nachbarn erläutern wir im Detail. Dazu betrachten wir jedes Bildelement; um unsere Notation beizubehalten, sprechen wir von 1×1 Blöcken im Gegensatz zu den Makroblöcken (16×16 Blöcke), denen eine Translation zugeordnet ist. Jedes zu schätzende Bildelement erhält man durch eine Linearkombination der $n = 9$ Hypothesen (8 Translationen der Nachbarn und diejenige des aktuellen Blocks).

$$\hat{\mathbf{s}} = \sum_{\nu=1}^{n} \mathbf{c}_\nu h_\nu \tag{3.1}$$

$[x, y, t]$ sei eine diskrete Position eines Bildelements im aktuellen Makroblock. Die 1×1 Hypothesen erhalten wir durch Evaluierung aller n Translationen relativ zu dieser Position.

$$\mathbf{c}_\nu = \mathbf{s}[x - \mathbf{\Delta}_{x_\nu}, y - \mathbf{\Delta}_{y_\nu}, t - \mathbf{\Delta}_{t_\nu}] \tag{3.2}$$

Auf diese Weise haben die Translationen der Nachbar-Makroblöcke Einfluß auf die Prädiktion des aktuellen Makroblocks.

3.2.2 Überlappende Bewegungsschätzung

Die örtlichen Abhängigkeiten zwischen den Makroblöcken zwingt uns nun, ein Translationsfeld eines Bildes zu schätzen. Wir benötigen einen Algorithmus, der die Translation des aktuellen Makroblocks und die der Nachbar-Makroblöcke gemeinsam optimiert. Dazu wird in [5] ein iterativer Algorithmus vorgeschlagen.

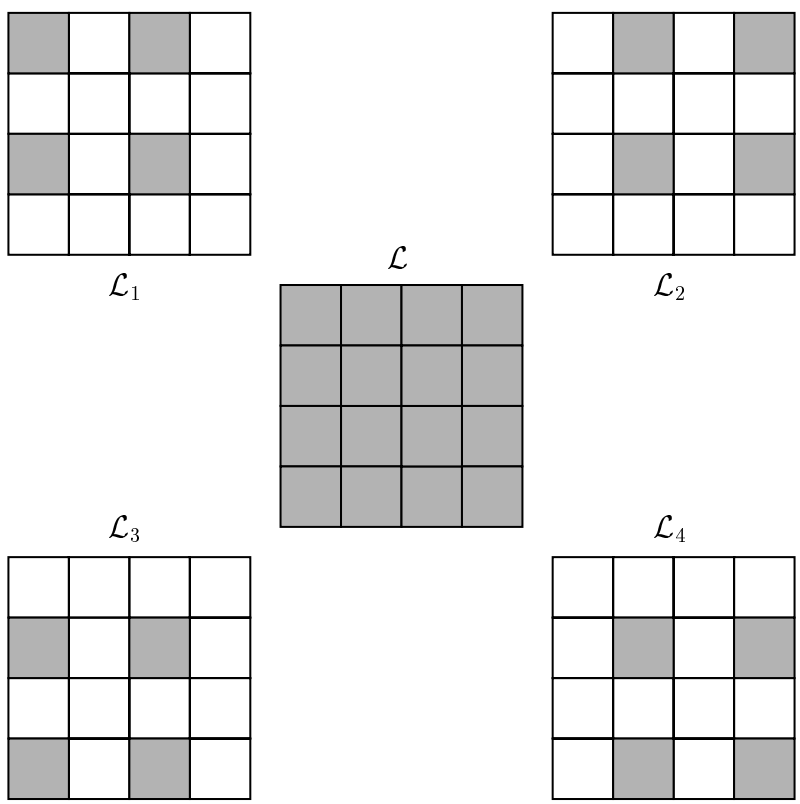


Abbildung 3.8: *Aufteilung der Blöcke in 4 Teilgitter bei der Prädiktion mit überlappenden Blöcken.*

Die Makroblock-Struktur eines Bildes weist eine Gitterstruktur auf. In unserem Modell berücksichtigen wir nur die direkten Nachbarn. Dies erlaubt uns die Aufteilung in vier Teilgitter, deren Makroblöcke keine direkten Nachbarn besitzen. **Abbildung 3.8** veranschaulicht diese Zerlegung. Es ist uns nun möglich, eine bedingte iterative Optimierung des Translationsfeldes mit Hilfe der *Iterated Conditional Modes* nach Besag [12] zu realisieren. Sind drei Teilgitter gegeben, so ist das vierte Teilgitter durch Rate-Distortion-optimales Blockmatching zu optimieren.

3.2.3 Optimale Gewichtskoeffizienten

Die optimalen Gewichtskoeffizienten h für jeden 1×1 Block haben die gleichen Bedingungen wie bei der Multihypothesen-Prädiktion nach **Kapitel 2** zu erfüllen. Die optimalen Interpolationskoeffizienten summieren sich für jedes Bildelement zu eins. Dies ist hier wünschenswert, da prädizierte Regionen ohne Bewegung dem Original entsprechen sollen. Bei der überlappenden Prädiktion haben wir die modifizierte Wiener-Lösung

$$h = {\sigma^2_{\mathbf{CC}}}^{-1} \left(\sigma^2_{\mathbf{CS}} - \frac{u^T {\sigma^2_{\mathbf{CC}}}^{-1} \sigma^2_{\mathbf{CS}} - 1}{u^T {\sigma^2_{\mathbf{CC}}}^{-1} u} u \right) \tag{3.3}$$

aus **Anhang A** für jedes Bildelement eines Makroblocks zu bestimmen. Die so erhaltenen Gewichtskoeffizienten lassen sich geometrisch sinnvoll zu einem „Fenster“ anordnen.

3.2.4 Entwurf des Prädiktors

Analog zu dem iterativen Entwurfsalgorithmus sind die Abhängigkeiten zwischen der überlappenden Prädiktion und dem „Fenster“ bei der Optimierung zu berücksichtigen. Diese Variante des *Generalized Lloyd Algorithm* benötigt neben einem initialen Codebuch für die Translationsparameter ein initiales „Fenster“. **Abbildung 3.10** zeigt ein solches „Fenster“. Werden 16×16 Blöcke mit obigem Modell prädiziert, so benötigen wir ein 48×48 „Fenster“. Das dargestellte initiale „Fenster“ läßt keine Blocküberlappung zu. Dies erkennt man daran, daß die Gewichte für die Nachbar-Makroblöcke null sind.

Abbildung 3.9 zeigt die Konvergenz der Lagrange-Kosten für die Trainingssequenzen, 16×16 Blöcke und $\lambda = 100$. Die Pseudo-1-Hypothesen befinden sich jeweils in den $m = 10$ vorherigen Bildern. Die Konvergenz dieser *Lloyd*-Variante ist ebenfalls bemerkenswert.

Abbildung 3.11 zeigt das „Fenster“ nach der ersten Iteration. Es ist deutlich zu erkennen, daß die Nachbar-Makroblöcke einen Einfluß auf die Prädiktion des aktuellen Makroblocks haben. Vor allem die Ränder der Blöcke unterliegen einem starken Einfluß der Nachbarn. Dies ist daran zu erkennen, daß die Interpolationskoeffizienten an den Rändern des aktuellen Blocks kleiner sind als die im Zentrum. Mit zunehmender Iteration prägt sich dieses Phänomen stärker aus.

Abbildung 3.12 zeigt das „Fenster“ im fünften Iterationsschritt. Das Verhalten der Lagrange-Kosten über den Iterationen weist bereits auf eine schnelle Konvergenz des Entwurfsalgorithmus hin. Die „Fenster“ im Iterationsschritt eins und fünf unterscheiden sich deshalb nur noch geringfügig.

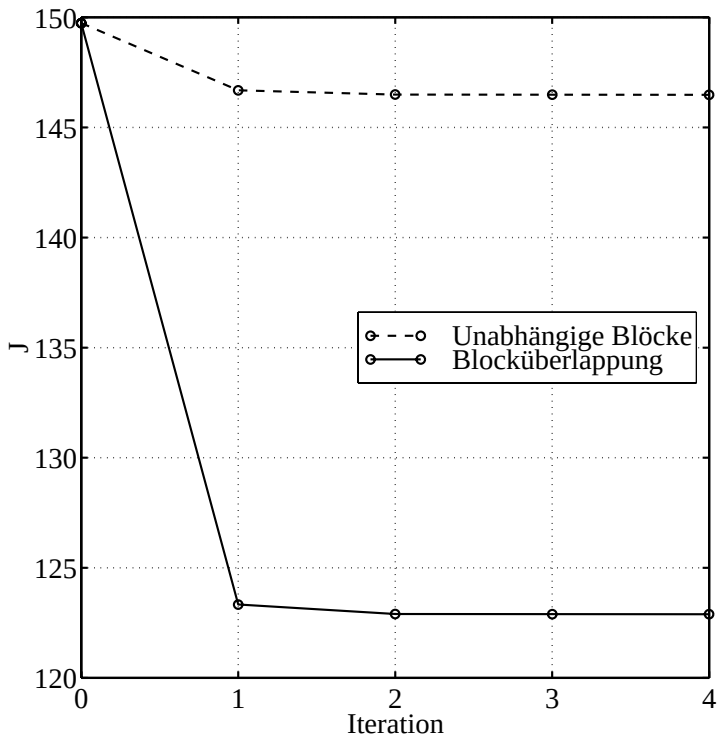


Abbildung 3.9: *Konvergenz der Lagrange-Kosten für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$, $m = 10$ *und* $n = 1$

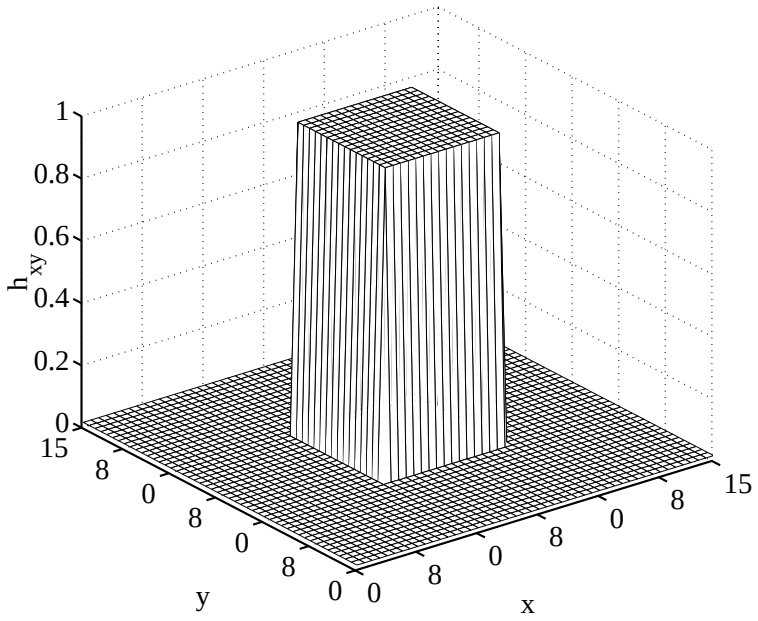


Abbildung 3.10: *Initiale Interpolationskoeffizienten für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$, $m = 10$ *und* $n = 1$

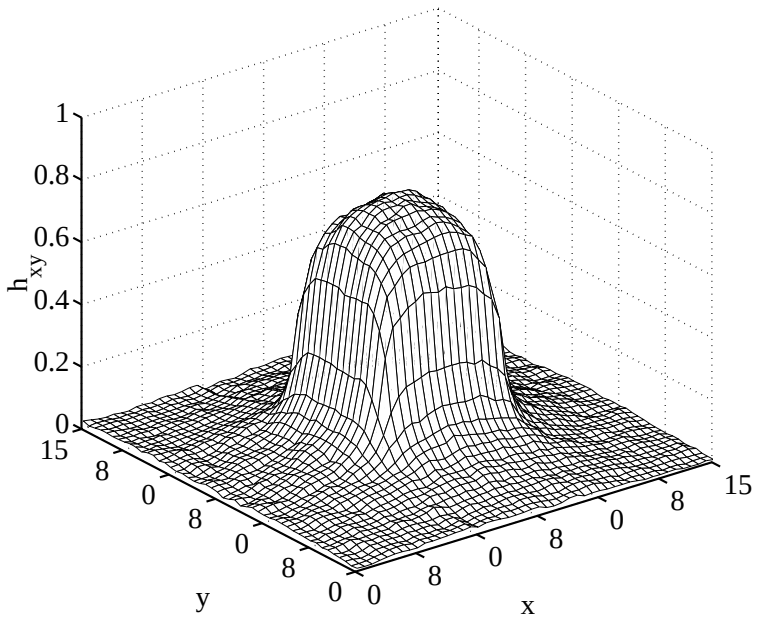


Abbildung 3.11: *Interpolationskoeffizienten bei der 1. Iteration für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$*,* $m = 10$ *und* $n = 1$

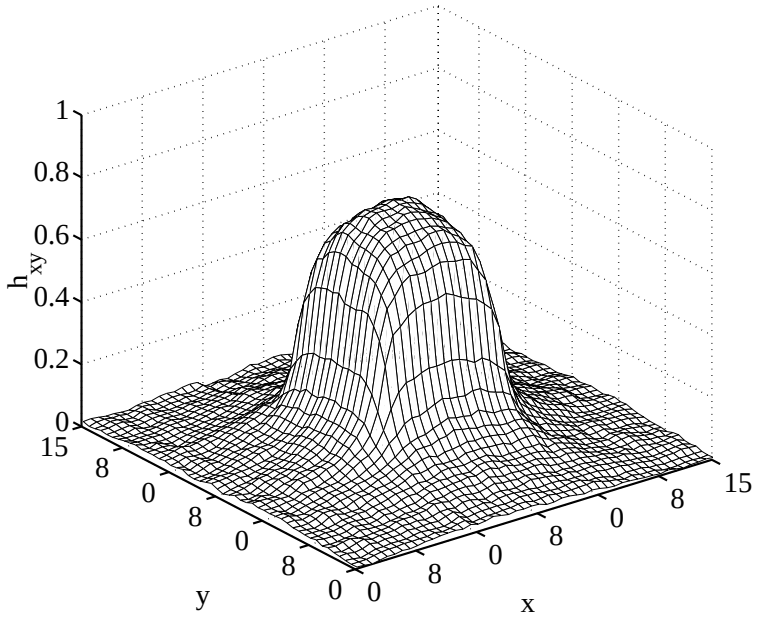


Abbildung 3.12: *Interpolationskoeffizienten bei der 5. Iteration für die Trainingssequenzen (QCIF, 7.5 fps),* 16×16 *Blöcke,* $\lambda = 100$*,* $m = 10$ *und* $n = 1$

Abbildungen 3.13 und 3.14 präsentieren das Rate-Distortion-Verhalten der Testsequenzen für 16×16 Blöcke bei Integer-Pel-Genauigkeit.

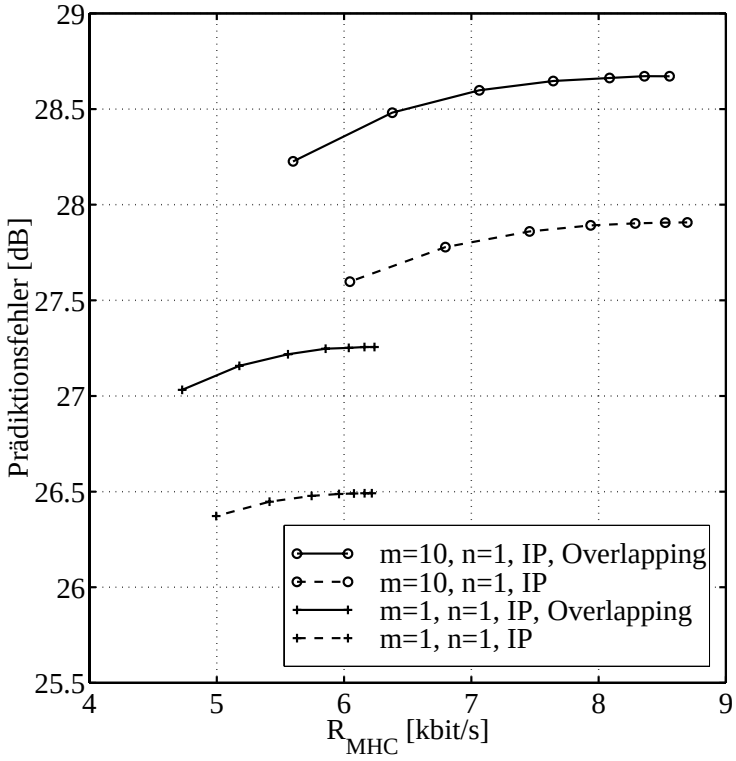


Abbildung 3.13: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke,* $n = 1$ *und Integer-Pel-Genauigkeit zum Verhalten der Blocküberlappung*

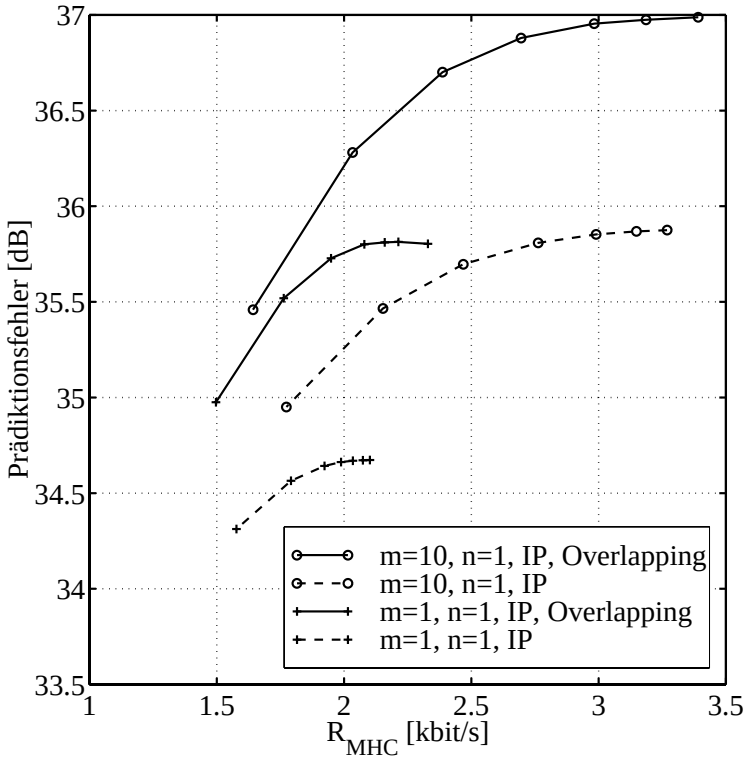


Abbildung 3.14: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke,* $n = 1$ *und Integer-Pel-Genauigkeit zum Verhalten der Blocküberlappung*

Wiederum ist festzustellen, daß der Multihypothesen-Effekt unabhängig von der Such-

raumgröße auftritt. Da wir für die Simulation ein sehr einfaches Multihypothesen-Modell verwendet haben, sind die Multihypothesen-Gewinne eher moderat. Der Suchraum-Gewinn von $m = 1$ nach $m = 10$ ist trotzdem zu beobachten.

3.3 Interblock-Prädiktion der Translation

Im letzten Abschnitt wurden die örtlichen Abhängigkeiten der Makroblöcke bei den Bildelementen berücksichtigt. Nun möchten wir die Abhängigkeiten über die Translationen der Makroblöcke einfließen lassen. Für diese Erweiterung wählen wir das Multihypothesen-Modell nach **Kapitel 2**. Die Interblock-Prädiktion der Translation gibt uns die Möglichkeit, die Bewegungsinformationen der örtlichen Nachbarblöcke für die Codierung der Translation des aktuellen Blocks heranzuziehen.

3.3.1 Translationsmodell

Zunächst entwickeln wir ein Translationsmodell. Dazu betrachten wir ein Video-Objekt, das eine Translation in örtlicher und zeitlicher Richtung besitzt. Diese Translation wird von allen Blöcken aus denen das Video-Objekt besteht nachvollzogen. Somit ist es möglich, die Translationsinformationen des aktuellen Blocks aus denen der direkten Nachbarn zu schätzen. Mit diesen heuristischen Überlegungen formulieren wir ein Modell für eine lokal gleichförmige Translation der Blöcke. Dazu definieren wir die geschätzte Translationsgeschwindigkeit des aktuellen Blocks $\hat{\mathbf{V}}$ als Quotient aus örtlicher und zeitlicher Verschiebung.

$$\hat{\mathbf{V}}_x = \frac{\mathbf{\Delta}_x}{\mathbf{\Delta}_t} \tag{3.4}$$

$$\hat{\mathbf{V}}_y = \frac{\mathbf{\Delta}_y}{\mathbf{\Delta}_t} \tag{3.5}$$

Die tatsächliche absolute Translation des aktuellen Blocks $\mathbf{\Delta}$ erhalten wir aus der differentiellen Translation des aktuellen Blocks $\mathbf{D}$ mit Hilfe der geschätzte Translationsgeschwindigkeit des aktuellen Blocks $\hat{\mathbf{V}}$.

$$\mathbf{\Delta}_x = \hat{\mathbf{V}}_x \mathbf{D}_t + \mathbf{D}_x \tag{3.6}$$

$$\mathbf{\Delta}_y = \hat{\mathbf{V}}_y \mathbf{D}_t + \mathbf{D}_y \tag{3.7}$$

$$\mathbf{\Delta}_t = \mathbf{D}_t \tag{3.8}$$

3.3.2 Translationsgeschwindigkeit des aktuellen Blocks

Nach unseren Überlegungen erhalten wir einen Schätzwert für die Translationsgeschwindigkeit des aktuellen Blocks $\hat{\mathbf{V}}$ aus den Translationsgeschwindigkeiten der örtlichen Nachbarn $\mathbf{V}^{\rho}$. Je nach Position des aktuellen Blocks im Bild unterscheiden wir nach **Abbildung 3.15** vier verschiedene Fälle.

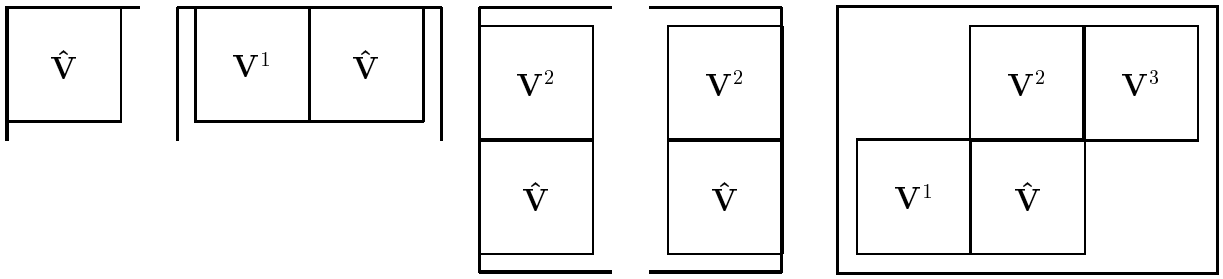


Abbildung 3.15: *Interblock-Prädiktion der Translationsgeschwindigkeit*

1. Der aktuelle Block ist in der linken oberen Ecke des Bildes. Für diesen Fall wird der Schätzwert zu null gesetzt.

$$\begin{aligned} \hat{\mathbf{V}}_x &= 0 \\ \hat{\mathbf{V}}_y &= 0 \end{aligned}$$

2. Der aktuelle Block befindet sich in der obersten Zeile des Bildes. Der Schätzwert entspricht hier der Translationsgeschwindigkeit des linken Nachbarn.

$$\begin{aligned} \hat{\mathbf{V}}_x &= \mathbf{V}_x^1 \\ \hat{\mathbf{V}}_y &= \mathbf{V}_y^1 \end{aligned}$$

3. Der aktuelle Block befindet sich am linken oder rechten Rand des Bildes. Der Schätzwert nimmt hier die Translationsgeschwindigkeit des oberen Nachbarn an.

$$\begin{aligned} \hat{\mathbf{V}}_x &= \mathbf{V}_x^2 \\ \hat{\mathbf{V}}_y &= \mathbf{V}_y^2 \end{aligned}$$

4. Der aktuelle Block befindet sich im Inneren des Bildes. Der Schätzwert berechnet sich komponentenweise aus dem *Median* [18] der Translationsgeschwindigkeiten der linken und oberen Nachbarn.

$$\begin{aligned} \hat{\mathbf{V}}_x &= \mathrm{Med}(\mathbf{V}_x^1, \mathbf{V}_x^2, \mathbf{V}_x^3) \\ \hat{\mathbf{V}}_y &= \mathrm{Med}(\mathbf{V}_y^1, \mathbf{V}_y^2, \mathbf{V}_y^3) \end{aligned}$$

3.3.3 Translationsgeschwindigkeit der Nachbarblöcke

Die für die Schätzung der Translationsgeschwindigkeit des aktuellen Blocks benötigten Translationsgeschwindigkeiten der Nachbarblöcke $\mathbf{V}^\rho$ determinieren wir aus den Translationsgeschwindigkeiten der n Hypothesen durch arithmetische Mittelung.

$$\begin{aligned} \mathbf{V}_x^\rho &= \frac{1}{n} \sum_{\nu=1}^{n} \frac{\mathbf{\Delta}_{x_\nu}^\rho}{\mathbf{\Delta}_{t_\nu}^\rho} \\ \mathbf{V}_y^\rho &= \frac{1}{n} \sum_{\nu=1}^{n} \frac{\mathbf{\Delta}_{y_\nu}^\rho}{\mathbf{\Delta}_{t_\nu}^\rho} \end{aligned}$$

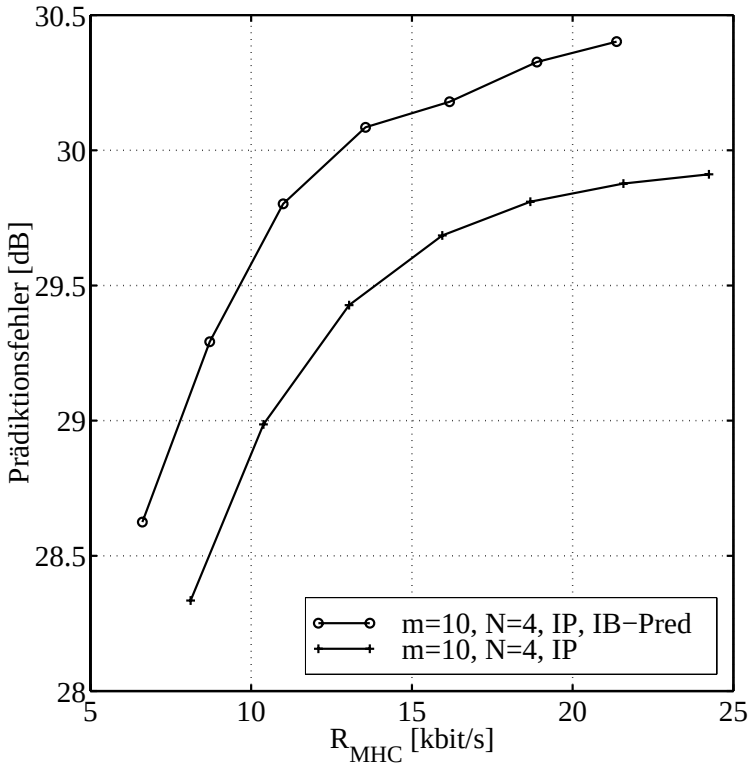


Abbildung 3.16: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke, Integer-Pel-Genauigkeit,* $m = 10$ *und* $N = 4$ *zum Verhalten der Interblock-Prädiktion.*

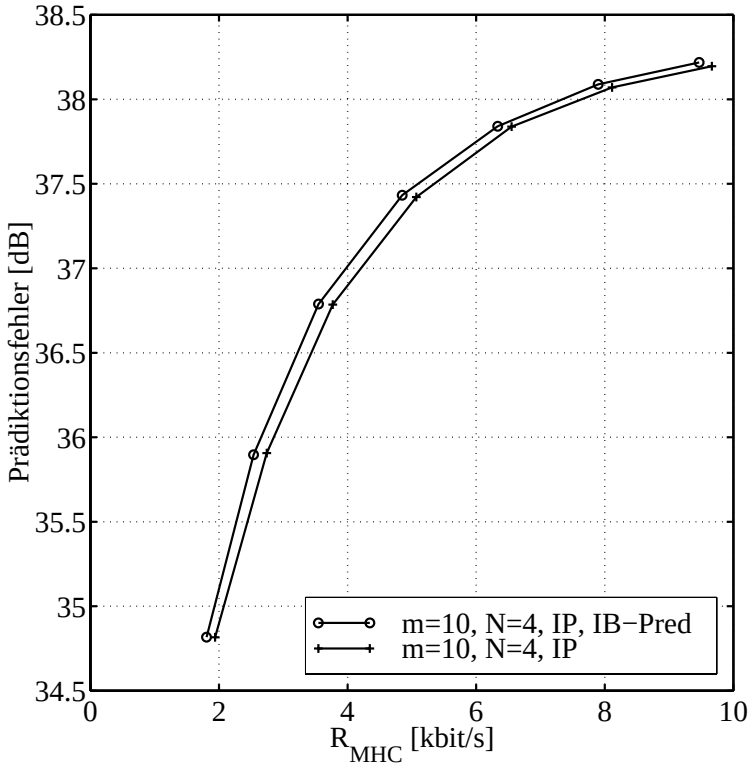


Abbildung 3.17: *Prädiktionsfehler über der Rate des Multihypothesencodes für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s),* 16×16 *Blöcke, Integer-Pel-Genauigkeit,* $m = 10$ *und* $N = 4$ *zum Verhalten der Interblock-Prädiktion.*

Abbildungen 3.16 und 3.17 stellen das Rate-Distortion-Verhalten der Interblock-Prädiktion der Translation dar. Da die Video-Objekte in der Testsequenz „Mother and Daughter“ sich sehr wenig bewegen, erzielen wir kaum einen Gewinn bei der Interblock-

Prädiktion der Translation. Die Testsequenz „Foreman“ besitzt hingegen eine hohe Bewegungsdynamik. Dementsprechend läßt sich durch die hier beschriebene Methode das Rate-Distortion-Verhalten verbessern. Wir bemerken weiterhin eine Unregelmäßigkeit des Rate-Distortion-Verhaltens bei der Testsequenz „Foreman“. Wir vermuten, daß die Schätzung der Translationsgeschwindigkeit aus den Translationsgeschwindigkeiten der Hypothesen nicht sehr zuverlässig ist. Eine weitere Untersuchung ist notwendig.

Kapitel 4

Schlußbemerkungen

Wir haben ein ratendynamisches Verfahren zur bewegungskompensierten Prädiktion untersucht, das jeden Block mit n Translationen codiert. Dieses Verfahren stellt eine Verallgemeinerung der bisher bekannten bewegungskompensierten Prädiktion dar. Im speziellen integriert diese Generalisierung sowohl die Methode der bidirektionalen Prädiktion als auch die der blocküberlappenden Prädiktion.

In diesem Rahmen haben wir n Hypothesen einer bewegungskompensierten Prädiktion linear überlagert. Dabei haben wir für jede der n Hypothesen den gleichen Suchraum zugelassen. Der Suchraum wird dabei durch die maximal zulässige Translation festgelegt.

In diesem Suchraum haben wir mit Hilfe des OHSA n statistisch abhängige Hypothesen gefunden, die wir optimal linear überlagern konnten. Dabei haben wir beobachtet, daß die n-Hypothese andere Eigenschaften als die $n+1$-Hypothese aufweist. Ein Transfer dieser Hypothesen ist also nicht zulässig.

Wir haben weiter festgestellt, daß für zunehmende Hypothesenzahl die Korrelation zwischen den Hypothesen und die Korrelation der einzelnen Hypothesen zum Originalblock *abnimmt*. Dabei erhöht sich absolut die Effizienz der Multihypothesen-Prädiktion. Jedoch nimmt für exponentiell zunehmende Hypothesenzahl die Effizienz nur unterproportional zu.

Der Multihypothesen-Effekt tritt unabhängig von der Suchraumgröße bzw. von der Blockgröße auf. Beide Parameter haben jedoch Einfluß auf die Prädiktionseffizienz.

Wir interpretieren die bewegungskompensierte Multihypothesen-Pädiktion als Vektorquantisierung und lösen das Problem des Prädiktorentwurfs mit Algorithmen zum Entwurf eines Vektorquantisierers. Dabei konnten wir beobachtet, daß die Größe des Suchraums Einfluß auf die optimalen Überlagerungskoeffizienten hat.

- Für sehr kleine Suchräume wird unter den n Hypothesen eine dominante Hypothese observiert. Einer dominanten Hypothese ist ein Überlagerungskoeffizient größer als $\frac{1}{n}$ zugeordnet.
- Für wachsende Suchraumgröße finden sich asymptotisch n gleichwertige Hypothesen. Dies ist daran zu erkennen, daß die optimalen Überlagerungskoeffizienten asymptotisch den Wert $\frac{1}{n}$ annehmen.

Die Größe des Suchraums hat auch Einfluß auf die Konvergenzgeschwindigkeit des Entwurfsalgorithmus. Kleine Blockgrößen stellen bei einem kleinen Suchraum die Konvergenz sicher.

Abschließend möchten wir noch auf die Möglichkeiten beim Einsatz der Multihypothesen-Prädiktion für Algorithmen zur Videocodierung hinweisen. Für hybride Schemata ist das Verhalten der Restfehlerbildcodierung bei Einsatz der Multihypothesen-Prädiktion zu untersuchen. Die Ergebnisse zur Multihypothesen-Prädiktion von Originalblöcken mit unterschiedlicher Größe deuten auf ein Schema zur Videocodierung mit variabler Blockgröße, das ohne Restfehlerbildcodierung realisierbar ist.

Anhang A

Wiener-Lösung mit Nebenbedingung

Die optimalen Interpolationskoeffizienten für jeden Schritt im iterativen Entwurfsalgorithmus ergeben sich aus der Lösung des *Wiener-Problems*

$$E\left\{\mathbf{S}\right\} = E\left\{\hat{\mathbf{S}}\right\} \tag{A.1}$$

$$\min_h E\left\{\left\|\mathbf{S} - \hat{\mathbf{C}}\right\|_2^2\right\} \tag{A.2}$$

unter der Nebenbedingung, daß sich die Interpolationskoeffizienten zu eins summieren.

$$u^T h = 1 \tag{A.3}$$

u kennzeichnet dabei den Eins-Vektor der jeweiligen Dimension.

$$u = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \tag{A.4}$$

Die erste Bedingung des Wiener-Problems nach **Gleichung A.1** fordert die Identität zwischen dem mittleren Block der Originalblöcke und dem mittleren Block der geschätzten Blöcke. Dabei ist der Erwartungswert für jedes Pel zu bestimmen.

$$\begin{aligned} E\left\{\mathbf{S}\right\} &= E\left\{\hat{\mathbf{S}}\right\} \\ &= E\left\{\mathbf{C}h\right\} \\ \mu_{\mathbf{S}} &= \mu_{\mathbf{C}} h \end{aligned} \tag{A.5}$$

Durch Auflösung der quadrierten l_2-Norm [19] in **Gleichung A.2** des Wiener-Problems ergeben sich die entsprechenden Korrelationsmatrizen.

$$\min_h E\left\{(\mathbf{S} - \mathbf{C}h)^T (\mathbf{S} - \mathbf{C}h)\right\}$$

$$\min_h E\left\{\mathbf{S}^T\mathbf{S} - 2h^T\mathbf{C}^T\mathbf{S} + h^T\mathbf{C}^T\mathbf{C}h\right\}$$

$$\min_h \left\{\varphi_{\mathbf{SS}} - 2h^T\varphi_{\mathbf{CS}} + h^T\varphi_{\mathbf{CC}}h\right\}$$

Dabei kennzeichnet $\varphi_{\mathbf{CS}}$ die *Kreuzkorrelation* zwischen den geschätzten Blöcken **C** und den Originalblöcken **S** bzw. $\varphi_{\mathbf{CC}}$ die *Autokorrelation* zwischen den geschätzten Blöcken **C** [15].

$$\varphi_{\mathbf{CS}} = E\left\{\mathbf{C}^T\mathbf{S}\right\} \tag{A.6}$$

Die Korrelationen können mit Hilfe der ersten Wiener-Bedingung in Kovarianzen transformiert werden.

$$\begin{aligned} \sigma^2_{\mathbf{CS}} &= E\left\{(\mathbf{C}-\mu_{\mathbf{C}})^T(\mathbf{S}-\mu_{\mathbf{S}})\right\} && \text{(A.7)} \\ &= \varphi_{\mathbf{CS}} - \mu^T_{\mathbf{C}}\mu_{\mathbf{S}} && \text{(A.8)} \end{aligned}$$

Die bei der Umformung der zweiten Wiener-Bedingung auftretenden Terme $\mu_{\mathbf{C}}h$ vereinfachen sich durch **Gleichung A.5** zu $\mu_{\mathbf{S}}$.

$$\min_h \left\{\sigma^2_{\mathbf{SS}} - 2h^T\sigma^2_{\mathbf{CS}} + h^T\sigma^2_{\mathbf{CC}}h\right\} \tag{A.9}$$

Die Hypothesen befinden sich alle im gleichen Suchraum. Nehmen wir eine Ortsinvarianz erster Ordnung für die Hypothesen an, so sind die Mittelwerte der Hypothesen identisch und stimmen mit dem Mittelwert der Originalblöcke überein.

$$\mu_{\mathbf{C}} = \mu_{\mathbf{S}}u^T \tag{A.10}$$

Mit dieser Annahme und mit Hilfe der zweiten Wiener-Bedingung ergibt sich die Nebenbedingung $u^Th = 1$ für das vorliegende Optimierungsproblem.

$$\mu_{\mathbf{C}}h = \mu_{\mathbf{S}}u^Th = \mu_{\mathbf{S}} \tag{A.11}$$

Das Wiener-Problem mit der Nebenbedingung aus **Gleichung A.3** läßt sich nun mit dem Lagrange-Parameter ε zu einem Minimierungsproblem ohne Nebenbedingung formulieren.

$$\min_h \left\{\sigma^2_{\mathbf{SS}} - 2h^T\sigma^2_{\mathbf{CS}} + h^T\sigma^2_{\mathbf{CC}}h + \varepsilon\left(h^Tu - 1\right)\right\} \tag{A.12}$$

Aus der notwendigen Bedingung für ein Minimum

$$-2\sigma^2_{\mathbf{CS}} + 2\sigma^2_{\mathbf{CC}}h + \varepsilon u = 0$$

resultiert für die Interpolationskoeffizienten

$$h = {\sigma^2_{\mathbf{CC}}}^{-1}\left(\sigma^2_{\mathbf{CS}} - \frac{\varepsilon}{2}u\right).$$

Die Nebenbedingung $u^Th = 1$ ist weiterhin zu erfüllen und spezifiziert den Lagrange-Parameter ε.

$$\frac{\varepsilon}{2} = \frac{u^T{\sigma^2_{\mathbf{CC}}}^{-1}\sigma^2_{\mathbf{CS}} - 1}{u^T{\sigma^2_{\mathbf{CC}}}^{-1}u}$$

Die optimalen Interpolationskoeffizienten für jeden Schritt im iterativen Entwurfsalgorithmus ergeben sich nun aus der *Wiener-Lösung mit Nebenbedingung.*

$$h = {\sigma^2_{\mathbf{CC}}}^{-1}\left(\sigma^2_{\mathbf{CS}} - \frac{u^T{\sigma^2_{\mathbf{CC}}}^{-1}\sigma^2_{\mathbf{CS}} - 1}{u^T{\sigma^2_{\mathbf{CC}}}^{-1}u}u\right) \tag{A.13}$$

Diese so erhaltenen Interpolationskoeffizienten erfüllen also die Nebenbedingung $u^Th = 1$ und lösen das Wiener-Problem.

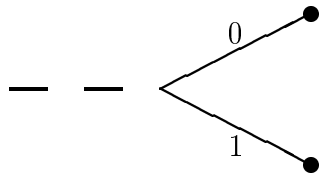
Anhang B

Huffman-Codierung

Die Huffman-Codierung stellt eine einfache und optimale Konstruktionsmethode zur Codierung einer ergodischen Markov-Quelle dar.

Schritt 0: Die k Indizes seien die Endknoten eines binären Baumes mit unbekannter Struktur. Ordne den Endknoten die Auftrittswahrscheinlichkeiten p_i für $i = 1, 2, \ldots, k$ der Indizes zu. Betrachte diese k Knoten als *aktiv*.

Schritt 1: Vereinige die beiden am wenigsten wahrscheinlichen Knoten mit einem binären Zweig wie folgt:



Setze diese beiden Knoten *passiv*, setze den neugeschaffenen Knoten *aktiv* und weise ihm die Summe der Wahrscheinlichkeiten der beiden verbundenen Knoten zu.

Schritt 2: Wenn nur noch ein aktiver Knoten übrig ist, sind wir an der Wurzel und halten an. Im anderen Fall gehe zu Schritt 1.

Abbildung B.1: *Huffman-Algorithmus zur Konstruktion eines binären präfixfreien Codes*

Der Huffman-Algorithmus nach **Abbildung B.1** [9] generiert einen binären präfixfreien Code, der folgende Eigenschaften besitzt:

1. Keine zwei Codeworte dürfen identisch sein.
2. Kein Codewort darf Präfix eines längeren Codewortes sein. Diese Forderung bedeutet, daß man ein Codewort erkennen kann, sobald dessen letztes Symbol empfangen wurde.

Das folgende Beispiel [20] verdeutlicht den Huffman-Algorithmus. Es werden fünf Indizes A, B, C, D, E mit den Wahrscheinlichkeiten $p_A = 0.264$, $p_B = 0.053$, $p_C = 0.108$, $p_D = 0.137$ und $p_E = 0.438$ codiert. In vier Schritten wird der Huffman-Baum nach dem Huffman-Algorithmus konstruiert.

1. Verbinde B und C zu einem Baum der Wahrscheinlichkeit $p_{BC} = 0.162$. p_{BC} bezeichnet dabei die Wahrscheinlichkeit, daß entweder Index B oder Index C auftritt.

2. Verbinde BC und D zu einem Baum der Wahrscheinlichkeit $p_{BCD} = 0.298$.

3. Verbinde A und BCD zu einem Baum der Wahrscheinlichkeit $p_{ABCD} = 0.562$.

4. Verbinde ABCD und E zu einem Baum der Wahrscheinlichkeit $p_{ABCDE} = 1.000$.

Die Konstruktion des Huffman-Baumes wird in **Abbildung B.2** veranschaulicht. Dabei werden die rechten Zweige mit dem Codesymbol „1“ und die linken Zweige mit dem Codesymbol „0“ gekennzeichnet. Die Vereinigung der beiden jeweils am wenigsten wahrscheinlichen Knoten mit einem binären Zweig ist durch die Schrittnummer gekennzeichnet.

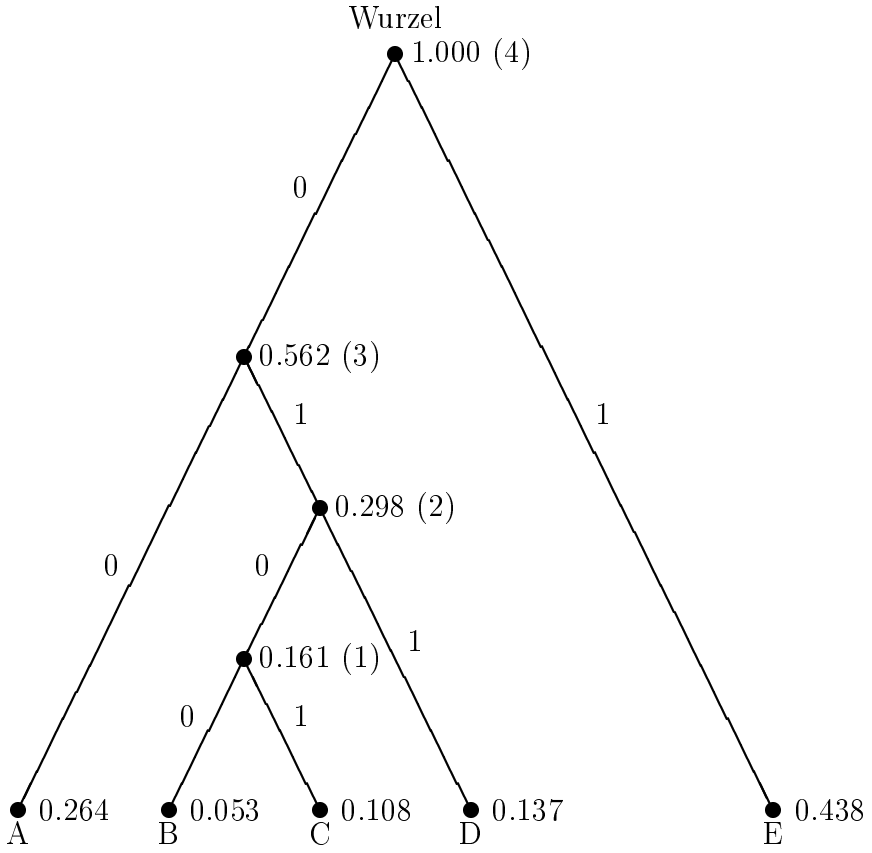


Abbildung B.2: *Konstruktion eines Huffman-Baums*

Man erhält die Codeworte indem man von der Wurzel des Huffman-Baumes zu den Endknoten (Indizes) läuft und die an den Zweigen vermerkten Codesymbole von links nach rechts notiert. **Tabelle B.1** zeigt die Zuordnung der Codeworte zu den Indizes.

Index	Wahrscheinlichkeit	Codewort
A	0.264	00
B	0.053	0100
C	0.108	0101
D	0.137	011
E	0.438	1

Tabelle B.1: *Zuordnung der Codeworte zu den Indizes*

Zur Beurteilung der Qualität des Huffman-Codes wird abschließend die Index-Entropie H mit der mittleren Codewortlänge L verglichen. Die Index-Entropie ist die untere Schranke für die mittlere Codewortlänge und wird für dieses Beispiel relativ gut erreicht.

$$H = -\sum_{i=1}^{k} p_i \log_2 p_i = 1.993\text{bit} \qquad L = \sum_{i=1}^{k} p_i l_i = 2.021\text{bit}$$

Anhang C

Simulationsbedingungen

Die Beurteilung basiert auf den Testsequenzen „Foreman“ und „Mother and Daughter“ mit jeweils einer Dauer von 10 s. Es werden die Testsequenzen in QCIF-Auflösung bei 7.5 Vollbildern pro Sekunde untersucht.

Zur Berechnung des Prädiktionsfehlers wird über die quadratischen Fehler zwischen Originalblöcken $\mathbf{s}_w$ und geschätzten Blöcken $\hat{\mathbf{s}}_w$ der Y-Komponente arithmetisch gemittelt.

$$D = \frac{1}{W} \sum_{w=1}^{W} \|\mathbf{s}_w - \hat{\mathbf{s}}_w\|^2 \tag{C.1}$$

W ist die Anzahl der Blöcke, die für die Schätzung der Distortion prädiziert werden. Dabei sind die Blöcke im ersten Bild der jeweiligen Sequenz nicht prädiziert. Bei der Berechnung des logarithmischen Prädiktionsfehlers wird die Distortion auf den quadratischen Maximalwert pro Bildelement normiert. Jeder Block enthält dabei s^2 Bildelemente.

$$PD = 10 \log_{10} \left(\frac{255^2 \cdot s^2}{D} \right) \tag{C.2}$$

In der Literatur wird häufig für jedes Bild der logarithmische Wert berechnet und für die Sequenz das arithmetische Mittel der einzelnen logarithmischen Werte ($\overline{PD}$) angegeben. Wird von logarithmischen Maßen zur Basis b ein arithmetischer Mittelwert gebildet, so werden die ursprünglichen Maße geometrisch gemittelt.

$$\frac{1}{K} \sum_{k=1}^{K} \log_b D[k] = \log_b \left(\prod_{k=1}^{K} D[k] \right)^{\frac{1}{K}} = \log_b D_g \tag{C.3}$$

Wir bevorzugen hingegen den arithmetischen Mittelwert nicht nur in Ortsrichtung, sondern auch in Zeitrichtung, d.h wir unterscheiden bei der Schätzung des Distortion nicht zwischen Ort und Zeit.

$$D_a = \frac{1}{K} \sum_{k=1}^{K} D[k] \tag{C.4}$$

Der geometrische Mittelwert ist dabei immer kleiner oder gleich dem arithmetischen Mittelwert [21]. **Abbildung C.1** zeigt eine geometrische Interpretation für arithmetischen und geometrischen Mittelwert.

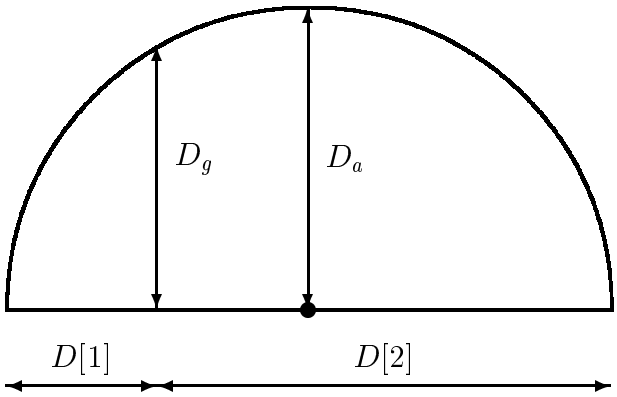


Abbildung C.1: *Veranschaulichung des arithmetischen Mittels* $D_a = \frac{1}{2}(D[1] + D[2])$ *und des geometrischen Mittels* $D_g = (D[1]D[2])^{\frac{1}{2}}$ *am Halbkreis mit dem Durchmesser* $D[1]+D[2]$. *Der geometrische Mittelwert ist immer kleiner oder gleich dem arithmetischen Mittelwert.*

Nehmen wir zum Beispiel eine Sequenz von K Bildern an, dessen erstes Bild dem Original entspricht ($D[1] = 0$) und dessen weitere Bilder stark verrauscht sind. Für den geometrischen Mittelwert erhält man $D_g = 0$ und somit für $\overline{PD} = \infty$. Das geometrische Mittel suggeriert in diesem Fall eine hohe Qualität vergleichbar mit der Originalsequenz, obwohl dies tatsächlich nur für das erste Bild zutrifft. Der arithmetische Mittelwert D_a der Sequenz mit dem ersten Bild ist hingegen nur um den Faktor $\frac{K-1}{K}$ kleiner als der arithmetische Mittelwert D'_a der Sequenz ohne das erste Bild. Der logarithmische Wert erhöht sich dabei nur um $10 \log_{10} \left(\frac{K}{K-1}\right) dB$ und ist für wachsendes K zu vernachlässigen.

Anhang D

Methode der n besten Matches

Für die Multihypothesen-Codierung eines Blockes $\mathbf{s}$ sind n Hypothesen $(\mathbf{c}_1, \ldots, \mathbf{c}_n)$ im Suchraum aufzufinden. In **Abschnitt 2.2.2** ist ein lokal optimaler Algorithmus zur Auswahl der Hypothesen vorgestellt. Im folgenden wollen wir das in **Abschnitt 2.1** vorgestellte Multihypothesen-Modell für die Multihypothesen-Prädiktion beibehalten, jedoch das Prinzip der n *besten Matches* anwenden.

Nach der Methode der n besten Matches sind im Suchraum n 1-Hypothesen zu finden, die man durch Rate-Distortion-optimales Blockmatching erhält. Rate-Distortion-optimales Blockmatching heißt, daß die 1-Hypothese $\mathbf{c}_1$ mit den geringsten Lagrange-Kosten gewählt wird.

$$\min_{\mathbf{c}_1} \left\{ \|\mathbf{s} - \mathbf{c}_1\|_2^2 + \lambda |\gamma(\mathbf{c}_1)| \right\}$$

Der Algorithmus funktioniert derart, daß sukzessiv der aktuelle optimale Match aus dem Suchraum entfernt und somit ein neuer Suchraum definiert wird, der den aktuellen optimalen Match nicht mehr enthält. Nach n Schritten ergeben sich die n besten Hypothesen. Nach dem Multihypothesen-Modell werden die so gefundenen und mit den Koeffizienten gewichteten Hypothesen summiert und bestimmen somit den geschätzten Block.

$$\hat{\mathbf{s}} = \sum_{\nu=1}^{n} \mathbf{c}_\nu h_\nu$$

Für das gegebene Multihypothesen-Modell ist die Methode der n besten Matches *nicht* optimal.

Abbildungen D.1 und D.2 zeigen Simulationsergebnisse für die Methode der n besten Matches im Vergleich zum lokal optimalen Algorithmus (OHSA). Die Simulationsbedingungen sind in **Abschnitt C** festgehalten. Die Ergebnisse sind mit $\lambda = 0$, Integer-Pel-Genauigkeit und gleichverteilten Gewichten $h_\nu = \frac{1}{n}$ erzielt.

Es ist festzustellen, daß kein Multihypothesen-Effekt auftritt. Für die Testsequenz „Mother and Daughter“ liegt der Prädiktionsfehler für alle untersuchten Hypothesenzahlen im Bereich der 1-Hypothese. Dies ist nachzuvollziehen, da effektiv nur eine Hypothese im Suchraum bestimmt wird. Für die Testsequenz „Foreman“ nimmt der Prädiktionsfehler für zunehmende Hypothesenzahlen zu. Dies ist damit zu erklären, daß mit zunehmender Hypothesenzahl auch qualitativ schlechtere Matches verwendet werden und der

Prädiktionsfehler sich somit im Vergleich zur 1-Hypothese verschlechtert. Daraus läßt sich schließen, daß n optimale Hypothesen nicht die Eigenschaften der 1-Hypothese besitzen.

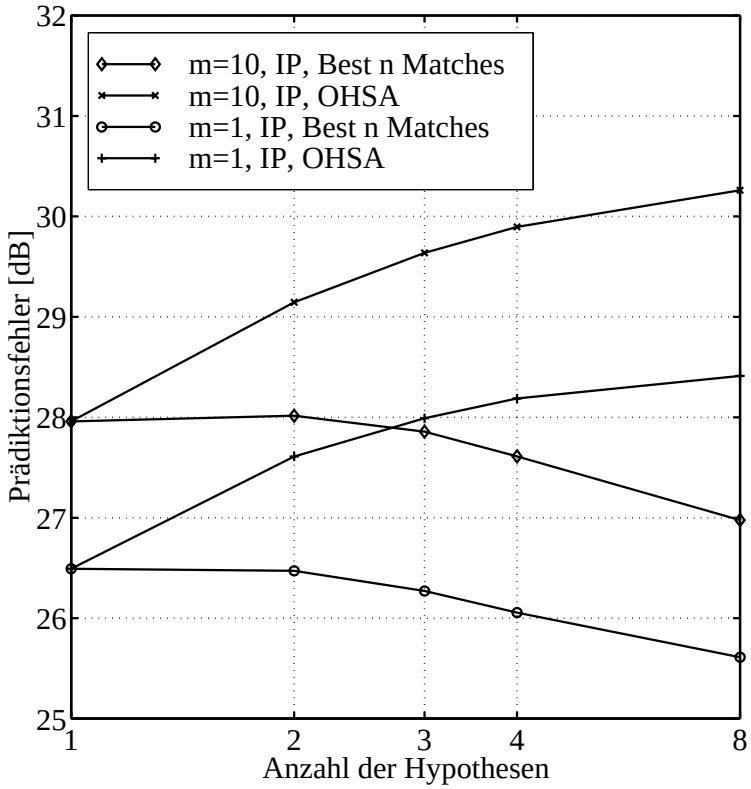


Abbildung D.1: *Prädiktionsfehler über die Anzahl der Hypothesen für die Testsequenz „Foreman“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcken*

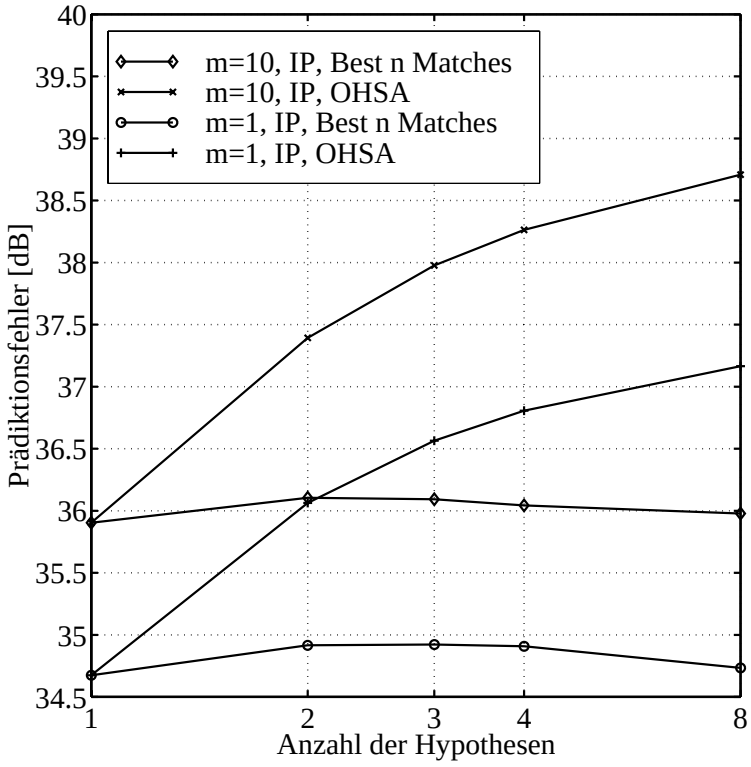


Abbildung D.2: *Prädiktionsfehler über die Anzahl der Hypothesen für die Testsequenz „Mother and Daughter“ (QCIF, 7.5 fps, 10 s) und* 16×16 *Blöcken*

Anhang E

Signaldarstellung und Systemkomponenten

$\mathcal{V}$ bezeichne die Menge aller zu prädizierenden Blöcke gleicher Größe. Dann sei die Quelle durch eine Zufallsvariable

$$\mathbf{S} : \mathcal{V} \rightarrow \{\mathbf{s}\} \tag{E.1}$$

beschrieben. Da die prädizierten Blöcke auch in der Menge der Blöcke gleicher Größe $\mathcal{V}$ enthalten sind, sind auch die prädizierten Blöcke durch eine Zufallsvariable

$$\hat{\mathbf{S}} : \mathcal{V} \rightarrow \{\hat{\mathbf{s}}\} \tag{E.2}$$

zu beschreiben. Alle möglichen Blöcke der Quelle sind gleich wahrscheinlich.

$$f_{\mathbf{S}}(\mathbf{s}) = const. \qquad \forall \mathbf{s} \tag{E.3}$$

Die Wahrscheinlichkeitsfunktion für eine abzählbare Menge $\mathcal{V}$ ist dabei durch die Wahrscheinlichkeiten der einzelnen Ereignisse definiert.

$$f_{\mathbf{S}}(\mathbf{s}) = P(\mathbf{S} = \mathbf{s}). \tag{E.4}$$

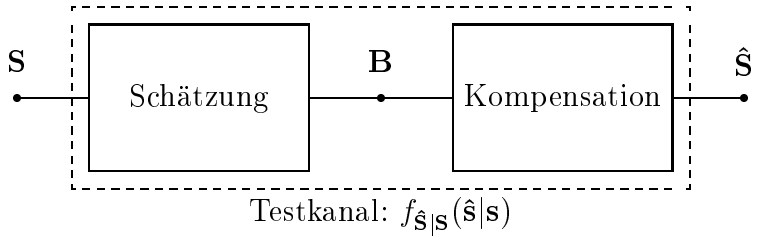


Abbildung E.1: *Modell des äquivalenten Testkanals*

Die bewegungskompensierte Prädiktion kann man als Testkanal auffassen und somit durch eine bedingte Wahrscheinlichkeitsfunktion $f_{\hat{\mathbf{S}}|\mathbf{S}}(\hat{\mathbf{s}}|\mathbf{s})$ beschreiben. **Abbildung E.1** veranschaulicht das Modell des äquivalenten Testkanals.

$\mathcal{B}$ bezeichnet die Menge aller Bewegungscodes. Der Bewegungscode sei beschrieben durch die Zufallsvariable

$$\mathbf{B} : \mathcal{B} \rightarrow \{\mathbf{b}\} \tag{E.5}$$

Der Schätzalgorithmus bildet die Blockquelle auf den Bewegungscode, der Bewegungskompensator den Bewegungscode auf die prädizierten Blöcke ab. Dabei besitzt der Bewegungskompensator die Eigenschaft einer eineindeutigen Abbildung.

Literaturverzeichnis

[1] ITU-T, *Video Coding for Low Bitrate Communication: Draft Recommendation H.263*, Okt. 1995.

[2] B. Girod, "The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences", *IEEE Journal on Selected Areas in Communications*, Band SAC-5, Num. 7, Aug. 1987.

[3] B. Girod, "Motion-Compensating Prediction with Fractional-Pel Accuracy", *IEEE Transactions on Communications*, Band 41, Num. 4, Apr. 1993.

[4] T. Wiegand, X. Zhang und B. Girod, "Motion-Compensating Long-Term Memory Prediction", in *Proceedings of the IEEE International Conference on Image Processing*, Okt. 1997.

[5] M.T. Orchard und G.J. Sullivan, "Overlapped Block Motion Compensation: An Estimation-Theoretic Approach", *IEEE Transactions on Image Processing*, Band 3, Num. 5, S. 693–699, Sept. 1994.

[6] S. Wu und A. Gersho, "Joint Estimation of Forward and Backward Motion Vectors for Interpolative Prediction of Video", *IEEE Transactions on Image Processing*, Band 3, Num. 5, S. 684–687, Sept. 1994.

[7] B. Girod, "Efficiency Analysis of Multi-Hypothesis Motion-Compensated Prediction for Video Coding", Telecommunications Institute, University of Erlangen-Nuremberg, 1997.

[8] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[9] R. Johannesson, *Informationstheorie - Grundlage der (Tele-) Kommunikation*, Addison-Wesley, Lund, 1992.

[10] C.E. Shannon, *The Collected Papers*, IEEE Press, New York, 1993.

[11] H. Everett III, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources", *Operations Research*, Band 11, S. 399–417, 1963.

[12] J. Besag, "On the Statistical Analysis of Dirty Pictures", *J. Roy. Statist. Soc. B*, Band 48, Num. 3, S. 259–302, 1986.

[13] P.A. Chou, T. Lookabaugh und R.M. Gray, "Entropy-Constrained Vector Quantization", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Band 37, S. 31–42, Jan. 1989.

[14] Y. Linde, A. Buzo und R.M. Gray, "An Algorithm for Vector Quantizer Design", *IEEE Transactions on Communications*, Band COM-28, S. 84–95, Jan. 1980.

[15] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1991.

[16] A. Gersho und R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Press, 1992.

[17] M.J. Sabin und R.M. Gray, "Global Convergence and Empirical Consistency of The Generalized Lloyd Algorithm", *IEEE Transactions on Information Theory*, Band 32, Num. 2, S. 148–155, Mářz 1986.

[18] A.M. Tekalp, *Digital Video Processing*, Prentice Hall, London, 1995.

[19] I.N. Bronstein und K.A Semendjajew, *Ergänzende Kapitel zum Taschenbuch der Mathematik*, Verlag Harri Deutsch, Frankfurt/Main, 1991.

[20] M.F. Barnsley und L.P. Hurd, *Fractal Image Compression*, A.K. Peters, Wellesley, Massachusetts, 1993.

[21] I.N. Bronstein und K.A. Semendjajew, *Taschenbuch der Mathematik*, Verlag Harri Deutsch, Frankfurt/Main, 1991.

[22] J.-R. Ohm, *Digitale Bildcodierung*, Springer-Verlag, Berlin, 1995.

[23] R. Rajagopalan und M.T. Orchard, "Jointly Optimal Forward-Backward Motion Compensation for Video Signals", Department of Electrical Engineering, Princeton University, 1996.

[24] J. Katto und M. Ohta, "An Analytical Framework for Overlapped Motion Compensation", in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1995, Band 5, S. 2189–2192.

[25] M. Ohta und S. Nogaki, "Hybrid Picture Coding with Wavelet Transform and Overlapped Motion-Compensated Interframe Prediction Coding", *IEEE Transactions on Signal Processing*, Band 41, Num. 12, S. 3416–3424, Dez. 1995.

[26] M.T. Orchard, "Predictive Motion-Field Segmentation for Image Sequence Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Band 3, Num. 1, S. 54–70, Feb. 1993.

[27] J. Katto, J. Ohki, S. Nogaki und M. Ohta, "A Wavelet Codec With Overlapped Motion Compensation for Very Low Bit-Rate Environment", *IEEE Transactions on Circuits and Systems for Video Technology*, Band 4, Num. 3, S. 328–338, Juni 1994.

[28] J.W. Woods, "Two-Dimensional Discrete Markovian Fields", *IEEE Transactions on Information Theory*, Band IT-18, Num. 2, S. 232–240, Mářz 1972.

[29] K. Rose, "A Mapping Approach to Rate-Distortion Computation and Analysis", *IEEE Transactions on Information Theory*, Band 40, Num. 6, S. 1939–1952, Nov. 1994.

[30] E.T. Jaynes, "On The Rationale of Maximum-Entropy Methods", *Proceedings of the IEEE*, Band 70, Num. 9, S. 939–952, Sept. 1982.