Analyse–Synthese–Codierung basierend auf dem Modell bewegter dreidimensionaler, gegliederter Objekte

Von dem Fachbereich Elektrotechnik
und Informationstechnik der Universität Hannover

zur Erlangung des akademischen Grades

Doktor–Ingenieur

genehmigte

Dissertation

von

Lic.–Ing. Geovanni Martínez–Castillo

geboren am 15. September 1965 in San José

1998

Referent: Prof. Dr.–Ing. H. G. Musmann
Korreferent: Prof. Dr.–Ing. K. Jobmann
Tag der Promotion: 13.02.1998

Gedruckt mit Unterstützung des Deutschen Akademischen Austauschdienstes

Kurzfassung

Analyse–Synthese–Codierung basierend auf dem Modell bewegter dreidimensionaler, gegliederter Objekte

In der vorliegenden Arbeit wird ein objektbasierter Analyse–Synthese–Coder für die Bewegtbildcodierung bei niedrigen Datenraten entwickelt, der auf dem Quellenmodell bewegter dreidimensionaler, gegliederter Objekte basiert. Dieses Quellenmodell beschreibt die realen Objekte durch Modellobjekte, die durch Form–, Bewegungs– und Texturparameter definiert sind. Die Modellobjekte können gegliedert sein, d.h. sie bestehen i.a. aus mehreren, miteinander durch Gelenke gekoppelten, starren Objektkomponenten.

Bildanalysealgorithmen schätzen die Parameter jedes Modellobjektes. Ausgehend von der geschätzten Objektsilhouette wird zunächst eine dreidimensionale starre Form berechnet. Sie wird dann in Objektkomponenten gegliedert, indem kleine Oberflächenelemente mit ähnlichen Bewegungsparametern zu einer Objektkomponente zusammengefaßt werden. Hierfür wurde ein robuster Bewegungsschätzer entwickelt, der die zur Schätzung verwendeten Meßwerte statistisch modelliert und Ausreißer in den Meßwerten erkennt. Die Form der Objektkomponenten wird im Verlauf der Bildsequenz ständig an den jeweiligen Bildinhalt angepaßt. Für die Schätzung der Position eines Gelenkes wird ein Verfahren auf der Basis einer Regression vorgestellt, das die Einschränkungen der Relativbewegung der durch das Gelenk gekoppelten Objektkomponenten ausnutzt. Für eine genauere Schätzung der Bewegungsparameter der Objektkomponenten werden die Gelenke in die Schätzung miteinbezogen. Die geschätzten Parameter eines jeden Modellobjektes werden nach ihrer objektbasierten Codierung übertragen. Am Empfänger wird aus den decodierten Parametern dann ein Bild durch eine objektbasierte Bildsynthese rekonstruiert.

Ein Vergleich mit aus der Literatur bekannten dreidimensionalen Quellenmodellen ergab, daß das neue Quellenmodell bei gleicher Bildqualität eine Reduktion der Datenrate um bis zu 20% im Vergleich zu dem Quellenmodell bewegter dreidimensionaler, starrer Objekte und um bis zu 10% im Vergleich zu dem Quellenmodell bewegter dreidimensionaler, flexibler Objekte erreicht. Ferner wird gezeigt, daß sich durch die objektbasierte Codierung aufgrund der Übertragung der Objektform die störenden Codierfehler eines blockbasierten Hybridcoders, wie Moskito– und Blockartefakte, vermeiden lassen.

Schlagwörter: Form– und Bewegungsschätzung, gegliederte Objekte, Bewegtbildcodierung

Abstract

Analysis/synthesis coding based on the model of moving three–dimensional articulated Objects

In the present study, an object–based analysis/synthesis coder is elaborated for coding moving images at low date rates. This coder is based on the source model of moving three–dimensional articulated objects. The model describes the real objects by means of model objects defined by shape, motion and texture parameters. The model objects may be articulated, i.e. they consist of several rigid object components linked to each other by joints.

Image analysis algorithms estimate the parameters of each model object. Using the estimated object silhouette, a three–dimensional rigid shape is first calculated. This is then articulated in object components by clustering small surface elements with similar motion parameters into an object component. For this purpose, a robust motion estimator has been developed which models the measured values statistically and detects outliers in them. The object component shape is continuously adapted to the corresponding image content in the course of the image sequence. For estimating the position of a joint, a linear regression–based procedure is introduced which makes use of relative motion constraints of the object components connected to the joints. For a more precise estimation of the motion parameters of an object component, joints are also taken into account in motion estimation. The estimated parameters of each model object are coded using object–based parameter coding. An image is then synthesized from the decoded parameters in the receiver by means of object–based image synthesis.

A comparison with previously known three–dimensional source models in professional literature showed that at the same image quality this new source model achieves a date rate reduction of up to 20% in comparison with the source model of moving three–dimensional rigid objects and by up 10% in comparison with the source model of moving three–dimensional flexible objects. It has been additionally demonstrated that object–based coding makes it possible to avoid the disturbing coding errors of an block–based hybrid coder like mosquito and blocking artifacts. This is achieved by transmitting the shape of the objects.

Keywords: shape and motion estimation, articulated objects, video coding

Vorwort

Die vorliegende Dissertation entstand aus meiner wissenschaftlichen Arbeit am Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung der Universität Hannover.

Dem Direktor des Instituts, Herrn Professor Dr.–Ing. H. G. Musmann, danke ich für die Anregung zum Thema der Arbeit. Er sorgte für hervorragende Arbeitsmöglichkeiten. Ihm gilt mein besonderer Dank für die Betreuung der Arbeit und die Übernahme des Hauptreferats.

Herrn Professor Dr.–Ing. K. Jobmann danke ich für die Übernahme des Koreferats.

All meinen ehemaligen Kollegen und Diplomanden, die durch Diskussionen und Anregungen zum Gelingen der Arbeit beigetragen haben, danke ich herzlich. Insbesondere gilt mein Dank Herrn Dipl.–Ing. M. Kampmann, Herrn Dipl.–Ing. J. Stauder, Herrn Dipl.–Ing. P. Gerken, Herrn Dipl.–Ing. M. Kirscht, Herrn Dr.–Ing. J. Ostermann, Herrn M.Sc. L. Zhang und Herrn Dipl.–Inform. R. Mech für die stets entgegengebrachte Hilfsbereitschaft und für die vielen konstruktiven Diskussionen.

Für die Hilfe bei der Fertigstellung des Manuskripts möchte ich insbesondere Herrn Dipl.–Ing. M. Kampmann danken.

Für die tatkräftige Unterstützung durch die Betreuung der Rechner danke ich Herrn Dipl.–Ing. O. Dehning, Herrn Dipl.–Ing. W. Wölker, Herrn Dipl.–Ing. M. Pahl, Herrn Dipl.–Ing. H. Münkel, Herrn Dipl.–Ing. M. Wollborn und Herrn Dipl.–Ing. L. Falkenhagen.

Dem Deutschen Akademischen Austauschdienst (DAAD) danke ich für die finanzielle Unterstützung während meiner wissenschaftlichen Arbeit an der Universität Hannover.

Zum Schluß möchte ich meinen Eltern und Geschwistern für die von ihnen aufgebrachte große Unterstützung danken.

Inhalt

1. Einleitung 1

2. Objektbasierte Analyse–Synthese–Codierung (OBASC) 7

3. Quellenmodelle für OBASC 13

3.1 Kameramodell 13

3.2 Beleuchtungsmodell 16

3.3 Szenenmodell 16

3.4 Objektmodell 17

3.4.1 Objektmodell bewegter dreidimensionaler, starrer Objekte 18

3.4.2 Objektmodell bewegter dreidimensionaler, flexibler Objekte 22

3.4.3 Objektmodell bewegter dreidimensionaler, gegliederter Objekte . 23

4. Bildanalyse für bewegte dreidimensionale, gegliederte Objekte 31

4.1 Konzept der objektbasierten Bildanalyse 31

4.2 Schätzung der dreidimensionalen Form gegliederter Objekte 35

4.2.1 Forminitialisierung 36

4.2.2 Objektgliederung 39

4.2.3 Formadaption 50

4.2.4 Automatische Schätzung der Position von Gelenken 55

4.3 Schätzung der dreidimensionalen Bewegung gegliederter Objekte 65

4.3.1 Modellierung des Schätzfehlers der Form für die Bewegungsschätzung 65

4.3.2 Verfahren zur Schätzung der dreidimensionalen Bewegung eines relativ kleinen starren Oberflächenelementes 70

4.3.3 Verfahren zur Schätzung der Parameter der dreidimensionalen Bewegung gegliederter Objekte 80

5. Bildsynthese für bewegte dreidimensionale, gegliederte Objekte . . . 91

6. Parametercodierung für bewegte dreidimensionale, gegliederte Objekte 93

7. Experimentelle Untersuchungen 99

7.1 Beschreibung der verwendeten Testsequenzen 99

7.2 Gütekriterium zur Bewertung der Bildqualität in einem Analyse–Synthese Coder 100

7.3 Vergleich der Bildqualität der Coder mit dreidimensionalen Quellenmodellen 100

8. Zusammenfassung 109

9. Literatur 115

Abkürzungen und Formelzeichen

A3D	Quellenmodell bewegter dreidimensionaler, gegliederter Objekte
b	aus der Objektsilhouette bestimmte Objektbreite für die Erzeugung eines Dreiecksnetzes
bl_k, bl_{k+1}	eine der beiden Chrominanzkomponenten des rauschbehafteten Videosignals $i(\boldsymbol{x})$ in den Bildern k bzw. $k+1$ zu den diskreten Zeitpunkten t_k bzw. t_{k+1}
B_B	Bildbreite (beim Bildformat CIF: 352 pel)
$\boldsymbol{B}'$	codierte Bewegungsparameter der MC–Objekte
$\hat{\boldsymbol{B}}_{j,i}$	Schätzwerte der Bewegungsparameter nach j Beobachtungspunkten bei der i–ten Iteration
$\boldsymbol{B}_m$	Bewegungsparameter des Objektes m
$\boldsymbol{B}'_m$	codierte Bewegungsparameter des Objektes m
$\boldsymbol{B}_{pos}$	$= \{\boldsymbol{G}, [\boldsymbol{R}]_0\}$: Ortsvektor $\boldsymbol{G}$ des Ursprunges eines Koordinatensystems und Verdrehung $[\boldsymbol{R}]_0$ dieses Koordinatensystems bezüglich des Weltkoordinatensystems
$\boldsymbol{BL}_m$	$= \{BL_m\}$: eine Komponente der Chrominanzwerte von $\boldsymbol{C}_m$
$c(\xi)^+$, $c(\xi)^-$	Funktionen, die die zur Kennzeichnung der Bezugs– bzw. Relativobjektkomponente eines Kugelgelenkes ξ verwendete Nummer δ angeben
CCITT	Commité Consultatif International de Téléphonie et Télégraphie
CIF	Common Intermediate Format mit der örtlichen Auflösung von 352x288 Bildpunkten und einer zeitlichen Abtastfrequenz von $f_{st} = 30\text{Hz}$
COST	Coopération europèenne dans le domaine de la Recherche Scientifique et Technique
$\boldsymbol{C}^{FB}$, $\boldsymbol{C}^{MC}$, $\boldsymbol{C}^{MF}$	Texturparameter der freigewordenen Bildbereiche, MC–Objekte bzw. MF–Objekte
$\boldsymbol{C}'$	codierte Texturparameter aller Objekte

$\boldsymbol{C}^{''}$	codierte Texturparameter der freigewordenen Bildbereiche $\boldsymbol{C}^{FB}$ und der MF–Objekte $\boldsymbol{C}^{MF}$
$\boldsymbol{C}_m$	Texturparameter des Objektes m bestehend aus den Luminanz– und Chrominanzwerten $\boldsymbol{L}_m = \{L_m\}$ bzw. $\boldsymbol{BL}_m = \{BL_m\}$ und $\boldsymbol{RL}_m = \{RL_m\}$
$\boldsymbol{C}_m^{'}$	codierte Texturparameter des Objektes m
$[\boldsymbol{C}]_{\Delta \boldsymbol{h}}$	Kovarianzmatrix eines Positionsfehlers $\Delta \boldsymbol{h}$ in der Bildebene
$[\boldsymbol{C}]_{\Delta \boldsymbol{H}}$	Kovarianzmatrix eines räumlichen Positionsfehlers $\Delta \boldsymbol{H}$
$\mathcal{C}^{(\delta)}$, $\mathcal{C}^{(0)}$	Objektkomponente δ bzw. Wurzelobjektkomponente
$\mathcal{C}^{(c(\xi)^+)}$, $\mathcal{C}^{(c(\xi)^-)}$	Bezugs– bzw. Relativobjektkomponente eines Gelenkes ξ
d	Abstand zwischen einer beliebig geformten Linie und ihrer Approximation durch ein Polygon
d_h	Abstand der Höhenlinien bei der Erzeugung eines Dreiecksnetzes
$d^*_{\max} = d^{MC*}_{\max}$	maximal erlaubter Abstand zwischen einer beliebig geformten Linie und ihrer Approximation durch ein Polygon bei MC–Objekten
$d^{*MF}_{\max}$	maximal erlaubter Abstand zwischen der geschätzten Form eines MF–Objetkes und der Formapproximation
dfd	bewegungskompensierte Bild–zu–Bild–Luminanzdifferenz (*d*isplaced *f*rame *d*ifference)
DCT	Diskrete Cosinus Transformation
$\mathfrak{D}_Z$	Z–Komponente der Position des Kontrollpunktes einer verdeckten Objektkomponente, der am nächsten zur Kamerabildebene liegt
$\boldsymbol{D}(\boldsymbol{h}_{k+1})$	$= (D_x(\boldsymbol{h}_{k+1}), D_x(\boldsymbol{h}_{k+1}))^T$: zweidimensionaler Displacementvektor eines Displacementvektorfeldes an der Stellt $\boldsymbol{h}_{k+1}$
$\boldsymbol{DFD}$	Meßvektor, dessen Komponenten die meßbaren bewegungskompensierten Bild–zu–Bild–Luminanzdifferenzen an den Beobachtungspunkten darstellen
$\bar{e}_{\boldsymbol{J}}$	mittlere Abweichung der Position eines Kugelgelenkes
fd	Bild–zu–Bild–Luminanzdifferenz (frame–difference)
f_{st}	zeitliche Abtastfrequenz (Bildwiederholfrequenz)
f_{sx}, f_{sy}	horizontale bzw. vertikale Abtastfrequenz

F	Brennweite der Kamera
F2D	Quellenmodell bewegter zweidimensionaler, flexibler Objekte
F3D	Quellenmodell bewegter dreidimensionaler, flexibler Objekte
$\boldsymbol{F}^{MC}$, $\boldsymbol{F}^{MF}$	Formparameter der MC–Objekte bzw. MF–Objekte
$\boldsymbol{F}'$	codierte Formparameter der MC– und MF–Objekte
$\boldsymbol{F}''$	codierte Formparameter der MF–Objekte und Update–Formparameter der MC–Objekte
$\boldsymbol{F}_m$	Formparameter des Objektes m
$\boldsymbol{F}'_m$	codierte Formparameter des Objektes m
$\boldsymbol{FD}$	Meßvektor, dessen Komponenten die meßbaren Bild–zu–Bild–Luminanzdifferenzen an den Beobachtungspunkten darstellen
$\boldsymbol{g}$	$= (\frac{\delta s}{\delta x}, \frac{\delta s}{\delta y})^T = (g_x, g_y)^T$: örtlicher Bildsignalgradient
$\overline{\boldsymbol{g}}$	$= (\overline{g}_x, \overline{g}_y)^T$: gemittelter örtlicher Bildsignalgradient
$\boldsymbol{G}$	Ortsvektor des Ursprunges eines Koordinatensystems im Weltkoordinatensystem
$\boldsymbol{G}_i$	$\boldsymbol{G}$ nach der i–ten Iteration
$\boldsymbol{G}'$	$\boldsymbol{G}$ nach der Bewegung
h	Objekttiefe für die Erzeugung eines Dreicknetzes
H_B	Bildhöhe (beim Bildformat CIF: 288 pel)
H.261	ITU–T Recomendation H.261: Video codec for audiovisual services at p x 64 kbit/s, März 1993
H.263	ITU–T Recommendation H.263: Video Coding for Narrow Telecommunication Channels at < 64 kbit/s, Dezember 1995
$\boldsymbol{h}$	$= (h_x, h_y)^T$: der Punkt mit dem Ortsvektor $\boldsymbol{H}$ wird mittels der Kamera auf den Punkt mit den Bildkoordinaten $\boldsymbol{h}$ in der Bildebene abgebildet
$\boldsymbol{h}_i$	$\boldsymbol{h}$ bei der i–ten Iteration
$\boldsymbol{h}'$	$\boldsymbol{h}$ nach der Bewegung
$\boldsymbol{H}$	Ortsvektor eines Oberflächenpunktes im Weltkoordinatensystem
$\boldsymbol{H}_i$	$\boldsymbol{H}$ bei der i–ten Iteration
$\boldsymbol{H}'$	$\boldsymbol{H}$ nach der Bewegung

$\boldsymbol{H}''$	$\boldsymbol{H}$ nach zwei aufeinanderfolgenden Bewegungen
$\boldsymbol{H}^*$	Ortsvektor eines Oberflächenpunktes im zugehörigen Koordinatensystem
$\boldsymbol{H}^{**}$	Ortsvektor eines Oberflächenpunktes einer Objektkomponente im Koordinatensystem der korrespondierenden Bezugsobjektkomponente
$\Delta\boldsymbol{h}$	$= (\Delta h_x, \Delta h_y)^T$: Vektor einer Koordinatendifferenz in der Bildebene
$\Delta\boldsymbol{H}$	$= (\Delta H_X, \Delta H_Y, \Delta H_Z)^T$: Vektor einer Ortskoordinatendifferenz im Weltkoordinatensystem
$i_k,\ i_{k+1}$	rauschbehaftetes Videosignal in den Bildern k bzw. $k+1$ zu dem diskreten Zeitpunkten t_k bzw. t_{k+1}, das sich jeweils aus der Luminanzkomponente l_k bzw. l_{k+1} und den beiden Chrominanzkomponenten bl_k bzw. bl_{k+1} und rl_k bzw. rl_{k+1} zusammensetzt
i_k'	synthetisiertes Modellbild im Bild k zu dem diskreten Zeitpunkt t_k, das aus den codierten und übertragenen Bewegungsparametersätzen $\boldsymbol{F}_k'$, $\boldsymbol{B}_k'$ und $\boldsymbol{C}_k'$ synthetisiert wird
ITU–T	Telecommunication Standardisation Sector of the International Telecommunication Union
$\boldsymbol{J}^{(\xi)}$	Ortsvektor des Kugelgelenkes ξ im Weltkoordinatensystem
$\boldsymbol{J}^{(\xi)'}$	$\boldsymbol{J}^{(\xi)}$ nach der Bewegung
$\boldsymbol{J}^{(\xi)*}$	Ortsvektor des Kugelgelenkes ξ im Koordinatensystems der zugehörigen Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$
$\boldsymbol{J}^{(0)}$	Ortsvektor des Ursprunges des Koordinatensystems der Wurzelobjektkomponente $\mathcal{C}^{(0)}$
$\boldsymbol{K}^{(j+1)}$	Korrekturvektor nach $j+1$ Beobachtungspunkten
$[\boldsymbol{K}]$	Abbildungsmatrix eines linearen Kameramodells
$\hat{l}_{A3D}$	Luminanzkomponente des synthetisierten Modellbildes, daß aus Modellobjekten des Quellenmodells A3D berechnet wurde
l_k, l_{k+1}	Luminanzkomponente des rauschbehafteten Videosignals $i(\boldsymbol{x})$ in den Bildern k bzw. $k+1$ zu den diskreten Zeitpunkten t_k bzw. t_{k+1}
l_k'	Luminanzkomponente des synthetisierten Modellbildes i_k'
$\boldsymbol{L}_m$	$= \{L_m\}$: eine Komponente der Luminanzwerte der Texturparameter $\boldsymbol{C}_m$

mb_{k+1}	Binärmaske zur Kennzeichnung der durch Bewegung veränderten Bildbereiche
mc_{k+1}	Ternärmaske zur Kennzeichnung der Objektsilhouetten, des freigewordenen Hintergrunds und des statischen Hintergrunds
n_k	Rauschkomponente des Videosignals i_k mit der Leistung σ_n^2
n_x, n_y	diskrete Bildkoordinaten
MC–Objekt	modellierbares Objekt (englisch: model compliance)
MF–Objekt	nicht modellierbares Objekt (englisch: model failure)
MPEG	Moving Picture Expert Group
NLIN,NPEL	Anzahl von Zeilen bzw. Anzahl von Bildpunkten je Zeile des digitalisierten Bildes
OBASC	Objektbasierte Analyse–Synthese–Codierung
$[\boldsymbol{O}]$	$= (\boldsymbol{o}^{(\vartheta-1)T},...,\boldsymbol{o}^{(0)T})^T$: Systemmatrix eines Gleichungssystems
pel	Maßeinheit zur Angabe von Bild– und Raumkoordinatendifferenzen als Vielfache der Bildpunktabstände auf dem Abtastraster
$p_{\boldsymbol{V}}(\boldsymbol{V})$	Verbundwahrscheinlichkeitsdichte der Luminanzfehler $\boldsymbol{V}$ an unterschiedlichen Orten in der Bildebene
$p_{\boldsymbol{DFD/R}}(\boldsymbol{DFD/R})$	bedingte Wahrscheinlichkeitsdichte der bewegungskompensierten Bild–zu–Bild–Luminanzdifferenzen $\boldsymbol{DFD}$ an den Beobachtungspunkten bei gegebenen Rotationsparametern $\boldsymbol{R}$
$p_{\boldsymbol{FD/B}}(\boldsymbol{FD/B})$	bedingte Wahrscheinlichkeitsdichte der Bild–zu–Bild–Luminanzdifferenzen $\boldsymbol{FD}$ an den Beobachtungspunkten bei gegebenen Parametern $\boldsymbol{B}$ der dreidimensionalen Bewegung.
PCM	Puls Code Modulation
PSNR	Peak Signal to Noise Ratio in dB, das die Leistung σ_e^2 eines Fehlersignals e, bezogen auf die fiktive Leistung 255^2, angibt: $\text{PSNR [dB]} = 10 \cdot \log_{10}(255^2/\sigma_e^2)$
$\boldsymbol{P}$	Ortsvektor eines Kontrollpunktes im Weltkoordinatensystem
$\boldsymbol{P}'$	$\boldsymbol{P}$ nach der Bewegung
$\boldsymbol{P}^*$	Ortsvektor eines Kontrollpunktes im zugehörigen Koordinatensystem
$[\boldsymbol{Q}]$	Kovarianzmatrix des Schätzfehlers der Bewegungsparameter

$[\boldsymbol{Q}^{(j+1)}]$	Kovarianzmatrix $[\boldsymbol{Q}]$ nach $j+1$ Beobachtungspunkten
rl_k, rl_{k+1}	eine der beiden Chrominanzkomponenten des rauschbehafteten Videosignals $i(\boldsymbol{x})$ in den Bildern k bzw. $k+1$ zu den diskreten Zeitpunkten t_k bzw. t_{k+1}
R, S, T	Koordinaten im Koordinatensystem eines starren Objektes
R	Datenrate für die Codierung eines Bildes
R_B, R_F, R_C	Datenrate für die Codierung der Bewegungs–, Form– bzw. Texturparameter eines Bildes
R3D	Quellenmodell bewegter dreidimensionaler, starrer Objekte
$[\boldsymbol{R}]$	Rotationsmatrix, die ein Koordinatensystem um die Winkel R_X, R_Y und R_Z in der angegebenen Reihenfolge dreht
$\boldsymbol{R}$	Rotationsvektor mit Rotationswinkeln R_X, R_Y und R_Z, die ein Koordinatensystem in der angegebenen Reihenfolge drehen
$\hat{\boldsymbol{R}}_{j,i}$	Schätzwert eines Rotationsvektors nach j Beobachtungspunkten bei der i–ten Iteration
$[\boldsymbol{R}]_0$	Rotationsmatrix, die die Verdrehung eines Koordinatensystems bezüglich des Weltkoordinatensystems durch drei aufeinanderfolgende Drehungen mit den Winkeln $\Re_X$, $\Re_Y$ und $\Re_Z$ beschreibt
$(\Re_X, \Re_Y, \Re_Z)^T$	Vektor mit Rotationswinkeln $\Re_X$, $\Re_Y$ und $\Re_Z$, die die Verdrehung eines Koordinatensystems bezüglich des Weltkoordinatensystems beschreiben. Das Koordinatensystem wird um diese Winkel in der angegebenen Reihenfolge gedreht
$[\boldsymbol{R}_l]$	linearisierte Rotationsmatrix
$[\boldsymbol{R}^{(u)}_{\mathcal{J}-1,i}]$	Bei der Iteration i geschätzter Update der Rotationsparameter
$\boldsymbol{RL}_m$	$= \{RL_m\}$: eine Komponente der Chrominanzwerte der Texturparameter $\boldsymbol{C}_m$
s_k	rauschfreie Luminanzkomponente im Bild k
s_r, s_s	Parameter der lokalen Verschiebung eines Kontrollpunktes $\boldsymbol{P}$
$\Delta\boldsymbol{s}$	$=\Delta\boldsymbol{s}_n+\Delta\boldsymbol{s}_f$: gesamter Luminanzfehler am Ort $\boldsymbol{h}$
$\Delta\boldsymbol{s}_n$, $\Delta\boldsymbol{s}_f$	Luminanzfehler aufgrund des Kamerarauschens bzw. des Schätzfehlers der Form am Ort $\boldsymbol{h}$ in der Bildebene

t_k	diskreter Zeitpunkt k
$T_t = \frac{1}{f_{st}}$	zeitliche Abtastintervallbreite (zeitl. Abstand zwischen aufeinanderfolgenden Bildern einer Fernsehbildsequenz)
T_x , T_y	$= f_{sx}^{-1}, f_{sy}^{-1}$: horizontale bzw. vertikale Abtastintervallbreite
TMN5	blockbasierter Hybridcoder gemäß dem Standard H.263
$\boldsymbol{T}$	$= (T_X, T_Y, T_Z)^T$: Translationsvektor, mit dem ein Koordinatensystem verschoben wird
$\boldsymbol{T}^{(u)}_{\mathcal{J}-1,i}$	Bei der i–ten Iteration geschätzter Update der Translationsparameter
$[\boldsymbol{U}]$	Kovarianzmatrix von $\boldsymbol{V}$. $\lvert[\boldsymbol{U}]\rvert$ stellt ihre Determinante dar
$\boldsymbol{W}^{(j)}$	$= \{\boldsymbol{H}^{(j)}, L^{(j)}, \boldsymbol{g}^{(j)}\}$: Beobachtungspunkt j mit der Position $\boldsymbol{H}^{(j)}$, dem Grauwert $L^{(j)}$ und dem Bildsignalgradienten $\boldsymbol{g}^{(j)}$
$\boldsymbol{W}^{(j)}_i$	$\boldsymbol{W}^{(j)}$ bei der i–ten Iteration
$\boldsymbol{W}^{(j)'}$	$\boldsymbol{W}^{(j)}$ Beobachtungspunkt nach der Bewegung
$\boldsymbol{V}$	Vektor mit Luminanzfehlern aus unterschiedlichen Orten in der Bildebene
x, y	horizontale bzw. vertikale Ortskoordinate in der Bildebene
X, Y , Z	Koordinaten im dreidimensionalen Weltkoordinatensystem
$\mathfrak{X}^{(\delta)}, \mathfrak{Y}^{(\delta)}, \mathfrak{Z}^{(\delta)}$	Koordinaten im Koordinatensystem der Objektkomponente $\mathfrak{C}^{(\delta)}$
σ_n^2	Varianz des Kamerarauschens
σ_X^2, σ_Y^2, σ_Z^2	Varianzen eines räumliche Positionsfehlers $\Delta\boldsymbol{H}$
$\sigma^2_{\Delta s_f}$, $\sigma^2_{\Delta s_n}$	Varianz des Luminanzfehlers $\Delta\boldsymbol{s}_f$ bzw. $\Delta\boldsymbol{s}_n$
$\sigma^2_{\Delta s}$	$= \sigma^2_{\Delta s_f} + \sigma^2_{\Delta s_n}$: Varianz des gesamten Luminanzfehlers
σ^2_{dfd}	Varianz des *dfd* der Beobachtungspunkte
$\overline{\sigma}^2_T$, $\overline{\sigma}^2_R$	mittlere Varianz der Schätzfehler der Translations– bzw. Rotationsparameter
β	Quotient von Objektbreite b zu maximaler Objekttiefe h für die Erzeugung eines Dreiecksnetzes
$\sigma^2_{\tilde{e}}$	Leistung der Synthesefehler
Ψ	Gerade, die durch die Position eines Kugelgelenkes verläuft
Ψ'	Ψ nach der Bewegung

$\frac{\delta h_x(\boldsymbol{H})}{\delta H_x}$	partielle Ableitung von $h_x(H_X, H_Y, H_Z)$ nach H_X

1. Einleitung

Seit der Einführung des diensteintegrierenden, digitalen Übertragungsnetzes ISDN (Integrated Services Digital Network) in Deutschland stehen jedem Teilnehmer mit dem ISDN–Basisanschluß über bestehende Telefonanschlußleitungen zwei sog. ”B–Kanäle” von je 64 kbit/s sowie ein ”D–Kanal” von 16 kbit/s in jeder Verkehrsrichtung zur Verfügung[75]. Da für die Sprachübertragung nur einer der B–Kanäle erforderlich ist, kann der andere z.B. für Bewegtbildübertragung genutzt werden, so daß das ISDN auch Bildfernsprechdienst (Bildtelefon) ermöglicht.

Die Übertragung digitaler Bewegtbildsequenzen in dem für den Bildfernsprechdienst vereinbarten Bildformat CIF[12] (Common Intermediate Format) bei einer Bildfolgefrequenz von 10Hz erfordert eine PCM–Datenrate (Puls Code Modulation) von 12.5 Mbit/s. Für die Übertragung über einen B–Kanal ist demzufolge eine Reduktion dieser Datenrate um den Faktor 200 erforderlich. Um diese hohe Datenreduktion erreichen zu können, sind Quellencodierverfahren zur Redundanz– und Irrelevanzreduktion notwendig. Bei der Redundanzreduktion werden statistische Abhängigkeiten ausgenutzt, die in Bewegtbildsequenzen sowohl innerhalb eines Bildes (örtlich) als auch zwischen Bildern (zeitlich) bestehen. Redundante Anteile des Bildsignals können beim Empfänger aus anderen Signalteilen rekonstruiert und daher von der Übertragung ausgeschlossen werden[68]. Zur Redundanzreduktion sind Quellenmodelle für die Signalquelle erforderlich. Je besser das Quellenmodell das statistische Verhalten eines Bildsignals beschreibt, umso größer ist die Datenreduktion. Bei der Irrelevanzreduktion werden Bildunterschiede zwischen Originalbild und übertragenem Bild zugelassen, die der Nachrichtenempfänger, d.h. das menschliche Auge, nicht wahrnehmen kann oder als unwesentlich beurteilt[22][73]. Zur Irrelevanzreduktion sind psychovisuelle Wahrnehmungsmodelle des Nachrichtenempfängers notwendig.

In den Jahren 1989 und 1995 wurden die Standardisierungsvorschläge H.261 bzw. H.263 von der ITU–T (früher: CCITT) verabschiedet, die die Übertragung einer digitalen Bewegtbildsequenz mit Datenraten von 32 kbit/s bis zu 2 Mbit/s ermöglichen[12][42]. Die zugehörigen Quellencodierverfahren basieren auf dem Konzept einer blockbasierten Hybridcodierung mit Bewegungskompensation und Transformationscodierung des verbleibenden Prädiktionsfehlers[4][23][25][54][80]. Den Standardisierungsvorschlägen H.261 und H.263 liegt das Quellenmodell translatorisch bewegter, quadratischer Blöcke zugrunde[5][20][57][64].

Um eine wirklichkeitsnähere Beschreibung der sich in der realen Welt bewegenden Objekte zu erreichen, wird im Quellenmodell der objektbasierten Analyse–Synthese–Codierung die Bewegung, Form und Oberflächentextur von sich bewegenden Objekten mittels eines Objektmodells beschrieben[20][56]. Zusätzlich kann das Quellenmodell noch ein Beleuchtungs–, Kamera– und Szenenmodell enthalten. Das Objektmodell beschreibt über ein Bewegungsmodell die möglichen Objektbewegungen, über ein Formmodell die möglichen Objektformen und über ein Oberflächenmodell die Oberflächentextur des Objektes.

Die objektbasierte Analyse–Synthese–Codierung (OBASC) unterteilt jedes Bild einer Bewegtbildsequenz in bewegte, beliebig geformte Objekte und ordnet jedem Objekt ein Modellobjekt zu, das die Form, Bewegung und Textur des Objektes durch Form–, Bewegungs– und Texturparameter beschreibt[56]. Die Texturparameter beinhalten hierbei die Luminanz– und Chrominanzwerte der Textur auf der Objektoberfläche. Die genauere Bedeutung der Form– und Bewegungsparameter hängt vom verwendeten Objektmodell ab. Die Wahl eines für die Codierung geeigneten Objektmodells und die zuverlässige Schätzung der Form– und Bewegungsparameter lassen sich als die Hauptprobleme in der OBASC benennen[14][56]. Für jedes bewegte Objekt werden die Form– und Bewegungsparameter übertragen und die Rekonstruktion unter Verwendung zuvor übertragener Texturparameter durchgeführt. In Bildbereichen mit nicht ausreichender Bildqualität müssen die Texturparameter und die zweidimensionale Form dieser Bereiche übertragen werden. Sie werden als nicht modellierbare Objekte oder MF–Objekte (englisch: Model Failure) bezeichnet. Da die Codierung dieser Texturparameter mindestens eine Datenrate von 1.1 bit/pel erfordert[65], sollten die MF–Objekte möglichst klein gehalten werden.

Die objektbasierte Analyse–Synthese–Codierung wurde bisher mit drei unterschiedlichen Objektmodellen realisiert; dem Objektmodell zweidimensionaler, flexibler Objekte (F2D, englisch: Flexible 2D Objects)[18][19][29][30][31][32][33], dem Objektmodell dreidimensionaler, starrer Objekte (R3D, englisch: Rigid 3D Objects)[58][59][60][65] und dem Objektmodell dreidimensionaler, flexibler Objekte (F3D, englisch: Flexible 3D Objects)[61][63][65]. Gemäß dem Objektmodell F2D werden die Texturparameter eines Modellobjektes aus einem realen Bild entnommen. Die Form eines Objektes wird als zweidimensional angenommen und durch seine Silhouette beschrieben. Die Bewegung wird durch ein Displacementvektorfeld dargestellt. Im Vergleich zu dem Objektmodell F2D beschreiben die Objektmodelle R3D und F3D die Objekte mittels bewegter, starrer bzw. flexibler dreidimensionaler Modellobjekte. Durch die dreidimensionale Beschreibung wird eine bessere Annäherung der Form und Bewegung der realen Objekte erreicht, als dies mit zweidimensionalen Modellobjekten möglich ist[20][65]. Die dreidimensionale Form eines Modellobjektes wird durch ein Dreiecksnetz aufgespannt. Die Stützpunkte des Dreiecksnetzes bilden die Parameter zur Beschreibung der Form. Um die Anzahl der Parameter zu reduzieren, wird die Form aus der Silhouette des korrespondierenden realen Objektes mittels einer Distanztransformation berechnet[62][65]. Die Bewegung eines starren Modellobjektes wird durch sechs Parameter beschrieben: drei Rotationswinkel und einen dreidimensionalen Translationsvektor. Durch Projektion eines realen Bildes auf das dreidimensionale Modellobjekt wird diesem Textur zugeordnet. Während das Objektmodell R3D nur die Bewegung von starren Objekten beschreiben kann, erlaubt das Objektmodell F3D auch eine lokale Verformung der Objekte tangential zur Objektoberfläche. Diese Verformung wird durch Verschiebungsparameter beschrieben. In vielen Anwendungsgebieten der Bewegtbildübertragung sind die realen Objekte gegliedert, d.h. sie bestehen aus mehreren miteinander gekoppelten Objektkomponenten mit

unterschiedlichen Bewegungen. Dabei können sich die Objektkomponenten gegenseitig verdecken. Das Objektmodell der starren Objekte (R3D) versagt, wenn sich die Objektkomponenten eines gegliederten Objektes unterschiedlich bewegen.

In dieser Arbeit wird ein objektbasierter Analyse–Synthese–Coder basierend auf dem Quellenmodell bewegter dreidimensionaler, gegliederter Objekte (A3D, englisch: Articulated 3D Objects) entwickelt. Gemäß diesem Quellenmodell können die realen Objekte gegliedert sein und die Objektkomponenten sich unterschiedlich bewegen und gegenseitig verdecken. In Gegensatz zu den Quellenmodellen R3D und F3D wird hier jede Objektkomponente eines realen Objektes durch ihren eigenen Satz von Bewegungs–, Form– und Texturparametern beschrieben. Die Form der Objektkomponenten wird dabei als starr angenommen. Die Objektkomponenten eines gegliederten Objektes sind untereinander durch Kugelgelenke gekoppelt. Ein Gelenk legt der Relativbewegung zweier gekoppelter Objektkomponenten Einschränkungen der Bewegungsmöglichkeiten auf und wird hier als Bestandteil der Form eines gegliederten Objektes betrachtet. Dazu ist jedes Gelenk eindeutig einer Objektkomponente zugeordnet und seine Position im Koordinatensystem der zugehörigen Objektkomponente dargestellt.

Im Vergleich zu den Quellenmodellen R3D und F3D soll mittels des erweiterten dreidimensionalen Form– und Bewegungsmodells eine genauere Nachbildung der Form und Bewegung realer Objekte insbesondere bei unterschiedlicher Bewegung der Objektkomponenten ermöglicht werden. Daraus soll sich eine effizientere Beschreibung eines zu codierenden Bildes aufgrund der aktuellen Bewegungs– und Formparameter sowie zuvor übertragener Texturparameter ergeben, als es mit den Quellenmodellen R3D und F3D möglich ist. Durch diese effizientere Beschreibung eines zu codierenden Bildes soll eine Reduzierung der erforderlichen Datenrate der Texturparameter erzielt werden. Da mit dem neuen Quellenmodell für jede Objektkomponente die Bewegungsparameter übertragen werden müssen, erhöht sich die zur Objektbeschreibung erforderliche Datenrate. Um die erforderliche Datenrate zur Beschreibung der Form eines gegliederten Objektes nicht wesentlich zu erhöhen, soll die Form der Objektkomponenten aus einem ersten starren Modellobjekt durch Aufteilung in Modellobjektkomponenten gewonnen werden. Die dreidimensionale Bewegung einer starren Objektkomponente läßt sich ebenfalls durch sechs Parameter beschreiben. Die Position eines Gelenkes bezogen auf die Koordinaten der zugehörigen Objektkomponente läßt sich durch einen Ortsvektor mit drei Komponenten darstellen. Die Gelenke sollen sowohl eine genauere Schätzung als auch eine effizientere Codierung der Bewegungsparameter der Objektkomponenten ermöglichen. Es ist letztlich zu prüfen, ob die Reduktion der Datenrate der Texturparameter trotz der zusätzlichen Datenrate für die Bewegung und Form einen Gewinn erbringt.

Zur Implementierung einer OBASC basierend auf dem Quellenmodell bewegter dreidimensionaler, gegliederter Objekte sind in der Bildanalyse Algorithmen für die Schätzung der Form– und Bewegungsparameter gegliederter Objekte erforderlich.

Für die Schätzung der Formparameter gegliederter Objekte soll ein Verfahren entwickelt werden, das die Formparameter der Objektkomponenten und die Position eines jeden Gelenkes bezüglich des Koordinatensystems der zugehörigen Objektkomponente schätzt. Das Verfahren zur Schätzung der Formparameter der Objektkomponenten soll aus folgenden Schritten bestehen: Forminitialisierung, Objektgliederung und Formadaption.
Bei der Forminitialisierung soll jedes reale Objekt zunächst durch ein starres Modellobjekt beschrieben werden. Die Form dieses starren Modellobjektes soll wie beim Quellenmodell R3D aus der realen Objektsilhouette unter Verwendung einer Hypothese berechnet werden.
Bei der Objektgliederung soll das starre Modellobjekt in Modellobjektkomponenten aufgeteilt werden. Dazu sollen Parameter der dreidimensionalen Bewegung von relativ kleinen Oberflächenelementen des starren bewegten Objektes ausgewertet werden, indem Oberflächenelementen mit ähnlichen Bewegungsparameterwerten zu einer Objektkomponente zusammengefaßt werden. Dabei sollen die in vorherigen Bildern erhaltenen Ergebnisse dieser Auswertung mitberücksichtigt werden. Um die Zuverlässigkeit der Schätzung der Bewegungsparameter von relativ kleinen Oberflächenelementen zu verbessern, soll ein Maximum–Likelihood–Schätzer entwickelt werden, der sowohl ein stochastisches Modell des Schätzfehlers der Form des starren Modellobjektes als auch Meßwerte benachbarter Oberflächenelemente bei der Bewegungsschätzung berücksichtigt. Weiterhin soll der Schätzer robust gegen Ausreißer in den Meßwerten sein. Im Falle gegenseitiger Verdeckung der Objektkomponenten sollen verdeckte und verdeckende Objektkomponenten unterschieden und die Form ihrer nicht sichtbaren Teile geschätzt werden. Diese Schätzung stellt ein sehr schwieriges Problem dar. Im Rahmen dieser Arbeit soll dafür nur ein möglichst einfacher Ansatz untersucht werden.
Bei der Formadaption soll die Form der Modellobjektkomponenten fortlaufend an die reale Form der Objektkomponenten angepaßt werden. Dazu soll ein Verfahren entwickelt werden, das die Form der Modellobjektkomponenten mittels der realen Objektsilhouetten anpaßt. Die Objektsilhouette soll wie beim Quellenmodell R3D im wesentlichen aus einer Änderungsdetektion bestimmt werden.
Zur Schätzung der Position eines Gelenkes bezogen auf das Koordinatensystem der zugehörigen Objektkomponente soll ein Verfahren auf Basis einer Regression entwickelt werden, das die Gelenkbindung ausnutzt. Gemäß der Gelenkbindung muß die Bewegung eines Gelenkes die Bewegungsgleichungen beider angeschlossenen Objektkomponenten erfüllen. Dabei soll vorausgesetzt werden, daß die Parameter der dreidimensionalen Form der Objektkomponenten bereits geschätzt worden sind.

Für eine genauere Schätzung der Bewegungsparameter gegliederter Objekte soll der oben beschriebene Maximum–Likelihood–Schätzer so erweitert werden, daß für jede Objektkomponente die von den Gelenken der Objektkomponenten auferlegten Einschränkungen der Bewegungsmöglichkeiten berücksichtigt werden. Dabei sollen Bewegungsparameter aus der Schätzung genommen werden, die aufgrund der eingeschränkten Bewegungsmöglichkeiten von anderen Bewegungsparametern abhängig sind.

Anschließend müssen sowohl die Algorithmen zur Bildsynthese als auch die Algorithmen zur effizienten Codierung der Textur–, Form– und Bewegungsparameter starrer Objekte für gegliederte Objekte erweitert werden. Zur Bewertung der Effizienz des vorgeschlagenen erweiterten Quellenmodells gegliederter Objekte muß die Gesamtdatenrate bestehend aus den Teildatenraten für die Form–, Bewegungs– und Texturparameter betrachtet werden, da sich teilweise Erhöhung und Reduzierung ergeben.

In Kapitel 2 werden die Grundlagen der objektbasierten Analyse–Synthese–Codierung erläutert. Dazu werden die Struktur der OBASC und ihre fünf Hauptkomponenten Bildanalyse, Bildsynthese, Parametercodierung, Parameterdecodierung und Parameterspeicher dargestellt.

In Kapitel 3 wird das neue Quellenmodell vorgestellt.

In Kapitel 4 werden die erweiterten Aufgaben der Bildanalyse erläutert. Die Bildanalyse ermittelt für jedes Objekt die zugehörenden Modellobjektparameter. Die Algorithmen für die Schätzung der Parameter der dreidimensionalen Form und Bewegung gegliederter Objekte werden daran anschließend beschrieben.

In Kapitel 5 werden die notwendigen Modifikationen der Bildsynthese beschrieben. Die Bildsynthese rekonstruiert unter Verwendung der von der Bildanalyse für jedes Objekt ermittelten Modellobjektparameter das zu übertragende Bild.

In Kapitel 6 werden die Erweiterungen der Codierungsverfahren für die geschätzten Parameter erläutert.

In Kapitel 7 werden die Ergebnisse der experimentellen Untersuchungen beschrieben. Die für die Untersuchungen ausgewählten Testsequenzen werden zunächst kurz vorgestellt. Die Diskussion und Bewertung der Codiereffizienz der unterschiedlichen dreidimensionalen Quellenmodelle bilden den Abschluß dieses Kapitels.

2. Objektbasierte Analyse–Synthese–Codierung (OBASC)

In diesem Kapitel werden kurz die Grundlagen der objektbasierten Analyse–Synthese–Codierung erläutert. Dazu wird zunächst das Konzept der OBASC erklärt und danach werden die fünf Hauptkomponenten Bildanalyse, Parametercodierung, Parameterdecodierung, Parameterspeicher und Bildsynthese beschrieben.

Die OBASC verfolgt das Ziel, mittels eines geeigneten Quellenmodells jedes Bild einer Bildsequenz effizient zu codieren. Dazu wird jedes Bild i_{k+1} zum Zeitpunkt t_{k+1} (kurz: Bild *k+1*) in sich einheitlich bewegende, topologisch zusammenhängende Objekte der Kamerabildebene unterteilt. Jedes Objekt der Kamerabildebene wird durch drei Parametersätze Textur, Form und Bewegung beschrieben. Die Texturparameter $\boldsymbol{C}_m$ beschreiben die Oberflächentextur, die Formparameter $\boldsymbol{F}_m$ die Form und die Bewegungsparameter $\boldsymbol{B}_m$ die Bewegung eines Objektes *m*.

Zur Erklärung des Konzepts der OBASC wird Bild 2.1 benutzt. Der Parameterspeicher enthält die übertragenen und decodierten Parametersätze $\boldsymbol{C}'$, $\boldsymbol{F}'$ und $\boldsymbol{B}'$ aller Objekte. Die übertragenen Parameter zur Aktualisierung des Parameterspeichers nach der Decodierung werden mit dem Index $()''$ bezeichnet. Der Parameterspeicher speichert sowohl im Sender als auch im

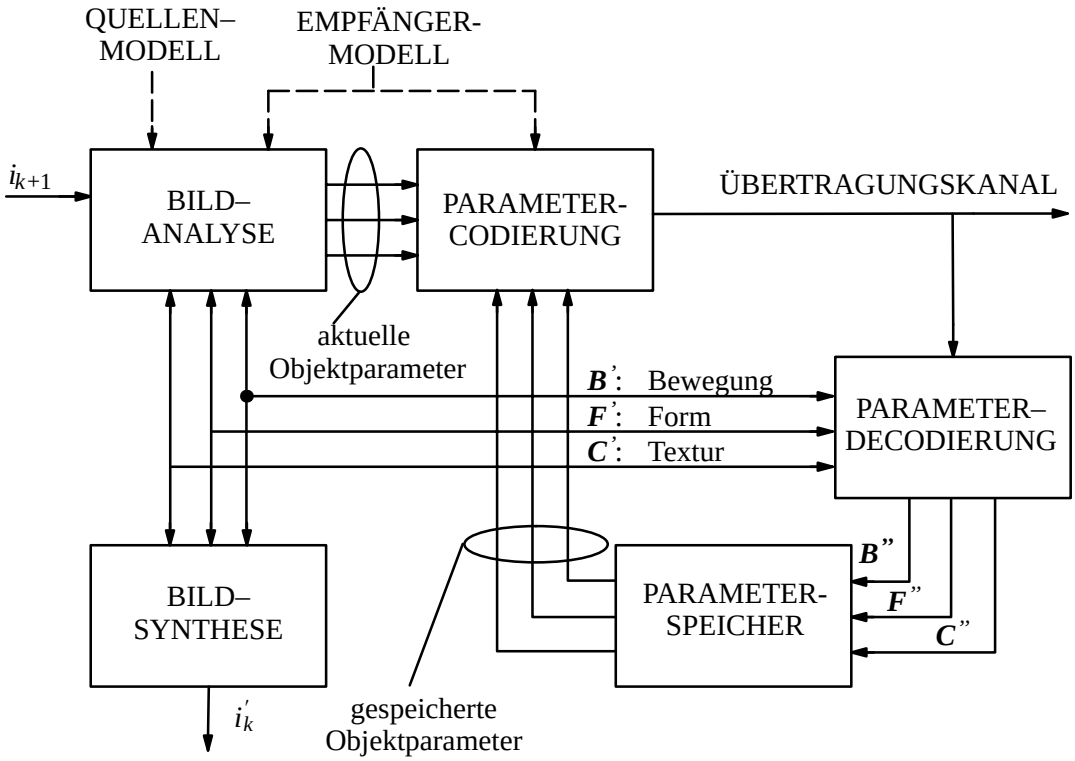


Bild 2.1 Blockdiagramm eines objektbasierten Analyse–Synthese–Coders.

Empfänger die gleichen Informationen. Mit den gespeicherten Parametersätzen wird mittels eines Algorithmus zur Bildsynthese das decodierte Bild i_k' synthetisiert. Dieses synthetisierte Bild i_k' wird beim Empfänger auf einem Monitor dargestellt.

Die aktuellen Parametersätze aller Objekte werden durch eine Bildanalyse automatisch geschätzt. Dazu werden das aktuelle Bild i_{k+1} und die für das Bild i_k ermittelten Parametersätze $\boldsymbol{C}'$, $\boldsymbol{F}'$ und $\boldsymbol{B}'$ berücksichtigt. Die aktuellen Parametersätze $\boldsymbol{C}_{m,k+1}$, $\boldsymbol{F}_{m,k+1}$ und $\boldsymbol{B}_{m,k+1}$ eines Objektes *m* stehen am Ausgang der Bildanalyse in unkomprimierter Form zur Verfügung.

Um die Bedeutung der Parametersätze zu veranschaulichen, werden sie mit Hilfe des Bildes 2.2 für den Fall starrer, translatorisch bewegter Objekte der Kamerabildebene beschrieben. Die Formparameter $\boldsymbol{F}_m$ können als Bildpunkte eines Binärbildes interpretiert werden, in dem die Berandung des Objektes *m* gekennzeichnet ist. Die Bewegungsparameter $\boldsymbol{B}_m$ eines Objektes *m* beschreiben die Bewegung dieses Objektes anhand der x– und y–Komponente des im gesamten Objekt gültigen Displacementvektors. Die Texturparameter $\boldsymbol{C}_m$ beschreiben die Luminanz– und Chrominanzwerte der Objektoberfläche des Objektes *m*. Aufgrund von relativen Objektbewegungen können die Texturparameter eines Objektes verdeckt oder sichtbar werden. In Bild 2.2 werden z.B. aufgrund der Bewegung des Objektes 2 bei Objekt 1 Bildbereiche frei, die bisher nicht sichtbar waren. Für diese Bildbereiche lagen zum Zeitpunkt t_k keine Texturparameter vor. Sie werden hier als freigewordene Bildbereiche bezeichnet. Genauso gehören zu den freigewordenen Bildbereichen alle bisher nicht sichtbaren Objektteile, die aufgrund von Objektdrehungen sichtbar werden, für die ebenfalls noch keine Texturparameter vorliegen. Freigewordene Bildbereiche aufgrund von Kamerabewegungen sind im Rahmen dieser Arbeit aufgrund der Annahme einer statischen Kamera ausgeschlossen (vgl. Abschnitt 3.1). Die Texturparameter der freigewordenen Bildbereiche eines Objektes *m* werden als $\boldsymbol{C}_m^{FB}$ bezeichnet. Die aktuellen Texturparameter eines Objektes *m* ergeben sich dabei zu

$$\boldsymbol{C}_{m,k+1} = \boldsymbol{C}_{m,k} \cup \boldsymbol{C}_{m,k+1}^{FB}. \tag{2.1}$$

Da sich die Bewegungs– und Formparameter i.allg. effizient codieren lassen[33] und die Codierung der Texturparameter bei guter Bildqualität mindestens eine Datenrate von 1.1 bit/pel erfordert[59][65], sollen zu einer effizienten Codierung der Parametersätze die Bewegungs– und Formparameter durch die Bildanalyse so bestimmt werden, daß möglichst große Teile des aktuellen Bild i_{k+1} durch schon für das Bild i_k übertragene Texturparameter, aktuelle noch zu übertragende Bewegungs– und Formparameter sowie Texturparameter der freigewordenen Bildbereiche beschrieben werden können. Aufgrund von Verletzungen der Annahmen des verwendeten Quellenmodells oder Fehlern bei der Schätzung der Parametersätze gibt es Bildbereiche, die nicht mit ausreichender Qualität beschrieben werden können. Eine Überprüfung der Modellbeschreibung des Bildes i_{k+1} am Ende der Bildanalyse detektiert diese Bildbereiche, wobei Bildbereiche mit kleinen Positions– und Formfehlern nicht berücksichtigt werden[32][65]. Da diese detektierten Bildbereiche nicht durch aktuelle noch zu übertra-

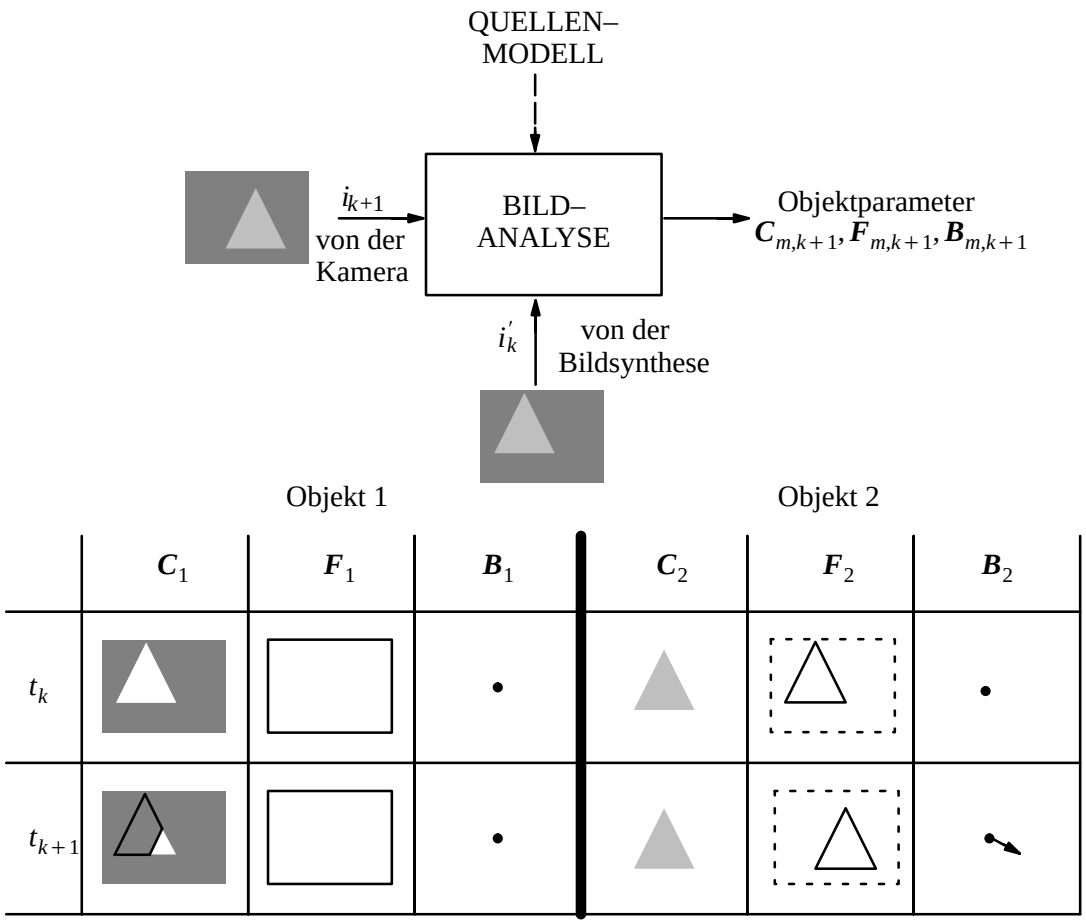


Bild 2.2 Darstellung der Bildanalyse am Beispiel starrer, translatorisch bewegter Objekte der Kamerabildebene. Die gestrichelten Linien bei der Darstellung der Formparameter $\boldsymbol{F}_2$ markieren die Bildbegrenzungen als Bezugspunkt zur Beschreibung der Lage dieses Objektes in der Kamerabildebene. Der schwarz umrandete Bereich bei der Darstellung der Texturparameter $\boldsymbol{C}_1$ beinhaltet diejenigen Texturparameter des Objektes 1, die durch die Bewegung des Objektes 2 sichtbar geworden sind.

gende Bewegungs– und Formparameter sowie schon für das Bild i_k übertragene Texturparameter beschrieben werden können, werden sie durch Form– und Texturparameter aus dem aktuellen Bild i_{k+1} beschrieben. Sie werden im folgenden als nicht modellierbare Objekte der Kamerabildebene oder kurz MF–Objekte (englisch: Model Failure) bezeichnet. Die verbleibenden Bildbereiche heißen modellierbare Objekte der Kamerabildebene oder kurz MC–Objekte (englisch: Model Compliance). Damit gehört jedes Objekt entweder zur Objektklasse MC oder MF. In Bild 2.2 sind z.B zwei MC–Objekte dargestellt.

Bei der Parametercodierung erfolgt eine effiziente Codierung der aktuellen von der Bildanalyse für jedes Objekt *m* geschätzten Parametersätze $\boldsymbol{C}_{m,k+1}$, $\boldsymbol{F}_{m,k+1}$ und $\boldsymbol{B}_{m,k+1}$. Wie in Bild 2.3 dargestellt, beinhaltet die Parametercodierung drei Teilcoder für die Codierung der Bewegungs–, Form– bzw. Texturparameter. Die Bewegungsparameter werden prädiktiv codiert. Die Formparameter werden unter Ausnutzung der aktuellen codierten Bewegungsparameter $\boldsymbol{B}'_{m,k+1}$ und der decodierten Formparameter $\boldsymbol{F}'_{m,k}$ aus dem zeitlich vorangegangenen Bild i_k ebenfalls prädiktiv codiert. Die Texturparameter werden durch Anwendung hybrider Verfahren codiert, die bewegungskompensierende Prädiktion und Transformationscodierung kombinieren.

Die Steuerung der Parameterübertragung entscheidet, welche der codierten Parametersätze eines jeden Objektes *m* zum Empfänger übertragen werden müssen. Diese Entscheidung wird von der zur Verfügung stehenden Datenrate und der Codermodeinformation gesteuert. Die Codermodeinformation wird von der Steuerung der Codiermodi erzeugt. Die Steuerung der Codiermodi wird durch die Objektklasse MC/MF und das Empfängermodell beeinflußt. Im Falle von MF–Objekten werden nur die Form– und Texturparameter übertragen, da Bewegungsparameter mit ausreichender Qualität für diese Bildbereiche nicht von der Bildanalyse bestimmt werden konnten. Für MC–Objekte werden nur die codierten Bewegungs– und Formparameter übertragen, da gemäß den Ergebnissen der Bildanalyse unter Verwendung

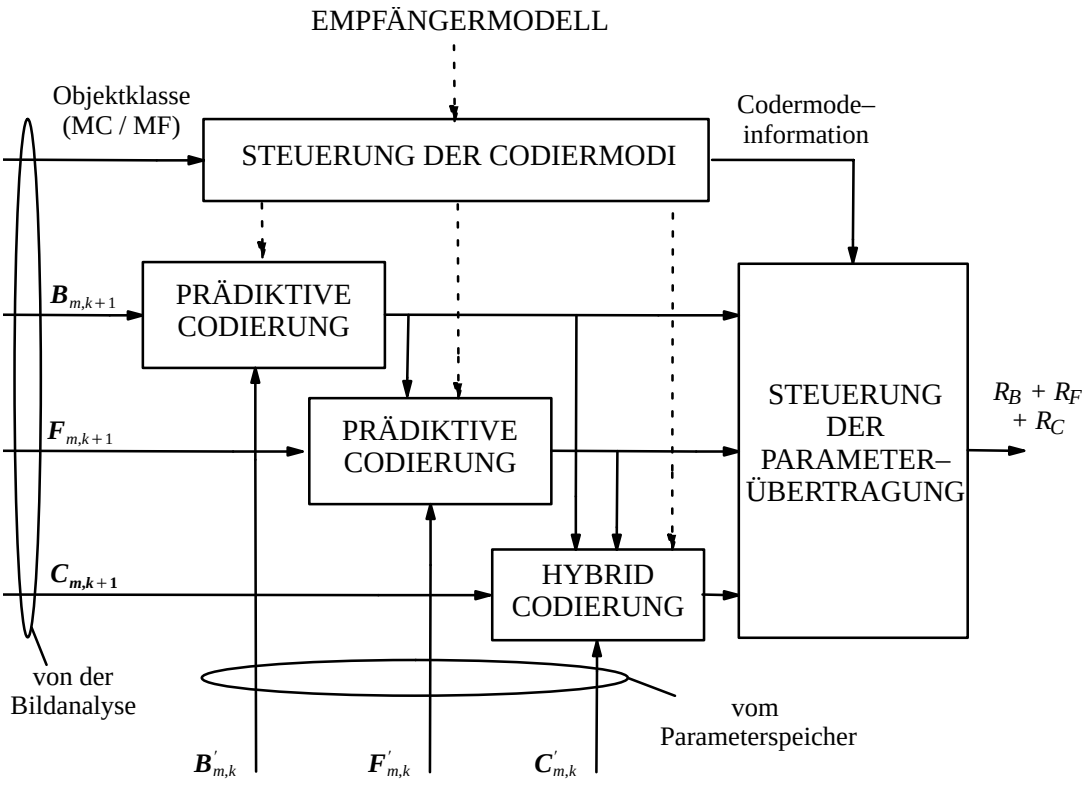


Bild 2.3 Parametercodierung (***B*** Bewegungs–, ***F*** Form– und ***C*** Texturparameter).

schon für das vorhergehende Bild übertragener Texturparameter und ohne Berücksichtigung aktueller kleiner Positions– und Formfehler eine gute Beschreibung des aktuellen Bildes i_{k+1} möglich ist. Zusätzlich müssen für die freigewordenen Bildbereiche der MC–Objekte Texturparameter übertragen werden. Die Form der freigewordenen Bildbereiche ergibt sich aus der Bewegungs– und Forminformation der MC–Objekte. Das Empfängermodell beschreibt die psychovisuelle Wahrnehmung des Nachrichtenempfängers[22][73]. Mittels des Empfängermodells können Sichtbarkeitsschwellen abgeschätzt werden, die zur Steuerung der in allen Teilcodern enthaltenen Quantisierer ausgenutzt werden, damit gerade keine für einen menschlichen Beobachter störenden Fehler sichtbar werden. Auf diese Weise erfolgt bei der Codierung eine Irrelevanzreduktion.

Abschließend werden die codierten Parameter übertragen. Als Datenrate R für die Codierung eines Bildes ergibt sich somit

$$R = R_B + R_F + R_C . \tag{2.2}$$

Die codierten und übertragenen Parameter werden sowohl vom Sender als auch vom Empfänger in der Parameterdecodierung decodiert (Bild 2.4). Die decodierten Parameter werden zur

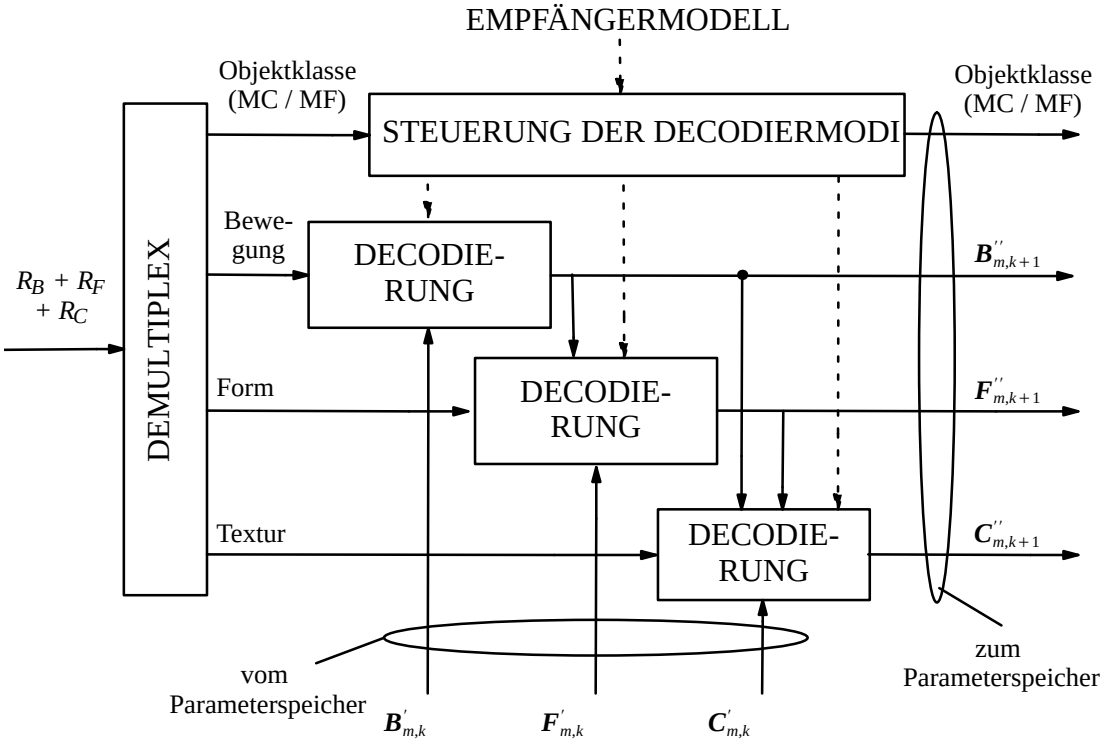


Bild 2.4 Parameterdecodierung (***B*** Bewegungs–, ***F*** Form– und ***C*** Texturparameter).

Aktualisierung der Parametersätze der MC–Objekte im Parameterspeicher verwendet. Mittels der decodierten Bewegungsparameter werden die MC–Objekte im Speicher an ihre neue

Position bewegt. Mittels der decodierten Formparameter der MC–Objekte erfolgt eine Aktualisierung der Form der entsprechenden Objekte im Speicher. Aus den Form– und Bewegungsparametern läßt sich die Form der freigewordenen Bildbereiche bestimmen. Die decodierten Texturparameter dieser Bildbereiche werden in die entsprechenden MC–Objekte eingetragen. Die Texturparameter der MC–Objekte werden an den durch die Formparameter der MF–Objekte bestimmten Stellen durch die Texturparameter der MF–Objekte ersetzt. Am Ausgang des Parameterspeichers stehen nur MC–Objekte mit ihren aktualisierten Bewegungs–, Form– und Texturparametern $\boldsymbol{B}'$, $\boldsymbol{F}'$ bzw. $\boldsymbol{C}'$ zur Verfügung.

3. Quellenmodelle für OBASC

Zur theoretischen Behandlung eines Nachrichtenübertragungssystems müssen Signalquelle, Kanal und Signalsinke durch mathematische Modelle beschrieben werden. Diese Modelle werden mit den Mitteln der Wahrscheinlichkeitslehre und Systemtechnik beschrieben und stellen häufig nur eine grobe Näherung der physikalischen Realität dar, da eine exakte mathematische Beschreibung teilweise nicht möglich oder unangemessen aufwendig ist. Im Rahmen dieser Arbeit wird ein neues Quellenmodell für die Signalquelle vorgestellt und zur Optimierung eines objektbasierten Codierverfahrens verwendet. Eine Modellierung der Signalsinke wird aus der Literatur[59] übernommen. Für die weiteren Untersuchungen wird der Übertragungskanal als ungestört angenommen.

In diesem Kapitel wird ein neues Quellenmodell für den objektbasierten Analyse–Synthese Coder vorgestellt. Für die Wahl der Parameter des Quellenmodells wird Rücksicht auf die Anwendung (hier die Bewegtbildcodierung bei niedrigen Datenraten) genommen. Dazu sollen die Parameter des Quellenmodells sowohl eine genauere Beschreibung der Signalquelle als auch eine effiziente Codierung ermöglichen. Mittels eines Quellenmodells werden die zeitlichen Bildsignaländerungen von einem Bild zum nächsten Bild einer Bewegtbildsequenz mathematisch beschrieben. Dazu wird angenommen, daß die zeitlichen Bildsignaländerungen im Kamerabild hauptsächlich durch sich bewegende Objekte in der durch eine Kamera aufgenommenen dreidimensionalen Welt hervorgerufen werden. Um den Zusammenhang zwischen den zeitlichen Bildsignaländerungen und den sich bewegenden Objekten mathematisch beschreiben zu können, wird der Vorgang der Abbildung von sich bewegenden Objekten in einer dreidimensionalen realen Welt auf das Kamerabild durch das Quellenmodell nachgebildet[45][56]. Als reale Welt wird hier die real existierende physikalische Welt bezeichnet. Sie läßt sich durch eine reale Szene mit realen sich bewegenden Objekten und einer realen Beleuchtung vollständig beschreiben. Die reale Welt wird durch die reale Kamera auf das reale Kamerabild abgebildet. Das Quellenmodell postuliert nun Annahmen über den Aufbau der realen Welt und die Abbildung der realen Welt auf das reale Kamerabild, die im Kamera–, Beleuchtungs–, Szenen– und Objektmodell festgelegt werden. Die künstliche Welt, die sich mittels des Quellenmodells aufbauen läßt, wird im Rahmen dieser Arbeit als Modellwelt bezeichnet. Sie wird durch eine Modellszene mit sich bewegenden Modellobjekten und eine Modellbeleuchtung beschrieben. Die Modellwelt wird durch die Modellkamera auf ein Modellbild abgebildet.

3.1 Kameramodell

Der Vorgang der Abbildung der dreidimensionalen, realen Welt auf das zweidimensionale, reale Bild wird durch das Kameramodell beschrieben[32][45][65]. Zur Beschreibung der Abbildung wird hier die Zentralprojektion verwendet, die eine hinreichend genaue Näherung der Abbildungseigenschaften einer realen Kamera darstellt[84]. Bei Annahme einer Zentralprojektion wird ein Punkt H auf der Oberfläche eines Objektes auf den Punkt h der Kamerabil-

debene abgebildet (Bild 3.1). Die Koordinaten des abgebildeten Punktes $\boldsymbol{h} = (h_x, h_y)^T$ in der Kamerabildebene ergeben sich in Abhängigkeit von der Brennweite F der Kamera und der Position des Punktes $\boldsymbol{H} = (H_X, H_Y, H_Z)^T$ im Weltkoordinatensystem (X, Y, Z) zu

$$\begin{aligned} h_x(\boldsymbol{H}) &= F \cdot \frac{H_X}{H_Z} \\ h_y(\boldsymbol{H}) &= F \cdot \frac{H_Y}{H_Z} \ . \end{aligned} \tag{3.1}$$

Es wird dabei angenommen, daß die Bildebene in der durch Z=F beschriebenen Raumebene parallel zur XY–Ebene liegt und daß sich der Brennpunkt der Kamera im Ursprung des Weltkoordinatensystems befindet. Eine Modellkamera arbeitet gemäß dem Kameramodell. Sie erzeugt daher unter Verwendung der Zentralprojektion Modellbilder der Modellwelt.

Die Abbildungsvorschrift gemäß Gl. (3.1) kann durch eine Taylorreihe im Arbeitspunkt H mit dem Ortsvektor $\boldsymbol{H}$ beschrieben werden

$$\begin{aligned} h_x(\boldsymbol{H} + \Delta\boldsymbol{H}) &= h_x(\boldsymbol{H}) + \frac{\delta h_x(\boldsymbol{H})}{\delta H_X}\Delta H_X + \frac{\delta h_x(\boldsymbol{H})}{\delta H_Y}\Delta H_Y + \frac{\delta h_x(\boldsymbol{H})}{\delta H_Z}\Delta H_Z + \varepsilon_x \\ h_y(\boldsymbol{H} + \Delta\boldsymbol{H}) &= h_y(\boldsymbol{H}) + \frac{\delta h_y(\boldsymbol{H})}{\delta H_X}\Delta H_X + \frac{\delta h_y(\boldsymbol{H})}{\delta H_Y}\Delta H_Y + \frac{\delta h_y(\boldsymbol{H})}{\delta H_Z}\Delta H_Z + \varepsilon_y. \end{aligned} \tag{3.2}$$

Dabei stellen $\varepsilon_x, \varepsilon_y$ die Terme höherer Ordnung dar. Bei Abbruch nach dem linearen Term sowie durch Auswertung der partiellen Ableitungen folgt

$$h(\boldsymbol{H} + \Delta\boldsymbol{H}) = h(\boldsymbol{H}) + \begin{bmatrix} \frac{F}{H_Z} & 0 & \frac{-F \cdot H_X}{H_Z^2} \\ 0 & \frac{F}{H_Z} & \frac{-F \cdot H_Y}{H_Z^2} \end{bmatrix} \cdot \Delta\boldsymbol{H} \tag{3.3}$$

$$\Delta\boldsymbol{h} = [\boldsymbol{K}] \cdot \Delta\boldsymbol{H}.$$

Unter einer digitalen Bewegtbildsequenz wird dann die zeitliche Folge örtlich abgetasteter Fernsehbilder verstanden (Bild 3.2), die durch die fortlaufende Abbildung der 3D–Szene auf die Bildebene unter Verwendung der Zentralprojektion entstehen[5]. Das Kamerasignal i_k besteht dabei aus drei Komponenten l_k, bl_k und rl_k, wobei l_k die Luminanzkomponente und bl_k bzw. rl_k jeweils eine der beiden Chrominanzkomponenten entsprechend der CCITT–Recommendation 601[11] darstellen. Hierbei wird angenommen, daß sich die Luminanzkomponente l_k additiv aus dem Nutzsignal s_k und dem Kamerarauschsignal n_k zusammensetzt[65]. Der zeitliche Abstand T_t zwischen aufeinanderfolgenden Bildern einer Sequenz ist der Kehr-

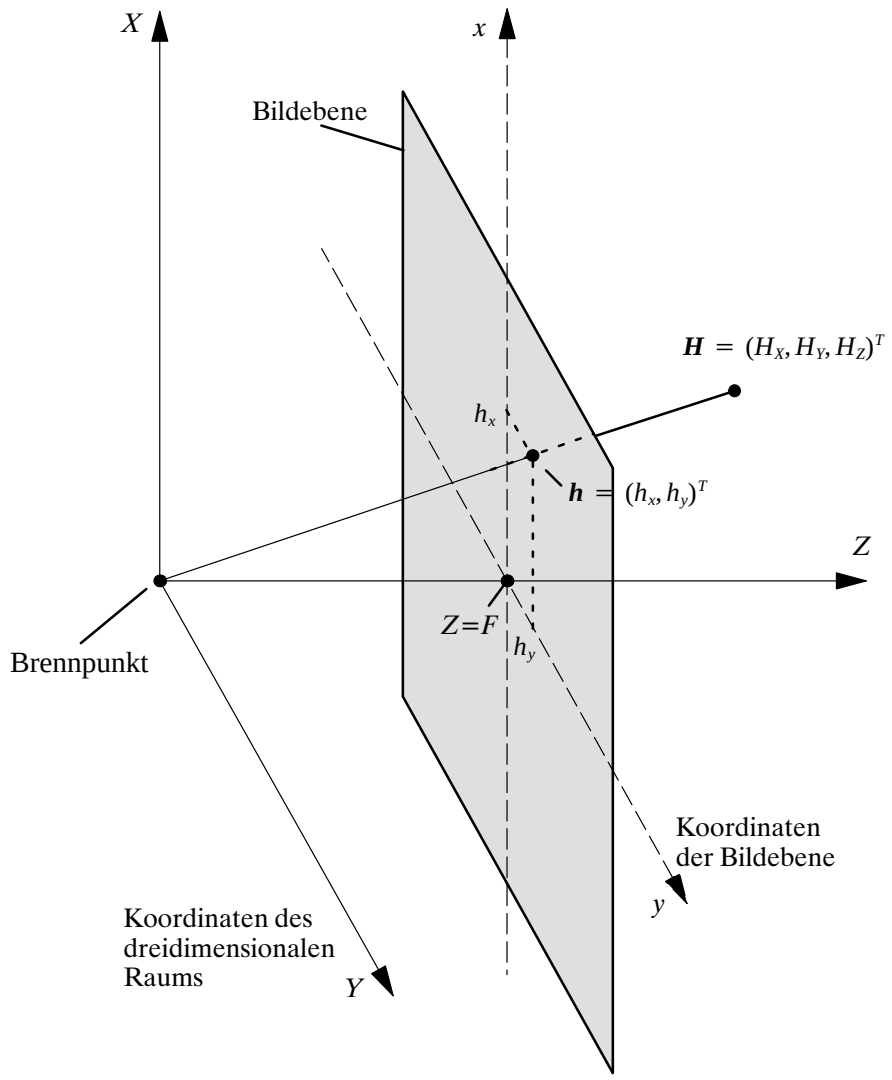


Bild 3.1 Abbildung der 3D–Szene auf die zweidimensionale Bildebene mittels Zentralprojektion.

wert der bei Bildaufnahme und –wiedergabe verwendeten Bildwiederholfrequenz f_{st}. Der horizontale und vertikale Abstand T_x bzw. T_y benachbarter Bildpunkte eines digitalen Fernsehbildes resultiert aus den örtlichen Abtastfrequenzen f_{sx}, f_{sy}. Aus T_x bzw. T_y ergeben sich mit den Bildabmessungen (Bildhöhe H_B und –breite B_B) die Anzahl *NLIN* von Zeilen je Bild und die Anzahl *NPEL* von Bildpunkten je Zeile der digitalen Bilder. Unter einem *Bildpunkt* wird ein Abtastwert $l_k[n_x, n_y]$ verstanden, der die Bildhelligkeit im Bild k, d.h. zum Zeitpunkt t_k, an den diskreten Ortskoordinaten n_x,n_y repräsentiert. Zur Beschreibung von Ortskoordinatendifferenzen als Vielfache der Bildpunktabstände auf dem Abtastraster wird die Maßeinheit ”pel” verwendet[5]. Zur Vereinheitlichung wird für das Raum– und das Bildkoordinatensystem die gleiche Einheit ”pel” benutzt.

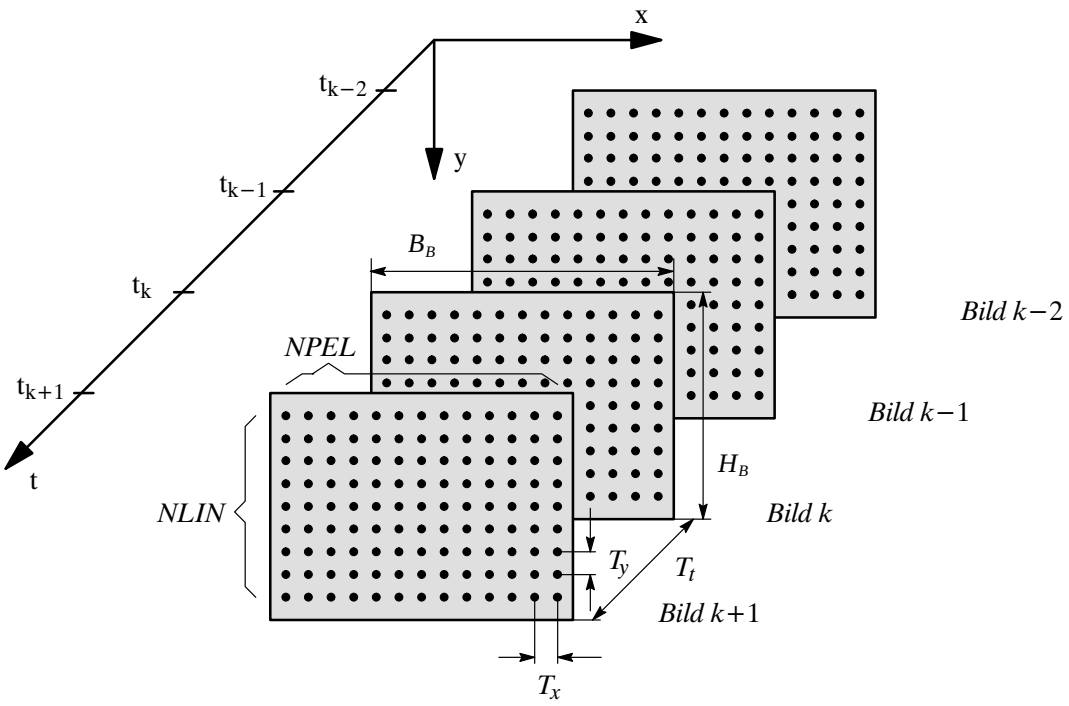


Bild 3.2 Darstellung einer digitalen Bildsequenz als zeitliche Folge örtlich abgetasteter Bilder (aus [5]).

Das hier beschriebene digitale Fernsehbild wird im folgenden als reales Bild bezeichnet. Das Modellbild entsteht entsprechend durch Abtastung und Quantisierung eines Bildes einer Modellkamera.

3.2 Beleuchtungsmodell

Mittels des Beleuchtungsmodells werden die zeitabhängigen Bildsignaländerungen im realen Kamerabild beschrieben, die durch die Änderung der Beleuchtung der realen Szene hervorgerufen werden [32][45][65]. Im Rahmen dieser Arbeit wird die reale Beleuchtung als diffus und zeitlich konstant beschrieben. Es wird daher angenommen, daß aufgrund der realen Beleuchtung keine Bildsignaländerungen in der Bildfolge hervorgerufen werden. Die durch gerichtete oder zeitliche Änderung der Beleuchtung entstehenden Einflüsse[66][82] werden damit bei diesem einfachen Beleuchtungsmodell nicht berücksichtigt.

3.3 Szenenmodell

Das Szenenmodell beschreibt die realen Objekte und ihre Lage in der realen Szene[65]. Die realen Objekte werden dabei mittels eines Objektmodells beschrieben und können nebeneinander liegen oder sich gegenseitig verdecken (Bild 3.3). Der Bildbereich, auf den das Objekt bei Berücksichtigung von Verdeckungen abgebildet wird, wird als Bildbereich des Objektes

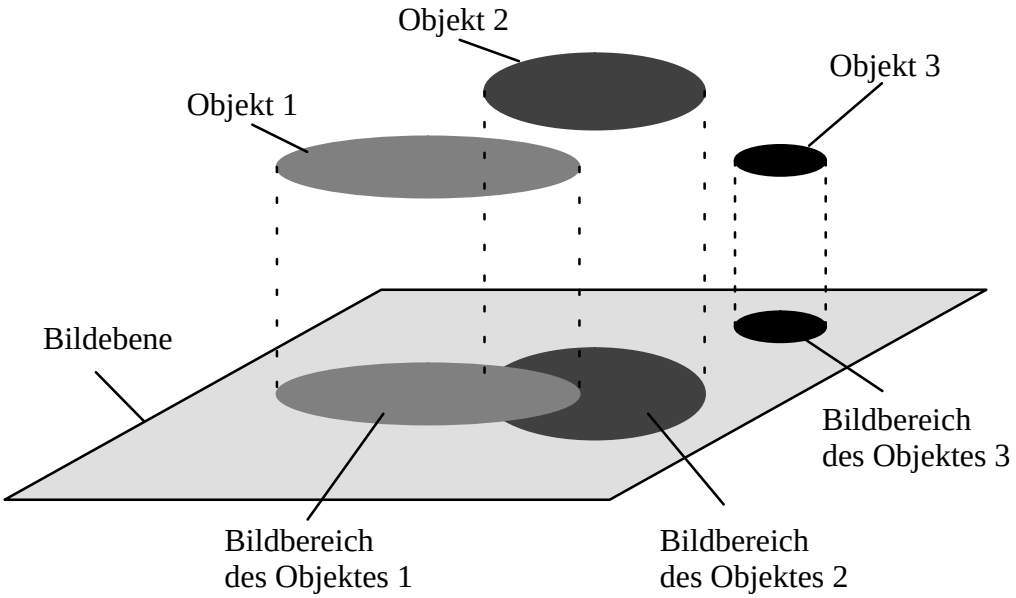


Bild 3.3 Szenenmodell. Objekt 1 verdeckt Objekt 2. Objekt 3 liegt neben Objekt 2.

bezeichnet. Der Bildbereich auf den ein Objekt ohne Berücksichtigung von Verdeckungen abgebildet werden würde, wird als Silhouette des Objektes gekennzeichnet.

Da die MF–Objekte gemäß Kapitel 2 nur zur Korrektur des Bildsignals in Teilen der Kamerabildebene verwendet werden, die nach der Abbildung einer Szene nicht mit ausreichender Qualität beschrieben worden sind, werden sie nicht als Teil einer Szene betrachtet und daher nicht durch das Objektmodell beschrieben.

3.4 Objektmodell

Mittels des Objektmodells werden die zeitlichen Bildsignaländerungen mathematisch beschrieben, die in der Kamerabildebene durch die sich bewegenden realen Objekte von einem realen Kamerabild k zum nächsten Kamerabild $k+1$ hervorgerufen werden. Das Objektmodell beschreibt über ein Formmodell die möglichen Objektformen, über ein Bewegungsmodell die möglichen Objektbewegungen und über ein Oberflächenmodell die Oberflächentextur der realen Objetke[65].

Die Form und Bewegung eines realen Objektes m wird durch die Parameter $\boldsymbol{F}_m$ bzw. $\boldsymbol{B}_m$ beschrieben. Sie werden hier als Form– bzw. Bewegungsparameter des Objektes m bezeichnet. Für das im Rahmen dieser Arbeit verwendete Oberflächenmodell wird die Oberfläche eines realen Objektes m als undurchsichtig und diffus reflektierend angenommen. Die Oberfläche eines realen Objektes wird durch die Texturparameter $\boldsymbol{C}_m$ beschrieben. Sie beinhalten

die Luminanz– und Chrominanzwerte $\boldsymbol{L}_m = \{L_m\}$ bzw. $\boldsymbol{BL}_m = \{BL_m\}$ und $\boldsymbol{RL}_m = \{RL_m\}$ der Objektoberfläche.

Aufgrund der bisher getroffenen Annahmen im Kamera–, Beleuchtungs– und Oberflächenmodell können die zeitlichen Veränderungen aufeinanderfolgender Bilder einer Bewegtbildsequenz allein durch die Bewegungs– und Formparameter der Objekte beschrieben werden. Die ortsabhängige Bild–zu–Bild–Luminanzdifferenz *fd* (englisch: frame difference) an der Stelle $\boldsymbol{h} = (h_x, h_y)^T$ von einem Bild k zum nächsten Bild k+1 wird daher als Funktion dieser Parameter dargestellt:

$$fd(\boldsymbol{h}) = s_{k+1}(\boldsymbol{h}) - s_k(\boldsymbol{h}) = s_k(f(\boldsymbol{B}, \boldsymbol{F}, \boldsymbol{h})) - s_k(\boldsymbol{h}). \tag{3.4}$$

Mit einer geeigneten Auswahl des Objektmodells soll diese Bild–zu–Bild–Luminanzdifferenz, mit möglichst wenigen, effizient zu codierenden Bewegungs– und Formparametern beschrieben werden.

In den folgenden Abschnitten werden die drei für diese Arbeit relevanten dreidimensionalen Quellenmodelle definiert. Nach der Beschreibung des Objektmodells bewegter dreidimensionaler, starrer Objekte[62][65] und bewegter dreidimensionaler, flexibler Objekte[63][65] folgt die Definition des neu entwickelten Objektmodells bewegter dreidimensionaler, gegliederter Objekte. Da alle Quellenmodelle die oben definierten Kamera–, Beleuchtungs– und Oberflächenmodelle beinhalten, werden hier nur die drei Objektmodelle erläutert.

3.4.1 Objektmodell bewegter dreidimensionaler, starrer Objekte

Bei diesem Objektmodell wird angenommen, daß die Form realer Objekte starr und dreidimensional ist und daß reale Objekte sich frei im dreidimensionalen Raum bewegen. Da von der realen Welt nur das reale Bild bekannt ist, erfolgt hier die Definition der dreidimensionalen Modellobjekte über die Bildbereiche der sich bewegenden, realen Objekte. Dabei stellt der Bildbereich eines sich bewegenden Objektes gemäß Abschnitt 3.3 den Bildbereich dar, auf den ein sich bewegendes reales Objekt abgebildet wird. Sie entsprechen den Objekten der Kamerabildebene von Kapitel 2. Jedem Bildbereich eines Objektes wird ein dreidimensionales starres Modellobjekt mit seinen Form–, Bewegungs– und Texturparametern zugeordnet.

In den folgenden Abschnitten erfolgt nun die noch ausstehende Definition des Form– und Bewegungsmodells.

3.4.1.1 Formmodell

Bei diesem Formmodell wird angenommen, daß die Form der realen Objekte starr ist. Das Formmodell beschreibt daher die Form eines realen Objektes durch die Form eines korrespondierenden dreidimensionalen starren Modellobjektes. Die Form eines Modellobjektes wird durch ein dreidimensionales starres Dreiecksnetz definiert, das durch N Kontrollpunkte $P^{(n)}$, $n=0,..,N-1$, aufgespannt wird (Bild 3.4). Die Position eines Kontrollpunktes $P^{(n)}$ wird durch den Ortsvektor $\boldsymbol{P}^{(n)*} = (P_R^{(n)}, P_S^{(n)}, P_T^{(n)})^T$ beschrieben. Dabei stellt (R, S, T) ein festes Objekt-

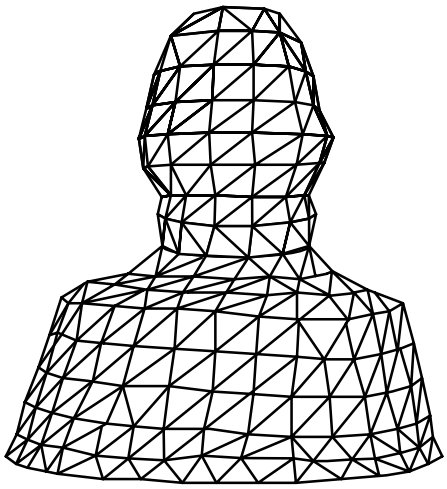


Bild 3.4 Dreiecksnetz eines Modellobjektes: Die Eckpunkte der Dreiecke sind die Kontrollpunkte $P^{(n)}$. Der sichtbare Oberflächenteil des Modellobjektes wird durch 300 Kontrollpunkte dargestellt.

koordinatensystem mit dem Ursprung in einem beliebigen Punkt G des Modellobjektes dar. Im weiteren kennzeichnet ein ()*, daß ein Ortsvektor () im Objektkoordinatensystem angegeben ist.

Falls das Modellobjekt in einer dreidimensionalen Modellwelt mit dem Weltkoordinatensystem (X, Y, Z) plaziert wird (Bild 3.5), läßt sich der Ortsvektor $\boldsymbol{H}$ eines beliebigen Modellobjektpunktes H bezüglich des Weltkoordinatensystems mit der Koordinatentransformation

$$\boldsymbol{H} = [\boldsymbol{R}]_0 \cdot \boldsymbol{H}^* + \boldsymbol{G} \tag{3.5}$$

bestimmen. Dabei ist $\boldsymbol{H}^*$ der Ortsvektor des Modellobjektpunktes H bezüglich des Objektkoordinatensystems, $\boldsymbol{G} = (G_x, G_y, G_z)^T$ gibt den Ortsvektor des Ursprungs des Objektkoordinatensystems bezüglich des Weltkoordinatensystems an und $[\boldsymbol{R}]_0$ stellt eine durch die drei Kardanwinkeln $\Re_X$, $\Re_Y$ und $\Re_Z$ beschriebene Rotationsmatrix gemäß

$$[\boldsymbol{R}]_0 = \begin{bmatrix} \cos\Re_Y\cos\Re_Z & \sin\Re_X\sin\Re_Y\cos\Re_Z - \cos\Re_X\sin\Re_Z & \cos\Re_X\sin\Re_Y\cos\Re_Z + \sin\Re_X\sin\Re_Z \\ \cos\Re_Y\sin\Re_Z & \sin\Re_X\sin\Re_Y\sin\Re_Z + \cos\Re_X\cos\Re_Z & \cos\Re_X\sin\Re_Y\sin\Re_Z - \sin\Re_X\cos\Re_Z \\ -\sin\Re_Y & \sin\Re_X\cos\Re_Y & \cos\Re_X\cos\Re_Y \end{bmatrix} \tag{3.6}$$

dar, welche die Verdrehung des Objektkoordinatensystems relativ zum Weltkordinatensystem durch drei aufeinanderfolgenden Drehungen beschreibt, wobei zuerst um die X–Achse mit dem Winkel $\Re_X$, dann um die Y–Achse mit dem Winkel $\Re_Y$ und abschließend um die Z–Achse mit dem Winkel $\Re_Z$ im Uhrzeigersinn gedreht wird.

Die drei Komponenten des Ortsvektors $\boldsymbol{G}$ des Ursprungs des Objektkoordinatensystems und die drei Winkel der Rotationsmatrix $[\boldsymbol{R}]_0$ definieren vollständig die Position $\boldsymbol{B}_{pos}$ eines Modellobjektes im Weltkoordinatensystem:

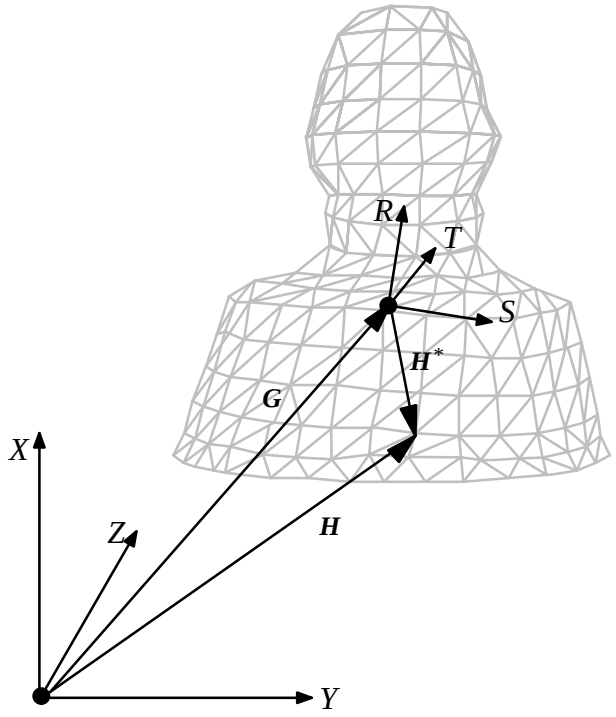


Bild 3.5 Das Objektkoordinatensystem (*R,S,T)* wird durch seinen Ursprung ***G*** und eine Rotationsmatrix $[\boldsymbol{R}]_0$ im Weltkoordinatensystem (*X,Y,Z*) festgelegt. $\boldsymbol{H}^*$ und $\boldsymbol{H}$ stellen die Ortsvektoren eines beliebigen Modellobjektpunktes H bezogen auf das Objektkoordinatensystem bzw. Weltkoordinatensystem dar.

$$\boldsymbol{B}_{pos} = \{\boldsymbol{G}, [\boldsymbol{R}]_0\}. \tag{3.7}$$

Gemäß Gl. (3.7) besitzt ein starres Modellobjekt 6 Freiheitsgrade im Weltkoordinatensystem.

3.4.1.2 Bewegungsmodell

Bei diesem Bewegungsmodell wird die dreidimensionale Bewegung eines Modellobjektes in der Modellwelt durch eine Rotation [***R***] und eine Translation ***T*** des Objektkoordinatensystems im Weltkoordinatensystem beschrieben (Bild 3.6). Ein Punkt H auf der Modellobjektoberfläche mit dem Ortsvektor ***H*** bewegt sich nach $\boldsymbol{H}'$ gemäß

$$\boldsymbol{H}' = [\boldsymbol{R}] \cdot (\boldsymbol{H} - \boldsymbol{G}) + \boldsymbol{G} + \boldsymbol{T}. \tag{3.8}$$

Die drei Kardanwinkel R_X, R_Y und R_Z der Rotationsmatrix [***R***] und die drei Komponenten T_X, T_Y und T_Z des Translationsvektors ***T*** stellen die Rotations– bzw. Translationsparameter des Objektes dar. Gleichung (3.8) verändert nicht die Position des Punktes $\boldsymbol{H}^*$ im Objektkoordinatensystem.

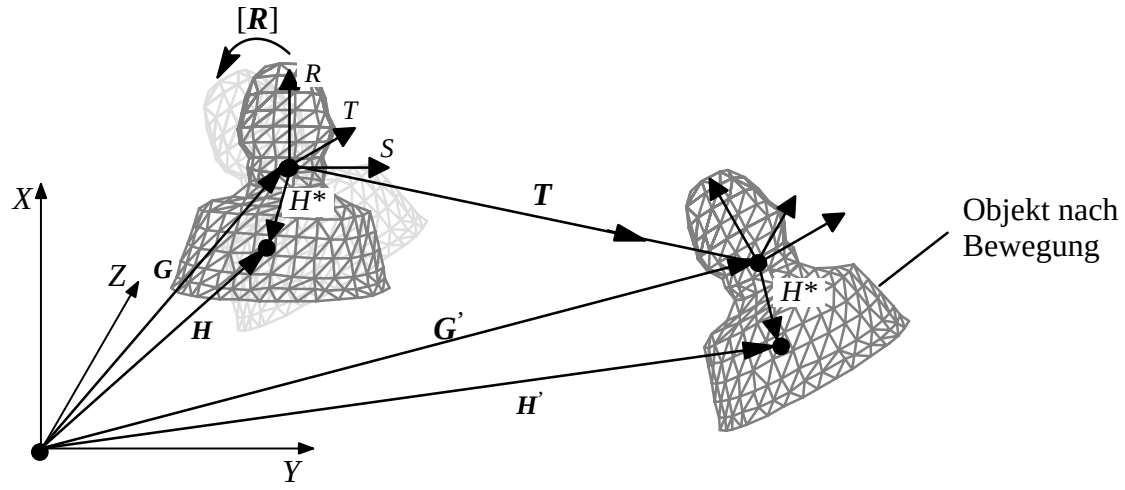


Bild 3.6 Bewegung eines Objektes gemäß Gl.(3.8). Das Objektkoordinatensystem wird mit der Rotationmatrix $[\boldsymbol{R}]$ gedreht und anschließend mit $\boldsymbol{T}$ nach $\boldsymbol{G}'$ verschoben. Der Punkt H bewegt sich von $\boldsymbol{H}$ nach $\boldsymbol{H}'$ gemäß Gl. (3.8).

Die aktuelle Postion $\boldsymbol{B}'_{pos}$ eines Modellobjektes im Weltkoordinatensystem ergibt sich aus der vorherigen Position $\boldsymbol{B}_{pos}$ sowie der Rotation $[\boldsymbol{R}]$ und der Translation $\boldsymbol{T}$ durch Einsetzen von Gl. (3.5) in Gl. (3.8) zu:

$$\boldsymbol{B}'_{pos} = \{\boldsymbol{G} + \boldsymbol{T} \;,\; [\boldsymbol{R}] \cdot [\boldsymbol{R}]_0\} = \{\boldsymbol{G}' \;,\; [\boldsymbol{R}] \cdot [\boldsymbol{R}]_0\}. \tag{3.9}$$

Aufgrund der bisher getroffenen Annahmen im Quellenmodell R3D kann die ortsabhängige Bilddifferenz an der Stelle $\boldsymbol{h} = (h_x, h_y)^T$ zwischen den beiden realen Bildern s_k und s_{k+1} als Funktion der Bewegungs– und Formparameter der Objekte beschrieben werden

$$fd(\boldsymbol{h}) = s_{k+1}(\boldsymbol{h}) - s_k(\boldsymbol{h}) = s_k\big(f(\boldsymbol{B}, \boldsymbol{F}, \boldsymbol{h})\big) - s_k(\boldsymbol{h}) \tag{3.10}$$

mit den Ortsvektoren der Kontrollpunkte als Formparameter

$$\boldsymbol{F} = \left\{\boldsymbol{F}_0, \boldsymbol{F}_1, .., \boldsymbol{F}_m = \left\{\boldsymbol{P}_m^{(0)}, \boldsymbol{P}_m^{(1)}, .., \boldsymbol{P}_m^{(N-1)}\right\}, .., \boldsymbol{F}_{\mathcal{M}-1}\right\} \tag{3.11}$$

und den Bewegungsparametern

$$\boldsymbol{B} = \left\{\boldsymbol{B}_0, \boldsymbol{B}_1, .., \boldsymbol{B}_m = \left\{T_X^{(m)}, T_Y^{(m)}, T_Z^{(m)}, R_X^{(m)}, R_Y^{(m)}, R_Z^{(m)}\right\}, .., \boldsymbol{B}_{\mathcal{M}-1}\right\}. \tag{3.12}$$

Dabei ist $\mathcal{M}$ die Anzahl der sich in der realen Szene bewegenden Objekte.

3.4.2 Objektmodell bewegter dreidimensionaler, flexibler Objekte

Ein flexibles Objekt unterscheidet sich von einem starren Objekt dadurch, daß es sich lokal verformen kann. Für das Objektmodell bewegter flexibler dreidimensionaler Objekte gelten daher das in Abschnitt 3.4 beschriebene Oberflächenmodell und das in Abschnitt 3.4.1 erläuterte Bewegungsmodell. Im folgenden Abschnitt erfolgt nun die Definition des Formmodells.

3.4.2.1 Formmodell

Bei diesem Formmodell wird angenommen, daß die realen Objekte flexibel sind. Als flexibles Objekt versteht man hierbei ein Objekt mit lokalen Verschiebungen auf der Objektoberfläche[63][65]. Die Verschiebungen auf der Objektoberfläche werden durch die Verschiebung der Kontrollpunkte in einer tangentialen Oberflächenebene modelliert (Bild 3.7). Für die

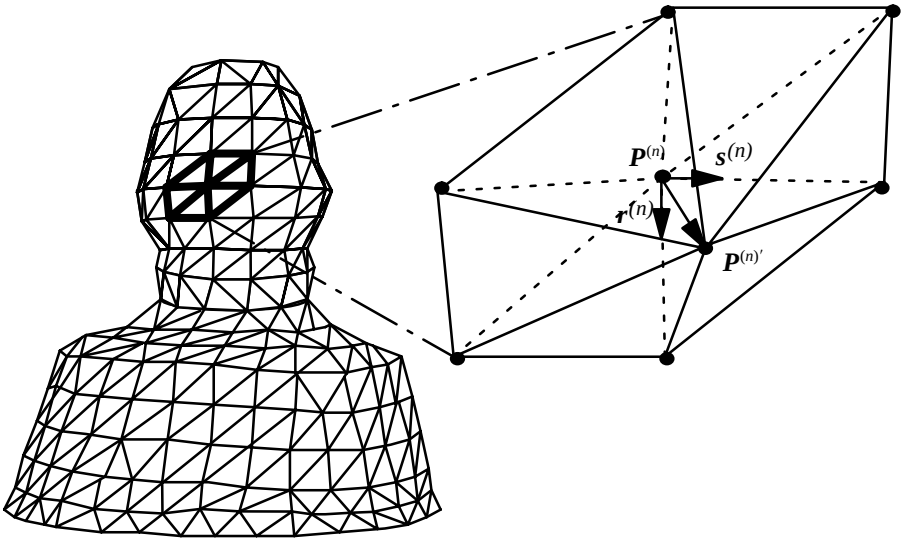


Bild 3.7 Flexible Verschiebung eines Kontrollpunkts $P^{(n)}$ mit dem Ortsvektor $\boldsymbol{P}^{(n)}$ nach $\boldsymbol{P}^{(n)'}$. Die Einheitsvektoren $\boldsymbol{r}^{(n)}$ und $\boldsymbol{s}^{(n)}$ spannen die Ebene auf, in der $\boldsymbol{P}^{(n)}$ verschoben wird. Schwarze Punkte markieren die Kontrollpunkte des Dreiecksnetzes.

Verschiebung eines Kontrollpunktes mit dem Ortsvektor $\boldsymbol{P}^{(n)}$ wird eine Tangentialebene durch die orthogonale Einheitsvektoren $\boldsymbol{r}^{(n)}$ und $\boldsymbol{s}^{(n)}$ aufgespannt. Die lokale Verschiebung eines Kontrollpunktes von $\boldsymbol{P}^{(n)}$ nach $\boldsymbol{P}^{(n)'}$ ergibt sich zu

$$\boldsymbol{P}^{(n)'} = \boldsymbol{P}^{(n)} + s_r^{(n)}\boldsymbol{r}^{(n)} + s_s^{(n)}\boldsymbol{s}^{(n)}, \tag{3.13}$$

wobei die Parameter der Verschiebung $(s_r^{(n)}, s_s^{(n)})^T$ die Größe der lokalen Verschiebung bestimmen. Ein flexibles Objekt mit N Kontrollpunkten benötigt gemäß Gl. (3.7) 6 Parameter zur Beschreibung seiner Position $\boldsymbol{B}_{pos}$ im Weltkoordinatensystem und gemäß Gl. (3.13) $2N$ Parameter zur Beschreibung der lokalen Verschiebungen. Es besitzt daher $6+2N$ Freiheitsgrade. Die Verschiebungsparameter $(s_r^{(n)}, s_s^{(n)})^T$ werden hier den Formparametern zugeordnet.

Aufgrund der bisher getroffenen Annahmen des Quellenmodells kann die ortsabhängige Bilddifferenz an der Stelle $\boldsymbol{h} = (h_x, h_y)^T$ zwischen den beiden realen Bildern s_k und s_{k+1} als Funktion der Bewegungs– und Formparameter der Objekte beschrieben werden

$$fd(\boldsymbol{h}) = s_{k+1}(\boldsymbol{h}) - s_k(\boldsymbol{h}) = s_k\big(f(\boldsymbol{B}, \boldsymbol{F}, \boldsymbol{h})\big) - s_k(\boldsymbol{h}) \tag{3.14}$$

mit den Ortsvektoren und Verschiebungsvektoren der Kontrollpunkte als Formparameter

$$\boldsymbol{F} = \left\{\boldsymbol{F}_0, \boldsymbol{F}_1, .., \boldsymbol{F}_m = \left\{\boldsymbol{P}_m^{(0)}, .., \boldsymbol{P}_m^{(N-1)}, s_{r,m}^{(0)}, .., s_{r,m}^{(N-1)}, s_{s,m}^{(0)}, .., s_{s,m}^{(N-1)}\right\}, .., \boldsymbol{F}_{\mathcal{M}_b - 1}\right\} \tag{3.15}$$

und mit den Bewegungsparametern

$$\boldsymbol{B} = \left\{\boldsymbol{B}_0, \boldsymbol{B}_1, .., \boldsymbol{B}_m = \left\{T_X^{(m)}, T_Y^{(m)}, T_Z^{(m)}, R_X^{(m)}, R_Y^{(m)}, R_Z^{(m)}\right\}, .., \boldsymbol{B}_{\mathcal{M}_b - 1}\right\}. \tag{3.16}$$

3.4.3 Objektmodell bewegter dreidimensionaler, gegliederter Objekte

Ein gegliedertes Objekt unterscheidet sich von einem starren Objekt dadurch, daß es aus einer endlichen Anzahl von starren Objektkomponenten besteht, die durch Gelenke, hier Kugelgelenke, untereinander gekoppelt sind. Jede Objektkomponente wird dabei durch ihre eigenen Bewegungs–, Form– und Texturparameter beschrieben. Ein Kugelgelenk schränkt die Relativbewegung zweier gekoppelter Objektkomponenten so ein, daß eine Objektkomponente relativ zur gekoppelten Objektkomponente nur um die Position des Kugelgelenkes rotieren kann. In der Literatur wird ein gegliedertes Objekt häufig als ”Mehrkörpersystem” bezeichnet[49][70][74][78][86][90].

Für das Objektmodell dreidimensionaler gegliederter Objekte gilt das in Abschnitt 3.4 beschriebene Oberflächenmodell. In den folgenden Abschnitten erfolgt nun die noch ausstehende Definition des Form– und Bewegungsmodells.

3.4.3.1 Formmodell

Bei diesem Formmodell wird angenommen, daß die realen Objekte gegliedert sind und Baumstruktur besitzen. Aufgrund der Annahme einer Baumstruktur sind die $\mathcal{N}$ Objektkomponenten eines gegliederten Objektes durch genau $\mathcal{N}$–1 Kugelgelenke miteinander verbunden. Ein Kugelgelenk verbindet dabei genau zwei Objektkomponenten und umgekehrt existiert maximal ein Kugelgelenk zwischen zwei Objektkomponenten. Ferner existiert nur ein einziger Weg von einer Objektkomponente zu einer anderen beliebigen Objektkomponente im gegliederten Objekt. Die Form eines gegliederten Objektes mit $\mathcal{N}$ Objektkomponenten und $\mathcal{N}$–1 Kugelgelenken wird durch ein gegliedertes Modellobjekt mit $\mathcal{N}$ Modellobjektkomponenten $\mathcal{C}^{(\delta)}$, δ=0,..,$\mathcal{N}$–1, und $\mathcal{N}$–1 Modellkugelgelenken $J^{(\xi)}$, ξ=1,..,$\mathcal{N}$–1, beschrieben. Die Form einer starren Modellobjektkomponente $\mathcal{C}^{(\delta)}$ wird durch ein starres Dreiecksnetz dargestellt, das durch N_δ Kontrollpunkte $P_\delta^{(n_\delta)}$, n_δ=0,..,$N_\delta - 1$, aufgespannt wird (Bild 3.8). Dabei wird

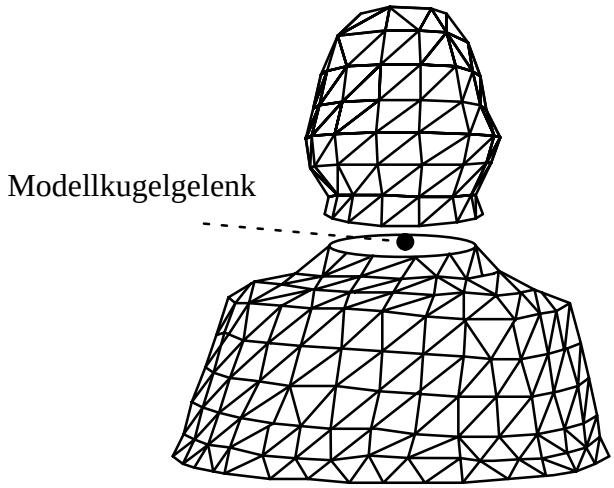


Bild 3.8 Gegliedertes Modellobjekt mit zwei starren Modellobjektkomponenten. Die Form einer starren Modellobjektkomponente wird durch ein starres Dreiecksnetz dargestellt. Die Eckpunkte der Dreiecke einer Objektkomponente stellen ihre Kontrollpunkte dar. Die Modellobjektkomponenten sind durch ein Modellkugelgelenk untereinander verbunden. Ein Modellkugelgelenk wird durch einen Punkt beschrieben, der zu beiden gekoppelten Objektkomponenten gehört.

der Ortsvektor $\boldsymbol{P}_{\delta}^{(n_\delta)^*}$ eines jeden Kontrollpunktes $P_{\delta}^{(n_\delta)}$ einer Objektkomponente $\mathcal{C}^{(\delta)}$ im Koordinatensystem $(\mathcal{X}^{(\delta)}, \mathcal{Y}^{(\delta)}, \mathcal{Z}^{(\delta)})$ dieser Objektkomponente dargestellt. Ein Modellkugelgelenk wird durch einen zu beiden gekoppelten Objektkomponenten gehörenden Punkt $J^{(\xi)}$ beschrieben, der sich zwischen beiden Objektkomponenten befindet (Bild 3.8) und im Koordinatensystem der innenliegenden der gekoppelten Objektkomponenten, der sog. Bezugsobjektkomponente des Kugelgelenkes $J^{(\xi)}$, dargestellt wird (Bild 3.9). Dabei ist die innenliegende Objektkomponente zweier durch ein Kugelgelenk $J^{(\xi)}$ gekoppelter Objektkomponenten die Objektkomponente, die im Baum näher an der Wurzelobjektkomponente liegt. Als Wurzelobjektkomponente wird hier die größte Objektkomponente eines gegliederten Objektes definiert. Sie wird als $\mathcal{C}^{(0)}$ gekennzeichnet. In Bild 3.9 ist die Bezugsobjektkomponente des Kugelgelenkes $J^{(1)}$ die Wurzelobjektkomponente des gegliederten Objektes. Der Ursprung des Koordinatensystems der Wurzelobjektkomponente wird in einem beliebigen Punkt $J^{(0)}$ dieser Objektkomponente mit dem Ortsvektor $\boldsymbol{J}^{(0)}$ bezüglich des Weltkoordinatensystems definiert. Die außenliegende Objektkomponente zweier durch ein Kugelgelenk $J^{(\xi)}$ gekoppelter Objektkomponenten wird hier als Relativobjektkomponente des Kugelgelenkes $J^{(\xi)}$ bezeichnet. Der Ursprung des Koordinatensystems dieser Objektkomponente wird im Kugelgelenk selbst definiert (Bild 3.9). Da die Position des Ursprunges des Koordinatensystems der Relativobjektkomponente im Weltkoordinatensystem durch die Position der korres-

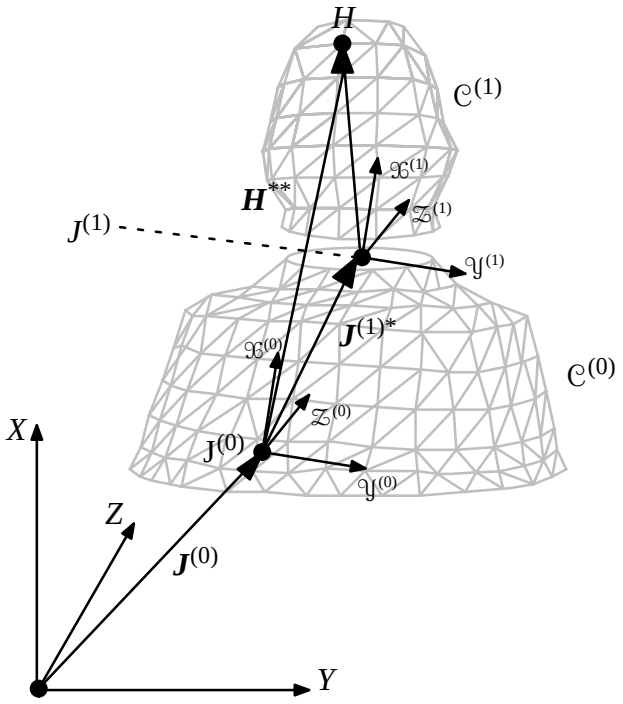


Bild 3.9 Ein Beispiel eines gegliederten Objektes mit zwei Objektkomponenten und einem Kugelgelenk. Die Objektkomponenten $\mathcal{C}^{(0)}$ und $\mathcal{C}^{(1)}$ stellen die Bezugs– und Relativobjektkomponenten des Kugelgelenkes $J^{(1)}$ dar. Die Position $\boldsymbol{J}^{(1)*}$ des Kugelgelenkes wird im Koordinatensystem $(\mathcal{X}^{(0)}, \mathcal{Y}^{(0)}, \mathcal{Z}^{(0)})$ der Bezugsobjektkomponente dargestellt. Die Bezugsobjektkomponente ist in diesem Beispiel auch die Wurzelobjektkomponente des gegliederten Objektes. Der Ursprung des Koordinatensystems $(\mathcal{X}^{(1)}, \mathcal{Y}^{(1)}, \mathcal{Z}^{(1)})$ der Relativobjektkomponente $\mathcal{C}^{(0)}$ wird im Kugelgelenk $J^{(1)}$ definiert. Der Ortsvektor $\boldsymbol{J}^{(0)}$ stellt den Ursprung des Koordinatensystems der Wurzelobjektkomponente $\mathcal{C}^{(1)}$ bezüglich des Weltkoordinatensystems dar. $\boldsymbol{H}^{**}$ stellt den Ortsvektor eines beliebigen Punktes H der Relativobjektobjektkomponente $\mathcal{C}^{(1)}$ bezüglich des Koordinatensystems der Bezugsobjektkomponente $\mathcal{C}^{(0)}$ dar.

pondierenden Bezugsobjektkomponente entsprechend Gl. (3.7) im Weltkoordinatensystem gegeben ist, besitzt die Relativobjektkomponente nur 3 Freiheitsgrade. Im Gegensatz zu einem starren Objekt, das 6 Freiheitsgrade besitzt, hat ein gegliedertes Objekt mit $\mathcal{N}$ Objektkomponenten insgesamt $3*\mathcal{N} + 3$ Freiheitsgrade. Im weiteren kennzeichnet ein ()**, daß der Ortsvektor () eines beliebigen Punktes der Relativobjektkomponente eines Kugelgelenkes $J^{(\xi)}$ im Koordinatensystem der korrespondierenden Bezugsobjektkomponente angegeben ist (Bild 3.9).

Für die Kennzeichnung der Objektkomponenten und Kugelgelenke eines gegliederten Objektes wird hier die reguläre Numerierung[86][89] verwendet (Bild 3.9). So erhalten alle Objektkomponenten eines Astes eines gegliederten Objektes beginnend von der Wurzelobjektkomponente nacheinander aufsteigende Nummern δ. Ein Kugelgelenk, das zwei Objektkomponenten verbindet, erhält die Nummer der korrespondierenden Relativobjektkomponente. Um in Abschnitt 3.4.3.2 eine einfache Formulierung des Bewegungsmodells eines gegliederten Objektes zu ermöglichen, werden hier die Funktionen $c(\xi)^+$ und $c(\xi)^-$ definiert. Sie geben die zur Kennzeichnung der Bezugs– bzw. Relativobjektkomponente eines Kugelgelenkes $J^{(\xi)}$ verwendeten Nummern δ an. Für das Gelenk $J^{(\xi=1)}$ in Bild 3.9 ist der Wert dieser Funktionen $c(\xi = 1)^+ = 0$ bzw. $c(\xi = 1)^- = 1$. Aufgrund der verwendeten regulären Numerierung folgt

$$c(\xi)^- = \xi. \tag{3.17}$$

Die Dreiecksnetze zweier durch ein Kugelgelenk gekoppelter Objektkomponenten werden durch gemeinsame Dreiecke flexibel miteinander verbunden (Bild 3.10). Ein gemeinsames

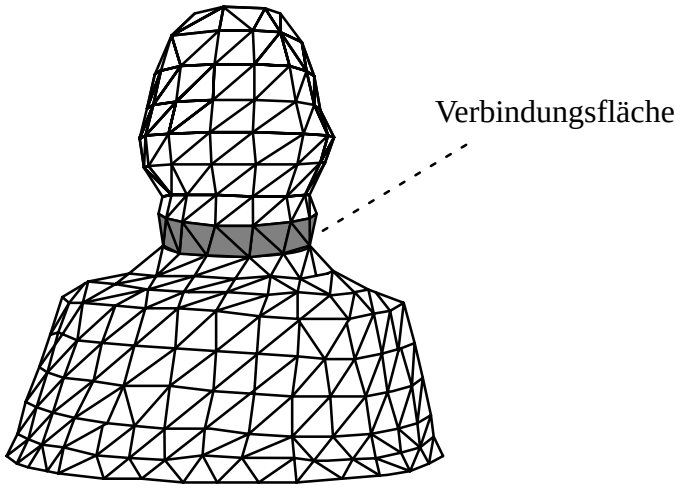


Bild 3.10 Verbindungsfläche zwischen Objektkomponenten. Sie besteht aus flexiblen Dreiecken, die Kontrollpunkte von beiden gekoppelten Objektkomponenten besitzen.

Dreieck besitzt Stützpunkte beider gekoppelter Objektkomponenten. Die gemeinsamen Dreiecke bilden eine Verbindungsfläche zwischen beiden Objektkomponenten. Mit der Relativbewegung der Objektkomponenten verformen sich die Dreiecke der Verbindungsfläche zwischen den Objektkomponenten. Diese Verformung kann unter Kenntnis der Form und Relativbewegung beider gekoppelter Objektkomponenten vollständig beschrieben werden.

3.4.3.2 Bewegungsmodell

Während sich die Wurzelobjektkomponente frei im dreidimensionalen Raum bewegen kann, kann die Relativobjektkomponente eines jeden Kugelgelenkes relativ zu der zugehörigen Bezugsobjektkomponente nur um die Position des Kugelgelenkes rotieren. Bei diesem Quellenmodell wird daher die dreidimensionale Bewegung eines gegliederten Objektes zunächst durch eine Rotation $[\boldsymbol{R}^{(0)}]$ und eine Translation $\boldsymbol{T}^{(0)}$ des Koordinatensystems der Wurzelobjektkomponente $\mathcal{C}^{(0)}$ im Weltkoordinatensystem beschrieben. Dann wird, beginnend von der Wurzelobjektkomponente, nacheinander die Bewegung der Relativobjektkomponente $\mathcal{C}^{(c(\xi)^-)}$ eines jeden Kugelgelenkes $J^{(\xi)}$ relativ zum Koordinatensystem der zugehörigen Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ durch eine Rotation $[\boldsymbol{R}^{(c(\xi)^-)}]$ ihres Koordinatensystems beschrieben (Bild 3.11).

Ein Punkt $H^{(0)}$ der Wurzelobjektkomponente $\mathcal{C}^{(0)}$ mit dem Ortsvektor $\boldsymbol{H}^{(0)}$ bewegt sich nach $\boldsymbol{H}^{(0)'}$ entsprechend Gl. (3.8)

$$\boldsymbol{H}^{(0)'} = [\boldsymbol{R}^{(0)}] \cdot (\boldsymbol{H}^{(0)} - \boldsymbol{J}^{(0)}) + \boldsymbol{J}^{(0)} + \boldsymbol{T}^{(0)}, \tag{3.18}$$

wobei $\boldsymbol{J}^{(0)}$ den Ursprung des Koordinatensystems der Wurzelobjektkomponente darstellt (Bild 3.11). Ein Punkt $H^{(c(\xi)^-)}$ der Relativobjektkomponente $\mathcal{C}^{(c(\xi)^-)}$ mit dem Ortsvektor $\boldsymbol{H}^{(c(\xi)^-)**}$ bezüglich des Koordinatensystems der Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ bewegt sich nach $\boldsymbol{H}^{(c(\xi)^-)**'}$ gemäß

$$\boldsymbol{H}^{(c(\xi)^-)**'} = [\boldsymbol{R}^{(c(\xi)^-)}] \cdot (\boldsymbol{H}^{(c(\xi)^-)**} - \boldsymbol{J}^{(\xi)*}) + \boldsymbol{J}^{(\xi)*}, \tag{3.19}$$

wobei $\boldsymbol{J}^{(\xi)*}$ die Position des Kugelgelenkes bezogen auf das Koordinatensystems der Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ darstellt (Bild 3.12).

Die beobachtete Translation $\boldsymbol{T}^{(c(\xi)^-)}$ der Relativobjektkomponente $\mathcal{C}^{(c(\xi)^-)}$ eines Kugelgelenkes $J^{(\xi)}$ im Weltkoordinatensystem ist von der Rotation $[\boldsymbol{R}^{(c(\xi)^+)}]$ und Translation $\boldsymbol{T}^{(c(\xi)^+)}$ der korrespondierenden Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ im Weltkoordinatensystem abhängig. Diese Abhängigkeit läßt sich dadurch bestimmen, daß das Kugelgelenk $J^{(\xi)}$ mit dem Ortsvektor $\boldsymbol{J}^{(\xi)}$ bezüglich des Weltkoordinatensystems die Bewegungsgleichung entsprechend (3.8) beider gekoppelter Objektkomponenten

$$\boldsymbol{H}^{(c(\xi)^+)'} = [\boldsymbol{R}^{(c(\xi)^+)}] \cdot (\boldsymbol{H}^{(c(\xi)^+)} - \boldsymbol{J}^{(c(\xi)^+)}) + \boldsymbol{J}^{(c(\xi)^+)} + \boldsymbol{T}^{(c(\xi)^+)} \tag{3.20}$$

$$\boldsymbol{H}^{(c(\xi)^-)'} = [\boldsymbol{R}^{(c(\xi)^-)}] \cdot (\boldsymbol{H}^{(c(\xi)^-)} - \boldsymbol{J}^{(\xi)}) + \boldsymbol{J}^{(\xi)} + \boldsymbol{T}^{(c(\xi)^-)} \tag{3.21}$$

erfüllen soll, d.h.

$$\begin{aligned}&[\boldsymbol{R}^{(c(\xi)^+)}] \cdot (\boldsymbol{J}^{(\xi)} - \boldsymbol{J}^{(c(\xi)^+)}) + \boldsymbol{J}^{(c(\xi)^+)} + \boldsymbol{T}^{(c(\xi)^+)} = \\ &[\boldsymbol{R}^{(c(\xi)^-)}] \cdot (\boldsymbol{J}^{(\xi)} - \boldsymbol{J}^{(\xi)}) + \boldsymbol{J}^{(\xi)} + \boldsymbol{T}^{(c(\xi)^-)}.\end{aligned} \tag{3.22}$$

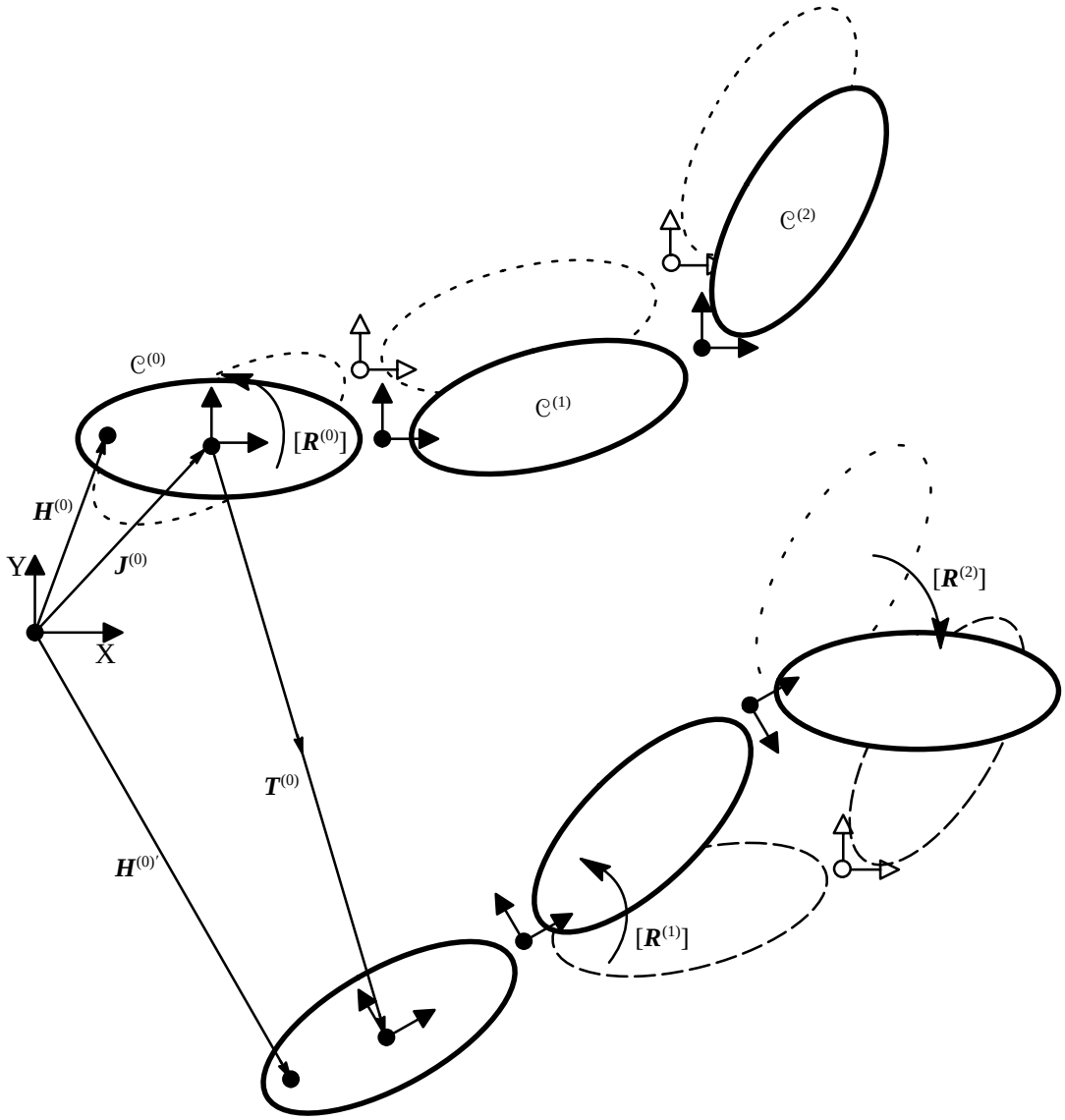


Bild 3.11 Bewegung eines gegliederten Objektes bestehend aus drei Objektkomponenten und 2 Kugelgelenken. In diesem Beispiel bewegen sich die Objektkomponenten nur parallel zur Kamerabildebene. Die Bewegung wird zunächst durch eine Rotation $[\boldsymbol{R}^{(0)}]$ und Translation $\boldsymbol{T}^{(0)}$ des Koordinatensystems der Wurzelobjektkomponente $\mathcal{C}^{(0)}$ im Weltkoordinatensystem, dann durch eine Rotation $[\boldsymbol{R}^{(1)}]$ des Koordinatensystems der Objektkomponente $\mathcal{C}^{(1)}$ und anschließend eine Rotation $[\boldsymbol{R}^{(2)}]$ des Koordinatensystems der Objektkomponente $\mathcal{C}^{(2)}$ beschrieben. - - - - - zeigt das gegliederte Objekt nach der Rotation $[\boldsymbol{R}^{(0)}]$, ——— nach der Rotation $[\boldsymbol{R}^{(0)}]$ und Translation $\boldsymbol{T}^{(0)}$, und - - - nach der Rotation $[\boldsymbol{R}^{(0)}]$,Translation $\boldsymbol{T}^{(0)}$und Rotation $[\boldsymbol{R}^{(1)}]$.

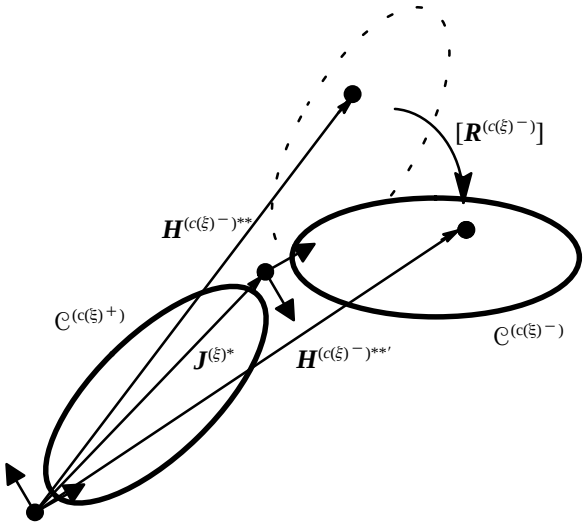


Bild 3.12 Bewegung der Relativobjektkomponente $\mathcal{C}^{(c(\xi)^-)}$ relativ zu der zugehörigen Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ gemäß Gl. (3.19). Das Koordinatensystem der Relativobjektkomponente wird mit der Rotationsmatrix $[\boldsymbol{R}^{(c(\xi)^-)}]$ gedreht. Der Punkt H der Relativobjektkomponente mit dem Ortsvektor $\boldsymbol{H}^{(c(\xi)^-)**}$ bezüglich des Koordinatensystems der Bezugsobjektkomponente bewegt sich dabei nach $\boldsymbol{H}^{(c(\xi)^-)**'}$ gemäß Gl. (3.19).

Diese Bedingung wird häufig in der Literatur als Gelenkbindung bezeichnet. $\boldsymbol{T}^{(c(\xi)^-)}$ ergibt sich dann zu

$$\boldsymbol{T}^{(c(\xi)^-)} = [\boldsymbol{R}^{(c(\xi)^+)}] \cdot (\boldsymbol{J}^{(\xi)} - \boldsymbol{J}^{(c(\xi)^+)}) + \boldsymbol{J}^{(c(\xi)^+)} + \boldsymbol{T}^{(c(\xi)^+)} - \boldsymbol{J}^{(\xi)}. \tag{3.23}$$

Dabei stellt $\boldsymbol{J}^{(c(\xi)^+)}$ den Ursprung des Koordinatensystems der Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ des Kugelgelenkes $J^{(\xi)}$ dar. Die drei Komponenten $T_X^{(c(\xi)^-)}$, $T_Y^{(c(\xi)^-)}$ und $T_Z^{(c(\xi)^-)}$ des Translationsvektors $\boldsymbol{T}^{(c(\xi)^-)}$ werden hier als abhängige Bewegungsparameter der Relativobjektkomponente $\mathcal{C}^{(c(\xi)^-)}$ eines Kugelgelenkes $J^{(\xi)}$ bezeichnet. Sie lassen sich mit Gl. (3.23) berechnen. Die drei Winkel $R_X^{(c(\xi)^-)}$, $R_Y^{(c(\xi)^-)}$ und $R_Z^{(c(\xi)^-)}$ der Rotationsmatrix $[\boldsymbol{R}^{(c(\xi)^-)}]$ stellen dabei die unabhängigen Bewegungsparameter dieser Objektkomponente dar.

Aufgrund der bisher getroffenen Annahmen des Quellenmodells kann die ortsabhängige Bilddifferenz an der Stelle $\boldsymbol{h} = (h_x, h_y)^T$ zwischen den beiden realen Bildern s_k und s_{k+1} als Funktion der Bewegungs– und Formparameter der Objektkomponenten interpretiert werden

$$fd(\boldsymbol{h}) = s_{k+1}(\boldsymbol{h}) - s_k(\boldsymbol{h}) = s_k\big(f(\boldsymbol{B},\ \boldsymbol{F},\ \boldsymbol{h})\big) - s_k(\boldsymbol{h}) \tag{3.24}$$

mit den Ortsvektoren der Kontrollpunkte und Gelenke als Formparameter

$$\boldsymbol{F} = \left\{\boldsymbol{F}_0, \boldsymbol{F}_1, .., \boldsymbol{F}_m = \left\{\boldsymbol{F}_m^{(0)}, \boldsymbol{F}_m^{(1)}, .., \boldsymbol{F}_m^{(\delta)} = \left\{\boldsymbol{P}_{m,\delta}^{(0)}, \boldsymbol{P}_{m,\delta}^{(1)}, .., \boldsymbol{P}_{m,\delta}^{(N_\delta - 1)}\right\}, .., \boldsymbol{F}_m^{(\mathcal{N}-1)}, \boldsymbol{J}_m^{(1)}, \boldsymbol{J}_m^{(2)}, .., \boldsymbol{J}_m^{(\mathcal{N}-1)}\right\}, .., \boldsymbol{F}_{\mathcal{M}_\flat - 1}\right\} \tag{3.25}$$

und mit den Bewegungsparametern

$$\boldsymbol{B} = \left\{\boldsymbol{B}_0, \boldsymbol{B}_1, .., \boldsymbol{B}_m = \left\{\boldsymbol{B}_m^{(0)} = \left\{T_{X,m}^{(0)}, T_{Y,m}^{(0)}, T_{Z,m}^{(0)}, R_{X,m}^{(0)}, R_{Y,m}^{(0)}, R_{Z,m}^{(0)}\right\}, \boldsymbol{B}_m^{(1)}, \boldsymbol{B}_m^{(2)}, .., \boldsymbol{B}_m^{(\delta)} = \left\{R_{X,m}^{(\delta)}, R_{Y,m}^{(\delta)}, R_{Z,m}^{(\delta)}\right\}, .., \boldsymbol{B}_m^{(\mathcal{N}-1)}\right\}, .., \boldsymbol{B}_{\mathcal{M}_\flat - 1}\right\}. \tag{3.26}$$

Dabei stellt $\boldsymbol{B}_m^{(0)}$ den Bewegungsparametersatz der Wurzelobjektkomponente des gegliederten Objektes *m* dar.

4. Bildanalyse für bewegte dreidimensionale, gegliederte Objekte

Um einen objektbasierten Analyse–Synthese Coder zu optimieren, wurde im Kapitel 3 das Quellenmodell bewegter, dreidimensionaler gegliederter Objekte vorgestellt. Gemäß diesem Quellenmodell wird jedes reale dreidimensionale gegliederte Objekt durch ein dreidimensionales gegliedertes Modellobjekt einer Modellwelt modelliert. Ein gegliedertes Modellobjekt *m* besteht aus einer endlichen Anzahl von Modellobjektkomponenten $\mathbb{C}_m^{(\delta)}$, $\delta=0,..,\mathcal{N}-1$, die durch Modellkugelgelenke $J_m^{(\xi)}$, $\xi=1,..,\mathcal{N}-1$, untereinander gekoppelt sind. Jede Modellobjektkomponente $\mathbb{C}_m^{(\delta)}$ wird durch ihren eigenen Satz von Bewegungs–, Form– und Texturparametern $\boldsymbol{F}_m^{(\delta)}$, $\boldsymbol{B}_m^{(\delta)}$ bzw. $\boldsymbol{C}_m^{(\delta)}$ beschrieben. Ein Modellkugelgelenk $J_m^{(\xi)}$ wird als Bestandteil der Form eines gegliederten Modellobjektes interpretiert und ihre Position $\boldsymbol{J}_m^{(\xi)}$ bezogen auf das Koordinatensystem der zugehörigen Bezugsmodellobjektkomponente dargestellt. In diesem Kapitel wird eine automatische Bildanalyse für die Schätzung der Parameter der dreidimensionalen Form der Modellobjektkomponenten eines gegliederten Modellobjektes einschließlich der Position der Modellkugelgelenke sowie für die Schätzung der Parameter der dreidimensionalen Bewegung einer jeden Modellobjektkomponente eines gegliederten Modellobjektes entwickelt. Hierbei werden als Parameter der dreidimensionalen Form $\boldsymbol{F}_m$ eines gegliederten Modellobjektes *m* die Parameter der dreidimensionalen Form aller seiner Modellobjektkomponenten und die Position aller seiner Modellkugelgelenke interpretiert. Die Parameter der dreidimensionalen Bewegung $\boldsymbol{B}_m$ eines gegliederten Modellobjektes *m* bestehen aus den Parametern der dreidimensionalen Bewegung aller seiner Modellobjektkomponenten. Anschließend setzen sich die Texturparameter $\boldsymbol{C}_m$ eines gegliederten Modellobjektes *m* aus den Texturparametern aller seiner Modellobjektkomponenten zusammen.

Für die Schätzung der Parameter unterteilt die Bildanalyse das aktuelle reale Bild in Objekte der Kamerabildebene. Hierbei stellt ein Objekt der Kamerabildebene den Bildbereich dar, auf den ein sich bewegendes, reales, gegliedertes Objekt abgebildet wird. Jedem Bildbereich eines gegliederten Objektes wird dann ein dreidimensionales gegliedertes Modellobjekt zugeordnet. In Abschnitt 4.1 wird ein Konzept für die Bildanalyse vorgestellt und ihre Teilaufgaben kurz beschrieben. Auf die Schätzung der Parameter der dreidimensionalen Form einer jeden Modellobjektkomponente eines gegliederten Modellobjektes wird in Abschnitt 4.2 eingegangen. Das Verfahren zur Schätzung der Position eines jeden Modellkugelgelenkes eines gegliederten Modellobjektes wird in Abschnitt 4.2.4 beschrieben. Das Verfahren zur Schätzung der Parameter der dreidimensionalen Bewegung der Modellobjektkomponenten eines gegliederten Modellobjektes wird in Abschnitt 4.3 beschrieben.

4.1 Konzept der objektbasierten Bildanalyse

Das Konzept der Bildanalyse wird mit Hilfe von Bild 4.1 beschrieben. Die Bildanalyse hat die Aufgabe, die aktuellen Parameter für jedes Modellobjekt zu schätzen. Zur Schätzung werden die bereits an den Empfänger übertragenen und in dem Parameterspeicher vorliegenden

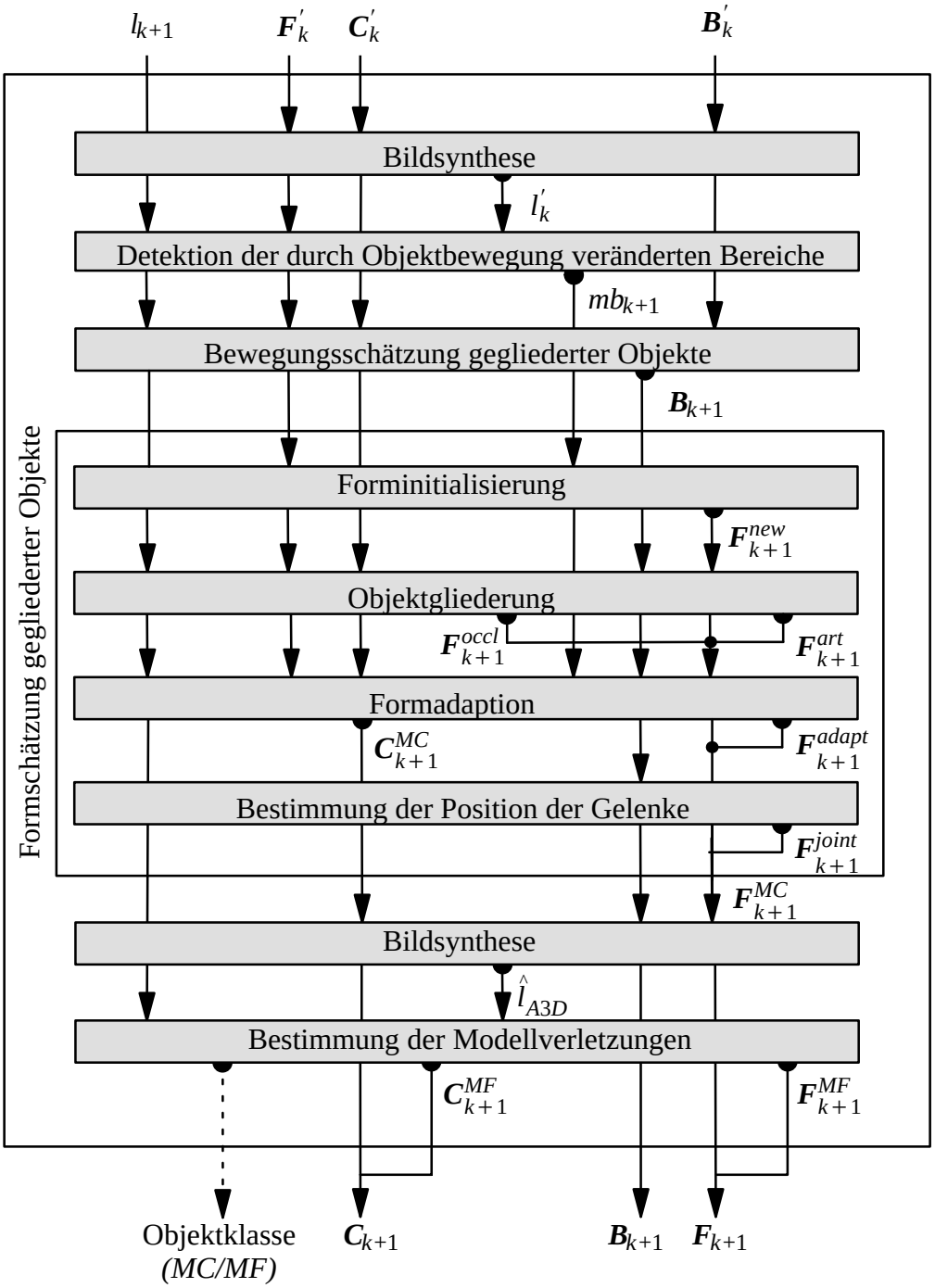


Bild 4.1 Bildanalyse für das Quellenmodell A3D. Halbkreise geben an, welche Parameter in einem Block erzeugt werden, Pfeilspitzen geben die ausgewerteten Parameter an.

Parametersätze sowie die Luminanzkomponente l_{k+1} des aktuellen Bildes i_{k+1} verwendet. Geht man davon aus, daß am Anfang der Bildsequenz für jedes sich bewegende gegliederte Objekt ein dreidimensionales starres Modellobjekt durch Forminitialisierung gemäß Abschnitt 4.2.1 erzeugt wurde, werden für die Bildanalyse basierend auf dem Quellenmodell A3D die folgenden Schritte zum Zeitpunkt *k+1* erforderlich: Bildsynthese, Detektion der durch Objektbewegung veränderten Bildbereiche, Bewegungsschätzung, Forminitialisierung, Objektgliederung, Formadaption, Bestimmung der Position der Gelenke und Bestimmung der Modellverletzungsbereiche.

Die Bildsynthese berechnet unter Verwendung der für jedes Objekt ermittelten Modellobjektparameter ein Modellbild (siehe Kapitel 5). In der Literatur wird dieses Modellbild häufig als synthetisches Bild oder als Rekonstruktion bezeichnet. Zu Beginn der Bildanalyse wird das synthetisierte Bild $l_k^{'}$ unter Verwendung der bereits an den Empfänger übertragenen Parametersätze $\boldsymbol{F}_k^{'}$, $\boldsymbol{C}_k^{'}$ und $\boldsymbol{B}_k^{'}$ erzeugt.

Durch Auswertung der Luminanzkomponenten des synthetisierten Bildes $l_k^{'}$ und des realen Bildes l_{k+1} erfolgt nun eine Detektion der durch Objektbewegung veränderten Bildbereiche (siehe Abschnitt 4.2.3). Die resultierende binäre Maske mb_{k+1} trägt die Kennzahl 1 (geändert) für Bildbereiche, deren Änderungen durch die Bewegung eines Objektes hervorgerufen worden sind. Die übrigen Bildbereiche werden mit der Kennzahl 0 (statisch) markiert. Jedes zusammenhängende Gebiet mit einer Kennzahl wird als Region bezeichnet. Die Regionen mit der Kennzahl 1 markieren sowohl die Objektsilhouette zum Zeitpunkt *k+1* als auch die seit dem Zeitpunkt *k* durch die Objektbewegung freigewordenen Bildbereiche.

Parallel zur Detektion der durch Objektbewegung veränderten Bildbereiche erfolgt für die Modellobjekte eine Bewegungsschätzung und –kompensation mit den geschätzten Bewegungsparametern $\boldsymbol{B}_{k+1}$. Bei der Bewegungsschätzung werden die Parameter der dreidimensionalen Bewegung für jede Objektkomponente eines gegliederten Objektes bestimmt (siehe Abschnitt 4.3).

Bei der Forminitialisierung wird für jede Region mit der Kennzahl 1, der noch nicht ein Modellobjekt zugeordnet worden ist, ein dreidimensionales starres Modellobjekt erzeugt. Dies ist dann der Fall, falls sich ein gegliedertes Objekt zum ersten Mal vom Zeitpunkt *k* bis zum aktuellen Zeitpunkt *k+1* bewegt. Die Form des starren Modellobjektes wird durch ein starres Dreiecksnetz dargestellt. Interpretiert man eine Region mit der Kennzahl 1 näherungsweise als eine Objektsilhouette, so wird das starre Dreiecksnetz aus dieser Region unter der Hypothese ellipsoider Form mit Hilfe einer Distanztransformation berechnet (siehe Abschnitt 4.2.1). Anschließend wird durch Aufprojektion des synthetisierten Bildes $l_k^{'}$ dem starren Modellobjekt Textur zugeordnet. $\boldsymbol{F}_{k+1}^{neu}$ stellen die Formparameter der geraden gefundenen sich bewegenden Objekte dar.

Da bei der Forminitialisierung die Form eines gegliederten Objektes durch ein einziges, dreidimensionales, starres Modellobjekt dargestellt wird, wird jedes starre Modellobjekt bei der Objektgliederung in Objektkomponenten aufgeteilt. Dazu werden benachbarte Dreiecke

des starren Modellobjektes, die ähnliche Parameter der dreidimensionalen Bewegung aufweisen, in Objektkomponenten zusammengefaßt (siehe Abschnitt 4.2.2). Die Objektkomponenten bleiben über gemeinsame Dreiecke miteinander flexibel verbunden. Jeder Objektkomponente wird eine Kennzahl zugeordnet. Für jeden Kontrollpunkt eines gegliederten Objektes wird die Kennzahl der korrespondierenden Objektkomponente in definierter Reihenfolge in eine lineare Liste $\boldsymbol{F}_{k+1}^{art}$ eingetragen. Die Liste $\boldsymbol{F}_{k+1}^{art}$ stellt die Objektgliederungsparameter dar. Im Falle gegenseitiger Verdeckung der Objektkomponenten wird bei der Objektgliederung die Form der verdeckenden Objektkomponente durch ein separates, geschlossenes Dreiecksnetz dargestellt und unter Vermeidung von räumlichen Überschneidungen vor den verdeckten Objektkomponenten plaziert. Das Dreiecksnetz wird aus der realen Silhouette der verdeckenden Objektkomponente unter der Hypothese ellipsoider Form bestimmt. Die reale Silhouette einer verdeckenden Objektkomponente wird aus einer lokalen Segmentierung eines Displacementvektorfeldes bestimmt. $\boldsymbol{F}_{k+1}^{occl}$ stellt die Formparameter der geraden gefundenen verdeckenden Objektkomponenten dar.

Da durch die Hypothese ellipsoider Form bei der Netzerzeugung die Form der Objektkomponenten nicht genau beschrieben wird, unterscheiden sich nach einer Rotation der Objektkomponenten im dreidimensionalen Raum die reale Objektsilhouette und die Modellobjektsilhouette. Bei der Formadaption wird daher die dreidimensionale Form der Objektkomponenten mittels der Objektsilhouette zum Zeitpunkt *k+1* angepaßt (siehe Abschnitt 4.2.3). Dazu werden die sichtbaren Stützpunkte der Modellobjektkomponenten an der Modellobjektsilhouette parallel zur Kamerabildebene so verschoben, daß die Objektsilhouette zum Zeitpunkt *k+1* und die Modellobjektsilhouette übereinstimmen. Die Objektsilhouette zum Zeitpunkt *k*+1 sowie die freigewordenen Bildbereiche werden aus einer Segmentierung der zugehörigen geänderten Region in der Maske mb_{k+1} mittels der geschätzten Bewegungsparametern bestimmt. Ferner wird die Form der verdeckenden Modellobjektkomponenten mittels ihrer eigenen Silhouetten zum Zeitpunkt *k*+1 angepaßt. Die Silhouette einer verdekkenden Objektkomponente zum Zeitpunkt *k+1* wird aus einer lokalen Segmentierung eines Displacementvektorfeldes bestimmt. Die Parameter $\boldsymbol{F}_{k+1}^{adapt}$ stellen die Parameter zur Formanpassung aller Objektkomponenten dar. Nach der Formadaption sind die aktuellen Formparameter aller Objektkomponenten bekannt. Zusätzlich werden die Texturparameter der Modellobjektkomponenten um die Texturparameter $\boldsymbol{C}_{k+1}^{FB}$ der freigewordenen Bildbereiche erweitert, so daß die Parameter $\boldsymbol{C}_{k+1}^{MC}$ auch bekannt sind.

Für je zwei benachbarte Objektkomponenten eines gegliederten Objektes wird unter Verwendung der Gelenkbindung einmalig die Position des Kugelgelenkes bestimmt (siehe Abschnitt 4.2.4). $\boldsymbol{F}_{k+1}^{joint}$ stellt die Position der Kugelgelenke dar. Damit ist die Formschätzung abgeschlossen. $\boldsymbol{F}_{k+1}^{MC}$ stellt die aktuellen Formparameter aller Modellobjekte dar.

Mit Hilfe der Parameter $\boldsymbol{F}_{k+1}^{MC}$, $\boldsymbol{C}_{k+1}^{MC}$ und $\boldsymbol{B}_{k+1}$ wird das Bild $\hat{l}_{A3D}$ synthetisiert. Den Abschluß der Bildanalyse für das Quellenmodell A3D bildet die Bestimmung der Modellverletzungen[59][65]. Bildbereiche des Bildes l_{k+1}, die im Bild $\hat{l}_{A3D}$ aufgrund der aktuellen Parameter-

sätze nicht mit ausreichender Qualität synthetisiert werden können, werden als MF–Objekte detektiert und durch ihre zweidimensionalen Form– und Texturparameter $\boldsymbol{F}_{k+1}^{MF}$ bzw. $\boldsymbol{C}_{k+1}^{MF}$beschrieben.

Am Ausgang der Bildanalyse stehen nun die aktuellen Parameter $\boldsymbol{B}_{k+1}$, $\boldsymbol{F}_{k+1}$ und $\boldsymbol{C}_{k+1}$ zur Verfügung. Die Formparameter $\boldsymbol{F}_{k+1}$ setzen sich aus den Formparametern $\boldsymbol{F}_{k+1}^{MC}$ und $\boldsymbol{F}_{k+1}^{MF}$ der MC–Objekte bzw. MF–Objekte zusammen. Die Texturparameter $\boldsymbol{C}_{k+1}$ werden durch die schon übertragenen Parameter $\boldsymbol{C}_{k}^{'}$ der MC–Objekte, die noch zu übertragenden Parameter $\boldsymbol{C}_{k+1}^{FB}$ der freigewordenen Bildbereiche sowie $\boldsymbol{C}_{k+1}^{MF}$ der MF–Objekte festgelegt.

4.2 Schätzung der dreidimensionalen Form gegliederter Objekte

Für die Schätzung der Parameter der dreidimensionalen Form gegliederter Objekte wird ein Verfahren entwickelt, das für jedes gegliederte Objekt die Parameter der dreidimensionalen Form der Objektkomponenten und die Position der Kugelgelenke bezüglich des Koordinatensystems der korrespondierenden Bezugsobjektkomponente schätzt. Das Verfahren zur Schätzung der Parameter der dreidimensionalen Form der Objektkomponenten besteht aus folgenden Schritten: Forminitialisierung, Objektgliederung und Formadaption. Bei der Forminitialisierung wird jedes reale Objekt zunächst durch ein starres Modellobjekt beschrieben (siehe Abschnitt 4.2.1). Die Form dieses starren Modellobjektes wird aus der realen Objektsilhouette unter Verwendung einer Hypothese berechnet. Bei der Objektgliederung wird das starre Modellobjekt in Modellobjektkomponenten aufgeteilt (siehe Abschnitt 4.2.2). Dazu werden die Parameter der dreidimensionalen Bewegung von Dreiecken des starren bewegten Objektes ausgewertet, indem Oberflächenelementen mit ähnlichen Bewegungsparameterwerten zu einer Objektkomponente zusammengefaßt werden. Dabei werden die in vorherigen Bildern erhaltenen Ergebnisse dieser Auswertung mitberücksichtigt. Im Falle gegenseitiger Verdeckung der Objektkomponenten werden verdeckte und verdeckende Objektkomponenten unterschieden und die Form der verdeckenden und verdeckten Objektkomponenten mit separaten Dreiecksnetzen dargestellt. Bei der Formadaption wird die Form der Modellobjektkomponenten fortlaufend an die reale Form der Objektkomponenten angepaßt (siehe Abschnitt 4.2.3). Dazu wird ein Verfahren entwickelt, das die Form der Modellobjektkomponenten mittels der realen Objektsilhouetten anpaßt. Die Objektsilhouette wird mittels einer Änderungsdetektion unter Verwendung der aktuellen Parameter der dreidimensionalen Bewegung der Modellobjektkomponenten bestimmt. Ferner wird die Form der verdeckenden Objektkomponenten mittels ihrer eigenen realen Silhouetten angepaßt. Dabei wird die reale Silhouette einer verdeckenden Objektkomponente aus einer lokalen Segmentierung eines Displacementvektorfeldes bestimmt. Zur Schätzung der Position eines Gelenkes wird anschließend in Abschnitt 4.2.4 ein Verfahren entwickelt, das die Gelenkbindung ausnutzt.

4.2.1 Forminitialisierung

Bei der Forminitialisierung wird jedes reale gegliederte Objekt durch ein starres Modellobjekt beschrieben. Die Form des starren Modellobjektes wird durch ein starres Dreiecksnetz aufgespannt. Dazu wird ein aus der Literatur bekanntes Verfahren verwendet [62][65]. Die Stützpunkte des Dreiecksnetzes bilden die Parameter zur Beschreibung der starren Form. Das starre Dreiecksnetz wird aus der Silhouette des korrespondierenden realen Objektes unter der Hypothese ellipsoider Form mit Hilfe einer Distanztransformation so berechnet, daß die Dreiecke im Raum etwa gleiche Kantenlängen aufweisen. Die Silhouette eines realen Objektes wird mit Hilfe einer Änderungsdetektion zwischen dem zuletzt übertragenen und dem aktuellen realen Bild bestimmt.

In Bild 4.2 werden die in [62][65] vorgeschlagenen fünf Schritte von einer Objektsilhouette (Bild 4.2a) zum starren Modellobjekt (Bild 4.2f) illustriert. In einem ersten Schritt wird auf die Silhouette des Objektes *m* eine aus der Ellipse abgeleitete Distanztransformation angewendet. Diese Distanztransformation ordnet jedem Punkt $\boldsymbol{h}$ der Objektsilhouette abhängig von seinem kleinsten euklidischen Abstand d zum Silhouettenrand folgende symmetrisch zur Bildebene liegende zwei Tiefenwerte:

$$w_{F\pm Z}(d) = F \pm \begin{cases} \frac{1}{\beta}\sqrt{d(b-d)} & \text{für } d < \frac{b}{2} \\ \frac{b}{2\beta} & \text{sonst} \end{cases} \tag{4.1}$$

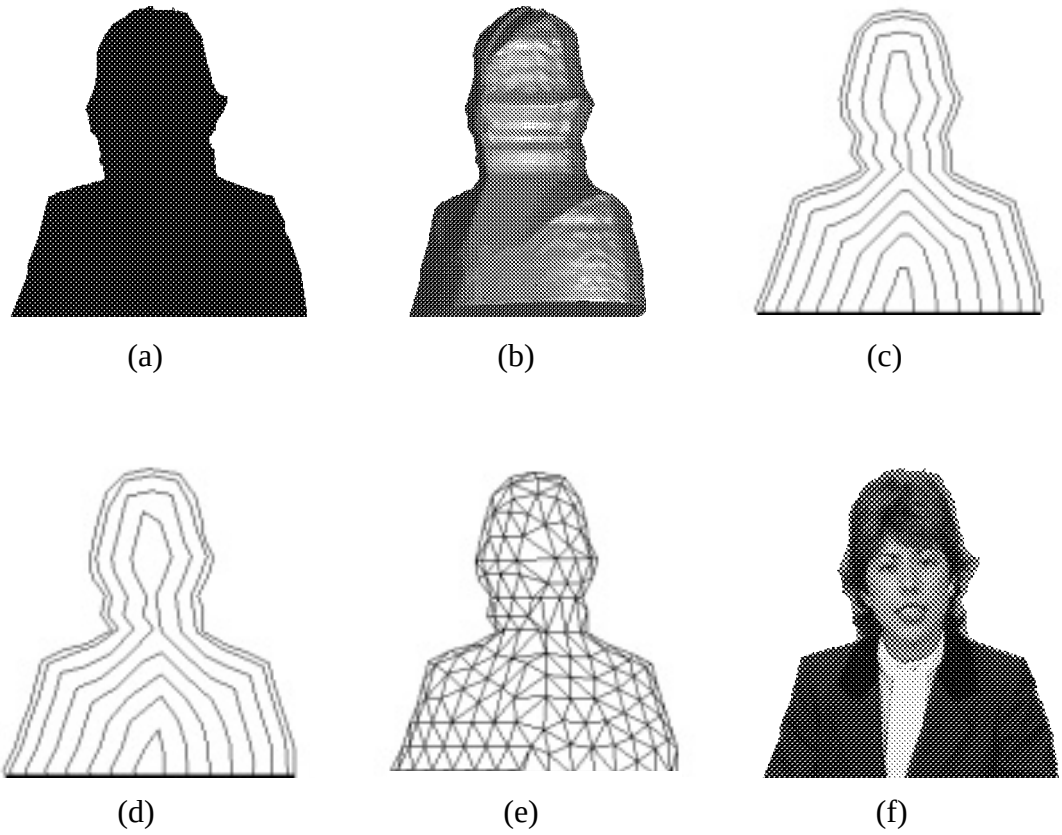
zu, wobei b die Objektbreite, h die Objekttiefe und $\beta = b/h$ darstellen (Bild 4.3a). Dabei wird die Objektbreite b aus der minimalen Ausdehnung der Objektsilhouette bestimmt[65]. Ein Wert β von 1.5 hat sich beim hier verwendeten Bildformat CIF mit einer Bildwiederholfrequenz von 10 Hz für die Modellierung vieler Objetke bewährt[65]. Bild 4.2b zeigt die aus Bild 4.2a erzeugte Form. Sie ist um 30^0 gedreht und beleuchtet dargestellt.

Diese Form wird nun mit Höhenlinien belegt. Alle Punkte einer Höhenlinie haben den gleichen Abstand zur Bildebene. Der Abstand zwischen zwei Höhenlinien wird auf der Oberfläche der Form gemessen. Ihr Abstand d_h ist weitgehend konstant (Bild 4.2c, Bild 4.3b). Beim hier verwendeten Bildformat CIF mit einer Bildwiederholfrequenz von 10 Hz hat sich für die Modellierung vieler Objekte $d_h = 19$ Bildpunkte als geeignet erwiesen[65]. Die Höhenlinien werden durch Polygone approximiert. Hierfür wird das gleiche Gütemaß d^*_{max} verwendet, mit dem später die Codierung der zweidimensionalen Formen erfolgt (vgl. Kapitel 7.3). Das Gütemaß d^*_{max} definiert den maximal erlaubten Abstand zwischen der Höhenlinie und dem die Höhenlinie approximierenden Polygon. Es wird das in [30] bestimmte optimale Gütemaß $d^*_{max} = 1.4$ pel für die Beschreibung der Silhouetten bewegter Objekte verwendet (Bild 4.2d). Falls benachbarte Eckpunkte eines Polygons weiter als $1.5 \cdot d_h$ voneinander entfernt

liegen, werden neue Polygonpunkte auf dem Polygon eingeführt, damit die Dreiecke des Dreiecksnetzes etwa gleiche Kantenlängen haben. Die so erzeugten Polygonpunkte dienen als Kontrollpunkte $\boldsymbol{P}^{(n)}$, über die ein Dreiecksnetz aufgespannt wird (Bild 4.2e).

Zum Abschluß wird das starre Dreiecksnetz in einer Modellwelt plaziert. Dazu wird der Ursprung des Koordinatensystems des Modellobjektes im Schwerpunkt seiner Kontrollpunkte

$$\boldsymbol{G} = (G_X, G_Y, G_Z)^T = \frac{1}{N}\sum_{n=0}^{N-1} \boldsymbol{P}^{(n)} \tag{4.2}$$



(a) (b) (c)

(d) (e) (f)

Bild 4.2 Verarbeitungsschritte von einer Objektsilhouette bis zum Modellobjekt. (a) Objektsilhouette, (b) beleuchtete und um 30^0 gedrehte dreidimensionale Form mit der geforderten Silhouette, (c) Approximation der Form durch Höhenlinien, (d) Polygone zur Approximation der Höhenlinien, (e) Dreiecksnetz mit den Polygonpunkten als Kontrollpunkte, (f) Modellobjekt: Dreiecksnetz mit Texturparametern. (aus [65])

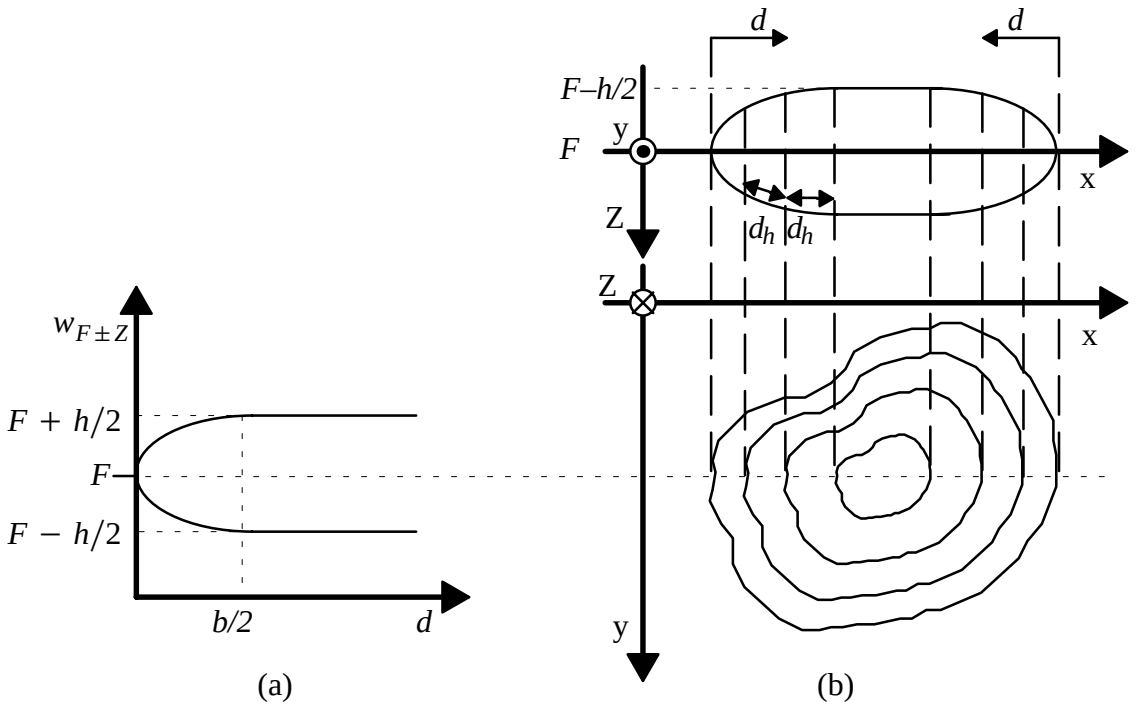


Bild 4.3 Erzeugung einer dreidimensionalen Form symmetrisch zur Bildebene *Z=F*. (a) Distanztransformation gemäß Gl. (4.1), *d* ist der kleinste Abstand zum Rand der Objektsilhouette, *b* die Objektbreite und $\beta=b/h=1.5$. (b) Schnitt durch ein Modellobjekt (oben) und Aufsicht auf die Höhenlinien vom Kamerabrennpunkt aus (unten). Für die Berechnung der Objekttiefe wird der Abstand *d* immer zu dem Objektrandpunkt gemessen, der den kleinsten Abstand hat. (aus [65])

definiert und bis auf die Ebene *Z=2F* verschoben. Dabei wird die Brennweite *F* auf 1000 pel gesetzt[45]. Um zu gewährleisten, daß das neue Modellobjekt nach der Verschiebung bei einer Abbildung in die Bildebene mittels des Kameramodells gemäß Abschnitt 3.1 die geforderte Größe hat, wird jeder Kontrollpunkt $\boldsymbol{P}^{(n)}$ durch die Transformation

$$\boldsymbol{P}^{(n)'} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \boldsymbol{P}^{(n)} + (0, 0, F)^T \tag{4.3}$$

in seine endgültige Position $\boldsymbol{P}^{(n)'}$ verschoben. Auf dieses Modellobjekt wird dann das auch dem Empfänger bekannte Bild s_k' projiziert. Dazu wird das Bild s_k' in die Bildebene der Modellkamera gelegt und jedem sichtbaren Dreieck ein zweidimensionaler dreiecksförmiger Bildausschnitt zugeordnet. Die Lage des zugehörigen Bildausschnittes eines Dreieckes wird durch die Projektion der drei Kontrollpunkte dieses Dreieckes in die Bildebene festgelegt[45]. Texturparameter, die auf diese Weise auf das Modellobjekt *m* projiziert werden, werden als seine Texturparameter $\boldsymbol{C}_m$ gekennzeichent.

Der Hintergrund liegt parallel zur Bildebene in sehr großer Entfernung zur Modellkamera. Dem zugehörigen Objekt HINTERGRUND wird bei der Initialisierung das erste Bild i_0' als Texturparameter zugewiesen.

4.2.2 Objektgliederung

Bei der Objektgliederung wird jedes starre Modellobjekt in Objektkomponenten aufgeteilt. Dazu befindet sich in der Literatur ein erster Ansatz[10], bei dem zunächst die drei Parameter (T_X, T_Y, R_Z) der insgesamt sechs Parameter $(T_X, T_Y, T_Z, R_X, R_Y, R_Z)$ der dreidimensionalen Bewegung eines jeden Dreieckes eines starren Modellobjektes geschätzt werden. Benachbarte Dreiecke mit ähnlichen Bewegungsparametern (T_X, T_Y, R_Z) werden dann zu Oberflächengebieten zusammengefaßt. Die Oberflächengebiete stellen die Objektkomponenten dar. Nachteil dieses Verfahren ist jedoch, daß die Zusammenfassung bei komplexeren Bewegungen der Objektkomponenten, wie z.B. bei einer Rotation um die X– oder Y–Koordinate, aufgrund des verwendeten, einfachen Bewegungsmodells versagt. In diesem Abschnitt wird daher ein Verfahren zur Objektgliederung entwickelt, das anstelle nur der drei Parameter (T_X, T_Y, R_Z) alle sechs Parameter $(T_X, T_Y, T_Z, R_X, R_Y, R_Z)$ der dreidimensionalen Bewegung eines jeden Dreieckes eines starren Modellobjektes auswertet. Zur Verbesserung der Zuverlässigkeit der Schätzung der Parameter der dreidimensionalen Bewegung eines Dreieckes wird in Abschnitt 4.3.2 ein Maximum–Likelihood–Schätzverfahren entwickelt. Weiterhin werden die in der Analyse vorheriger Bilder erhaltenen Ergebnisse der Zusammenfassung von Dreiecken bei der Analyse des aktuellen Bildes mittels eines zeitlichen Gedächtnisses berücksichtigt. Im Falle gegenseitiger Verdeckung der Objektkomponenten werden bei der Objektgliederung zusätzlich die verdeckten und verdeckenden Objektkomponenten unterschieden und die Form der verdeckenden und verdeckten Objektkomponenten mit separaten Dreiecksnetzen dargestellt.

Zur Objektgliederung werden in einem ersten Schritt die Parameter der dreidimensionalen Bewegung der einzelnen Dreiecke gemäß Abschnitt 4.3.2 geschätzt. Dann werden die Dreiecke, für welche ähnliche Parameter geschätzt worden sind, zu Oberflächengebieten zusammengefaßt. Damit ein Dreieck mit den geschätzten Bewegungsparametern $(\hat{T}^{(1)}_{ML,X}, \hat{T}^{(1)}_{ML,Y}, \hat{T}^{(1)}_{ML,Z}, \hat{R}^{(1)}_{ML,X}, \hat{R}^{(1)}_{ML,Y}, \hat{R}^{(1)}_{ML,Z})$ mit einem Nachbardreieck mit den geschätzten Bewegungsparametern $(\hat{T}^{(2)}_{ML,X}, \hat{T}^{(2)}_{ML,Y}, \hat{T}^{(2)}_{ML,Z}, \hat{R}^{(2)}_{ML,X}, \hat{R}^{(2)}_{ML,Y}, \hat{R}^{(2)}_{ML,Z})$ zusammengefaßt wird, müssen die Bedingungen

$$\begin{aligned} \sqrt{(\hat{T}^{(1)}_{ML,X} - \hat{T}^{(2)}_{ML,X})^2 + (\hat{T}^{(1)}_{ML,Y} - \hat{T}^{(2)}_{ML,Y})^2 + (\hat{T}^{(1)}_{ML,Z} - \hat{T}^{(2)}_{ML,Z})^2} &< th_T \\ \sqrt{(\hat{R}^{(1)}_{ML,X} - \hat{R}^{(2)}_{ML,X})^2 + (\hat{R}^{(1)}_{ML,Y} - \hat{R}^{(2)}_{ML,Y})^2 + (\hat{R}^{(1)}_{ML,Z} - \hat{R}^{(2)}_{ML,Z})^2} &< th_R \end{aligned} \tag{4.4}$$

erfüllt sein. Für CIF–Bewegtbildsequenzen mit 10Hz Bildwiederholfrequenz werden die festen Schwellwerte $th_T = 0.5$ pel und $th_R = 0.5$ grad verwendet. Um die geschätzten Bewe-

gungsparameter zweier benachbarter Dreiecke gemäß Gl. (4.4) vergleichen zu können, müssen jeweils die Koordinatensysteme beider Dreiecke, siehe Gl. (4.48), im gleichen Punkt $\boldsymbol{G}$ definiert sein.

Im idealen Fall stellen die so gefundenen Oberflächengebiete schon die Objektkomponenten dar. Aufgrund von Schätzfehlern in den Bewegungsparametern der Dreiecke sowie unzureichender Bewegung der Objektkomponenten stellen die erhaltenen Oberflächengebiete i.a. nur Teile der Objektkomponenten dar (Bild 4.4). Um die Objektkomponenten vollständig zu

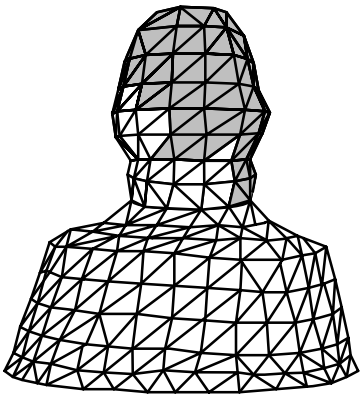


Bild 4.4 Oberflächengebiet entstanden aus der Zusammenfassung von benachbarten Dreiecken mit ähnlichen dreidimensionalen Bewegungsparametern. Seine Dreiecke sind hellgrau dargestellt. Das Oberflächengebiet stellt jedoch nur ein Teil der Objektkomponente ”Kopf” der Person gemäß Bild 4.2.f dar.

bestimmen, werden die in der Analyse vorheriger Bilder erhaltenen Ergebnisse der Zusammenfassung von Dreiecken bei der Analyse des aktuellen Bildes berücksichtigt. Dazu besitzt jedes Dreieck eines starren Modellobjektes ein Gedächtnis für die Zugehörigkeit zu mehreren Oberflächengebieten. Ein während der Analyse des aktuellen Bildes ermitteltes Oberflächengebiet erhält ein neues Kennzeichen, wenn es sich mit keinem der bereits gekennzeichneten Oberflächengebiete überlappt (Bild 4.5). Dieses Kennzeichen wird dann im Gedächtnis der zugehörigen Dreiecke eingetragen. Bei Überlappung verfügen die Oberflächengebiete über gemeinsame Dreiecke. Im Falle von Überlappung wird das Kennzeichen des bereits gekennzeichneten Oberflächengebietes auch in die Dreiecke der während der Analyse des aktuellen Bildes erhaltenen Oberflächengebietes eingetragen (Bild 4.6). Aus Stabilitätsgründen werden nur Erweiterungen zugelassen, die eine Vergrößerung der Gesamtfläche bis zu 30% ergeben, d.h.

$$\frac{\mathcal{F}_{new}}{\mathcal{F}} \leq th_{\mathcal{V}}. \tag{4.5}$$

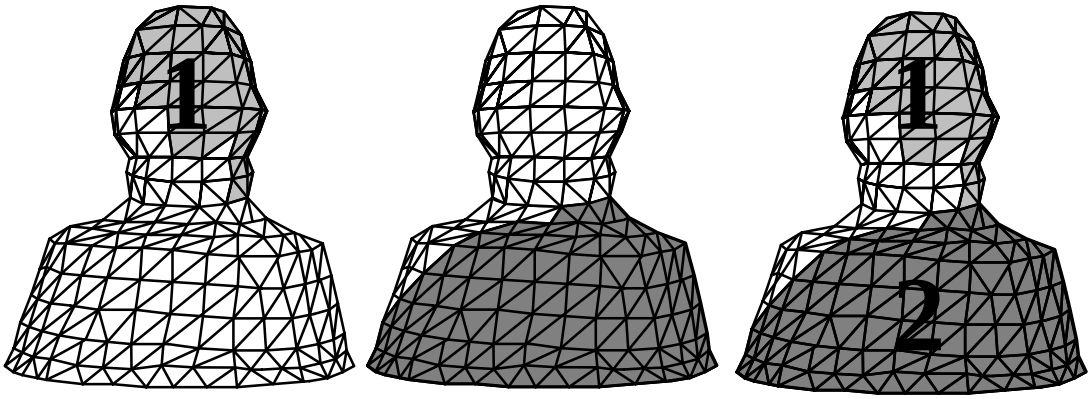


Bild 4.5 Ein Beispiel für das Kennzeichnen eines während der Analyse des aktuellen Bildes *k+1* ermittelten Oberflächengebietes. a) Einziges Oberflächengebiet, das bis zum Zeitpunkt *k* gekennzeichnet worden ist. Es erhielt das Kennzeichen 1. Seine Dreiecke sind hellgrau dargestellt. b) Oberflächengebiet, das während der Analyse des aktuellen Bildes *k+1* ermittelt wurde. Seine Dreiecke sind dunkelgrau dargestellt. c) Das während der Analyse des aktuellen Bildes ermittelte Oberflächengebiet erhält das neue Kennzeichen 2, da es sich mit dem bereits gekennzeichneten Oberflächengebiet nicht überlappt.

Hierbei ist $th_{\mathcal{F}} = 0.3$ und $\mathcal{F}$ stellt die Gesamtfläche des zu erweiternden bereits gekennzeichneten Oberflächengebietes und $\mathcal{F}_{new}$ die Gesamtfläche aller Dreiecke, die durch die Erweiterung neu hinzugefügt werden, dar. Aufgrund der Homogenität des Dreiecksnetzes gemäß Abschnitt 4.2.1 wird die Gesamtfläche eines Oberflächengebietes als Summe der Dreiecke definiert, die das entsprechende Oberflächengebiet darstellen.

Ein gekennzeichnetes Oberflächengebiet wird erst dann als Modellobjektkomponente detektiert, wenn es in zwei aufeinanderfolgenden Bildern seine Gesamtfläche nicht verändert hat und eine Mindestgröße von 10% der Gesamtfläche des sichtbaren Oberflächenteils des starren Modellobjektes überschreitet (Bild 4.7). Falls eine Objektkomponente detektiert wird, werden die entsprechenden Stützpunkte des ursprünglichen starren Modellobjektes der neuen Objektkomponente zugeordnet. Die Dreiecksnetze zweier auf diese Weise entstandener Objektkomponenten bleiben über gemeinsame Dreiecke flexibel miteinander verbunden (Bild 3.10).

Zur Erkennung einer gegenseitigen Verdeckung von Objektkomponenten wird die Übereinstimmung $\mathcal{U}$ der Berandung ihrer Silhouetten überprüft. Es wird folgendes Kriterium verwendet: Falls weniger als 20% der Berandung der Silhouette einer Modellobjektkomponente $\mathcal{C}^{(\delta)}$ mit der Berandung der Modellobjektsilhouette zusammenfällt, d.h.

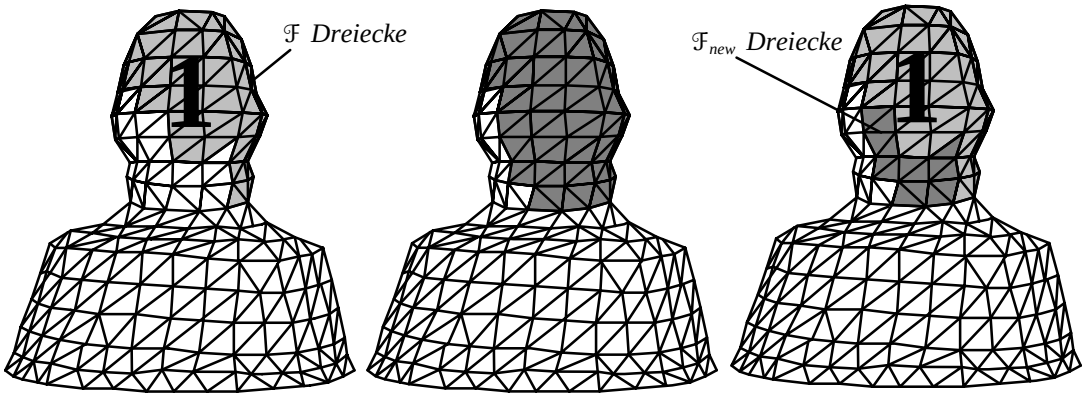


Bild 4.6 Ein Beispiel für die Erweiterung eines bereits gekennzeichneten Oberflächengebietes mittels eines während der Analyse des aktuellen Bildes *k+1* ermittelten Oberflächengebietes. a) Bereits mit 1 gekennzeichnetes Oberflächengebiet. Seine $\mathcal{F}$ Dreiecke sind hellgrau dargestellt. b) Ein während der Analyse des aktuellen Bildes *k+1* ermitteltes Oberflächengebiet. Seine Dreiecke sind dunkelgrau dargestellt. c) Erweitertes Oberflächengebiet. Das bereits eingetragene Kennzeichen 1 wird für alle Dreiecke des erweiterten Oberflächengebietes übernommen. Die $\mathcal{F}_{new}$ neu hinzugefügten Dreiecke sind dunkelgrau dargestellt.

$$\mathcal{U} = \left\lceil \frac{L_{\mathcal{G}}^{(\delta)}}{L^{(\delta)}} \right\rceil \leq th_{\mathcal{U}} \tag{4.6}$$

erfüllt wird, wird die Objektkomponente als verdeckend erkannt (Bild 4.8). Dabei ist $th_{\mathcal{U}} = 0.2$ und $L^{(\delta)}$ stellt die Länge der Berandung der Silhouette der Modellobjektkomponente dar. $L_{\mathcal{G}}^{(\delta)}$ ist die Länge der gemeinsamen Berandung der Modellobjektsilhouette und der Silhouette der Modellobjektkomponente.

Im Falle gegenseitiger Verdeckung der Objektkomponenten wird die Form der verdeckenden Objektkomponente durch ein separates, geschlossenes Dreiecksnetz dargestellt und unter Vermeidung von räumlichen Überschneidungen vor den verdeckten Modellobjektkomponenten plaziert. Das Dreiecksnetz wird aus der realen Silhouette der verdeckenden Objektkomponente gemäß Abschnitt 4.2.1 unter der Hypothese ellipsoider Form bestimmt. Die polygonförmige Berandung der Silhouette einer verdeckenden Modellobjektkomponente (Bild 4.9.a) stimmt jedoch nicht exakt mit der Berandung der realen Silhouette der verdeckenden Objektkomponente (Bild 4.9.b) überein. Die reale Silhouette der verdeckenden Objektkomponente wird daher vor der Netzerzeugung aus einer lokalen Segmentierung eines Displacementvektorfeldes in Regionen mit einheitlichem Betrag der Displacementvektoren bestimmt. Der zu segmentierende Bildbereich wird aus der Silhouette der verdeckenden Modellobjektkompo-

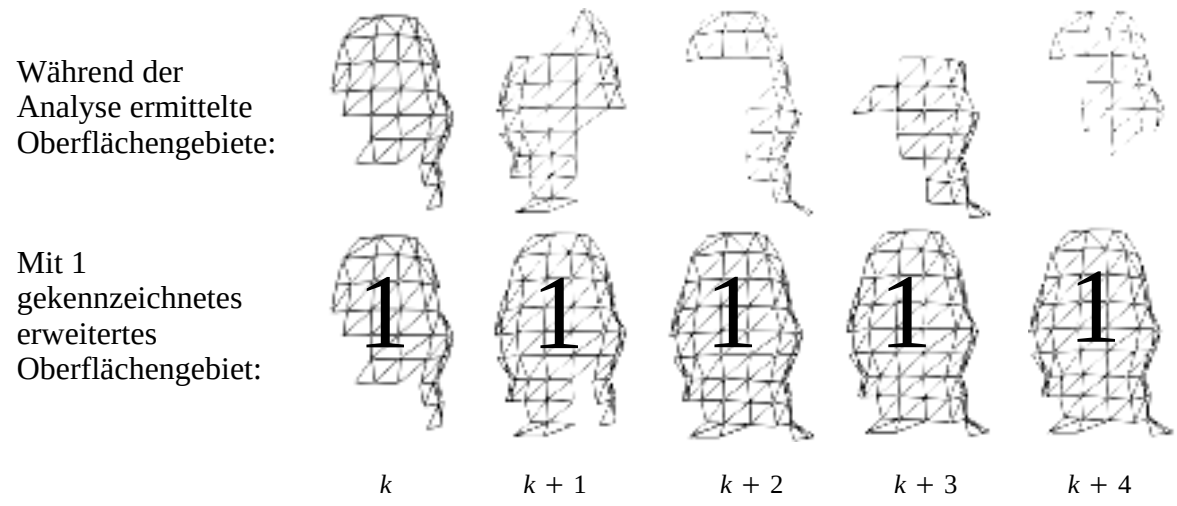


Bild 4.7 Ein Beispiel für die Detektion eines gekennzeichneten Oberflächengebietes als Objektkomponente. Das mit 1 gekennzeichnete Oberflächengebiet wird zum Zeitpunkt *k+4* als Objektkomponente detektiert, da es sowohl in Bild *k+3* als auch in Bild *k+4* seine Gesamtfläche nicht verändert und eine feste Mindestgröße überschreitet hat.

nente ermittelt. Dazu wird die Silhouette der verdeckenden Modellobjektkomponente durch Dilatationsoperationen vergrößert. Der auf diese Weise entstandene Bildbereich (Bild 4.9.c) wird dann segmentiert. Dabei wird angenommen, das sich die reale Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* vollständig innerhalb dieses Bildbereiches befindet. Der Teil des zu segmentierenden Bildbereiches, der nicht zu dieser Silhouette gehört, wird im folgenden als bewegter Hintergrund bezeichnet. Zur Definition eines Displacementvektorfeldes wird ein Oberflächenpunkt H mit dem Ortsvektor $\boldsymbol{H}_k = (H_{X,k}, H_{Y,k}, H_{Z,k})^T$ eines sich bewegenden Objektes betrachtet. Aufgrund der Bewegung des Objektes vom Zeitpunkt k bis zum Zeitpunkt $k+1$ bewegt sich H von $\boldsymbol{H}_k$ nach $\boldsymbol{H}_{k+1} = (H_{X,k+1}, H_{Y,k+1}, H_{Z,k+1})^T$. Dann bewegt sich der in die Bildebene abgebildete Punkt von $\boldsymbol{h}_k = (h_{x,k}, h_{y,k})^T$ nach $\boldsymbol{h}_{k+1} = (h_{x,k+1}, h_{y,k+1})^T$. Dem bewegten Punkt $\boldsymbol{h}_{k+1}$ wird nun die Bildkoordinate $\boldsymbol{h}_k$ über einen Displacementvektor $\boldsymbol{D}(\boldsymbol{h}_{k+1}) = (D_x(\boldsymbol{h}_{k+1}), D_y(\boldsymbol{h}_{k+1}))^T$ zugeordnet:

$$\boldsymbol{h}_k = \boldsymbol{h}_{k+1} - \boldsymbol{D}(\boldsymbol{h}_{k+1}) \,. \tag{4.7}$$

Die Zuordnung eines individuellen Displacementvektors mit 1/8 Pel Auflösung zu jeder auf dem Abtastraster liegenden Bildkoordinate in der Bildebene führt auf ein Displacementvektorfeld. Das Displacementvektorfeld wird aus dem vorherigen Bild s_k und dem aktuellen Bild s_{k+1} mittels eines hierarchischen Blockmatchers bestimmt[5][6]. Dabei kann der Displacementvektor an einer beliebigen Bildkoordinate durch seinen Betragswert $|\boldsymbol{D}| = \sqrt{D_x^2 + D_y^2}$

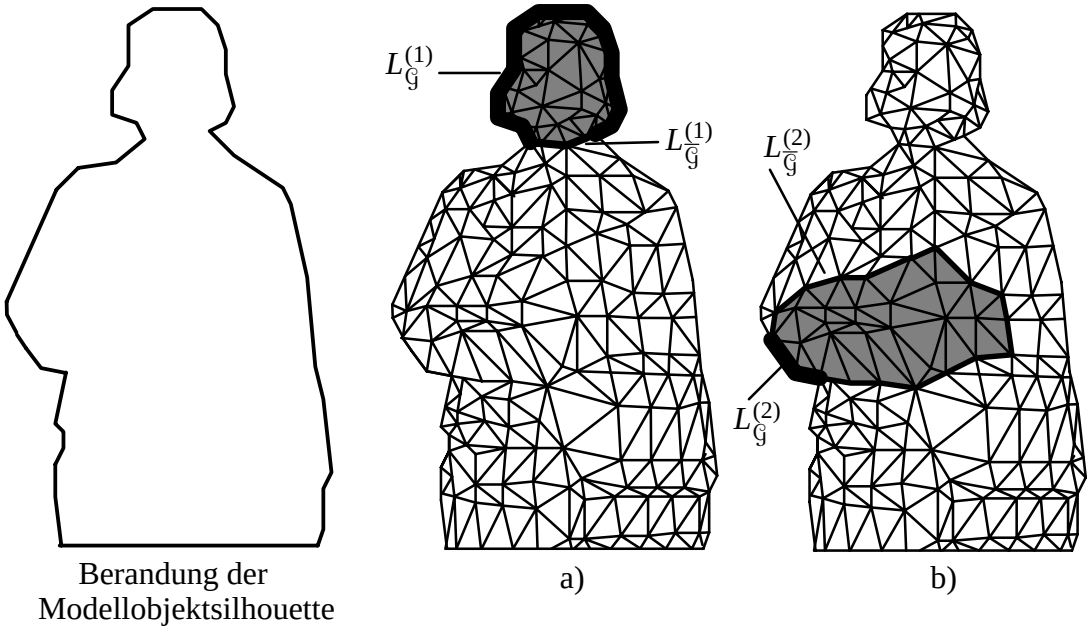


Bild 4.8 Zwei Beispiele für die Detektion gegenseitiger Verdeckung von Objektkomponenten. $L^{(\delta)}$ stellt die Länge der Berandung der Silhouette der Modellobjektkomponente $C^{(\delta)}$ dar. $L_{G}^{(\delta)}$ und $L_{\bar{G}}^{(\delta)}$ sind die Länge der gemeinsamen bzw. nicht gemeinsamen Berandung der Modellobjektsilhouette und der Silhouette der Modellobjektkomponente. Dann gilt $L^{(\delta)} = L_{G}^{(\delta)} + L_{\bar{G}}^{(\delta)}$. Die Dreiecke der zu überprüfenden Objektkomponenten sind in dunkelgrau dargestellt. a) Aufgrund von $\Theta = 0.75$ gemäß Gl. (4.6) wird keine gegenseitige Verdeckung detektiert. b) Aufgrund von $\Theta = 0.18$ gemäß Gl. (4.6) wird gegenseitige Verdeckung detektiert.

und seinen Winkel $\alpha_{\boldsymbol{D}}$ dargestellt werden. Es soll der Teil des Displacementvektorfeldes, der mit dem zu segmentierenden Bildbereich zusammenfällt, als lokales Displacementvektorfeld bezeichnet werden.

Das lokale Displacementvektorfeld wird nun in die reale Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* und den bewegten Hintergrund segmentiert. Dazu werden sowohl die Betragswerte der Displacementvektoren an den Bildkoordinaten der realen Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* als auch die Betragswerte an den Bildkoordinaten des bewegten Hintergrundes jeweils durch eine Gaußverteilung modelliert. Dabei wird zusätzlich vorausgesetzt, daß die Betragswerte voneinander statistisch unabhängig sind. Unter der Annahme, daß jeweils die Varianz beider Gaußverteilungen klein ist, d.h. daß die Betragswerte einer Region annähernd gleich groß sind, sowie unter der Annahme, daß die Mittelwerte beider Gaußverteilungen unterschiedlich sind, wird anschließend das lokale Displacementvektorfeld durch Auswertung der Betragswerte mittels eines

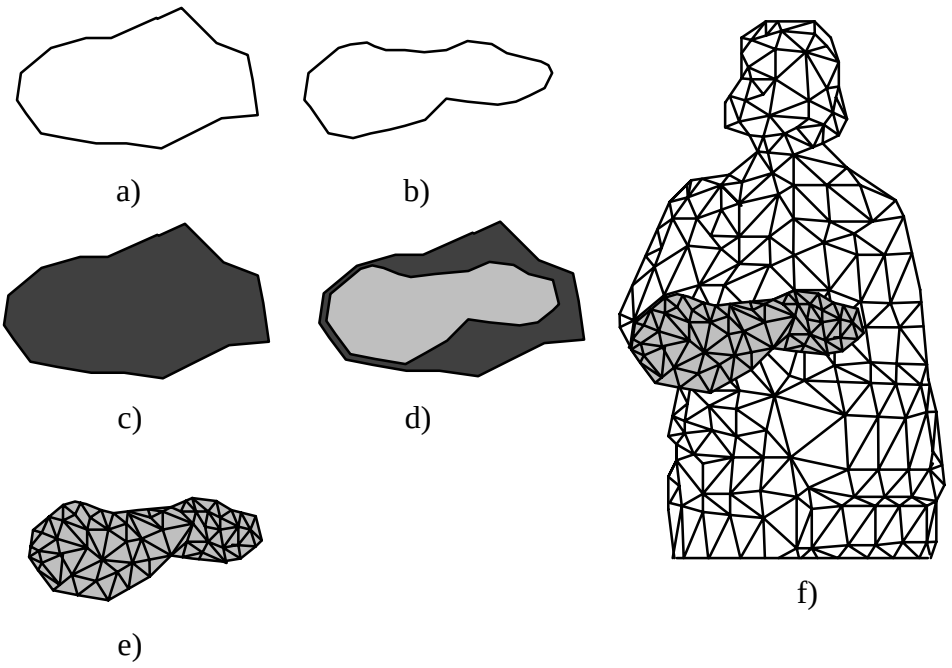


Bild 4.9 Ein Beispiel für die Erzeugung eines separates Dreiecknetzes für eine verdeckende Objektkomponente und seine Plazierung vor der verdeckten Objektkomponente. a) Polygonförmige Berandung der Silhouette einer verdekkenden Modellobjektkomponente. b) Berandung der realen Silhouette der verdeckenden Objektkomponente. c) Zu segmentierender Bereich d) Ergebnis der Segmentierung eines Displacementvektorfeldes innerhalb des zu segmentierenden Bereiches in Regionen mit einheitlichem Betrag des Displacementvektors. Die größte Region stellt die Silhouette der verdeckenden Objektkomponente dar. e) Erzeugung eines Dreiecksnetzes aus der segmentierten Silhouette. f) Plazierung des Dreiecksnetzes vor der verdeckten Modellobjektkomponente unter Vermeidung von räumlichen Überschneidungen.

Schwellwertes in die reale Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* und den bewegten Hintergrund segmentiert. Der optimale Schwellwert th_{opt} wird durch die Maximierung einer Likelihood–Funktion gemäß einem aus der Literatur bekannten Maximum–Likelihood–Schwellwertverfahren basierend auf sogennanten Mischverteilungsmodellen bestimmt [7][16][17][24][48][51][77]. Bildkoordinaten des lokalen Displacementvektorfeldes, deren Betragswerte kleiner gleich als th_{opt} sind, werden der Klasse $\mathcal{K}_1$, sonst der Klasse $\mathcal{K}_2$ zugeordnet. Anschließend werden benachbarte Bildkoordinaten des lokalen Displacementvektorfeldes mit der selben Klassenzugehörigkeit in Regionen zusammengefaßt. Im idealen Fall entstehen auf diese Weise nur zwei geschlossene Regionen, die der realen Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* und dem bewegten Hintergrund entsprechen. Bei Abweichungen der diskreten Betragswerte des lokalen Displacementvektorfeldes vom Modell oder bei zu hohen Varianzen können jedoch mehr als zwei Regionen entstehen. Um eine Segmentierung in zwei Regionen zu erreichen, wird zunächst

die größte Region einer jeden Klasse beibehalten. Dann wird eine jede kleine Region ohne Rücksicht auf die Klassenzugehörigkeit ihrer Bildkoordinaten derjeniger Nachbarregion zugeordnet, mit der sie die meisten Berührungspunkte aufweist. Zur Verfeinerung der Segmentierung werden örtliche Luminanzgradienten des Bildes zum Zeitpunkt *k+1* ausgewertet[7][46][71]. Unter der Annahme, daß die Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* den Großteil der Bildfläche innerhalb des zu segmentierenden Bildbereiches abdeckt, wird anschließend die größte gefundene Region als die reale Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* interpretiert (Bild 4.9.d).

Aus der segmentierten Silhouette wird dann unter der Hypothese ellipsoider Form ein geschlossenes Dreiecksnetz gemäß Abschnitt 4.2.1 erzeugt (Bild 4.9.e). Dabei kommt der Schwerpunkt des neuen Dreiecknetzes entsprechend Gl. (4.2) bei Z=F, d.h. in der Bildebene, zu liegen. Daher wird zum Abschluß die neue, verdeckende Modellobjektkomponente so in der Modellwelt plaziert, daß sie vor der verdeckten Objektkomponente zu liegen kommt, sich nicht mit der verdeckten Modellobjektkomponente räumlich überschneidet und bei Abbildung in die Bildebene weiterhin die geforderte Größe hat (Bild 4.9.f). Dazu wird jeder Kontrollpunkt $P_{\delta}^{(n_\delta)}$ mit dem Ortsvektor $\boldsymbol{P}_{\delta}^{(n_\delta)}$ zu seiner endgültigen Position $\boldsymbol{P}_{\delta}^{(n_\delta)'}$ gemäß

$$\boldsymbol{P}_{\delta}^{(n_\delta)'} = \begin{bmatrix} \frac{1}{F}\cdot\left(\mathbb{D}_Z - \frac{h}{2}\right) & 0 & 0 \\ 0 & \frac{1}{F}\cdot\left(\mathbb{D}_Z - \frac{h}{2}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \boldsymbol{P}_{\delta}^{(n_\delta)} + \begin{bmatrix} 0 \\ 0 \\ \left(\mathbb{D}_Z - F - \frac{h}{2}\right) \end{bmatrix} \tag{4.8}$$

verschoben. Dabei stellen $\mathbb{D}_Z$ die Z–Komponente des am nächsten zur Kamerabildebene liegenden Kontrollpunktes der verdeckten Modellobjektkomponente und *h* die Tiefe der neuen verdeckenden Modellobjektkomponente dar (Bild 4.10). Die Brennweite F wird gemäß Abschnitt 4.2.1 auf 1000 pel gesetzt. Auf die neue verdeckende Modellobjektkomponente wird dann Textur projiziert. Dazu wird das auch dem Empfänger bekannte Bild s_k' in die Bildebene der Modellkamera gelegt und jedem sichtbaren Dreieck der neuen verdeckenden Objektkomponente ein zweidimensionaler dreiecksförmiger Bildausschnitt zugeordnet. Die Lage des zugehörigen Bildausschnittes eines Dreieckes wird durch die Projektion der drei Kontrollpunkte dieses Dreieckes in die Bildebene festgelegt[45]. Luminanzen und Chrominanzen, die auf diese Weise auf die neue verdeckende Modellobjektkomponente projiziert werden, werden als ihre Texturparameter gekennzeichnet.

Experimentelle Untersuchungen

Das im Rahmen dieser Arbeit entwickelte Verfahren zur Objektgliederung wurde anhand von realen Testsequenzen getestet. In den Bildern 4.11 und 4.12 sind je ein Bild dieser Testsequenzen gezeigt. Die Testsequenzen *Claire*[13] (Bild 4.11.a), *Miss America*[9] (Bild 4.11.b) und *Akiyo*[53] (Bild 4.11.c) zeigen jeweils eine Person, deren Objektkomponenten Kopf und Schulter sich unterschiedlich im dreidimensionalen Raum bewegen. Die Testsequenzen *Gio-*

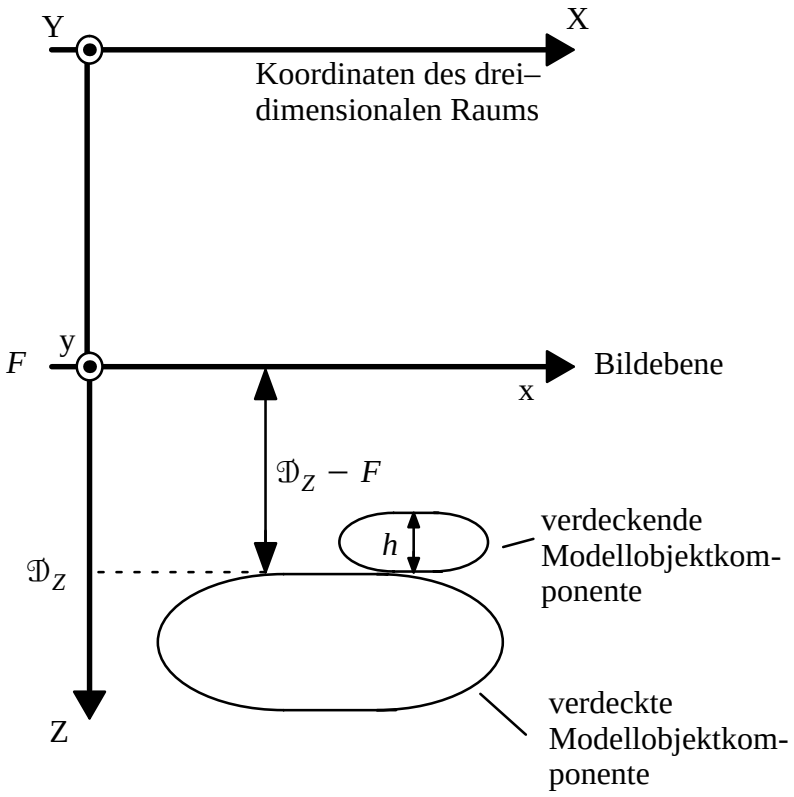


Bild 4.10 Ein Beispiel für die Plazierung der verdeckenden Modellobjektkomponente unter Vermeidung von räumlichen Überschneidungen vor der verdeckten Objektkomponente.

vanni[37] (Bild 4.12.a) und *Goldbeck*[38] (Bild 4.12.b) zeigen eine Person bestehend aus den Objektkomponenten Kopf, Oberkörper und Arme. Alle Objektkomponenten bewegen sich dabei unterschiedlich im dreidimensionalen Raum. Ferner bewegt sich der rechte Unterarm als verdeckende Objektkomponente vor dem Oberkörper. Die Testsequenz *Anis & Markus*[36] (Bild 4.12.c) zeigt jeweils die Objektkomponenten Kopf und Schulter von zwei Personen. Köpfe und Schulter beider Personen bewegen sich unterschiedlich im dreidimensionalen Raum. Alle diese Testsequenzen lagen im "Common Intermediate Format" (CIF) mit einer reduzierten Bildfolgefrequenz von f_{st}=10 Hz vor. Jedes Bild besteht dabei aus 288 Zeilen und 352 Bildpunkten pro Zeile für die Luminanzkomponente und aus 144 Zeilen und 176 Bildpunkten pro Zeile für die beiden Chrominanzkomponenten. Dabei werden die Amplituden der Abtastwerte für alle Komponenten mit 8 bit quantisiert, die Abtastfrequenz der Luminanzkomponente beträgt f_{sx}=6.75 MHz.

Bilder 4.11.a, 4.11.b und 4.11.c zeigen ferner die Ergebnisse der Objektgliederung für die Testsequenzen *Claire*, *Miss America* bzw. *Akiyo*. Das Modedellobjekt wurde nach 6, 7 bzw. 9 Bildern gegliedert. Eine wirklichkeitsnahe Objektgliederung des ursprünglichen, bei der Forminitialisierung erzeugten, starren Modellobjektes wurde jeweils ohne Vorwissen des

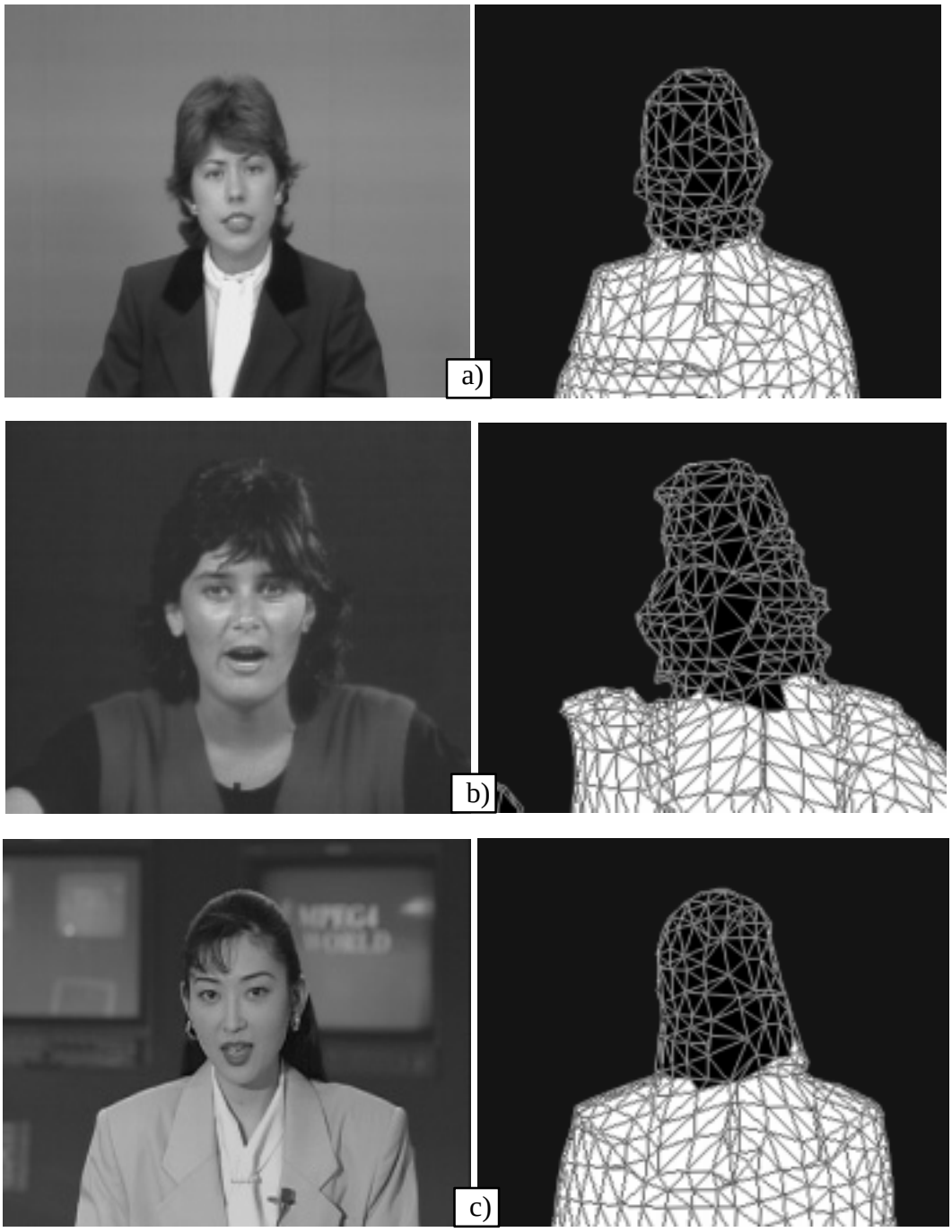


Bild 4.11 Ergebnisse des Verfahrens zur Objektgliederung. Links: a) Originalbild der Sequenz *Claire*, b) Originalbild der Sequenz *Miss America* und c) Originalbild der Sequenz *Akiyo*. Rechts: Dreiecksnetz des korrespondierenden gegliederten Modellobjektes. Die Dreiecke einer Modellobjektkomponente sind mit dem selben Grauwert dargestellt.

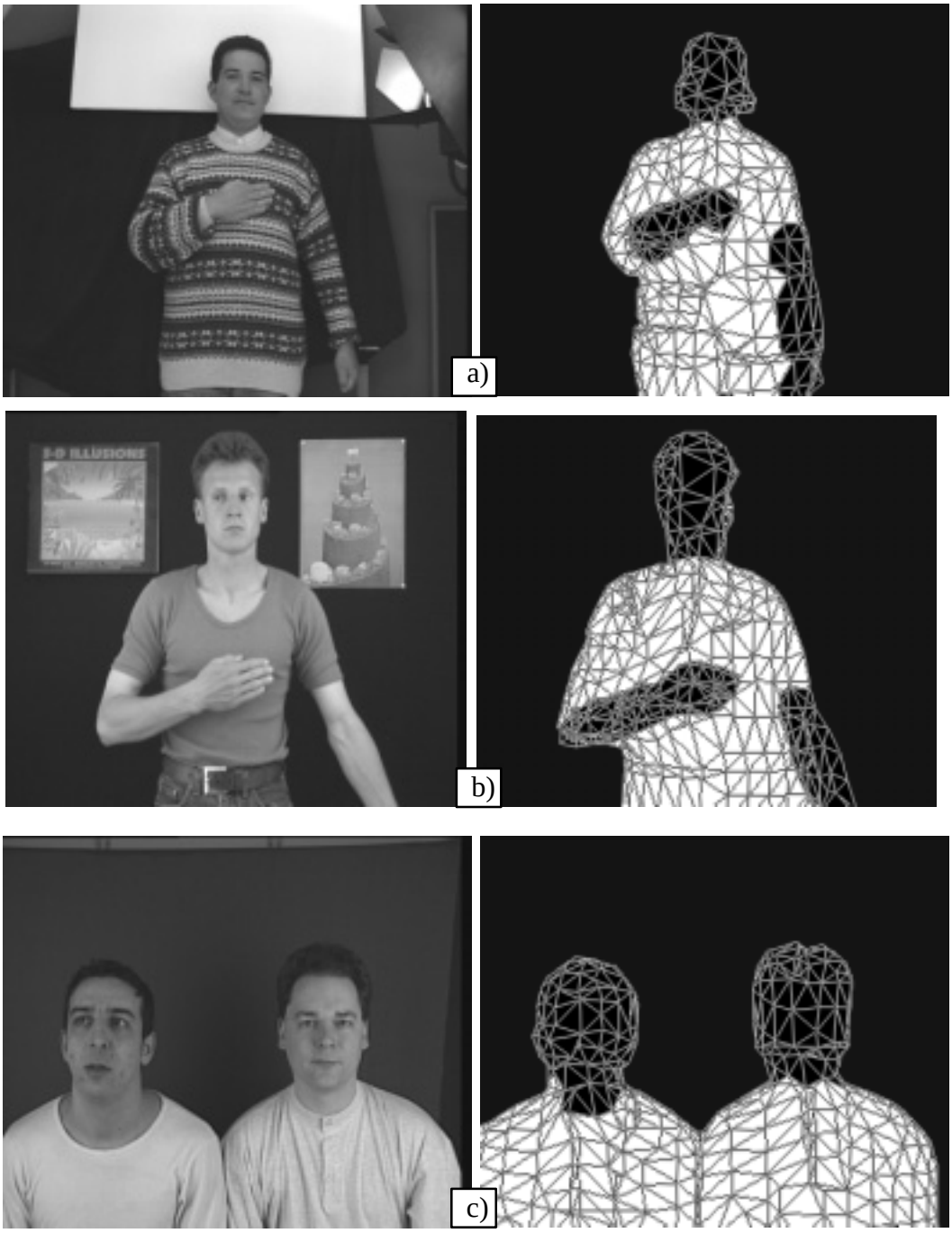


Bild 4.12 Ergebnisse des Verfahrens zur Objektgliederung. Links: a) Originalbild der Sequenz *Giovanni*, b) Originalbild der Sequenz *Goldbeck* und c) Originalbild der Sequenz *Anis&Markus*. Rechts: Dreiecksnetz des korrespondierenden gegliederten Modellobjektes. Die Dreiecke einer Modellobjektkomponente sind mit dem selben Grauwert dargestellt.

Szeneninhaltes erreicht. Ungenauigkeiten treten insbesondere am Übergang zwischen Kopf und Schulter auf, wo aufgrund von Schätzfehlern der Parameter der dreidimensionalen Bewegung Dreiecke der Schulter dem Kopf zugeordnet wurden. Die Ergebnisse der Objektgliederung für die Testsequenzen *Giovanni* und *Goldbeck* zeigen Bilder 4.12.a und 4.12.b. Für die Testsequenz *Giovanni* wurden der Kopf und der rechte Arm nach 6 Bildern gefunden. Dabei wurde der rechte Arm als verdeckend detektiert. Für die Objektkomponente rechter Arm wurde daher ein separates Dreiecksnetz erzeugt und vor der verdeckten Modellobjektkomponente Oberkörper plaziert. Der linke Arm wurde nach 8 Bildern gefunden. Für die Testsequenz *Goldbeck* wurde der Kopf nach 4 Bildern und der linke sowie der rechte Arm nach 8 Bildern gefunden. Dabei wurde durch eine Fehldetektion auch der linke Arm als verdeckend detektiert. Bild 4.12.c zeigt das Ergebnis der Objektgliederung für die Testsequenz *Anis & Markus*. Die Köpfe von beiden Personen wurden nach 8 Bildern gegliedert.

Für die hier verwendeten Testsequenzen wurde das Modellobjekt im Durchschnitt nach 8 Bildern in die Objektkomponenten gegliedert.

4.2.3 Formadaption

Da durch die Hypothese ellipsoider Form bei der Netzerzeugung die Form der Objektkomponenten eines gegliederten Objektes nicht genau beschrieben wird, unterscheiden sich nach einer Rotation der Objektkomponenten im dreidimensionalen Raum die reale Objektsilhouette und die Modellobjektsilhouette. In diesem Abschnitt wird daher ein Verfahren zur Formadaption entwickelt, das die Form der Modellobjektkomponenten im Laufe der Bildsequenz an die reale Objektsilhouette anpaßt. Dazu werden die sichtbaren Stützpunkte der Modellobjektkomponenten so an den Rand der Modellobjektsilhouette parallel zur Kamerabildebene verschoben, daß die reale Objektsilhouette und die Modellobjektsilhouette übereinstimmen. Die reale Objektsilhouette wird aus einer Änderungsdetektion zwischen dem zuletzt übertragenen Bild und dem aktuellen realen Bild unter Verwendung der geschätzten Bewegungsparameter der Modellobjektkomponenten bestimmt. Im Falle gegenseitiger Verdeckung der Objektkomponenten wird zusätzlich die Form der verdeckenden Modellobjektkomponenten mittels ihrer realen Silhouetten angepaßt. Die reale Silhouette einer verdeckenden Objektkomponente wird aus einer lokalen Segmentierung eines Displacementvektorfeldes bestimmt. Dabei werden Segmentierungsergebnisse des vorangegangenen Bildes bei der aktuellen Segmentierung berücksichtigt.

Formadaption ohne Verdeckung

Zur Formadaption ohne Berücksichtigung gegenseitiger Verdeckung der Objektkomponenten wird die Form der Modellobjektkomponenten im Laufe der Bildsequenz an die reale Objektsilhouette angepaßt. Dazu wird ein aus der Literatur bekanntes Verfahren[65] verwendet, das aus drei Schritten besteht.

Im ersten Schritt wird eine binäre Maske mb_{k+1} bestimmt, in der die durch Objektbewegung veränderten Bildbereiche von dem zuletzt synthetisierten Bild $l_k^{'}$ zu dem aktuellen Bild l_{k+1} mit der Kennzahl 1 und die übrigen Bereiche mit der Kennzahl 0 markiert sind (Bild 4.13).

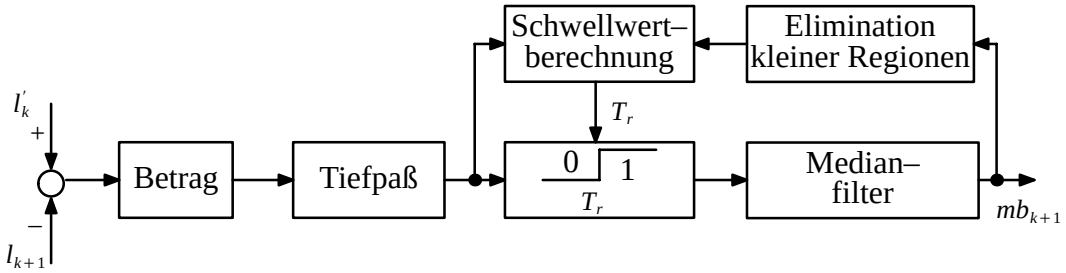


Bild 4.13 Blockdiagramm des Verfahrens zur Änderungsdetektion (aus [65]).

Jede Region mit der Kennzahl 1 wird als die Silhouette eines sich bewegenden Objektes inklusive des durch die aktuelle Bewegung freigewordenen Bildbereiches interpretiert. Zur Detektion der durch Objektbewegung veränderten Bildbereiche wird angenommen, daß die sich bewegenden Objekte zu deutlichen zeitlichen Änderungen in den Bildern führen und daß die lokale Beleuchtung diffus und zeitlich konstant ist. Zunächst wird bildpunktweise die absolute Differenz der Bilder l'_k und l_{k+1} berechnet. Das resultierende Differenzbild wird dann mit einem $3 \cdot 3$ Mittelwertfilter zur Reduzierung des Bildrauschens tiefpaßgefiltert. Das tiefpaßgefilterte Differenzbild wird im Anschluß daran mit dem rauschabhängigen Schwellwert T_r binarisiert. Bildpunkte mit Differenzwerten größer T_r werden mit der Kennzahl 1, die übrigen mit der Kennzahl 0 markiert. Nach der Schwellwertoperation erfolgt eine Medianfilterung zur Elimination einzelner mit der Kennzahl 1 markierter Punkte. Die resultierende Maske stellt mb_{k+1} dar. Der rauschabhängige Schwellwert T_r wird für das erste Bildpaar der Bildsequenz initialisiert und anschließend aus dem Bildpaar l'_k und l_{k+1} für die Anwendung auf das Bildpaar l'_{k+1} und l_{k+2} berechnet. Zur Berechnung von T_r wird das tiefpaßgefilterte Differenzbild in den statischen Bereichen durch eine Gaußverteilung approximiert und der Schwellwert so festgelegt, daß das Integral von 0 bis T_r über die Gaußverteilung den Wert 0.98 hat. Die für die Schwellwertbestimmung als statisch deklarierten Bereiche werden durch Elimination kleiner Regionen aus der Maske mb_{k+1} ermittelt. Die Elimination kleiner Regionen ordnet jeder Region, die kleiner als 0,5‰ der Bildfläche ist, die Kennzahl der sie umgebenden Region zu.

Nach der Bestimmung der Maske mb_{k+1} erfolgt im zweiten Schritt der Formadaption eine Segmentierung jeder Region mit der Kennzahl 1 in Objektsilhouette zum Zeitpunkt *k+1* und in die freigewordenen Bildbereiche. Dazu wird zunächst aus der Form der Modellobjektkomponenten und den geschätzten Bewegungsparametern ein Displacementvektorfeld mit den Vektoren

$$\boldsymbol{D}(\boldsymbol{h}_{k+1}) = (\boldsymbol{h}_{k+1} - \boldsymbol{h}_k)^T = F \cdot \left(\frac{H_{x,k+1}}{H_{z,k+1}} - \frac{H_{x,k}}{H_{z,k}} \; , \; \frac{H_{y,k+1}}{H_{z,k+1}} - \frac{H_{y,k}}{H_{z,k}} \right)^T \tag{4.9}$$

berechnet[65]. Der Zusammenhang zwischen $\boldsymbol{H}_k$ und $\boldsymbol{H}_{k+1}$ ist über die Bewegungsgleichung (3.8) und die geschätzten Bewegungsparameter gegeben. Dieses Displacementvektorfeld wird nun an den Rändern der Modellobjektsilhouette so weit extrapoliert, daß für alle als geändert gekennzeichneten Punkte der Maske mb_{k+1} ein Displacementvektor zur Verfügung steht. Da innerhalb der Maske mb_{k+1} die Silhouette eines jeden Objektes vor und nach seiner Bewegung enthalten und eine Zuordnung des Objekts vor und nach der Bewegung über das Displacementvektorfeld gegeben ist, ist die aktuelle Objektsilhouette durch die Menge der Endpunkte all der Displacementvektoren innerhalb des korrespondierenden geänderten Gebiets gegeben, deren Fußpunkte ebenfalls innerhalb der Maske mb_{k+1} zu liegen kommen. Geänderte Punkte, die aufgrund dieser Operation nicht zu der Silhouette eines Objekts gehören, kennzeichnen die freigewordenen Bildbereiche. In der Maske mc_{k+1}, der sog. Objektmaske, sind diese Regionen, nämlich die Silhouetten der realen Objekte, die Gebiete der freigewordenen Bildbereiche und der statische Hintergrund, unterschiedlich markiert.

In dem letzten Schritt der Formadaption wird nun die Silhouette jeder Modellobjektkomponente durch Verschiebung ihrer Randkontrollpunkte an die Silhouette des korrespondierenden realen Objektes angepaßt. Die Randkontrollpunkte werden gemäß [65] in Richtung auf den Rand der Objektsilhouette in der Maske mc_{k+1} verschoben. Dabei ist die Richtung der Verschiebung durch den Mittelwert der Normalvektoren der Dreiecke, zu denen der Randkontrollpunkt gehört, gegeben. Bild 4.14 zeigt, wie ein Kontrollpunkt zur Anpassung an die neue Objektsilhouette verschoben wird. Durch die Veränderung der Form der Modellobjektkomponenten ergibt sich gleichzeitig eine Anpassung an die Form der realen Objektkomponenten.

Da die Verschiebung der Randkontrollpunkte auch zu einer Verzerrung der Textur der Modellobjektkomponenten führt, können mit diesem Verfahren sinnvoll nur kleine Unterschiede zwischen den Silhouetten der Modellobjektkomponenten und der geschätzten Objektsilhouette ausgeglichen werden. Liegen die Unterschiede zwischen der Silhouetten des Modellobjektes und der geschätzten Objektsilhouette oberhalb von drei Prozent der Fläche der Objektsilhouette, wird ein neues Modellobjekt gemäß Abschnitt 4.2.1 erzeugt. Falls das Modellobjekt schon gegliedert wurde, werden vor dem Abbau des Modellobjektes die Silhouetten aller Modellobjektkomponenten gespeichert. Nach der Erzeugung des neuen Modellobjektes wird es mit Hilfe der gesicherten Silhouetten der Objektkomponenten wieder gegliedert, so daß die schon erzielte Aufteilung nach einem Wiederaufbau des gegliederten Objektes nicht verloren geht.

Formadaption bei teilweiser Verdeckung

Im Falle gegenseitiger Verdeckung der Objektkomponenten wird die Form der verdeckenden Modellobjektkomponenten mittels ihrer realen Silhouetten angepaßt. Zur Bestimmung der realen Silhouette einer verdeckenden Objektkomponente wird keine Änderungsdetektion

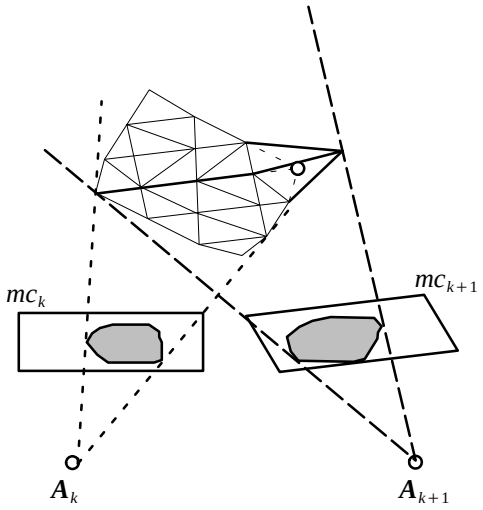


Bild 4.14 Modellobjektanpassung: Die Objektdrehung ist hier als Bewegung des Kamerabrennpunkts von A_k nach A_{k+1} dargestellt. Die geschätzten Silhouetten der realen Objekte sind in den Masken mc_k und mc_{k+1} wiedergegeben. Die Anpassung erfolgt durch Verschiebung der Kontrollpunkte senkrecht zur Modellobjektoberfläche. (aus [65])

durchgeführt, da dieses Verfahren versagt, wenn sich die verdeckten Objektkomponenten ebenfalls bewegen. Es wird daher in diesem Abschnitt ein Verfahren zur Schätzung der realen Silhouetten von verdeckenden Objektkomponenten entwickelt, das diese Silhouetten aus einer lokalen Segmentierung eines Displacementvektorfeldes bestimmt.

Das Verfahren zur Bestimmung der realen Silhouette einer verdeckenden Objektkomponente besteht aus vier Schritten[7]. Im ersten Schritt wird die verdeckende Modellobjektkomponente mittels ihrer aktuellen geschätzten Bewegungsparameter bewegt und auf diese Weise ihre Bewegung kompensiert. Im zweiten Schritt wird ein zu segmentierender Bildbereich definiert, in dem ein Displacementvektorfeld zur Bestimmung der Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* lokal segmentiert wird. Dazu wird die Silhouette der bewegungskompensierten verdeckenden Modellobjektkomponente (Bild 4.15.a) durch Dilatationsoperationen vergrößert und der auf diese Weise entstandener Bildbereich als zu segmentierender Bildbereich betrachtet (Bild 4.15.b). Dadurch werden Segmentierungsergebnisse des vorangegangenen Bildes bei der aktuellen Segmentierung berücksichtigt. Es wird angenommen, das sich die reale Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* vollständig innerhalb dieses zu segmentierenden Bildbereiches befindet. Im dritten Schritt wird ein Displacementvektorfeld aus dem vorherigen und aktuellen Bild s_k bzw. s_{k+1} mittels eines hierarchischen Blockmatchers bestimmt[5][6]. Der Teil des Displacementvektorfeldes, der mit dem zu segmentierenden Bildbereich zusammenfällt, wird als lokales Displacementvektorfeld bezeichnet. Im vierten Schritt wird die reale Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* aus einer Segmentierung dieses lokalen

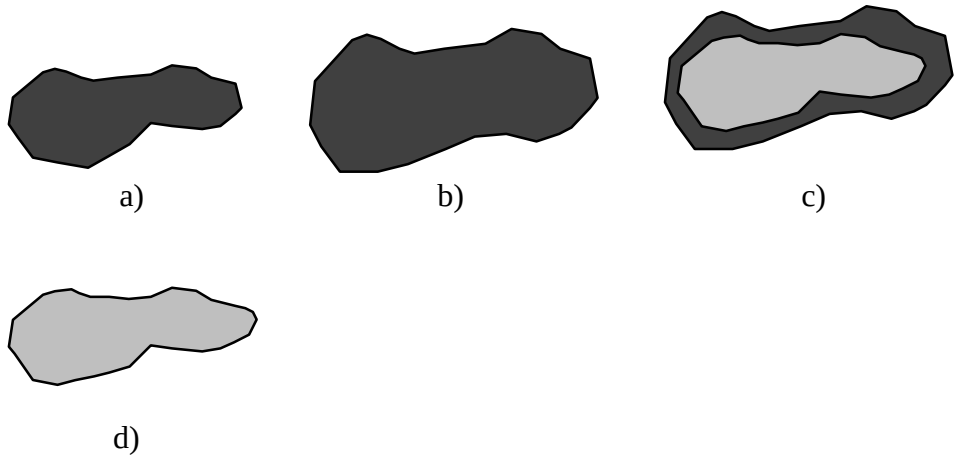


Bild 4.15 Ein Beispiel für die Bestimmung der realen Silhouette einer verdeckenden Objektkomponente zum Zeitpunkt *k+1*. a) Silhouette der verdeckenden Modellobjektkomponente zum Zeitpunkt *k+1*. b) Zu segmentierender Bildbereich. c) Ergebnisse der Segmentierung des Displacementvektorfeldes. d) Die größte Region stellt die reale Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* dar.

Displacementvektorfeldes in Regionen mit einheitlichem Betrag des Displacementvektors bestimmt (Bild 4.15.c). Zur Segmentierung wird das in Abschnitt 4.2.2 entwickelte Maximum–Likelihood–Schwellwertverfahren basierend auf Mischverteilungsmodellen verwendet. Unter der Annahme, daß die reale Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* innerhalb des zu segmentierenden Bildbereiches die größte Bildfläche bedeckt, wird die größte segmentierte Region als die reale Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1* interpretiert (Bild 4.15.d).

Aufgrund von Fehlern bei der Schätzung des Displacementvektorfeldes oder unzureichender Relativbewegung der verdeckenden und verdeckten Objektkomponenten können große Abweichungen zwischen der gerade geschätzten Silhouette und der realen Silhouette der verdekkenden Objektkomponente zum Zeitpunkt *k+1* auftreten. Um zu verhindern, daß eine stark abweichende Silhouette zur Formadaption verwendet wird, werden die Silhouette der bewegungskompensierten Modellobjektkomponente (Bild 4.16.a) und die gerade geschätzte Silhouette (Bild 4.16.b) verglichen. Im Falle größerer Unterschiede wird die gerade geschätzte Silhouette zur Formadaption verworfen. Eine geschätzte Silhouette wird zur Formadaption einer verdeckenden Modellobjektkomponenten nur dann zugelassen, wenn folgende Ungleichung

$$\frac{\mathcal{N}_{\overline{\mathcal{G}}}}{\mathcal{N}_{\mathcal{G}}} \leq th_a \tag{4.10}$$

erfüllt wird (Bild 4.16.c). Dabei stellen $\mathcal{N}_{\mathbb{G}}$ und $\mathcal{N}_{\overline{\mathbb{G}}}$ die Anzahl der Bildpunkte der gemeinsamen bzw. nicht gemeinsamen Bildbereiche der Silhouette der bewegungskompensierten verdeckenden Modellobjektkomponente und der gerade geschätzten Silhouette dar. Ein Wert th_a von 0.15 hat sich beim hier verwendeten Bildformat CIF mit einer Bildwiederholfrequenz von 10 Hz bewährt.

In dem letzten Schritt der Formadaption wird nun die Silhouette der verdeckenden Modellobjektkomponente durch Verschiebung ihrer Randkontrollpunkte an die geschätzte Silhouette angepaßt. Die Randkontrollpunkte werden wie im vorangegangenen Abschnitt in Richtung auf den Rand der geschätzten Silhouette verschoben (Bild 4.14). Liegen die Unterschiede zwischen der Silhouette der verdeckenden Modellobjektkomponente und der geschätzten Objektsilhouette oberhalb von drei Prozent der Fläche der verdeckenden Modellobjektkomponente, wird eine neue verdeckende Modellobjektkomponente gemäß Abschnitt 4.2.1 erzeugt und wiederum vor den verdeckten Objektkomponenten unter Vermeidung räumlicher Überschneidungen plaziert.

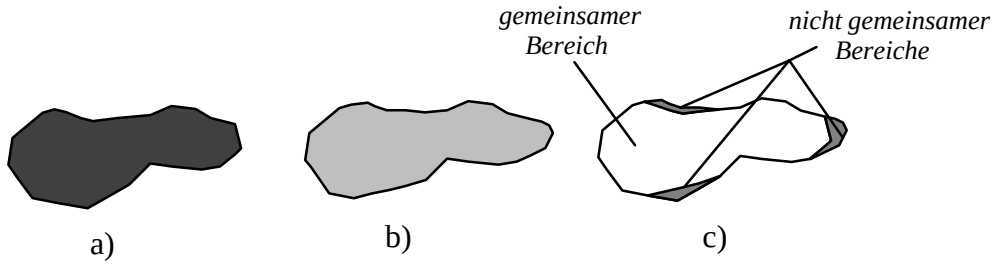


Bild 4.16 Ein Beispiel für die Zulassung einer geschätzten Silhouette zur Formadaption. a) Silhouette der verdeckenden Modellobjektkomponente zum Zeitpunkt *k+1*. b) Geschätzte Silhouette der verdeckenden Objektkomponente zum Zeitpunkt *k+1*. c) Da die Unterschiede beider Silhouetten nicht eine maximale Größe gemäß Gl. (4.10) überschreitet, wird die gerade geschätzte Silhouette zur Formadaption zugelassen.

Nachdem die neue Form aller Objektkomponenten geschätzt worden ist, werden die Texturparameter der freigewordenen Bildbereiche $\boldsymbol{C}_{k+1}^{FB}$ bei den Modellobjekten bestimmt, so daß nun die Texturparameter

$$\boldsymbol{C}_{k+1}^{MC} = \boldsymbol{C}_{k}^{MC} \cup \boldsymbol{C}_{k+1}^{FB} \tag{4.11}$$

bekannt sind.

4.2.4 Automatische Schätzung der Position von Gelenken

Das Quellenmodell A3D beschreibt jedes gegliederte Objekt in der realen Welt mittels eines gegliederten Modellobjektes. Die Modellobjekte werden in einer dreidimensionalen Modell-

welt plaziert. Ein gegliedertes Modellobjekt besteht aus einer endlichen Anzahl von Modellobjektkomponenten, die durch Modellkugelgelenke untereinander gekoppelt sind. Jede Modellobjektkomponente wird durch ihren eigenen Satz von Bewegungs–, Form– und Texturparametern beschrieben. Die Modellkugelgelenke werden als Bestandteile der Form eines gegliederten Modellobjektes interpretiert. Die Schätzung der Parameter der dreidimensionalen Form der Objektkomponenten eines gegliederten Objektes wurde schon in den Abschnitten 4.2.1, 4.2.2 und 4.2.3 beschrieben. Auf die Schätzung der Position der Modellkugelgelenke wird in diesem Abschnitt eingegangen. Die Position eines Modellkugelgelenkes wird dabei durch drei Koordinaten dargestellt. Sie wird einmalig geschätzt und im Koordinatensystem der zugehörigen Modellobjektkomponente dargestellt.

Zur Schätzung der Position eines Modellkugelgelenkes (kurz: Kugelgelenk) sind in der Literatur Verfahren bekannt, die zunächst die Position des Kugelgelenkes in der Bildebene in mehreren aufeinanderfolgenden Bildern schätzen. Aus den ermittelten Korrespondenzen zwischen den gefundenen Bildpunkten wird dann die Position des Kugelgelenkes im Raum bestimmt[26][85]. In realen Bildsequenzen läßt sich jedoch die Position der Kugelgelenke in der Bildebene nicht genau bestimmen[43]. Im Rahmen dieser Arbeit wird daher ein Schätzverfahren für die Position eines Kugelgelenkes im Raum entwickelt, das durch Auswertung der in vorangegangenen Bildern geschätzten Parametern der dreidimensionalen Bewegung beider angekoppelten Objektkomponente mittels der Gelenkbindung die Position des Kugelgelenkes bestimmen kann.

Dieses Schätzverfahren wird mit Hilfe von Bild 4.17 beschrieben. Bild 4.17 zeigt zwei durch ein Kugelgelenk $J^{(\xi)}$ gekoppelte Objektkomponenten und ihre Positionen vom Zeitpunkt *k–2* bis zum aktuellen Zeitpunkt *k*. Die Objektkomponenten werden gemäß Abschnitt 3.4.3.1 als $\mathcal{C}^{(c(\xi)^+)}$ und $\mathcal{C}^{(c(\xi)^-)}$ bezeichnet. Die Funktionen $c(\xi)^+$ und $c(\xi)^-$ geben die zur Kennzeichnung der Bezugs– bzw. Relativobjektkomponente eines Kugelgelenkes $J^{(\xi)}$ verwendeten Nummern an. Als Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ wird diejenige der beiden angekoppelten Objektkomponenten ausgewählt, die im Baum näher an der Wurzelobjektkomponente liegt. Die Wurzelobjektkomponente stellt die größte Objektkomponente des gegliederten Objektes dar. Zur Schätzung der Position des Kugelgelenkes wird angenommen, daß die aufeinanderfolgenden Bewegungsparametersätze pro Objektkomponente, die ihre Bewegung vom Zeitpunkt $k-\eta, \eta \geq 2$, bis zum aktuellen Zeitpunkt *k* beschreiben, bekannt sind. Die Schätzung dieser Parametersätze wird in Abschnitt 4.3.3 behandelt. Für die Bezugs– und Relativobjektkomponente von Bild 4.17 werden diese Bewegungsparametersätze vom Zeitpunkt *k–2* bis zum Zeitpunkt *k* als $\boldsymbol{B}^{(c(\xi)^+)}_{k-2\to k-1}$, $\boldsymbol{B}^{(c(\xi)^+)}_{k-1\to k}$ bzw. $\boldsymbol{B}^{(c(\xi)^-)}_{k-2\to k-1}$, $\boldsymbol{B}^{(c(\xi)^-)}_{k-1\to k}$ bezeichnet. Das Kennzeichen $k-\eta \to k-\eta+1$ deutet darauf hin, daß die Bewegungsparameter die Bewegung der korrespondierenden Objektkomponente vom Zeitpunkt $k-\eta$ bis zum nächsten Zeitpunkt $k-\eta+1$ beschreiben. Die korrespondierenden Translationsvektoren und Rotationsmatrixen werden für die Bezugs– und Relativobjektkomponente als $\{\boldsymbol{T}^{(c(\xi)^+)}_{k-2\to k-1}, [\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}]\}$, $\{\boldsymbol{T}^{(c(\xi)^+)}_{k-1\to k}, [\boldsymbol{R}^{(c(\xi)^+)}_{k-1\to k}]\}$ bzw. $\{\boldsymbol{T}^{(c(\xi)^-)}_{k-2\to k-1}, [\boldsymbol{R}^{(c(\xi)^-)}_{k-2\to k-1}]\}$, $\{\boldsymbol{T}^{(c(\xi)^-)}_{k-1\to k}, [\boldsymbol{R}^{(c(\xi)^-)}_{k-1\to k}]\}$ bezeichnet. Um die Beschreibung des Schätzverfahrens zu vereinfachen, wird zunächst die Position $\boldsymbol{J}^{(\xi)}_{k-2}$ des

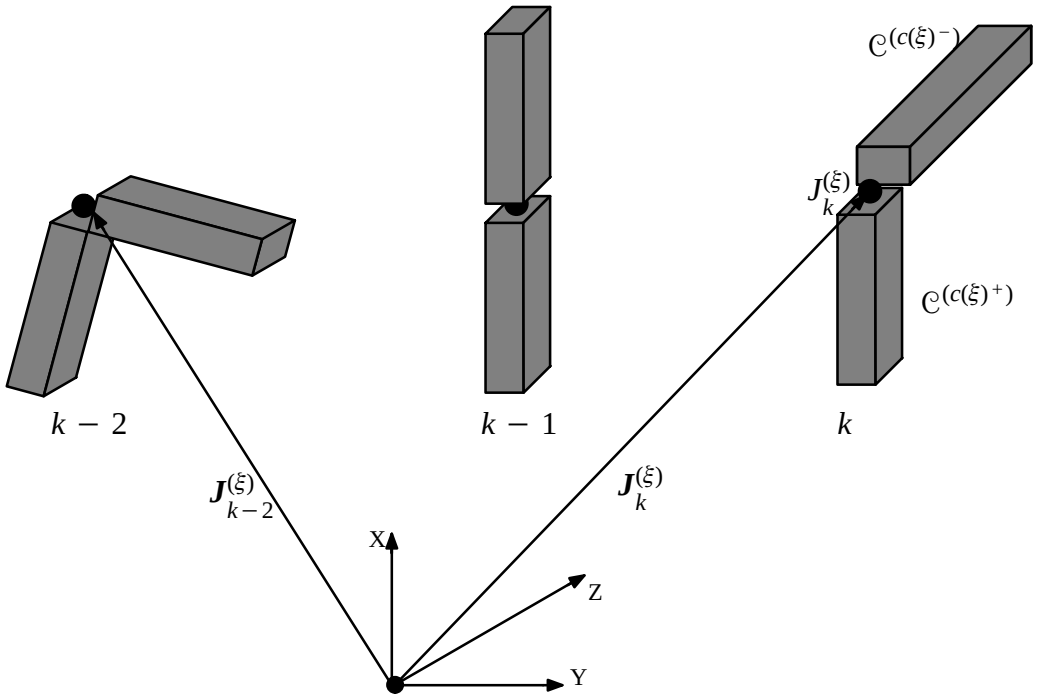


Bild 4.17 Zwei durch ein Kugelgelenk $\boldsymbol{J}^{(\xi)}$ gekoppelte Objektkomponenten und ihre Positionen vom Zeitpunkt *k–2* bis zum Zeitpunkt *k*. Die Position $\boldsymbol{J}_{k-2}^{(\xi)}$ des Kugelgelenkes zum Zeipunkt *k–2* wird zunächst bestimmt. Aus dieser geschätzten Position sowie den bekannten Bewegungsparametern der Objektkomponenten wird dann die Position $\boldsymbol{J}_k^{(\xi)}$ des Kugelgelenkes zum aktuellen Zeitpunkt *k* berechnet.

Kugelgelenkes zum Zeitpunkt *k–2* bestimmt. Aus dieser geschätzten Position sowie den bekannten Bewegungsparametern der Objektkomponenten wird dann die Position des Kugelgelenkes zum aktuellen Zeitpunkt *k* berechnet und gemäß Abschnitt 3.4.3.1 im Koordiantensystem der Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ dargestellt.

Für die Berechnung der für die Schätzung der Position des Kugelgelenkes erforderlichen Gesamtbewegung der Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ vom Zeitpunkt $k-\eta$ bis zum Zeitpunkt $k-\gamma$, $0 \le \gamma \le \eta - 1$, wird die Bewegungsgleichung entsprechend (3.8)

$$\boldsymbol{H}_{k-\eta+1}^{(c(\xi)^+)} = [\boldsymbol{R}_{k-\eta\to k-\eta+1}^{(c(\xi)^+)}] \cdot (\boldsymbol{H}_{k-\eta}^{(c(\xi)^+)} - \boldsymbol{G}_{k-\eta}^{(c(\xi)^+)}) + \boldsymbol{G}_{k-\eta}^{(c(\xi)^+)} + \boldsymbol{T}_{k-\eta\to k-\eta+1}^{(c(\xi)^+)} \tag{4.12}$$

verwendet. Dabei stellen $\boldsymbol{H}_{k-\eta}^{(c(\xi)^+)}$ und $\boldsymbol{H}_{k-\eta+1}^{(c(\xi)^+)}$ die Positionen eines beliebigen Oberfächenpunktes *H* der Bezugsobjektkomponente vor bzw. nach der Bewegung der Bezugsobjektkomponente vom Zeitpunkt $k-\eta$ bis zum nächsten Zeitpunkt $k-\eta+1$ dar. $\boldsymbol{G}_{k-\eta}^{(c(\xi)^+)}$ stellt die

Position des Koordinatensystems der Bezugsobjektkomponente zum Zeitpunkt $k-\eta$ dar. Die Gesamtbewegung der Bezugsobjektkomponente vom Zeitpunkt $k-\eta$ bis zum Zeitpunkt $k-\gamma$ ergibt sich durch wiederholte Anwendung der Gleichung (4.12) und Umformung zu

$$\boldsymbol{H}_{k-\gamma}^{(c(\xi)^+)} = [\boldsymbol{R}_{k-\eta\to k-\gamma}^{(c(\xi)^+)}] \cdot (\boldsymbol{H}_{k-\eta}^{(c(\xi)^+)} - \boldsymbol{G}_{k-\eta}^{(c(\xi)^+)}) + \boldsymbol{G}_{k-\eta}^{(c(\xi)^+)} + \boldsymbol{T}_{k-\eta\to k-\gamma}^{(c(\xi)^+)} \tag{4.13}$$

mit

$$[\boldsymbol{R}_{k-\eta\to k-\gamma}^{(c(\xi)^+)}] = [\boldsymbol{R}_{k-\gamma-1\to k-\gamma}^{(c(\xi)^+)}] \cdot [\boldsymbol{R}_{k-\gamma-2\to k-\gamma-1}^{(c(\xi)^+)}] \cdot \ldots \cdot [\boldsymbol{R}_{k-\eta\to k-\eta+1}^{(c(\xi)^+)}] \tag{4.14}$$

$$\begin{aligned}
\boldsymbol{T}_{k-\eta\to k-\gamma}^{(c(\xi)^+)} = {} & \boldsymbol{T}_{k-\gamma-1\to k-\gamma}^{(c(\xi)^+)} + \boldsymbol{G}_{k-\gamma-1}^{(c(\xi)^+)} - \boldsymbol{G}_{k-\eta}^{(c(\xi)^+)} + \\
& [\boldsymbol{R}_{k-\gamma-1\to k-\gamma}^{(c(\xi)^+)}] \cdot [\boldsymbol{R}_{k-\gamma-2\to k-\gamma-1}^{(c(\xi)^+)}] \cdot \ldots \cdot [\boldsymbol{R}_{k-\eta\to k-\eta+1}^{(c(\xi)^+)}] \cdot \boldsymbol{G}_{k-\eta}^{(c(\xi)^+)} + \\
& [\boldsymbol{R}_{k-\gamma-1\to k-\gamma}^{(c(\xi)^+)}] \cdot \Big(\boldsymbol{T}_{k-\gamma-2\to k-\gamma-1}^{(c(\xi)^+)} + \boldsymbol{G}_{k-\gamma-2}^{(c(\xi)^+)} - \boldsymbol{G}_{k-\gamma-1}^{(c(\xi)^+)} + \\
& [\boldsymbol{R}_{k-\gamma-2\to k-\gamma-1}^{(c(\xi)^+)}] \cdot \Big(\boldsymbol{T}_{k-\gamma-3\to k-\gamma-2}^{(c(\xi)^+)} + \boldsymbol{G}_{k-\gamma-3}^{(c(\xi)^+)} - \boldsymbol{G}_{k-\gamma-2}^{(c(\xi)^+)} + \\
& [\boldsymbol{R}_{k-\gamma-3\to k-\gamma-2}^{(c(\xi)^+)}] \cdot \Big(\boldsymbol{T}_{k-\gamma-4\to k-\gamma-3}^{(c(\xi)^+)} + \boldsymbol{G}_{k-\gamma-4}^{(c(\xi)^+)} - \boldsymbol{G}_{k-\gamma-3}^{(c(\xi)^+)} + \ldots \Big) \Big) \Big) .
\end{aligned} \tag{4.15}$$

Die Gesamtbewegung der Relativobjektkomponente $\mathcal{C}^{(c(\xi)^-)}$ läßt sich ähnlich berechnen. Dabei tritt an die Stelle von $c(\xi)^+$ gemäß Gl. (4.13), (4.14) und (4.15) nun $c(\xi)^-$.

Zur Bestimmung der Position des Gelenkes zum Zeitpunkt *k–2* wird die Gelenkbindung verwendet. Aufgrund der Gelenkbindung wird das Kugelgelenk $J^{(\xi)}$ mit dem Ortsvektor $\boldsymbol{J}_{k-2}^{(\xi)}$ bezüglich des Weltkoordinatensystems zum Zeitpunkt *k–2* die Bewegungsgleichung (4.12) beider angekoppelten starren Objektkomponenten

$$\boldsymbol{H}_{k-1}^{(c(\xi)^+)} = [\boldsymbol{R}_{k-2\to k-1}^{(c(\xi)^+)}] \cdot (\boldsymbol{H}_{k-2}^{(c(\xi)^+)} - \boldsymbol{G}_{k-2}^{(c(\xi)^+)}) + \boldsymbol{G}_{k-2}^{(c(\xi)^+)} + \boldsymbol{T}_{k-2\to k-1}^{(c(\xi)^+)} \tag{4.16}$$

$$\boldsymbol{H}_{k-1}^{(c(\xi)^-)} = [\boldsymbol{R}_{k-2\to k-1}^{(c(\xi)^-)}] \cdot (\boldsymbol{H}_{k-2}^{(c(\xi)^-)} - \boldsymbol{G}_{k-2}^{(c(\xi)^-)}) + \boldsymbol{G}_{k-2}^{(c(\xi)^-)} + \boldsymbol{T}_{k-2\to k-1}^{(c(\xi)^-)} \tag{4.17}$$

erfüllen, d.h.

$$\begin{aligned}
& [\boldsymbol{R}_{k-2\to k-1}^{(c(\xi)^+)}] \cdot (\boldsymbol{J}_{k-2}^{(\xi)} - \boldsymbol{G}_{k-2}^{(c(\xi)^+)}) + \boldsymbol{G}_{k-2}^{(c(\xi)^+)} + \boldsymbol{T}_{k-2\to k-1}^{(c(\xi)^+)} = \\
& [\boldsymbol{R}_{k-2\to k-1}^{(c(\xi)^-)}] \cdot (\boldsymbol{J}_{k-2}^{(\xi)} - \boldsymbol{G}_{k-2}^{(c(\xi)^-)}) + \boldsymbol{G}_{k-2}^{(c(\xi)^-)} + \boldsymbol{T}_{k-2\to k-1}^{(c(\xi)^-)}.
\end{aligned} \tag{4.18}$$

Dabei beschreiben $\boldsymbol{T}_{k-2\to k-1}^{(c(\xi)^+)}$ und $[\boldsymbol{R}_{k-2\to k-1}^{(c(\xi)^+)}]$ die Bewegung der Bezugsobjektkomponente vom Zeitpunkt *k–2* bis zum Zeitpunkt *k–1*. Entsprechend beschreiben $\boldsymbol{T}_{k-2\to k-1}^{(c(\xi)^-)}$ und $[\boldsymbol{R}_{k-2\to k-1}^{(c(\xi)^-)}]$ die Bewegung der Relativobjektkomponente vom Zeitpunkt *k–2* bis zum Zeitpunkt *k–1*. $\boldsymbol{G}_{k-2}^{(c(\xi)^+)}$ und $\boldsymbol{G}_{k-2}^{(c(\xi)^-)}$ sind die Ursprünge des Koordinatensystems der Bezugs– bzw. Relativobjektkomponente zum Zeitpunkt *k–2*. Durch Umformung der Gleichung (4.18) folgt

$$
\begin{aligned}
&[\boldsymbol{R}^{(c(\xi)^-)}_{k-2\to k-1}] \cdot \boldsymbol{G}^{(c(\xi)^-)}_{k-2} - \\
&+ [\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}] \cdot \boldsymbol{G}^{(c(\xi)^+)}_{k-2} + \\
&+ \boldsymbol{G}^{(c(\xi)^+)}_{k-2} - \boldsymbol{G}^{(c(\xi)^-)}_{k-2} + \\
&+ \boldsymbol{T}^{(c(\xi)^+)}_{k-2\to k-1} - \boldsymbol{T}^{(c(\xi)^-)}_{k-2\to k-1} \quad = \quad \left[[\boldsymbol{R}^{(c(\xi)^-)}_{k-2\to k-1}] - [\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}]\right] \cdot \boldsymbol{J}^{(\xi)}_{k-2}.
\end{aligned}
\tag{4.19}
$$

Interpretiert man die linke Seite der Gleichung (4.19) als Spaltenvektor $\boldsymbol{E}^{(\xi)}_{k-2\to k-1}$, folgt

$$
\boldsymbol{E}^{(\xi)}_{k-2\to k-1} \quad = \quad \left[[\boldsymbol{R}^{(c(\xi)^-)}_{k-2\to k-1}] - [\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}]\right] \cdot \boldsymbol{J}^{(\xi)}_{k-2}. \tag{4.20}
$$

Die Determinante von $\left[[\boldsymbol{R}^{(c(\xi)^-)}_{k-2\to k-1}] - [\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}]\right]$ ergibt sich, wie in [15] bewiesen, dabei zu

$$
\left\|\left[[\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}] - [\boldsymbol{R}^{(c(\xi)^-)}_{k-2\to k-1}]\right]\right\| = 0. \tag{4.21}
$$

Das Gleichungssystem (4.20) weist daher für $\boldsymbol{J}^{(\xi)}_{k-2}$ keine eindeutige Lösung, sondern mehrere Lösungen auf. *Die Lösungen der Gl. (4.20) stellen dabei eine Gerade $\Psi_{k-2\to k-1}$ im Raum dar, die durch die Position $\boldsymbol{J}^{(\xi)}_{k-2}$ des Kugelgelenkes zum Zeitpunkt k–2 verläuft*. Um den Beweis zu vereinfachen, wird das Kugelgelenk zum Zeitpunkt *k–2* zunächst als starr angenommen und die Relativobjektkomponente zusammen mit der Bezugsobjektkomponente $\mathbb{C}^{(c(\xi)^-)}$ um den Schwerpunkt $\boldsymbol{G}^{(c(\xi)^+)}_{k-2}$ der Bezugsobjektkomponente $\mathbb{C}^{(c(\xi)^+)}$ mit der Rotation $[\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}]$ und der Translation $\boldsymbol{T}^{(c(\xi)^+)}_{k-2\to k-1}$ gemäß Gl. (4.16) bewegt (Bild 4.18). Bezugsobjektkomponente und Kugelgelenk liegen nach der Bewegung an der richtigen Position des Zeitpunktes *k–1*. Um die richtige Position der Relativobjektkomponente zum Zeitpunkt *k–1* zu erhalten muß die Relativobjektkomponente nachträglich um den Ort des Kugelgelenkes mit der Rotation $[\boldsymbol{R}^{(c(\xi)^-)}_{k-2\to k-1}] \cdot [\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}]^{-1}$ gedreht werden. Ein Oberflächenpunkt $H^{(c(\xi)^-)}_{k-2}$ der Relativobjektkomponente mit dem Ortsvektor $\boldsymbol{H}^{(c(\xi)^-)}_{k-2}$ bewegt sich dann zunächst nach $\boldsymbol{H}^{(c(\xi)^-)'}$ gemäß

$$
\boldsymbol{H}^{(c(\xi)^-)'} = [\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}] \cdot (\boldsymbol{H}^{(c(\xi)^-)}_{k-2} - \boldsymbol{G}^{(c(\xi)^+)}_{k-2}) + \boldsymbol{G}^{(c(\xi)^+)}_{k-2} + \boldsymbol{T}^{(c(\xi)^+)}_{k-2\to k-1} \tag{4.22}
$$

und anschließend nach $\boldsymbol{H}^{(c(\xi)^-)}_{k-1}$ gemäß

$$
\boldsymbol{H}^{(c(\xi)^-)}_{k-1} = [\boldsymbol{R}^{(c(\xi)^-)}_{k-2\to k-1}] \cdot [\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}]^{-1} \cdot (\boldsymbol{H}^{(c(\xi)^-)'} - \boldsymbol{J}^{(\xi)}_{k-1}) + \boldsymbol{J}^{(\xi)}_{k-1} \tag{4.23}
$$

wobei

$$
\boldsymbol{J}^{(\xi)}_{k-1} = [\boldsymbol{R}^{(c(\xi)^+)}_{k-2\to k-1}] \cdot (\boldsymbol{J}^{(\xi)}_{k-2} - \boldsymbol{G}^{(c(\xi)^+)}_{k-2}) + \boldsymbol{G}^{(c(\xi)^+)}_{k-2} + \boldsymbol{T}^{(c(\xi)^+)}_{k-2\to k-1} \tag{4.24}
$$

die Position des Kugelgelenkes zum Zeitpunkt *k–1* darstellt.

Die Gelenkbindung Gl. (4.18) wird anstatt mit den Bewegungsgleichungen (4.16) und (4.17) mit den Bewegungsgleichungen (4.16) und (4.23) ausgewertet. Es werden nun alle Raumpunkte zum Zeitpunkt *k–2* bestimmt, die wie das Kugelgelenk die Bewegungsgleichung Gl.

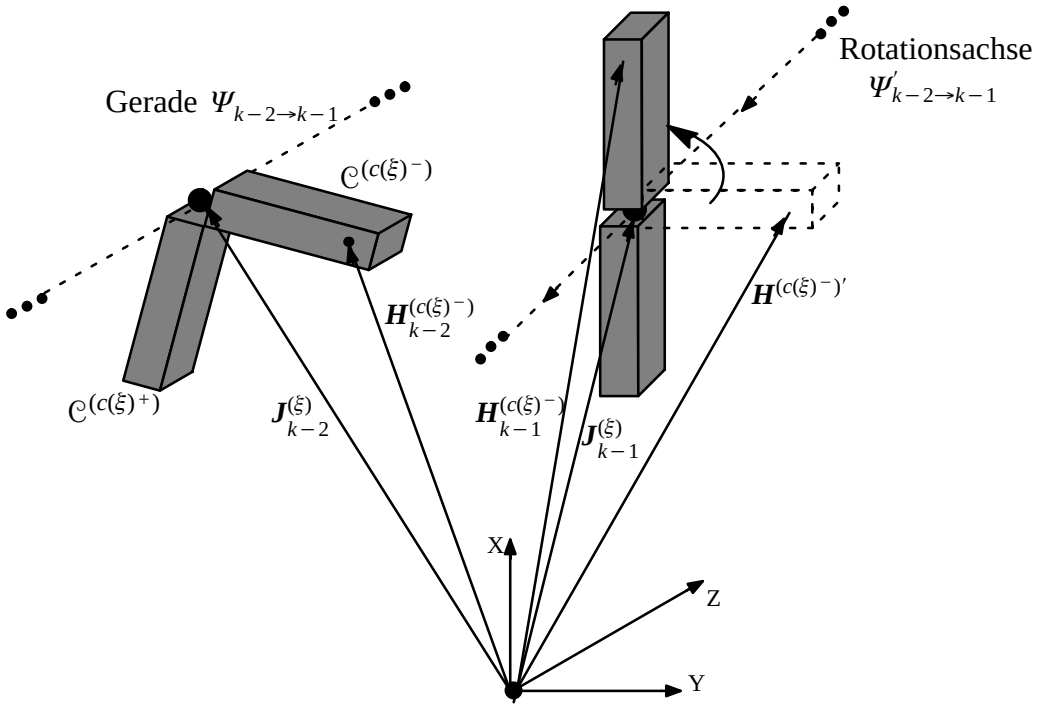


Bild 4.18 Bewegung der Relativobjektkomponente $\mathcal{C}^{(c(\xi)^-)}$ starr mit der Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ und anschließende Rotation der Relativobjektkomponente um die Position $\boldsymbol{J}_{k-1}^{(\xi)}$ des Kugelgelenkes zum Zeitpunkt *k+1*.

(4.16) und Gl. (4.23) erfüllen. Gl. (4.23) beschreibt eine Rotation, bei der alle Raumpunkte der Rotationsachse $\Psi'_{k-2\to k-1}$ unverändert bleiben. Für diese Raumpunkte gilt mit Gl. (4.22), daß

$$\boldsymbol{H}_{k-1}^{(c(\xi)^-)} = \boldsymbol{H}^{(c(\xi)^-)'} = [\boldsymbol{R}_{k-2\to k-1}^{(c(\xi)^+)}] \cdot (\boldsymbol{H}_{k-2}^{(c(\xi)^-)} - \boldsymbol{G}_{k-2}^{(c(\xi)^+)}) + \boldsymbol{G}_{k-2}^{(c(\xi)^+)} + \boldsymbol{T}_{k-2\to k-1}^{(c(\xi)^+)}. \qquad (4.25)$$

Ein Vergleich der Gl. (4.25) mit Gl. (4.16) zeigt, daß die Raumpunkte der Rotationsachse $\Psi'_{k-2\to k-1}$ die gleiche Translation und Rotation ausführen wie die Raumpunkte der Gl. (4.16) und somit die Bedingung der Gelenkbindung erfüllen. Der Zusammenhang der gesuchten Gerade $\Psi_{k-2\to k-1}$ zum Zeitpunkt *k–2* und der Rotationsachse $\Psi'_{k-2\to k-1}$ ist durch Gl. (4.22) bestimmt. Damit ist der Beweis erbracht.

Um eine eindeutige Position des Kugelgelenkes bestimmen zu können, werden zwei Bewegungen der Bezugs– und Relativobjektkomponente ausgewertet, die Bewegung vom Zeitpunkt *k–2* bis zum Zeitpunkt *k–1* und die Bewegung vom Zeitpunkt *k–2* bis zum Zeitpunkt *k*. Aus den zwei Bewegungen ergeben sich zwei Geradengleichungen für die möglichen

Positionen des Kugelgelenkes zum Zeitpunkt *k–2*, deren Schnittpunkt die gesuchte Position der Kugelgelenke ergeben. Mit Hilfe der Gl. (4.14) und Gl. (4.15) läßt sich folgende neue Gelenkbindung entsprechend Gl. (4.19) und (4.20)

$$\begin{aligned}&[\boldsymbol{R}_{k-2\to k}^{(c(\xi)^-)}]\cdot\boldsymbol{G}_{k-2}^{(c(\xi)^-)}-\\&+[\boldsymbol{R}_{k-2\to k}^{(c(\xi)^+)}]\cdot\boldsymbol{G}_{k-2}^{(c(\xi)^+)}+\\&+\boldsymbol{G}_{k-2}^{(c(\xi)^+)}-\boldsymbol{G}_{k-2}^{(c(\xi)^-)}+\\&+\boldsymbol{T}_{k-2\to k}^{(c(\xi)^+)}-\boldsymbol{T}_{k-2\to k}^{(c(\xi)^-)}\end{aligned}\quad=\quad\left[[\boldsymbol{R}_{k-2\to k}^{(c(\xi)^-)}]-[\boldsymbol{R}_{k-2\to k}^{(c(\xi)^+)}]\right]\cdot\boldsymbol{J}_{k-2}^{(\xi)}\tag{4.26}$$

$$\boldsymbol{E}_{k-2\to k}^{(\xi)}\quad=\quad\left[[\boldsymbol{R}_{k-2\to k}^{(c(\xi)^-)}]-[\boldsymbol{R}_{k-2\to k}^{(c(\xi)^+)}]\right]\cdot\boldsymbol{J}_{k-2}^{(\xi)}\tag{4.27}$$

schreiben. Die Lösungen der Gleichung (4.27) liegen in diesem Fall auf einer Gerade $\Psi_{k-2\to k}$ zum Zeitpunkt *k–2*, die durch die Position $\boldsymbol{J}_{k-2}^{(\xi)}$ des Kugelgelenkes verläuft. Dabei können die Geraden $\Psi_{k-2\to k-1}$ und $\Psi_{k-2\to k}$ parallel verlaufen.

Im Falle nicht paralleler Geraden $\Psi_{k-2\to k-1}$ und $\Psi_{k-2\to k}$ stellt ihr Schnittpunkt die Position des Kugelgelenkes zum Zeitpunkt *k–2* dar (Bild 4.19). Aufgrund von Meßfehlern der Bewegungsparameter wird die Bedingung des Schnittpunktes $\boldsymbol{E}_{k-2\to k-1}^{(\xi)}=\boldsymbol{E}_{k-2\to k}^{(\xi)}$ nicht exakt erfüllt. Die Position des Kugelgelenkes wird daher durch gleichzeitiges Lösen der Gleichungen (4.20) und (4.27) mit dem Residuum $\boldsymbol{r}=(r_0,r_1,...,r_5)^T$ bestimmt,

$$\begin{bmatrix}\boldsymbol{E}_{k-2\to k-1}^{(\xi)}\\\boldsymbol{E}_{k-2\to k}^{(\xi)}\end{bmatrix}=\begin{bmatrix}[\boldsymbol{R}_{k-2\to k-1}^{(c(\xi)^-)}]-[\boldsymbol{R}_{k-2\to k-1}^{(c(\xi)^+)}]\\[\boldsymbol{R}_{k-2\to k}^{(c(\xi)^-)}]-[\boldsymbol{R}_{k-2\to k}^{(c(\xi)^+)}]\end{bmatrix}\cdot\boldsymbol{J}_{k-2}^{(\xi)}+\boldsymbol{r}\tag{4.28}$$

$$\boldsymbol{E}^{(\xi)}=[\boldsymbol{R}^{(\xi)}]\cdot\boldsymbol{J}_{k-2}^{(\xi)}+\boldsymbol{r}.\tag{4.29}$$

Das resultierende überbestimmte Gleichungsystem (4.29) wird mit der Methode der kleinsten Fehlerquadrate[76] durch Minimierung des Residuums gelöst

$$|\boldsymbol{r}|^2=\boldsymbol{r}^T\cdot\boldsymbol{r}\underset{\boldsymbol{J}_{k-2}^{(\xi)}}{\rightarrow}\min.\tag{4.30}$$

Die Position des Kugelgelenkes ergibt sich zu

$$\hat{\boldsymbol{J}}_{k-2}^{(\xi)}=\left([\boldsymbol{R}^{(\xi)}]^T\cdot[\boldsymbol{R}^{(\xi)}]\right)^{-1}\cdot[\boldsymbol{R}^{(\xi)}]^T\cdot\boldsymbol{E}^{(\xi)}.\tag{4.31}$$

Die Position des Kugelgelenkes zum Zeitpunkt *k* ergibt sich zu

$$\hat{\boldsymbol{J}}_{k}^{(\xi)}=[\boldsymbol{R}_{k-2\to k}^{(c(\xi)^+)}]\cdot(\hat{\boldsymbol{J}}_{k-2}^{(\xi)}-\boldsymbol{G}_{k-2}^{(c(\xi)^+)})+\boldsymbol{G}_{k-2}^{(c(\xi)^+}+\boldsymbol{T}_{k-2\to k}^{(c(\xi)^+)}.\tag{4.32}$$

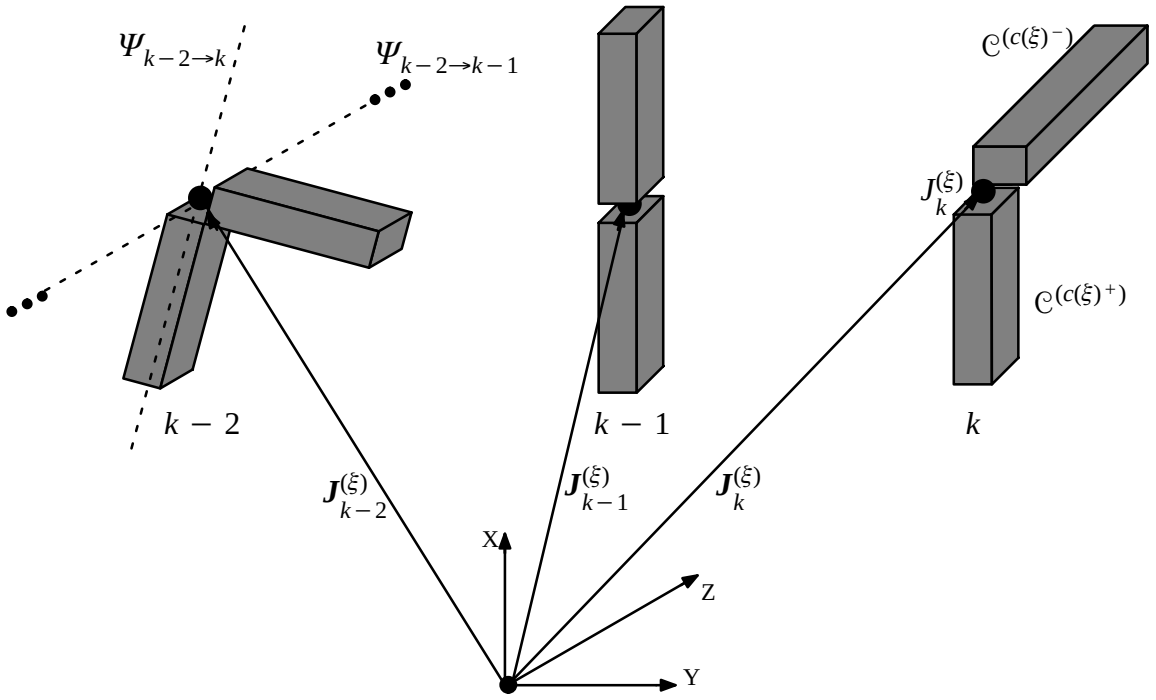


Bild 4.19 Die Position $J_{k-2}^{(\xi)}$ des Kugelgelenkes zum Zeitpunkt *k–2* wird durch den Schnittpunkt der Geraden $\Psi_{k-2\to k-1}$ und $\Psi_{k-2\to k}$ dargestellt.

Um sowohl die Genauigkeit des Schätzverfahrens zu verbesseren als auch das Problem paralleler Geraden zu vermeiden, werden η Bewegungen, wobei $\eta > 2$, ausgewertet

$$\begin{pmatrix} \boldsymbol{E}_{k-\eta\to k-\eta+1}^{(\xi)} \\ \boldsymbol{E}_{k-\eta\to k-\eta+2}^{(\xi)} \\ \boldsymbol{E}_{k-\eta\to k-\eta+3}^{(\xi)} \\ \vdots \\ \boldsymbol{E}_{k-\eta\to k}^{(\xi)} \end{pmatrix} = \begin{bmatrix} [\boldsymbol{R}_{k-\eta\to k-\eta+1}^{(c(\xi)^-)}] - [\boldsymbol{R}_{k-\eta\to k-\eta+1}^{(c(\xi)^+)}] \\ [\boldsymbol{R}_{k-\eta\to k-\eta+2}^{(c(\xi)^-)}] - [\boldsymbol{R}_{k-\eta\to k-\eta+2}^{(c(\xi)^+)}] \\ [\boldsymbol{R}_{k-\eta\to k-\eta+3}^{(c(\xi)^-)}] - [\boldsymbol{R}_{k-\eta\to k-\eta+3}^{(c(\xi)^+)}] \\ \vdots \\ [\boldsymbol{R}_{k-\eta\to k}^{(c(\xi)^-)}] - [\boldsymbol{R}_{k-\eta\to k}^{(c(\xi)^+)}] \end{bmatrix} \cdot \boldsymbol{J}_{k-\eta}^{(\xi)} + \boldsymbol{r} \tag{4.33}$$

$$\boldsymbol{E}^{(\xi)} = [\boldsymbol{R}^{(\xi)}] \cdot \boldsymbol{J}_{k-\eta}^{(\xi)} + \boldsymbol{r}, \tag{4.34}$$

wobei $\boldsymbol{r} = (r_0, r_1, ..., r_{3\eta-1})^T$ und $\boldsymbol{J}_{k-\eta}^{(\xi)}$ die Position des Kugelgelenkes zum Zeitpunkt $k - \eta$ darstellt. Die Position des Gelenkes zum aktuellen Zeitpunkt *k* ergibt sich zu

$$\boldsymbol{J}_{k}^{\hat{(\xi)}} = [\boldsymbol{R}_{k-\eta\rightarrow k}^{(c(\xi)^+)}] \cdot (\boldsymbol{J}_{k-\eta}^{\hat{(\xi)}} - \boldsymbol{G}_{k-\eta}^{(c(\xi)^+)}) + \boldsymbol{G}_{k-\eta}^{(c(\xi)^+)} + \boldsymbol{T}_{k-\eta\rightarrow k}^{(c(\xi)^+)}. \tag{4.35}$$

Experimentelle Untersuchungen

Zur Beurteilung der Genauigkeit des im Rahmen dieser Arbeit entwickelten Verfahrens zur Schätzung der Position eines Kugelgelenkes wird als Gütekriterium die mittlere Abweichung zwischen der korrekten Position und der geschätzten Position eines Kugelgelenkes gemessen. Dazu wird das Schätzverfahren auf synthetisch erzeugte Bildsequenzen mit bekannten Parametern angewendet. Jede dieser Sequenzen besteht aus N_s Bildern und zeigt ein sich bewegendes gegliedertes Objekt bestehend aus zwei durch ein Kugelgelenk gekoppelten starren Objektkomponenten. Die Form– und Texturparamter sowie die Position aller Objektkomponenten des gegliederten Objektes sind zu jedem Zeitpunkt einer Sequenz bekannt. Dabei sind die Form– und Texturparameter sowie die Position aller Objektkomponenten des gegliederten Objektes zum Zeitpunkt *0* für alle Testsequenzen identisch.

Zur Erzeugung einer solchen Bildsequenz wird zunächst ein starres Modellobjekt gemäß Abschnitt 4.2.1 aus der Testsequenz *Claire* im CIF–Format generiert (Bild 4.2.a bis 4.2.f). Der sichtbare Oberflächenteil dieses Modellobjektes wird durch 280 Stützpunkte und 491 Dreiecke dargestellt. Dabei beträgt die mittlere Größe der zu einem Dreieck korrespondierenden Bildfläche etwa 65 pel^2. Dieses starre Modellobjekt wird dann manuell in die Objektkomponenten Kopf und Schulter gegliedert. Dazu werden die entsprechenden Stützpunkte des ursprünglich starren Modellobjektes der Objektkomponente Schulter und die restlichen Stützpunkte der Objektkomponente Kopf zugeordnet (Bild 3.10). Die Dreiecksnetze der auf diese Weise entstandenen Objektkomponenten bleiben über gemeinsame Dreiecke flexibel miteinander verbunden. Dabei wird der sichtbare Oberflächenteil der Objektkomponente Schulter durch 170 Stützpunkte und 319 Dreiecke dargestellt. Entsprechend wird der sichtbare Oberflächenteil der Objektkomponente Kopf durch 110 Stützpunkte und 172 Dreiecke dargestellt. Die Position $\boldsymbol{J}^{(1)}$ des Kugelgelenkes wird im Schwerpunkt der gemeinsamen Dreiecke definiert und im Koordinatensystem der Objektkomponente Schulter dargestellt (Bild 3.9). Der Ursprung des Koordinatensystem der Objektkomponente Kopf wird am Kugelgelenk definiert. Das erste Bild einer synthetisch erzeugten Sequenz zum Zeitpunkt *0* entsteht durch die Synthese eines Modellbildes gemäß Kapitel 5. Vor der Synthese der restlichen Bilder vom Zeitpunkt *1* bis zum Zeitpunkt $N_s - 1$ wird jeweils die Objektkomponente Schulter zunächst bewegt und anschließend die Objektkomponente Kopf um das Kugelgelenk rotiert. Dabei nimmt jede Komponente der Translationsparameter Werte im Wertebereich [–16 pel, +16 pel] an. Entsprechend nimmt jede Komponente der Rotationsparameter Werte im Wertebereich [–10 grad, +10 grad] an. Auf die N_s Bilder der synthetischen Sequenz wird abschließend weißes Gaußsches Rauschen der Varianz σ_n^2 addiert.

Aus einer synthetisch erzeugten Sequenz wird die Position des Kugelgelenkes zum Zeitpunkt *0* geschätzt. Dazu werden zunächst die η aufeinanderfolgenden Bewegungsparametersätze pro Objektkomponente geschätzt, die ihre Bewegung vom Zeitpunkt *0* bis zum Zeitpunkt η

beschreiben. Die Schätzung dieser Parametersätze wird in Abschnitt 4.3.3 behandelt. Unter Verwendung dieser η Bewegungsparametersätzen wird dann die Position $\boldsymbol{J}_0^{(1)}$ des Kugelgelenkes zum Zeitpunkt 0 gemäß Abschnitt 4.2.4 bestimmt. Dabei ergibt sich die Abweichung zwischen der korrekten Position $\boldsymbol{J}_0^{(1)}$ und der geschätzten Position $\hat{\boldsymbol{J}}_0^{(1)}$ eines Kugelgelenkes zu

$$e_{\boldsymbol{J}_0} = |\boldsymbol{J}_0^{(1)} - \hat{\boldsymbol{J}}_0^{(1)}|. \tag{4.36}$$

Bei einer bestimmten Leistung des Bildrauschens σ_n^2 wird eine Vielzahl von synthetisch erzeugten Sequenzen generiert, aus denen jeweils die Position des Kugelgelenkes zum Zeitpunkt *0* auf die gleiche Weise geschätzt wird. Bild 4.20 zeigt die mittlere Abweichung der Position des Kugelgelenkes zum Zeitpunkt *0* über der Leistung des Bildrauschens. Für höhere

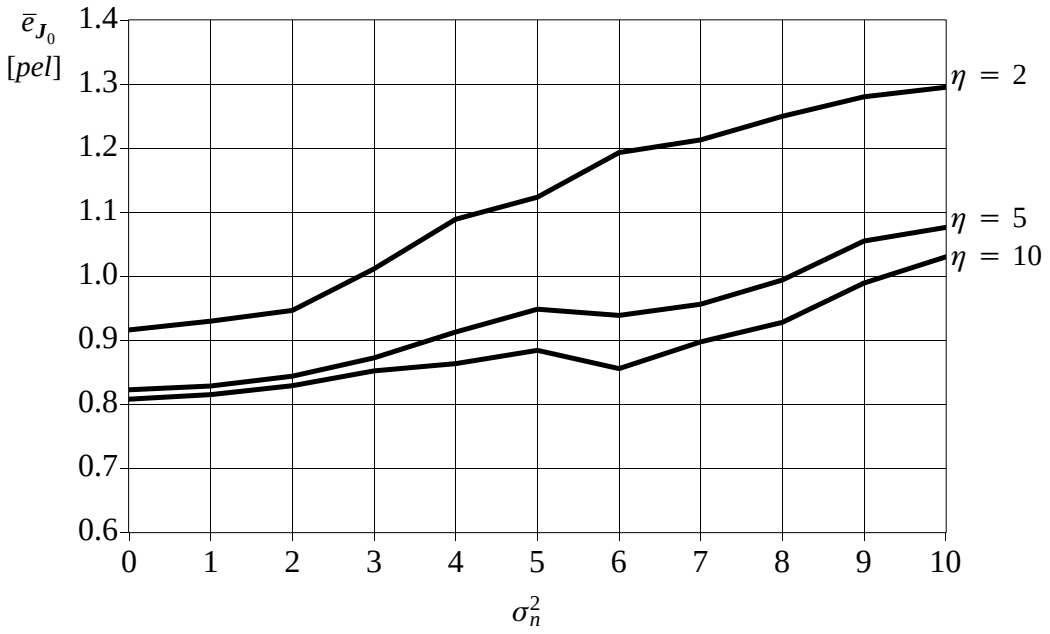


Bild 4.20 Mittlere Abweichung $\bar{e}_{\boldsymbol{J}_0}$ der Position des Kugelgelenkes gemäß Gl. (4.36) als Funktion der Varianz des Bildrauschens. Die verschiedenen Kurven ergeben sich durch unterschiedliche Auswahl der Anzahl η der zur Schätzung der Position des Kugelgelenkes verwendeten Bewegungsparametersätze pro Objektkomponente.

Leistung des Bildrauschens steigt die mittlere Abweichung der Position des Kugelgelenkes. Dies ist auf die steigende Varianz der Schätzfehler der Bewegungsparameter bei zunehmender Varianz des Bildrauschens zurückzuführen (vgl. Abschnitt 4.3.3). Die Genauigkeit des Schätzverfahrens wird durch die Auswertung von mehr als zwei Bewegungsparametersätzen

pro Objektkomponente verbessert. Für eine Leistung des Bildrauschens von 6 pel^2, d.h. 40 dB gemessen in PSNR (Peak Signal to Noise Ratio), wird die mittlere Abweichung der Position des Kugelgelenkes durch Auswertung von 5 Bewegungsparametersätzen pro Objektkomponente um 20% verbessert. Da eine weitere Verbesserung der Genauigkeit aus der Auswertung von mehr als 5 Bewegungsparametersätzen pro Objektkomponente klein ist, werden im weiteren für die Schätzung der Position eines Kugelgelenkes nur 5 Bewegungsparametersätze pro Objektkomponente verwendet. Bei einer Leistung des Bildrauschens von 40 dB wird eine mittlere Abweichung der Position des Kugelgelenkes von 0.94 pel erreicht. Dies entspricht 0.26% bezogen auf die Bildbreite B_B eines Bildes in dem hier verwendeten CIF–Format (Bild 3.2).

4.3 Schätzung der dreidimensionalen Bewegung gegliederter Objekte

Das Quellenmodell A3D beschreibt jedes gegliederte Objekt in der realen Welt mittels eines gegliederten Modellobjektes. Ein gegliedertes Modellobjekt besteht aus einer endlichen Anzahl von Modellobjektkomponenten, die durch Modellkugelgelenke untereinander gekoppelt sind. Jede Modellobjektkomponente wird durch ihren eigenen Satz von Bewegungs–, Form– und Texturparameter beschrieben. Die Modellkugelgelenke werden als Bestandteil der Form eines gegliederten Modellobjektes interpretiert und ihre Position bezogen auf das Koordinatemsystem der zugehörigen Modellobjektkomponente dargestellt. Dabei legt ein Kugelgelenk der Relativbewegung zweier gekoppelter Objektkomponenten Einschränkungen der Bewegungsmöglichkeiten auf. Die Schätzung der Parameter der dreidimensionalen Form der Objektkomponenten eines gegliederten Objektes einschließlich der Position der Kugelgelenke wurde schon in Abschnitt 4.2 beschrieben. Die Schätzung der Bewegungsparameter der Objektkomponenten soll in diesem Abschnitt behandelt werden. Da die Schätzung der Bewegungsparameter von Fehlern der vorausgegangenen Formschätzung beeinträchtigt wird, wird in Abschnitt 4.3.1 zunächst ein Modell des Schätzfehlers der dreidimensionalen Form entwikkelt. Im Abschnitt 4.3.2 wird zunächst ein Maximum– Likelihood–Schätzer für die Parameter der dreidimensionalen Bewegung von relativ kleinen starren planaren Oberflächenelementen entwickelt. Dieses Schätzverfahren wird dann in Abschnitt 4.3.3 für die Schätzung der Bewegungsparameter gegliederter dreidimensionaler Objekte erweitert.

4.3.1 Modellierung des Schätzfehlers der Form für die Bewegungsschätzung

Ein reales Objekt wird bei der Forminitialisierung gemäß Abschnitt 4.2.1 durch ein starres Modellobjekt beschrieben. Die Form des starren Modellobjektes wird durch ein dreidimensionales starres Dreiecksnetz dargestellt, das unter der Hypothese ellipsoider Form erzeugt wird. Die Form eines Modellobjektes stimmt daher mit der Form des korrespondierenden realen Objektes nicht exakt überein. Als Schätzfehler der Form wird hier der räumliche

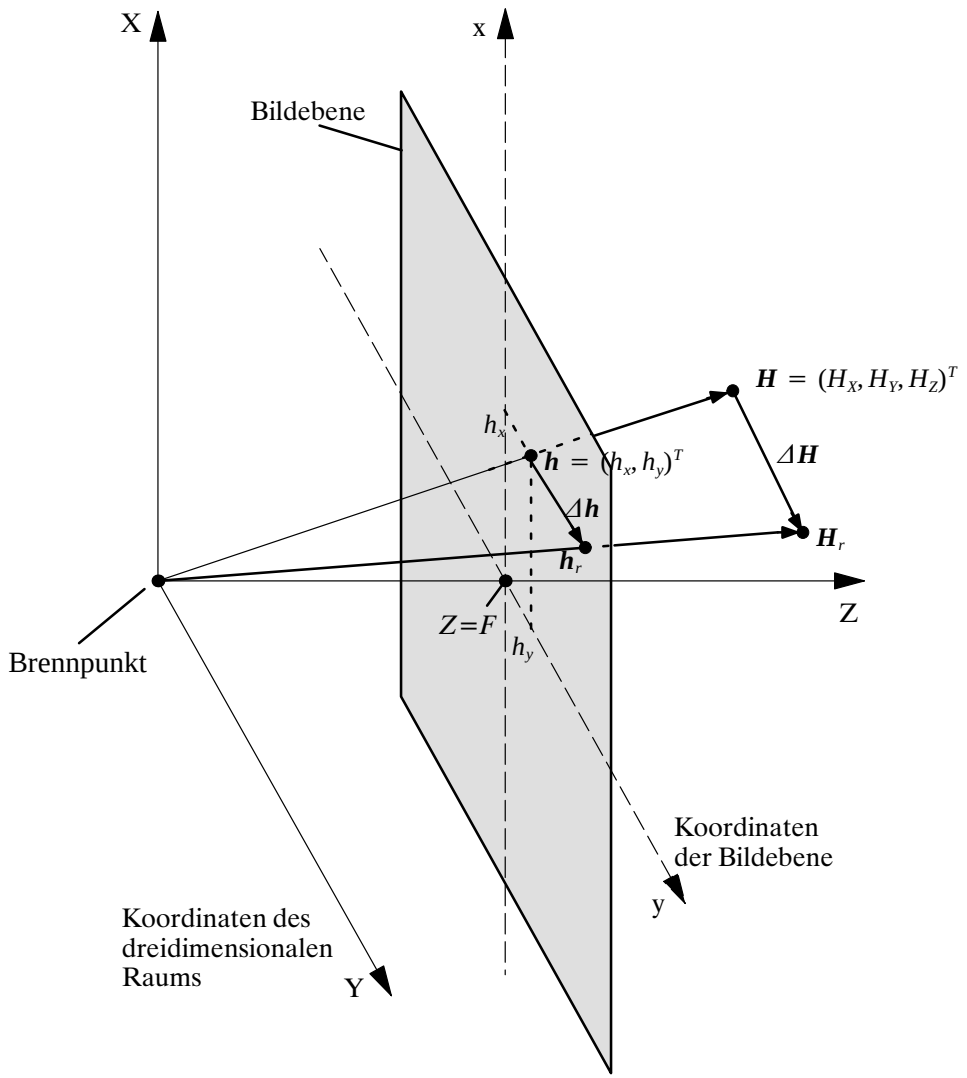


Bild 4.21 Abbildung eines Formfehlers $\Delta\boldsymbol{H}$ eines dreidimensionalen Modellobjektes in einen Positionsfehler $\Delta\boldsymbol{h}$ in der Bildebene. $\boldsymbol{H}_r$ ist der Ortsvektor eines beliebigen Oberflächenpunktes eines realen Objektes und $\boldsymbol{H}$ stellt den Ortsvektor des korrespondierenden Punktes auf der Oberfläche des Modellobjektes dar.

Positionsfehler $\Delta\boldsymbol{H}$ definiert. Er entspricht der Differenz der Vektoren des Oberflächenpunktes $\boldsymbol{H}_r$ des realen Objektes und des Ortsvektors $\boldsymbol{H}$ des korrespondierenden Punktes auf der Oberfläche des Modellobjektes (Bild 4.21)

$$\Delta\boldsymbol{H} = \boldsymbol{H}_r - \boldsymbol{H}. \tag{4.37}$$

Diese räumlichen Positionsfehler beeinflussen die Bewegungsschätzung, da diese zeitliche Luminanzdifferenzen auswertet und durch die Positionsfehler zusätzliche Luminanzdifferenzen erzeugt werden. In diesem Abschnitt wird daher zunächst ein Modell des Schätzfehlers

der Form entwickelt, um damit die zusätzlichen Luminanzfehler in der Bildebene zu berechnen. Diese werden in Abschnitt 4.3.2 in der Bewegungsschätzung berücksichtigt. Damit soll auch beim Vorhandensein von Formfehlern eine möglichst genaue Bewegungsschätzung ermöglicht werden.

Die drei Komponenten ΔH_X, ΔH_Y und ΔH_Z des Positionsfehlers $\Delta \boldsymbol{H}$ werden hier als stationäre, mittelwertfreie, Gauß'sche Zufallsprozesse mit den Varianzen σ_X^2, σ_Y^2 bwz. σ_Z^2 beschrieben. Sie werden dabei als unkorreliert angenommen. Die Kovarianzmatrix des räumlichen Positionsfehlers ergibt sich zu

$$E[\Delta \boldsymbol{H} \cdot \Delta \boldsymbol{H}^T] = [\boldsymbol{C}]_{\Delta \boldsymbol{H}} = \begin{bmatrix} \sigma_X^2 & 0 & 0 \\ 0 & \sigma_Y^2 & 0 \\ 0 & 0 & \sigma_Z^2 \end{bmatrix}. \tag{4.38}$$

Der räumliche Positionsfehler $\Delta \boldsymbol{H}$ am Ort $\boldsymbol{H}$ wird nun auf einen Positionsfehler $\Delta \boldsymbol{h}$ am Ort $\boldsymbol{h}$ in der Bildebene umgerechnet (Bild 4.21). Dazu wird die Abbildungsvorschrift des Kameramodells gemäß Gl. (3.3) verwendet. Daraus folgt folgender linearer Zusammenhang zwischen dem räumlichen Positionsfeher $\Delta \boldsymbol{H}$ und den zugehörigen Positionsfehler $\Delta \boldsymbol{h}$ am Ort $\boldsymbol{h}$ in der Bildebene

$$\Delta \boldsymbol{h} = \begin{bmatrix} \frac{F}{H_Z} & 0 & \frac{-F \cdot H_X}{{H_Z}^2} \\ 0 & \frac{F}{H_Z} & \frac{-F \cdot H_Y}{{H_Z}^2} \end{bmatrix} \cdot \Delta \boldsymbol{H} = [\boldsymbol{K}] \cdot \Delta \boldsymbol{H}\,, \tag{4.39}$$

wobei F die bekannte Brennweite der Kamera ist. Die zwei Komponenten Δh_x, Δh_y des Positionfehlers $\Delta \boldsymbol{h}$ in der Bildebene stellen ebenfalls zwei stationäre, mittelwertfreie, Gauß'sche Zufallsprozesse mit der Kovarianzmatrix

$$[\boldsymbol{C}]_{\Delta \boldsymbol{h}} = E[\Delta \boldsymbol{h} \cdot \Delta \boldsymbol{h}^T] = [\boldsymbol{K}] \cdot [\boldsymbol{C}]_{\Delta \boldsymbol{H}} \cdot [\boldsymbol{K}]^T \tag{4.40}$$

$$= \begin{bmatrix} \left(\frac{\sigma_X^2 F^2}{H_Z^2} + \frac{\sigma_Z^2 F^2 H_X^2}{H_Z^4}\right) & \frac{\sigma_Z^2 F^2 H_X H_Y}{H_Z^4} \\ \frac{\sigma_Z^2 F^2 H_X H_Y}{H_Z^4} & \left(\frac{\sigma_Y^2 F^2}{H_Z^2} + \frac{\sigma_Z^2 F^2 H_Y^2}{H_Z^4}\right) \end{bmatrix}$$

dar. Anschließend wird von dem Positionsfehler $\Delta \boldsymbol{h}$ auf einen Luminanzfehler $\Delta s_f = (s(\boldsymbol{h}_r) - s(\boldsymbol{h}))$ in der Bildebene geschlossen. Dazu wird das Bildsignal durch eine Taylorreihe im Arbeitspunkt $\boldsymbol{h}$ mit Abbruch nach dem linearen Term beschrieben[6]. Der Zusammenhang zwischen dem Positionsfeher $\Delta \boldsymbol{h}$ und dem zugehörigen Luminanzfehler Δs_f am Ort $\boldsymbol{h}$ ergibt sich zu

$$\Delta s_f = \boldsymbol{g}^T \cdot \Delta \boldsymbol{h}, \tag{4.41}$$

wobei $\boldsymbol{g}^T = (\frac{\delta s}{\delta x}, \frac{\delta s}{\delta y}) = (g_x, g_y)$ den örtlichen Bildsignalgradienten darstellt. Gl. (4.41) beschreibt dabei die Änderung des Luminanzsignals zwischen den Orten $\boldsymbol{h}_r$ und $\boldsymbol{h}$ im Bild. Der Luminazfehler $\Delta s_{\boldsymbol{f}}$ stellt ebenfalls einen stationären, mittelwertfreien, Gauß'schen Zufallsprozeß mit der Varianz $\sigma^2_{\Delta s_f}$ dar.

$$\begin{aligned} \sigma^2_{\Delta s_f} &= E[\Delta s_{\boldsymbol{f}} \cdot \Delta s_{\boldsymbol{f}}^T] = \boldsymbol{g}^T \cdot [\boldsymbol{C}]_{\Delta \boldsymbol{h}} \cdot \boldsymbol{g} \\ &= \boldsymbol{g}^T \cdot [\boldsymbol{K}] \cdot [\boldsymbol{C}]_{\Delta \boldsymbol{H}} \cdot [\boldsymbol{K}]^T \cdot \boldsymbol{g} \\ &= \frac{\sigma_Z^2}{H_Z^2} \cdot \left((g_x h_x + g_y h_y)^2 + \frac{F^2}{\sigma_Z^2} (g_x^2 \sigma_X^2 + g_y^2 \sigma_Y^2) \right) \end{aligned} \tag{4.42}$$

$\sigma^2_{\Delta s_f}$ nach Gl. (4.42) ist abhängig von den Ortskoordinaten $\boldsymbol{h}$ und den lokalen Luminanzgradienten $\boldsymbol{g}$ (Bild 4.22). Dabei wird der Luminanzfehler $\sigma^2_{\Delta s_f}$ in die Richtung des Gradienten größer und ist minimal in der Mitte des Bildes.

Zusätzlich zum Formfehler führt auch das Kamerarauschen zu einem Luminanzfehler Δs_n am Ort $\boldsymbol{h}$ in der Bildebene. Im Rahmen dieser Arbeit wird das Kamerarauschen am Ort $\boldsymbol{h}$ durch einen mittelwertfreien, Gauß'schen Zufallsprozeß mit der Varianz $\sigma^2_{\Delta s_n}$ beschrieben. Dabei werden Bildrauschen und Formfehler als statistisch unabhängig angenommen und der gesamte Luminanzfehler Δs am Ort $\boldsymbol{h}$ additiv aus dem Luminanzfehler aufgrund von Kamerarauschen und dem Luminanzfehler aufgrund von Fehlern der dreidimensionalen Form des starren Modellobjektes betrachtet. Die Varianz des gesamten Luminanzfehlers ergibt sich dann zu

$$\sigma^2_{\Delta s} = \sigma^2_{\Delta s_f} + \sigma^2_{\Delta s_n}. \tag{4.43}$$

Die Verbundwahrscheinlichkeitdichte der Luminanzfehler an $\mathcal{J}$ unterschiedlichen Orten $\boldsymbol{h}^{(j)}$, $0..j..\mathcal{J}–1,$ in der Bildebene ergibt sich zu

$$p_{\boldsymbol{V}}(\boldsymbol{V}) = \frac{1}{\sqrt{(2\pi)^{\mathcal{J}} |[\boldsymbol{U}]|}} \cdot \exp\left(-\frac{1}{2} \cdot \boldsymbol{V}^T \cdot [\boldsymbol{U}]^{-1} \cdot \boldsymbol{V} \right) \tag{4.44}$$

wobei

$$\boldsymbol{V} = \begin{bmatrix} \Delta s^{(\mathcal{J}-1)} \\ \Delta s^{(\mathcal{J}-2)} \\ \Delta s^{(\mathcal{J}-3)} \\ \vdots \\ \Delta s^{(0)} \end{bmatrix} \tag{4.45}$$

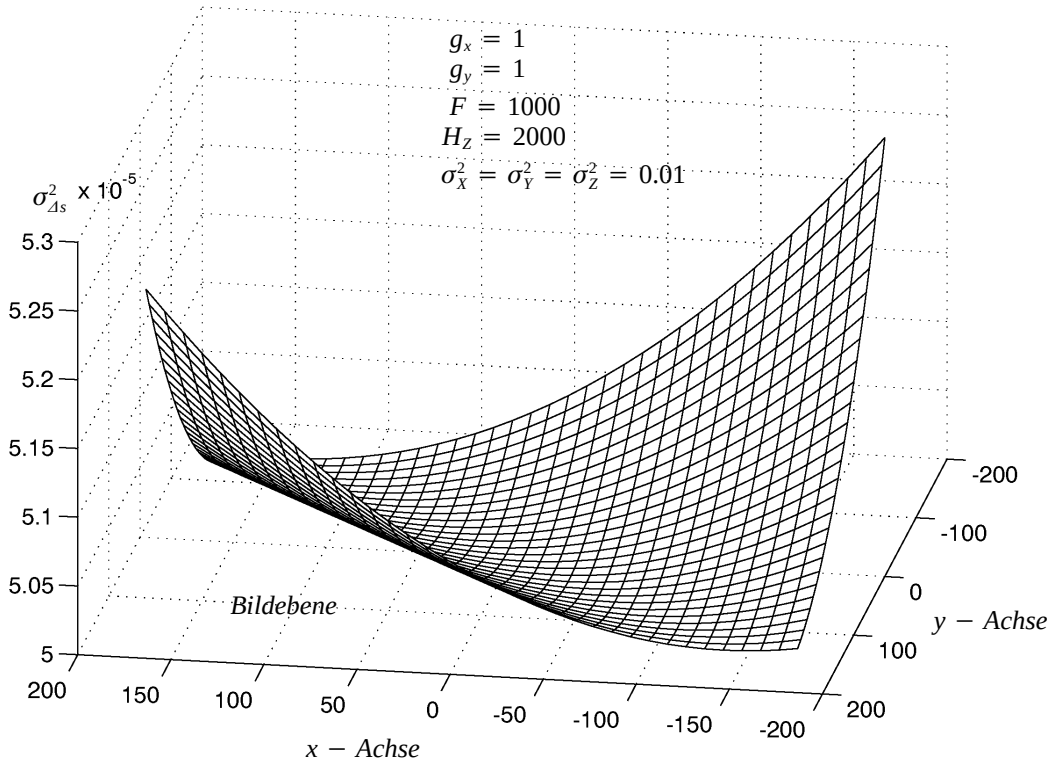


Bild 4.22 Varianz des Luminanzfehlers aufgrund von Schätzfehlern der dreidimensionalen Form in Abhängigkeit der Bildkoordinaten und des Luminanzgradienten. Es wurde ein konstanter Luminanzgradient von $\boldsymbol{g} = (1, 1)$ angenommen .

der Vektor mit den $\mathcal{J}$ Luminanzfehlern ist und $|[\boldsymbol{U}]|$ die Determinante der Kovarianzmatrix $[\boldsymbol{U}]$ darstellt. Die Kovarianzmatrix $[\boldsymbol{U}]$ ist gemäß

$$E[\boldsymbol{V} \cdot \boldsymbol{V}^T] = [\boldsymbol{U}] = \begin{bmatrix} \sigma^2_{\Delta s^{(\mathcal{J}-1)}} & 0 & 0 & . & . & 0 \\ 0 & \sigma^2_{\Delta s^{(\mathcal{J}-2)}} & 0 & . & . & 0 \\ 0 & 0 & \sigma^2_{\Delta s^{(\mathcal{J}-3)}} & . & . & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\Delta s^{(0)}} \end{bmatrix} \tag{4.46}$$

definiert, wobei die $\mathcal{J}$ Luminanzfehler als statistisch unabhängig angenommen werden.

4.3.2 Verfahren zur Schätzung der dreidimensionalen Bewegung eines relativ kleinen starren Oberflächenelementes

Bei der Objektgliederung gemäß Abschnitt 4.2.2 wird ein starres Modellobjekt in Objektkomponenten aufgeteilt. Dazu werden benachbarte Dreiecke des starren Modellobjektes, die ähnliche Parameter der dreidimensionalen Bewegung aufweisen, zu Objektkomponenten zusammengefaßt. In diesem Abschnitt wird für die Objektgliederung ein Verfahren zur Schätzung der Parameter der dreidimensionalen Bewegung eines Dreieckes eines starren Modellobjektes entwickelt. Dabei wird das in Abschnitt 4.3.1 entwickelte Modell des Schätzfehlers der Form berücksichtigt. Zur Verbesserung der Zuverlässigkeit werden Meßwerte benachbarter Dreiecke in die Bewegungsschätzung miteinbezogen. Ferner werden Ausreißer in den Meßwerten detektiert und von der Bewegungsschätzung ausgeschlossen.

Um das Verfahren zur Schätzung der Parameter der dreidimensionalen Bewegung eines Dreieckes später auch für größere Oberflächenelemente verwenden zu können, wird es hier allgemein für Oberflächenelemente bestehend aus mindestens einem Dreieck beschrieben. Bild 4.23 zeigt ein beliebiges Oberflächenelement $\mathcal{S}$ eines Modellobjektes, dessen Parameter der dreidimensionalen Bewegung $\boldsymbol{B}^{(\mathcal{S})} = (T_X^{(\mathcal{S})}, T_Y^{(\mathcal{S})}, T_Z^{(\mathcal{S})}, R_X^{(\mathcal{S})}, R_Y^{(\mathcal{S})} R_X^{(\mathcal{S})})^T$ geschätzt werden sollen. Ein Oberflächenelement wird durch eine Untermenge von $N_{\mathcal{S}}$ Kontrollpunkten $P_{\mathcal{S}}^{(n_{\mathcal{S}})}$, $0..n_{\mathcal{S}}..N_{\mathcal{S}} - 1$, eines starren Modellobjektes beschrieben.

Zur Beschreibung der dreidimensionalen Bewegung eines Oberflächenelementes $\mathcal{S}$ wird ein im Schwerpunkt $G^{(\mathcal{S})}$ des Oberflächenelementes festes Koordinatensystem mit dem Ursprung

$$\boldsymbol{G}^{(\mathcal{S})} = (G_X^{(\mathcal{S})}, G_Y^{(\mathcal{S})}, G_Z^{(\mathcal{S})}) = \frac{1}{N_{\mathcal{S}}} \cdot \sum_{n_{\mathcal{S}}=0}^{N_{\mathcal{S}}-1} \boldsymbol{P}_{\mathcal{S}}^{(n_{\mathcal{S}})} \tag{4.47}$$

bezüglich des Weltkoordinatensystems definiert. Die Bewegung wird dann durch eine Drehung $[\boldsymbol{R}^{(\mathcal{S})}]$ und Translation $\boldsymbol{T}^{(\mathcal{S})} = (T_X^{(\mathcal{S})}, T_Y^{(\mathcal{S})}, T_Z^{(\mathcal{S})})^T$ dieses Koordinatensystems im Weltkoordinatensystem beschrieben. Ein Punkt H auf der Oberfläche des Oberflächenelementes mit dem Ortsvektor $\boldsymbol{H}$ bewegt sich nach $\boldsymbol{H}'$ gemäß

$$\boldsymbol{H}' = [\boldsymbol{R}^{(\mathcal{S})}] \cdot (\boldsymbol{H} - \boldsymbol{G}^{(\mathcal{S})}) + \boldsymbol{G}^{(\mathcal{S})} + \boldsymbol{T}^{(\mathcal{S})} \tag{4.48}$$

Dabei läßt sich die Rotationsmatrix $[\boldsymbol{R}^{(\mathcal{S})}]$ unter Verwendung der Rotationsparameter $R_X^{(\mathcal{S})}$, $R_Y^{(\mathcal{S})}$ und $R_Z^{(\mathcal{S})}$ gemäß Gl. (3.6) bestimmen.

Zur Bestimmung der Parameter der dreidimensionalen Bewegung gemäß Gl. (4.48) befinden sich in der Literatur zwei grundlegende Ansätze[1]. Bei dem ersten Ansatz werden die Bewegungsparameter aus örtlichen und zeitlichen Luminanzdifferenzen an Beobachtungspunkten geschätzt[45][5]. Bei dem zweiten Ansatz werden zuerst in einem Bild Merkmale wie Kanten, Ecken oder Linienstrukturen ermittelt und die korrespondierenden Merkmale im nächsten Bild gesucht. Aus den so ermittelten Korrespondenzen wird dann die 3D–Bewegung ge-

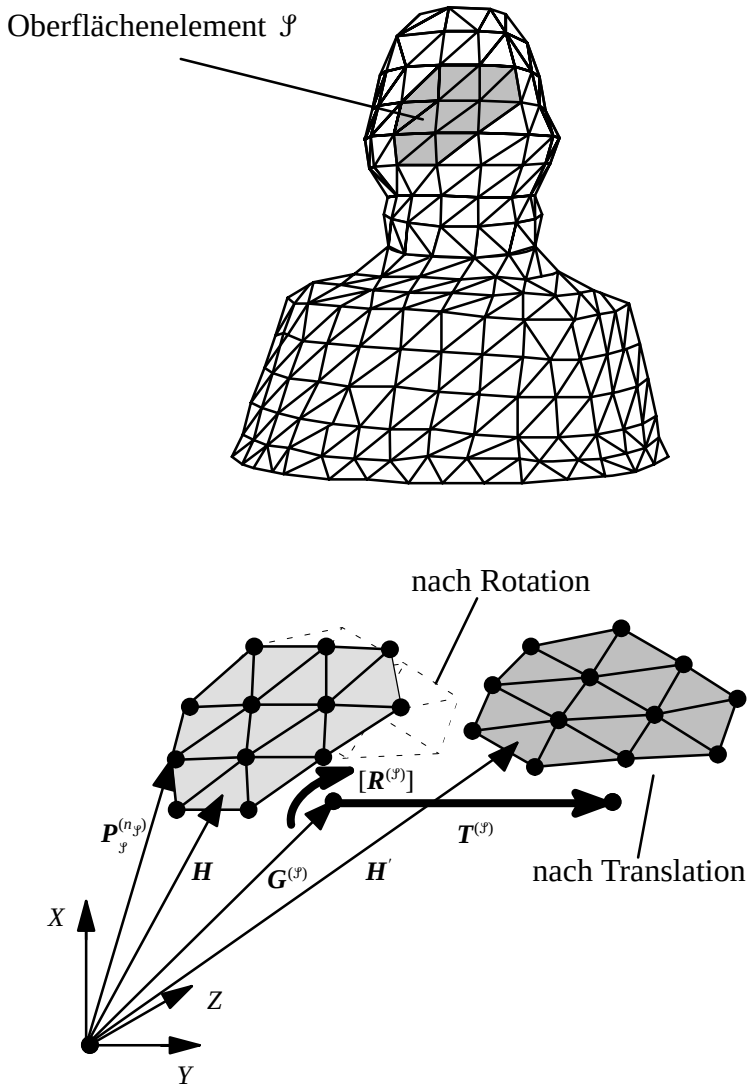


Bild 4.23 Bewegung eines Oberflächenelementes $\mathcal{S}$.

schätzt[34][87]. Im Rahmen dieser Arbeit wird für die Bestimmung der Parameter der dreidimensionalen Bewegung der erste Ansatz verfolgt. Hierbei werden die Bewegungsparameter durch einen Maximum–Likelihood–Schätzer bestimmt.

Für die Bewegungsschätzung wird ein Oberflächenelement durch einen Satz von J Beobachtungspunkte $\boldsymbol{W}^{(j)}, 0..j..\mathcal{J}-1$ repräsentiert. Ein Beobachtungspunkt $\boldsymbol{W}^{(j)}$ liegt auf dem Oberflächenelement an der Stelle $\boldsymbol{H}^{(j)} = (H_X^{(j)}, H_Y^{(j)}, H_Z^{(j)})^T$ und trägt den Luminanzwert $L^{(j)}$ des Oberflächenelementes an dieser Stelle. Die Projektion eines Beobachtungspunktes $\boldsymbol{W}^{(j)}$ in der Bildebene läßt sich mit Gl. (3.1) bestimmen. Sie wird als Bildposition eines Beobachtungs-

punktes bezeichnet und mit $\boldsymbol{h}^{(j)} = (h_x, h_y)^T$ gekennzeichnet. Der dort beobachtbare örtliche Luminanzgradient sei $\boldsymbol{g}^{(j)} = (g_x, g_y)^T$. Zur Bestimmung des örtlichen Luminanzgradienten wird der Sobel–Operator verwendet[45]. Um den Einfluß des Kamerarauschens bei realen Bildsequenzen klein zu halten, werden als Beobachtungspunkte Bildpunkte mit einem großen örtlichen Bildsignalgradienten

$$|\boldsymbol{g}^{(j)}| > T_g \tag{4.49}$$

ausgewählt[65]. Dabei sorgt gleichzeitig diese Forderung nach großen Bildsignalgradienten dafür, daß die Bewegungsparameter mit hoher Genauigkeit geschätzt werden können[3].

Ein Maximum–Likelihood–Schätzverfahren ermittelt diejenigen Parameter der dreidimensionalen Bewegung $\hat{\boldsymbol{B}}$ als Schätzwert, bei denen die bedingte Wahrscheinlichkeitsdichte der zeitlichen Luminanzdifferenzen an den Beobachtungspunkten $p_{\boldsymbol{FD}/\hat{\boldsymbol{B}}}(\boldsymbol{FD}/\hat{\boldsymbol{B}})$ maximal ist,

$$p_{\boldsymbol{FD}/\hat{\boldsymbol{B}}}(\boldsymbol{FD}/\hat{\boldsymbol{B}}_{ML}) \geq p_{\boldsymbol{FD}/\hat{\boldsymbol{B}}}(\boldsymbol{FD}/\hat{\boldsymbol{B}}) \quad \text{für alle } \hat{\boldsymbol{B}}. \tag{4.50}$$

Dabei ist $\boldsymbol{FD} = (fd(\boldsymbol{h}_k^{(\vartheta-1)}), fd(\boldsymbol{h}_k^{(\vartheta-2)}), ..., fd(\boldsymbol{h}_k^{(0)}))^T$ der Meßvektor, dessen Komponenten die meßbaren Luminazdifferenzen an den Beobachtungspunkten darstellen.

Statt das Maximum von $p_{\boldsymbol{FD}/\hat{\boldsymbol{B}}}(\boldsymbol{FD}/\hat{\boldsymbol{B}})$ zu bestimmen, kann man wegen der Monotonie der ln–Funktion auch das Maximum von $\ln\ p_{\boldsymbol{FD}/\hat{\boldsymbol{B}}}(\boldsymbol{FD}/\hat{\boldsymbol{B}})$ bestimmen. Dies hat bei den hier verwendeten Wahrscheinlichkeitsdichtefunktionen mit exponentiellen Verläufen rechentechnische Vorteile. Unter der Annahme der Existenz genau eines Maximums der Wahrscheinlichkeitsdichte gilt für den optimalen Schätzwert nach der Maximum–Likelihood–Schätzung

$$\left.\frac{\partial\ \ln\ p_{\boldsymbol{FD}/\hat{\boldsymbol{B}}}(\boldsymbol{FD}/\hat{\boldsymbol{B}})}{\partial\hat{\boldsymbol{B}}}\right|_{\hat{\boldsymbol{B}}=\hat{\boldsymbol{B}}_{ML}} = 0. \tag{4.51}$$

Für die Bestimmung der bedingten Wahrscheinlichkeitsdichte $p_{\boldsymbol{FD}/\hat{\boldsymbol{B}}}(\boldsymbol{FD}/\hat{\boldsymbol{B}})$ wird zunächst ein Beobachtungspunkt betrachtet, der sich vom Zeitpunkt t_k zum Zeitpunkt t_{k+1} in der Bildebene von $\boldsymbol{h}_k$ nach $\boldsymbol{h}_{k+1}$ bewegt. Bei einer lokalen Beschreibung des Bildsignals s_{k+1} durch eine Taylorreihe zweiter Ordnung im Arbeitspunkt $\boldsymbol{h}_k$ und unter der Annahme, daß die Luminanzdifferenz

$$fd(\boldsymbol{h}_k) = s_{k+1}(\boldsymbol{h}_k) - s_k(\boldsymbol{h}_k) \tag{4.52}$$

zwischen den Bildern s_{k+1} und s_k an der Stelle $\boldsymbol{h}_k$ ausschließlich auf eine Bewegung des Oberflächenelementes zurückzuführen ist[81], d.h.

$$s_{k+1}(\boldsymbol{h}_{k+1}) = s_k(\boldsymbol{h}_k)\,, \tag{4.53}$$

ergibt sich nach [6] folgender Zusammenhang

$$fd(\boldsymbol{h}_k) = s_{k+1}(\boldsymbol{h}_k) - s_k(\boldsymbol{h}_k) = -\bar{\boldsymbol{g}} \cdot (\boldsymbol{h}_{k+1} - \boldsymbol{h}_k). \tag{4.54}$$

Dabei stellt $\bar{\boldsymbol{g}}$ den aus den beiden Bildern gemittelten örtlichen Luminanzgradienten

$$\bar{\boldsymbol{g}} = (\bar{g}_x, \bar{g}_y) = \frac{1}{2}\left(\boldsymbol{g}_k(\boldsymbol{h}_k) + \boldsymbol{g}_{k+1}(\boldsymbol{h}_k)\right) \tag{4.55}$$

dar. Die Berechnung der örtlichen Bildsignalgradienten $\boldsymbol{g}_k(\boldsymbol{h}_k)$, $\boldsymbol{g}_{k+1}(\boldsymbol{h}_k)$ wird wiederum mittels des Sobel–Operators ausgeführt. Gleichung (4.54) stellt einen Zusammenhang zwischen der unbekannten Position $\boldsymbol{h}_{k+1}$ des Beobachtungspunktes $\boldsymbol{W}$ nach der Bewegung und den bekannten Größen fd, $\bar{\boldsymbol{g}}$ und $\boldsymbol{h}_k$ her.

Werden nun die Bildkoordinaten $\boldsymbol{h}$ in Gl. (4.54) durch Raumkoordinaten gemäß Gl. (3.1) ersetzt, folgt

$$fd(\boldsymbol{h}_k) = -F \cdot \bar{\boldsymbol{g}} \cdot \left(\frac{H_{X,k+1}}{H_{Z,k+1}} - \frac{H_{X,k}}{H_{Z,k}},\ \frac{H_{Y,k+1}}{H_{Z,k+1}} - \frac{H_{Y,k}}{H_{Z,k}}\right)^T. \tag{4.56}$$

Die Position $\boldsymbol{H} = (H_{X,k}, H_{Y,k}, H_{Z,k})$ des Beobachtungspunktes $\boldsymbol{W}_k$ ist bekannt. Die Positionen $\boldsymbol{H}_k$ und $\boldsymbol{H}_{k+1}$ sind über die Bewegungsgleichung (4.48)

$$\boldsymbol{H}_{k+1} = [\boldsymbol{R}^{(\mathcal{P})}] \cdot (\boldsymbol{H}_k - \boldsymbol{G}^{(\mathcal{P})}) + \boldsymbol{G}^{(\mathcal{P})} + \boldsymbol{T}^{(\mathcal{P})} \tag{4.57}$$

verbunden. Durch Einsetzen in Gl. (4.56) ergibt sich eine nichtlineare Gleichung mit den bekannten Größen $fd(\boldsymbol{h}_k)$, $\boldsymbol{H}_k$, $\bar{\boldsymbol{g}}$ und der Brennweite F sowie den unbekannten sechs Bewegungsparametern. Für kleine Objektbewegungen kann Gl. (4.56) linearisiert werden[45]. Dazu wird die Rotationsmatrix $[\boldsymbol{R}^{(\mathcal{P})}]$ gemäß Gl. (3.6) unter der Annahme kleiner Rotationswinkel $\{R_X, R_Y, R_Z\}$ sowie $\sin\omega \approx \omega$, $\sin^2\omega \approx 0$ und $\cos\omega \approx 1$ für kleine ω durch

$$[\boldsymbol{R}_l^{(\mathcal{P})}] = \begin{bmatrix} 1 & -R_z & R_Y \\ R_Z & 1 & -R_X \\ -R_Y & R_X & 1 \end{bmatrix} \tag{4.58}$$

approximiert. Durch Einsetzen von Gl. (4.57) und Gl. (4.58) in Gl. (4.56) und Sortieren nach den sechs Bewegungsparametern ergibt sich die Beobachtungsgleichung

$$
\begin{aligned}
fd(\boldsymbol{h}) = &+ F \cdot \bar{g}_x / H_Z \cdot T_X \\
&+ F \cdot \bar{g}_y / H_Z \cdot T_Y \\
&- [(H_X \bar{g}_x + H_Y \bar{g}_y) F / H_Z^2 + df / H_Z] \cdot T_Z \\
&- [[H_X \bar{g}_x (H_Y - G_Y) + H_Y \bar{g}_y (H_Y - G_Y) \\
&\quad + H_Z \bar{g}_y (H_Z - G_Z)] F / H_Z^2 \\
&\quad + fd / H_Z (H_Z - G_Y)] \cdot R_X \\
&+ [[H_Y \bar{g}_y (H_X - G_X) + H_X \bar{g}_x (H_X - G_X) \\
&\quad + H_Z \bar{g}_x (H_Z - G_Z)] F / H_Z^2 \\
&\quad + fd / H_Z (H_X - G_X)] \cdot R_Y \\
&- [\bar{g}_x (H_Y - G_Y) - \bar{g}_y (H_X - G_X)] F / H_Z \cdot R_Z
\end{aligned}
\tag{4.59}
$$

In Gl. (4.59) wird der Index k zur Vereinfachung der Darstellung ausgelassen. Interpretiert man die Koeffizienten der 6 Bewegungsparameter als Spaltenvektor $\boldsymbol{o}$, so kann Gl. (4.59) in der Form

$$fd(\boldsymbol{h}) = \boldsymbol{o}^T \cdot \boldsymbol{B}^{(\mathcal{P})} \tag{4.60}$$

geschrieben werden. Zur Berücksichtigung des Luminanzfehlers an dem Beobachtungspunkt, der auf das Kamerarauschen und den Schätzfehler der dreidimensionalen Form des starren Modellobjektes zurückzuführen ist (vgl. Abschnitt 4.3), wird Gl. (4.60) um Δs erweitert

$$fd(\boldsymbol{h}) = \boldsymbol{o}^T \cdot \boldsymbol{B}^{(\mathcal{P})} + \Delta s. \tag{4.61}$$

Dieser Luminanzfehler wird auch in der Literatur als Meßfehler oder Residuum bezeichnet[5][65]. Durch Aufstellen der Gl. (4.61) an $\mathcal{J}$, $\mathcal{J}$>6, Beobachtungspunkten $\boldsymbol{W}^{(j)}$ entsteht folgendes Gleichungsystem in Vektor–Matrix–Form,

$$\boldsymbol{FD} = [\boldsymbol{O}] \cdot \boldsymbol{B}^{(\mathcal{P})} + \boldsymbol{V} \tag{4.62}$$

bzw.

$$\boldsymbol{V} = \boldsymbol{FD} - [\boldsymbol{O}] \cdot \boldsymbol{B}^{(\mathcal{P})} \tag{4.63}$$

mit $\boldsymbol{FD} = (fd(\boldsymbol{h}_k^{(\mathcal{J}-1)}), fd(\boldsymbol{h}_k^{(\mathcal{J}-2)}), ..., fd(\boldsymbol{h}_k^{(0)}))^T$, der Matrix $[\boldsymbol{O}] = (\boldsymbol{o}^{(\mathcal{J}-1)T}, \boldsymbol{o}^{(\mathcal{J}-2)T}, ..., \boldsymbol{o}^{(0)T})^T$, $\boldsymbol{B}^{(\mathcal{P})} = (T_X, T_Y, T_Z, R_X, R_Y, R_Z)^T$ und $\boldsymbol{V} = (\Delta s^{(\mathcal{J}-1)}, \Delta s^{(\mathcal{J}-2)}, ..., \Delta s^{(0)})^T$.

Aus Gl. (4.63) und der Kovarianzmatrix Gl. (4.46) ergibt sich die bedingte Wahrscheinlichkeit $p_{\boldsymbol{FD}/\hat{\boldsymbol{B}}}(\boldsymbol{FD}/\hat{\boldsymbol{B}})$ gemäß Gl. (4.44) zu

$$p_{\boldsymbol{FD}/\hat{\boldsymbol{B}}}(\boldsymbol{FD}/\hat{\boldsymbol{B}}) = \frac{1}{\sqrt{(2\pi)^{\mathcal{J}} |[\boldsymbol{U}]|}} \exp\left(-\frac{1}{2} \cdot \left(\boldsymbol{FD} - [\boldsymbol{O}] \cdot \hat{\boldsymbol{B}}\right)^T [\boldsymbol{U}]^{-1} \left(\boldsymbol{FD} - [\boldsymbol{O}] \cdot \hat{\boldsymbol{B}}\right) \right) \tag{4.64}$$

Der Maximum–Likelihood–Schätzwert $\hat{\boldsymbol{B}}_{ML}$ erhält man durch Einsetzen von Gl. (4.64) in Gl. (4.51)

$$\frac{\partial}{\partial \hat{\boldsymbol{B}}}\left(\left(\boldsymbol{FD} - [\boldsymbol{O}] \cdot \hat{\boldsymbol{B}}\right)^T [\boldsymbol{U}]^{-1}\left(\boldsymbol{FD} - [\boldsymbol{O}] \cdot \hat{\boldsymbol{B}}\right)\right)\Big|_{\hat{\boldsymbol{B}} = \hat{\boldsymbol{B}}_{ML}} = 0 \tag{4.65}$$

Die Lösung der Gl. (4.65) ergibt sich zu

$$\hat{\boldsymbol{B}}_{ML} = \left([\boldsymbol{O}]^T [\boldsymbol{U}]^{-1} [\boldsymbol{O}]\right)^{-1} [\boldsymbol{O}]^T [\boldsymbol{U}]^{-1} \boldsymbol{FD}. \tag{4.66}$$

Durch Ersetzen von Gl. (4.62) in Gl. (4.66) folgt der Schätzfehler der Parameter der dreidimensionalen Bewegung

$$\hat{\boldsymbol{B}}_{ML} - \boldsymbol{B}^{(\mathcal{P})} = \left([\boldsymbol{O}]^T [\boldsymbol{U}]^{-1} [\boldsymbol{O}]\right)^{-1} [\boldsymbol{O}]^T [\boldsymbol{U}]^{-1} \boldsymbol{V}. \tag{4.67}$$

Mit Gl. (4.66) ergibt sich die Kovarianzmatrix des Schätzfehlers der Parameter der dreidimensionalen Bewegung zu

$$[\boldsymbol{Q}]_{ML} = E\left[(\hat{\boldsymbol{B}}_{ML} - \boldsymbol{B}^{(\mathcal{P})}) \cdot (\hat{\boldsymbol{B}}_{ML} - \boldsymbol{B}^{(\mathcal{P})})^T\right] = \left([\boldsymbol{O}]^T [\boldsymbol{U}]^{-1} [\boldsymbol{O}]\right)^{-1}. \tag{4.68}$$

Der Maximum–Likelihood–Schätzer gemäß Gl. (4.66) kann auch rekursiv formuliert werden. Die Beobachtungspunkte werden dann nicht parallel gemäß Gl. (4.66) ausgewertet, sondern beginnend vom Beobachtungspunkt $\boldsymbol{W}^{(0)}$ nacheinander berechnet. Nach $j + 1$ Beobachtungspunkten ergibt sich Gl. (4.66) zu

$$\hat{\boldsymbol{B}}_{j+1} = \left([\boldsymbol{O}^{(j+1)}]^T [\boldsymbol{U}^{(j+1)}]^{-1} [\boldsymbol{O}^{(j+1)}]\right)^{-1} [\boldsymbol{O}^{(j+1)}]^T [\boldsymbol{U}^{(j+1)}]^{-1} \boldsymbol{FD}^{(j+1)}, \tag{4.69}$$

wobei

$$\boldsymbol{FD}^{(j+1)} = \begin{bmatrix} fd^{(j+1)} \\ fd^{(j)} \\ fd^{(j-1)} \\ fd^{(j-2)} \\ \vdots \\ fd^{(0)} \end{bmatrix} = \begin{bmatrix} fd^{(j+1)} \\ \boldsymbol{FD}^{(j)} \end{bmatrix} \tag{4.70}$$

$$[\boldsymbol{O}^{(j+1)}] = \begin{bmatrix} \boldsymbol{o}^{(j+1)T} \\ [\boldsymbol{O}^{(j)}] \end{bmatrix} \tag{4.71}$$

$$\boldsymbol{V}^{(j+1)} = \begin{bmatrix} v^{(j+1)} \\ \boldsymbol{V}^{(j)} \end{bmatrix} \tag{4.72}$$

$$[\boldsymbol{U}^{(j+1)}] = \begin{bmatrix} \sigma^2_{\Delta s(j+1)} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{\Delta s(j)} & 0 & . & . & 0 \\ 0 & 0 & \sigma^2_{\Delta s(j-1)} & . & . & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\Delta s(0)} \end{bmatrix} = \begin{bmatrix} \sigma^2_{\Delta s(j+1)} & 0 & 0 & . & . & 0 \\ 0 & & & & & \\ 0 & & & & & \\ . & & [\boldsymbol{U}^{(j)}] & & & \\ . & & & & & \\ 0 & & & & & \end{bmatrix}. \tag{4.73}$$

Durch Umformung der Gl. (4.69) ergeben sich die Gleichungen des rekursiven Maximum–Likelihood–Schätzers zu

$$\hat{\boldsymbol{B}}_{j+1} = \hat{\boldsymbol{B}}_j + \boldsymbol{K}^{(j+1)} \cdot \left(fd^{(j+1)} - \boldsymbol{o}^{(j+1)T} \cdot \hat{\boldsymbol{B}}_j \right) \tag{4.74}$$

$$\boldsymbol{K}^{(j+1)} = [\boldsymbol{Q}^{(j)}] \cdot \boldsymbol{o}^{(j+1)} \cdot \left(\boldsymbol{o}^{(j+1)T} \cdot [\boldsymbol{Q}^{(j)}] \cdot \boldsymbol{o}^{(j+1)} + \sigma^2_{\Delta s^{(j+1)}} \right)^{-1} \tag{4.75}$$

$$[\boldsymbol{Q}^{(j+1)}] = [\boldsymbol{Q}^{(j)}] - \boldsymbol{K}^{(j+1)} \cdot \boldsymbol{o}^{(j+1)T} \cdot [\boldsymbol{Q}^{(j)}]. \tag{4.76}$$

Dabei stellt $\hat{\boldsymbol{B}}_j$ den mit dem vorherigen Beobachtungspunkt $\boldsymbol{W}^{(j)}$ erhaltenen Schätzwert dar. $\boldsymbol{K}^{(j+1)}$ wird in der Literatur als Korrekturvektor bezeichnet und $[\boldsymbol{Q}^{(j+1)}]$ stellt die neue Kovarianzmatrix des Schätzfehlers dar. Gl. (4.74), Gl. (4.75) und Gl. (4.76) entsprechen den Gleichungen eines Kalman–Filters[8][55]. Anstelle einer Initialisierung des rekursiven Maximum–Likelihood–Schätzers mit dem ersten Beobachtungspunkt $\boldsymbol{W}^{(0)}$ kann die Initialisierung unter Verwendung von Gl. (4.66) und Gl. (4.68) mit mehreren Beobachtungspunkten begonnen werden.

Da der Zusammenhang gemäß Gl. (4.54) i. allg. nur für kleine Differenzen $(\boldsymbol{h}_{k+1} - \boldsymbol{h}_k)$ hinreichend genau gültig ist, soll das rekursive Bewegungsschätzverfahren zusätzlich als Iterationsalgorithmus ausgelegt werden. Dazu wird das Oberflächenelement gemäß Gl. (4.48) mit den geschätzten Bewegungsparametern bewegt und auf diese Weise die Bewegung kompensiert. Im nächsten Iterationsschritt erfolgt eine erneuerte Schätzung der Bewegungsparameter, die zur Aktualisierung der in der vorherigen Iteration geschätzten Bewegungsparameter dient. Dabei tritt an die Stelle von *fd* gemäß Gl. (4.61) nun die sog. *dfd* (englisch: *d*isplaced *f*rame *d*ifference), die in Folge der Bewegungskompensation in jedem Iterationsschritt verkleinert wird. In Folge der Bewegungskompensation verschiebt sich ein Beobachtungspunkt von $\boldsymbol{h}_k^{(j)}$ nach $\boldsymbol{h}_i^{(j)}$, wobei $\boldsymbol{h}_i^{(j)}$ die Position eines jeden Beobachtungspunktes in der Bildebene bei der *i*–ten Iteration darstellt. Das Iterationsverfahren wird beendet, wenn die Ungleichung

$$\left| \frac{1}{\vartheta} \sum_{a=0}^{\vartheta-1} \left(dfd\left(\boldsymbol{h}_i^{(a)}, \hat{\boldsymbol{B}}_{\vartheta-1,i-1} \right) \right)^2 \right| - \left| \frac{1}{\vartheta} \sum_{a=0}^{\vartheta-1} \left(dfd\left(\boldsymbol{h}_{i+1}^{(a)}, \hat{\boldsymbol{B}}_{\vartheta-1,i} \right) \right)^2 \right| < T_h \tag{4.77}$$

erfüllt wird. Dabei stellen $\hat{\boldsymbol{B}}_{\vartheta-1,i-1}$ und $\hat{\boldsymbol{B}}_{\vartheta-1,i}$ die initialen Bewegungsparameter in der *i*–ten bzw. *(i+1)*–ten Iteration dar. Für die Berechnung der Gesamtbewegung eines Oberflächenelementes $\mathcal{S}$ wird die Bewegungsgleichung

$$\boldsymbol{H}_{i+1} = [\boldsymbol{R}^{(u)}_{\vartheta-1,i}] \cdot (\boldsymbol{H}_i - \boldsymbol{G}_i) + \boldsymbol{G}_i + \boldsymbol{T}^{(u)}_{\vartheta-1,i} \tag{4.78}$$

verwendet. Dabei stellt $\boldsymbol{T}^{(u)}_{\vartheta-1,i}$ der bei der *i*–ten Iteration geschätzte Update der Translationsparameter und $[\boldsymbol{R}^{(u)}_{\vartheta-1,i}]$ der bei der *i*–ten Iteration geschätzte Update der Rotationsmatrix dar. $\boldsymbol{G}_i$ beschreibt die aktuelle Position des Schwerpunktes des Oberflächenelementes bei der *i*–ten Iteration.

Die Gesamtbewegung eines Oberflächenelementes $\mathcal{S}$ ergibt sich nach *i* Iterationen durch wiederholte Anwendung der Gleichung (4.78) und Umformung zu

$$\boldsymbol{H}_{i+1} = [\boldsymbol{R}_{\vartheta-1,i}] \cdot (\boldsymbol{H}_1 - \boldsymbol{G}_1) + \boldsymbol{G}_1 + \boldsymbol{T}_{\vartheta-1,i}, \tag{4.79}$$

mit

$$[\boldsymbol{R}_{\vartheta-1,i}] = [\boldsymbol{R}^{(u)}_{\vartheta-1,i}] \cdot [\boldsymbol{R}^{(u)}_{\vartheta-1,i-1}] \cdot \ldots \cdot [\boldsymbol{R}^{(u)}_{\vartheta-1,1}] \tag{4.80}$$

$$\begin{aligned} \boldsymbol{T}_{\vartheta-1,i} = \boldsymbol{T}^{(u)}_{\vartheta-1,i} + \boldsymbol{G}_i - \boldsymbol{G}_1 + [\boldsymbol{R}^{(u)}_{\vartheta-1,i}] \cdot [\boldsymbol{R}^{(u)}_{\vartheta-1,i-1}] \cdot \ldots \cdot [\boldsymbol{R}^{(u)}_{\vartheta-1,1}] \cdot \boldsymbol{G}_1 + \\ [\boldsymbol{R}^{(u)}_{\vartheta-1,i}] \cdot \left(\boldsymbol{T}^{(u)}_{\vartheta-1,i-1} + \boldsymbol{G}_{i-1} - \boldsymbol{G}_i + \right. \\ \left. [\boldsymbol{R}^{(u)}_{\vartheta-1,i-1}] \cdot \left(\boldsymbol{T}^{(u)}_{\vartheta-1,i-2} + \boldsymbol{G}_{i-2} - \boldsymbol{G}_{i-1} + \ldots \right) \right). \end{aligned} \tag{4.81}$$

Der Gl. (4.59) liegen Modellannahmen zugrunde, wie beispielsweise die Annahme, daß ein Oberflächenelement starr ist. In realen Bildszenen sind diese Annahmen nicht immer erfüllt. Auf diese Weise kann ein Beobachtungspunkt auftreten, dessen Meßwert die Modellannahmen nicht erfüllt und der hier als Ausreißer bezeichnet wird. Zur Verbesserung der Konvergenz des Schätzverfahrens sollen Ausreißer unter den Beobachtungspunkten für jede Iteration *i* detektiert und für diese Iteration aus der Bewegungsschätzung ausgeschlossen werden[35][65][91]. Dazu wird hier die rekursive Verarbeitung des Maximum–Likelihood–Schätzers gemäß Gl. (4.74), Gl. (4.75) und Gl. (4.76) ausgenutzt. Ein Beobachtungspunkt $\boldsymbol{W}_i^{(j+1)}$ wird genau dann als Ausreißer detektiert, wenn die neuen geschätzten Bewegungsparameter $\hat{\boldsymbol{B}}_{j+1,i}$ die mittlere quadratische *dfd* gemittelt über alle Beobachtungspunkte vergrößern anstelle zu verkleinern, d.h.

$$\frac{1}{\vartheta}\sum_{a=0}^{\vartheta-1}\left(dfd\left(\boldsymbol{h}_i^{(a)},\hat{\boldsymbol{B}}_{j,i}\right)\right)^2 \leq \frac{1}{\vartheta}\sum_{a=0}^{\vartheta-1}\left(dfd\left(\boldsymbol{h}_i^{(a)},\hat{\boldsymbol{B}}_{j+1,i}\right)\right)^2 \tag{4.82}$$

erfüllt wird.

Falls der Beobachtungspunkt $\boldsymbol{W}_i^{(j+1)}$ als Ausreißer detektiert wurde, wird der rekursive Maximum–Likelihood–Schätzer auf dem Stand des Beobachtungspunktes $\boldsymbol{W}_i^{(j)}$ zurückgesetzt. Anschließend wird mit dem nächsten Beobachtungspunkt $\boldsymbol{W}_i^{(j+2)}$ fortgefahren.

Die Auswertung der Ungleichung (4.82) ist sehr rechenintensiv. Um den Rechenaufwand bei der Detektion von Ausreißern möglichst klein zu halten, wird vor der Auswertung des Kriteriums gemäß Gl. (4.82) ein aus der Literatur bekanntes Kriterium überprüft, das weniger rechenaufwendig ist. Bei diesem Kriterium wird angenommen, daß für jede *i*–te Iteration Ausreißer durch einen großen Wert des *dfd* erkannt werden können[27][65]. Ein Beobachtungspunkt $\boldsymbol{W}_i^{(j+1)}$ wird dabei als Ausreißer detektiert, wenn die Bedingung

$$|dfd(\boldsymbol{h}_i^{(j+1)},\hat{\boldsymbol{B}}_{J-1,i-1})| > \sigma_{dfd} \tag{4.83}$$

mit

$$\sigma_{dfd} = \sqrt{\frac{1}{\vartheta}\sum_{a=0}^{\vartheta-1}\left(dfd(\boldsymbol{h}_i^{(a)},\hat{\boldsymbol{B}}_{\vartheta-1,i-1})\right)^2} \tag{4.84}$$

erfüllt wird. Anderenfalls wird das Kriterium gemäß Gl. (4.82) ausgewertet. Im Vergleich zu Gl. (4.82), wo die Bewegungskompensation für jeden Beobachtungspunkt mit $\hat{\boldsymbol{B}}_{j,i}$ und $\hat{\boldsymbol{B}}_{j+1,i}$ ausgeführt werden muß, wird in diesem Fall nur eine Bewegungskompensation für alle Beobachtungspunkte erforderlich.

Besteht das Oberflächenelement aus genau einem Dreieck, werden zur weiteren Verbesserung der Konvergenz des Schätzverfahrens auch die Beobachtungspunkte direkt benachbarter Dreiecke in die Schätzung miteinbezogen. Bild 4.24 zeigt die benachbarten Dreiecke eines beliebigen Dreieckes eines starren Modellobjektes, dessen Parameter der dreidimensionalen Bewegung geschätzt werden sollen. Benachbarte Dreiecke sind diejenige, die mindestens einen Kontrollpunkt mit dem zu untersuchenden Dreieck teilen.

Experimentelle Untersuchungen

Für die Beurteilung des Verfahrens zur Schätzung der dreidimensionalen Bewegung eines relativ kleinen starren Oberflächenelementes wird als Gütekriterium die Konvergenzwahrscheinlichkeit zu den korrekten Bewegungsparametern gemessen. Je größer die Konvergenzwahrscheinlichkeit, desto zuverlässiger ist das Schätzverfahren. Dabei wird als Referenz auf ein aus der Literatur bekanntes Gradientenverfahren[45][62][65] Bezug genommen. Die Schätzverfahren werden auf synthetisch erzeugte Bildpaare angewendet.

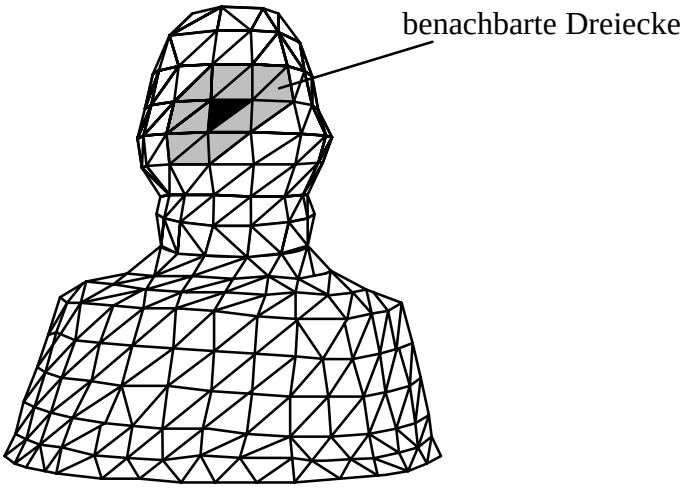


Bild 4.24 Zur Bewegungsschätzung eines Dreieckes werden Beobachtungspunkte benachbarter Dreiecke in die Schätzung miteinbezogen.

Zur Erzeugung eines Bildpaares wird zunächst ein starres Modellobjekt gemäß Abschnitt 4.2.1 aus der Testsequenz *Claire* im CIF–Format generiert. Der sichtbare Oberflächenteil dieses Modellobjektes wird durch 280 Stützpunkte und 491 Dreiecke dargestellt. Dabei beträgt die mittlere Größe der zu einem Dreieck korrespondierenden Bildfläche etwa 65 pel^2. Das erste Bild entsteht durch die Synthese eines Modellbildes gemäß Kapitel 5. Vor der Synthese des zweiten Bildes wird zunächst auf die Position $\boldsymbol{P}^{(n)} = (P_X^{(n)}, P_Y^{(n)}, P_Z^{(n)})^T$ eines jeden Stützpunktes $P^{(n)}$ des starren Modellobjektes ein Gaußverteilter Positionsfehler $\Delta\boldsymbol{P} = (\Delta P_X, \Delta P_Y, \Delta P_Z)^T$ der Varianz $\sigma_X^2 = \sigma_Y^2 = \sigma_Z^2 = 0.01 pel^2$ addiert. Auf diese Weise entsteht ein Fehler der dreidimensionalen Form. Das starre Modellobjekt wird dann mit den Bewegungsparametern (T_X=2pel, T_Y=2pel, T_Z=2pel, R_X=1grad, R_Y=1grad, R_Z=1grad) bewegt. Auf das Testbildpaar wird abschließend weißes Gaußsches Rauschen mit einer Leistung gemessen in PSNR von 40 dB addiert. In den folgenden Experimenten wird nun auf das starre Modellobjekt, welches zum ersten Bild korrespondiert, zurückgegriffen.

Für jedes sichtbare Dreieck des starren Modellobjektes erfolgt nun gemäß Abschnitt 4.3.2 die Schätzung der Parameter der dreidimensionalen Bewegung durch Auswertung des zweiten Bildes. Für jedes sichtbare Dreieck des Modellobjektes ergeben sich der maximale Fehler $e_{max,T}$ der Translationsparameter und der maximale Fehler $e_{max,R}$ der Rotationsparameter zu

$$
\begin{aligned}
e_{max,T} &= \max\left\{|T_X - \hat{T}_{ML,X}| \ , \ |T_Y - \hat{T}_{ML,Y}| \ , \ |T_Z - \hat{T}_{ML,Z}| \right\} \\
e_{max,R} &= \max\left\{|R_X - \hat{R}_{ML,X}| \ , \ |R_Y - \hat{R}_{ML,Y}| \ , \ |R_Z - \hat{R}_{ML,Z}| \right\},
\end{aligned}
\tag{4.85}
$$

wobei $\hat{\boldsymbol{A}}_{ML} = (\hat{T}_{ML,X}, \hat{T}_{ML,Y}, \hat{T}_{ML,Z}, \hat{R}_{ML,X}, \hat{R}_{ML,Y}, \hat{R}_{ML,Z})$ die geschätzten Bewegungsparameter darstellen. Für die Bewegungsparameter eines Dreiecks wird Konvergenz angenommen, wenn die Bedingung

$$
e_{max,T} \overset{!}{<} 0.5 \text{ pel} \quad ; \quad e_{max,R} \overset{!}{<} 0.5 \text{ grad} \tag{4.86}
$$

erfüllt ist. $e_{max,T}$ und $e_{max,R}$ definieren den Konvergenzbereich und wurden hier heuristisch zu 0.5 pel bzw. 0.5 grad gesetzt. Die Konvergenzwahrscheinlichkeit wird hierbei als

$$
p = \frac{N_D}{491} \tag{4.87}
$$

definiert. Dabei stellt N_D die Anzahl der Dreiecke dar, für die die Gl. (4.86) erfüllt ist.

Es werden nun mehrere Bildpaare mit einer Leistung des Bildrauschens von 40 dB erzeugt. Für jedes Bildpaar wird dann die Konvergenzwahrscheinlichkeit gemäß Gl. (4.87) gemessen. Für das im Rahmen dieser Arbeit entwickelte Verfahren zur Schätzung der Parameter der dreidimensionalen Bewegung eines Dreieckes ergab sich eine mittlere Konvergenzwahrscheinlichkeit gemäß Gl. (4.87) von 0.8. Verwendet man zur Schätzung der Parameter der dreidimensionalen Bewegung eines Dreiecks ein aus der Literatur bekanntes Gradientenverfahren[45][62][65], ergibt sich eine mittlere Konvergenzwahrscheinlichkeit von 0.1.

4.3.3 Verfahren zur Schätzung der Parameter der dreidimensionalen Bewegung gegliederter Objekte

Für die Bestimmung der Parameter der dreidimensionalen Bewegung eines gegliederten Objektes befinden sich in der Literatur zwei Ansätze[2]. Bei dem ersten Ansatz werden mittels eines Schätzverfahrens alle sechs Bewegungsparameter einer jeden Objektkomponente eines gegliederten Objektes ohne Berücksichtigung der Gelenke bestimmt[44][47]. Bei dem zweiten Ansatz werden für jede Objektkomponente diejenigen Bewegungsparameter aus der Schätzung herausgenommen, die aufgrund der von den Gelenken der Objektkomponenten auferlegten Einschränkungen der Bewegungsmöglichkeiten von anderen Bewegungsparametern abhängig sind[26][50][85]. Damit soll die Genauigkeit der Schätzung verbessert werden. In dieser Arbeit wird der zweite Ansatz verfolgt. Für die Schätzung aller sechs Bewegung-

sparameter einer Objektkomponente wird das in Abschnitt 4.3.2 entwickelte Maximum–Likelihood–Schätzverfahren für die Parameter der dreidimensionalen Bewegung von Oberflächenelementen verwendet. Für die Schätzung nur der unabhängigen Bewegungsparameter einer Objektkomponente wird unter Berücksichtigung der eingeschränkten Bewegungsmöglichkeiten der Objektkomponenten das in Abschnitt 4.3.2 entwickelte Maximum–Likelihood–Schätzverfahren so erweitert, daß nur die unabhängigen Bewegungsparameter bestimmt werden.

Verfahren zur Schätzung der Parameter der dreidimensionalen Bewegung gegliederter Objekte ohne Berücksichtigung von Gelenken

Dieses Verfahren wird benötigt, um einen Vergleich zwischen der Bewegungsschätzung ohne Berücksichtigung von Gelenken mit der Bewegungsschätzung unter Berücksichtigung von Gelenken durchzuführen. Ferner ist es zur Bestimmung der Position der Gelenke gemäß Abschnitt 4.2.4 erforderlich. Zur Schätzung der Parameter der dreidimensionalen Bewegung eines gegliederten Objektes ohne Berücksichtigung der Gelenke werden alle sechs Bewegungsparameter für jede Objektkomponente eines gegliederten Objektes geschätzt. Zur Schätzung der sechs Parameter der dreidimensionalen Bewegung einer Objektkomponente $\mathcal{C}^{(\delta)}$ wird hier das in Abschnitt 4.3.2 beschriebene Maximum–Likelihood–Schätzverfahren für die Parameter der dreidimensionalen Bewegung von Oberflächenelementen verwendet. Dazu wird als Oberflächenelement $\mathcal{S}$ die Gesamtoberfläche der Objektkomponente $\mathcal{C}^{(\delta)}$ betrachtet. Beobachtungspunkte aus allen jeweils sichtbaren Dreiecken der Objektkomponente werden dabei für die Bewegungsschätzung entnommen.

Verfahren zur Schätzung der Parameter der dreidimensionalen Bewegung gegliederter Objekte unter Berücksichtigung von Gelenken

Bei der Schätzung der Parameter der dreidimensionalen Bewegung gegliederter Objekte unter Berücksichtigung von Gelenken werden nur die unabhängigen Bewegungsparameter für jede Objektkomponente eines gegliederten Objektes bestimmt. Für ein gegliedertes Objekt mit Baumstruktur, dessen Objektkomponenten durch Kugelgelenke gekoppelt sind, werden für die Wurzelobjektkomponente sechs Bewegungsparameter und für die restlichen Objektkomponenten jeweils nur die Rotationsparameter geschätzt. Gemäß Abschnitt 3.4.3.1 wird als Wurzel der Baumstruktur die größte Objektkomponente eines gegliederten Objektes definiert. Sie wird als $\mathcal{C}^{(0)}$ gekennzeichnet.

Zur Schätzung aller sechs Bewegungsparameter der Wurzelobjektkomponente $\mathcal{C}^{(0)}$ wird das in Abschnitt 4.3.2 beschriebene Maximum–Likelihood–Schätzverfahren für die Parameter der dreidimensionalen Bewegung von Oberflächenelementen verwendet. Dazu wird als Oberflächenelement $\mathcal{S}$ die Gesamtoberfläche der Wurzelobjektkomponente $\mathcal{C}^{(0)}$ betrachtet. Beobachtungspunkte aus allen jeweils sichtbaren Dreiecken der Wurzelobjektkomponente werden dabei in die Schätzung miteinbezogen.

Nach der Schätzung der sechs Bewegungsparameter der Wurzelobjektkomponente werden jeweils nur die Rotationsparameter der restlichen Objektkomponenten von der Wurzelobjektkomponente beginnend nacheinander geschätzt. Die Schätzung der Rotationsparameter einer Objektkomponente wird mit Hilfe von Bild 4.25 und Bild 4.26 beschrieben. Bild 4.25 zeigt

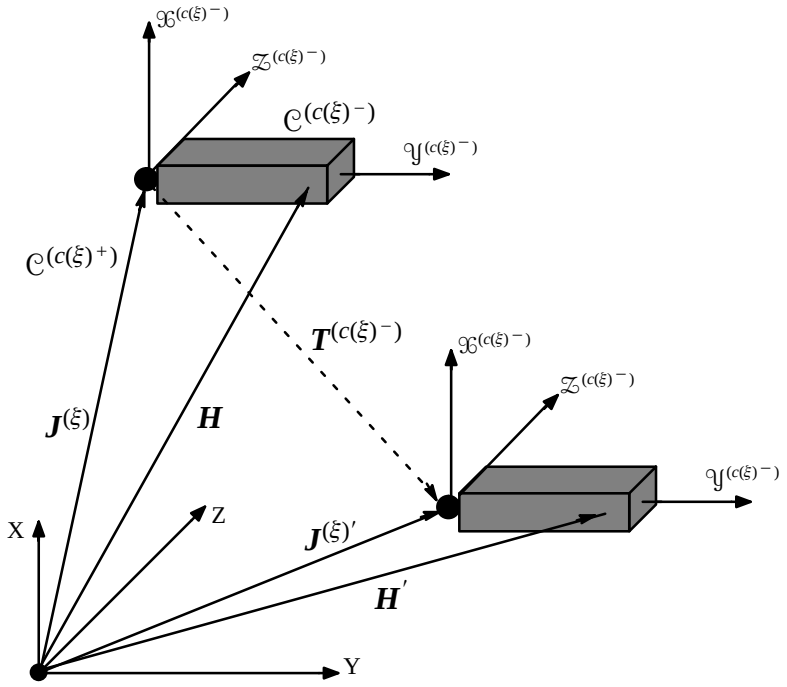


Bild 4.25 Zur Schätzung der Rotationsparameter der Relativobjektkomponente $\mathcal{C}^{(c(\xi)^-)}$ wird zunächst diese Objektkomponente mit ihren abhängigen Translationsparametern $\boldsymbol{T}^{(c(\xi)^-)}$ verschoben. Diese Translationsparameter werden mittels der Gelenkbindung aus den Bewegungsparametern und der Position der Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ berechnet. Dabei verschiebt sich ein Oberflächenpunkt H mit dem Ortsvektor $\boldsymbol{H}$ nach $\boldsymbol{H}'$.

zwei durch ein Kugelgelenk $J^{(\xi)}$ gekoppelte Objektkomponenten eines gegliederten Objektes. Gemäß Abschnitt 3.4.3.1 werden die zwei durch ein Kugelgelenk $J^{(\xi)}$ gekoppelten Objektkomponenten eines gegliederten Objektes als $\mathcal{C}^{(c(\xi)^+)}$ und $\mathcal{C}^{(c(\xi)^-)}$ bezeichnet. Die Funktionen $c(\xi)^+$ und $c(\xi)^-$ geben die zur Kennzeichnung der Bezugs– bzw. Relativobjektkomponente eines Kugelgelenkes $J^{(\xi)}$ verwendeten Nummern δ an. Als Bezugsobjektkomponente wird diejenige der beiden angekoppelten Objektkomponenten ausgewählt, die im Baum näher an der Wurzelobjektkomponente liegt. Im Koordinatensystem der Bezugsobjektkomponente

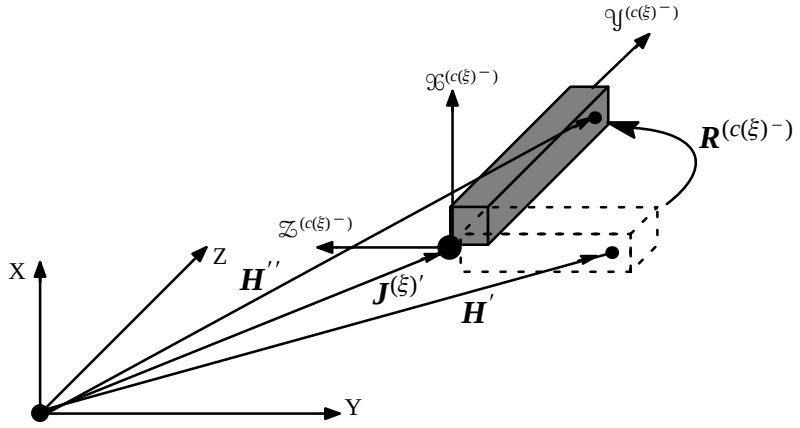


Bild 4.26 Zur Schätzung der Rotationsparameter einer Objektkomponente $\mathcal{C}^{(c(\xi)^-)}$ wird die Bewegung dieser Objektkomponente nach ihrer Verschiebung mit den abhängigen Translationsparametern $\boldsymbol{T}^{(c(\xi)^-)}$ durch eine Rotation $\boldsymbol{R}^{(c(\xi)^-)}$ ihres Koordinatensystems beschrieben. Dabei verschiebt sich ein Beobachtungspunkt *H* mit dem Ortsvektor $\boldsymbol{H}'$ nach $\boldsymbol{H}''$.

$\mathcal{C}^{(c(\xi)^+)}$ wird die Position des Kugelgelenkes $J^{(\xi)}$ dargestellt. Das Koordinatensystem $(\mathcal{X}^{(c(\xi)^-)}, \mathcal{Y}^{(c(\xi)^-)}, \mathcal{Z}^{(c(\xi)^-)})$ der anderen angekoppelten Objektkomponente, der sog. Relativobjektkomponente $\mathcal{C}^{(c(\xi)^-)}$, hat den Ursprung am Kugelgelenk $J^{(\xi)}$ mit dem Ortsvektor $\boldsymbol{J}^{(\xi)}$.

Die Schätzung der Rotationsparameter einer Objektkomponente $\mathcal{C}^{(c(\xi)^-)}$ erfolgt in zwei Schritten. Im ersten Schritt werden zunächst die abhängigen Translationsparameter $\boldsymbol{T}^{(c(\xi)^-)}$ mittels der Gelenkbindung gemäß Gl. (3.23) berechnet. Die Objektkomponente wird mit den berechneten Translationsparametern verschoben (Bild 4.25). Ein Punkt *H* auf der Oberfläche der Relativobjektkomponente mit dem Ortsvektor $\boldsymbol{H}$ bewegt sich daher nach $\boldsymbol{H}'$ gemäß

$$\begin{aligned} \boldsymbol{H}' &= \boldsymbol{H} + \boldsymbol{T}^{(c(\xi)^-)} \\ \boldsymbol{H}' &= \boldsymbol{H} + \left([\boldsymbol{R}^{(c(\xi)^+)}] \cdot (\boldsymbol{J}^{(\xi)} - \boldsymbol{J}^{(c(\xi)^+)}) + \boldsymbol{J}^{(c(\xi)^+)} + \boldsymbol{T}^{(c(\xi)^+)} - \boldsymbol{J}^{(\xi)}\right) \\ \Rightarrow \boldsymbol{T}^{(c(\xi)^-)} &= \boldsymbol{H}' - \boldsymbol{H} = [\boldsymbol{R}^{(c(\xi)^+)}] \cdot (\boldsymbol{J}^{(\xi)} - \boldsymbol{J}^{(c(\xi)^+)}) + \boldsymbol{J}^{(c(\xi)^+)} + \boldsymbol{T}^{(c(\xi)^+)} - \boldsymbol{J}^{(\xi)}, \end{aligned} \tag{4.88}$$

wobei $\boldsymbol{T}^{(c(\xi)^+)}$ und $[\boldsymbol{R}^{(c(\xi)^+)}]$ den Translationsvektor bzw. die Rotationsmatrix der Bezugsobjektkomponente darstellen. $\boldsymbol{J}^{(c(\xi)^+)}$ ist der Ursprung des Koordinatensystems der Bezugsobjektkomponente. Da die Bezugsobjektkomponente $\mathcal{C}^{(c(\xi)^+)}$ näher an der Wurzelobjektkomponente liegt, wird hierbei angenommen, daß $\boldsymbol{T}^{(c(\xi)^+)}$ und $[\boldsymbol{R}^{(c(\xi)^+)}]$ in einem vorherigen Schritt bestimmt worden sind und auch $\boldsymbol{J}^{(c(\xi)^+)}$ bekannt ist. Aus Gl. (4.88) ergibt sich die neue Position $\boldsymbol{J}^{(\xi)'}$ des Kugelgelenkes, indem $\boldsymbol{H}' = \boldsymbol{J}^{(\xi)'}$ und $\boldsymbol{H} = \boldsymbol{J}^{(\xi)}$ eingesetzt werden

$$\boldsymbol{J}^{(\xi)'} = [\boldsymbol{R}^{(c(\xi)^+)}] \cdot (\boldsymbol{J}^{(\xi)} - \boldsymbol{J}^{(c(\xi)^+)}) + \boldsymbol{J}^{(c(\xi)^+)} + \boldsymbol{T}^{(c(\xi)^+)}. \tag{4.89}$$

Im zweiten Schritt werden die Rotationsparameter $\boldsymbol{R}^{(c(\xi)^-)}$ der Objektkomponente $\mathcal{C}^{(c(\xi)^-)}$ geschätzt. Dazu wird der in Abschnitt 4.3.2 entwickelte Maximum–Likelihood–Schätzer für Parameter der dreidimensionalen Bewegung eines Oberflächenelementes so modifiziert, daß nur die Rotationsparameter bestimmt werden. Als Oberflächenelement $\mathcal{S}$ wird die Gesamtoberfläche der Objektkomponente $\mathcal{C}^{(c(\xi)^-)}$ betrachtet. Dabei werden Beobachtungspunkte aus allen Dreiecken der Objektkomponente für die Bewegungsschätzung verwendet. Da die Translation der Objektkomponente im ersten Schritt bereits kompensiert worden ist, wird die Bewegung der Objektkomponente nur durch eine Rotation $\boldsymbol{R}^{(c(\xi)^-)}$ ihres Koordinatensystems beschrieben (Bild 4.26). Ein Punkt H auf der Oberfläche dieser Objektkomponente mit dem Ortsvektor $\boldsymbol{H}'$ bewegt sich nach $\boldsymbol{H}''$ gemäß

$$\boldsymbol{H}'' = [\boldsymbol{R}^{(c(\xi)^-)}] \cdot (\boldsymbol{H}' - \boldsymbol{J}^{(\xi)'}) + \boldsymbol{J}^{(\xi)'}. \tag{4.90}$$

Dabei stellt $[\boldsymbol{R}^{(c(\xi)^-)}]$ die Rotationsmatrix dar, die mit den zu schätzenden Rotationswinkeln des Vektors $\boldsymbol{R}^{(c(\xi)^-)}$ gemäß Gl. (3.6) aufgestellt wird. Die Schätzung der Rotationswinkel wird nachfolgend beschrieben.

Ein Beobachtungpunkt $\boldsymbol{W}_k^{(j)}$ an der Stelle $\boldsymbol{H}_k^{(j)}$ zum Zeitpunkt k befindet sich nach der Kompensation gemäß Gl. (4.88) an der Stelle $\boldsymbol{H}_k^{(j)'}$. Dabei verändert sich seine Projektion in der Bildebene von $\boldsymbol{h}_k^{(j)}$ nach $\boldsymbol{h}_k^{(j)'}$. Der Zusammenhang zwischen der *dfd* an der Stelle $\boldsymbol{h}_k^{(j)'}$ und dem korrespondierenden Beobachtungspunkt $\boldsymbol{W}_{k+1}^{(j)}$ an der Stelle $\boldsymbol{H}_{k+1}^{(j)}$ ergibt sich gemäß Gl. (4.56) bei Weglassen des Indexes j zu

$$dfd(\boldsymbol{h}_k', \boldsymbol{T}^{(c(\xi)^-)}) = -F \cdot \bar{\boldsymbol{g}} \cdot \left(\frac{H_{X,k+1}}{H_{Z,k+1}} - \frac{H'_{X,k}}{H'_{Z,k}}, \frac{H_{Y,k+1}}{H_{Z,k+1}} - \frac{H'_{Y,k}}{H'_{Z,k}} \right)^T. \tag{4.91}$$

Dabei wird $\bar{\boldsymbol{g}}$ gemäß Gl. (4.55) an der Stelle $\boldsymbol{h}_k'$ gemessen. Die Positionen $\boldsymbol{H}_k'$ und $\boldsymbol{H}_{k+1}$ sind über die Bewegungsgleichung (4.90)

$$\boldsymbol{H}_{k+1} = [\boldsymbol{R}^{(c(\xi)^-)}] \cdot (\boldsymbol{H}_k' - \boldsymbol{J}^{(\xi)'}) + \boldsymbol{J}^{(\xi)'} \tag{4.92}$$

verbunden. Durch Einsetzen von Gl. (4.92) in Gl. (4.91) unter Annahme kleiner Rotationswinkel gemäß Gl. (4.58) ergibt sich die *dfd* in Abhängigkeit von den unbekannten Rotationsparametern $\boldsymbol{R}^{(c(\xi)^-)}$ bei Weglassen des Indexes k zu

$$
\begin{aligned}
dfd(\boldsymbol{h}', \boldsymbol{T}^{(c(\xi)^-)}) = & -[[H'_X\bar{g}_x(H'_Y - J'_Y) + H'_Y\bar{g}_y(H'_Y - J'_Y) \\
& + H'_Z\bar{g}_y(H'_Z - J'_Z)]F/H'^2_Z \\
& + dfd(\boldsymbol{h}', \boldsymbol{T}^{(c(\xi)^-)})/H'_Z(H'_Z - J'_Y)] \cdot R_X \\
& + [[H'_Y\bar{g}_y(H'_X - J'_X) + H'_X\bar{g}_x(H'_X - J'_X) \\
& + H'_Z\bar{g}_x(H'_Z - J'_Z)]F/H'^2_Z \\
& + dfd(\boldsymbol{h}', \boldsymbol{T}^{(c(\xi)^-)})/H'_Z(H'_X - J'_X)] \cdot R_Y \\
& - [\bar{g}_x(H'_Y - J'_Y) - \bar{g}_y(H'_X - J'_X)]F/H'_Z \cdot R_Z
\end{aligned}
\tag{4.93}
$$

Interpretiert man nun die Koeffizienten der 3 Rotationsparameter als Spaltenvektor $\boldsymbol{o}$, so geht Gl. (4.93) in

$$
dfd(\boldsymbol{h}', \boldsymbol{T}^{(c(\xi)^-)}) = \boldsymbol{o}^T \cdot \boldsymbol{R}^{(c(\xi)^-)} \tag{4.94}
$$

über. Zur Berücksichtigung des Luminanzfehlers an dem Beobachtungspunkt, der auf das Kamerarauschen und den Schätzfehler der dreidimensionalen Form des starren Modellobjektes zurückzuführen ist (vgl. Abschnitt 4.3), wird Gl. (4.94) um Δs erweitert

$$
dfd(\boldsymbol{h}', \boldsymbol{T}^{(c(\xi)^-)}) = \boldsymbol{o}^T \cdot \boldsymbol{R}^{(c(\xi)^-)} + \Delta s. \tag{4.95}
$$

Durch Aufstellen der Gl. (4.95) für $\mathcal{J}$ Beobachtungspunkte, $\mathcal{J}$>3, entsteht folgendes überbestimmtes Gleichungssystem

$$
\boldsymbol{DFD} = [\boldsymbol{O}] \cdot \boldsymbol{R}^{(c(\xi)^-)} + \boldsymbol{V} \tag{4.96}
$$

mit $\boldsymbol{DFD} = (dfd(\boldsymbol{h}_k^{(\mathcal{J}-1)'}, \boldsymbol{T}^{(c(\xi)^-)}), dfd(\boldsymbol{h}_k^{(\mathcal{J}-2)'}, \boldsymbol{T}^{(c(\xi)^-)}), ..., dfd(\boldsymbol{h}_k^{(0)'}, \boldsymbol{T}^{(c(\xi)^-)}))^T$, der Matrix $[\boldsymbol{O}] = (\boldsymbol{o}^{(\mathcal{J}-1)T}, \boldsymbol{o}^{(\mathcal{J}-2)T}, ..., \boldsymbol{o}^{(0)T})^T$, $\boldsymbol{R}^{(c(\xi)^-)} = (R_X, R_Y, R_Z)^T$ und $\boldsymbol{V} = (\Delta s^{(\mathcal{J}-1)}, \Delta s^{(\mathcal{J}-2)}, ..., \Delta s^{(0)})^T$.

Aus Gl. (4.96) und der gleichen Annahmen wie in Abschnitt 4.3.1 ergibt sich die bedingte Wahrscheinlichkeit $p_{\boldsymbol{DFD}/\hat{\boldsymbol{R}}}(\boldsymbol{DFD}/\hat{\boldsymbol{R}})$ entsprechend Gl. (4.64) zu

$$
p_{\boldsymbol{DFD}/\hat{\boldsymbol{R}}}(\boldsymbol{DFD}/\hat{\boldsymbol{R}}) = \frac{1}{\sqrt{(2\pi)^{\mathcal{J}}|[\boldsymbol{U}]|}} \exp\Big(-\frac{1}{2} \cdot \Big(\boldsymbol{DFD} - [\boldsymbol{O}] \cdot \hat{\boldsymbol{R}}\Big)^T [\boldsymbol{U}]^{-1} \Big(\boldsymbol{DFD} - [\boldsymbol{O}] \cdot \hat{\boldsymbol{R}}\Big)\Big). \tag{4.97}
$$

Die Gleichungen des rekursiven Maximum–Likelihood–Schätzers nach $j + 1$ Beobachtungspunkten ergeben sich durch Einsetzen der Gl. (4.97) in

$$
\frac{\partial \ \ln \ p_{\boldsymbol{DFD}/\hat{\boldsymbol{R}}}(\boldsymbol{DFD}/\hat{\boldsymbol{R}})}{\partial \hat{\boldsymbol{R}}} = 0 \tag{4.98}
$$

und Umformung gemäß Gl. (4.70), (4.71), (4.72) und (4.73) zu

$$\hat{\boldsymbol{R}}_{j+1} = \hat{\boldsymbol{R}}_j + \boldsymbol{K}^{(j+1)} \cdot \left(dfd(\boldsymbol{h}^{(j+1)'}, \boldsymbol{T}^{(c(\xi)^-)}) - \boldsymbol{o}^{(j+1)T} \cdot \hat{\boldsymbol{R}}_j \right) \tag{4.99}$$

$$\boldsymbol{K}^{(j+1)} = [\boldsymbol{Q}^{(j)}] \cdot \boldsymbol{o}^{(j+1)} \cdot \left(\boldsymbol{o}^{(j+1)T} \cdot [\boldsymbol{Q}^{(j)}] \cdot \boldsymbol{o}^{(j+1)} + \sigma^2_{\Delta s^{(j+1)}} \right)^{-1} \tag{4.100}$$

$$[\boldsymbol{Q}^{(j+1)}] = [\boldsymbol{Q}^{(j)}] - \boldsymbol{K}^{(j+1)} \cdot \boldsymbol{o}^{(j+1)T} \cdot [\boldsymbol{Q}^{(j)}], \tag{4.101}$$

wobei die Initialisierung wie in Abschnitt 4.3.2 mit mehreren Beobachtungspunkten begonnen wird.

Aufgrund der Approximation gemäß Gl. (4.91) wird die Schätzung der Rotationsparameter iterativ durchgeführt. Die Position eines jeden Beobachtungspunktes in der Bildebene bei der *i*–ten Iteration wird nun als $\boldsymbol{h}_i^{(j)'}$ gekennzeichnet. Das Iterationsverfahren wird beendet, wenn die Ungleichung

$$\left| \frac{1}{\mathcal{G}} \sum_{a=0}^{\mathcal{G}-1} \left(dfd\left(\boldsymbol{h}_i^{(a)'}, \{\boldsymbol{T}^{(c(\xi)^-)}, \hat{\boldsymbol{R}}_{\mathcal{G}-1,i-1}\} \right) \right)^2 \right| - \left| \frac{1}{\mathcal{G}} \sum_{a=0}^{\mathcal{G}-1} \left(dfd\left(\boldsymbol{h}_{i+1}^{(a)'}, \{\boldsymbol{T}^{(c(\xi)^-)}, \hat{\boldsymbol{R}}_{\mathcal{G}-1,i}\} \right) \right)^2 \right| < T_h, \tag{4.102}$$

erfüllt ist, wobei $\{\boldsymbol{T}^{(c(\xi)^-)}, \hat{\boldsymbol{R}}_{\mathcal{G}-1,i-1}\}$ und $\{\boldsymbol{T}^{(c(\xi)^-)}, \hat{\boldsymbol{R}}_{\mathcal{G}-1,i}\}$ die initialen Bewegungsparameter in der *i*–ten bzw. *(i+1)*–ten Iteration darstellen. Für die Berechnung der Gesamtrotation $\boldsymbol{R}^{(c(\xi)^-)}$ einer Objektkomponente wird die Bewegungsgleichung

$$\boldsymbol{H}'_{i+1} = [\hat{\boldsymbol{R}}^{(u)}_{\mathcal{G}-1,i}] \cdot (\boldsymbol{H}'_i - \boldsymbol{J}^{(\xi)'}) + \boldsymbol{J}^{(\xi)'} \tag{4.103}$$

verwendet. Dabei stellt $[\hat{\boldsymbol{R}}^{(u)}_{\mathcal{G}-1,i}]$ den bei der *i*–ten Iteration geschätzten Update der Rotationsmatrix dar. Die Gesamtrotation einer Objektkomponente ergibt sich nach *i* Iterationen durch wiederholte Anwendung der Gleichung (4.103) und Umformung zu

$$\boldsymbol{H}'_{i+1} = [\hat{\boldsymbol{R}}_{\mathcal{G}-1,i}] \cdot (\boldsymbol{H}'_i - \boldsymbol{J}^{(\xi)'}) + \boldsymbol{J}^{(\xi)'} \tag{4.104}$$

mit

$$[\hat{\boldsymbol{R}}_{\mathcal{G}-1,i}] = [\hat{\boldsymbol{R}}^{(u)}_{\mathcal{G}-1,i}] \cdot [\hat{\boldsymbol{R}}^{(u)}_{\mathcal{G}-1,i-1}] \cdot \ldots \cdot [\hat{\boldsymbol{R}}^{(u)}_{\mathcal{G}-1,1}]. \tag{4.105}$$

Für die Detektion von Ausreißern werden ähnliche Regeln wie in Abschnitt 4.3.2 angewendet. Dazu werden zwei Kriterien nacheinander überprüft. Bei dem ersten Kriterium wird ein Beobachtungspunkt $\boldsymbol{W}_i^{(j+1)'}$ als Ausreißer detektiert und eliminiert, wenn die Bedingung

$$|dfd(\boldsymbol{h}_i^{(j+1)'}, \{\boldsymbol{T}^{(c(\xi)^-)}, \hat{\boldsymbol{R}}_{\mathcal{G}-1,i-1}\})| > \sigma_{dfd} \tag{4.106}$$

mit

$$\sigma_{dfd} = \sqrt{\frac{1}{\vartheta}\sum_{a=0}^{\vartheta-1}\left(dfd(\boldsymbol{h}_i^{(a)'}, \{\boldsymbol{T}^{(c(\xi)^-)}, \hat{\boldsymbol{R}}_{\vartheta-1,i-1}\})\right)^2} \tag{4.107}$$

erfüllt wird. Anderenfalls wird das zweite Kriterium ausgewertet. Beim zweiten Kriterium wird ein Beobachtungspunkt $\boldsymbol{W}_i^{(j+1)'}$ genau dann als Ausreißer detektiert, wenn

$$\frac{1}{\vartheta}\sum_{a=0}^{\vartheta-1}\left(dfd\left(\boldsymbol{h}_i^{(a)'}, \{\boldsymbol{T}^{(c(\xi)^-)}, \hat{\boldsymbol{R}}_{j,i}\}\right)\right)^2 \leq \frac{1}{\vartheta}\sum_{a=0}^{\vartheta-1}\left(dfd\left(\boldsymbol{h}_i^{(a)'}, \{\boldsymbol{T}^{(c(\xi)^-)}, \hat{\boldsymbol{R}}_{j+1,i}\}\right)\right)^2 \tag{4.108}$$

erfüllt ist.

Falls der Beobachtungspunkt $\boldsymbol{W}_i^{(j+1)'}$ als Ausreißer detektiert wurde, wird der rekursive Maximum–Likelihood–Schätzer auf dem Stand des Beobachtungspunktes $\boldsymbol{W}_i^{(j)'}$ zurückgesetzt. Anschließend wird mit dem nächsten Beobachtungspunkt $\boldsymbol{W}_i^{(j+2)'}$ fortgefahren.

Experimentelle Untersuchungen

Zur Beurteilung der Genauigkeit des im Rahmen dieser Arbeit entwickelten Verfahrens zur Schätzung der dreidimensionalen Bewegung gegliederter Objekte unter Berücksichtigung von Gelenken wird als Gütekriterium die mittlere Varianz der Schätzfehler der Translations– und die mittlere Varianz der Schätzfehler der Rotationsparameter je Objektkomponente gemessen. Je kleiner die Varianz der Translations– und die Varianz der Rotationsparameter sind, desto genauer ist das Schätzverfahren. Dabei wird als Referenz auf das Verfahren zur Schätzung der dreidimensionalen Bewegung gegliederter Objekte ohne Berücksichtigung von Gelenken Bezug genommen. Die Schätzverfahren werden auf synthetisch erzeugte Bildpaare angewendet. Ein Bildpaar zeigt ein gegliedertes Objekt bestehend aus zwei durch ein Kugelgelenk gekoppelten starren Objektkomponenten, das sich vom Zeitpunkt *0* bis zum Zeitpunkt *1* bewegt. Für alle Bildpaare sind die Form– und Texturparameter sowie die Position des gegliederten Objektes zu beiden Zeitpunkten bekannt. Die Form– und Texturparameter sowie die Position des gegliederten Objektes zum Zeitpunkt *0* sind für alle Bildpaare gleich.

Zur Erzeugung eines Bildpaares wird entsprechend den experimentellen Untersuchungen in Abschnitt 4.2.4 zunächst ein starres Modellobjekt generiert und in die Objektkomponenten Schulter $\mathcal{C}^{(0)}$ und Kopf $\mathcal{C}^{(1)}$ gegliedert. Dabei wird der sichtbare Oberflächenteil der Objektkomponente Schulter $\mathcal{C}^{(0)}$ durch 170 Stützpunkte und 319 Dreiecke dargestellt. Entsprechend wird der sichtbare Oberflächenteil der Objektkomponente Kopf $\mathcal{C}^{(1)}$ durch 110 Stützpunkte und 172 Dreiecke dargestellt. Die mittlere Größe der zu einem Dreieck korrespondierenden Bildfläche beträgt etwa $65\,\mathrm{pel}^2$. Die Position des Kugelgelenkes $\boldsymbol{J}^{(1)}$ wird im Schwerpunkt der gemeinsamen Dreiecke der Objektkomponenten definiert und im Koordinatensystem der Objektkomponente Schulter dargestellt (Bild 3.9). Der Ursprung des Koordinatensystem der Objektkomponente Kopf wird am Kugelgelenk definiert. Das erste Bild des Bildpaares entsteht durch die Synthese eines Modellbildes gemäß Kapitel 5. Vor der Synthese des zweiten Bildes eines Bildpaares wird zunächst die Objektkomponente Schulter bewegt und anschlie-

ßend die Objektkomponente Kopf um das Kugelgelenk rotiert. Dabei nimmt jede Komponente der Translationsparameter Werte im Wertebereich [–16 pel, +16 pel] an. Entsprechend nimmt jede Komponente der Rotationsparameter Werte im Wertebereich [–10 grad, +10 grad] an. Auf das Testbildpaar wird abschließend weißes Gaußsches Rauschen der Varianz σ_n^2 addiert.

Für ein Bildpaar werden die Bewegungsparameter beider gekoppelter Objektkomponenten, die jeweils ihre Bewegung vom Zeitpunkt *0* bis zum Zeitpunkt *1* beschreiben, nun bestimmt. Dazu wird auf das gegliederte Modellobjekt, welches zum ersten Bild korrespondiert, zurückgegriffen und für jede Objektkomponente die Parameter der dreidimensionalen Bewegung gemäß Abschnitt 4.3.2 durch Auswertung des zweiten Bildes geschätzt. Dabei ergibt sich der absolute Schätzfehler je Komponente der Translations– und Rotationsparameter einer Objektkomponente $\mathbb{C}^{(\delta)}$, $\delta : 0, 1$, zu

$$\begin{aligned} e_{T_X}^{(\delta)} &= |T_X^{(\delta)} - \hat{T}_X^{(\delta)}| \text{ pel} \\ e_{T_Y}^{(\delta)} &= |T_Y^{(\delta)} - \hat{T}_Y^{(\delta)}| \text{ pel} \\ e_{T_Z}^{(\delta)} &= |T_Z^{(\delta)} - \hat{T}_Z^{(\delta)}| \text{ pel} \end{aligned} \tag{4.109}$$

$$\begin{aligned} e_{T_X}^{(\delta)} &= |R_X^{(\delta)} - \hat{R}_X^{(\delta)}| \text{ grad} \\ e_{T_Y}^{(\delta)} &= |R_Y^{(\delta)} - \hat{R}_Y^{(\delta)}| \text{ grad} \\ e_{T_Z}^{(\delta)} &= |R_Z^{(\delta)} - \hat{R}_Z^{(\delta)}| \text{ grad,} \end{aligned} \tag{4.110}$$

wobei $\hat{\boldsymbol{B}}^{(\delta)} = (\hat{T}_X^{(\delta)}, \hat{T}_Y^{(\delta)}, \hat{T}_Z^{(\delta)}, \hat{R}_X^{(\delta)}, \hat{R}_Y^{(\delta)}, \hat{R}_Z^{(\delta)})^T$ die geschätzten Bewegungsparameter der Objektkomponente $\mathbb{C}^{(\delta)}$ darstellen.

Bei einer bestimmten Leistung des Bildrauschens σ_n^2 wird eine Vielzahl von synthetisch erzeugten Bildpaaren generiert, aus denen die Bewegungsparameter der gekoppelten Objektkomponenten auf die gleiche Weise bestimmt werden. Dabei ergeben sich die mittlere Varianz der Schätzfehler der Translations– und die mittlere Varianz der Schätzfehler der Rotationsparameter je Objektkomponente zu

$$\bar{\sigma}_T^2 = \frac{1}{6} \cdot \sum_{\delta=0}^{1} \left(\sigma_{T_X}^{(\delta)2} + \sigma_{T_Y}^{(\delta)2} + \sigma_{T_X}^{(\delta)2} \right) \text{ pel}^2 \tag{4.111}$$

$$\bar{\sigma}_R^2 = \frac{1}{6} \cdot \sum_{\delta=0}^{1} \left(\sigma_{R_X}^{(\delta)2} + \sigma_{R_Y}^{(\delta)2} + \sigma_{R_X}^{(\delta)2} \right) \text{ grad}^2, \tag{4.112}$$

wobei $\sigma_{T_X}^{(\delta)2}$, $\sigma_{T_Y}^{(\delta)2}$ und $\sigma_{T_Z}^{(\delta)2}$ jeweils die Varianz jeder Komponente der Schätzfehler der Translationsparameter einer Objektkomponente $\mathcal{C}^{(\delta)}$ darstellen. Entsprechend stellen $\sigma_{R_X}^{(\delta)2}$, $\sigma_{R_Y}^{(\delta)2}$ und $\sigma_{R_Z}^{(\delta)2}$ jeweils die Varianz jeder Komponente der Schätzfehler der Rotationsparameter einer Objektkomponente $\mathcal{C}^{(\delta)}$ dar.

Die Bilder 4.27 und 4.28 zeigen die mittlere Varianz der Translations– bzw. Rotationsparameter gemäß Gl. (4.111) bzw. Gl. (4.112) über der Leistung des Bildrauschens. Dabei steigt die Varianz der Translations– und die Varianz der Rotationsparamter bei zunehmender Leistung des Bildrauschens. Für eine Leistung des Bildrauschens von $\sigma_n^2 = 6$ pel^2, d.h. 40 dB gemessen in PSNR, ergibt sich eine Verbesserung der Varianz der Translationsparameter von 30% und eine Verbesserung der Varianz der Rotationsparameter von 35%

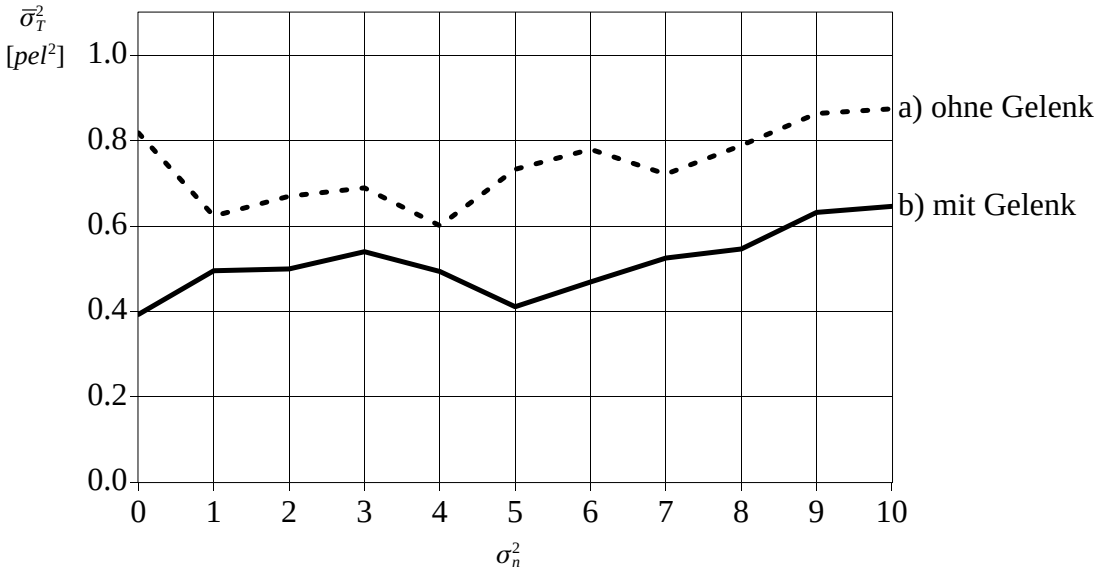


Bild 4.27 Mittlere Varianz $\bar{\sigma}_T^2$ des Schätzfehlers der Translationsparameter gemäß Gl. (4.111) als Funktion der Leistung des Bildrauschens σ_n^2 : a) ohne Berücksichtigung des Kugelgelenkes und b) unter Berücksichtigung des Kugelgelenkes.

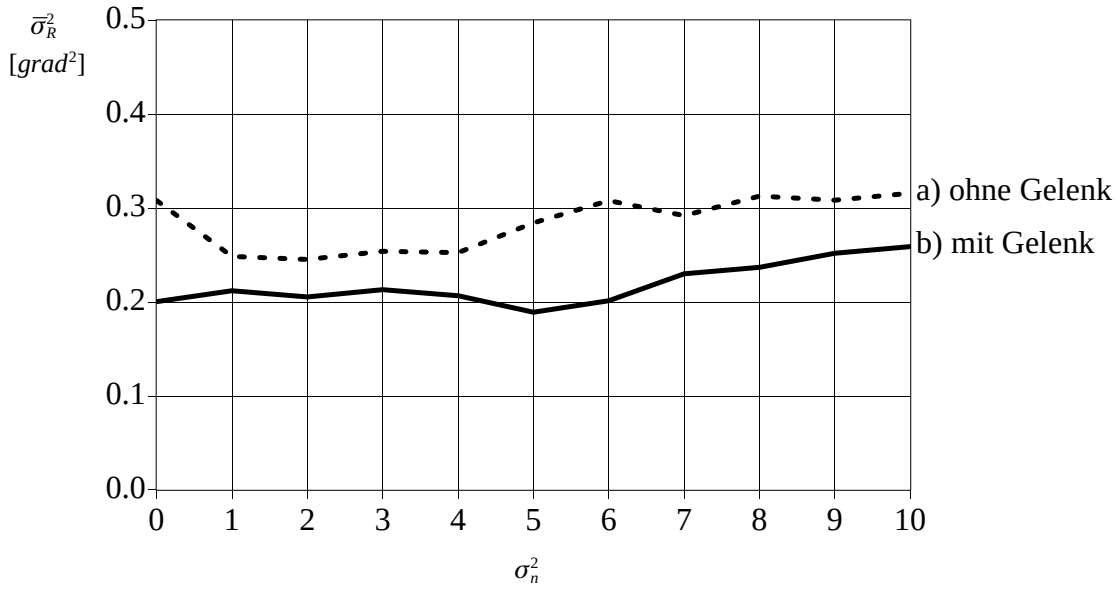


Bild 4.28 Mittlere Varianz $\overline{\sigma}_R^2$ des Schätzfehlers der Rotationsparameter gemäß Gl. (4.112) als Funktion der Leistung des Bildrauschens: a) ohne Berücksichtigung des Kugelgelenkes und b) unter Berücksichtigung des Kugelgelenkes.

5. Bildsynthese für bewegte dreidimensionale, gegliederte Objekte

Die Bildsynthese erzeugt unter Verwendung der von der Bildanalyse für jedes Objekt ermittelten Modellobjektparameter ein Modellbild der Modellwelt. Die Bildsynthese wird in zwei Schritten durchgeführt[65]: Texturemapping und dann Aliasingkompensation. Bei dem Texturemapping wird zunächst das Modellbild mit den Texturparametern des Modellobjektes HINTERGRUND initialisiert. Anschließend werden die Texturparameter einer jeden Modellobjektkomponente eines gegliederten Modellobjektes in die Bildebene projiziert. Bei der Aliasingkompensation wird das durch das Texturemapping entstehende Aliasing unterdrückt.

Für das Texturemapping muß jeder sichtbare Oberflächenpunkt einer Modellobjektkomponente durch Zentralprojektion gemäß Gl. (3.1) in die Modellbildebene abgebildet werden. Da die Berechnung der Zentralprojektion sehr rechenaufwendig ist, wird hier eine affine Transformation verwendet, mit deren Hilfe die Textur eines Dreiecks des Modellobjektes auf ein Dreieck in der Bildebene abgebildet wird[34][45][88]. Die Texturparameter eines Dreiecks des Modellobjektes liegen gemäß Abschnitt 4.2 als zweidimensionale Texturbilder vor. Die Positionen der Kontrollpunkte eines Dreieckes sind für jedes Texturbild bekannt. Mittels des Kameramodells gemäß Gl. (3.1) werden die aktuellen Koordinaten der Kontrollpunkte in der Modellbildebene bestimmt. Die Parameter der affinen Transformation werden nun so berechnet, daß die Position jedes Kontrollpunktes in dem Texturbild exakt auf die korrespondierende Position in der Modellbildebene abgebildet wird. Bei dem Texturemapping wird dann für jeden Bildpunkt des Dreiecks im Modellbild mittels der Transformation die entsprechende Position im Texturbild bestimmt. Im allgemeinen Fall stimmen die Bildpunktraster von Modellbild und Texturbild nicht überein. Der gesuchte Texturwert wird daher mit Hilfe der bilinearen Interpolation[68] aus den Texturwerten der vier benachbarten Bildpunkte im Texturbild berechnet.

Bevor der berechnete Texturwert eines Bildpunktes im Modellbild eingetragen wird, muß festgestellt werden, ob der darzustellende Oberflächenpunkt sichtbar ist. Ein Oberflächenpunkt einer Objektkomponente ist nicht sichbar, wenn er entweder durch eine andere Objektkomponente verdeckt wird oder nach einer Drehung der Objektkomponente im dreidimensionalen Raum nicht mehr sichtbar ist. Um die Sichtbarkeit der Oberflächenpunkte überprüfen zu können, wird ein ”Z–Buffer” für jeden Bildpunkt des Modellbildes definiert[69]. Der Z–Buffer wird mit ∞ initialisiert. Im Z–Buffer wird die Z–Koordinate des Ortsvektors des darzustellenden Oberflächenpunktes eingetragen. Soll die Sichtbarkeit eines Oberflächenpunktes im Texturbild getestet werden, so wird seine Z–Koordinate mit der im Z–Buffer des Bildpunktes eingetragenen Z–Koordinate verglichen. Ist die Z–Koordinate des darzustellenden Oberflächenpunktes kleiner als die eingetragene, so ist der darzustellende Oberflächenpunkt sichtbar und der berechnete Texturwert wird im Bildpunkt des Modellbildes und die

Z–Koordinate in den zugehörigen Z–Buffer eingetragen. Im anderen Fall ist der darzustellenden Oberflächenpunkt verdeckt und es erfolgt keine Änderung des Modellbildes und des Z–Buffers.

Bei dem Texturemapping gibt es zwei Ursachen für Aliasing. Aliasing aufgrund einer Bandbreitenskalierung der Texturparameter bei Vergrößerung des Winkels zwischen der Kameraachse und der Oberflächennormalen eines Dreiecks nach einer Drehung einer Modellobjektkomponente im dreidimensionalen Raum[66][65] und Aliasing am Übergang einer Modellobjektkomponente zu statischem Hintergrund oder zu einer verdeckten Modellobjektkomponente aufgrund der diskreten Abtastwerte des Modellbilds und einer nicht modellierten Apertur der realen Kamera[65]. Zur Reduktion des Aliasing wird ein Zwischenbild mit erhöhter Auflösung berechnet und anschließend bandbegrenzt und unterabgetastet[79]. Durch Einbeziehen der Silhouette der Modellobjektkomponenten in die Aliasingkompensation können die Störungen durch Aliasing am Rand einer Modellobjektkomponente noch weiter unterdrückt werden[83][67].

6. Parametercodierung für bewegte dreidimensionale, gegliederte Objekte

In diesem Kapitel wird ein Verfahren zur effizienten Codierung der von der Bildanalyse ermittelten Parametersätze beschrieben. Die Bewegungsparameter der Modellobjekte (MC–Objekte) werden gleichförmig quantisiert und in PCM–Darstellung übertragen. Zur Codierung der Formparameter der Modellobjekte sowie der Formparameter der MF–Objekte werden Konturcodiertechniken verwendet. Die Texturparameter der MF–Objekte sowie der freigewordenen Bildbereiche werden unter Verwendung hybrider Verfahren codiert, die bewegungskompensierende Prädiktion und Transformationsverfahren vereinen und bei der Codierung Kenntnis der Form der zu codierenden Gebiete mitberücksichtigen. Anschließend wird eine Steuerung zur Übertragung der codierten Parametersätze behandelt.

Verfahren zur Codierung der Bewegungsparameter gegliederter Objekte

Bei der Codierung der Parameter der dreidimensionalen Bewegung gegliederter Objekte müssen nur die unabhängigen Bewegungsparameter einer jeden Objektkomponente eines gegliederten Objektes quantisiert, codiert und übertragen werden. Für ein gegliedertes Objekt mit Baumstruktur, dessen Objektkomponenten durch Kugelgelenke gekoppelt sind, müssen für die Wurzelobjektkomponente sechs Bewegungsparameter und für die restlichen Objektkomponenten jeweils nur die Rotationsparameter quantisiert, codiert und übertragen werden.

Jeder der sechs Bewegungsparameter der Wurzelobjektkomponente wird nach [65] gleichförmig quantisiert. Tabelle 6.1 zeigt die Quantisiererstufenzahl für jeden der Parameter der dreidimensionalen Bewegung der Wurzelobjektkomponente. Zur Quantisierung der Rotationsparameter für die restlichen Objektkomponenten wird ebenfalls ein 9 bit gleichförmiger Quantisierer gemäß Tabelle 6.1 verwendet. Die quantisierten Bewegungsparameter werden in PCM–Darstellung übertragen.

Tabelle 6.1: Stufenzahl in bit für die Quantisierung der Bewegungsparameter der Wurzelobjektkomponente.

Bewegungsparameter	T_X	T_Y	T_Z	R_X	R_Y	R_Z
Quantisiererstufenzahl (bit)	8	8	6	9	9	9

Die Übertragung nur der drei unabhängigen Bewegungsparameter statt aller sechs Bewegungsparameter einer jeden Objektkomponente eines gegliederten Objektes ermöglicht eine Reduktion der erforderlichen Datenrate für die Übertragung der Parameter der dreidimensionalen Bewegung eines gegliederten Objektes um

$$(\mathcal{N} - 1) \cdot (b_{T_X} + b_{T_Y} + b_{T_Z}) \cdot bit, \tag{6.1}$$

wobei $\mathcal{N}$ die Anzahl der Objektkomponenten eines gegliederten Objektes und b_{T_X}, b_{T_Y} und b_{T_Z} die Quantisiererstufenzahl für die Translationsparameter T_X, T_Y bzw. T_Z gemäß Tabelle 6.1 darstellt.

Verfahren zur Codierung der Formparameter gegliederter Objekte

Für die Übertragung der Form eines gegliederten Objektes wird einmalig seine Silhouette, die Silhouette der verdeckenden Objektkomponenten, für jeden Stützpunkt die Kennzahl für die Zugehörigkeit zu einer Objektkomponente und die drei Komponenten des Ortsvektors der Position eines jeden Gelenkes codiert. Für die fortlaufende Formadaption eines gegliederten Objektes gemäß Abschnitt 4.2.3 ist die Übertragung seiner realen Silhouette sowie der realen Silhouetten der verdeckenden Objektkomponenten auch für jedes folgende Bild der Sequenz erforderlich. Ferner muß die zweidimensionale Form der MF–Objekte übertragen werden. Die Codierung einer Silhouette erfolgt wie beim Quellenmodell R3D durch eine prädiktive Codierung der Positionen der Stützpunkte am Rande der Silhouette. Die Kennzahlen der Stützpunkte werden mittels einer Lauflängencodierung codiert. Die drei Komponenten des Ortsvektors der Position eines Gelenkes werden gleichförmig quantisiert und in PCM–Darstellung übertragen. Die zweidimensionale Form der MF–Objekte in der Bildebene wird mittels einer Polygon–Spline–Approximation codiert.

Da sowohl Empfänger als auch Sender aus einer Silhouette ein Dreiecksnetz erzeugen, welches einen Polygonzug der berandenden Kontrollpunkte, sog. Randkontrollpunkte, als Berandung hat, wird hier die Codierung einer Silhouette nach [65] durch die Positionen der Randkontrollpunkte durchgeführt. Dazu wird nach der Codierung der Position eines ersten Randkontrollpunktes in der Bildebene die Position des jeweils nächsten Randkontrollpunktes relativ zu seinem Vorgänger entropiecodiert. Um die Position der Randkontrollpunkte möglichst effizient codieren zu können, werden statistische Abhängigkeiten zwischen den Positionen der Randkontrollpunkte aufeinanderfolgender Bilder für die Codierung und Übertragung der Position der Randkontrollpunkte des aktuellen Bildes ausgenutzt. Dazu wird anstatt der Position der Randkontrollpunkte folgender Bilder der für die Formadaption gemäß Abschnitt 4.2.3 ermittelte Verschiebungsvektor für jeden Randkontrollpunkt codiert und übertragen. Da die Richtung dieses Verschiebevektors für jeden Randkontrollpunkt gemäß Abschnitt 4.2.3 durch den Mittelwert der Normalvektoren der Dreiecke, zu denen der Randkontrollpunkt gehört, gegeben ist, muß für die Verschiebung nur der Betrag des Verschiebungsvektors codiert werden. Da die Silhouette auf dem diskreten Abtastraster der Bildebene definiert ist, kann der Betrag der Verschiebung nur diskrete Werte annehmen. Daher wird diese Verschiebung verlustlos codiert.

Zur Übertragung der Information zur Gliederung eines starren Modellobjektes in Modellobjektkomponenten, wird zunächst jeder Modellobjektkomponente eine Kennzahl zugeordnet. Jedem Kontrollpunkt wird dann die Kennzahl der zugehörigen Modellobjektkomponente zugewiesen. Die Kennzahl eines jeden Kontrollpunktes eines gegliederten Modellobjektes wird dann in definierter Reihenfolge in eine lineare Liste eingetragen, die anschließend mittels

einer Lauflängencodierung codiert und übertragen wird. Die Reihenfolge der Stützpunkte ist dem Empfänger bekannt.

Die Position eines jeden Kugelgelenkes bezogen auf das Koordinatensystem der zugehörigen Bezugsobjektkomponente muß einmalig quantisiert, codiert und übertragen werden. Zur Quantisierung jeder der drei Komponenten der dreidimensionalen Position eines Kugelgelenkes wird ein gleichförmiger Quantisierer verwendet. Zur Bestimmung der Quantisiererstufenzahl bei dem hier verwendeten Bildformat CIF werden zunächst Annahmen über den Szeneninhalt getroffen. Es wird angenommen, daß sich in der Szene ein gegliedertes Objekt bestehend aus zwei durch ein Kugelgelenk gekoppelten Objektkomponenten befindet. Das Kugelgelenk hat die Koordinaten (0,0,2F) im Weltkoordiantensystem. Die maximale Ausdehnung einer Objektkomponente wird gleich der Bildhöhe H_B eines Bildes in pel angenommen (siehe Bild 3.2). Dabei wird die maximale Ausdehnung einer Objektkomponente als der maximale Abstand zwischen dem Ursprung des Koordinatensystems der Objektkomponente und einem beliebigen Punkt dieser Objektkomponente definiert. Die drei Komponenten der Position eines Kugelgelenkes im Koordinatensystem der Bezugsobjektkomponente werden nun so quantisiert, daß sich eine maximale Abweichung der Position des Kugelgelenkes entsprechend Gl. (4.36) von 0.1 pel ergibt. Dabei ergibt sich für die Quantisierung der drei Komponenten der Position eines Kugelgelenkes ein Quantisierer mit 13 bit (Tabelle 6.2).

Tabelle 6.2: Stufenzahl in bit für die Quantisierung der drei Komponenten der Position eines Kugelgelenkes bezogen auf das Koordinatensystem der zugehörigen Bezugsobjektkomponente.

Komponenten der Position eines Gelenkes	$\boldsymbol{J}X$	$\boldsymbol{J}Y$	$\boldsymbol{J}Z$
Quantisiererstufenzahl (bit)	13	13	13

Die Formparameter $\boldsymbol{F}_{k+1}^{MF}$ der MF–Objekte sind durch die Berandung ihrer zweidimensionalen Form in der Bildebene gegeben. Zur Codierung der Formparameter $\boldsymbol{F}_{k+1}^{MF}$ der MF–Objekte wird hier das in [32] vorgeschlagene Verfahren übernommen. Bei diesem Verfahren wird zunächst die Berandung eines MF–Objektes durch ein Polygon approximiert. Dazu werden zunächst vier Berandungspunkte als Initialpolygonpunkte der Approximation ermittelt[32]. Für die weitere Approximation wird zur Bewertung des Approximationsfehlers der maximale euklidische Abstand $d_{\max}$ zwischen der geschätzten Berandung und der approximierten Berandung verwendet. Für jeden Polygonabschnitt wird nun überprüft, ob der Abstand zwischen dem bereits ermittelten Polygonzug und der tatsächlichen Berandung einen zulässigen Approximationsfehler $d_{\max}^{*MF}$ überschreitet oder nicht. Wenn der Abstand $d_{\max}$ für einen Polygonabschnitt größer als der zulässige Approximationsfehler $d_{\max}^{*MF}$ ist, wird ein weiterer Polygonpunkt eingefügt und damit die Approximation der Berandung verfeinert. Die Verfeinerung der Approximation durch das Einfügen weiterer Polygonpunkte wird solange weitergeführt, bis für jeden Polygonabschnitt der zulässige Approximationsfehler $d_{\max}^{*MF}$ unterschritten wird.

Anschließend wird für diese Polygonpunkte eine Splineapproximation[28] berechnet. Die so ermittelten Polygon– und Splineapproximationen werden abschließend verglichen. Für alle Kurvenabschnitte, in den die Splineapproximation das vorgegebene Fehlermaß $d_{\max}^{*MF}$ einhält, wird die Splineapproximation verwendet, in den übrigen Kurvenabschnitten die Polygonapproximation. Die zu codierenden Parameter sind die Polygonpunktkoordinaten und der Kurventyp zwischen zwei Polygonpunkten. Nach der Codierung der Position eines ersten Polygonpunktes in der Bildebene wird die Position des jeweils nächsten Polygonpunktes relativ zu seinem Vorgänger entropiecodiert. Für die Codierung des Kurventyps aller Kurvenabschnitte wird eine Lauflängencodierung verwendet. Eine Codierung der Formparameter mittels zeitlicher Prädiktion entfällt, da die MF–Objekte nur für einen Zeitpunkt *k* definiert sind.

Verfahren zur Codierung der Texturparameter gegliederter Objekte

Bei einer objektbasierten Analyse–Synthese Codierung basierend auf dem Quellenmodell bewegter dreidimensionaler, gegliederter Objekte müssen die Texturparameter für die MF–Objekte, die freigewordenen Objektbereiche und das erste zu decodierende Bild nach einem Szenenschnitt codiert und übertragen werden. Zur Codierung dieser Texturparameter wird das in [65] vorgeschlagene Verfahren übernommen. Eine Ausnutzung der Eigenschaften der menschlichen visuellen Wahrnehmung wird in dieser Arbeit bei der Codierung der Texturparameter nicht implizit vorgenommen.

Die Texturparameter der MF–Objekte und der freigewordenen Bildbereiche werden prädiktiv codiert. Für die Codierung der Texturparameter der MF–Objekte erfolgt eine zeitliche Prädiktion durch das bewegungs– und formkompensierte Modellbild. Für die Codierung der Texturparameter der freigewordenen Bildbereiche erfolgt adaptiv eine örtliche Prädiktion aus benachbarten Bildbereichen oder eine zeitliche Prädiktion aus einem Hintergrundspeicher. Anschließend wird der Prädiktionsfehler codiert und übertragen.

Die Gebiete, für den Prädiktionsfehler innerhalb einer objektbasierten Analyse–Synthese–Codierung basierend auf dem Quellenmodell bewegter dreidimensionaler, gegliederter Objekte übertragen werden müssen, sind i.allg. beliebig berandet. Zur Codierung des Prädiktionsfehlers beliebig berandeter Gebiete wird hier eine objektangepaßte Transformationscodierung basierend auf der DCT [21][32][72] eingesetzt. Für die objektangepaßte Transformationscodierung basierend auf der DCT erfolgt zuerst unter Berücksichtigung der Form der zu codierenden Gebiete eine Berechnung der Basisvektoren sowohl beim Sender als auch beim Empfänger. Unter Verwendung dieser Basisvektoren wird anschließend der Prädiktionsfehler beim Sender transformiert und die Koeffizienten dieser Transformation berechnet. Diese Koeffizienten werden dann quantisiert, codiert und an den Empfänger übertragen.

Im Rahmen dieser Arbeit wird angenommen, daß dem Empfänger die Texturparameter des kompletten ersten Bildes nach einem Szenenschnitt bekannt sind. Diese Texturparameter

können jedoch gemäß eines bereits bestehenden standardisierten Verfahrens wie zum Beispiel H.263[42], MPEG–1[40] oder MPEG–2[39] übertragen werden.

Verfahren zur Steuerung der Parametercodierung

Die Steuerung der Parametercodierung entscheidet, welche Parameter eines jeden Objektes zum Empfänger übertragen werden müssen. Diese Entscheidung hängt davon ab, ob ein Objekt als modellierbares (MC)–Objekt oder nicht modellierbares (MF)–Objekt gekennzeichnet wurde.

Die im Rahmen dieser Arbeit verwendete Steuerung der Parametercodierung wurde in [65] vorgeschlagen. Sie unterscheidet zwischen drei verschiedenen Codermodi für die MF– und MC–Objekte und die freigewordenen Bildbereiche. Für die MF–Objekte werden nur die codierten Form– und Texturparameter, nicht aber die Bewegungsparameter übertragen (Mode 1). Dagegen werden für die MC–Objekte nur die codierten Bewegungs– und Formparameter, nicht aber die Texturparameter übertragen (Mode 2). Zur Detektion von Modellverletzungen wird hier das in [65] vorgeschlagene Verfahren übernommen. Gemäß der in diesem Verfahren bei der Detektion der Modellverletzungen getroffenen Annahmen über das Empfängermodell sind die durch die fehlende Übertragung der Texturparameter auftretenden geometrischen Verzerrungen innerhalb der MC–Objekte ohne Bedeutung für die subjektive Bildqualität[32][59][65]. Deshalb werden für diese Bildbereiche keine Texturparameter übertragen. Die Berandung der freigewordenen Bildbereiche ist durch die Kenntnis der Form– und Bewegungsparameter der MC–Objekte eindeutig festgelegt. Für diese Bildbereiche werden die Texturparameter codiert und übertragen (Mode 3). Da ein Szenenschnitt nicht durch das Quellenmodell beschrieben werden kann, wird er als Spezialfall eines MF–Objektes betrachtet (Mode 4). Im Falle eines Szenenschnitts müssen dem Empfänger die Texturparameter eines kompletten Bilds übertragen werden. Im Mode 4 sind die Formparameter bekannt und nur Texturparameter müssen unter Verwendung eines bereits bestehenden, standardisierten Verfahrens wie zum Beispiel H.263[42], MPEG–1[40] oder MPEG–2[39] übertragen werden.

Die Steuerung der Parametercodierung überträgt prinzipiell zuerst die Bewegungsparameter der MC–Objekte, dann die Formparameter aller Objekte und abschließend die Texturparameter der MF–Objekte und der freigewordenen Bildbereiche. Diese Reihenfolge der Parameterübertragung ermöglicht eine prädiktive Codierung der Parametersätze.

7. Experimentelle Untersuchungen

Der im Rahmen dieser Arbeit entwickelte objektbasierte Analyse–Synthese Coder basierend auf dem Quellenmodell bewegter dreidimensionaler, gegliederter Objekte (A3D) wurde auf einer Rechenanlage implementiert und anhand von international ausgewählten Bildsequenzen getestet. Seine Effizienz wurde mit den Codern basierend auf den Quellenmodellen starrer dreidimensionaler Objekte (R3D) und flexibler dreidimensionaler Objekte (F3D) verglichen. Die hierfür ausgewählten Testsequenzen werden in Abschnitt 7.1 vorgestellt. Für die Bewertung der Effizienz wird ein aus der Literatur bekanntes, meßbares Gütekriterium verwendet, welches in Abschnitt 7.2 kurz beschrieben wird. Eine Diskussion und Bewertung der Effizienz der unterschiedlichen dreidimensionalen Quellenmodelle wird in Abschnitt 7.3 vorgenommen.

7.1 Beschreibung der verwendeten Testsequenzen

Um eine Vergleichbarkeit der Codierergebnisse mit anderen Veröffentlichungen zu gewährleisten[32][65], wurden für die Untersuchungen des in dieser Arbeit entwickelten Codierverfahrens die Testsequenzen *Claire*[13] und *Miss America* [9] verwendet. Diese Testsequenzen wurden von internationalen Expertengruppen (”European COST 211bis Simulation Subgroup” und ”CCITT Study Group XV”) als typische Bildfernsprechsequenzen mit Kopf– und Schulterbewegungen zum Testen von Bildfernsprechcodecs ausgewählt. In Bild 7.1 ist

a) b)

Bild 7.1 Für die Codierung verwendete Testsequenzen (Originalbilder):
a) *Claire* Bild 33 und b) *Miss America* Bild 8.

je ein Bild der zwei Testsequenzen gezeigt. Da die sprechende Person als ein gegliedertes Objekt bestehend aus den gekoppelten Objektkomponenten Kopf und Schulter betrachtet werden kann, eignen sich diese Testsequenzen zum Testen des im Rahmen dieser Arbeit entwickelten objektbasierten Coders basierend auf dem Quellenmodell A3D. Alle Testse-

quenzen wurden im "Common Intermediate Format" (CIF) mit einer reduzierten Bildfolgefrequenz von f_{st}=10 Hz verwendet. Dieses Bildformat wurde für Bildtelefonanwendungen mit einer Datenrate von 64 kbit/s bis 2Mbit/s vereinbart [41]. Jedes Bild besteht dabei aus 288 Zeilen und 352 Bildpunkten pro Zeile für die Luminanzkomponente und aus 144 Zeilen und 176 Bildpunkten pro Zeile für die beiden Chrominanzkomponenten. Dabei werden die Amplituden der Abtastwerte für alle Komponenten mit 8 bit quantisiert, die Abtastfrequenz der Luminanzkomponente beträgt f_{sx}=6.75 MHz. Sofern nicht anders angegeben, stellen sämtliche Meßergebnisse in diesem Kapitel jeweils einen Mittelwert über die zwei Testsequenzen dar.

7.2 Gütekriterium zur Bewertung der Bildqualität in einem Analyse–Synthese Coder

Um eine objektive Bewertung der Codierergebnisse unterschiedlicher Verfahren zu ermöglichen, sind Gütekriterien erforderlich, mit denen die Bildqualität der codierten Bildsequenzen als meßbare Zahl beschrieben werden kann. Das hier verwendete aus der Literatur bekannte Gütekriterium ist die auf Vollaussteuerung bezogene Codierfehlerleistung PSNR (englisch: *P*eak *S*ignal to *N*oise *R*atio) gemäß

$$\mathrm{PSNR[dB]} \;=\; 10\log_{10}\left(\frac{255^2}{\sigma_e^2}\right). \tag{7.1}$$

Dabei stellt σ_e^2 die Leistung der Synthesefehler, d.h. den mittleren quadratischen Fehler zwischen den Abtastwerten des Luminanzsignals des realen Bildes l_k und des korrespondierenden codierten Bildes $l_k^{'}$ dar. σ_e^2 wird durch Mittelung der Synthesefehler über alle Bildpunkte aller Bilder einer Bildfolge berechnet.

7.3 Vergleich der Bildqualität der Coder mit dreidimensionalen Quellenmodellen

In diesem Abschnitt wird die Bildqualität codierter Bildsequenzen der objektbasierten Analyse–Synthese Coder basierend auf den Quellenmodellen A3D, R3D sowie F3D miteinander verglichen. Dabei wird als Referenz auf die codierten Bildsequenzen des blockbasierten Hybridcoders TMN5 gemäß dem Standard H.263 Bezug genommen[42]. Zur Beurteilung der Bildqualität wird das Gütekriterium gemäß Abschnitt 7.2 verwendet. Es wird zunächst für die drei objektbasierten Coder eine optimierte Einstellung der Parametercodierung vorgeschlagen. Anschließend wird die Bildqualität der objektbasierten Coder untereinander sowie mit den Ergebnissen der Referenz TMN5 verglichen.

Aufteilung der Datenrate auf die Parametersätze

Zur Optimierung der Parametercodierung eines objektbasierten Analyse–Synthese Coders wird bei einer vorgegebenen Bildqualität gemäß Gl. (7.1) die Gesamtdatenrate durch eine geeignete Aufteilung auf die Parametersätze minimiert. Die Gesamtdatenrate besteht dabei

aus den Teildatenraten für die Übertragung der Bewegungs–, Form– und Texturparameter. Zur Minimierung der Gesamtdatenrate werden der zulässige Fehler bei der Approximation der Objektberandung ($d_{\max}^{MF}$) der MF–Objekte und die Quantisiererstufenzahl der Bewegungsparameter der MC–Objekte als freie Parameter optimiert[32][65]. Die Optimierungsgröße $d_{\max}^{MF}$ gibt dabei an, um wieviele Bildpunkte die Originalobjektberandung und die approximierte Objektberandung maximal voneinander abweichen dürfen. Die Approximation der Berandung der MF–Objekte beeinflußt wesentlich den erforderlichen Datenaufwand für die Codierung der Texturparameter. Je genauer die Berandung dieser Objekte approximiert wird, desto weniger Bildpunkte müssen durch Texturinformation nachgearbeitet werden. Um eine Vergleichbarkeit der Codiereffizienz der unterschiedlichen dreidimensionalen Quellenmodelle zu gewährleisten, wird hier für alle objektbasierten Coder dieselbe Einstellung der Parametercodierung verwendet. Dazu wird nach [65] ein Maß von $d_{\max}^{*MF} = 2.5$ pel verwendet und für die Quantisierung der Translations– und Rotationsparameter die Quantisiererstufenzahl gemäß Tabelle 6.1 benutzt. Ferner wird für die Berandungsapproximation der MC–Objekte ein Maß von $d_{\max}^{*MC} = 1.4$ pel verwendet[65]. Für die Quantisierung der Verschiebungsparameter des Quellenmodells F3D wird ein 32–stufiger Quantisierer benutzt[65]. Für die Quantisierung der drei Koordinaten der Position eines Kugelgelenkes wird die Quantisiererstufenzahl gemäß Tabelle 6.2 verwendet.

Die Testsequenzen wurden mit den vorgeschlagenen Einstellungen der Parametercodierung für die drei objektbasierten Coder codiert. Tabelle 7.1 zeigt die Aufteilung der Datenrate auf die Parametersätze Bewegung und Form für die Quellenmodelle A3D, R3D und F3D und gibt die Fläche der Bildbereiche an, für die Texturparameter übertragen werden müssen. Im Vergleich zu den Quellenmodellen R3D und F3D erfordert das Quellenmodell A3D 50% weniger Datenrate für die Codierung der Bewegungsparameter. Dies läßt sich darauf zurückführen, daß mittels der Quellenmodelle R3D und F3D die Szene mit mehreren Modellobjekten beschrieben wird, für die jeweils Translations– und Rotationsparameter übertragen werden müssen[65]. Das erfordert insgesamt 200 bit/Bild. Das Quellenmodell A3D beschreibt die Szene mit einem einzigen gegliederten Modellobjekt bestehend aus zwei Modellobjektkomponenten (Bild 4.11.a und 4.11.b). Für die Wurzelobjektkomponente Schulter werden Rotations– und Translationsparameter übertragen, für die Objektkomponente Kopf nur die Rotationsparameter. Das erfordert insgesamt 100 bit/Bild. Für die drei Quellenmodelle sind für die Codierung der Objektsilhouette 400 bit/Bild erforderlich. Für die Quellenmodelle R3D und F3D erfordert die Codierung der Formparameter der MC–Objekte insgesamt 500 bit/Bild bzw. 950 bit/Bild. Die Codierung der Formparameter erfordert für das Quellenmodell F3D 450 bit/Bild mehr, da die flexiblen Verschiebungsparameter übertragen werden müssen. Für das Quellenmodell A3D sind zur Übertragung einer Kennzahl je Stützpunkt, mit der die Zugehörigkeit zu einer Modellobjektkomponente festgelegt wird, und der drei Koordinaten der Position des Kugelgelenkes zusätzlich 50 bit/Bild erforderlich. Die Codierung der Formparameter der MC–Objekte erfordert somit insgesamt 450 bit/Bild für das Quellenmodell A3D. Aufgrund der besseren Beschreibung der Form und Bewegung der realen Objekte bei

Verwendung des Quellenmodells A3D ergibt sich eine effizientere Beschreibung eines zu codierenden Bildes aufgrund der aktuellen Bewegungs– und Formparameter sowie zuvor übertragener Texturparameter, als es mit dem Quellenmodell R3D möglich ist. Es entstehen daher kleinere und weniger MF–Objekte als beim Quellenmodell R3D. Dies führt zum einen zu einer geringeren Datenrate für die Codierung der Formparameter der MF–Objekte und zum anderen zu einer geringeren Bildfläche, für die Texturparameter übertragen werden müssen. Ein Vergleich mit dem Quellenmodell F3D zeigt, daß das Quellenmodell A3D effizienter ist, obwohl bei beiden Quellenmodellen die Gesamtfläche, für die Texturparameter übertragen werden müssen, gleich ist. Dies ist darauf zurückzuführen, daß beim Quellenmodell F3D zusätzlich die flexiblen Verschiebungsparameter übertragen werden müssen. Dabei beträgt die Fläche, für die Texturparameter übertragen werden müssen, im Mittel 4% der Bildfläche für das Quellenmodell R3D und im Mittel 3% der Bildfläche für die Quellenmodelle A3D und F3D. Für die Codierung der Texturparameter bei guter Bildqualität muß mindestens eine Datenrate von 1.1 bit pro Bildpunkt aufgewendet werden [65][72].

Tabelle 7.1: Aufteilung der Datenrate auf die Bewegungs– und Formparameter für die Quellenmodelle R3D, F3D und A3D.

Quellen–modell	Bewegung: R_B [bit/Bild]	Form der MC–Objekte: $R_{F,MC}$ [bit/Bild]	Form der MF–Objekte: $R_{F,MF}$ [bit/Bild]	Flächenanteil für neue Texturparameter [% der Bildfläche]
R3D	200	500	1150	4%
F3D	200	950	1000	3%
A3D	100	450	1000	3%

Vergleich der Bildqualität anhand eines objektiven Gütekriteriums

Zum Vergleich der Bildqualität der codierten Bildsequenzen der unterschiedlichen Coder wird hier das Gütekriterium gemäß Abschnitt 7.2 verwendet. Dabei wird die Leistung der Synthesefehler durch Mittelung über alle Bildpunkte aller Bilder einer Bildfolge berechnet. Um Schwankungen in der zur Codierung eines Bildes benötigten Datenrate ausgleichen zu können, wird ein Pufferspeicher, dessen Größe jeweils der im Mittel für die Codierung eines Bildes verfügbaren Datenmenge entspricht, verwendet. Ferner wird für die Übertragung des ersten Bildes eine beliebig große zur Verfügung stehende Datenrate angenommen. Einschwingvorgänge des Coders nach einem Szenenschnitt bleiben damit unberücksichtigt.

Bild 7.2 zeigt das Gütekriterium PSNR der objektbasierten Analyse–Synthese–Coder und des blockbasierten Coders TMN5. Bezüglich dieses Gütekriteriums übertrifft der blockbasierte Hybridcoder die objektbasierten Analyse–Synthese–Coder um etwa 1.5 bis 2 dB. Der große Qualitätsvorsprung wird dadurch hervorgerufen, daß der blockbasierte Hybridcoder unabhängig von der Bedeutung der Synthesefehler für einen menschlichen Beobachter für alle Bereiche, in denen ein Synthesefehler vorliegt, Texturparameter überträgt. Der objektbasierte Coder läßt aber solche Synthesefehler, die aufgrund kleiner Positions– und Formfehler, sog. geometrischer Verzerrungen, entstehen, bewußt zu und überträgt hier keine Texturparameter.

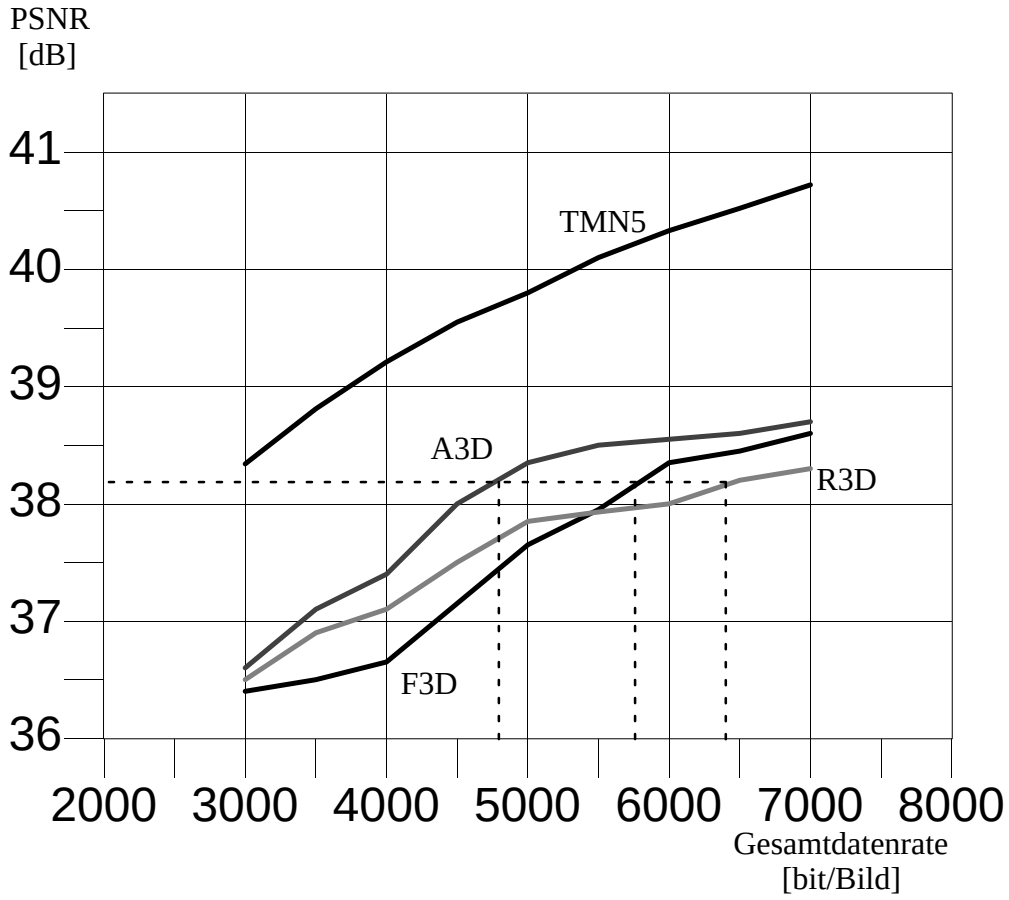


Bild 7.2 PSNR des blockbasierten Hybridcoders TMN5 und der objektbasierten Analyse–Synthese Coder basierend auf den Quellenmodellen A3D, R3D und F3D.

Diese geometrischen Verzerrungen mindern die subjektive Bildqualität nicht, verkleinern aber das PSNR. In [65] wird daher ein geeigneteres Maß für die Bewertung der Bildqualität von codierten Bildsequenzen der objektbasierten Coder vorgeschlagen, das die Synthesefehler abhängig von ihrer Bedeutung für einen menschlichen Beobachter gewichtet. Dazu werden nur die Synthesefehler solcher Bildpunkte gemittelt, die nicht zu den Bereichen geometrischer Verzerrungen gehören. Dieses Maß ist in der Literatur als PSNR* bekannt. Da der blockbasierte Coder TMN5 das Konzept der geometrischen Verzerrungen nicht berücksichtigt, wurde hier auf ein solches Maß verzichtet.

Im Gegensatz zum blockbasierten Hybridcoder TMN5 ist bei den objektbasierten Analyse–Synthese Codern bei steigender Gesamtdatenrate nur ein geringer Anstieg des PSNR zu beobachten. Dies begründet sich darin, daß ein objektbasierter Analyse–Synthese Coder die mehr zur Verfügung stehende Datenrate für eine bessere Codierung der Texturparameter

nutzt, was nur zu einer geringen Verbesserung des PSNR führt. Um das PSNR zu erhöhen, könnte die mehr zur Verfügung stehende Datenrate auch für eine bessere Codierung der Form– und Bewegungsparameter genutzt werden. Dies wurde in dieser Arbeit nicht untersucht.

Ein Vergleich des Quellenmodells A3D mit den Quellenmodellen R3D und F3D zeigt, daß das Quellenmodell A3D effizienter arbeitet, besonders bei Datenraten um 5000 bit/Bild. Aufgrund der besseren Beschreibung von Form und Bewegung der Objekte wird eine Verringerung des Synthesefehlers gegenüber dem Quellenmodell R3D erreicht. Die zusätzliche Übertragung der flexiblen Verschiebungsparameter beim Quellenmodell F3D führt dazu, daß das Quellenmodell A3D auch gegenüber diesem Quellenmodell F3D effizienter arbeitet. Ein Vergleich des Quellenmodells F3D mit dem Quellenmodell R3D zeigt, daß das Quellenmodell F3D nur bei höheren Bildqualitäten effizienter als das Quellenmodell R3D arbeitet.

Obwohl das Quellenmodell A3D weniger Freiheitsgrade in der Formbeschreibung als das Quellenmodell F3D hat, arbeitet das Quellenmodell A3D effizienter, da beim Quellenmodell F3D die flexiblen Formparameter zusätzlich übertragen werden müssen.

Vergleicht man nun die Effizienz der Quellenmodelle nicht bei gleicher Datenrate, sondern bei gleicher Bildqualität PSNR=38.25 dB, so zeigt Bild 7.2, daß bei Verwendung der Quellenmodelle R3D und F3D etwa 6400 bit/Bild bzw. 5600 bit/Bild benötigt werden, während bei Verwendung des Quellenmodells A3D nur etwa 5000 bit/Bild aufgewendet werden müssen. Der erzielte Codiergewinn mit dem Quellenmodell A3D begründet sich in der genaueren Nachbildung der Form und Bewegung realer Objekte, die mit diesem Quellenmodell insbesondere bei untereinander unterschiedlicher Bewegung der Objektkomponenten möglich ist.

Vergleich der Bildqualität anhand subjektiver Gütekriterien

Neben den objektiven Kriterien für die Bildqualität ist für einen menschlichen Beobachter zur Beurteilung der Bildqualität vor allem der subjektive Eindruck von Bedeutung. Die Bilder 7.3 und 7.4 zeigen die Codierergebnisse für die objektbasierten Coder und als Referenz für den blockbasierten Hybridcoder TMN5.

Die störendsten Codierfehler des blockbasierten Hybridcoders TMN5 bei einer Datenrate von 5000 bit/Bild sind die Block– und Moskitoeffekte. Die Moskitoeffekte werden dadurch verursacht, daß den Blöcken, die mehrere Bildbereiche unterschiedlich bewegter Objekte enthalten, nur ein Displacementvektor zugeordnet wird. Die Bewegungskompensation verursacht deshalb Synthesefehler, die anschließend codiert werden müssen. Aufgrund der niedrigen Datenrate ist aber die Codierung des Syntesefehlers verlustbehaftet, so daß sichtbare Codierfehler in Form von Moskitoeffekten entstehen. Bei einem objektbasierten Coder treten Moskitoeffekte nicht auf (Bild 7.5), da an den Objekträndern aufgrund der geschätzten Formparameter keine Fehler durch die Bewegungskompensation auftreten und die Texturparameter sowohl für die freigewordenen Bildbereiche als auch für die MF–Objekte mit einer Datenrate von mindestens 1.1 bit/pel übertragen werden. Blockeffekte entstehen vor allem dann, wenn für benachbarte Blöcke unterschiedliche Displacementvektoren geschätzt wurden und der entstehende Synthesefehler in Form eines Versatzes der Texturparameter an den Blockgren-

a) b)

c) d)

Bild 7.3 Codierergebnisse zu Bild 7.1.a für verschiedene Quellenmodelle: a) blockbasierter Hybridcoder TMN5 mit 50 kbit/s, b) R3D mit 64 kbit/s, c) F3D mit 56 kbit/s und d) A3D 50 kbit/s. Testsequenz *Claire*.

zen nicht durch zusätzlich übertragene Texturparameter kompensiert wird. Es entstehen daher im rekonstruierten Bild sichtbare horizontal und vertikal ausgerichtete Linien, sog. Blockeffekte, die die subjektive Bildqualität stören. Diese störenden Codierfehler treten beim objektbasierten Coder nicht auf (Bild 7.6), da aufgrund der Bewegungsbeschreibung entsprechend Gl. (4.9) jedem Bildpunkt ein eigener Displacementvektor zugeordnet werden kann.

a) b)

c) d)

Bild 7.4 Codierergebnisse zu Bild 7.1.b für verschiedene Quellenmodelle: a) blockbasierter Hybridcoder TMN5 mit 50 kbit/s, b) R3D mit 64 kbit/s, c) F3D mit 56 kbit/s und d) A3D 50 kbit/s. Testsequenz *Miss America*.

a)

b)

Bild 7.5 Veranschaulichung der Moskitoeffekte an einem Ausschnitt aus Bild 7.3: a) blockbasierter Hybridcoder TMN5 mit 50kbit/s, b) objektbasierter Coder basierend auf dem Quellenmodell A3D mit 50 kbit/s.

a) b)

Bild 7.6 Veranschaulichung der Blockeffekte an einem Ausschnitt aus Bild 7.4: a) blockbasierter Hybridcoder TMN5 mit 50kbit/s, b) objektbasierter Coder basierend auf dem Quellenmodell A3D mit 50 kbit/s.

8. Zusammenfassung

Im Rahmen dieser Arbeit wurde für die Bewegtbildcodierung ein objektbasierter Analyse–Synthese–Coder entwickelt, der auf dem Quellenmodell bewegter dreidimensionaler, gegliederter Objekte (A3D) basiert. Ziel war es, mit den zu codierenden und zu übertragenen Parametern des neuen Quellenmodells eine effizientere Beschreibung und Codierung einer Bildfolge zu erreichen. Dafür mußten Algorithmen für die Schätzung der Parameter der dreidimensionalen Form und Bewegung gegliederter Objekte entwickelt werden.

Das entwickelte Verfahren zur Schätzung der Formparameter gegliederter Objekte besteht aus den Schritten Forminitialisierung, Objektgliederung, Formadaption sowie Schätzung der Position eines jeden Gelenkes bezüglich des Koordinatensystems der zugehörigen Objektkomponente.

Bei der Forminitialisierung wird jedes reale Objekt durch ein starres Modellobjekt beschrieben. Die Form des starren Modellobjektes wird durch ein starres Dreiecksnetz aufgespannt. Die Stützpunkte des Dreiecksnetzes bilden die Parameter zur Beschreibung der starren Form. Das starre Dreiecksnetz wird wie beim Quellenmodell R3D aus der effizient zu codierenden Silhouette des korrespondierenden realen Objektes unter der Hypothese ellipsoider Form mit Hilfe einer Distanztransformation so berechnet, daß die Dreiecke im Raum etwa gleiche Kantenlängen aufweisen. Die Silhouette eines realen Objektes wird mit Hilfe einer Änderungsdetektion zwischen dem zuletzt übertragenen und dem aktuellen realen Bild bestimmt. Für typische Bildtelefonanwendungen wird der sichtbare Oberflächenteil eines Modellobjektes durch 200 bis 400 Stützpunkte und 350 bis 700 Dreiecke dargestellt[65]. Beim hier verwendeten Bildformat CIF (352*288 pel^2) beträgt dabei die mittlere Größe der zu einem Dreieck korrespondierenden Bildfläche etwa 65 pel^2.

Bei der Objektgliederung wird jedes starre Modellobjekt in Objektkomponenten aufgeteilt. Zur Objektgliederung wird die Annahme starrer Objektkomponenten ausgenutzt. Dazu werden benachbarte Dreiecke des starren Modellobjektes, die ähnliche Parameter der dreidimensionalen Bewegung aufweisen, in Oberflächengebiete zusammengefaßt. Im idealen Fall stellen die so gefundenen Oberflächengebiete schon die Objektkomponenten dar. Aufgrund von Schätzfehlern in den Bewegungsparametern der Dreiecke sowie unzureichender Bewegung der Objektkomponenten stellen die erhaltenen Oberflächengebiete i.a. nur Teile der Objektkomponenten dar. Um die Objektkomponenten vollständig zu bestimmen, werden die in der Analyse vorheriger Bilder erhaltenen Ergebnisse der Zusammenfassung von Dreiecken bei der Analyse des aktuellen Bildes berücksichtigt. Dazu besitzt jedes Dreieck eines starren Modellobjektes ein Gedächtnis für die Zugehörigkeit zu mehreren Oberflächengebieten. Ein während der Analyse des aktuellen Bildes ermitteltes Oberflächengebiet erhält ein neues Kennzeichen, wenn es sich mit keinem der bereits gekennzeichneten Oberflächengebiete überlappt. Im Falle von Überlappung wird das bereits eingetragene Kennzeichen für alle Dreiecke des auf diese Weise erweiterten Oberflächengebietes übernommen. Aus Stabilitätsgründen werden nur Erweiterungen zugelassen, die eine Vergrößerung der Gesamtfläche bis

zu 30% ergeben. Ein gekennzeichnetes Oberflächengebiet wird als Modellobjektkomponente detektiert, wenn es in zwei aufeinanderfolgenden Bildern seine Gesamtfläche nicht verändert hat und eine feste Mindestgröße von 10% der Gesamtfläche des sichtbaren Oberflächenteils des starren Modellobjektes überschreitet. Falls eine Objektkomponente detektiert wird, werden die entsprechenden Stützpunkte des ursprünglichen starren Modellobjektes der neuen Objektkomponente zugeordnet. Die Dreiecksnetze zweier auf diese Weise entstandenen Objektkomponenten bleiben über gemeinsame Dreiecke flexibel untereinander verbunden.

Die für die Objektgliederung erforderliche Schätzung der Parameter der dreidimensionalen Bewegung eines Dreiecks erfolgt mittels eines Maximum–Likelihood–Schätzverfahrens. Das Verfahren ermittelt diejenigen Parameter der dreidimensionalen Bewegung als Schätzwert, bei denen die bedingte Wahrscheinlichkeit der zeitlichen Luminanzdifferenzen an Beobachtungspunkten maximal ist. Als Beobachtungspunkte werden Bildpunkte mit einem hohen örtlichen Luminanzgradienten ausgewählt. Die bedingte Wahrscheinlichkeit wird dabei als Funktion der zu schätzenden Bewegungsparameter, der gemessenen Luminanzdifferenzen sowie der Kovarianzmatrix der Luminanzfehler beschrieben. Der Luminanzfehler an jedem Beobachtungspunkt setzt sich additiv aus einem Luminanzfehler aufgrund von Kamerarauschen und einem weiteren Luminanzfehler aufgrund von Fehlern der dreidimensionalen Form des starren Modellobjektes zusammen. Die Kovarianzmatrix der Luminanzfehler wird gewonnen, indem sowohl das Kamerarauschen als auch der Fehler der dreidimensionalen Form durch stationäre, mittelwertfreie, Gauß'sche Zufallsprozesse beschrieben werden. Um die Konvergenzwahrscheinlichkeit bei der Bestimmung der Parameter der dreidimensionalen Bewegung eines Dreiecks zu erhöhen, werden Beobachtungspunkte benachbarter Dreiecke in die Schätzung miteinbezogen. Zur weiteren Erhöhung der Konvergenzwahrscheinlichkeit werden Ausreißer unter den Beobachtungspunkten detektiert und aus der Bewegungsschätzung ausgeschlossen. Dazu arbeitet der Maximum–Likelihood–Schätzer rekursiv. In jedem Schritt der Rekursion wird unter Verwendung eines neu hinzugenommenen Beobachtungspunktes und der schon verarbeiteten Beobachtungspunkte ein neuer, verbesserter Satz von Bewegungsparametern bestimmt. Dann wird die mittlere quadratische Luminanzdifferenz über alle Beobachtungspunkte nach der Bewegungskompensation des Dreieckes und seiner benachbarten Dreiecke mittels des alten und neuen Satzes von Bewegungsparametern berechnet. Ist die mit dem neuen Satz von Bewegungsparametern berechnete, mittlere quadratische Luminanzdifferenz größer, so wird der korrespondierende Beobachtungspunkt als Ausreißer betrachtet und von der Bewegungsschätzung ausgeschlossen. Ein Vergleich mit einem aus der Literatur bekannten Gradientenverfahren zur Bestimmung der Parameter der dreidimensionalen Bewegung eines Dreickes eines starren Modellobjektes ergibt beim hier verwendeten Bildformat CIF sowie einer mittleren Größe der zu einem Dreieck korrespondierenden Bildfläche von 65 pel^2 eine Erhöhung der Konvergenzwahrscheinlichkeit von 10% auf 80%.

Zur Erkennung einer gegenseitigen Verdeckung von Objektkomponenten wird die Übereinstimmung ihrer Silhouetten überprüft. Für die betrachteten Testsequenzen wird folgendes Kriterium verwendet. Falls weniger als 20% der Silhouette einer Modellobjektkomponente

mit der Modellobjektsilhouette zusammenfällt, wird die Objektkomponente als verdeckend erkannt. Im Falle gegenseitiger Verdeckung der Objektkomponenten muß die Form der verdeckenden Objektkomponente durch ein zusätzliches separates, geschlossenes Dreiecksnetz dargestellt und unter Vermeidung von räumlichen Überschneidungen vor den verdeckten Objektkomponenten plaziert werden. Das Dreiecksnetz wird aus der realen Silhouette der verdeckenden Objektkomponente unter der Hypothese ellipsoider Form bestimmt. Aufgrund von Fehlern bei der oben beschriebenen Zusammenfassung benachbarter Dreiecke zu Oberflächengebieten stimmt die polygonförmige Berandung eines Oberflächengebietes nicht exakt mit der realen Silhouette der verdeckenden Objektkomponente überein. Die reale Silhouette der verdeckenden Objektkomponente wird daher vor der Netzerzeugung aus einer lokalen Segmentierung eines Displacementvektorfeldes in Regionen mit einheitlichem Betrag des Displacementvektors bestimmt. Das Displacementvektorfeld wird aus dem vorherigen und aktuellen Bild geschätzt. Zur Segmentierung wird ein aus der Literatur bekanntes Maximum–Likelihood–Schwellwertverfahren basierend auf Mischverteilungsmodellen verwendet. Zur Verfeinerung der Segmentierung werden zusätzlich örtliche Luminanzgradienten des aktuellen Bildes ausgewertet.

Da durch die Hypothese ellipsoider Form bei der Netzerzeugung die Form der Objektkomponenten nicht genau beschrieben wird, unterscheiden sich nach einer Rotation der Objektkomponenten im dreidimensionalen Raum die reale Objektsilhouette und die Modellobjektsilhouette. Die Formadaption paßt daher die Form der Modellobjektkomponenten im Laufe der Bildsequenz an die reale Objektsilhouette an. Dazu werden die sichtbaren Stützpunkte der Modellobjektkomponenten an der Modellobjektsilhouette parallel zur Kamerabildebene so verschoben, daß die reale Objektsilhouette und die Modellobjektsilhouette übereinstimmen. Die reale Objektsilhouette wird wie beim Quellenmodell R3D aus einer Änderungsdetektion zwischen dem zuletzt übertragenen Bild und dem aktuellen realen Bild unter Verwendung der geschätzten Bewegungsparameter der Modellobjektkomponenten bestimmt. Die Form der verdeckenden Modellobjektkomponenten wird ebenfalls mittels ihrer realen Silhouetten angepaßt. Die reale Silhouette einer verdeckenden Objektkomponente wird wiederum aus einer lokalen Segmentierung eines Displacementvektorfeldes mittels des oben beschriebenen Verfahrens bestimmt. Um die Zuverlässigkeit der Segmentierung zu verbessern, werden Segmentierungsergebnisse des vorangegangenen Bildes bei der aktuellen Segmentierung mitberücksichtigt.

Unter Kenntnis der aktuellen Position zweier gekoppelter Objektkomponenten eines gegliederten Objektes im Raum und mindestens zweier aufeinanderfolgender Bewegungsparametersätze pro Objektkomponente, die ihre Bewegung von den vorangegangenen Zeitpunkten bis zum aktuellen Zeitpunkt beschreiben, läßt sich unter Verwendung der Gelenkbindung die Position des Gelenkes bezogen auf das Koordinatensystem der zugehörigen Objektkomponente eindeutig berechnen. Die sechs Bewegungsparameter einer Objektkomponente werden mit dem oben erwähnten Maximum–Likelihood–Schätzer bestimmt, wobei Beobachtungspunkte aus allen Dreiecken der Objektkomponente entnommen werden. Um die Zuverlässig-

keit des Verfahrens zur Schätzung der Position eines Gelenkes bei fehlerbehafteten Bewegungsparametern zu verbessern, werden mehr als zwei Bewegungsparametersätze pro Objektkomponente verwendet. Die Position des Gelenkes wird dann mittels linearer Regression geschätzt. Beim Bildformat CIF, einer Leistung des Kamerarauschens gemessen in PSNR von 40 dB sowie bekannter exakter Form und Position zweier gekoppelter Objektkomponenten wird für das Gelenk unter Verwendung von fünf Bewegungsparametersätzen je Objektkomponente eine mittlere Abweichung der Position des Gelenkes von 0.26% bezogen auf die Bildbreite eines Bildes in dem hier verwendeten Bildformat CIF erreicht.

Das Verfahren zur Schätzung der Parameter der dreidimensionalen Form gegliederter Objekte wurde anhand international ausgewählter Bildfernsprechsequenzen sowie zusätzlicher Testsequenzen getestet, die Kopf–, Schulter– und Armbewegungen von Personen zeigen. Die Sequenzen lagen im Format CIF mit einer reduzierten Bildfolgefrequenz von 10Hz vor. Die experimentellen Untersuchungen zeigen eine wirklichkeitsnahe Aufteilung eines starren Modellobjektes in Modellobjektkomponenten. Die Objektkomponenten wurden dabei nach durchschnittlich acht Bildern detektiert.

Zur Verbesserung der Genauigkeit der Bewegungsschätzung eines gegliederten Objektes wurde ein Verfahren entwickelt, das für jede Objektkomponente unter Berücksichtigung der von den Gelenken der Objektkomponente auferlegten Einschränkungen der Bewegungsmöglichkeiten nur die unabhängigen Bewegungsparameter schätzt. Hierbei sind die Rotationsparameter die unabhängigen Bewegungsparameter einer Objektkomponente. Dazu wird das gegliederte Objekt durch einen Graphen mit Baumstruktur beschrieben, dessen Wurzel die größte Objektkomponente darstellt. Zunächst werden die sechs Bewegungsparameter der Wurzelobjektkomponente mittels des oben erwähnten Maximum–Likelihood–Schätzers bestimmt. Danach werden jeweils nur die Rotationsparameter der restlichen Objektkomponenten von der Wurzelobjektkomponente beginnend nacheinander geschätzt. Die Schätzung der Rotationsparameter einer Objektkomponente erfolgt in zwei Schritten. Im ersten Schritt werden zunächst die abhängigen Translationsparameter mittels der Gelenkbindung aus den Bewegungsparametern und der Position der vorangegangenen Bezugsobjektkomponente berechnet. Die Objektkomponente wird mit den berechneten Translationsparametern verschoben. Im zweiten Schritt werden die Rotationsparameter mittels des oben beschriebenen Maximum–Likelihood–Schätzers bestimmt. Beim Bildformat CIF, einer Leistung des Kamerarauschens gemessen in PSNR von 40 dB sowie bekannter exakter Form und Position eines gegliederten Objektes bestehend aus zwei durch ein Kugelgelenk gekoppelten Objektkomponenten ergibt sich unter Ausnutzung der Gelenkbindung und für einen Wertebereich je Komponente der Translations– und Rotationsparamter von $\pm$ 16pel bzw. $\pm$ 10grad eine Verbesserung der mittleren Varianz der Schätzfehler der Translationsparameter und der mittleren Varianz der Schätzfehler der Rotationsparameter je Objektkomponente von 30% bzw. 35%.

Im Vergleich zum Quellenmodell R3D erfordert das Quellenmodell A3D für jedes Objekt die zusätzliche, aber nur einmalige Übertragung einer Kennzahl je Stützpunkt, mit der die Zugehörigkeit zu einer Objektkomponente festgelegt wird, und der drei Komponenten des Ortsvektors der Position eines jeden Gelenkes bezogen auf das Koordinatensystem der zugehörigen Objektkomponente. Ferner ist die laufende Übertragung der drei Translationsparameter der Wurzelobjektkomponente und der drei Rotationsparameter je Objektkomponente des gegliederten Objektes für jedes Bild einer Sequenz erforderlich. Für die im Rahmen dieser Arbeit für die Codierung ausgewählten internationalen Bildfernsprechsequenzen mit Kopf– und Schulterbewegungen im Format CIF mit einer reduzierten Bildfolgefrequenz von 10Hz verringert sich jedoch durch den Einsatz des Quellenmodells A3D die Größe der nicht modellierbaren Bildbereiche (MF–Objekte) und somit die Fläche, für die Texturparameter übertragen werden müssen, von 4% auf weniger als 3% der Bildfläche.

Die Parametercodierung wurde so eingestellt, daß sich eine möglichst hohe Bildqualität bei einer möglichst niedrigen Gesamtdatenrate ergibt. Die Gesamtdatenrate besteht dabei aus den Teildatenraten für die Übertragung der Bewegungs–, Form– und Texturparameter. Die Bewegungsparameter werden gleichförmig quantisiert und in dieser ersten Version in PCM–Darstellung übertragen. Für die Übertragung der drei Translationsparameter der Wurzelobjektkomponente sind 22 bit/Bild und für die Übertragung der drei Rotationsparameter je Objektkomponente 27 bit/Bild erforderlich. Für die Übertragung der Form eines gegliederten Objektes wird einmalig seine Silhouette, für jeden Stützpunkt die Kennzahl für die Zugehörigkeit zu einer Objektkomponente und die drei Komponenten des Ortsvektores der Position eines jeden Gelenkes codiert. Dabei ist für die fortlaufende Formadaption eines gegliederten Objektes gemäß dem oben beschriebenen Verfahren die Übertragung seiner realen Silhouette auch für jedes folgende Bild der Sequenz erforderlich. Die Codierung der Silhouette erfolgt wie beim Quellenmodell R3D durch eine prädiktive Codierung der Positionen der Stützpunkte am Rand der Silhouette. Dies erfordert eine Datenrate von etwa 400 bit/Bild. Die Kennzahlen der Stützpunkte werden mittels einer Lauflängencodierung codiert. Die Reihenfolge der Stützpunkte ist dem Empfänger bekannt. Die drei Komponenten des Ortsvektors der Position eines Gelenkes werden gleichförmig quantisiert und in dieser ersten Version in PCM–Darstellung übertragen. Insgesamt erfordert die Codierung der Formparameter des Quellenmodells A3D eine Datenrate von etwa 450 bit/Bild.

Für MF–Objekte werden Form– und Texturparameter wie beim Coder für das Quellenmodell R3D codiert und übertragen. Die Formparameter der MF–Objekte werden mittels einer Polygon–Spline–Approximation codiert. Das erfordert eine Datenrate von etwa 1000 bit/Bild. Die Texturparameter der MF–Objekte und der durch die Bewegung der Objekte freiwerdenden Bildbereiche werden prädiktiv mittels einer formangepaßten, Diskreten Cosinus–Transformation und Quantisierung der Koeffizienten codiert. Die Codierung der Texturparameter erfordert dabei mindestens eine Datenrate von 1.1 bit/pel[65].

Eine Bewertung der verschiedenen dreidimensionalen Quellenmodelle wurde anhand eines meßbaren Gütekriteriums für die Bildqualität vorgenommen. Als Gütekriterium wurde dabei die Codierfehlerleistung gemessen in PSNR verwendet. Bei einem PSNR von 38 dB und dem angegebenen Format CIF erreicht ein Coder mit dem Quellenmodell A3D eine Datenrate von 50 kbit/s. Bei gleicher Bildqualität erzielt dabei ein Coder mit den Quellenmodell F3D und mit dem Quellenmodell R3D eine Datenrate von 56 kbit/s bzw. 64 kbit/s. Trotz der zusätzlich zu übertragenen Informationen für die Bewegung und Form realer Objekte wurde mit dem Quellenmodell A3D ein Codiergewinn erzielt. Dieser begründet sich in der genaueren Nachbildung der Form und Bewegung realer Objekte, die mit diesem Quellenmodell insbesondere bei untereinander unterschiedlicher Bewegung der Objektkomponenten möglich ist.

Ein subjektiver Vergleich der codierten Sequenzen des im Rahmen dieser Arbeit vorgestellten objektbasierten Analyse–Synthese–Coders mit denen des blockbasierten Coders gemäß dem Standard H.263 zeigt bei einer Datenrate von 5000 bit/Bild, daß sich durch die objektbasierte Codierung aufgrund der Übertragung der Objektform die störenden Codierfehler eines blockbasierten Hybridcoders wie Moskito– und Blockartefakte im gesamten Bild vermeiden lassen.

9. Literatur

[1] J. Aggarwal and N. Nandhakumar, ”On the Computation of Motion from Sequences of Images–A Review”, Proceedings of the IEEE, Vol. 76, No. 8, August 1988.

[2] J. Aggarwal, Q. Cai, W. Liao and B. Sabata, ”Articulated and Elastic Non–rigid Motion: A Review”, Proc. IEEE Workshop on Motion of Non–Rigid and Articulated Objects, Austin, Texas, S. 2–14, November 1994.

[3] S. Beyer, ”Displacement–Schätzverfahren für Fernsehsignale mit minimaler Schätzfehlervarianz”, Fortschritt–Berichte VDI, Reihe 10, No. 51, Düsseldorf: VDI–Verlag, 1986.

[4] D. Biere, W. Meurer, T. Kummerow, ”Video–Codec und Endgeräte für das ISDN–Bildtelefon”, Nachrichtentechnische Zeitschrift (NTZ), Vol. 42, No. 3, S. 136–145, 1989.

[5] M. Bierling, ”Displacement Estimation by Hierarchical Blockmatching”, 3rd SPIE Symposium on Visual Communications and Image Processing, Cambridge, USA, S. 942–951, November 1988.

[6] M. Bierling, ”Hierarchische Displacementschätzung zur Bewegungskompensation in digitalen Fernsehbildsequenzen”, Dissertation, Universität Hannover, Fortschritt–Berichte VDI, Reihe 10, No. 179, Düsseldorf: VDI–Verlag, 1991.

[7] A. Bouagina, ”Fortlaufende 3D–Formschäzung gegliederter 3D–Modellobjekte bei gegenseitiger Verdeckung in Bewegtbilsequenzen”, Studienarbeit, Universität Hannover, 1996.

[8] K. Brammer aud G. Siffling, ”Kalman–Bucy–Filter”, Münschen: Oldenbourg Verlag, 1985.

[9] British Telecom Research Laboratories, ”Testsequence Miss America”, CIF, 10 Hz, 50 Bilder, Martlesham, Großbritannien.

[10] H. Busch, ”Subdividing non rigid 3D objects into quasi rigid parts”, IEE 3rd International Conference on Image Processing and Applications, Warwick, UK, IEE Publ. 307, S. 1–4, Juli 1989.

[11] CCITT, ”Draft revision of recommendation 601, encoding parameters of digital television for studios”, Question 25/11, Study Programmes 25G/11, 25H/11, Doc. 11/214–E, September 1983.

[12] CCITT Recommendation H.261, ”Video Codec for Audiovisual Services at $p * 64$ kbit/s”, Dezember 1989.

[13] Centre National d’Etudes des Telecommunication, ”Testsequence Claire”, CIF, 10 Hz, 156 Bilder, Paris, Frankreich.

[14] N. Diehl, ”Object–Oriented Motion Estimation and Segmentation in Image Sequences”, Signal Processing: Image Communication, Vol. 3, No. 1, S. 23–56, Februar 1991.

[15] R. Farhoud, ”Verfahren zur Schätzung der dreidimensionalen Position der Gelenke gegliederter Objekte für die Bewegtbildcodierung”, Studienarbeit, Universität Hannover, 1996.

[16] J. Frerichs, ”Verfahren zur Detektion kleiner bewegter Objektkomponenten bei teilweiser gegenseitiger Verdeckung für die Bewegtbildcodierung”, Studienarbeit, Universität Hannover, 1996.

[17] C. Galsbey, ”An Analysis of Histogram–Based–Thresholding Algorithms”, CVGIP: Graphical Models and Image Processing, Vol. 55, S. 532–537, November 1993.

[18] P. Gerken, ”Object–Based Analysis–Synthesis–Coding of Image Sequences at Very Low Bitrates”, COST 211ter European workshop on new techniques for coding of video signals at very low bitrates, Hannover, Germany, No. 1.1, Dezember 1993.

[19] P. Gerken, ”Object–Based Analysis–Synthesis Coding of Image Sequences at Very Low Bit Rates”, IEEE Transactions on Circuits and Systems in Video Technology: Special Issue on Very Low Bit Rate Video Coding, Vol.4, No. 3, S. 228–235, Juni 1994.

[20] P. Gerken, ”Quellenmodell des MPEG4–Video–Verifikationsmodells und weiterführende Quellenmodelle”, Fernseh– und Kinotechnik, Vol. 50, No. 8/9, S. 589–594, August 96.

[21] M. Gilge, ”Regionenorientierte Transformationscodierung in der Bildkommunikation”, Dissertation, RWTH Aachen, Fortschritt–Berichte VDI, Reihe 10, No. 128, Düsseldorf: VDI–Verlag, 1990.

[22] B. Girod, ”Ein Modell der menschlichen visuellen Wahrnehmung zur Irrelevanzreduktion von Fernsehluminanzsignalen”, Dissertation, Universität Hannover, Fortschritt–Berichte VDI, Reihe 10, No. 84, Düsseldorf: VDI–Verlag, 1988.

[23] B. Girod, E. Steinbach, N. Färber, ”Comparison of the H.263 and H.261 Video Compression Standards”, First International Symposium on Photonics Technologies and Systems for Voice, Video, and Data Communication, Pennsylvania, Oktober 1995.

[24] J. Goldbeck, ”Automatische Modellierung bewegter 3D–Objekte in Bildfolgen unter Berücksichtigung möglicher Verdeckungen”, Studienarbeit, Universität Hannover, 1993.

[25] K. Grotz, J. Mayer and G. Süßmeier, ”A 64 kbit/s Videophone Codec with Forward Analysis and Control”, Signal Processing: Image Communication, Vol. 1, No. 2, S. 103–116, Oktober 1989.

[26] R. Holt, A. Netravali, T. Huang and R. Qian, ”Determining Articulated Motion from Perspective Views: A Decomposition Approach”, Proc. IEEE Workshop on Motion of Non–Rigid and Articulated Objects, Austin, Texas, S. 126–137, November 1994.

[27] M. Hötter, R. Thoma, ”A Selfadaptive, Model Based Segmentation Algorithm”, Picture Coding Symposium (PCS ’88), Turin, Italien, No. 2.2, September 1988.

[28] M. Hötter, ”Predictive Contour Coding for an Object–Oriented Analysis–Synthesis Coder”, IEEE International Symposium on Information Theory, San Diego, California, USA, S. 75, Januar 1990.

[29] M. Hötter, ”Object–Oriented Analysis–Synthesis Coding Based on the Model of Flexible 2D–Objects”, Picture Coding Symposium (PCS ’90), Cambridge, USA, No. 7.1, März 1990.

[30] M. Hötter, ”Object–Oriented Analysis–Synthesis Coding Based on Moving Two–Dimensional Objects”, Signal Processing: Image Communication, Vol. 2, No. 4, S. 409–428, Dezember 1990.

[31] M. Hötter, ”Optimization of an Object–Oriented Analysis–Synthesis Coder Based on the Model of Flexible 2D–Objects”, Picture Coding Symposium (PCS ’91), Tokyo, Japan, No. 10.4, September 1991.

[32] M. Hötter, ”Objektorientierte Analyse–Synthese–Codierung basierend auf dem Modell bewegter, zweidimensionaler Objekte”, Dissertation, Universität Hannover, Fortschritt–Berichte VDI, Reihe 10, No. 217, Düsseldorf: VDI–Verlag, 1992.

[33] M. Hötter, ”Optimization and Efficiency of an Object–Oriented Analysis–Synthesis Coder”, IEEE Transactions on Circuits and Systems for video technology, Vol. 4, No. 2, S. 181–194, April 1994.

[34] T. Huang and A. N. Netravali, ”Motion and Structure from Feature Correspondences: A Review”, Proceedings of the IEEE, Vol. 82, No. 2, February 1994.

[35] P. Huber, ”Robust Statistics”, New York: Wiley–Verlag, 1981.

[36] Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, ”Testsequence Anis&Markus”, CIF, 10 Hz, 50 Bilder, Hannover, Deutschland.

[37] Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, ”Testsequence Giovanni”, CIF, 10 Hz, 50 Bilder, Hannover, Deutschland.

[38] Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, ”Testsequence Goldbeck”, CIF, 10 Hz, 50 Bilder, Hannover, Deutschland.

[39] ISO/IEC DIS 13818 (Internationaler Standard MPEG–2).

[40] ISO/IEC IS 11172 (Internationaler Standard MPEG–1).

[41] ITU–T Recommendation H.261, ”Video codec for audiovisual services at p x 64 kbit/s”, März 1993.

[42] ITU–T Recommendation H.263, ”Video Coding for Narrow Telecommunication Channels at < 64 kbit/s”, Dezember 1995.

[43] I. Kakadiaris, D. Metaxas and R. Bajcsy, ”Active part–decomposition, shape und motion estimation of articulated objects: A physics–based approach”, Proceedings CVPR, S. 980–984, 1994.

[44] M. Kampmann and J. Ostermann, ”Automatic Adaption of a Face Model in a Layered Coder with an Object–based Analysis–Synthesis Layer and a Knowledge–based Layer”, Signal Processing: Image Communication, Vol. 9, No. 3, S. 201–220, März 1997.

[45] F. Kappei, ”Modellierung und Rekonstruktion bewegter dreidimensionaler Objekte in einer Fernsehbildfolge”, Dissertation, Universität Hannover, 1989.

[46] T. Kaßbaum, ”Korrektur von Änderungsmasken durch Auswertung lokaler Konturinformation”, Studienarbeit, Universität Hannover, 1989.

[47] R. Koch, ”Dynamic 3D scene analysis through synthesis feedback control”, IEEE T–PAMI, Vol. 15, No. 6, S. 556–568, Juni 1993.

[48] T. Kurita, N. Otsu and N. Abdelmalek, ”Maximum Likelihood Thresholding Based on Population Mixture Models”, Pattern Recognition, Vol. 25, No. 10, S. 1231–1240, März 1992.

[49] G. Leister, ”Beschreibung und Simulation von Mehrkörpersystemen mit geschlossenen kinematischen Schleifen”, Dissertation, Universität Stuttgart, Fortschritt–Berichte VDI, Reihe 11, No. 167, Düsseldorf: VDI–Verlag, 1992.

[50] G. Martínez, ”3D Motion Estimation of Articulated 3D Objects for Object–Based Analysis–Synthesis Coding (OBASC)”, International Workshop on Coding Techniques for Very Low Bit–rate Video, Tokyo, Japan, No. G–1, November 1995.

[51] G. Martínez, ”Shape Estimation of Articulated 3D Objects considering mutual occlusions for object–based analysis–synthesis coding (OBASC)”, Picture Coding Symposium 96 (PCS’96), Melbourne, Australia, S. 141–146, März 1996.

[52] G. Martínez, ”Shape Estimation of Articulated 3D Objects for Object–Based Analysis–Synthesis Coding (OBASC)”, Signal Processing: Image Communications, Vol. 9, No. 3, S. 175–199, März 1997.

[53] Matsushita Electric Ind., ”Testsequence Akiyo”, CIF, 10 Hz, 100 Bilder, Japan.

[54] F. May, ”Algorithmen und Realisierung eines ISDN–Bildtelefon–Codecs”, Nachrichtentechnische Zeitschrift (NTZ), Vol. 42, No. 3, S. 130–133, 1989.

[55] J. Mendel, ”Lessons in estimation theory for signal processing, communications and control”, Englewood Cliffs, New Jersey: Prentice Hall, 1995.

[56] H. Musmann, M. Hötter, J. Ostermann, ”Object–Oriented Analysis–Synthesis Coding of Moving Images”, Signal Processing: Image Communication, Vol. 1, No. 2, S. 117–138, Oktober 1989.

[57] H. Musmann, ”A Layered Coding System for Very Low Bit Rate Video Coding”, Signal Processing: Image Communication, Vol. 7, No. 4, S. 267–278, November 1995.

[58] J. Ostermann, ”Modelling of 3D Moving Objects for an Analysis–Synthesis Coder”, SPIE/SPSE Symposium on Sensing and Reconstruction of 3D Objects and Scenes, Santa Clara, California, Proc. SPIE 1260, Februar 1990.

[59] J. Ostermann, H. Li, ”Detection and Coding of Model Failures in an Analysis–Synthesis Coder Based on Moving 3D–Objects”, 3rd International Workshop on 64 kbit/s Coding of Moving Video, Rotterdam, The Netherlands, No. 4.4, September 1990.

[60] J. Ostermann, H. Li, ”Coding of Color Parameters in an Analysis–Synthesis Coder Based on Moving 3D–Objects”, Picture Coding Symposium (PCS ’91), Tokyo, Japan, No. 10.5, September 1991.

[61] J. Ostermann, ”An Analysis–Synthesis Coder Based on Moving Flexible 3D–objects”, Picture Coding Symposium (PCS ’93), Lausanne, Suisse, No. 2.8, März 1993.

[62] J. Ostermann, ”Object–Based Analysis–Synthesis Coding Based on the Source Model of Moving Rigid 3D–Objects”, Signal Processing: Image Communication, No. 6, S. 143–161, 1994.

[63] J. Ostermann, ”Object–Based Analysis–Synthesis Coding (OBASC) Based on the Source Model of Moving Flexible 3D–Objects”, IEEE Trans. on Image Processing, Vol. 3, No. 5, September 1994.

[64] J. Ostermann, ”Differences Between an Object–Based Analysis–Synthesis Coder and a Block–Based Hybrid Coder”, Asilomar Conference on Signals, Systems & Computers, Asilomar, Kalifornien, Oktober 1995.

[65] J. Ostermann, ”Analyse–Synthese–Codierung basierend auf dem Modell bewegter, dreidimensionaler Objekte”, Dissertation, Universität Hannover, Fortschritt–Berichte VDI, Reihe 10, No. 391, Düsseldorf: VDI–Verlag, 1995.

[66] D. Pearson, ”Texture mapping in model–based image coding”, Signal Processing: Image Communication, Vol. 2, No. 4, S. 377–395, Dezember 1990.

[67] S. Pipereit, ”Untersuchung von Antialiasingverfahren für die Bildsynthese”, Studienarbeit, Universität Hannover, 1990.

[68] W. Pratt, ”Digital Image Processing”, New York: Wiley Interscience, 1987.

[69] D. Rogers, ”Procedural elements for computer graphics”, Singapore: McGraw–Hill, 1985.

[70] M. Salecker, ”Zur Dynamik hybrider Mehrkörpersysteme Theorie und symbolische Beschreibung”, Dissertation, Universität Karlsruhe, 1990.

[71] H. Scheffler, ”Erkennung von Schatten bewegter Objekte in Bildfolgen”, Studienarbeit, Universität Hannover, 1990.

[72] H. Schiller, M. Hötter, ”Investigations on Colour coding in an object–oriented analysis–synthesis coder”, Signal Processing: Image Communication, Vol. 5, No. 4, S. 319–326, Oktober 1993.

[73] B. Schmale, ”HDTV–Teilbandcodierung mit bildpunktweiser Quantisierung”, Dissertation, Universität Hannover, Fortschritt–Berichte VDI, Reihe 10, No. 394, Düsseldorf: VDI–Verlag, 1995.

[74] K. Schmoll, ”Modularer Aufbau von Mehrkörpersystemen unter Verwendung der Relativkinematik”, Dissertation, Universität Stuttgart, Fortschritt–Berichte VDI, Reihe 18, No. 57, Düsseldorf: VDI–Verlag, 1988.

[75] H. Schoen, ”The Deutsche Bundespost on its Way Towards the ISDN”, Sonderdruck aus der Zeitschrift für das Post und Fernmeldewesen, No. 6, Juni 1984.

[76] H. Schwarz, ”Numerische Mathematik”, Stuttgart: B.G. Teubner, 1988.

[77] S. Sclove, ”Population Mixture Models and Clustering Algorithms”, Commun. Statist.–Theor. Meth., A6(5), S. 417–434, 1977.

[78] A. Shabana, ”Dynamics of Multibody Systems”, USA: John Wiley & Sons, 1989.

[79] S. Smarsli, ”Algorithmus zur Unterdrückung von Aliasingfehlern in synthetisierten Bewegtbildfolgen”, Studienarbeit, Universität Hannover, 1990.

[80] W. Speltacker and G. Weick, ”Entwicklung von Endgeräten und ihre Nutzung”, Nachrichtentechnische Zeitschrift (NTZ), Vol. 42, No. 3, S. 146–149, 1989.

[81] J. Stauder, ”An illumination estimation method for 3D object–based analysis–synthesis coding”, COST 211ter European workshop on new techniques for coding of video signals at very low bitrates, Hannover, Germany, No. 4.5, Dezember 1993.

[82] J. Stauder, ”Estimation of point light source parameters for object–based coding”, Signal Processing: Image Communication, Vol. 7, No. 4.6, S. 355–379, November 1995.

[83] D. Thalman, ”Image Synthesis: Theory and Practice”, Tokyo: Springer, 1987.

[84] R. Tsai and T. Huang, ”Estimation three–dimensional motion parameter of a rigid planar patch”, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP–29, No. 6, S. 1147–1152, Dezember 1981.

[85] J. Web. and J. Aggarwal, ”Structure from motion of rigid and jointed objects”, in Proceedings of the Seventh International Joint Conference on Artifitial Intelligence, Vancouver, S. 686–691, August 1981.

[86] H. Wei, ”Dynamische Simulation hybrider Mehrkörpersysteme mit beliebiger Systemstruktur”, Dissertation, Universität Karlsruhe, Fortschritt–Berichte VDI, Reihe 18, No. 108, Düsseldorf: VDI–Verlag, 1992.

[87] J. Weng, T.S. Huang, N. Ahuja, ”Motion and structure from two perspective views: Algorithms, error analysis and error estimation”, IEEE T–PAMI, Vol. 11, No. 5, S. 451–476, 1989.

[88] B. Welsh, ”Model–based coding of images”, Ph.D.–Thesis, Essex University, 1991.

[89] J. Wittenburg, ”Dynamics of Systems of Multi–Bodies”, Teuber–Verlag, 1977.

[90] U. Wolz, ”Dynamik von Mehrkörpersystemen – Theorie und Symbolische Programmierung”, Dissertation, Universität Karlsruhe, Fortschritt–Berichte VDI, Reihe 11, No. 75, Düsseldorf: VDI–Verlag, 1985.

[91] X. Zhuang, ”A highly robust estimator through partially likelihood function modeling and its application in computer vision”, IEEE T–PAMI, Vol. 14, No. 1, S. 19–34, Januar 1992.

Lebenslauf

Geovanni Martínez–Castillo
geboren am 15.09.1965 in San José, Costa Rica
als Sohn des Busunternehmers Carlos Martínez–Piedra
und der Hausfrau Alicia Castillo–Solano

1972 – 1978	Grundschule Jesús Jiménez–Zamora in Cartago
1978 – 1983	Gymnasium San Luis Gonzaga in Cartago
1983 – 1987	Studium der Elektrotechnik an der Universität von Costa Rica, Abschluß ”Baccalaureate in Electrical Engineering”
1987 – 1989	Fortsetzung des Studiums der Elektrotechnik an der Universität von Costa Rica, Abschluß ”Licentiate in Electrical Engineering” mit Auszeichnung
1987 – 1991	Dozent am Fachbereich Elektrotechnik des Nationalen Instituts für Technologie (I.T.C.R)
1989 – 1991	Dozent am Fachbereich Elektrotechnik der Universität von Costa Rica (U.C.R)
04/1991 – 10/1991	Deutschsprachkurs am Goethe Institut, Göttingen
10/1991 – 04/1997	Doktorand des Instituts für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover Stipendiat des Deutsches Akademischen Austauschdienst (DAAD) seit 04/1991