BM3D-verfeinerte Zwischenbildsynthese für unterschiedlich aufgelöste Referenzansichten

Dipl.-Ing. Thomas Richter, Dipl.-Ing. (FH) Eugen Wige, Prof. Dr.-Ing. André Kaup; {richter, wige, kaup}@LNT.de
Lehrstuhl für Multimediakommunikation und Signalverarbeitung,
Universität Erlangen-Nürnberg, Cauerstr. 7, 91058 Erlangen

Kurzfassung

Die Synthese unbekannter Zwischenansichten mit Hilfe von Tiefenkarten ist eine Kernaufgabe in der 3D-Videosignalverarbeitung. Um ein hochqualitatives Zwischenbild generieren zu können, werden gewöhnlich mehrere hochaufgelöste Referenzansichten verwendet. Prinzipiell werden dabei die beiden benachbarten Kameraansichten mit Hilfe der entsprechenden Tiefeninformation in die Bildebene der Zielansicht projiziert und anschließend zusammengefügt. Im Normalfall bleiben nach dem Zusammenfügen der beiden Projektionsergebnisse nur noch sehr kleine Lochstrukturen übrig, die mit unterschiedlichen Interpolationsmethoden geschlossen werden können. Um die Kosten im Bezug auf das genutzte Kameraarray zu senken, könnten einige der hochauflösenden Kameras durch niedrigaufgelöste ersetzt werden. Diese niedrigauflösenden Kameras sind zwar billiger in der Anschaffung, liefern im Gegenzug aber auch eine deutlich schlechtere Bildqualität und beeinflussen damit das Ergebnis der Zwischenbildsynthese, besonders in Bildbereichen, die nur in der niedrigaufgelösten Referenz sichtbar sind. In diesem Beitrag wird die Fragestellung behandelt, wie ein Zwischenbild hoher Qualität aus zwei Kameraansichten generiert werden kann, die stark unterschiedliche Auflösungen und damit Bilder unterschiedlicher visueller Qualität liefern. Das Hauptaugenmerk dieser Arbeit liegt auf einem Nachverarbeitungsschritt, basierend auf dem BM3D-Verfahren, mit dem Ziel, Bildbereiche, die aus der niedrigaufgelösten Referenzansicht eingefügt wurden, an ihre hochaufgelöste Umgebung anzupassen. Die Simulationsergebnisse zeigen, dass die BM3D-Verfeinerung von synthetisierten Zwischenbildern zu einem PSNR-Gewinn von bis zu 0,51 dB auf dem gesamten Bild und 5,69 dB bezogen auf die Verfeinerungspositionen führt. Auch die visuelle Bildqualität lässt sich durch die vorgeschlagene Nachverarbeitung deutlich steigern.

1. Einführung

Depth-Image-Based Rendering (DIBR) [1] ist ein Standardansatz um eine Zwischenansicht aus benachbarten Referenzansichten zu generieren. Zusätzlich zu den Referenzbildern wird dabei die Tiefeninformation der Referenzkameras ausgenutzt um neue dazwischenliegende Kameraansichten zu synthetisieren. Free-Viewpoint-Television (FTV) und 3D-Video (3DV) bilden dabei nur zwei mögliche Anwendungsgebiete für DIBR. Während der Betrachter durch FTV seinen eigenen Blickwinkel wählen kann, bekommt er mit 3DV einen räumlichen Szeneneindruck mit Hilfe von stereoskopischen oder autostereoskopischen Displaytechniken.

Neben dem Normalfall, bei dem die Referenzansichten gleiche Auflösung besitzen, ist auch ein Kamerasetup denkbar, bei dem benachbarte Referenzkameras unterschiedliche örtliche Auflösung besitzen. Es gibt unterschiedliche Gründe, warum ein solches Setup in vielen Szenarien eine sinnvolle Alternative darstellen kann. Neben den geringeren Kosten für niedrigaufgelöste Kameras bedeutet eine geringe Ortsauflösung auch typischerweise eine geringere Komplexität bezüglich der notwendigen Datenübertragung. Allerdings wird bei der Synthese unbekannter Zwischenbilder die Information aus der niedrigaufgelösten Nachbaransicht für alle Bildbereiche benötigt, die in der hochauflösenden Referenzansicht entweder nicht sichtbar sind oder auf Grund leichter Schwankungen in der Tiefenkarte nicht korrekt abgebildet werden können. Das Einfügen dieser niedrigaufgelösten Bildbereiche führt zu deutlich sichtbaren Artefakten und damit zu einer Verschlechterung der Bildqualität im gewünschten Zwischenbild.

Um hochaufgelöste Zwischenbilder aus unterschiedlich aufgelösten Referenzansichten zu synthetisieren wurde ein Ansatz in [2] vorgestellt. Dabei werden beide Referenzansichten in die Bildebene der gewünschten Zielansicht projiziert. Anschließend werden hohe Frequenzen im Projektionsergebnis der hochauflösenden Referenzansicht gesucht. Das Ergebnis dieser Frequenzdetektion bildet die Basis für eine pixelweise Gewichtungsfunktion. Damit wird der Einfluss der niedrigaufgelösten Referenzansicht in texturierten Bildbereichen geringer gewichtet als in glatten Gebieten. Allerdings führt dieser Ansatz zu einem Verlust an Bildschärfe im Falle einer fehlerhaften oder ungenauen Hochfrequenzdetektion.

Die Grundidee dieser Arbeit besteht darin, die niedrigaufgelösten Bildbereiche als rauschüberlagerte Version der unbekannten hochaufgelösten Information zu betrachten. Basierend darauf, wurde bereits das NLM-Verfahren (Non-Local-Means) [3] angewendet, um die niedrigfrequenten Bildbereiche an ihre hochaufgelöste Umgebung anzupassen [4]. In dieser Arbeit wird der BM3D-Ansatz (Block Matching and 3D Filtering) [5] für diese Problemstellung untersucht.

Der Beitrag ist wie folgt gegliedert. In Kapitel 2 werden die Grundlagen der Zwischenbildsynthese basierend auf

DIBR behandelt. Kapitel 3 beschreibt die Grundidee des BM3D-Algorithmus, sowie seine Anwendung zur Anpassung niederfrequenter Bildbereiche an hochaufgelöste Nachbarschaften. Simulationsergebnisse werden in Kapitel 4 gezeigt. Die Arbeit schließt mit einer Schlussfolgerung in Kapitel 5 ab.

2. Zwischenbildsynthese

Zwischenbildsynthese [6] ist ein generelles Konzept, um unbekannte Kameraperspektiven aus benachbarten Referenzansichten zu schätzen und besteht typischerweise aus drei Schritten. Zuerst werden sowohl die linke als auch die rechte Referenzansicht in die Bildebene der gewünschten Zwischenansicht projiziert. Anschließend werden die beiden Projektionsergebnisse zu einem synthetisierten Bild zusammengefügt. Zum Schluss wird ein Füllalgorithmus verwendet, um diejenigen Bildbereiche zu schließen, für die keine korrespondierende Bildinformation in einer der beiden Referenzansichten gefunden wurde. Die einzelnen Schritte der Zwischenbildsynthese werden im Folgenden detailliert beschrieben.

Mit Hilfe der folgenden Gleichung lässt sich ein Bildpunkt (u_r, v_r) einer Referenzansicht in dreidimensionale Weltkoordinaten (x_w, y_w, z_w) überführen:

$$\begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} = \mathbf{R}_r^{-1} \cdot \left(z_r \cdot \mathbf{A}_r^{-1} \begin{pmatrix} u_r \\ v_r \\ 1 \end{pmatrix} - t_r \right), \qquad (1)$$

wobei der Index *r* für die Referenzansicht steht, die Matrix **A** die intrinsischen Kameraparameter und die Matrix **R** und der Vektor t jeweils die Rotation und die Verschiebung der Kamera bezüglich des Ursprungs des 3D-Weltkoordinatensystems beschreiben. Der Skalar z bezeichnet den physikalischen Tiefenwert und lässt sich aus der zugehörigen Tiefenkarte wie folgt berechnen:

$$\frac{1}{z} = \frac{d_r(u_r, v_r)}{255} \cdot \left(\frac{1}{z_{near}} - \frac{1}{z_{far}} \right) + \frac{1}{z_{far}}, \qquad (2)$$

wobei der entsprechende Tiefenwert mit $d_r(u_r, v_r)$ beschrieben wird und der Tiefenbereich der Szene durch z_{near} und z_{far} definiert ist. Anschließend wird der daraus resultierende Punkt in 3D-Weltkoordinaten durch

$$\begin{pmatrix} u_v \\ v_v \\ 1 \end{pmatrix} = \frac{1}{z_v} \cdot \mathbf{A}_v \cdot \left(\mathbf{R}_v \cdot \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} + t_v \right) \qquad (3)$$

auf die Bildebene der gewünschten Zielansicht projiziert. Hierbei beschreibt der Index *v* die gewünschte virtuelle Zwischenansicht. Durch Anwenden von (1) bis (3) für beide benachbarte Referenzansichten erhält man zwei Schätzungen für die gesuchte Kameraperspektive. Die gegebenen Referenzansichten seien im Folgenden mit

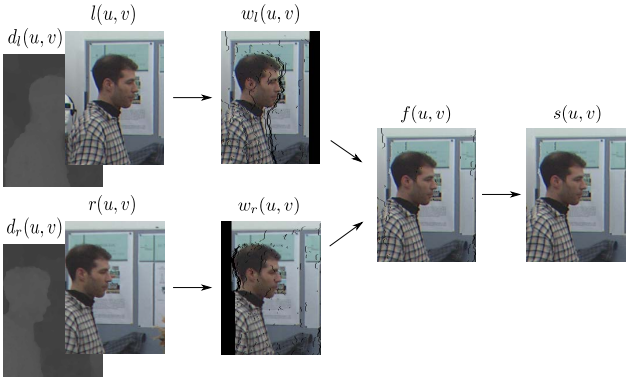

Bild 1 Prinzipieller Ablauf der Zwischenbildsynthese, am Beispiel eines Ausschnitts der *book-arrival* Testsequenz für Referenzkameras mit **identischer** örtlicher Auflösung

$l(u,v)$ und $r(u,v)$ bezeichnet. Die daraus resultierenden Syntheseergebnisse werden mit $w_l(u,v)$ beziehungsweise $w_r(u,v)$ beschrieben. Im Anschluss an die Abbildung der Referenzansichten auf die Bildebene der Zielperspektive werden die entstandenen Projektionsergebnisse kombiniert. Im Falle zweier hochaufgelöster Referenzansichten kann die Kombination an einer Bildposition (u,v) beschrieben werden als:

$$f(u,v) = \begin{cases} w_l(u,v), & falls\ (u,v) \in \overline{H}_l \cap H_r \\ w_r(u,v), & falls\ (u,v) \in H_l \cap \overline{H}_r \\ \alpha \cdot w_l(u,v) + (1-\alpha) \cdot w_r(u,v), & falls\ (u,v) \in \overline{H}_l \cap \overline{H}_r \\ 0, & falls\ (u,v) \in H_l \cap H_r \end{cases}, \qquad (4)$$

wobei α einen Gewichtungsfaktor im Bereich $0 \le \alpha \le 1$ beschreibt, H die Menge derjenigen Bildpunkte kennzeichnet, die durch die entsprechende Referenzansicht nicht abgebildet werden können (H: engl. hole) und $\overline{H}$ die Komplementärmenge von H bezeichnet. Im Falle von Referenzansichten mit gleicher örtlicher Auflösung wird der Gewichtungsfaktor typischerweise zu $\alpha = 0.5$ gewählt, um einen gleichmäßigen Einfluss benachbarter Kameraperspektiven gewährleisten zu können. In (4) wird außerdem gezeigt, dass es in der Zwischenansicht, auf Grund von Aufdeckungen oder fehlerhaften Tiefenwerten, Bildpositionen geben kann, für die weder in der linken noch in der rechten Ansicht entsprechende Information gefunden werden kann. Diese Gebiete werden abschließend mit einem Lochfüllalgorithmus geschlossen. Dafür wird beispielhaft auf [7] verwiesen.

In Bild 1 wird der Prozess der Zwischenbildsynthese für identisch aufgelöste Referenzansichten visualisiert. Zwei Referenzansichten $l(u,v)$ und $r(u,v)$ werden genutzt, um mit Hilfe der entsprechenden Tiefeninformation aus $d_r(u,v)$ und $d_l(u,v)$ die beiden Syntheseergebnisse $w_l(u,v)$ und $w_r(u,v)$ zu generieren. Anschließend werden die synthetisierten Bilder zu $f(u,v)$ kombiniert und übrig gebliebene Lochstrukturen gefüllt. Für den Fall, dass die beiden benachbarten Referenzansichten unterschiedliche örtliche Auflösungen haben, ist der ent-

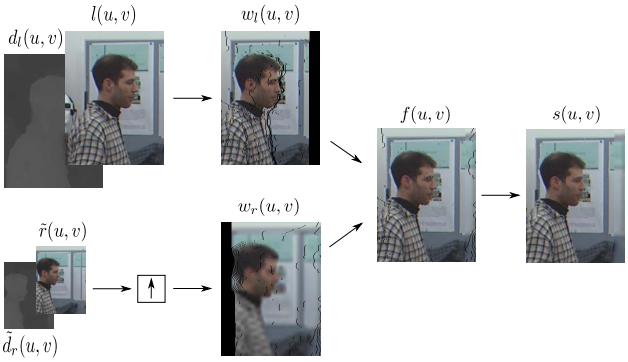

Bild 2 Prinzipieller Ablauf der Zwischenbildsynthese, am Beispiel eines Ausschnitts der *book-arrival* Testsequenz, für Referenzkameras mit **unterschiedlichen** örtlichen Auflösungen

sprechende Syntheseablauf in Bild 2 dargestellt. Wieder besteht der Algorithmus prinzipiell aus drei Teilschritten. Im dargestellten Szenario wird die Szene mit einer hoch aufgelösten Kamera von links und einer niedrig aufgelösten Kamera von rechts aufgenommen. Die niedrige örtliche Auflösung wird dabei mit einer Tilde gekennzeichnet. Anschließend werden wieder beide Ansichten in die Zielperspektive projiziert, zusammengefügt und übrig gebliebene Löcher gefüllt. Um den Einfluss der niedrig aufgelösten Ansicht auf das kombinierte Syntheseergebnis zu minimieren, wird $\alpha = 1$ gesetzt. Die Bildinformation aus der niedrig aufgelösten Ansicht wird also nur für die Bildbereiche verwendet, die durch die hochaufgelöste Nachbaransicht nicht abgebildet werden können. Das Einfügen dieser interpolierten und dadurch verschwommenen Bildbereichen führt allerdings zu stark wahrnehmbaren visuellen Artefakten im gewünschten Zwischenbild. Der Verlust an visueller Bildqualität, der durch den Einfluss der niedrigaufgelösten Referenzansicht verursacht wird, ist in Bild 3 verdeutlicht.

Im folgenden Kapitel wird gezeigt, wie diese verschwommenen Gebiete mit Hilfe des BM3D-Verfahrens an ihre hochaufgelöste Nachbarschaft angepasst und dadurch visuell störende Artefakte deutlich reduziert werden können.

3. BM3D-verfeinerte Kombination der Syntheseergebnisse

Ausgehend von der gegebenen Problemstellung, niederfrequente Bildbereiche an ihre hochaufgelöste Umgebung anzupassen, wird im Folgenden das BM3D-Verfahren [5] als effektive Nachverarbeitungsmöglichkeit vorgestellt. Das Grundprinzip des BM3D-Algorithmus, der als aktueller Stand der Technik zur Rauschunterdrückung angesehen wird, besteht aus zwei Iterationen. Beide Iterationen beinhalten im Wesentlichen die Schritte Gruppierung, 3D-Transformation, Filterung und Aggregation. Das jeweilige Eingangsbild wird dabei blockweise abgearbeitet. Der erste Anlauf basiert auf einem Schwellwertverfahren im Frequenzbereich und liefert eine Vorschätzung, die als Initialisierung für den zweiten Anlauf verwendet wird. Die zweite Iteration berechnet mit einer Wiener-Filterung das

Bild 3 Visueller Vergleich eines synthetisierten Ausschnitts der *alt-moabit* Testsequenz für Referenzbilder identischer (links) und unterschiedlicher örtlicher Auflösung (rechts)

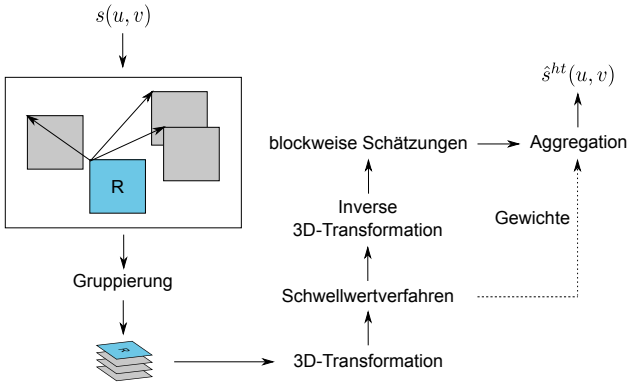

Bild 4 Blockdiagramm der Verarbeitungsschritte in der ersten Iteration des BM3D-Algorithmus

finale entrauschte Bild. Beide Iterationen, sowie die Verwendung des Verfahrens im vorliegenden Anwendungsfall werden im Folgenden beschrieben.

Die einzelnen Verarbeitungsschritte der ersten Iteration des BM3D-Verfahrens werden in Bild 4 visualisiert. Als Eingangsbild erhält der Algorithmus das Ausgabebild der Zwischenbildsynthese $s(u,v)$, bestehend aus nieder- und hochfrequenten Anteilen. Der aktuell verarbeitete Block wird als Referenzblock bezeichnet und ist in Bild 4 gesondert mit dem Buchstaben „R“ gekennzeichnet. Für diesen aktuellen Block werden ähnliche Blöcke im Bild $s(u,v)$ gesucht und inklusive dem Referenzblock zu einem 3D-Würfel gruppiert. Ein Block wird dabei als ähnlich eingestuft, falls sein Abstand zum Referenzblock kleiner als ein vordefinierter Schwellwert τ^{ht} ist. Die durch die Ähnlichkeitssuche ermittelten Blöcke werden nach ihrer Ähnlichkeit absteigend in einem Würfel $\mathbf{S}^{ht}$ sortiert und mit Hilfe einer 3D-Transformation in den Frequenzbereich gebracht. Die Frequenzkoeffizienten werden anschließend einem Schwellwertverfahren unterzogen. Folglich werden alle Koeffizienten, die kleiner als ein Schwellwert λ^{ht} sind, im Frequenzbereich zu Null gesetzt. Dies führt im vorliegenden Anwendungsfall zu einer groben Glättung der Übergänge zwischen niedrig- und hochaufgelösten Bildbereichen. Formal gesehen lässt sich der zurücktransformierte Würfel $\hat{\mathbf{S}}^{ht}$ der initialen Schätzung ausdrücken als:

$$\hat{\mathbf{S}}^{ht} = T_{3D}^{-1}(\Upsilon(T_{3D}((\mathbf{S}^{ht})))) , \qquad (5)$$

wobei $\mathbf{S}^{ht}$ den gruppierten Würfel im Ortsbereich mit

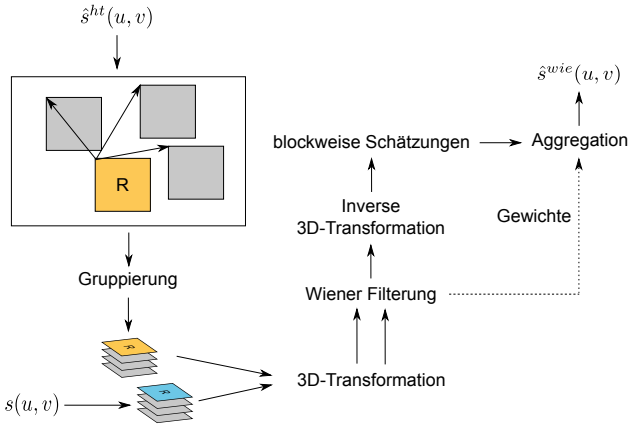

Bild 5 Blockdiagramm der Verarbeitungsschritte in der zweiten Iteration des BM3D-Algorithmus

Blöcken aus dem Syntheseergebnis $s(u,v)$ beschreibt, Y ein Schwellwertoperator ist und die 3D-Transformation durch T_{3D} dargestellt wird. Nach der Rücktransformation werden die geschätzten Blöcke im Würfel $\hat{\mathbf{S}}^{ht}$ wieder an ihre ursprünglichen Bildpositionen geschrieben. Weil sich die einzelnen Blöcke eines Würfels typischerweise überlappen, können durch die inverse Transformation mehrere Schätzwerte für eine Pixelposition generiert werden. Diese multiplen Schätzungen werden zum Abschluss der ersten Iteration des BM3D-Verfahrens zu einem Schätzergebnis kombiniert. Der Aggregationsschritt bildet dazu ein gewichtetes Mittel an allen Pixelpositionen, an denen es auf Grund von Überlappungen der Ähnlichkeitsblöcke zu mehreren Schätzergebnissen kam. Die Gewichte sind dabei invers proportional zur Anzahl der Frequenzkoeffizienten, die durch die Schwellwertbildung nicht zu Null gesetzt wurden. Diese Art der Gewichtsdefinition basiert auf der Idee, Bildpunkten aus stärker verrauschten Blöcken weniger Gewicht zu geben. Das Ausgabebild des ersten Schritts wird mit $\hat{s}^{ht}(u,v)$ bezeichnet.

Das Blockdiagramm der zweiten Iteration des BM3D-Verfahrens ist in Bild 5 dargestellt. In diesem Anlauf wird die Suche nach ähnlichen Blöcken nicht auf dem Ursprungsbild $s(u,v)$, sondern auf der initialen Schätzung $\hat{s}^{ht}(u,v)$ ausgeführt. Die so gefundenen ähnlichen Blöcke, deren Abstand zum aktuellen Referenzblock kleiner als ein Schwellwert τ^{wie} ist, werden dann sowohl im Syntheseergebnis $s(u,v)$ als auch in der initialen Schätzung $\hat{s}^{ht}(u,v)$ zu je einem 3D-Würfel gruppiert. Die beiden daraus resultierenden Würfel werden mit $\mathbf{S}^{wie}$ und $\hat{\mathbf{S}}^{wie}_{ht}$ bezeichnet. Die anschließende Schätzung des entrauschten Würfels basiert auf der Wiener-Filterung. Die dazu benötigten Filterkoeffizienten lassen sich mit

$$\mathbf{K} = \frac{\left|T_{3D}(\hat{\mathbf{S}}^{ht}_{wie})\right|^2}{\left|T_{3D}(\hat{\mathbf{S}}^{ht}_{wie})\right|^2 + \sigma^2} \tag{6}$$

berechnen, wobei der Würfel **K** die berechneten Filterkoeffizienten enthält und die Standardabweichung des angenommenen Rauschens durch den Parameter σ gekennzeichnet wird. Mit Hilfe von

$$\hat{\mathbf{S}}^{wie} = T_{3D}^{-1}(\mathbf{K} \cdot T_{3D}(\mathbf{S}^{wie})) \tag{7}$$

wird der final entrauschte Würfel $\hat{\mathbf{S}}^{wie}$ durch Transformation, Filterung und Rücktransformation des Würfels $\mathbf{S}^{wie}$ berechnet. Abschließend werden, wie im ersten Anlauf, die einzelnen Blöcke an ihre Ursprungskoordinaten zurückgeschrieben und bei multiplen Schätzungen entsprechend gewichtet gemittelt. Die zweite Iteration des BM3D-Verfahrens liefert damit das finale Ausgangsbild $\hat{s}^{wie}(u,v)$.

In diesem konkreten Anwendungsfall sind nur die verfeinerten Bildpunkte von Interesse, die ursprünglich aus der niedrig aufgelösten Referenzansicht in das synthetisierte Zwischenbild eingefügt wurden. Deshalb wird abschließend das BM3D-verfeinerte Zwischenbild $s^{BM3D}(u,v)$ nach

$$s^{BM3D}(u,v) = \begin{cases} s(u,v), & falls\ (u,v) \in \overline{H_l} \\ \hat{s}^{wie}(u,v), & falls\ (u,v) \in H_l \end{cases} \tag{8}$$

gebildet, wobei nur Bildpunkte aus dem entrauschten Bild $\hat{s}^{wie}(u,v)$ eingefügt werden, die nicht mit Hilfe der hochaufgelösten Referenzansicht dargestellt werden können.

4. Simulationsergebnisse

In diesem Kapitel wird die Leistungsfähigkeit des vorgestellten BM3D-Nachverarbeitungsschritts untersucht. Der Ansatz wurde jeweils für die ersten 20 Bilder der Multiview Testsequenzen *book-arrival*, *alt_moabit, kendo* und *pantomime* [8]-[10] getestet. Tabelle 1 gibt einen Überblick über die gewählten Kameraansichten für alle verwendeten Testsequenzen. Die Auflösung der niedrig aufgelösten Referenzansicht wurde mit den Faktoren 2, 4 und 6 in jeweils beiden örtlichen Dimensionen verringert. Die Ergebnisse der vorgestellten BM3D-Verfeinerung werden im Folgenden gegen die Syntheseergebnisse für Referenzansichten mit sowohl identischen als auch unterschiedlichen örtlichen Auflösungen verglichen. Im Falle von unterschiedlichen Auflösungen erfolgt der Vergleich gegen nicht verfeinerte Zwischenbilder und Synthesebilder, die mit Hilfe des NLM- Verfahren verfeinert wurden. Die gezeigten Ergebnisse für eine NLM-Nachverarbeitung sind allerdings nicht direkt mit [4] vergleichbar, da für einen fairen Vergleich zwischen BM3D und NLM der Originalalgorithmus nach [3] verwendet wurde. Im Gegensatz dazu verwendet der in [4] vorgestellte NLM-Ansatz zusätzliche Intelligenz, wie beispielsweise eine geringere Gewichtung von Bildpunkten die ursprünglich aus der niedrig aufgelösten Referenzansicht eingefügt wurden.

Für alle gezeigten Simulationsergebnisse wurde eine Standardabweichung von $\sigma = 40$ angenommen. Die Breite des Suchfensters wurde auf 21 gesetzt. Die Blockgröße wurde für den ersten Anlauf auf 8 und für den zweiten Anlauf des BM3D-Ansatzes auf 16 gesetzt.

Tabelle 1 Übersicht über die gewählten linken, mittleren und rechten Kameraansichten für alle verwendeten Testsequenzen

	Linke Ansicht	Zwischenansicht	Rechte Ansicht
book-arrival	7	6	5
alt_moabit	6	5	4
kendo	2	3	4
pantomime	39	40	41

Tabelle 2 Mittlere PSNR Ergebnisse für alle Testsequenzen und Unterabtastfaktoren in dB; **PSNR auf gesamtem synthetisierten Zwischenbild ausgewertet**

	↓	*book-arrival*	*alt_moabit*	*kendo*	*pantomime*
$PSNR_{FR}$	-	32,99	29,82	34,48	33,89
$PSNR_{u}$	2	32,95	29,75	34,43	33,89
$PSNR_{NLM}$	2	33,02	29,79	34,48	33,91
$PSNR_{BM3D}$	2	**33,03**	**29,79**	**34,49**	**33,92**
$PSNR_{u}$	4	32,66	29,45	34,17	33,82
$PSNR_{NLM}$	4	32,83	29,68	34,30	33,86
$PSNR_{BM3D}$	4	**32,90**	**29,75**	**34,38**	**33,90**
$PSNR_{u}$	6	32,39	29,20	33,89	33,74
$PSNR_{NLM}$	6	32,65	29,57	34,09	33,81
$PSNR_{BM3D}$	6	**32,74**	**29,71**	**34,24**	**33,87**

Tabelle 3 Mittlere PSNR Ergebnisse für alle Testsequenzen und Unterabtastfaktoren in dB; **PSNR nur auf Bildpunkten ausgewertet, die nicht durch die hochaufgelöste Ansicht abgebildet werden können**

	↓	*book-arrival*	*alt_moabit*	*kendo*	*pantomime*
$PSNR_{FR}$	-	31,82	29,65	32,48	28,63
$PSNR_{u}$	2	31,11	27,35	31,15	29,10
$PSNR_{NLM}$	2	32,31	28,50	32,36	31,17
$PSNR_{BM3D}$	2	**32,52**	**28,57**	**32,70**	**32,56**
$PSNR_{u}$	4	28,24	22,92	27,12	25,51
$PSNR_{NLM}$	4	29,78	25,86	28,62	27,44
$PSNR_{BM3D}$	4	**30,52**	**27,56**	**29,91**	**30,16**
$PSNR_{u}$	6	26,54	20,90	24,80	23,17
$PSNR_{NLM}$	6	28,17	24,17	26,31	25,12
$PSNR_{BM3D}$	6	**29,00**	**26,59**	**27,88**	**28,31**

Bild 6 Visueller Vergleich eines unverfeinerten (oben), NLM-verfeinerten (Mitte) und BM3D-verfeinerten (unten) Zwischenbildes am Beispiel eines Ausschnitts der *alt_moabit* Testsequenz und einem Unterabtastfaktor von 6

Tabelle 2 zeigt den PSNR-Vergleich für alle untersuchten Testsequenzen und verschiedenen Unterabtastfaktoren. Dabei beschreibt $PSNR_{FR}$ das resultierende PSNR für identisch hochaufgelöste Referenzbilder, $PSNR_{u}$ das PSNR für die unverfeinerte Synthese aus unterschiedlich aufgelösten Nachbaransichten und $PSNR_{NLM}$ beziehungsweise $PSNR_{BM3D}$ die PSNR-Ergebnisse für synthetisierte Zwischenbilder mit NLM- oder BM3D-Verfeinerung. Die Werte sind jeweils über alle 20 betrachteten Bilder gemittelt. Verglichen mit den unverfeinerten Zwischenbildern liefert die BM3D-Verfeinerung einen maximalen Gewinn von 0,51 dB für die Testsequenz *alt_moabit* und einen Unterabtastfaktor von 6. Hierbei ergibt die unverfeinerte Zwischenbildsynthese einen PSNR-Wert von 29,20 dB während die vorgestellte BM3D-Verfeinerung einen PSNR-Wert von 29,71 dB erreicht. Gemittelt über alle Testsequenzen, erzielt die BM3D-Verfeinerung für diesen Unterabtastfaktor einen mittleren Gewinn von 0,34 dB gegenüber unverfeinerten Zwischenbildern. Der mittlere Verlust gegenüber synthetisierten Bildern bei identisch hochaufgelösten Nachbarbildern liegt bei 0,16 dB. Auch gegenüber der NLM-Verfeinerung liefert der BM3D-Ansatz bessere Ergebnisse für alle betrachteten Simulationen. Für größer werdende Unterabtastfaktoren sinken die PSNR-Werte der

verfeinerten Zwischenbilder deutlich langsamer als die der nicht verfeinerten Syntheseergebnisse. Deshalb führt der Verfeinerungsschritt für größere Unterabtastfaktoren auch zu größeren Gewinnen gegenüber den unverfeinerten Zwischenbildern.

Weil im Normalfall nur ein geringer Anteil an Bildpunkten verfeinert werden muss, sind die PSNR-Gewinne natürlich deutlich höher, wenn die Qualität nur auf den Bildbereichen ausgewertet wird, die nicht mit Hilfe der hochaufgelösten Nachbaransicht abgebildet werden können. Für diesen Fall sind die entsprechenden PSNR-Werte in Tabelle 3 aufgelistet. Hierfür erreicht die BM3D-Verfeinerung einen maximalen Gewinn von 5,69 dB für die *alt_moabit* Testsequenz und eine Reduktion der Auflösung um Faktor 6, verglichen mit unverfeinerten Synthesebildern. Der maximale Gewinn von 3,19 dB gegenüber der NLM-Verfeinerung wird für die *pantomime* Sequenz erreicht. Bild 6 zeigt abschließend einen visuellen Vergleich zwischen unverfeinerten und nachverarbeiteten Zwischenbildern. Im unverfeinerten Syntheseergebnis sind stark wahrnehmbare Artefakte zu sehen (oben). Diese Artefakte können mit Hilfe des NLM-Ansatzes (Mitte) reduziert aber nicht komplett entfernt werden. Der BM3D-Ansatz liefert sowohl gegenüber dem unverfeinerten Zwischenbild als auch gegenüber der NLM-Nachverarbeitung eine deutlich bessere Bildqualität (unten).

5. Schlussfolgerung

Wird ein Zwischenbild aus zwei Referenzbildern mit unterschiedlichen örtlichen Auflösungen generiert, so wird die Qualität des Syntheseergebnisses stark von der niedrig aufgelösten Nachbaransicht beeinflusst. Das BM3D-Verfahren, das ursprünglich zur Entrauschung von Bildern entwickelt wurde, bildet einen effektiven Ansatz um niedrig aufgelöste Bildbereiche an ihre hochaufgelöste Umgebung anzupassen. Die Simulationsergebnisse haben gezeigt, dass durch die BM3D-Nachverarbeitung ein PSNR-Gewinn von bis zu 0,51 dB bezogen auf das gesamte Bild und 5,69 dB bezogen auf die Verfeinerungspositionen möglich ist. Auch die visuelle Qualität lässt sich durch die vorgeschlagene Nachverarbeitung deutlich erhöhen.

6. Literatur

[1] C. Fehn: „Depth-Image-Based Rendering (DIBR), Compression and Transmission for a new Approach on 3D-TV“, in Proc. SPIE Electronic Imaging - Stereoscopic Displays and Virtual Reality Systems XI, vol. 5291, 93-104, 2004

[2] S. Lee, S. Lee, J. Lee, H. Wey: „View Synthesis for Mixed Resolution Multiview 3D Videos“, in Proc. 3DTV-Conference: The True Vision – Capture, Transmission and Display of 3D Video, 1-4, 2011

[3] A. Buades, J.-M. Morel: „A Non-Local Algorithm for Image Denoising“, in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 60-65, 2005

[4] T. Richter, M. Schöberl, J. Seiler, T. Tröger, A. Kaup: „Mixed-Resolution View Synthesis Using Non-Local Means Refined Image Merging“, Proc. SPIE Electronic Imaging – Stereoscopic Displays and Applications XXXIII, vol. 8288, 2012

[5] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian: „Image Denoising by Sparse 3D Transform-domain Collaborative Filtering“, IEEE Trans. Image Proc., vol. 16, no. 8, 2080-2095, 2007

[6] D. Tian, P.-L. Lai, P. Lopez, C. Gomila: „View Synthesis Techniques for 3D Video“, in Proc. SPIE Electronic Imaging – Applications of Digital Image Processing XXXII, vol. 7443, 2009

[7] K.-J. Oh, S. Yea, Y.-S. Ho: „Hole-Filling Method Using Depth Based In-Painting for View Synthesis in free Viewpoint Television (FTV) and 3D Video“, in Proc. Picture Coding Symposium, 233-236, 2009

[8] I. Feldmann, M. Müller, F. Zilly, R. Tanger, K. Müller, A. Smolic, P. Kauff, T. Wiegand: „HHI Test Material for 3D Video“, ISO/IEC JTC1/SC29/WG11, Document M15413, Archamps, France, 2008

[9] M. Tanimoto, T. Fujii, M. Tehrani, M. Wildeboer, N. Fukushima, H. Furihata: „Moving Multiview Camera Test Sequences for MPEG-FTV“, ISO/IEC JTC1/SC29/WG11, Document M16922, Xian, China, 2009

[10] M. Tanimoto, T. Fujii, N. Fukushima: „1D Parallel Test Sequences for MPEG-FTV“, ISO/IEC JTC1/SC29/WG11, Document M15378, Archamps, 2008