Very Low Bit-Rate Video Coding Using 3-D Models

Der Technischen Fakultät der
Friedrich-Alexander-Universität Erlangen-Nürnberg
zur Erlangung des Grades

Doktor-Ingenieur

vorgelegt von

Peter Eisert

Erlangen, 2000

Als Dissertation genehmigt von
der Technischen Fakultät der
Friedrich-Alexander-Universität
Erlangen-Nürnberg

Tag der Einreichung:	09. Oktober 2000
Tag der Promotion:	21. November 2000
Dekan:	Prof. Dr. H. Meerkamm
Berichterstatter:	Prof. Dr.-Ing. B. Girod
	Prof. Dr. G. Greiner

Berichte aus der Kommunikations- und Informationstechnik

Band 20

Peter Eisert

Very Low Bit-Rate Video Coding Using 3-D Models

D 29 (Diss. Universität Erlangen-Nürnberg)

Shaker Verlag
Aachen 2001

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Eisert, Peter:
Very Low Bit-Rate Video Coding Using 3-D Models /
Peter Eisert. Aachen : Shaker, 2001
(Berichte aus der Kommunikations- und Informationstechnik ; Bd. 20)
Zugl.: Erlangen-Nürnberg, Univ., Diss., 2000
ISBN 3-8265-8308-6

Printed in Germany.

ISBN 3-8265-8308-6
ISSN 1432-489X

Shaker Verlag GmbH • P.O. BOX 1290 • D-52013 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • eMail: info@shaker.de

Acknowledgments

I would like to thank my supervisor, Prof. Dr.–Ing. Bernd Girod, who gave me the opportunity to join his group and to work in a stimulative and supportive atmosphere. I would also like to thank Prof. Dr. Günther Greiner for his interest in my work and for reviewing my thesis.

During my doctoral studies, I had the pleasure to interact with a large number of excellent researchers and skilled engineers. I am very thankful to my colleagues and friends Joachim Eggers, Niko Färber, Markus Flierl, Frank Hartung, Achim Hummel, Marcus Magnor, Gerhard Runze, Marion Schabert, Wolfgang Sörgel, Eckehard Steinbach, Alex Stenger, Klaus Stuhlmüller, Jonathan Su, Matthias Teschner, Lutz Trautmann, and Thomas Wiegand for the many stimulating discussions, joint work, and proof-reading.

Furthermore, I would like to thank the many others around the Telecommunications Laboratory in Erlangen who made my stay here so enjoyable. I am especially grateful to Ursula Arnold for her invaluable administrative support.

Last but not least, I would like to thank my friends and my family for their valuable suggestions and support.

Contents

Notation

a	aspect ratio
$\bar{a}$	average aspect ratio
$\mathbf{A}_{rb}$	system matrix for rigid body motion estimation
$\mathbf{A}_{rb,i}$	rigid body system matrix for frame i
$\mathbf{A}_{int}$	system matrix for internal camera parameter estimation
$\mathbf{A}_{int,i}$	internal camera parameter estimation system matrix for frame i
$\mathbf{A}_{cal}$	system matrix for camera calibration
$\mathbf{A}_{fb}$	system matrix for flexible body motion estimation
$\mathbf{b}$	vector of intensity differences between two frames
$\mathbf{c}_i$	i-th B-spline control point
$\hat{\mathbf{c}}_i$	estimate of the i-th B-spline control point position
c_{amb}	relative intensity of ambient light
c^R_{amb}, c^G_{amb}, c^B_{amb}	RGB components of the relative ambient light intensity
c^1_{amb}, c^2_{amb}, c^3_{amb}	transformed relative intensity values of ambient light
c_{dir}	relative intensity of directional light
c^R_{dir}, c^G_{dir}, c^B_{dir}	RGB components of the relative directional light intensity
c^1_{dir}, c^2_{dir}, c^3_{dir}	transformed relative intensity values of directional light
$\mathbf{c}_{DCT}$	DCT coefficients
$\mathbf{C}$	system matrix for illumination estimation
C_B, C_R	chrominance components of the video signal
C_x, C_y	displacement of the optical axis from the center of the CCD sensor
$\mathbf{d} = [d_x \;\; d_y]^T$	2-D image displacement vector
$\hat{\mathbf{d}}$	estimated 2-D image displacement vector
D	distortion
D_{lens}	diameter of camera lens
D_{DFD}	distortion after motion compensation
D_{REC}	distortion of reconstructed frame
E	illuminance
f	focal length
Δf	estimated focal length changes
f_x, f_y	scaled focal length
$\hat{f}_x$	estimate of f_x
$\mathbf{F}$, $\mathbf{F}_\Delta$	matrices specifying inequality constraints for the LSI
$\mathbf{FAP}$	vector of K facial animation parameters
FAP^k	k-th facial animation parameter
$\mathbf{FAP}_1$, $\mathbf{FAP}_2$	facial animation parameters of previous and current frame, respectively
$\widehat{\mathbf{FAP}}_2$	facial animation parameter estimate for the current frame

$\mathbf{\Delta FAP}$	estimated facial animation parameter changes
FAP^k_{min}	minimum value for FAP k
FAP^k_{max}	maximum value for FAP k
ΔFAP^k_{min}	minimum changes for FAP k
ΔFAP^k_{max}	maximum changes for FAP k
$\mathbf{g}$, $\mathbf{g}_\Delta$	vectors specifying inequality constraints for the LSI
h	macroblock header
i, j, k	indices
I	γ-predistorted pixel intensity
I^R,I^G,I^B	RGB components of the γ-predistorted intensity
I_1, I_2	intensities of the previous and current frame, respectively
$\hat{I}_2$	estimated intensities of the current (model) frame
$\bar{I}$	average pixel intensity $\bar{I} = \frac{(I_2+\hat{I}_2)}{2}$
I'	linear pixel intensity
I'^R,I'^G,I'^B	RGB components of the linear intensity
I'_{tex}	linear pixel intensity of the texture map
$I_{tex,eye}$	intensity of the texture map in the eye region
$I_{tex,shade}$	modulating texture map in the eye region
K	number of FAPs
$\mathbf{l} = [l_x \;\; l_y \;\; l_z]^T$	illuminant direction
L	luminance
$\mathbf{m}^k_i$	head model deformation vector for control point i corresponding to the k-th FAP
$\hat{\mathbf{m}}^k_i$	motion-compensated vector $\mathbf{m}^k_i$
$\mathbf{n} = [n_x \;\; n_y \;\; n_z]^T$	surface normal
N	number of frames
N_x, N_y	image width and height in pixels
$N_{j,i}$	triangular B-spline basis functions
$\mathbf{o}^k$	center of rotation for the k-th FAP
$\hat{\mathbf{o}}^k$	estimate of $\mathbf{o}^k$
p, q	surface gradients
q	image geometry $\frac{f_y}{f_x}$
$\hat{q}$	estimate of image geometry
Δq	estimated changes in q
Q	DCT quantizer parameter
r	radius
$R(\mathbf{n})$, $R(p,q)$	reflectance
$\mathbf{R}$	rotation matrix
$\mathbf{R}_1$, $\mathbf{R}_2$	rotation matrix for the previous and the current frame, respectively

$\mathbf{R}_{12}$	relative rotation matrix between the previous and the current frame
$\hat{\mathbf{R}}_{12}$	estimate of the relative rotation matrix between the previous and the current frame
$\Delta\mathbf{R}$	matrix of estimated rotation parameters
$\mathbf{R}_{FAP^k}$	rotation matrix for the k-th FAP
R_x, R_y, R_z	Euler angles
R_{motion}	rate for motion vectors
R_{REC}	bit-rate for macroblock reconstruction
s_k	rotation angle scaling factor for the k-th FAP
s_x, s_y	scaling factors relating frame buffer coordinates with image coordinates
$\mathbf{t} = [t_x \;\; t_y \;\; t_z]^T$	translation vector
$\mathbf{t}_1$, $\mathbf{t}_2$	translation vector for the previous and the current frame, respectively
$\mathbf{t}_{12}$	relative translation vector between the previous and the current frame
$\hat{\mathbf{t}}_{12}$	estimate of the relative translation vector between the previous and the current frame
$\Delta\mathbf{t}$	estimated translation parameters
t'_z	estimated and corrected z-component of translation
$t_{z,o}$	original correct z-component used in calibration experiments
$\mathbf{t}_x$, $\mathbf{t}_y$, $\mathbf{t}_z$	deformation vector for 3-D object points
$\mathbf{T}$	matrix containing the deformation vectors $\mathbf{t}_x$, $\mathbf{t}_y$, and $\mathbf{t}_z$
$\mathbf{v}_j$	j-th vertex
$\mathbf{x} = [x \;\; y \;\; z]^T$	camera coordinates of a point in the 3-D scene
$\mathbf{x}_1$, $\mathbf{x}_2$	3-D object point coordinates corresponding to previous and the current frame, respectively
$\hat{\mathbf{x}}_2$	estimated 3-D object point coordinates for the current frame
$\mathbf{x}_c$	object center
$\mathbf{x}_w = [x_w \;\; y_w \;\; z_w]^T$	world coordinates of a point in the 3-D scene
$\mathbf{x}_0 = [x_0 \;\; y_0 \;\; z_0]^T$	local object coordinates
$\mathbf{x}_l = [x_l \;\; y_l \;\; z_l]^T$	local light source coordinates
$\mathbf{X} = [X \;\; Y]^T$	frame buffer coordinates centered around the optical axis
$\mathbf{X}_1$, $\mathbf{X}_2$	image coordinates of the previous and the current frame
$\mathbf{X}' = [X' \;\; Y']^T$	undistorted frame buffer coordinates
$\mathbf{X}_f = [X_f \;\; Y_f]^T$	frame buffer coordinates
$\mathbf{X}_i = [X_i \;\; Y_i]^T$	undistorted 2-D coordinate in the image plane
$\mathbf{X}_{dist}$	distorted 2-D coordinate in the image plane
$\mathbf{X}_d = [X_d \;\; Y_d]^T$	distorted 2-D pixel coordinates
$\mathbf{X}_0 = [X_0 \;\; Y_0]^T$	location of optical axis in pixel coordinates

$\mathcal{I}$	index list containing the three vertex indices of a triangle
$\mathcal{I}_j$	index list containing the control point indices that influence vertex j
α	angle between the optical axis and the viewing direction
Δ	picture reference parameter
γ	gamma value
κ_i	parameters describing the radial lens distortion of the camera
$\hat{\kappa}_i$	estimate of radial lens distortion parameters
$\Delta\kappa_i$	estimated radial lens distortion parameters
$\lambda_i(\mathbf{x})$	barycentric coordinate
λ_{motion}, λ_{mode}	Lagrange multipliers
φ_{height}	height angle of the camera
$\bar{\varphi}_{height}$	average height angle of the camera
φ'_{height}	estimated height angle after correction
φ'_o	original correct height angle used in calibration experiments
$\mathbf{\Omega} = [\Omega_x \ \ \Omega_y \ \ \Omega_z]^T$	axis of rotation with $\lvert\mathbf{\Omega}\rvert = 1$
$\mathbf{\Omega}^k$	axis of rotation for the k-th FAP
Φ	threshold for angle between illuminant direction and surface normal
σ	standard deviation
Θ	rotation angle, specifying rotation around $\mathbf{\Omega}$
Θ^k	rotation angle corresponding to the k-th FAP

Abbreviations and Acronyms

2-D	Two-dimensional
3-D	Three-dimensional
AU	Action Unit, facial expression parameter of the FACS system
bps	Bits per second
BRDF	Bidirectional Reflection Distribution Function
CCD	Charge Coupled Device
CIE	Commission Internationale de l'Eclairage
CIF	Common Intermediate Format (352 × 288 pixels)
CRT	Cathode Ray Tube
DCT	Discrete Cosine Transform
DFD	Displaced Frame Difference
DOF	Degrees of Freedom
FACS	Facial Action Coding System
FAP	Facial Animation Parameter
FDP	Facial Definition Parameter
fps	Frames per second
GOB	Group of Blocks
Hz	Hertz (one cycle per second)
I-frame	Intra-Coded Frame
IDCT	Inverse Discrete Cosine Transform
ISDN	Integrated Services Digital Network
ITU	International Telecommunications Union
ITU-T	International Telecommunications Union, Telecommunications Standardization Sector
JPEG	Joint Photographic Experts Group
kbps	kilo bits per second
KLT	Karhunen-Loève Transform
LPC	Linear Predictive Coding
LS	Least-Squares
LSI	Least-Squares with Inequality Constraints
MAC	Model-Aided Coding
Mbps	Mega bits per second
MBC	Model-Based Coding
MCP	Motion Compensated Prediction
MSE	Mean Squared Error
MPEG	Moving Pictures Expert Group
NNLS	Non-Negative Least Squares

NURBS	Non-Uniform Rational B-Splines
OBASC	Object-Based Analysis-Synthesis Coder
P-frame	Predicted Frame
PAL	Phase Alternating Line
PDA	Personal Digital Assistant
PSNR	Peak Signal-to-Noise Ratio
PSTN	Public Switched Telephone Network
QCIF	Quarter Common Intermediate Format (176×144 pixels)
RD	Rate-Distortion
SAD	Sum of Absolute Differences
SfM	Structure-from-Motion
SNHC	Synthetic and Natural Hybrid Coding, part of MPEG-4 standard
SSD	Sum of Squared Differences
TMN-10	Test model near term, version 10, of the H.263 standard
TV	Television
VLC	Variable Length Code
VOP	MPEG-4 Video Object Plane