

Reihe 10

Informatik/
Kommunikation

Nr. 871

Dipl.-Ing. Thorsten Laude,
Langenhagen

Konturbasierte multidirektionale Intra-Prädiktion für die Videocodierung

The logo for the Institute for Information Processing (tnt) consists of the lowercase letters 'tnt' in a bold, sans-serif font. The letters are dark grey and are positioned above a horizontal line that spans the width of the letters. Below the line, the letters 'tnt' are repeated in a smaller, lighter grey font, creating a layered effect.

Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

Konturbasierte multidirektionale Intra-Prädiktion für die Videocodierung

Von der Fakultät für Elektrotechnik und Informatik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor-Ingenieur

genehmigte

Dissertation

von

Dipl.-Ing. Thorsten Laude

geboren am 29. Dezember 1988 in Hannover

2020

Hauptreferent: Prof. Dr.-Ing. Jörn Ostermann
Korreferent: Prof. Dr.-Ing. Jens-Rainer Ohm
Vorsitzender: Prof. Dr.-Ing. Bodo Rosenhahn

Tag der Promotion: 6. November 2020

Fortschritt-Berichte VDI

Reihe 10

Informatik/
Kommunikation

Dipl.-Ing. Thorsten Laude,
Langenhagen

Nr. 871

Konturbasierte
multidirektionale
Intra-Prädiktion für
die Videocodierung



Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

Laude, Thorsten

Konturbasierte multidirektionale Intra-Prädiktion für die Videocodierung

Fortschr.-Ber. VDI Reihe 10 Nr. 871. Düsseldorf: VDI Verlag 2021.

160 Seiten, 36 Bilder, 12 Tabellen.

ISBN 978-3-18-387110-0, ISSN 0178-9627,

€ 57,00/VDI-Mitgliederpreis € 51,30.

Keywords: Videocodierung – HEVC – Intra-Prädiktion – maschinelles Lernen – Gauß-Prozesse – Deep Learning – neuronale Netzwerke – stochastische Prozesse

In dieser Arbeit werden zwei Verfahren zur Verbesserung der Intra-Prädiktion für die Videocodierung vorgeschlagen. Der erste Beitrag in dieser Arbeit besteht aus einem stochastischen Konturmodell zur Modellierung und Extrapolation von Konturen, die im Referenzbereich detektiert werden. Für die Modellierung wird ein Gauß-Prozess verwendet. Für die Kontur-extrapolation wird eine multivariate Gauß-Verteilung formuliert. Der zweite Beitrag in dieser Arbeit ist ein auf neuronalen Netzwerken basierendes Verfahren zur Abtastwertprädiktion. Mit den neuronalen Netzwerken werden die benachbarten Referenzabtastwerte sowie das Ergebnis der Konturmodellierung und -extrapolation als Eingabedaten verarbeitet, um eine Prädiktion der Abtastwerte des zu codierenden Blocks zu erzeugen. Die Codierungseffizienz des Videocoders HEVC wird um bis zu 5% gesteigert.

Bibliographische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter www.dnb.de abrufbar.

Bibliographic information published by the Deutsche Bibliothek

(German National Library)

The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at www.dnb.de.

© VDI Verlag GmbH · Düsseldorf 2021

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe (Fotokopie, Mikrokopie), der Speicherung in Datenverarbeitungsanlagen, im Internet und das der Übersetzung, vorbehalten.

Als Manuskript gedruckt. Printed in Germany.

ISSN 0178-9627

ISBN 978-3-18-387110-0

VORWORT

Diese Dissertation habe ich als wissenschaftlicher Mitarbeiter des Instituts für Informationsverarbeitung (TNT) der Gottfried Wilhelm Leibniz Universität Hannover geschrieben. Meinem Doktorvater Professor Jörn Ostermann danke ich für die exzellente Betreuung, seine Ideen und Anregungen, das Hauptreferat zu dieser Dissertation, stets beste Arbeitsbedingungen und alles, was ich wissenschaftlich und andersweitig gelernt habe. Schon während meines Studiums gab er mir die Gelegenheit, am TNT zu forschen und an internationalen Standardisierungsaktivitäten teilzunehmen.

Professor Bodo Rosenhahn danke ich für zahlreiche wissenschaftliche Diskussionen und dafür, dass er als Vorsitzender der Prüfungskommission den reibungslosen Ablauf der Promotion unter den Bedingungen der Corona-Pandemie ermöglicht hat. Professor Jens-Rainer Ohm danke ich für die Übernahme des Korreferats.

Meinen ehemaligen Arbeitskolleginnen und -kollegen, von denen viele inzwischen enge Freunde sind, danke ich für ihre Hilfsbereitschaft, ihren Rat und für viele schöne gemeinsame Erlebnisse. Hierzu zählen unter anderem Hendrik Hachmann, der darüber hinaus ein ausgezeichnete Trauzeuge war, Issi Zell, Felix Kuhnke, Bastian Wandt, Stella Graßhof, Jan Voges, Marco Munderloh, der während meiner Zeit am TNT mein Mentor war, Holger Meuel, Benjamin Spitschan, Aron Sommer, Matthias Reso und Matthias Schuh, der das Institut technisch am Laufen hält und stets für eine gute Atmosphäre sorgt. Sontje Ihler danke ich ebenfalls für ihre Freundschaft und zahlreiche Diskussionen. Wegen ihnen war die Zeit am TNT eine sehr gute und schöne Zeit.

Matthias Narrosche danke ich, dass er mich zur Videocodierung gebracht hat. Solveig Behr, Doris Jaspers-Göring, Pia Bank, Hilke Brodersen, Thomas Wehberg und Martin Pahl danke ich für die kompetente administrative und technische Unterstützung.

Den Mitgliedern des Arbeitssaals Dachkammer danke ich für die unvergessliche Zeit und die gemeinsame Unterstützung während des Studiums. Hierzu zählen unter anderem Niklas Briest, Moritz Wallat, Malte John sowie Felix Burghardt.

Meiner Frau Jennifer Laude danke ich für ihre Liebe, ihre unermüdliche Unterstützung und auch für ihr Verständnis für lange Arbeitszeiten während der Fertigstellung der Dissertation. Meinen Eltern Birgit und Burkhard Laude danke ich dafür, dass sie mich während meines gesamten

Werdegangs unterstützt haben. Ohne sie wäre diese Arbeit nicht möglich gewesen. Meinen Schwiegereltern Petra und Norbert Lübke danke ich für die Aufnahme in ihre Familie.

INHALTSVERZEICHNIS

1	EINLEITUNG	1
1.1	Motivation	1
1.2	Stand der Forschung	4
1.3	Ungelöste Probleme	6
1.4	Ziele der Arbeit	7
1.5	Aufbau der Arbeit	9
2	GRUNDLAGEN	10
2.1	Videocodierung	10
2.2	Maschinelles Lernen	27
2.3	Konturdetektion	49
3	VERFAHREN ZUR MODELLIERUNG VON KONTUREN	53
3.1	Konturdetektion	53
3.2	Konturglättung	56
3.3	Modellierung der Kontur	57
3.4	A-Priori-Gauß-Prozess	58
3.5	Posterior-Gauß-Prozess	59
3.6	Konturextrapolation	63
3.7	Einbettung in das Gesamtsystem	65
4	VERFAHREN ZUR ABTASTWERTPRÄDIKTION MITTELS MASCHINELLEN LERNENS	67
4.1	Einordnung in das Gesamtsystem	67
4.2	Datenbasis	67
4.3	Architekturen	73
4.4	Training	78
5	EXPERIMENTELLE UNTERSUCHUNG UND BEWERTUNG	84
5.1	Integration in einen Bildcodec	85
5.1.1	Mehrwert des vorgeschlagenen Konturmodells	89
5.1.2	Mehrwert der vorgeschlagenen Abtastwertprädiktion	93
5.1.3	Einordnung von Effizienz und Komplexität	96
5.2	Integration in einen Videocodec	101
5.2.1	Polynomielles Konturmodell und Abtastwertprädiktion mit neuronalen Netzwerken	109
5.2.2	Stochastisches Konturmodell und Abtastwertprädiktion mit neuronalen Netzwerken	112
6	ZUSAMMENFASSUNG	119
	LITERATUR	123

FORMELZEICHEN

VIDEOCODIERUNG

r	Datenrate
H	Entropie
$S_{\text{horizontal}}$	Horizontale Chroma-Unterabtastung
I	Informationsgehalt
$\Xi\Phi\Psi$	Konstanten zur Modellierung der Prädiktionsfehlerkoeffizientendatenrate
λ	Lagrange-Faktor für die Rate-Distortion-Kosten
S_Y	Luma-Abtastrate
S_{Max}	Maximaler Signalwert
C	Prädiktionsfehlerkoeffizienten
R	Rate
C_{RD}	Rate-Distortion-Kosten
S_{vertikal}	Vertikale Chroma-Unterabtastung
D	Verzerrung, engl. Distortion

KONTURMODELLIERUNG

Θ_0	Anfangswerte der Hyperparameter der Kovarianzfunktion
\mathbf{b}	Beobachtungen des zu modellierenden Prozesses („Ground Truth“)
\mathbf{f}	Beobachtungen für rauschbehaftete Trainingspunkte
$\bar{\mathbf{f}}$	Beobachtungen für rauschfreie Trainingspunkte
\mathbf{f}_*	Beobachtungen für Testpunkte
I	Bild, engl. Image

s	Blockgröße
$ \dots $	Determinantenoperator
\mathcal{I}	Einheitsmatrix
\mathcal{GP}	Gauß-Prozess
Φ	Gradientenrichtungen für die Konturdetektion
Θ	Hyperparameter der Kovarianzfunktion
c_{x_p}	Konfidenzfunktion im extrapolierten Konturpunkt x_p
G_{skel}	Konturenskelett
\mathbf{K}	Konturbild
$k(x_p, x_q)$	Kovarianzfunktion
$K(A, B)$	Kovarianzmatrix für die Zufallsvariablen A und B
Σ	Kovarianzmatrix inklusive unkorreliertem weißen Rauschen
ξ	Längsskalierungsfaktor für $k_{\text{SE}}(x_p, x_q)$
\mathcal{N}	Normal-Verteilung
$\tilde{\mathbf{b}}$	Prior-Beobachtungen für den modellierten Prozess (Prädiktion)
\mathbf{p}	Prädiktionen (Posterior-Beobachtungen)
K_{post}	Posterior-Kovarianzmatrix
Σ_{post}	Posterior-Kovarianzmatrix mit Rauschen
ζ	Skalierungsfaktor für $k_{\text{SE}}(x_p, x_q)$
K_{Sobel}	Sobel-Operator
$k_{\text{SE}}(x_p, x_q)$	<i>Squared Exponential</i> -Kernel bzw. -Kovarianzfunktion
σ_{p_i}	Standardabweichung der Prädiktion für den Punkt x_{p_i}
x_{*j}, \mathbf{x}_*	Testpunkte
T	Tiefpassfilter für die Konturdetektion
x_i, \mathbf{x}	Trainingspunkte

FORMELZEICHEN

ϵ	unkorrelliertes weißes Rauschen
L	untere Dreiecksmatrix nach Cholesky-Zerlegung
σ_{pi}^2	Varianz der Prädiktion für den Punkt x_{pi}
σ_n^2	Varianz des unkorrelierten weißen Rauschens ϵ
$p(\dots)$	Wahrscheinlichkeitsdichte für das Argument
$P(\dots)$	Wahrscheinlichkeit für das Argument
x_{pi}, \mathbf{x}_p	x-Werte der Posterior-Prädiktionen
y_i, \mathbf{y}	y-Werte der Konturpixel
y_{pi}, \mathbf{x}_p	y-Werte der Posterior-Beobachtungen

MASCHINELLES LERNEN

a	Aktivierungsfunktion
o	Aktivierungssignal eines Neurons
α	Aktualisierungsschrittweise beim Gradientenabstiegsverfahren mit Momentum
N	Anzahl an verwendeten Trainingsbeispielen
\mathbf{b}	Biasvektor einer Schicht eines neuronalen Netzwerks
i	Eingangssignal eines Neurons
e	Eulersche Zahl
e_{gen}	Generalisierungsfehler
w	Gewicht eines neuronalen Netzwerks
\mathbf{W}	Gewichtsmatrix einer Netzwerkschicht
\mathbf{W}_F	Gewichtsmatrix einer Faltungsschicht
b	Gleichanteil, engl. bias, eines Neurons
$\nabla_c C$	Gradient von C nach c
g_i	Gradient zum Zeitpunkt i

β_1, β_2	Hyperparameter des Adam-Verfahrens
l	Index der Schicht eines neuronalen Netzwerkes
$\mathcal{O}()$	Komplexität
c	Konfiguration eines neuronalen Netzwerkes
C	Kostenfunktion für das Training eines neuronalen Netzwerkes
v_i	Leistung zum Zeitpunkt i
η	Lernrate
N^*	Minibatch-Größe
m_i	Mittelwert zum Zeitpunkt i
μ	Momentum
e_{Test}	Testfehler
e_{Training}	Trainingsfehler
$o(i)$	Übertragungsfunktion eines Neurons

AKRONYME

Adam	Adaptive Moment Estimation
AI	all intra
AV ₁	AOMedia Video 1
AVC	Advanced Video Coding
BAC	binäre arithmetische Codierung
BD	Bjøntegaard Delta
Bins	binäre Symbole vor der Entropiecodierung
Bits	binäre Symbole nach der Entropiecodierung
BVI	Bristol Vision Institute
CABAC	Context-adaptive Binary Arithmetic Coding
CAVLC	Context-adaptive Variable Length Coding
CIE	Commission Internationale de l'Éclairage
CoMIC	Contour-based Multidirectional Intra Coding
CAE	Convolutional Auto Encoder
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CTB	Coding Tree Block
CTC	Common Test Conditions
CTU	Coding Tree Unit
CU	Coding Unit
DC	Direct Current
DCT	Discrete Cosine Transform
DPB	Decoded Picture Buffer

DPCM	Differential Pulse Code Modulation
DST	Discrete Sine Transform
GOP	Group of Pictures
GPU	Graphics Processing Unit
HDF5	Hierarchical Data Format 5
HEVC	High Efficiency Video Coding
HLS	High-level Syntax
HM	High Efficiency Video Coding (HEVC) Test Model
IBC	Intra Block Copy
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
IEC	International Electrotechnical Commission
ISO	International Standardization Organisation
JBIG	Joint Bi-level Image Experts Group
JCT-VC	Joint Collaborative Team on Video Coding
JEM	Joint Exploration Model
JPEG	Joint Photographic Experts Group
JVET	Joint Video Experts Team
KLT	Karhunen-Loève-Transformation
KODIM	Kodak Image Dataset
lreLU	Leaky Rectifier Linear Unit
MIP	Matrix-weighted Intra Prediction
MPEG	Moving Picture Experts Group
MPM	Most Probable Modes
MSE	Mean Squared Error

AKRONYME

MV	Motion Vector
NAL	Network Abstraction Layer
NMS	Non-maximum Suppression
PB	Prediction Block
PCA	Principal Component Analysis
Pel	picture element
PSNR	Peak Signal-to-Noise Ratio
PU	Prediction Unit
QP	Quantisierungsparameter
RAISE	Raw Image Dataset
RD	Rate-Distortion
RDO	Rate-Distortion Optimization
ReLU	Rectifier Linear Unit
RGB	Rot-Grün-Blau
RNN	Recurrent Neural Network
SATD	Sum of Absolute Transformed Differences
SSIM	Structural Similarity
TB	Transform Block
TPU	Tensor Processing Unit
TU	Transform Unit
USC-SIPI	University of Southern California - Signal and Information Processing Institute
VCEG	Video Coding Experts Group
VTM	Versatile Test Model
WPP	Wavefront Parallel Processing

ABSTRACT

The amount of transmitted video data is growing faster than the channel capacity available for this purpose. This leads to the necessity of a continuous improvement of the coding methods for the used video codecs. Modern video codecs are generally based on the principle of hybrid coding, i.e. the combination of a prediction with a transformation coding of the prediction error. The prediction methods can be roughly divided into intra and inter prediction. In this work, two methods are proposed for improving intra prediction.

The first contribution in this thesis is a stochastic contour model for modeling and extrapolation of contours detected in the reference area. A Gaussian process is used for the modeling. Expectations of typically occurring contour shapes were taken into account by choosing the *Squared Exponential Kernel* as covariance function of the prior of the Gaussian process. The posterior Gaussian process resulted from the prior Gaussian process by optimizing the hyperparameters of the covariance function for each contour. A multivariate Gaussian distribution was formulated for the contour extrapolation. The second contribution in this thesis is a neural network-based method for sample value prediction. The neural networks are used to process the adjacent reference sample values and the results of contour modeling and extrapolation as input data to generate a prediction of the sample values of the block to be coded. The contours are required for the sample value prediction. The neural networks were designed with an auto-encoder architecture and trained using a SATD cost function.

The coding efficiency of the video codec HEVC was increased by up to 5%. Averaged over all 55 test sequences, the *All Intra* configuration resulted in BD-rates of -0.54% for high bit rates and -1.0% for low bit rates. Compared to the methods from our own prior works, which were already better than related works from the literature, an additional increase in coding efficiency of 0.21 percentage points for high bit rates and 0.27 percentage points for low bit rates was achieved.

Keywords – video coding, HEVC, intra prediction, machine learning, Gaussian process.

KURZFASSUNG

Die zur Übertragung von Videos benötigte Übertragungskapazität wächst schneller als die hierfür zur Verfügung stehende Kanalkapazität. Hieraus entsteht die Notwendigkeit einer stetigen Verbesserung der Codierungsverfahren für die verwendeten Videocodecs. Moderne Videocodecs beruhen in der Regel auf dem Prinzip der Hybridcodierung, also der Kombination von einer Prädiktion mit einer Transformationscodierung des Prädiktionsfehlers. Die Prädiktionsverfahren können grob in Intra- und Inter-Prädiktion unterschieden werden. Für die Verbesserung der Intra-Prädiktion werden in dieser Arbeit zwei Verfahren vorgeschlagen.

Der erste Beitrag in dieser Arbeit besteht aus einem stochastischen Konturmodell zur Modellierung und Extrapolation von Konturen, die im Referenzbereich detektiert werden. Für die Modellierung wird ein Gauß-Prozess verwendet. Die Erwartungen an typischerweise vorkommende Konturverläufe werden durch die Wahl des *Squared Exponential Kernels* als Kovarianzfunktion des Prior-Gauß-Prozesses berücksichtigt. Der Posterior-Gauß-Prozess ergibt sich aus dem Prior-Gauß-Prozess durch die Optimierung der Hyperparameter der Kovarianzfunktion für jede Kontur. Für die Konturextrapolation wird eine multivariate Gauß-Verteilung formuliert. Der zweite Beitrag in dieser Arbeit ist ein auf neuronalen Netzwerken basierendes Verfahren zur Abtastwertprädiktion. Mit den neuronalen Netzwerken werden die benachbarten Referenzabtastwerte sowie das Ergebnis der Konturmodellierung und -extrapolation als Eingabedaten verarbeitet, um eine Prädiktion der Abtastwerte des zu codierenden Blocks zu erzeugen. Die Konturen werden für die Abtastwertprädiktion benötigt. Die neuronalen Netzwerke wurden mit einer Autoencoder-Architektur entworfen und mittels einer SATD-Kostenfunktion trainiert.

Die Codierungseffizienz des Videocodecs HEVC wurde um bis zu 5% gesteigert. Gemittelt über alle 55 Testsequenzen ergaben sich für die *All Intra*-Konfiguration BD-Raten von $-0,54\%$ für hohe Datenraten und in Höhe von $-1,0\%$ für niedrige Datenraten. Verglichen mit den Verfahren aus eigenen Vorarbeiten, welche bereits besser waren als verwandte Arbeiten aus der Literatur, wurde eine zusätzliche Steigerung der Codierungseffizienz um 0,21 Prozentpunkte für hohe Datenraten und um 0,27 Prozentpunkte für niedrige Datenraten erzielt.

Stichworte – Videocodierung, HEVC, Intra-Prädiktion, maschinelles Lernen, Gauß-Prozesse