Fortschritt-Berichte VDI

VDI

Reihe 10

Informatik/
Kommunikation

Dipl.-Ing. Holger Meuel,
Hannover

Nr. 865

Analysis of Affine Motion-Compensated Prediction and its Application in Aerial Video Coding

tnt

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

Analysis of Affine Motion-Compensated Prediction and its Application in Aerial Video Coding

Von der Fakultät für Elektrotechnik und Informatik

der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor-Ingenieur

(abgekürzt: Dr.-Ing.)

genehmigte

Dissertation

von

Dipl.-Ing. Holger Meuel

geboren am 27. Februar 1983 in Lübeck

2019

Hauptreferent:	Prof. Dr.-Ing. Jörn Ostermann
Korreferent:	Prof. Dr.-Ing. André Kaup
Vorsitzender:	Prof. Dr.-Ing. Hans-Georg Musmann

Tag der Promotion:	5. August 2019

Fortschritt-Berichte VDI

Reihe 10

Informatik/
Kommunikation

Dipl.-Ing. Holger Meuel,
Hannover

Nr. 865

Analysis of Affine Motion-Compensated Prediction and its Application in Aerial Video Coding

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

Meuel, Holger
Analysis of Affine Motion-Compensated Prediction and its Application in Aerial Video Coding
Fortschr.-Ber. VDI Reihe 10 Nr. 865. Düsseldorf: VDI Verlag 2019.
168 Seiten, 53 Bilder, 9 Tabellen.
ISBN 978-3-18-386510-9, ISSN 0178-9627,
€ 62,00/VDI-Mitgliederpreis € 55,80.

Keywords: video coding – affine motion-compensated prediction (MCP) – simplified affine motion-compensated prediction – rate-distortion theory – aerial surveillance – region of interest-(ROI-) based video coding – global motion compensation (GMC) – moving object detection – long-term mosaicking – radial distortion compensation

This work deals with affine motion-compensated prediction (MCP) in video coding. Using the rate-distortion theory and the displacement estimation error caused by inaccurate motion parameter estimation, the minimum bit rate for encoding the prediction error is derived. Similarly, a 4-parameter simplified affine model as considered for the upcoming video coding standard VVC is analyzed. Both models provide valuable information about the minimum bit rate for encoding the prediction error as a function of the motion estimation accuracy.

Although the bit rate in MCP can be reduced by using a motion model capable of describing the motion in the scene, the total video bit rate may remain high. Thus, a codec independent coding system is proposed for aerial videos, which exploits the planarity of such sequences. Only new emerging areas and moving objects in each frame are encoded. From these, the decoder reconstructs a mosaic, from which video frames are extracted again. This system outperforms HEVC by 90 %.

Finally, a fully automatic in-loop radial distortion compensation for the generation of long-term mosaics from 1500 aerial frames is proposed.

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.

Als Manuskript gedruckt. Printed in Germany.

ISSN 0178-9627

ISBN 978-3-18-386510-9

Acknowledgement

This thesis was written during my time at the Institut für Informationsverarbeitung (TNT) of the Gottfried Wilhelm Leibniz Universität Hannover.

My special thank goes to Prof. Dr.-Ing. Jörn Ostermann who provided the possibility to work at the institute. He continuously supported me financially and scientifically. Particularly, I would like to thank for the intense and valuable discussions and supervision during the development of this thesis and of course for the evaluation of my thesis as first examiner. I also would like to thank Prof. Dr.-Ing. André Kaup for being the second examiner of this thesis, his helpful comments and the discussions at several opportunities. I also cordially thank Prof. Dr.-Ing. Hans-Georg Musmann for taking over the chair of the examination board and his continuous scientific support during my time at the TNT. For the inspiring discussions I like to thank Prof. Dr.-Ing. Bodo Rosenhahn who offered friendly support at all times.

Moreover, I especially thank all my colleagues. In particular, I owe my deep gratitude to Dr.-Ing. Marco Munderloh and Dr.-Ing. Ulrike Pestel-Schiller. Thanks for the continuous support in any matter from the very beginning until the defense of my thesis in word and deed! My deep appreciation also goes to my room mate Yiqun Liu who supported me relentlessly in any issue. I like to specially thank Stephan Ferenz, Hendrik Hachmann, Florian Kluger, Hanno Ackermann, Ph.D., Dr.-Ing. Aron Sommer, Dr.-Ing. Karsten Vogt, Stella Graßhof, Benjamin Spitschan, Dr.-Ing. Christian Becker, and Yasser Samayoa for plenty of discussions, general and mathematical support, and their encouragement. My acknowledgment also goes to my former room mate Julia Schmidt for her help and advice in business and personal matters. Thanks for all the fruitful cooperations resulting in publications, scientific and personal development and finally this thesis. For their sedulous support I like to thank Matthias Schuh, Doris Jaspers-Göring, Hilke Brodersen, Melanie Huch and the entire former office staff. For their administrative and constant support my acknowledgment goes to Dr.-Ing. Martin Pahl and Thomas Wehberg. Thanks for the great and inspiring time!

I thank my sister Sylvia Nissen for her good wishes and thoughts and my parents Ingrid and Dr. rer. nat. Bernd Meuel for enabling me to study electrical engineering as a basis for this thesis.

Finally, my deepest gratitude goes to my wife Dr. rer. nat. Katharina Neuhäuser for her untiring magnificent support without this thesis would not have been finalized. Thanks for encouraging me over the entire time, the substantial support and for always lighting up my life! I also owe my gratitude to Katharina's parents Emma and Prof. Dr. rer. nat. Hartmut Neuhäuser for their unconditional support in any matter and for integrating me into their family like a son.

This work is dedicated to my wife.

Contents

Abbreviations and Symbols

Abbreviations

AV1	AOMedia Video 1
AVC	Advanced Video Coding (H.264, MPEG-4 part 10)
AWGN	Additive white Gaussian noise
B-frame	Bidirectionally predicted frame
B	Byte
BD	Bjøntegaard delta
BD-PSNR	Bjøntegaard delta PSNR
BD-rate	Bjøntegaard delta rate
CABAC	Context-adaptive binary arithmetic coding
CCR	Comparison category rating (also known as double stimulus comparison or pair comparison method)
CIF	Common Intermediate Format, CIF video sequences have a resolution of 352 × 288 pel and are recorded at 30 fps
CMOS	Complementary metal-oxide-semiconductor
Codec	Coder-decoder
CRF	Corner response function
CTU	Coding tree unit
DCT	Discrete cosine transform
DoF	Degree of Freedom
DPCM	Differential pulse-code modulation
DVB	Digital Video Broadcasting
DVB-C/-C2	Digital Video Broadcasting – Cable (1st/2nd generation)
DVB-S/-S2	Digital Video Broadcasting – Satellite (1st/2nd generation)
DVB-T/-T2	Digital Video Broadcasting – Terrestrial (1st/2nd generation)
FP	False positive (detections)

Fps	Frames per second
GMC	Global motion compensation
GME	Global motion estimation
GOF	Group of frames (for in-loop radial distortion compensation)
GUI	Graphical user interface
HD	High definition (HD resolution equals 1920 × 1080 pel)
HEVC	High Efficiency Video Coding (H.265, MPEG-H part 2)
HM	HEVC Test Model
I-frame	Intra-coded frame
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
ITU	International Telecommunication Union, former: The International Telegraph and Telephone Consultative Committee (CCITT, from French: Comité Consultatif International Téléphonique et Télégraphique), former International Telegraph Union (ITU)
ITU-T	ITU Telecommunication Standardization Sector
JCT-VC	Joint Collaborative Team on Video Coding
JEM	Joint Exploration Model of JVET
JVET	Joint Video Exploration Team (on Future Video Coding) of ITU-T VCEG and ISO/IEC MPEG founded in October 2015, later transitioned into Joint Video Experts Team (also abbreviated by JVET) in April 2018
kbit	Kilobit
KLT	Kanade-Lucas-Tomasi feature tracker
LD	Low-delay
LDP	Low-delay p
MB	Megabyte
Mbit	Megabit
MC	Motion compensation
MCP	Motion-compensated prediction
ME	Motion estimation
MO	Moving object
MPEG	Motion Picture Experts Group
MPEG-4 ASP	MPEG-4 Advanced Simple Profile
MSE	Mean squared error

MV	Motion vector
MVP	Motion vector prediction
NA	New area
P-frame	Predicted frame
PCM	Pulse-code modulation
Pdf	Probability density function
Pel	Picture element (also known as pixel)
PSD	Power spectral density
PSNR	Peak signal-to-noise ratio
QCIF	Quarter CIF, QCIF video sequences have a resolution of 176 × 144 pel and are recorded at 30 fps
RA	Random-access profile
RANSAC	Random sample consensus
RD	Rate-distortion
RDC	Radial distortion compensation
RDO	Rate-distortion optimization
ROI	Region of interest
ROI-MO	Region of interest – moving object
ROI-NA	Region of interest – new area
ROI-PSNR	PSNR of ROI areas
SAD	Sum of absolute differences
SEI	Supplemental enhancement information
SfM	Structure from motion
SIFT	Scale-Invariant Feature Transform
SNR	Signal-to-noise ratio
s	Second
SSD	Sum of squared differences
TCS	Temporally consistent superpixel
TP	True positive (detections)
TV	Television
UAV	Unmanned aerial vehicle
VCEG	Video Coding Experts Group
VOD	Video on demand
VVC	Versatile Video Coding
x265	Open-source HEVC video encoder software
Y	Luminance component

Symbols

a, b	Parameter of the simplified affine model
$\mathtt{A}$	Affine matrix of size 2 × 2
A	Auxiliary variable
$\mathtt{A}_f$	Fully affine matrix of size 2 × 3
a_{ij}	Entries of the fully affine matrix, $i = \{1,2\}$, $j = \{1,2,3\}$
a_i	Entries of the simplified affine matrix, $i = \{a,b,c,f\}$
α	Drop rate of an exponential isotropic (autocorrelation) function
α_x, α_y	Drop rates of exponential (autocorrelation) functions in x- and y-direction
B_{CRF}	Maximum number of feature points per frame
$b_k(\boldsymbol{n})$	Binarized image intensity differences of the frame k
c	Parameter of the simplified affine model (translation in x-direction)
$\boldsymbol{C} = (C_x,C_y,C_z)^\top$	Position of the camera in world coordinates
$c_{\mathrm{size,max}}$, $c_{\mathrm{shape,max}}$	Maximum allowed size and shape change in in-loop radial distortion compensation
c_x, c_y	Thresholds which limit rotations around the x- and y-axis, respectively, in in-loop radial distortion compensation
$\boldsymbol{d}$	Motion vector
D	Maximum allowed average distortion (rate-distortion theory)
$d(u;v)$	General distortion measure between symbols u and v (rate-distortion theory)
d_f	Minimum feature distance
$d_k(\boldsymbol{n})$	Image intensity differences of the frame k
$\boldsymbol{d} = (d_x,d_y)^\top$	Displacement vector
$\boldsymbol{d}_i = (d_{i,x},d_{i,y})^\top$	Displacement of the i-th feature
$\hat{\boldsymbol{d}}$	Estimate of $\boldsymbol{d}$
$\bar{d}$	Average distortion (rate-distortion theory)
${}^{\mathrm{simp}}D$	Distortion using a simplified affine model (rate-distortion theory)
$\Delta x'$, $\Delta y'$	Displacement estimation error in horizontal (x-) and vertical (y-) direction of the fully affine model

$\varDelta x'_{\text{mod}}, \varDelta y'_{\text{mod}}$	Displacement estimation error caused by an inappropriate motion model in horizontal (x-) and vertical (y-) direction
$\varDelta x'_{\text{s}}, \varDelta y'_{\text{s}}$	Displacement estimation error in horizontal (x-) and vertical (y-) direction of the simplified affine model
δ	Dirac delta function
$\boldsymbol{d}'$	Motion vector (for transmission) with limited accuracy
e	Prediction error signal
$E(\cdot)$	Expectation value of $(\cdot)$
$e_k(\boldsymbol{n})$	Binarized image intensity differences of the frame k after erosion
$e_{ij,\text{mod}}$	Error terms caused by the motion model, $i = \{1,2\}$, $j = \{1,2,3\}$
e'	Quantized prediction error signal (residuum)
e_{q}	Quantization error
e_i	Error terms (perturbations of a, b, c, f) of the simplified affine model, $i = \{a,b,c,f\}$
e_{ij}	Error terms (perturbations of a_{ij}) of the fully affine model with $i = \{1,2\}, j = \{1,2,3\}$
ϵ	Arbitrarily small error (rate-distortion theory)
f	Frequency (rate-distortion theory)
f	Parameter of the simplified affine model (translation in y-direction)
$\boldsymbol{f}_{i,k}$	Position of the i-th feature in the frame k
f_c	Focal length
$\boldsymbol{g}_{k-1}$	Holds the temporal derivatives of I
$h_{11}, \ldots, h_{33}$	Elements of $\mathtt{H}$
$\mathtt{H}$	Homography matrix of size 3 × 3
H_{G}	Entropy of a memoryless, time-discrete, amplitude-continuous Gaussian source
i, j	Counter variables
$I(\boldsymbol{n})$	Image intensity at the position $\boldsymbol{n}$
$I_k(\boldsymbol{n})$	Image intensities of the frame k
i_{RDC}	Number of iterations for in-loop radial distortion compensation
I_x,I_y	Partial derivatives of I

k	Frame index
k_{ang}	Constant value in the small-angle approximation
κ_1	Radial distortion parameter
$\kappa_{1,l}$	Radial distortion parameter of group of frames with index l
k_{H}	Harris weighting factor
K	Number of code symbols (rate-distortion theory)
$\mathtt{K}$	Camera calibration matrix of size 3×3
l	Counter variable (for groups of frames in in-loop radial distortion compensation)
L	Number of source symbols emitted by source U (rate-distortion theory)
λ_1,λ_2	Eigenvalues of Harris corner matrix $\mathtt{M}$
Λ	Two-dimensional (2D) spatial frequency vector $\Lambda := (\omega_x, \omega_y)$
m, n	Counter variables
$\mathtt{M}$	Harris corner matrix
M_{CRF}	Minimum distance between feature points
n_{RDC}	Number of frames in a group of frames
$\boldsymbol{n} = (x,y)^\top$	Point on the image plane in image coordinates
$\frac{\boldsymbol{n}_\text{s}}{d_\text{s}}$	Surface normal vector, with d_s being the distance between the camera center and the surface
N_x,N_y	Number of sensor elements in x- and y-direction
$N(f)$	Distortion of a single source in rate-distortion theory
$\mathcal{N}(m_\text{G}; v_\text{G})$	Follows a Gaussian distribution with mean m_G and variance v_G
$N_\text{P}(n_\text{G})$	Power of Gaussian noise n_G
n_G	Gaussian noise
n_{mos}	Frame distance (long-term mosaicking)
ω_x, ω_y	Spatial frequencies in x- and y-direction
$\boldsymbol{p} = (x_c,y_c)^\top$	Point on the image plane in sensor coordinates
$\tilde{\boldsymbol{p}} = (x_d,y_d)^\top$	Point on the image plane with lens distortion
$\boldsymbol{p}_k$	Point on the image plane of camera $\boldsymbol{C}_k$
$\hat{\boldsymbol{p}}_k$	Estimate of $\boldsymbol{p}_k$ through affine motion compensation
$\boldsymbol{P} = (X,Y,Z)^\top$	Point in world coordinates
$\tilde{\boldsymbol{P}} = (X_\text{c},Y_\text{c},Z_\text{c})^\top$	Point in camera coordinates

$p_{\Delta X',\Delta Y'}(\Delta x',\Delta y')$	2D probability density function of the displacement estimation error (of the fully affine model)
${}^{\mathrm{simp}}p_{\Delta X'_{\mathrm{s}},\Delta Y'_{\mathrm{s}}}(\Delta x'_{\mathrm{s}},\Delta y'_{\mathrm{s}})$	2D probability density function of the displacement estimation error using a simplified affine model
$p(\cdot)$	Probability density function of $(\cdot)$
$p_{\bullet}(\cdot)$	General form of a probability density function of the random process • with the observations $(\cdot)$
$P(\Lambda)$	Fourier transform of the displacement estimation error
$\boldsymbol{q}(q_1,q_2)^{\top}, q$	Projective components of the homography
r, r_{d}	Radii of $\boldsymbol{p}$ and $\tilde{\boldsymbol{p}}$ to the center of distortion
$r_{11}\ldots r_{33}$	Elements of $\mathtt{R}$
$r_k(\boldsymbol{n})$	Pel-wise motion detection results of the frame k
$R(D)$	Bit rate R as a function of the distortion D (rate-distortion theory)
${}^{\mathrm{simp}}R\left({}^{\mathrm{simp}}D\right)$	Bit rate R as a function of the distortion D using a simplified affine model (rate-distortion theory)
$R_{\mathrm{G}}(D)$	Bit rate R_{G} of a Gaussian source as a function of the distortion D (rate-distortion theory)
R_{ss}	Autocorrelation function of the video signal s
$R_{ss,\mathrm{iso}}$	Isotropic autocorrelation function of the video signal s
$\rho_{ss,x}, \rho_{ss,y}$	Autocorrelation coefficients of the video signal s in x- and y-direction
$\mathtt{R} = \mathtt{R}_{\theta}\mathtt{R}_{\gamma}\mathtt{R}_{\beta}$	Camera orientation matrix of size 3 × 3
s	Video signal
s_{s}	Scaling parameter of the simplified affine model
$s_{\mathrm{w}}, s_{\mathrm{h}}$	Width and height of the camera sensor
s_x, s_y	Width and height of one pel on the image sensor
$\hat{s}$	Predicted signal
s'	Reconstructed video signal
s^*	Preprocessed signal
$\sigma^2_{\Delta x'}, \sigma^2_{\Delta y'}$	Variances of $\Delta x'$ and $\Delta y'$ of the fully affine model
$\sigma^2_{\Delta x'_{\mathrm{s}}}, \sigma^2_{\Delta y'_{\mathrm{s}}}$	Variances of $\Delta x'_{\mathrm{s}}$ and $\Delta y'_{\mathrm{s}}$ of the simplified affine model
$\sigma^2_{e_{ij}}$	Variances of the error terms e_{ij}, $i = \{1{,}2\}$, $j = \{1{,}2{,}3\}$
$\sigma^2_{e_{ij,\mathrm{mod}}}$	Variance of the error terms $e_{ij,\mathrm{mod}}$, $i = \{1{,}2\}$, $j = \{1{,}2{,}3\}$, representing the motion model error

σ_u^2	Variance of the source symbols u
σ_x, σ_y	Standard deviations of x and y
S_{CRF}	Threshold of corner response function
S_{ee}	Power spectral density of the prediction error e
${}^{\mathrm{simp}}S_{ee}$	Power spectral density of the prediction error e using a simplified affine model
$S(f)$	Power spectral density
S_{ss}	Power spectral density of the video signal s
t	Time
$\boldsymbol{t}$	Translation vector component of a homography
Θ	Parameter that generates the function $R(D)$ by taking on all positive real values (rate-distortion theory)
θ	Rotation parameter of the simplified affine model
$\theta_x, \theta_y, \theta_z$	Rotation angles (of the camera)
T_b, T_r	Binarization and erosion thresholds of the noise filter
$u_1, u_2, \ldots, u_L$	Sequence of (unquantized) source symbols (rate-distortion theory)
$\breve{u}$	One specific source symbol (rate-distortion theory)
u,v,$\boldsymbol{u}$, $\boldsymbol{v}$	Arbitrary feature indices and positions
U	Time-discrete, amplitude-continuous source (rate-distortion theory)
$v_1, v_2, \ldots, v_L$	Sequence of (quantized) code symbols (rate-distortion theory)
$\breve{v}$	One specific code symbol (rate-distortion theory)
$\mathtt{W}_x, \mathtt{W}_y, \mathtt{W}_z$	Skew-symmetric matrices induced by rotation around the X-, Y-, and Z-axis
W	Search window
W_{H}	Window in the Harris corner detector
W_s	Bandwidth of signal s (rate-distortion theory)
x, y	Coordinates in x- and y-direction (in pel)
$\hat{x}$, $\hat{y}$	Perturbed x- and y-value
$\hat{x}'$, $\hat{y}'$	Perturbed x'- and y'-value
$\hat{x}_{\mathrm{s}}'$, $\hat{y}_{\mathrm{s}}'$	Perturbed x_s'- and y_s'-coordinates of the simplified affine model
$\hat{x}_{\mathrm{s}}$, $\hat{y}_{\mathrm{s}}$	Perturbed x- and y-value of the simplified affine model
x', y'	Projected/transformed x- and y-coordinates
x_{s}', y_{s}'	Projected/transformed x- and y-coordinates of the simplified affine model

Abstract

Motion-compensated prediction is used in video coding standards like *High Efficiency Video Coding* (HEVC) as one key element of data compression. Commonly, a purely translational motion model is employed. In order to also cover non-translational motion types like rotation or scaling (zoom) contained in aerial video sequences such as captured from unmanned aerial vehicles, an affine motion model can be applied.

In this work, a model for affine motion-compensated prediction in video coding is derived by extending a model of purely translational motion-compensated prediction. Using the rate-distortion theory and the displacement estimation error caused by inaccurate affine motion parameter estimation, the minimum required bit rate for encoding the prediction error is determined. In this model, the affine transformation parameters are assumed to be affected by statistically independent estimation errors, which all follow a zero-mean Gaussian distributed probability density function (pdf). The joint pdf of the estimation errors is derived and transformed into the pdf of the location-dependent displacement estimation error in the image. The latter is related to the minimum required bit rate for encoding the prediction error. Similar to the derivations of the fully affine motion model, a four-parameter simplified affine model is investigated. It is of particular interest since such a model is considered for the upcoming video coding standard *Versatile Video Coding* (VVC) succeeding HEVC. As the simplified affine motion model is able to describe most motions contained in aerial surveillance videos, its application in video coding is justified. Both models provide valuable information about the minimum bit rate for encoding the prediction error as a function of affine estimation accuracies.

Although the bit rate in motion-compensated prediction can be considerably reduced by using a motion model which is able to describe motion types occurring in the scene, the total video bit rate may remain quite high, depending on the motion estimation accuracy. Thus, at the example of aerial surveillance sequences, a codec independent region of interest- (ROI-) based aerial video coding system is proposed that exploits the characteristic of such sequences. Assuming the captured scene to be planar, one frame can be projected into another using global motion compensation. Consequently, only new emerging areas have to be encoded. At the decoder, all new areas are registered into a so-called mosaic. From this, reconstructed frames are extracted and concatenated as a video sequence. To also preserve moving objects in the reconstructed video, local motion is detected and encoded in addition to the new areas. The proposed general ROI coding system was evaluated for very low and low bit rates between 100 and 5000 kbit/s for aerial sequences of HD resolution. It is able to reduce the bit rate by 90 % compared to common HEVC coding of similar quality. Subjective tests confirm that the overall image quality of the ROI coding system exceeds that of a common HEVC encoder especially at very low bit rates below 1 Mbit/s.

To prevent discontinuities introduced by inaccurate global motion estimation—as may be caused by radial lens distortion—a fully automatic in-loop radial distortion compensation is proposed. For this purpose, an unknown radial distortion compensation parameter that is constant for a group of frames is jointly estimated with the global motion. This parameter is optimized to minimize the distortions of the projections of frames in the mosaic. By this approach, the global motion compensation was improved by 0.27 dB and discontinuities in the frames extracted from the mosaic are diminished. As an additional benefit, the generation of long-term mosaics becomes possible, constructed by more than 1500 aerial frames with unknown radial lens distortion and without any calibration or manual lens distortion compensation.

Keywords: video coding, affine motion-compensated prediction (MCP), simplified affine motion-compensated prediction, rate-distortion theory, aerial surveillance, global motion compensation (GMC), region of interest- (ROI-) based aerial video coding, moving object detection, long-term mosaicking, radial distortion compensation

Kurzfassung

Bewegungskompensierte Prädiktion wird in Videocodierstandards wie *High Efficiency Video Coding* (HEVC) als ein Schlüsselelement zur Datenkompression verwendet. Typischerweise kommt dabei ein rein translatorisches Bewegungsmodell zum Einsatz. Um auch nicht-translatorische Bewegungen wie Rotation oder Skalierung (Zoom) beschreiben zu können, welche beispielsweise in von unbemannten Luftfahrzeugen aufgezeichneten Luftbildvideosequenzen enthalten sind, kann ein affines Bewegungsmodell verwendet werden.

In dieser Arbeit wird aufbauend auf einem rein translatorischen Bewegungsmodell ein Modell für affine bewegungskompensierte Prädiktion hergeleitet. Unter Verwendung der Raten-Verzerrungs-Theorie und des Verschiebungsschätzfehlers, welcher aus einer inexakten affinen Bewegungsschätzung resultiert, wird die minimal erforderliche Bitrate zur Codierung des Prädiktionsfehlers hergeleitet. Für die Modellierung wird angenommen, dass die sechs Parameter einer affinen Transformation durch statistisch unabhängige Schätzfehler gestört sind. Für jeden dieser Schätzfehler wird angenommen, dass die Wahrscheinlichkeitsdichteverteilung einer mittelwertfreien Gaußverteilung entspricht. Aus der Verbundwahrscheinlichkeitsdichte der Schätzfehler wird die Wahrscheinlichkeitsdichte des ortsabhängigen Verschiebungsschätzfehlers im Bild berechnet. Letztere wird schließlich zu der minimalen Bitrate in Beziehung gesetzt, welche für die Codierung des Prädiktionsfehlers benötigt wird. Analog zur obigen Ableitung des Modells für das voll-affine Bewegungsmodell wird ein vereinfachtes affines Bewegungsmodell mit vier Freiheitsgraden untersucht. Ein solches Modell wird derzeit auch im Rahmen der Standardisierung des HEVC-Nachfolgestandards *Versatile Video Coding* (VVC) evaluiert. Da das vereinfachte Modell bereits die meisten in Luftbildvideosequenzen vorkommenden Bewegungen abbilden kann, ist der Einsatz des vereinfachten affinen Modells in der Videocodierung gerechtfertigt. Beide Modelle liefern wertvolle Informationen über die minimal benötigte Bitrate zur Codierung des Prädiktionsfehlers in Abhängigkeit von der affinen Schätzgenauigkeit.

Zwar kann die Bitrate mittels bewegungskompensierter Prädiktion durch Wahl eines geeigneten Bewegungsmodells und akkurater affiner Bewegungsschätzung stark reduziert werden, die verbleibende Gesamtbitrate kann allerdings dennoch relativ hoch sein. Deshalb wird am Beispiel von Luftbildvideosequenzen ein *Regionen-von-Interesse-* (ROI-) basiertes Codiersystem vorgeschlagen, welches spezielle Eigenschaften solcher Sequenzen ausnutzt. Unter der Annahme, dass eine aufgenommene Szene planar ist, kann ein Bild durch globale Bewegungskompensation in ein anderes projiziert werden. Deshalb müssen vom aktuellen Bild prinzipiell nur noch neu im Bild erscheinende Bereiche codiert werden. Am Decoder werden alle neuen Bildbereiche in einem gemeinsamen Mosaikbild registriert, aus dem schließlich die Einzelbilder der Videosequenz rekonstruiert werden können. Um auch lokale Bewegungen abzubilden, werden bewegte Objekte detektiert und zusätzlich zu neuen Bildbereichen als ROI codiert. Die Leistungsfähigkeit des ROI-Codiersystems wurde insbesondere für sehr niedrige und niedrige Bitraten von 100 bis 5000 kbit/s für Bilder in HD-Auflösung evaluiert. Im Vergleich zu einer gewöhnlichen HEVC-Codierung kann die Bitrate um 90 % reduziert werden. Durch subjektive Tests wurde bestätigt, dass das ROI-Codiersystem insbesondere für sehr niedrige Bitraten von unter 1 Mbit/s deutlich leistungsfähiger in Bezug auf Detailauflösung und Gesamteindruck ist als ein herkömmliches HEVC-Referenzsystem.

Um Diskontinuitäten in den rekonstruierten Videobildern zu vermeiden, die durch eine durch Linsenverzeichnungen induzierte ungenaue globale Bewegungsschätzung entstehen können, wird eine automatische Radialverzeichnungskorrektur vorgeschlagen. Dabei wird ein unbekannter, jedoch über mehrere Bilder konstanter Korrekturparameter gemeinsam mit der globalen Bewegung geschätzt. Dieser Parameter wird derart optimiert, dass die Projektionen der Bilder in das Mosaik möglichst wenig verzerrt werden. Daraus resultiert eine um 0.27 dB verbesserte globale Bewegungskompensation, wodurch weniger Diskontinuitäten in den aus dem Mosaik rekonstruierten Bildern entstehen. Dieses Verfahren ermöglicht zusätzlich die Erstellung von Langzeitmosaiken aus über 1500 Luftbildern mit unbekannter Radialverzeichnung und ohne manuelle Korrektur.

Stichwörter: Videocodierung, affine bewegungskompensierte Prädiktion, vereinfachte affine bewegungskompensierte Prädiktion, Raten-Verzerrungs-Theorie, Luftbildüberwachung, globale Bewegungskompensation, Regionen-von-Interesse-(ROI-) basierte Luftbildcodierung, Bewegtobjektdetektion, Langzeitmosaikerstellung, Radialverzeichnungskorrektur