



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 103 40 407 A1** 2005.04.07

(12)

Offenlegungsschrift

(21) Aktenzeichen: **103 40 407.4**

(22) Anmeldetag: **02.09.2003**

(43) Offenlegungstag: **07.04.2005**

(51) Int Cl.7: **G06T 9/00**

(71) Anmelder:
**Fraunhofer-Gesellschaft zur Förderung der
angewandten Forschung e.V., 80686 München, DE**

(74) Vertreter:
**Schoppe, Zimmermann, Stöckeler & Zinkler, 82049
Pullach**

(72) Erfinder:
**Marpe, Detlev, 12161 Berlin, DE; Schwarz, Heiko,
13187 Berlin, DE; Wiegand, Thomas, 10789 Berlin,
DE**

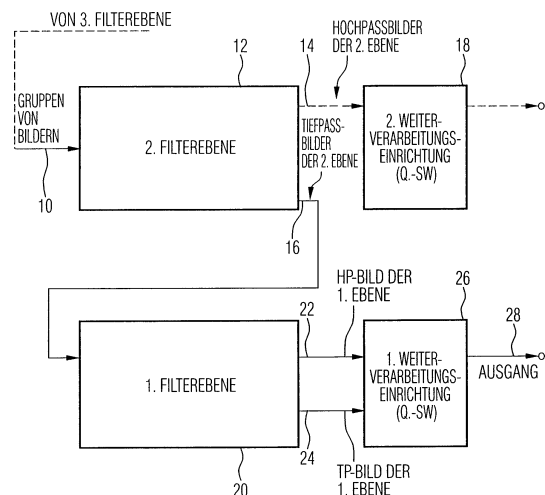
(56) Für die Beurteilung der Patentfähigkeit in Betracht
gezogene Druckschriften:
**Taubmann, D. (July 2003) "Successive refinement
of
video: fundamental issues, past efforts and new
directions". International Symposium on Visual
Communications and Image Processing
(VCIP2003),
SPIE volume 5150, pp.791-805;
Secker, A. and Taubmann, D. (October 2001)
"Motion-
compensated highly scalable video compression
using
an adaptive 3D wavelet transform based on lifting",
Proceedings of the IEEE International Conference
on Image Processing (ICIP) volume 2, pages 1092-
1032;**

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Prüfungsantrag gemäß § 44 PatG ist gestellt.

(54) Bezeichnung: **Vorrichtung und Verfahren zum Codieren einer Gruppe von aufeinanderfolgenden Bildern und
Vorrichtung und Verfahren zum Decodieren eines codierten Bildsignals**

(57) Zusammenfassung: Zum Codieren einer Gruppe von aufeinanderfolgenden Bildern wird eine Filterbank-Zerlegung verwendet, die eine zweite Filterebene zum Erzeugen eines ersten Hochpassbildes der zweiten Ebene und eines ersten Tiefpassbildes der zweiten Ebene aus dem ersten und dem zweiten Bild der Gruppe von Bildern und zum Erzeugen eines zweiten Hochpassbildes der zweiten Ebene und eines zweiten Tiefpassbildes der zweiten Ebene aus dem dritten und dem vierten Bild der ursprünglichen Gruppe von Bildern aufweist. Der Codierer umfasst ferner eine erste Filterebene zum Erzeugen eines ersten Hochpassbildes der ersten Ebene und eines ersten Tiefpassbildes der ersten Ebene aus dem ersten und dem zweiten Tiefpassbild der zweiten Ebene sowie eine Weiterverarbeitungseinrichtung zum Weiterverarbeiten des ersten Hochpassbildes der ersten Ebene und des ersten Tiefpassbildes der ersten Ebene, um ein codiertes Bildsignal zu erhalten, wobei die erste Weiterverarbeitungseinrichtung einen Quantisierer umfasst, der eine Quantisierer-Schrittweite hat. Das codierte Bildsignal stellt eine Basis-Skalierungsschicht im Hinblick auf eine zeitliche Skalierung dar. Durch die gruppenweise Bildverarbeitung und Zerlegung in Hochpass- und Tiefpasssignale wird ferner eine SNR-Skalierbarkeitsimplementierung ermöglicht.



Beschreibung

[0001] Die vorliegende Erfindung bezieht sich auf Videocodier/Decodieralgorithmen und insbesondere auf Videocodier/Decodieralgorithmen, die konform zum internationalen Standard ISO/IEC 14496-10 sind, wobei dieser Standard auch als H.264/AVC bezeichnet wird.

[0002] Der Standard H.264/AVC ist Ergebnis eines Videostandardisierungs-Projekts der ITU-T-Videocodier-Expertengruppe (VCEG) und der ISO/IEC-Bewegbild-Expertengruppe (MPEG). Die Hauptziele dieses Standardisierungsvorhabens bestehen darin, ein übersichtliches Videocodier-Konzept mit sehr gutem Kompressionsverhalten zu schaffen, und gleichzeitig eine Netzwerk-freundliche Videodarstellung zu erzeugen, die sowohl Anwendungen mit "Konversationscharakter", wie beispielsweise die Video-Telefonie, sowie Anwendungen ohne Konversationscharakter (Speichern, Rundfunk, Stream-Übertragung) umfassen.

Stand der Technik

[0003] Es existiert neben dem oben-zitierten Standard ISO/IEC 14496-10 auch eine Vielzahl von Publikationen die sich auf den Standard beziehen. Lediglich beispielhaft wird auf "The Emerging H.264-AVC Standard", Ralf Schäfer, Thomas Wiegand und Heiko Schwarz, EBU Technical Review, Januar 2003, verwiesen. Zudem umfasst die Fachveröffentlichung "Overview of the H.264/AVC Video Coding Standard", Thomas Wiegand, Gary J. Sullivan, Gesle Bjontegaard und Ajay Lohtra, IEEE Transactions on Circuits and Systems for Video Technology, July 2003, sowie die Fachveröffentlichung "Context-based adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard", Detlev Marpe, Heiko Schwarz und Thomas Wiegand, IEEE Transactions on Circuits and Systems for Video Technology, September 2003, einen detaillierten Überblick verschiedener Aspekte des Video-Codier-Standards.

[0004] Zum besseren Verständnis wird jedoch nachfolgend anhand der **Fig. 9** bis **11** ein Überblick über den Videocodierer/Decodierer-Algorithmus gegeben.

[0005] **Fig. 9** zeigt einen kompletten Aufbau eines Videocodierers, welcher allgemein aus zwei unterschiedlichen Stufen besteht. Allgemein gesagt, erzeugt die erste Stufe, die prinzipiell Video-bezogen arbeitet, Ausgangsdaten, die schließlich von einer zweiten Stufe, die in **Fig. 9** mit **80** bezeichnet ist, einer Entropy-Codierung (Entropy Coding) unterzogen werden. Die Daten sind Daten **81a**, quantisierte Transformations-Koeffizienten **81b** sowie Bewegungsdaten **81c**, wobei diese Daten **81a**, **81b**, **81c** dem Entropie-Codierer **80** zugeführt werden, um ein codiertes Videosignal am Ausgang des Entropie-Codierers **80** zu erzeugen.

[0006] Im Einzelnen wird das Eingangsvideosignal (Input Video Signal) in Makroblöcke aufgeteilt, bzw. gesplittet, wobei jeder Makroblock 16×16 Pixel hat. Dann wird die Zuordnung der Makroblöcke zu Slice-Gruppen und Slices (slice = Scheibe) ausgewählt, wonach jeder Makroblock jedes Slices durch das Netz von Betriebsblöcken, wie sie in **Fig. 8** gezeigt sind, verarbeitet wird. Es sei darauf hingewiesen, dass eine effiziente Parallel-Verarbeitung von Makroblöcken möglich ist, wenn verschiedene Slices in einem Videobild sind. Die Zuordnung der Makroblöcke zu Slice-Gruppen und Slices wird mittels einer Blocks Codiersteuerung (Coder Control) **82** in **Fig. 8** durchgeführt. Es existieren verschiedene Slices, die folgendermaßen definiert sind:

I-Slice: Der I-Slice ist ein Slice, in dem alle Makroblöcke des Slices unter Verwendung einer Intra-Prediction codiert werden.

P-Slice: Zusätzlich zu dem Codiertypen des I-Slices können bestimmte Makroblöcke des P-Slices ebenfalls unter Verwendung einer Inter-Prädiktion mit zumindest einem Bewegungs-Kompensierten-Prädiktionssignal (Motion Compensated Prediction Signal) pro Prädiktionsblock codiert werden.

B-Slice: Zusätzlich zu den Codiertypen, die im P-Slice verfügbar sind, können bestimmte Makroblöcke des B-Slices ebenfalls unter Verwendung einer Inter-Prädiktion mit zwei Bewegungs-Kompensierten-Prädiktions-signalen pro Prädiktionsblock codiert werden.

[0007] Die obigen drei Codiertypen sind sehr ähnlich zu denen in früheren Standards, jedoch mit der Ausnahme der Verwendung von Referenzbildern, wie es nachfolgend beschrieben wird. Die folgenden zwei Codiertypen für Slices sind im Standard H.264/AVC neu:

SP-Slice: Er wird auch als Schalt-P-Slice bezeichnet, der so codiert wird, dass ein effizientes Umschalten zwischen unterschiedlichen vorkodierten Bildern möglich wird.

SI-Slice: Der SI-Slice wird auch als Schalt-I-Slice bezeichnet, der eine genaue Anpassung des Makroblocks in einem SP-Slice für einen direkten beliebigen Zugriff und für Fehler-Wiederherstellungszwecke erlaubt.

[0008] Insgesamt sind Slices eine Sequenz von Makroblöcken, die in der Reihenfolge eines Raster-Scans

verarbeitet werden, wenn nicht eine ebenfalls im Standard definierte Eigenschaft der flexiblen Makroblock-Anordnung FMO (FMO = Flexible Macro Block Ordering) verwendet wird. Ein Bild kann in einen oder mehrere Slices aufgeteilt werden, wie es in **Fig. 11** dargestellt ist. Ein Bild ist daher eine Sammlung von einem oder mehreren Slices. Slices sind in dem Sinn selbständig voneinander, da ihre Syntaxelemente aus dem Bitstrom analysiert (geparst) werden können, wobei die Werte der Abtastwerte in dem Bereich des Bildes, der durch den Slice dargestellt wird, korrekt decodiert werden können, ohne dass Daten von anderen Slices benötigt werden, vorausgesetzt, dass verwendete Referenzbilder sowohl im Codierer als auch im Decodierer identisch sind. Bestimmte Informationen von anderen Slices können jedoch nötig sein, um das Deblocking-Filter über Slice-Grenzen hinweg anzuwenden.

[0009] Die FMO-Eigenschaft modifiziert die Art und Weise, wie Bilder in Slices und Makroblöcke partitioniert werden, indem das Konzept der Slice-Gruppen verwendet wird. Jede Slice-Gruppe ist ein Satz von Makroblöcken, die durch eine Makroblock-zu-Slice-Gruppen-Abbildung definiert ist, die durch den Inhalt eines Bildparametersatzes und durch bestimmte Informationen von Slice-Headern spezifiziert ist. Diese Makroblock-zu-Slice-Gruppen-Abbildung besteht aus einer Slice-Gruppen-Identifikationszahl für jeden Makroblock in dem Bild, wobei spezifiziert wird, zu welcher Slice-Gruppe der zugeordnete Makroblock gehört. Jede Slice-Gruppe kann in eine oder mehrere Slices partitioniert werden, so dass ein Slice eine Sequenz von Makroblöcken innerhalb derselben Slice-Gruppe hat, die in der Reihenfolge einer Rasterabtastung innerhalb des Satzes von Makroblöcken einer speziellen Slice-Gruppe verarbeitet wird.

[0010] Jeder Makroblock kann in einem von mehreren Codiertypen abhängig von dem Slice-Codiertyp übertragen werden. In all den Slice-Codiertypen werden die folgenden Typen einer Intra-Codierung unterstützt, die als $\text{intra}_{-4 \times 4}$ oder $\text{intra}_{-16 \times 16}$ bezeichnet werden, wobei zusätzlich ein Chroma-Prädiktions-Modus und auch ein $\text{I}_{\text{-PCM}}$ Prädiktionsmodus unterstützt werden.

[0011] Der $\text{intra}_{-4 \times 4}$ Modus basiert auf der Prädiktion von jedem 4×4 Chroma-Block separat und ist gut geeignet zum Codieren von Teilen eines Bildes mit herausragenden Details. Der $\text{intra}_{-16 \times 16}$ -Modus führt auf der anderen Seite eine Prädiktion des gesamten 16×16 -Chroma-Blocks durch und ist mehr geeignet zum Codieren von "weichen" Bereichen eines Bildes.

[0012] Zusätzlich zu diesen zwei Chroma-Prädiktions-Typen wird eine getrennte Chroma-Prädiktion durchgeführt. Als Alternative für $\text{intra}_{-4 \times 4}$ und $\text{intra}_{-16 \times 16}$ erlaubt der $\text{I}_{-4 \times 4}$ -Codiertyp, dass der Codierer einfach die Prädiktion sowie die Transformationscodierung überspringt und statt dessen die Werte der codierten Abtastwerte direkt überträgt. Der $\text{I}_{\text{-PCM}}$ -Modus dient den folgenden Zwecken: Er ermöglicht es dem Codierer, die Werte der Abtastwerte präzise darzustellen. Er liefert eine Art und Weise, um die Werte von sehr anormalem Bildinhalt ohne Datenvergrößerung genau darzustellen. Er ermöglicht es ferner, für die Anzahl von Bits eine harte Grenze vorzugeben, die ein Codierer für eine Makroblockhandhabung haben muss, ohne dass die Codiereffizienz leidet.

[0013] Im Gegensatz zu früheren Videocodierstandards (nämlich H.263 plus und MPEG-4 Visual), wo die Intra-Prädiktion im Transformationsbereich durchgeführt worden ist, wird die Intra-Prädiktion bei H.264/AVC immer im Raumbereich (Spatial Domain) durchgeführt, und zwar indem auf benachbarte Abtastwerte von vorher codierten Blöcken bezug genommen wird, die links, bzw. oberhalb des zu präzisierenden Blocks liegen (**Fig. 10**). Dies kann in bestimmten Umgebungen, bei denen Übertragungsfehler auftreten, eine Fehlerfortpflanzung mit sich bringen, wobei diese Fehlerfortpflanzung aufgrund der Bewegungskompensation (Motion Compensation) in intercodierten Makroblöcken stattfindet. Daher kann ein begrenzter Intra-Codiermodus signalisiert werden, der eine Prädiktion nur von intra-codierten benachbarten Makroblöcken ermöglicht.

[0014] Wenn der $\text{intra}_{-4 \times 4}$ -Modus verwendet wird, wird jeder 4×4 -Block aus räumlich benachbarten Abtastwerten vorhergesagt. Die 16 Abtastwerte des 4×4 -Blocks werden unter Verwendung von vorher decodierten Abtastwerten in benachbarten Blöcken vorhergesagt. Für jeden 4×4 -Block kann einer von 9 Prädiktionsmodi verwendet werden. Zusätzlich zur "DC-Prädiktion (wo ein Wert verwendet wird, um den gesamten 4×4 -Block vorher zu sagen), werden 8 Richtungs-Prädiktions-Modi spezifiziert. Diese Modi sind geeignet, um Richtungsstrukturen in einem Bild, wie beispielsweise Kanten bei verschiedenen Winkeln vorherzusagen.

[0015] Zusätzlich zu den Intra-Makroblock-Codiertypen werden verschiedene prädiktive oder bewegungskompensierte Codiertypen als P-Makroblocktypen spezifiziert. Jeder P-Makroblock-Typ entspricht einer spezifischen Aufteilung des Makroblocks in die Blockformen, die für eine bewegungskompensierte Prädiktion verwendet werden. Aufteilungen mit Luma-Blockgrößen von 16×16 , 16×8 , 8×16 und 8×8 Abtastwerten werden durch die Syntax unterstützt. Im Falle von Aufteilungen von 8×8 Abtastwerten wird ein zusätzliches Syn-

taxelement für jede 8×8-Aufteilung übertragen. Dieses Syntaxelement spezifiziert, ob die entsprechende 8×8-Aufteilung ferner in Aufteilungen von 8×4, 4×8 oder 4×4 Luma-Abtastwerten und entsprechenden Chroma-Abtastwerten weiter partitioniert wird.

[0016] Das Prädiktionssignal für jeden prädiktiv-codierten M×M-Lumablock wird erhalten, indem ein Bereich des entsprechenden Referenzbildes, der durch einen Translations-Bewegungsvektor und einen Bildreferenz-index spezifiziert ist, verschoben wird. Wenn somit ein Makroblock unter Verwendung von vier 8×8-Aufteilungen codiert wird, und wenn jede 8×8-Aufteilung ferner in vier 4×4-Aufteilungen aufgeteilt wird, kann eine maximale Menge an 16 Bewegungsvektoren für einen einzigen P-Makroblock im Rahmen des sogenannten Bewegungsfelds bzw. Motion Field übertragen werden.

[0017] Der Quantisierungsparameter-Slice QP wird verwendet, um die Quantisierung der Transformationskoeffizienten bei H.264/AVC festzulegen. Der Parameter kann 52 Werte annehmen. Diese Werte sind so angeordnet, dass eine Zunahme von 1 im Hinblick auf den Quantisierungsparameter eine Erhöhung der Quantisierung schrittweise um etwa 12 % bedeutet. Dies bedeutet, dass eine Erhöhung des Quantisierungsparameters um 6 eine Zunahme der Quantisierer-Schrittweite um genau einen Faktor von 2 mit sich bringt. Es sei darauf hingewiesen, dass eine Änderung der Schrittgröße um etwa 12 % ebenfalls in etwa eine Reduktion der Bitrate um etwa 12 % bedeutet.

[0018] Die quantisierten Transformationskoeffizienten eines Blocks werden allgemein in einem Zick-Zack-Weg abgetastet und unter Verwendung von Entropie-Codierverfahren weiter verarbeitet. Die 2×2-DC-Koeffizienten der Chroma-Komponente werden in Raster-Scan-Reihenfolge abgetastet und alle Invers-Transformations-Operationen innerhalb H.264/AVC können unter Verwendung von nur Additionen und Shift-Operationen von 16-Bit-Ganzzahlwerten implementiert werden. Auf ähnliche Art und Weise werden nur 16-Bit-Speicherzugriffe für eine gute Implementierung der Vorwärtstransformationen und des Quantisierverfahrens im Codierer benötigt.

[0019] Bezugnehmend auf **Fig. 9** wird das Eingangsbild zunächst Bild für Bild in einer Videosequenz, jeweils für jedes Bild gesehen, in die Makroblöcke mit 16×16 Pixeln aufteilt. Hierauf wird jedes Bild einem Subtrahierer **84** zugeführt, der das ursprüngliche Bild von einem Bild subtrahiert, das von einem Decodierer **85** geliefert wird, der im Encodierer enthalten ist. Das Subtraktionsergebnis, also die Restsignale im Raum-Bereich (Spatial Domain), werden nunmehr transformiert, skaliert und quantisiert (Block **86**), um die quantisierten Transformationskoeffizienten auf der Leitung **81b** zu erhalten. Zur Erzeugung des Subtraktionssignals, das in den Subtrahierer **84** eingespeist wird, werden die quantisierten Transformationskoeffizienten zunächst wieder skaliert und invers transformiert (Block **87**), um einem Addierer **88** zugeführt zu werden, dessen Ausgang das Deblocking-Filter **89** speist, wobei am Ausgang des Deblocking-Filters das Ausgangsvideosignal, wie es z.B. ein Decodierer decodieren wird, z.B. zu Kontrollzwecken überwacht werden kann (Ausgang **90**).

[0020] Unter Verwendung des decodierten Ausgangssignals am Ausgang **90** wird dann eine Bewegungsschätzung (Motion Estimation) in einem Block **91** durchgeführt. Zur Bewegungsschätzung in Block **91** wird, wie es aus **Fig. 9** ersichtlich ist, ein Bild des ursprünglichen Input-Videosignals zugeführt. Der Standard erlaubt zwei verschiedene Bewegungsschätzungen, nämlich eine Vorwärts-Bewegungs-Schätzung und eine Rückwärts-Bewegungs-Schätzung. Bei der Vorwärts-Bewegungs-Schätzung wird die Bewegung des aktuellen Bilds im Hinblick auf das vorhergehende Bild abgeschätzt. Dagegen wird bei der Rückwärts-Bewegungs-Schätzung die Bewegung des vergangenen Bilds unter Verwendung des aktuellen Bilds abgeschätzt. Die Ergebnisse der Bewegungsschätzung (Block **91**) werden einem Bewegungskompensations-Block (Motion Compensation) **92** zugeführt, der insbesondere dann, wenn ein Schalter **93** auf den Inter-Prädiktions-Modus geschaltet ist, wie es in **Fig. 9** der Fall ist, eine bewegungs-kompensierte Inter-Prädiktion durchführt. Steht der Schalter **93** dagegen auf Intra-Frame-Prädiktion, so wird eine Intra-Frame-Prädiktion unter Verwendung eines Blocks **490** durchgeführt. Hierzu werden die Bewegungsdaten nicht benötigt, da für eine Intra-Frame-Prädiktion keine Bewegungskompensation ausgeführt wird.

[0021] Der Bewegungsschätzungsblock **91** erzeugt Bewegungsdaten bzw. Bewegungsfelder, wobei Bewegungsdaten bzw. Bewegungsfelder, die aus Bewegungsvektoren (Motion Vectors) bestehen, vom Decodierer übertragen werden, damit eine entsprechende inverse Prädiktion, also Rekonstruktion unter Verwendung der Transformationskoeffizienten und der Bewegungsdaten durchgeführt werden kann. Es sei darauf hingewiesen, dass im Falle einer Vorwärts-Prädiktion der Bewegungsvektor aus dem unmittelbar vorhergehenden Bild bzw. auch aus mehreren vorhergehenden Bildern berechnet werden kann. Darüber hinaus sei darauf hingewiesen, dass im Falle einer Rückwärts-Prädiktion ein aktuelles Bild unter Verwendung des unmittelbar angrenzenden zukünftigen Bildes und natürlich auch unter Verwendung von weiteren zukünftigen Bildern berechnet werden

kann.

[0022] Nachteilig an dem in **Fig. 9** dargestellten Videocodierer-Konzept ist dass es keine einfache Skalierbarkeitsmöglichkeit bietet. Wie es in der Technik bekannt ist, versteht man unter dem Ausdruck "Skalierbarkeit" ein Codierer/Decodier-Konzept, bei dem der Codierer einen skalierten Datenstrom liefert. Der skalierte Datenstrom umfasst eine Basis-Skalierungsschicht sowie eine oder mehrere Erweiterungs-Skalierungsschichten. Die Basis-Skalierungsschicht umfasst eine Darstellung des zu codierenden Signals allgemein gesagt mit geringerer Qualität, jedoch auch mit geringerer Datenrate. Die Erweiterungs-Skalierungsschicht enthält eine weitere Darstellung des Videosignals, die typischerweise zusammen mit der Darstellung des Videosignals in der Basis-Skalierungsschicht eine Darstellung mit verbesserter Qualität im Hinblick auf die Basis-Skalierungsschicht liefert. Dagegen hat die Erweiterungs-Skalierungsschicht selbstverständlich einen eigenen Bitbedarf, so dass die Anzahl der Bits zur Darstellung des zu codierenden Signals mit jeder Erweiterungsschicht zunimmt.

[0023] Ein Decodierer wird je nach Ausgestaltung bzw. nach Möglichkeit entweder nur die Basis-Skalierungsschicht decodieren, um eine vergleichsweise qualitativ schlechte Darstellung des durch das codierte Signal dargestellten Bildsignals zu liefern. Mit jeder "Hinzunahme" einer weiteren Skalierungsschicht kann der Decodierer jedoch schrittweise die Qualität des Signals (zu Lasten der Bitrate und der Verzögerung) verbessern.

[0024] Je nach Implementierung und nach Übertragungskanal von einem Codierer zu einem Decodierer wird immer wenigstens die Basis-Skalierungsschicht übertragen, da die Bitrate der Basis-Skalierungsschicht typischerweise so gering ist, dass auch ein bisher begrenzter Übertragungskanal ausreichend sein wird. Erlaubt der Übertragungskanal nicht mehr Bandbreite für die Anwendung, so wird nur die Basis-Skalierungsschicht, nicht aber eine Erweiterungs-Skalierungsschicht übertragen. Dies hat zur Folge, dass der Decodierer lediglich eine niederqualitative Darstellung des Bildsignals erzeugen kann. Im Vergleich zum unskalierten Fall, bei dem die Datenrate so hoch gewesen wäre, dass eine Übertragung das Übertragungssystem überhaupt nicht möglich gewesen wäre, ist die niederqualitative Darstellung von Vorteil. Erlaubt der Übertragungskanal die Übertragung von einer oder mehrerer Erweiterungsschichten, so wird der Codierer auch eine oder mehrere Erweiterungsschichten zum Decodierer übertragen, so dass dieser je nach Anforderung schrittweise die Qualität des ausgegebenen Videosignals erhöhen kann.

[0025] Im Hinblick auf die Codierung von Videosequenzen kann man zwei unterschiedliche Skalierungen unterscheiden. Die eine Skalierung ist die zeitliche Skalierung, dahingehend, dass z.B. nicht alle Video-Einzelbilder einer Videosequenz übertragen werden, sondern dass – zur Reduzierung der Datenrate – beispielsweise nur jedes zweite Bild, jedes dritte Bild, jedes vierte Bild, etc. übertragen wird.

[0026] Die andere Skalierung ist die SNR-Skalierbarkeit (SNR = Signal to Noise Ratio), bei der jede Skalierungsschicht, also sowohl die Basis-Skalierungsschicht als auch die erste, zweite, dritte... Erweiterungs-Skalierungsschicht sämtliche zeitlichen Information umfasst, jedoch mit einer unterschiedlichen Qualität. So hätte die Basis-Skalierungsschicht zwar eine niedrige Datenrate, jedoch ein geringes Signal/Rausch-Verhältnis, wobei dieses Signal/Rausch-Verhältnis dann, mit Hinzunahme jeweils einer Erweiterungs-Skalierungsschicht schrittweise verbessert werden kann.

[0027] Das in **Fig. 9** dargestellte Codierer-Konzept ist dahingehend problematisch, dass es darauf basiert, dass lediglich Restwerte durch den Subtrahierer **84** erzeugt werden, und dann weiter verarbeitet werden. Diese Restwerte werden aufgrund von Prädiktionsalgorithmen berechnet, und zwar in der in

[0028] **Fig. 9** gezeigten Anordnung, die unter Verwendung der Blöcke **86, 87, 88, 89, 92, 93, 94** und **84** eine geschlossene Schleife bildet, wobei in der geschlossenen Schleife ein Quantisierungs-Parameter eingeht, und zwar in den Blöcken **86, 87**. Würde nunmehr eine einfache SNR-Skalierbarkeit dahingehend implementiert werden, dass z.B. jedes prädizierte Restsignal zunächst mit einer groben Quantisierer-Schrittweite quantisiert wird, und dann, schrittweise, unter Verwendung von Erweiterungsschichten mit feineren Quantisierungs-Schrittweiten quantisiert werden würde, so würde dies folgende Konsequenzen haben. Aufgrund der inversen Quantisierung und der Prädiktion insbesondere im Hinblick auf die Bewegungsschätzung (Block **91**) und die Bewegungskompensation (Block **92**), die unter Verwendung des ursprünglichen Bildes einerseits und des quantisierten Bildes andererseits stattfinden, ergibt sich ein "Auseinanderlaufen" der Quantisierungs-Schrittweiten" sowohl im Encoder als auch im Decoder. Dies führt dazu, dass die Erzeugung der Erweiterungs-Skalierungsschichten auf Encoder-Seite sehr problematisch wird. Ferner wird die Verarbeitung der Erweiterungs-Skalierungsschichten auf der Decoderseite zumindest im Hinblick auf die im Standard H.264/AVC definierten Elemente unmöglich. Grund hierfür ist die anhand von **Fig. 9** dargestellte geschlossene

Schleife im Video-Encoder, in der die Quantisierung enthalten ist.

[0029] Das derzeit standardisierte Codierer/Decodierer-Konzept ist somit im Hinblick auf den Skalierbarkeitsgedanken wenig flexibel.

Aufgabenstellung

[0030] Die Aufgabe der vorliegenden Erfindung besteht darin, ein flexibleres Konzept zum Codieren/Decodieren von Bildsignalen zu schaffen.

[0031] Diese Aufgabe wird durch eine Vorrichtung zum Codieren einer Gruppe von aufeinander folgenden Bildern gemäß Patentanspruch 1, eine Vorrichtung zum Decodieren eines codierten Signals gemäß Patentanspruch 17, ein Verfahren zum Codieren einer Gruppe von aufeinander folgenden Bildern gemäß Patentanspruch 20, ein Verfahren zum Decodieren eines Bildsignals gemäß Patentanspruch 21, eine Filtervorrichtung gemäß Patentanspruch 22, eine Invers-Filtervorrichtung gemäß Patentanspruch 23, ein Verfahren zum Filtern gemäß Patentanspruch 24, ein Verfahren zum Invers-Filtern gemäß Patentanspruch 25 oder ein Computer-Programm gemäß Patentanspruch 26 gelöst.

[0032] Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, dass die im Hinblick auf den Skalierbarkeitsgedanken problematische geschlossene Schleife im Video-Encoder dadurch durchbrochen werden kann, dass einerseits von der Einzelbild-weisen Bildverarbeitung weggegangen wird und statt dessen eine Verarbeitung von Bildern gruppenweise durchgeführt wird. Die Gruppe von Bildern wird unter Verwendung eines Lifting-Schemas unter Verwendung von mehreren Filterebenen jeweils in Hochpaßbilder und Tiefpaßbilder zerlegt, um dann nicht mehr Restsignale einzelner Bilder, wie im bekannten Standard, sondern Hochpaßbilder bzw. Tiefpaßbilder zu transformieren, zu skalieren, zu quantisieren und dann einer Entropie-Codierung zuzuführen. Im Videocodierer wird somit gewissermaßen eine Subband-Filterbank vorzugsweise implementiert als Wavelet-Filter vorgeschaltet, wobei die Filterzerlegung die geschlossene Schleife des standardisierten Decodierers durchbricht, so dass eine zeitliche oder SNR-Skalierbarkeit einfach implementierbar ist. Es werden nicht mehr Restsignale in räumlicher Hinsicht weiterverarbeitet, also transformiert, skaliert, quantisiert und dann entropie-codiert, sondern Restsignale, also Hochpass-Signale in zeitlicher Hinsicht, da eine Hochpass-Tiefpass-Filterung durch die verschiedenen Filterebenen in zeitlicher Hinsicht, also über die Gruppe von Bildern betrachtet, durchgeführt wird.

[0033] Die vorliegende Erfindung liefert somit vorzugsweise eine SNR-skalierbare Erweiterung des H.264/AVC-Video-Standard. Zu diesem Zweck, also um eine effiziente SNR-skalierbare Bitstromdarstellung einer Videosequenz zu erreichen, wird die zeitliche Abhängigkeit zwischen Bildern unter Verwendung eines Subband-Lösungsansatzes, der mit offener Schleife, also ohne die problematische geschlossene Schleife auskommt, codiert. In diesem Codierer/Decodierer-Szenario werden die meisten Komponenten von H.264/AVC wie im Standard spezifiziert verwendet, wobei nur wenig Änderungen im Hinblick auf die Subband-Codierer-Struktur benötigt werden.

[0034] Bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung umfasst jede Filterebene einen Rückwärts-Prädiktor einerseits sowie einen Vorwärts-Prädiktor andererseits. Der Rückwärts-Prädiktor führt eine Rückwärts-Bewegungs-Kompensation (Backward Motion Compensation) durch. Der Vorwärts-Prädiktor führt eine Vorwärts-Bewegungskompensation durch. Das Ausgangssignal des Rückwärts-Prädiktors wird von einem Bild der ursprünglichen Gruppe von Bildern subtrahiert, um ein Hochpass-Bild zu erhalten. Das Hochpass-Bild wird dem Vorwärts-Prädiktor als Eingangssignal zugeführt, um ausgangsseitig ein vorwärtsprädiziertes Signal zu erhalten, das dem Bildsignal, das das andere Bild darstellt, hinzu addiert wird, um ein Tiefpass-Bild zu erhalten, das dann, mittels einer niedrigeren Filterebene wieder in ein Hochpass-Signal und ein Tiefpass-Signal zerlegt wird, wobei das Tiefpass-Signal die Gemeinsamkeiten von zwei betrachteten Bildern umfasst, während das Hochpass-Signal die Unterschiede von zwei betrachteten Bildern umfasst.

[0035] Eine Filterebene benötigt zumindest die Verarbeitung einer Gruppe von zwei Bildern. Existieren mehrere Filterebenen, so wird beispielsweise dann, wenn zwei Filterebenen verwendet werden, eine Gruppierung von vier Bildern benötigt. Existieren drei Filterebenen, so wird eine Gruppierung von acht Bildern benötigt. Existieren dagegen vier Filterebenen, so sollten 16 Bilder in eine Gruppe von Bildern gruppiert werden und gemeinsam verarbeitet werden. Es sei darauf hingewiesen, dass die Gruppe von Bildern beliebig groß gewählt werden kann, dass jedoch eine Gruppe mit zumindest zwei Filterebenen zumindest vier Bilder umfassen sollte. Je nach Anwendung werden hohe Gruppengrößen bevorzugt, wobei jedoch dann entsprechend größere Subband-Filterbanken zur Zerlegung auf Codierseite und zur Zusammensetzung auf Decodierseite nötig sind.

[0036] Erfindungsgemäß wird somit vor der eigentlichen Videocodierung gemäß dem Standard H.264/AVC eine zeitliche Subband-Codierung der Videosequenzen vorgenommen, so dass die Quantisierung unter Verwendung einer Quantisierungs-Schrittweite aus der in **Fig. 9** dargestellten geschlossenen Schleife herausgenommen wird, dahingehend, dass nunmehr eine einfache SNR-Skalierbarkeit erreichbar ist. Nunmehr können einzelne Skalierungsschichten ohne weiteres unter Verwendung einzelner unterschiedlicher Quantisierer-Schrittweiten erzeugt werden. Die erfindungsgemäße Zerlegung basiert auf der Lifting-Darstellung einer Filterbank. Diese Lifting-Darstellung der zeitlichen Subband-Zerlegungen erlaubt die Verwendung von bekannten Verfahren zur Bewegungs-kompensierten Prädiktion. Darüber hinaus können die meisten anderen Komponenten eines hybriden Videocodierers, wie beispielsweise H.264/AVC, ohne Modifikation verwendet werden, während nur ein paar wenige Teile verändert werden müssen.

Ausführungsbeispiel

[0037] Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend beziehungsweise auf die beiliegenden Zeichnungen detailliert erörtert. Es zeigen:

[0038] **Fig. 1** ein Blockschaltbild eines erfindungsgemäßen Codierers zum Codieren einer Gruppe von aufeinander folgenden Bildern;

[0039] **Fig. 2** ein Blockschaltbild eines erfindungsgemäßen Decodierers zum Decodieren eines Bildsignals;

[0040] **Fig. 3** ein Blockschaltbild eines erfindungsgemäßen Decodierers gemäß einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung mit vier Ebenen;

[0041] **Fig. 4** ein Blockschaltbild zur Veranschaulichung der Lifting-Zerlegung einer zeitlichen Subband-Filterbank;

[0042] **Fig. 5** ein Blockschaltbild zur Veranschaulichung der Zeit-Skalierung;

[0043] **Fig. 6** ein Blockschaltbild zur Veranschaulichung der SNR-Skalierung;

[0044] **Fig. 7** ein Übersichtsdiagramm zur Veranschaulichung der zeitlichen Verlegung einer Gruppe von beispielsweise 8 Bildern;

[0045] **Fig. 8** eine bevorzugte zeitliche Platzierung von Tiefpass-Bildern für eine Gruppe von 16 Bildern;

[0046] **Fig. 9** ein Übersichts-Blockschaltbild zur Veranschaulichung der grundsätzlichen Codierstruktur für einen Codierer gemäß dem Standard H.264/AVC für einen Makroblock;

[0047] **Fig. 10** eine Kontext-Anordnung, die aus zwei benachbarten Pixelelementen A und B links bzw. oberhalb eines gegenwärtigen Syntaxelements C besteht; und

[0048] **Fig. 11** eine Darstellung der Aufteilung eines Bildes in Slices.

[0049] **Fig. 1** zeigt eine Vorrichtung zum Codieren einer Gruppe von aufeinanderfolgenden Bildern, wobei die Gruppe wenigstens ein erstes, zweites, drittes und viertes Bild aufweist. Die Gruppe von Bildern wird in einen Eingang **10** einer zweiten Filterebene **12** eingespeist. Die zweite Filterebene ist ausgebildet, um aus der Gruppe von Bildern Hochpass-Bilder an einem ersten Ausgang **14** und Tiefpass-Bilder an einem zweiten Ausgang **16** der Filterebene **12** zu erzeugen. Im einzelnen ist die zweite Filterebene **12** ausgebildet, um ein erstes Hochpass-Bild der zweiten Ebene und ein erstes Tiefpass-Bild der zweiten Ebene aus dem ersten und zweiten Bild der Gruppe von Bildern und ein zweites Hochpass-Bild der zweiten Ebene und ein zweites Tiefpass-Bild der zweiten Ebene aus dem dritten und vierten Bild der Gruppe von Bildern zu erzeugen.

[0050] Bei dem beispielhaft beschriebenen Fall, bei dem die Gruppe von Bildern vier Bilder hat, erzeugt die zweite Filterebene **12** somit ausgangsseitig am Hochpass-Ausgang **14** zwei Hochpass-Bilder der zweiten Ebene, und erzeugt ausgangsseitig am Tiefpassausgang **16** zwei Tiefpass-Bilder der zweiten Ebene. Optional werden die Hochpass-Bilder der zweiten Ebene einer zweiten Weiterverarbeitungseinrichtung zugeführt. Die zweite Weiterverarbeitungseinrichtung ist ausgebildet, um das erste Hochpass-Bild der zweiten Ebene sowie das zweite Hochpass-Bild der zweiten Ebene weiterzuverarbeiten, wobei die zweite Weiterverarbeitungseinrichtung **18** einen Quantisierer umfasst, der eine Quantisierungs-Schrittweite (QP) umfasst. Beispielhaft ist die

zweite Weiterverarbeitungseinrichtung **18** ausgebildet, um die Funktionalitäten der Blöcke **86**, **80** von **Fig. 9** durchzuführen. Das Ausgangssignal der zweiten Weiterverarbeitungseinrichtung, also eine quantisierte und gegebenenfalls entropie-codierte Darstellung der Hochpass-Bilder der zweiten Ebene wird im Falle einer SNR-Skalierbarkeit in den ausgangsseitigen Bitstrom geschrieben.

[0051] Im Falle einer Zeit-Skalierbarkeit wird dieses Signal bereits die erste Erweiterungs-Skalierungsschicht dar. Erzeugt ein Encodierer lediglich eine Basis-Skalierungsschicht, so wird keine Weiterverarbeitung der Hochpass-Bilder nötig sein, weshalb die Verbindung zwischen dem Block **12** und dem Block **18** gestrichelt eingezeichnet ist.

[0052] Die beiden Tiefpass-Bilder der zweiten Ebene, die an dem Tiefpass-Ausgang **16** der zweiten Filterebene ausgegeben werden, werden in einen Eingang einer ersten Filterebene **20** eingespeist. Die erste Filterebene **20** ist ausgebildet, um ein erstes Hochpass-Bild der ersten Ebene und ein erstes Tiefpass-Bild der ersten Ebene aus dem ersten und dem zweiten Tiefpass-Bild der zweiten Ebene zu erzeugen. Das erste Hochpass-Bild der ersten Ebene, das durch die Filterebene **20** erzeugt wird, wird an einem Hochpass-Ausgang **22** der ersten Filterebene ausgegeben. Das erste Tiefpass-Bild der ersten Ebene wird an einem Tiefpass-Ausgang **24** der ersten Filterebene ausgegeben.

[0053] Beide Signale werden in einer ersten Weiterverarbeitungseinrichtung zur Weiterverarbeitung des ersten Hochpass-Bildes der ersten Ebene und des ersten Tiefpass-Bildes der ersten Ebene eingespeist, um an einem Ausgang **28** ein codiertes Bildsignal zu erhalten, wobei die erste Weiterverarbeitungseinrichtung **26**, wie es bereits anhand von der zweiten Weiterverarbeitungseinrichtung **18** dargestellt worden ist, einen Quantisierer umfasst, der eine Quantisierer-Schrittweite hat. Vorzugsweise sind die Quantisierer-Schrittweiten der zweiten Weiterverarbeitungseinrichtung **18** und der ersten Weiterverarbeitungseinrichtung **26** identisch.

[0054] Das Ausgangssignal der ersten Weiterverarbeitungseinrichtung **26** umfasst somit das Hochpass-Bild der ersten Ebene und das Tiefpass-Bild der ersten Ebene und stellt somit im Sinne einer zeitlichen Skalierbarkeit die Basis-Skalierungsschicht dar.

[0055] Im Sinne einer SNR-Skalierbarkeit stellt das codierte Bildsignal am Ausgang **28** zusammen mit den kodierten Hochpassbildern der zweiten Ebene am Ausgang der zweiten Weiterverarbeitungseinrichtung **18** das Ausgangssignal dar, das eine Basis-Skalierungsschicht ist.

[0056] **Fig. 2** zeigt eine Vorrichtung zum Dekodieren eines kodierten Signals, und zwar des Signals, das an dem Ausgang **28** von **Fig. 1** als codiertes Bildsignal ausgegeben wird. Wie es ausgeführt worden ist, umfasst das codierte Bildsignal am Ausgang **28** eine quantisierte und entropie-kodierte Darstellung des ersten Hochpass-Bildes und des ersten Tiefpass-Bildes der ersten Ebene. Diese Informationen, also das Signal an dem Ausgang **28** wird in eine Invers-Weiterverarbeitungseinrichtung **30** eingespeist, die unter Verwendung und Kenntnis der von der ersten Weiterverarbeitungseinrichtung **26** von **Fig. 6** benutzten Quantisierungs-Schrittweite eine Invers-Weiterverarbeitung durchführt, wobei diese Invers-Weiterverarbeitung zum Beispiel eine Entropie-Decodierung, eine inverse Quantisierung sowie eine Rücktransformation etc. umfassen wird, wie es in der Technik bekannt ist, und wie es durch einen Codierer bzw. die im Codierer verwendete Weiterverarbeitungseinrichtung vorgegeben ist.

[0057] An einem Ausgang der Invers-Weiterverarbeitungseinrichtung **30** liegt dann eine rekonstruierte Version des ersten Hochpass-Bildes sowie eine rekonstruierte Version des ersten Tiefpass-Bildes vor. Diese beiden rekonstruierten Versionen des ersten Hochpass-Bildes und des ersten Tiefpass-Bildes werden in eine Invers-Filterebene **32** eingespeist, wobei die Invers-Filterebene **32** ausgebildet ist, um die rekonstruierte Version des ersten Hochpass-Bildes und die rekonstruierte Version des ersten Tiefpass-Bildes invers zu filtern, um eine rekonstruierte Version eines ersten Tiefpass-Bildes und eines zweiten Tiefpass-Bildes einer nächsthöheren Ebene, also der zweiten Ebene zu erhalten. Im Falle einer zeitlichen Skalierbarkeit stellt diese rekonstruierte Version des ersten und des zweiten Tiefpass-Bildes der beiden Ebenen die Basisschicht dar. Im Falle einer SNR-Skalierbarkeit wird ferner auch das Ausgangssignal der zweiten Weiterverarbeitungseinrichtung **18** von **Fig. 1** einer Invers-Weiterverarbeitung unter Verwendung der entsprechenden Quantisierungs-Schrittweite unterzogen, um dann, zusammen mit der rekonstruierten Version des ersten und des zweiten Tiefpass-Bildes der zweiten Ebene in eine Invers-Filterebene nächster Ordnung eingespeist zu werden, wie es nachfolgend noch beispielsweise beziehungsweise auf **Fig. 3** näher erörtert wird.

[0058] Wavelet-basierte Videocodier-Algorithmen, bei denen Lifting-Implementierungen für die Wavelet-Analyse und für die Wavelet-Synthese eingesetzt werden, sind in J.-R. Ohm, „Complexity and delay analysis of

MCTF interframe wavelet structures", ISO/IECJTC1/WG11 Doc.M8520, July 2002, beschrieben. Anmerkungen zur Skalierbarkeit finden sich auch in D. Taubman, „Successive refinement of video: fundamental issues, past efforts and new directions", Proc. of SPIE (VCIP' 03), vol. 5150, pp. 649–663, 2003, wobei hierfür jedoch erhebliche Änderungen an Codiererstrukturen nötig sind. Erfindungsgemäß wird dagegen ein Codierer/Decodierer-Konzept erreicht, das einerseits die Skalierbarkeitmöglichkeit hat und das andererseits auf standardkonformen Elementen, insbesondere z.B. für die Bewegungskompensation, aufbauen kann.

[0059] Bevor detailliert auf einen Encoder/Decoder-Aufbau anhand von **Fig. 3** eingegangen wird, sei zunächst anhand von **Fig. 4** ein grundsätzliches Lifting-Schema auf Seiten des Codierers bez. ein Invers-Lifting-Schema auf Seiten des Decodierers dargestellt. Detaillierte Ausführungen zu den Hintergründen der Kombination von Lifting-Schemen und Wavelet-Transformationen finden sich in W. Sweldens, „A customdesign construction of biorthogonal wavelets", J. Appl. Comp. Harm. Anal., vol. 3 (no. 2), pp. 186–200, 1996 und I. Daubechies und W. Sweldens, "Factoring wavelet transforms into lifting steps", J. Fourier Anal. Appl., vol. 4 (no.3), pp. 247–269, 1998. Im allgemeinen besteht das Lifting-Schema aus drei Schritten, dem Polyphasen-Zerlegungsschritt, dem Prädiktions-Schritt und dem Update-Schritt (Aktualisierungsschritt), wie es anhand des Codierers in **Fig. 1a** dargestellt ist. Der Polyphasen-Zerlegungsschritt wird durch einen ersten Bereich I dargestellt, der Prädiktions-Schritt wird durch einen zweiten Bereich II dargestellt, und der Aktualisierungsschritt wird durch einen dritten Bereich III dargestellt.

[0060] Der Zerlegungs-Schritt umfasst eine Aufteilung des eingangsseitigen Datenstroms in eine identische erste Kopie für einen unteren Zweig **40a** sowie eine identische Kopie für einen oberen Zweig **40b**. Ferner wird die identische Kopie des oberen Zweigs **40b** um eine Zeitstufe (z^{-1}) verzögert, so dass ein Abtastwert s_{2k+1} mit einem ungeradzahlgigen Index k zum gleichen Zeitpunkt wie ein Abtastwert mit einem geraden Index s_{2k} durch einen jeweiligen Dezimierer bzw. Downsampler **42a**, **42b** läuft. Der Dezimierer **42a** bzw. **42b** reduziert die Anzahl der Samples im oberen bzw. im unteren Zweig **40b**, **40a** durch Eliminieren jedes jeweils zweiten Abtastwerts.

[0061] Der zweite Bereich II, der sich auf den Prädiktion-Schritt bezieht, umfasst einen Prädiktionoperator **43** sowie einen Subtrahierer **44**. Der dritte Bereich, also der Aktualisierungsschritt umfasst einen Aktualisierungs-Operator **45** sowie einen Addierer **46**. Ausgangsseitig existieren noch zwei Normierer **47**, **48**, zum Normieren des Hochpass-Signals h_k (Normierer **47**) und zum Normieren des Tiefpass-Signals l_k durch den Normierer **48**.

[0062] Im einzelnen führt die Polyphasenzerlegung dazu, dass die geradzahlgigen und ungeradzahlgigen Abtastwerte eines gegebenen Signals $s[k]$ getrennt werden. Da die Korrelationsstruktur typischerweise eine Lokalcharakteristik zeigt, sind die geraden und ungeraden Polyphasenkomponenten hochkorreliert. Daher wird in einem anschließenden Schritt eine Prädiktion (P) der ungeraden Abtastwerte unter Verwendung der geraden Abtastwerte durchgeführt. Der entsprechende Prädiktions-Operator (P) für jeden ungeraden Abtastwert:

$$s_{\text{odd}}[k] = s[2k + 1]$$

ist eine lineare Kombination der benachbarten geraden Abtastwerte:

$$s_{\text{even}}[k] = s[2k]$$

d.h.

$$P(s_{\text{even}})[k] = \sum_l p_l s_{\text{even}}[k + l]$$

[0063] Als Ergebnis des Prädiktions-Schritts werden die ungeradzahlgigen Abtastwerte durch ihre entsprechenden Prädiktions-Restwerte

$$h[k] = s_{\text{odd}}[k] - P(s_{\text{even}})[k]$$

ersetzt. Es sei darauf hingewiesen, dass der Prädiktionsschritt äquivalent zum Durchführen eines Hochpass-Filters einer Zwei-Kanal-Filterbank ist, wie es in I. Daubechies und W. Sweldens, „Factoring wavelet transforms into lifting steps", J. Fourier Anal. Appl. vol 4 (no.3), pp. 247–269, 1998 dargelegt ist.

[0064] Im dritten Schritt des Lifting-Schemas wird eine Tiefpass-Filterung durchgeführt, indem die geraden

Abtastwerte

$s_{\text{even}}[k]$

durch eine lineare Kombination der Prädiktions-Restwerte $h[k]$ ersetzt werden. Der entsprechende Aktualisierungs-Operator U ist gegeben durch

$$U(h)[k] = \sum_l u_l h[k+l]$$

[0065] Durch Ersetzen der geraden Abtastwerte mit

$$l[k] = s_{\text{even}}[k] + U(h)[k]$$

kann das gegebene Signal $s[k]$ schließlich durch $l(k)$ und $h(k)$ dargestellt werden, wobei jedes Signal jedoch die halbe Abtastrate hat. Da sowohl der Aktualisierungs-Schritt als auch der Prädiktions-Schritt vollständig invertierbar sind, kann die entsprechende Transformation als kritisch abgetastete Perfekt-Rekonstruktions-Filterbank interpretiert werden. In der Tat kann gezeigt werden, dass jegliche biorthogonale Familie von FIR-Filtern durch eine Sequenz von einem oder mehreren Prädiktions-Schritten und einem oder mehreren Aktualisierungs-Schritten realisiert werden kann. Für eine Normierung der Tiefpaß- und Hochpass-Komponenten werden, wie es ausgeführt worden ist, die Normierer **47** und **48** mit geeignet gewählten Skalierungsfaktoren F_l und F_h versorgt.

[0066] Das Invers-Lifting-Schema, das der Synthese-Filterbank entspricht, ist in **Fig. 4**, auf der rechten Seite, gezeigt. Es besteht einfach aus der Anwendung des Prädiktions- und Aktualisierungs-Operators in umgekehrter Reihenfolge und mit umgekehrten Vorzeichen, gefolgt von der Rekonstruktion unter Verwendung der geraden und ungeraden Polyphasenkomponenten. Im einzelnen umfasst der in **Fig. 4** rechts gezeigte Decodierer somit wieder einen ersten Decodiererbereich I, einen zweiten Decodiererbereich II sowie einen dritten Decodiererbereich III. Der erste Decodiererbereich macht die Wirkung des Aktualisierungs-Operators **45** rückgängig. Dies geschieht dadurch, dass das durch einen weiteren Normierer 50 zurück-normierte Hochpass-Signal dem Aktualisierungs-Operator **45** zugeführt wird. Das Ausgangssignal des decodiererseitigen Aktualisierungs-Operators **45** wird dann nun, im Gegensatz zum Addierer **46** in **Fig. 4**, einem Subtrahierer **52** zugeführt. Entsprechend wird mit dem Ausgangssignal des Prädiktors **43** vorgegangen, dessen Ausgangssignal nunmehr nicht, wie auf Codiererseite einem Subtrahierer zugeführt wird, sondern dessen Ausgangssignal nunmehr einem Addierer **53** zugeführt wird. Nunmehr findet ein Upsampling des Signals in jedem Zweig um den Faktor **2** statt (Blöcke **54a**, **54b**). Hierauf wird der obere Zweig um einen Abtastwert in die Zukunft geschoben, was äquivalent zum Verzögern des unteren Zweigs ist, um dann eine Addition der Datenströme auf dem oberen Zweig und dem unteren Zweig in einem Addierer **55** durchzuführen, um das rekonstruierte Signal s_k am Ausgang der Synthese-Filterbank zu erhalten.

[0067] Durch den Prädiktor **43** bzw. den Aktualisierer **45** können verschiedene Wavelets implementiert werden. Falls das sogenannte Haar-Wavelet implementiert werden soll, sind der Prädiktions-Operator und der Aktualisierungs-Operator durch folgende Gleichung gegeben

$$P_{\text{Haar}}(s_{\text{even}})[k] = s[2k] \text{ und } U_{\text{Haar}}(h)[k] = \frac{1}{2}h[k],$$

derart, dass

$$h[k] = s[2k+1] - s[2k] \text{ und } l[k] = s[2k] + \frac{1}{2}h[k] = \frac{1}{2}(s[2k] + s[2k+1])$$

dem nicht-normierten Hochpaß- bzw. Tiefpaß-(Analyse-)Ausgangssignal des Haar-Filters entsprechen.

[0068] Im Falle des 5/3-Biorthogonal-Spline-Wavelets haben das Tiefpaß- und das Hochpaß-Analyse-Filter dieses Wavelets **5** bzw. **3** Filtertaps, wobei die entsprechende Skalierungsfunktion ein B-Spline der Ordnung **2** ist. In Codieranwendungen für nicht-bewegte Bilder (Still-Images, wie beispielsweise JPEG 2000) wird dieses Wavelet für ein zeitliches Subband-Codierschema verwendet. In einer Lifting-Umgebung sind die entsprechenden Prädiktions- und Aktualisierungs-Operatoren der 5/3-Transformation folgendermaßen gegeben

$$P_{5/3}(s_{\text{even}})[k] = \frac{1}{2}(s[2k] + s[2k+2]) \text{ und } U_{5/3}(h)[k] = \frac{1}{4}(h[k] + h[k-1]).$$

[0069] **Fig. 3** zeigt ein Blockschaltbild einer erfindungsgemäßen Codierer/Decodierer-Struktur mit beispiel-

haften vier Filterebenen sowohl auf Seiten des Codierers als auch auf Seiten des Decodierers. Aus **Fig. 3** ist zu ersehen, dass die erste Filterebene, die zweite Filterebene, die dritte Filterebene und die vierte Filterebene bezogen auf den Codierer identisch sind. Die Filterebenen bezogen auf den Decodierer sind ebenfalls identisch. Auf Codiererseite umfasst jede Filterebene als zentrale Elemente einen Rückwärts-Prädiktor M_{i0} **60** sowie einen Vorwärts-Prädiktor M_{i1} **61**. Der Rückwärts-Prädiktor **60** entspricht prinzipiell dem Prädiktor **43** von **Fig. 4**, während der Vorwärts-Prädiktor **61** dem Aktualisierer **45** von **Fig. 4** entspricht.

[0070] Im Unterschied zu **Fig. 4** sei darauf hingewiesen, dass sich **Fig. 4** auf einen Strom von Abtastwerten bezieht, bei denen ein Abtastwert einen ungeradzahligem Index $2k+1$ hat, während ein anderer Abtastwert einen geradzahligem Index $2k$ hat. Die Notation in **Fig. 3** bezieht sich jedoch, wie es bereits anhand von **Fig. 1** dargelegt worden ist, auf eine Gruppe von Bildern anstatt auf eine Gruppe von Abtastwerten. Hat ein Bild beispielsweise eine Anzahl von Abtastwerten bzw. Pixeln, so wird dieses Bild insgesamt eingespeist. Dann wird das nächste Bild eingespeist etc. Es existieren somit nicht mehr ungeradzahlige und geradzahlige Abtastwerte, sondern ungeradzahlige und geradzahlige Bilder. Erfindungsgemäß wird das für ungeradzahlige und geradzahlige Abtastwerte beschriebene Lifting-Schema auf ungeradzahlige bzw. geradzahlige Bilder, von denen jedes eine Vielzahl von Abtastwerten hat, angewendet. Aus dem abtastwert-weisen Prädiktor **43** von **Fig. 4** wird nunmehr die Rückwärts-Bewegungskompensations-Prädiktion **60**, während aus dem abtastwert-weisen Aktualisierer **45** die Bildweise Vorwärts-Bewegungskompensations-Prädiktion **61** wird.

[0071] Es sei darauf hingewiesen, dass die Bewegungsfiler, die aus Bewegungsvektoren bestehen, und die Koeffizienten für die Blöcke **60** und **61** darstellen, jeweils für zwei aufeinander bezogene Bilder berechnet und als Seiteninformationen vom Codierer zum Decodierer übertragen werden. Von wesentlichem Vorteil beim erfindungsgemäßen Konzept ist jedoch die Tatsache, dass die Elemente **91**, **92**, wie sie anhand von **Fig. 9** beschrieben sind und im Standard H.264/AVC standardisiert sind, ohne weiteres dazu verwendet werden können, um sowohl die Bewegungsfelder M_{i0} als auch die Bewegungsfelder M_{i1} zu berechnen. Für das erfindungsgemäße Konzept muss daher kein neuer Prädiktor/Aktualisierer eingesetzt werden, sondern es kann der bereits bestehende, untersuchte und auf Funktionalität und Effizienz überprüfte im Videostandard genannte Algorithmus für die Bewegungskompensation in Vorwärtsrichtung oder in Rückwärtsrichtung eingesetzt werden.

[0072] Insbesondere zeigt die in **Fig. 3** dargestellt allgemeine Struktur der verwendeten Filterbank eine zeitliche Zerlegung des Videosignals mit einer Gruppe von 16 Bildern, die an einem Eingang **64** eingespeist werden. Die Zerlegung ist eine dyadische zeitliche Zerlegung des Videosignals, wobei bei dem in **Fig. 3** gezeigten Ausführungsbeispiel mit 4 Ebenen $2^4 = 16$ Bilder, also eine Gruppengröße von 16 Bildern benötigt wird, um auf der Darstellung mit der kleinsten zeitlichen Auflösung, also auf den Signalen am Ausgang **28a** und am Ausgang **28b** anzukommen. Werden daher 16 Bilder gruppiert, so führt dies zu einer Verzögerung von 16 Bildern, was das in **Fig. 3** gezeigte Konzept mit vier Ebenen für interaktive Anwendungen eher problematisch macht. Wird daher auf interaktive Anwendungen abgezielt, so wird es bevorzugt, kleinere Gruppen von Bildern zu bilden, wie beispielsweise vier oder acht Bilder zu gruppieren. Dann wird die Verzögerung entsprechend reduziert, so dass auch der Einsatz für interaktive Anwendungen möglich wird. In Fällen, in denen Interaktivität nicht benötigt wird, beispielsweise zu Speicherzwecken etc. kann die Anzahl der Bilder in einer Gruppe, also die Gruppengröße, entsprechend erhöht werden, beispielsweise auf **32**, **64**, etc. Bilder.

[0073] Erfindungsgemäß wird es bevorzugt, die interaktive Anwendung des Haar-basierten bewegungskompensierten Lifting-Schemas zu verwenden, das aus einer Rückwärts-Bewegungskompensations-Prädiktion (M_{i0}), wie in H.264/AVC besteht, und das ferner einen Aktualisierungsschritt umfasst, der eine Vorwärtsbewegungskompensation (M_{i1}) umfasst. Sowohl der Prädiktions-Schritt als auch der Aktualisierungsschritt verwenden den Bewegungskompensationsprozess, wie er in H.264/AVC dargestellt ist. Ferner wird es bevorzugt, nicht nur die Bewegungskompensation zu verwenden, sondern auch das in **Fig. 9** mit dem Bezugszeichen **89** bezeichnet Deblocking-Filter **89** einzusetzen.

[0074] Die zweite Filterebene umfasst wieder Downsampler **66a**, **66b**, einen Subtrahierer **69**, einen Rückwärts-Prädiktor **67**, einen Vorwärts-Prädiktor **68** sowie einen Addierer **70** und, wie es bereits anhand von **Fig. 1** dargestellt worden ist, die Weiterverarbeitungseinrichtung **18**, um an einem Ausgang der Weiterverarbeitungseinrichtung **18**, wie es bereits anhand von **Fig. 1** dargestellt worden ist, das erste und das zweite Hochpass-Bild der zweiten Ebene auszugeben, während am Ausgang des Addierers **70** das erste und das zweite Tiefpass-Bild der zweiten Ebene ausgegeben werden.

[0075] Der erfindungsgemäße Codierer in **Fig. 3** umfasst zusätzlich eine dritte Ebene sowie eine vierte Ebene, wobei in den Eingang **64** der vierten Ebene eine Gruppe von 16 Bildern eingespeist wird. An einem Hochpass-Ausgang **72** der vierten Ebene, der auch als HP4 bezeichnet ist, werden mit einem Quantisierungs-Pa-

parameter Q quantisierte und entsprechend weiterverarbeitete acht Hochpass-Bilder ausgegeben. Entsprechend werden an einem Tiefpass-Ausgang **73** der vierten Filterebene acht Tiefpass-Bilder ausgegeben, die in einem Eingang **74** der dritten Filterebene eingespeist wird. Die Ebene ist wiederum wirksam, um an einem Hochpass-Ausgang **75**, der auch mit HP3 bezeichnet ist, vier Hochpass-Bilder zu erzeugen und um an einem Tiefpass-Ausgang **76** vier Tiefpass-Bilder zu erzeugen, die in den Eingang **10** der zweiten Filterebene eingespeist werden und zerlegt werden, wie es anhand von **Fig. 3** bzw. anhand von **Fig. 1** dargelegt worden ist.

[0076] Es sei besonders darauf hingewiesen, dass die durch eine Filterebene verarbeitete Gruppe von Bildern nicht unbedingt Videobilder sein müssen, die von einer ursprünglichen Videosequenz stammen, sondern auch Tiefpass-Bilder sein können, die von einer nächsthöheren Filterebene an einem Tiefpass-Ausgang der Filterebene ausgegeben worden sind.

[0077] Ferner sei darauf hingewiesen, dass das in **Fig. 3** gezeigte Codierer-Konzept für sechzehn Bilder ohne weiteres auf acht Bilder reduziert werden kann, wenn einfach die vierte Filterebene weggelassen wird und die Gruppe von Bildern in den Eingang **74** eingespeist wird. Genauso kann das in **Fig. 3** gezeigte Konzept auch ohne weiteres auf eine Gruppe von zweiunddreißig Bildern erweitert werden, indem eine fünfte Filterebene hinzugefügt wird, und indem die dann sechzehn Hochpass-Bilder an einem Hochpass-Ausgang der fünften Filterebene ausgegeben werden und die sechzehn Tiefpass-Bilder am Ausgang der fünften Filterebene in den Eingang **64** der vierten Filterebene eingespeist werden.

[0078] Auf Decodiererseite wird ebenfalls das baumartige Konzept der Codiererseite angewendet, jedoch nun nicht mehr, wie auf Codiererseite von der höheren Ebene zur niedrigeren Ebene, sondern nunmehr, auf der Decodiererseite, von der niedrigeren Ebene zur höheren Ebene. Hierzu wird von einem Übertragungsmedium, das schematisch als Network Abstraction Layer **100** bezeichnet ist, der Datenstrom empfangen und der empfangene Bitstrom wird zunächst einer Invers-Weiterverarbeitung unter Verwendung der Invers-Weiterverarbeitungseinrichtungen **30a**, **30b** unterzogen, um eine rekonstruierte Version des ersten Hochpass-Bildes der ersten Ebene am Ausgang der Einrichtung **30a** und eine rekonstruierte Version des Tiefpass-Bildes der ersten Ebene am Ausgang des Blocks **30b** von **Fig. 3** zu erhalten. Dann wird, in Analogie zur rechten Hälfte von **Fig. 4**, zunächst die Vorwärts-Bewegungskompensations-Prädiktion mittels des Prädiktors **61** rückgängig gemacht, um dann das Ausgangssignal des Prädiktors **61** von der rekonstruierten Version des Tiefpass-Signals zu subtrahieren (Subtrahierer **101**).

[0079] Das Ausgangssignal des Subtrahierers **101** wird in einen Rückwärts-Kompensations-Prädiktor **60** eingespeist, um ein Prädiktionsergebnis zu erzeugen, das in einem Addierer **102** zur rekonstruierten Version des Hochpass-Bildes addiert wird. Hierauf werden beide Signale, also die Signale im unteren Zweig **103a**, **103b** auf die doppelte Abtastrate gebracht, und zwar unter Verwendung der Upsampler **104a**, **104b**, wobei dann das Signal auf dem oberen Zweig je nach Implementierung verzögert bzw. „beschleunigt“ wird. Es sei darauf hingewiesen, dass das Upsampling durch die Brücke **104a**, **104b** einfach durch Einfügen von einer Anzahl von Nullen, die gleich der Anzahl von Samples für ein Bild entspricht, durchgeführt wird. Die Verschiebung um die Verzögerung eines Bildes durch das mit z^{-1} gezeigte Element im oberen Zweig **103b** gegenüber dem unteren Zweig **103a** bewirkt, dass die Addition durch einen Addierer **106** dazu führt, dass ausgangsseitig bezüglich des Addierers **106** die beiden Tiefpass-Bilder der zweiten Ebene nacheinander vorliegen.

[0080] Die rekonstruierte Version des ersten und des zweiten Tiefpass-Bildes der zweiten Ebene werden dann in das decodiererseitige Invers-Filter der zweiten Ebene eingespeist und dort, zusammen mit den übertragenen Hochpaß-Bildern der zweiten Ebene wieder durch die identische Implementierung der Invers-Filterbank kombiniert, um an einem Ausgang **108** der zweiten Ebene eine Folge von vier Tiefpass-Bildern der dritten Ebene zu haben. Die vier Tiefpass-Bilder der dritten Ebene werden in einer Invers-Filterebene der dritten Ebene mit den übertragenen Hochpaß-Bildern der dritten Ebene kombiniert, um an einem Ausgang **110** des Invers-Filters der dritten Ebene acht Tiefpass-Bilder der vierten Ebene in aufeinanderfolgendem Format zu haben. Diese acht Tiefpass-Bilder der dritten Ebene werden dann, in einem Invers-Filter der vierten Ebene mit den acht Hochpaß-Bildern der vierten Ebene, die vom Übertragungsmedium **100** über den Eingang HP4 empfangen werden, wieder wie anhand der ersten Ebene besprochen, kombiniert, um an einem Ausgang **112** des Invers-Filters der vierten Ebene eine rekonstruierte Gruppe von sechzehn Bildern zu erhalten.

[0081] In jeder Stufe der Analyse-Filterbank werden somit zwei Bilder, also entweder ursprüngliche Bilder oder Bilder die Tiefpaß-Signale darstellen und in einer nächsthöheren Ebene erzeugt worden sind, in ein Tiefpass-Signal und in ein Hochpass-Signal zerlegt. Das Tiefpass-Signal kann als Darstellung der Gemeinsamkeiten der Eingangsbilder betrachtet werden, während das Hochpass-Signal als Darstellung der Unterschiede zwischen den Eingangsbildern betrachtet werden kann. In der entsprechenden Stufe der Synthese-Filterbank

werden die beiden Eingangsbilder unter Verwendung des Tiefpass-Signals und des Hochpass-Signals wieder rekonstruiert. Da im Syntheseschritt die inversen Operationen des Analyseschritts durchgeführt werden, garantiert die Analyse/Synthese-Filterbank (ohne Quantisierung selbstverständlich) eine perfekte Rekonstruktion.

[0082] Die einzigen auftretenden Verluste kommen aufgrund der Quantisierung in den Weiterverarbeitungseinrichtungen z.B. **26a**, **26b**, **18** vor. Wird sehr fein quantisiert, so wird ein gutes Signal-Rausch-Verhältnis erreicht. Wird dagegen sehr grob quantisiert, so wird ein relativ schlechtes Signal-Rausch-Verhältnis, jedoch bei niedriger Bitrate, also bei niedrigem Bitbedarf, erreicht.

[0083] Ohne SNR-Skalierbarkeit könnte bereits mit dem in **Fig. 3** dargestellten Konzept zumindest eine Zeit-Skalierungssteuerung implementiert werden. Hierzu wird bezugnehmend auf **Fig. 5** eine Zeit-Skalierungs-Steuerung **120** eingesetzt, die ausgebildet ist, um eingangsseitig die Hochpaß- bzw. Tiefpass-Ausgänge bzw. die Ausgänge der Weiterverarbeitungseinrichtungen (**26a**, **26b**, **18** ...) zu erhalten, um aus diesen Teildatenströmen TP1, HP1, HP2, HP3, HP4 einen skalierten Datenstrom zu erzeugen, der in einer Basis-Skalierungsschicht die weiterverarbeitete Version des ersten Tiefpass-Bildes und des ersten Hochpass-Bildes hat. In einer ersten Erweiterungs-Skalierungsschicht könnte dann die weiterverarbeitete Version der zweiten Hochpass-Bilder untergebracht sein. In einer zweiten Erweiterungs-Skalierungsschicht könnten dann die weiterverarbeiteten Versionen der Hochpass-Bilder dritter Ebene untergebracht sein, während in einer dritten Erweiterungs-Skalierungsschicht die weiterverarbeiteten Versionen der Hochpass-Bilder vierter Ebene eingebracht sind. Damit könnte ein Decodierer allein aufgrund der Basis-Skalierungsschicht bereits eine zeitlich gering qualitative Sequenz von Tiefpass-Bildern niedriger Ebene erzeugen, also pro Gruppe von Bildern, zwei Tiefpass-Bilder erster Ebene. Mit Hinzunahme jeder Erweiterungs-Skalierungsschicht kann die Anzahl der rekonstruierten Bilder pro Gruppe immer verdoppelt werden. Die Funktionalität des Decodierers wird typischerweise von einer Skalierungs-Steuerung gesteuert, die ausgebildet ist, um zu erkennen, wie viel Skalierungsschichten im Datenstrom enthalten sind bzw. wie viele Skalierungsschichten vom Decodierer beim Decodieren berücksichtigt werden sollen.

[0084] Nachfolgend wird bezugnehmend auf **Fig. 6** auf eine SNR-Skalierungsfunktionalität eingegangen, wie sie für die vorliegende Erfindung bevorzugt wird. Die Erklärung der SNR-Skalierbarkeit gemäß **Fig. 6** wird unter Verwendung der aus **Fig. 3** ausgegebenen Signale HP4, HP3, HP2, HP1, TP1 durchgeführt. Ein mit einer SNR-Skalierbarkeit ausgestatteter Codierer umfasst neben den anhand von **Fig. 3** dargestellten Elementen zusätzlich eine Invers-Weiterverarbeitungseinrichtung **140**, die prinzipiell die mit Q^{-1} bezeichneten Elemente auf Decodiererseite von **Fig. 3** enthalten wird. Es sei jedoch darauf hingewiesen, dass die Invers-Weiterverarbeitungseinrichtung **140** auf Codiererseite vorgesehen ist, um die SNR-Skalierbarkeit zu schaffen. Die Invers-Weiterverarbeitungseinrichtung **140** erzeugt eine inversquantisierte und invers-weiterverarbeitete Darstellung der Tiefpass-Bilder erster Ebene TP1, der Hochpass-Bilder erster Ebene HP1, der Hochpass-Bilder zweiter Ebene, der Hochpass-Bilder dritter Ebene und der Hochpaßbilder vierter Ebene.

[0085] Diese invers-quantisierten Videosignale werden einem Komponenten-Subtrahierer **142** zugeführt, der zugleich das erste Tiefpass-Bild TP1, das erste Hochpass-Bild HP1, die Hochpass-Bilder zweiter Ebene, die Hochpass-Bilder dritter Ebene und die Hochpass-Bilder vierter Ebene vor der Quantisierung, also am Ausgang der Subtrahierer **69** bzw. Addierer **70** von **Fig. 3** umfasst. Hierauf wird eine komponentenweise Subtraktion durchgeführt, d.h. es wird beispielsweise das erste invers-quantisierte Tiefpass-Bild erster Ebene von dem ersten Tiefpass-Bild erster Ebene vor der Quantisierung subtrahiert etc. um jeweilige Quantisierungsfehler für das erste Tiefpass-Bild erster Ebene F_{TP1} , das erste Hochpass-Bild erster Ebene F_{HP1} etc., zu erhalten. Diese Quantisierungsfehler, die wegen der Tatsache vorhanden sind, dass ursprünglich eine Quantisierung der ersten Quantisierungsschrittweite Qu -SW durchgeführt worden ist, wird in eine Weiterverarbeitungseinrichtung **144** eingespeist, die eine Weiterverarbeitung vorzugsweise sämtlicher Quantisierungs-Fehlersignale F unter Verwendung einer zweiten Quantisierer-Schrittweite durchführt, die kleiner als die erste Quantisierer-Schrittweite ist. Während das Ausgangssignal von der Einrichtung **144** eine erste Erweiterungs-Skalierungsschicht darstellt, stellt das Eingangssignal in die Invers-Weiterverarbeitungseinrichtung **140** die Basis-Skalierungsschicht dar, die, ebenso wie die erste Erweiterungs-Skalierungsschicht in das Medium **100** bzw. logisch ausgedrückt in die Network Abstraction Layer übertragen werden kann, wo je nach Ausführungsform noch zusätzliche Signalformatmanipulationen etc. durchgeführt werden können.

[0086] Zum Erzeugen einer zweiten Erweiterungs-Skalierungsschicht wird, aufbauend auf **Fig. 6**, dahingehend vorgegangen, dass die erste Erweiterungs-Skalierungsschicht, aufgrund des Blocks **144** wieder unter Verwendung der zweiten Quantisierer-Schrittweite invers-weiterverarbeitet wird. Das Ergebnis wird dann zu den unter Verwendung der ersten Quantisierer-Schrittweite invers-quantisierten Daten hinzuaddiert, um eine

decodierte Darstellung des Bildsignals auf der Basis der Basis-Skalierungsschicht und der ersten Erweiterungs-Skalierungsschicht zu erzeugen. Diese Darstellung wird dann wieder in einen Komponenten-Subtrahierer eingespeist, um den noch verbleibenden Fehler aufgrund der Quantisierung unter Verwendung der ersten Quantisierer-Schrittweite und der zweiten Quantisierer-Schrittweite zu berechnen. Dann soll das daraus erhaltene Fehlersignal wieder unter Verwendung einer dritten Quantisierer-Schrittweite, die kleiner als die erste und die zweite Quantisierer-Schrittweite ist, erneut quantisiert werden, um eine zweite Erweiterungs-Skalierungsschicht zu erzeugen.

[0087] Es sei darauf hingewiesen, dass jede Skalierungsschicht nicht nur die weiterverarbeiteten Tiefpaß-/Hochpaß-Bilder der verschiedenen Ebene haben kann, sondern zusätzlich, als Seiteninformationen, das Vorwärts-Bewegungs-Feld M_{i0} sowie das Rückwärts-Bewegungs-Feld M_{i1} haben kann. Ferner sei darauf hingewiesen, dass für jeweils zwei aufeinanderfolgende Bilder ein Vorwärts-Bewegungs-Feld, sowie ein Rückwärts-Bewegungs-Feld erzeugt werden, wobei diese Bewegungsfelder bzw. allgemein gesagt Prädiktionsinformationen vom Encoder zum Decoder in Form von Seiteninformationen zusätzlich zu den weiterverarbeiteten Hochpaß/Tiefpass-Bildern übertragen werden.

[0088] Nachfolgend werden der Prädiktionsschritt und der Aktualisierungsschritt des Analyse- und des Synthese-Prozesses detaillierter beschrieben. Die Bewegungsfelder M_{i0} und M_{i1} spezifizieren allgemein die Bewegung zwischen zwei Bildern unter Verwendung eines Teilsatzes der P-Slice-Syntax von H264/AVC. Für die Bewegungsfelder M_{i0} , die von den Prädiktionsschritten verwendet werden, wird es bevorzugt, einen INTRA-artigen makroblock-basierten Rückzugsmodus aufzunehmen, in dem die (bewegungskompensierten) Prädiktions-signale für einen Makroblock durch ein 4×4 -Array von Luma-Transformationskoeffizientenpegeln und zwei 2×2 -Arrays von Chroma-Transformationskoeffizientenpegeln ähnlich zum INTRA 16×16 -Makroblock-Modus bei H264/AVC spezifiziert wird, wobei alle AC-Koeffizienten auf Null gesetzt sind. In den Bewegungsfeldern M_{i1} , die für die Aktualisierungsschritte verwendet werden, wird dieser Modus nicht verwendet.

[0089] Nachfolgend wird der allgemeine Bewegungskompensations-Prädiktionsprozeß dargestellt, der von den Prädiktions- bzw. Aktualisierungs-Schritten sowohl auf der Analyseseite als auch auf der Synthese-seite verwendet wird.

[0090] Eingegeben in diesen Prozeß werden ein Referenzbild R, ein Quantisierungsparameter QP (falls benötigt) und ein blockweises Bewegungsfeld M, wobei die folgenden Eigenschaften vorliegen: Für jeden Makroblock des bewegungskompensierten Prädiktionsbildes P spezifiziert das Bewegungsfeld M einen Makroblock-Modus, der gleich $P_{-16 \times 16}$, $P_{-16 \times 8}$, $P_{-8 \times 16}$, $P_{-8 \times 8}$ oder INTRA sein kann.

[0091] Wenn der Makroblockmodus gleich $P_{-8 \times 8}$ ist, wird für jeden 8×8 -Sub-Makroblock ein entsprechender Sub-Makroblock-Modus spezifiziert ($P_{-8 \times 8}$, $P_{-8 \times 4}$, $P_{-4 \times 8}$, $P_{-4 \times 4}$).

[0092] Wenn der Makroblockmodus gleich INTRA ist, ist die Erzeugung des Prädiktions-signal-spezifiziert durch ein 4×4 -Array von Luminanz-Koeffizientenpegeln und zwei 2×2 -Arrays von Chrominanz-Koeffizienten-pegeln.

[0093] Andernfalls ist die Erzeugung des Prädiktions-signal-spezifiziert durch einen Bewegungsvektor mit Viertel-Abtastwert-Genauigkeit für jede Makroblock- oder Sub-Makroblock-Aufteilung.

[0094] Mit dem Referenzbild R und der Bewegungsfeldbeschreibung M wird das bewegungs-kompensierte Prädiktions-signal P, das dann subtrahiert bzw. addiert wird, auf Makroblock-Art und -Weise aufgebaut, wie es nachfolgend beschrieben wird:

Wenn der Makroblockmodus, der in M spezifiziert ist, nicht = INTRA ist, wird für jede Makroblock-Aufteilung oder für jede Sub-Makroblock-Aufteilung folgendes ausgeführt: Die Luma- und Chroma-Abtastwerte des Bildes, die durch die entsprechende Makroblock- oder Sub-Makroblock-Aufteilung umfasst werden, werden durch eine Viertel-Abtastwert-genaue bewegungs-kompensierte Prädiktion gemäß dem Standort ISO/IEC 14496-10 AVC, Doc. JVT-G050r1, May 2003, erhalten.

$$p[i,j] = \text{Mint}(r, i - m_x, j - m_y),$$

[0095] In der vorstehenden Gleichung ist $[m_x, m_y]^T$ der Bewegungsvektor des betrachteten Makroblocks bzw. des betrachteten Sub-Makroblocks, gegeben durch M. $r[]$ ist das Array von Luma- oder Chroma-Abtastwerten des Referenzbildes N. Darüber hinaus stellt $M_{\text{int}}()$ den Interpolationsprozess dar, der für die bewegungs-kompensierte Prädiktion in H264/AVC spezifiziert ist, jedoch mit der Ausnahme, dass das Clippen auf das Intervall

[0;255] entfernt ist.

[0096] Andernfalls (der Makroblock-Modus ist INTRA), gilt das folgende:

Das gegebene 4×4-Array aus Luminanz-Transformationskoeffizienten-Pegeln wird als Array von DC-Luma-Koeffizienten-Pegeln für den INTRA_{16×16}-Makroblock-Modus in H.264/AVC behandelt, wobei der Invers-Skalierungs-Transformations-Prozeß gemäß H.264/AVC verwendet wird, und zwar unter Verwendung des gegebenen Quantisierungsparameters QP, wobei angenommen wird, dass alle AC-Transformationskoeffizienten-Pegel auf Null gesetzt sind. Als Ergebnis wird ein 16×16-array res[] aus Residual-Luma-Abtastwerten (Rest-Luma-Abtastwerten) erhalten. Die Luma-Abtastwerte des Prädiktions-Bildes P, die sich auf den betrachteten Makroblock beziehen, werden folgendermaßen berechnet:

$$p[i,j] = 128 + res[i,j].$$

[0097] Hier sei angemerkt, dass für jeden 4×4-Luma-Block das erhaltene Prädiktionssignal P[] konstant ist und eine Approximierung des Durchschnitts des ursprünglichen 4×4-Luma-Blocks darstellt.

[0098] Für jede Chrominanz-Komponente wird das gegebene 2×2-Array aus Chrominanz-Transformations-Koeffizienten-Pegeln als Array mit DC-Chroma-Koeffizienten-Pegeln behandelt, wobei der Invers-Skalierungs/Transformationsprozeß für Chroma-Koeffizienten gemäß dem Standard verwendet wird, und zwar unter Verwendung des gegebenen Quantisierungsparameters QP, wobei angenommen wird, dass alle AC-Transformationskoeffizienten-Pegel gleich Null sind. Als Ergebnis wird ein 8×8-array res[] aus Rest-Chroma-Abtastwerten erhalten. Die Chroma-Abtastwerte des Prädiktionsbildes P, die sich auf den betrachteten Makroblock beziehen, werden folgendermaßen berechnet:

$$p[i,j] = 128 + res[i,j]$$

[0099] Es sei darauf hingewiesen, dass für jeden 4×4-Chroma-Block das erhaltene Prädiktionssignal p[] konstant ist und eine Approximierung des Durchschnitts des ursprünglichen 4×4-Chroma-Blocks darstellt.

[0100] Nach dem Erzeugen des gesamten Prädiktionssignals-Bildes P wird vorzugsweise das Deblocking-Filter, wie es gemäß dem Standard definiert ist, auf dieses Bild angewendet, wobei die Ableitung der Grenzfiterstärke nur auf den Makroblock-Modi (Informationen über INTRA) basiert und insbesondere auf den Bewegungsvektoren, die in der Bewegungsbeschreibung M spezifiziert sind, basiert. Ferner wird das Clipping auf das Intervall [0;255] entfernt.

[0101] Wie aus der obigen Beschreibung zu sehen ist, ist der allgemeine Prozeß des Erzeugens von bewegungs-kompensierten Prädiktionsbildern nahezu identisch zu dem Rekonstruktionsprozeß von P-Slices gemäß H.264/AVC. Die folgenden Unterschiede sind jedoch zu sehen:

Das Clipping auf das Intervall [0;225] in den Verfahren der bewegungskompensierten Prädiktion und des Deblockings wird entfernt.

[0102] Ferner wird eine vereinfachte Intra-Modus-Rekonstruktion ohne Intra-Prädiktion durchgeführt, wobei ferner alle AC-Transformationskoeffizienten-Pegel auf Null gesetzt sind. Ferner wird eine vereinfachte Rekonstruktion für die bewegungskompensierten Prädiktionsmodi ohne Residual-Informationen durchgeführt.

[0103] Nachfolgend wird auf den Prädiktionsschritt auf der Analyseseite (Codiererseite) Bezug genommen. Es seien zwei Eingangsbilder A und B sowie das Bewegungsfeld M₁₀ gegeben, das die blockweise Bewegung des Bilds B bezüglich des Bilds A darstellt. Ferner sei ein Quantisierungsparameter QP gegeben, wobei die folgenden Operationen durchgeführt werden, um ein Residualbild H zu erhalten:

Ein Bild P, das eine Prädiktion des Bildes B darstellt, wird erhalten, indem die vorstehend beschriebenen Verfahren aktiviert werden, wobei das Referenzbild A die Bewegungsfeldbeschreibung M₁₀ und der Quantisierungsparameter QP als Eingabe verwendet werden.

[0104] Das Residualbild H wird erzeugt durch:

$$h[i,j] = b[i,j] - p[i,j].$$

wobei h[], b[] und p[] die Luma- oder Chroma-Abtastwerte eines der Bilder H, B, bzw. P darstellen.

[0105] Nachfolgend wird auf den Aktualisierungsschritt auf Analyseseite (Codiererseite) eingegangen. Hierzu

wird das Eingangsbild A, das Restbild H, das in dem Prädiktionsschritt erhalten worden ist, sowie das Bewegungsfeld M_{i1} gegeben, das die blockweise Bewegung des Bilds A bezüglich des Bilds B darstellt, wobei ferner die folgenden Operationen durchgeführt werden, um ein Bild L zu erhalten, das das zeitliche Tiefpass-Signal darstellt.

[0106] Ein Bild P wird erzeugt, indem der vorstehend beschriebene Prozeß aufgerufen wird, wobei jedoch das Bild H das Referenzbild darstellt und ferner die Bewegungsfeld-Beschreibung M_{i1} als Eingabe verwendet werden.

[0107] Das Tiefpassbild L wird durch folgende Gleichung erzeugt:

$$l[i,j] = a[i,j] + (p[i,j] >> 1),$$

wobei $l[]$, $a[]$ und $p[]$ die Luma- oder Chroma-Abtastwert-Arrays der Bilder L, A bzw. P darstellen. Ferner sei darauf hingewiesen, dass der Operator $>>$ eine Verschiebung um ein Bit nach rechts, also eine Halbierung des Wertes darstellt.

[0108] Nachfolgend wird auf den Aktualisierungsschritt auf Synthese Seite (Decodiererseite) eingegangen. Gegeben seien ein Tiefpassfilter L, das Restbild H sowie das Bewegungsfeld M_{i1} , wobei die folgenden Operationen durchgeführt werden, um das rekonstruierte Bild A zu erhalten:

Ein Bild P wird erzeugt durch Aufrufen des vorstehend beschriebenen Prozesses, wobei H als Referenzbild verwendet wird, und wobei die Bewegungsfeld-Beschreibung M_{i1} als Eingabe verwendet wird. Das rekonstruierte Bild A wird erzeugt durch:

$$a[i,j] = l[i,j] - (p[i,j] >> 1).$$

wobei $a[]$, $l[]$ und $p[]$ die Abtastwerte-Arrays der Bilder A, L bzw. P darstellen.

[0109] Nachfolgend wird auf den Prädiktionsschritt auf Syntheseseite (Decodiererseite) eingegangen. Gegeben sei das Restbild H, das rekonstruierte Bild A, das in dem Aktualisierungsschritt erhalten wird, sowie das Bewegungsfeld M_{i0} , wobei folgende Operationen durchgeführt werden, um das rekonstruierte Bild B zu erhalten: Ein Bild P, das eine Prädiktion des Bildes B darstellt, wird erzeugt durch Aufrufen des vorstehend beschriebenen Verfahrens, wobei das Bild A als Referenzbild verwendet wird, wobei die Bewegungsfeld-Beschreibung M_{i0} verwendet wird, wobei der Quantisierungsparameter QP als Eingabe verwendet wird.

[0110] Das rekonstruierte Bild B wird erzeugt durch

$$b[i,j] = h[i,j] + p[i,j],$$

wobei $b[]$, $h[]$ und $p[]$ die Abtastwert-Arrays der Bilder B, H bzw. P darstellen.

[0111] Durch Kaskadieren der grundsätzlich paarweisen Bildzerlegungsstufen wird eine dyadische Baumstruktur erhalten, die eine Gruppe von 2^n Bildern in 2^{n-1} Restbilder und ein einziges Tiefpaß- (oder Intra-) Bild zerlegt, wie es anhand von **Fig. 7** für eine Gruppe von 8 Bildern dargestellt ist. Insbesondere zeigt **Fig. 7** das Hochpass-Bild HP1 der ersten Ebene am Ausgang **22** des Filters erster Ebene sowie das Tiefpass-Bild erster Ebene am Ausgang **24** des Filters erster Ebene. Die beiden Tiefpass-Bilder TP 2 am Ausgang **16** des Filters zweiter Ebene sowie die beiden Hochpass-Bilder, die durch die zweite Weiterverarbeitungseinrichtung **18** von **Fig. 3** verarbeitet werden, sind als zweite Ebene in **Fig. 7** gezeigt.

[0112] Die Tiefpass-Bilder dritter Ebene liegen am Ausgang **76** des Filters dritter Ebene an, während die Hochpass-Bilder dritter Ebene am Ausgang **75** in weiterverarbeiteter Form vorliegen. Die Gruppe von 8 Bildern könnte ursprüngliche 8 Videobilder umfassen, wobei dann der Codierer von **Fig. 3** ohne vierte Filterebene eingesetzt werden würde. Ist die Gruppe von 8 Bildern dagegen eine Gruppe von 8 Tiefpass-Bildern, wie sie am Ausgang **73** des Filters vierter Ebene eingesetzt werden, so kann der erfindungsgemäße Codierer ebenfalls eingesetzt werden.

[0113] Allgemein gesagt werden für eine Gruppe von 2^n Bildern ($2^{n+1}-2$) Bewegungsfeldbeschreibungen, (2^n-1) Restbilder sowie ein einziges Tiefpaß- (oder Intra-) Bild übertragen.

[0114] Die Bewegungsfeld-Beschreibungen werden unter Verwendung eines Teilsatzes der H.264/AVC-Sli-

geschichtsyntax, die die folgende Syntaxelemente aufweist, codiert:

- slice header A (mit geänderter Bedeutung bestimmter Elemente)
- slice data (subset)

o macroblock layer (subset)

- mb_type (P_16x16, P_16x8, P_8x16, P_8x8, INTRA)
- if(mb_type == P_8x8)
 - sub_mb_type (P_8x8, P_8x4, P_4x8, P_4x4)
- if(mb_type != INTRA)
 - mb_qp_delta
 - residual blocks (only LUMA_DC and CHROMA_DC)
 - else
 - motion vector differences

o end_of_slice_flag

- rbsp_slice_trailing_bits

[0115] Die Bewegungsvector-Prädiktoren werden wie im Standard spezifiziert abgeleitet.

[0116] Bezüglich der Restbilder (Hochpass-Signale) sei darauf hingewiesen, dass die Restbilder unter Verwendung eines Teilsatzes der H.264/AVC-Slice-Schicht-Syntax codiert werden, die folgende Syntaxelemente umfasst:

- slice header A (mit geänderter Bedeutung bestimmter Elemente)
- slice data (subset)

o macroblock layer (subset)

- coded_block_pattern
- mb_qp_delta
- residual blocks

o end_of_slice_flag

- rbsp_slice_trailing_bits

[0117] Im Hinblick auf die Intra-Bilder (Tiefpaß-Signale) sei darauf hingewiesen, dass dieselben allgemein unter Verwendung der Syntax des Standards codiert werden. In der einfachsten Form werden die Tiefpass-Bilder von jeder Bildergruppe unabhängig als Intra-Bilder codiert. Die Codier-Effizienz kann verbessert werden, wenn die Korrelationen zwischen den Tiefpass-Bildern aufeinanderfolgender Bildergruppen ausgenutzt werden. Somit werden in einer allgemeineren Form die Tiefpass-Bilder als P-Bilder unter Verwendung rekonstruierter Tiefpass-Bilder vorhergehender Bildergruppen als Referenzen codiert. Intra- (IDR-)Bilder werden in regulären Intervallen eingesetzt, um zufällige Zugriffspunkte zu schaffen. Die Tiefpass-Bilder werden decodiert und rekonstruiert, wie es im Standard spezifiziert ist, und zwar unter Verwendung der Deblocking-Filter-Operation.

[0118] Wie es bereits ausgeführt worden ist, schafft die vorliegende Erfindung eine SNR-Skalierbarkeit. Die Struktur in Form einer offenen statt einer geschlossenen Schleife des Subband-Lösungsansatzes liefert die Möglichkeit, eine SNR-Skalierbarkeit effizient und in weiten Stücken standardkonform unterzubringen. Es wird eine SNR-Skalierbare-Erweiterung erreicht, bei der die Basisschicht so codiert wird, wie es vorstehend be-

schrieben wird, und bei der die Erweiterungs-Schichten aus Verbesserungsbildern für die Subband-Signale bestehen, die selber wieder wie für die Residual-Bild-Syntax vorgesehen codiert werden.

[0119] Auf der Codierseite werden Rekonstruktions-Fehlerbilder zwischen den ursprünglichen Subband-Bildern, die durch die Analysefilterband erzeugt werden, und den rekonstruierten Subband-Bildern, die nach dem Decodieren der Basisschicht oder einer vorherigen Erweiterungsschicht erzeugt werden, erzeugt. Diese Rekonstruktions-Fehlerbilder werden unter Verwendung eines kleineren Quantisierungs-Parameters als in der Basis-Skalierungsschicht oder einer oder mehreren vorherigen Erweiterungs-Skalierungsschichten quantisiert und codiert, und zwar unter Verwendung der Restbild-Syntax, die vorstehend dargelegt worden ist. Auf der Decodierseite können die Subband-Darstellung der Basisschicht und die Verbesserungssignale verschiedener Erweiterungsschichten unabhängig voneinander decodiert werden, wobei die schließliche Erweiterungsschicht-Subbanddarstellung erhalten wird, indem die Basisschicht-Rekonstruktion und rekonstruierte Verbesserungssignale der Erweiterungsschichten für alle zeitlichen Subbänder aufaddiert werden.

[0120] Es sei darauf hingewiesen, dass Verhaltensverluste im Vergleich zu nicht-skalierten Anwendungen vergleichsweise klein sind, wenn die Quantisierungsparameter um einen Wert von 6 von einer Schicht zur nächsten verringert werden. Diese Halbierung der Quantisierungsschrittweite resultiert etwa in einer Verdopplung der Bitrate von einer Erweiterungsskalierungsschicht zur nächsten Erweiterungsskalierungsschicht bzw. von der Basis-Skalierungsschicht zur ersten Erweiterungsskalierungsschicht.

[0121] An dieser Stelle sei darauf hingewiesen, dass die anhand der **Fig. 5** und **6** dargestellten verschiedenen Skalierungen, nämlich der Zeitskalierung und der SNR-Skalierung kombiniert werden können, dahingehend, dass ein zeitlich skaliertes Datenstrom selbstverständlich noch SNR-skaliert sein kann.

[0122] Nachfolgend werden noch auf bestimmte Betriebscodierer Steuerungsmöglichkeiten eingegangen. Zunächst werden bevorzugte Quantisierungsparameter dargestellt. Wenn die Bewegung vernachlässigt wird und wenn die Bitverschiebung nach rechts im Aktualisierungsschritt, wie sie vorstehend dargelegt worden ist, durch eine reelwertige Multiplikation mit einem Faktor $1/2$ ersetzt wird, kann der grundsätzliche Zwei-Kanal-Analyse-Schritt normiert werden, indem die Hochpaß-Abtastwerte des Bilds H mit einem Faktor $1/\sqrt{2}$ multipliziert werden, und indem die Hochpaß-Abtastwerte mit einem Faktor $\sqrt{2}$ multipliziert werden.

[0123] Vorzugsweise wird diese Normierung in der Realisierung der Analyse-Filterbank und der Synthese-Filterbank vernachlässigt, um den Bereich der Abtastwerte nahezu konstant zu halten. Sie wird jedoch berücksichtigt während der Quantisierung der zeitlichen Subbänder. Für die grundsätzliche Zwei-Kanal-Analyse-Synthese-Filterbank kann dies ohne weiteres erreicht werden, indem das Tiefpass-Signal mit der Hälfte der Quantisierungsschrittweite quantisiert wird, die zum Quantisieren des Hochpass-Signals verwendet wird. Dies führt zu folgender Quantisierer-Auswahl für die spezifizierte dyadische Zerlegungsstruktur einer Gruppe von $2n$ Bildern: Es sei $QP_{L(n)}$ der Quantisierungsparameter, der zum Codieren des Tiefpass-Bildes verwendet wird, das nach der n -ten Zerlegungsstufe erhalten wird.

[0124] Die Quantisierungsparameter, die zum Codieren der Hochpass-Bilder verwendet werden, die nach der i -ten Zerlegungsstufe erhalten werden, werden folgendermaßen berechnet

$$QP_{H(i)} = QP_{L(n)} + 3 \cdot (n + 2 - i)$$

[0125] Der Quantisierungsparameter $QP_{intra(i)}$, der zum Quantisieren der Intra-Prädiktionssignale der Bewegungsfeld-Beschreibung $M_{(i-1)0}$ verwendet wird, welche in der i -ten Zerlegungsstufe verwendet werden, werden aus dem Quantisierungsparameter QP_{Hi} für die Hochpass-Bilder abgeleitet, die in dieser Zerlegungsstufe erzeugt werden, durch folgende Gleichung:

$$QP_{INTRA(i)} = QP_{H(i)} - 6.$$

[0126] Die Bewegungsfeldbeschreibungen M_{i0} und M_{i1} , die in dem Prädiktionsschritt einerseits und dem Aktualisierungsschritt andererseits verwendet werden, werden vorzugsweise unabhängig voneinander geschätzt. Nachfolgend wird der Prozeß zum Schätzen der Bewegungsfeldbeschreibung M_{i0} , der im Prädiktionsschritt verwendet wird, beschrieben. Der Prozeß zum Schätzen von M_{i1} wird erhalten, indem die ursprünglichen Bilder A und B ausgetauscht werden, und indem der INTRA-Modus aus dem Satz von möglichen Macroblock-Modi entfernt wird.

[0127] Die Bilder A und B seien gegeben, die entweder ursprüngliche Bilder sind, oder Bilder, die Tiefpaß-Si-

gnale darstellen, die in einer vorherigen Analyse-Stufe erzeugt werden. Ferner werden die entsprechenden Arrays von Luma-Abtastwerten $a[]$ und $b[]$ bereitgestellt. Die Bewegungsbeschreibung M_{i0} wird in einer Makroblock-weisen Art und Weise folgendermaßen geschätzt:

Für alle möglichen Makroblock- und Sub-Makroblock-Teilungen eines Makroblocks i innerhalb des Bildes B werden die zugeordneten Bewegungsvektoren

$$m_i = [m_x, m_y]^T$$

durch Minimieren des Lagrange-Funktional

$$m_i = \arg \min_{m \in S} \{ D_{SAD}(i, m) + \lambda \cdot R(i, m) \}$$

bestimmt,

wobei der Verzerrungsterm folgendermaßen gegeben ist:

$$D_{SAD}(i, m) = \sum_{(x,y) \in P} |b[x, y] - a[x - m_x, y - m_y]|$$

[0128] Hierbei spezifiziert S den Bewegungsvektor-Suchbereich innerhalb des Referenzbildes A . P ist der Bereich, der durch die betrachtete Makroblock-Teilung oder Sub-Makroblock-Teilung überstrichen wird. $R(i, m)$ spezifiziert die Anzahl von Bits, die benötigt werden, um alle Komponenten des Bewegungsvektors m zu übertragen, wobei ein fester Lagrange-Multiplizierer ist.

[0129] Die Bewegungssuche schreitet zunächst über alle Ganzzahl-Abtastwert-genauen Bewegungsvektoren in dem gegebenen Suchbereich S fort. Dann, unter Verwendung des besten Ganzzahl-Bewegungsvektors werden die 8 umgebenden Halb-Abtastwertgenauen Bewegungsvektoren getestet. Schließlich wird unter Verwendung des besten Halb-Abtastwert-genauen Bewegungsvektors die 8 umgebenden Viertel-Abtastwert-genauen Bewegungsvektoren getestet. Für die Halb- und die Viertel- Abtastwert-genaue Bewegungsvektor-Verbesserung wird der Term

$$a[x - m_x, y - m_y]$$

als Interpolationsoperator interpretiert.

[0130] Die Modus Entscheidung für den Makroblock-Modus und den Sub-Makroblock-Modus folgt grundsätzlich demselben Lösungsansatz. Von einem gegebenen Satz von möglichen Makroblock- oder Sub-Makroblock-Modi S_{mode} wird der Modus p_i der das folgende Lagrange-Funktional minimiert, ausgewählt:

$$p_i = \arg \min_{p \in S_{\text{mode}}} \{ D_{SAD}(i, p) + \lambda \cdot R(i, p) \}.$$

[0131] Der Verzerrungsterm ist folgendermaßen gegeben:

$$D_{SAD}(i, p) = \sum_{(x,y) \in P} |b[x, y] - a[x - m_x[p, x, y], y - m_y[p, x, y]]|,$$

wobei P den Makroblock oder Sub-Makroblock-Bereich spezifiziert, und wobei $n[p, x, y]$ der Bewegungsvektor ist, der dem Makroblock- oder Sub-Makroblock-Modus p und der Teilung oder Sub-Makroblock-Teilung zugeordnet ist, die die Luma-Position (xy) umfaßt.

[0132] Der Ratenterm $R(ip)$ stellt die Anzahl von Bits dar, die der Wahl des Codiermodus p zugeordnet sind. Für die Bewegungskompensierten Codiermodi umfasst derselbe die Bits für den Makroblock-Modus (falls zutreffend), den bzw. die Sub-Makroblock-Modi (falls zutreffend) und den bzw. die Bewegungsvektoren. Für den Intra-Modus umfasst derselbe die Bits für den Makroblock-Modus und die Arrays von quantisierten Luma- und Chroma-Transformation-Koeffizienten Pegeln.

[0133] Der Satz von möglichen Sub-Makroblock-Modi wird durch

$$\{P_{8 \times 8}, P_{8 \times 4}, P_{4 \times 8}, P_{4 \times 4}\}$$

gegeben. Der Satz von möglichen Makroblock-Modi wird durch

$$\{P_{16 \times 16}, P_{16 \times 8}, P_{8 \times 16}, P_{8 \times 8}, \text{INTRA}\}$$

gegeben, wobei der Intra-Modus nur verwendet wird, wenn eine Bewegungsfeldbeschreibung M_{i0} , die für den Prädiktionsschritt verwendet wird, geschätzt wird.

[0134] Der Lagrange-Multiplizierer λ wird abhängig von dem Basisschicht-Quantisierungs-Parameter für das bzw. die Hochpass-Bilder Q_{Phi} der Zerlegungsstufe, für die das Bewegungsfeld geschätzt wird, gemäß folgender Gleichung gesetzt:

$$\lambda = 0.33 \cdot 2^{(QP_{\text{Hi}}/3 - 4)}$$

[0135] Die grundsätzliche Zwei-Kanal-Analyse Filterbank zerlegt zwei Eingangsbilder A, B in ein Tiefpass-Bild L und ein Hochpass-Bild H. Gemäß der Notation in diesem Dokument verwendet das Tiefpass-Bild L das Koordinatensystem des ursprünglichen Bildes A. Wird somit eine perfekte (fehlerfreie) Bewegungskompensation angenommen, sind die Bilder A und L identisch.

[0136] Die Zerlegungsstruktur, die in **Fig. 4** dargelegt ist, wird erhalten, wenn in allen Zerlegungsstufen die ganzzahligen Eingangsbilder bei zeitlichen Abtastpositionen 0, 2, 4,... als Eingangsbilder A behandelt werden, und wenn die ungeradzahligen Eingangsbilder bei Abtastpositionen 1, 3, 5, ... als Eingangsbilder B behandelt werden. Dieses Schema ermöglicht eine optimale zeitliche Skalierbarkeit. Der zeitliche Abstand zwischen den Bildern, die in jeder Zwei-Kanal-Analyse Filterbank zerlegt werden, wird jedoch um einen Faktor 2 von einer Zerlegungsstufe zur nächsten erhöht. Es ist ferner bekannt, dass der Wirkungsgrad der bewegungskompensierten Prädiktion abnimmt, wenn der zeitliche Abstand zwischen dem Referenzbild und dem zu prädizierenden Bild zunimmt.

[0137] Erfindungsgemäß wird es bevorzugt, Zerlegungsschemen zu realisieren, bei denen der zeitliche Abstand zwischen den Bildern, die durch die Zwei-Kanal-Filterbank zerlegt werden, um einen Faktor kleiner als 2 von einer Zerlegungsstufe zur nächsten erhöht wird. Diese Schemen liefern jedoch nicht das Merkmal der optimalen zeitlichen Skalierbarkeit, da die Abstände zwischen benachbarten Tiefpaß-Filtern in den meisten Zerlegungsstufen variieren.

[0138] Erfindungsgemäß wird das in **Fig. 8** gezeigte Zerlegungsschema verwendet, von dem ausgegangen wird, dass es einen vernünftigen Kompromiß zwischen zeitlicher Skalierbarkeit und Codier-Wirkungsgrad ermöglicht. Die Sequenz der ursprünglichen Bilder wird als Sequenz von Eingangsbildern A, B, A, B, A, B... A, B behandelt. Somit liefert dieses Schema eine Stufe mit optimaler zeitlicher Skalierbarkeit (gleicher Abstand zwischen den Tiefpass-Bildern). Die Sequenz von Tiefpass-Bildern, die als Eingangs-Signal in alle folgenden Zerlegungsstufen verwendet werden, werden als Sequenzen von Eingangsbildern B, A, A, B, B, A ... A, B behandelt, wodurch die Abstände zwischen den Tiefpass-Bildern, die zerlegt werden in dem folgenden Zwei-Kanal-Analyse-Schema, kleingehalten werden, wie es in **Fig. 8** zu sehen ist.

[0139] Bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung werden die Bewegungsfelder M_{i0} und M_{i1} , die in dem Prädiktions-Schritt bzw. den Aktualisierungsschritt verwendet werden, unabhängig voneinander geschätzt und codiert. Dies kann möglicherweise dazu führen, dass die Bit-Rate, die benötigt wird, um die Bewegungsparameter zu übertragen, erhöht wird. Möglicherweise existiert jedoch auch ein negativer Einfluß auf die Konnektivität dieser zwei Bewegungsfelder, die einen wichtigen Einfluß auf den Codierwirkungsgrad des Sub-Band-Lösungsansatzes haben kann. Daher wird es bevorzugt, dass im Sinne der Verbesserung des Codierer-Wirkungsgrades die Bewegungsfelder M_{i1} , die im Aktualisierungsschritt verwendet werden, nicht unabhängig voneinander geschätzt und codiert werden, sondern von den Bewegungsfeldern M_{i0} , die in den Prädiktionsschritten berechnet werden, abgeleitet werden, dahingehend, dass sie noch eine blockweise Bewegung darstellen, die mit der H.264/AVC-Spezifikation kompatibel ist. Als zusätzlichen Seiteneffekt wird dies ebenfalls die Komplexität reduzieren, die für den Aktualisierungsschritt benötigt wird.

[0140] Die gegenwärtige Analyse/Synthese-Struktur stellt eine Lifting-Darstellung der einfachen Haar-Filter bei einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung dar. Dieses Schema wird bei einem anderen Ausführungsbeispiel auf eine Lifting-Darstellung mit bi-orthogonalen 5/3-Filtern erweitert, was zu einer bi-direktionalen Bewegungskompensierten Prädiktion führt. Der bevorzugteste Lösungsansatz besteht in einem adaptiven Umschalten zwischen dem Lifting-Darstellungen der Haar-Filter und der 5/3-Filter auf einer Blockbasis, wobei vorzugsweise ferner die bewegungskompensierte Prädiktion, wie sie für B-Slices in H.264/AVC spezifiziert ist, eingesetzt werden kann.

[0141] Zusätzlich wird es bevorzugt, die Größe der Gruppen von Bildern für die zeitliche Subbandzerlegung adaptiv auszuwählen.

[0142] Da die Verwendung von mehreren Referenzbildern das Verhalten von Prädiktions-Basierten Videocodierschemen beträchtlich verbessert hat, wird auch eine Aufnahme dieses Lösungsansatzes in ein Subband-

schema bevorzugt.

[0143] Ferner wird die Verwendung eines geeigneten Bitzuordnungsalgorithmus bevorzugt, der SNR-Schwankungen innerhalb einer Gruppe von Bildern reduziert, die bei bestimmten Testsequenzen auftreten können.

[0144] Ferner wird es bevorzugt, neue Techniken zur Transformations-Koeffizienten-Codierung zu verwenden, die die SNR-Skalierbarkeit verbessern und einen zusätzlichen Grad an räumlicher Skalierbarkeit liefern.

[0145] Abhängig von den Gegebenheiten kann das erfindungsgemäße Verfahren zum Codieren, zum Decodieren, zum Filtern bzw. zum inversen Filtern in Hardware oder in Software implementiert werden. Die Implementierung kann auf einem digitalen Speichermedium, insbesondere einer Diskette oder CD mit elektronisch auslesbaren Steuersignalen erfolgen, die so mit einem programmierbaren Computersystem zusammenwirken können, dass das Verfahren ausgeführt wird. Allgemein besteht die Erfindung somit auch in einem Computer-Programm-Produkt mit einem auf einem Maschinen-lesbaren Träger-gespeicherten Programmcode zur Durchführung des erfindungsgemäßen Verfahrens, wenn das Computer-Programm-Produkt auf einem Rechner abläuft. In anderen Worten ausgedrückt, kann die Erfindung somit auf einem Computer-Programm mit einem Programmcode zur Durchführung des Verfahrens realisiert werden, wenn das Computer-Programm auf einem Computer abläuft.

Patentansprüche

1. Vorrichtung zum Codieren einer Gruppe von aufeinanderfolgenden Bildern, wobei die Gruppe mindestens ein erstes, ein zweites, ein drittes und ein viertes Bild aufweist, mit folgenden Merkmalen:
einer zweiten Filterebene (**12**) zum Erzeugen eines ersten Hochpassbildes der zweiten Ebene und eines ersten Tiefpassbildes der zweiten Ebene aus dem ersten und dem zweiten Bild und zum Erzeugen eines zweiten Hochpassbildes der zweiten Ebene und eines zweiten Tiefpassbildes der zweiten Ebene aus dem dritten Bild und dem vierten Bild;

einer ersten Filterebene (**20**) zum Erzeugen eines ersten Hochpassbildes der ersten Ebene und eines ersten Tiefpassbildes der ersten Ebene aus dem ersten und aus dem zweiten Tiefpassbild der zweiten Ebene; und
einer ersten Weiterverarbeitungseinrichtung (**26**) zum Weiterverarbeiten des ersten Hochpassbildes der ersten Ebene und des ersten Tiefpassbildes der ersten Ebene, um ein codiertes Bildsignal (**28**) zu erhalten, wobei die erste Weiterverarbeitungseinrichtung (**26**) einen Quantisierer umfasst, der eine Quantisierer-Schrittweite hat.

2. Vorrichtung nach Anspruch 1, bei der die zweite Filterebene (**12**) oder die erste Filterebene (**20**) gemäß einem Lifting-Schema ausgebildet ist.

3. Vorrichtung nach Anspruch 1 oder 2, bei der die zweite Filterebene (**12**) oder die erste Filterebene (**20**) ausgebildet sind, um ein Tiefpassbild so zu erzeugen, dass es Gemeinsamkeiten aus zwei gefilterten Bildern umfasst, und um ein Hochpassbild so zu erzeugen, dass es Unterschiede aus zwei gefilterten Bildern umfasst.

4. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die zweite Filterebene folgende Merkmale aufweist:

einen ersten Prädiktor (**67**) zum Erzeugen eines ersten Prädiktionsbildes für das erste Bild basierend auf dem zweiten Bild;

einem zweiten Prädiktor (**68**) zum Erzeugen eines zweiten Prädiktionsbildes für das zweite Bild basierend auf dem ersten Bild;

einen ersten Kombiner (**69**) zum Kombinieren des ersten Prädiktionsbildes mit dem ersten Bild; und

einem zweiten Kombiner (**70**) zum Kombinieren des zweiten Prädiktionsbildes mit dem zweiten Bild.

5. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die erste Filtereinrichtung folgende Merkmale aufweist:

einen ersten Prädiktor (**60**) zum Erzeugen eines ersten Prädiktionsbildes für das erste Tiefpassbild der zweiten Ebene basierend auf dem zweiten Tiefpassbild der zweiten Ebene;

einen zweiten Prädiktor (**61**) zum Erzeugen eines zweiten Prädiktionsbildes für das zweite Tiefpassbild der zweiten Ebene basierend auf dem ersten Tiefpassbild der zweiten Ebene;

einem ersten Kombiner (**44**) zum Kombinieren des ersten Prädiktionsbildes mit dem ersten Bild; und einem zweiten Kombiner (**46**) zum Kombinieren des zweiten Prädiktionsbildes mit dem zweiten Bild.

6. Vorrichtung nach Anspruch 4 oder 5, bei der die erste Kombinationseinrichtung (**44, 69**) ein Subtrahierer

ist, und die zweite Kombinationseinrichtung (**70, 46**) ein Addierer ist, und bei der ein Ausgangssignal des ersten Kombinierers (**44, 69**) ein Hochpassbild ist, und bei der ein Ausgangssignal des zweiten Kombinierers ein Tiefpassbild ist.

7. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die erste Filterebene oder die zweite Filterebene ausgebildet ist, um eine Bewegungskompensation durchzuführen.

8. Vorrichtung nach Anspruch 4 oder 5, bei der der erste Prädiktor ausgebildet ist, um eine Rückwärtsbewegungskompensation auszuführen, und bei der der zweite Prädiktor ausgebildet ist, um eine Vorwärtsbewegungskompensation auszuführen.

9. Vorrichtung nach Anspruch 4 oder 5, bei der der erste oder der zweite Prädiktor ausgebildet ist, um ein Prädiktionsbild unter Verwendung von Prädiktionsinformationen zu erzeugen, wobei die Prädiktionsinformationen in einem Seitenkanal übertragbar sind.

10. Vorrichtung nach einem der vorhergehenden Ansprüche, die ferner folgende Merkmale aufweist: eine Zeit-Skalierbarkeitssteuerung (**120**) zum Eintragen einer weiterverarbeitenden Version des ersten Hochpassbildes der ersten Ebene und des ersten Tiefpassbildes der ersten Ebene in einer Basis-Skalierungsschicht, und zum Eintragen einer weiterverarbeiteten Version des ersten Hochpassbildes der zweiten Ebene und des zweiten Hochpassbildes der zweiten Ebene in einer Erweiterungs-Skalierungsschicht.

11. Vorrichtung nach einem der vorhergehenden Ansprüche, die ferner folgende Merkmale aufweist: eine Signal/Rausch-Abstand-Skalierungsverarbeitungseinrichtung, die folgende Merkmale aufweist: eine Einrichtung zum Eintragen einer weiterverarbeiteten Version des ersten Hochpassbildes der zweiten Ebene und des zweiten Hochpassbildes der zweiten Ebene sowie einer weiterverarbeiteten Version des ersten Hochpassbildes der ersten Ebene und einer weiterverarbeiteten Version des ersten Tiefpassbildes der ersten Ebene in eine Basis-Skalierungsschicht, wobei der Basis-Skalierungsschicht zumindest eine Quantisierungs-Schrittweite zugeordnet ist; eine Einrichtung (**140**) zum Invers-Weiterverarbeiten der weiterverarbeiteten Versionen, die in der Basis-Skalierungsschicht enthalten sind, wobei die Einrichtung zum Invers-Weiterverarbeiten einen Requantisierer umfasst, der ausgebildet ist, um aufgrund der Quantisierungs-Schrittweite zu arbeiten, die der Basis-Skalierungsschicht zugeordnet ist; eine Einrichtung (**142**) zum Subtrahieren der Inversweiterverarbeiteten Versionen von entsprechenden ursprünglich erzeugten Hochpaß- bzw. Tiefpass-Bildern, um komponentenweise Quantisierungs-Fehlersignale zu erhalten; und eine Einrichtung (**144**) zum Weiterverarbeiten der Quantisierungs-Fehlersignale unter Verwendung einer weiteren Quantisierer-Schrittweite, die kleiner ist als eine Quantisierer-Schrittweite, die der Basis-Skalierungsschicht zugeordnet ist, und zum Eintragen der weiterverarbeiteten Versionen der Quantisierungs-Fehlersignale in eine Erweiterungsschicht, der die weitere Quantisierer-Schrittweite zugeordnet ist.

12. Vorrichtung nach einem der vorhergehenden Ansprüche, die ferner eine zweite Weiterverarbeitungseinheit (**18**) zum Weiterverarbeiten des ersten Hochpassbildes der zweiten Ebene und des zweiten Hochpassbildes der zweiten Ebene aufweist, und bei der die erste Weiterverarbeitungseinrichtung (**26a, 26b**) und die zweite Weiterverarbeitungseinrichtung (**18**) ausgebildet sind, um Weiterverarbeitungen gemäß dem Standard H.264/AVC durchzuführen.

13. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der das erste, das zweite, das dritte und das vierte Bild ein erstes, ein zweites, ein drittes bzw. ein viertes Tiefpassbild einer dritten Ebene sind, die unter Verwendung einer dritten Filtereinrichtung aus einer Gruppe von acht Bildern erzeugt worden sind.

14. Vorrichtung nach Anspruch 13, bei der die acht Bilder wiederum acht Tiefpassbilder sind, die unter Verwendung einer vierten Filterebene aus einer Gruppe von sechzehn Bildern erzeugt worden sind.

15. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die zweite Filterebene oder die erste Filterebene eine Einrichtung zum Verzögern eines Bildes um die Dauer eines Bildes und eine Einrichtung (**42a, 42b**) zum Eliminieren eines Bildes aus einem entsprechenden Signal umfasst.

16. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Gruppe von Bildern mit dem ersten, dem zweiten, dem dritten und dem vierten Bild aus einer ursprünglichen Gruppe von Bildern abgeleitet ist, bei der das zweite Bild sequenzmäßig vor dem ersten Bild auftritt, wobei die Gruppe aus der ursprünglichen

Gruppe mittels einer Umstellungseinrichtung erzeugbar ist.

17. Vorrichtung zum Decodieren eines codierten Signals, das durch Filtern einer Gruppe von aufeinanderfolgenden Bildern, die mindestens ein erstes, ein zweites, ein drittes oder ein viertes Bild umfassen, abgeleitet ist, durch Filtern zum Erzeugen eines ersten Hochpassbildes einer zweiten Ebene und eines ersten Tiefpassbildes der zweiten Ebene aus dem ersten und dem zweiten Bild und zum Erzeugen eines zweiten Hochpassbildes der zweiten Ebenen und eines zweiten Tiefpassbildes der zweiten Ebene aus dem dritten und dem vierten Bild, durch Filtern zum Erzeugen eines ersten Hochpassbildes der ersten Ebene und eines ersten Tiefpassbildes der ersten Ebene aus dem ersten und dem zweiten Tiefpassbild der zweiten Ebene und durch Weiterverarbeiten des ersten Hochpassbildes der ersten Ebene und des ersten Tiefpassbildes der ersten Ebene, um ein codiertes Bildsignal zu erhalten, mit folgenden Merkmalen:

einer Einrichtung (30) zum Invers-Weiterverarbeiten einer weiterverarbeiteten Version des ersten Hochpassbildes der ersten Ebene und einer weiterverarbeiteten Version des ersten Tiefpassbildes der ersten Ebene unter Verwendung einer Quantisierer-Schrittweite, um eine rekonstruierte Version des ersten Hochpassbildes und um eine rekonstruierte Version des ersten Tiefpassbildes zu erhalten; und
einer Invers-Filterebene (32) zum Invers-Filtern der rekonstruierten Version des ersten Hochpassbildes und der rekonstruierten Version des ersten Tiefpassbildes, um eine rekonstruierte Version eines ersten Tiefpassbildes und eines zweiten Tiefpassbildes einer nächsthöheren Ebene zu erhalten.

18. Vorrichtung nach Anspruch 17, die ferner folgende Merkmale aufweist:

eine weitere Invers-Weiterverarbeitungseinrichtung zum Invers-Weiterverarbeiten eines ersten Hochpassbildes der zweiten Ebene und eines zweiten Hochpassbildes der zweiten Ebene, um eine rekonstruierte Version des ersten Hochpassbildes der zweiten Ebene und des zweiten Hochpassbildes der zweiten Ebene zu erhalten; und

eine weitere Invers-Filterebene zum Invers-Filtern der rekonstruierten Version des ersten Tiefpassbildes der nächsthöheren Ebene und des zweiten Tiefpassbildes der nächsthöheren Ebene mit der rekonstruierten Version des ersten Hochpassbildes der zweiten Ebene und der rekonstruierten Version des zweiten Hochpassbildes der zweiten Ebene, um rekonstruierte Versionen von vier Bildern einer nächsthöheren Ebene zu erhalten, die ursprüngliche Bilder oder Tiefpassbilder darstellen.

19. Vorrichtung nach Anspruch 17 oder 18, bei der das codierte Signal ein skaliertes Signal mit einer Basisschicht und einer Erweiterungsschicht ist, wobei die Basisschicht unter Verwendung einer Quantisierer-Schrittweite weiterverarbeitete Bilder und unter Verwendung einer weiteren Quantisierer-Schrittweite weiterverarbeitete Quantisierungs-Fehlersignale für die Bilder aufweist, mit ferner folgenden Merkmalen:

einer Einrichtung zum Invers-Weiterverarbeiten der Quantisierungs-Fehlersignale, um eine rekonstruierte Version der Quantisierungs-Fehlersignale zu erhalten; und

eine Einrichtung zum Addieren der rekonstruierten Version der Quantisierungs-Fehlersignale zu einer rekonstruierten Version der Signale der Basisschicht, um ein decodiertes Ausgangssignal zu erhalten, das eine höhere Qualität hat als ein decodiertes Ausgangssignal allein aufgrund der Basisschicht.

20. Verfahren zum Codieren einer Gruppe von aufeinanderfolgenden Bildern, wobei die Gruppe mindestens ein erstes, ein zweites, ein drittes und ein viertes Bild aufweist, mit folgenden Schritten:

Erzeugen (12) eines ersten Hochpassbildes der zweiten Ebene und eines ersten Tiefpassbildes der zweiten Ebene aus dem ersten und dem zweiten Bild und Erzeugen eines zweiten Hochpassbildes der zweiten Ebene und eines zweiten Tiefpassbildes der zweiten Ebene aus dem dritten Bild und dem vierten Bild;

Erzeugen (20) eines ersten Hochpassbildes der ersten Ebene und eines ersten Tiefpassbildes der ersten Ebene aus dem ersten und aus dem zweiten Tiefpassbild der zweiten Ebene; und

Weiterverarbeiten (26) des ersten Hochpassbildes der ersten Ebene und des ersten Tiefpassbildes der ersten Ebene, um ein codiertes Bildsignal (28) zu erhalten, wobei die erste Weiterverarbeitungseinrichtung (26) einen Quantisierer umfasst, der eine Quantisierer-Schrittweite hat.

21. Verfahren zum Decodieren eines codierten Signals, das durch Filtern einer Gruppe von aufeinanderfolgenden Bildern, die mindestens ein erstes, ein zweites, ein drittes oder ein viertes Bild umfassen, abgeleitet ist, durch Filtern zum Erzeugen eines ersten Hochpassbildes einer zweiten Ebene und eines ersten Tiefpassbildes der zweiten Ebene aus dem ersten und dem zweiten Bild und zum Erzeugen eines zweiten Hochpassbildes der zweiten Ebenen und eines zweiten Tiefpassbildes der zweiten Ebene aus dem dritten und dem vierten Bild, durch Filtern zum Erzeugen eines ersten Hochpassbildes der ersten Ebene und eines ersten Tiefpassbildes der ersten Ebene aus dem ersten und dem zweiten Tiefpassbild der zweiten Ebene und durch Weiterverarbeiten des ersten Hochpassbildes der ersten Ebene und des ersten Tiefpassbildes der ersten Ebene, um ein codiertes Bildsignal zu erhalten, mit folgenden Schritten:

Invers-Weiterverarbeiten (**30**) einer weiterverarbeiteten Version des ersten Hochpassbildes der ersten Ebene und einer weiterverarbeiteten Version des ersten Tiefpassbildes der ersten Ebene unter Verwendung einer Quantisierer-Schrittweite, um eine rekonstruierte Version des ersten Hochpassbildes und um eine rekonstruierte Version des ersten Tiefpassbildes zu erhalten; und
 Invers-Filtern (**32**) der rekonstruierten Version des ersten Hochpassbildes und der rekonstruierten Version des ersten Tiefpassbildes, um eine rekonstruierte Version eines ersten Tiefpassbildes und eines zweiten Tiefpassbildes einer nächsthöheren Ebene zu erhalten.

22. Filtervorrichtung zum Filtern eines ersten und eines zweiten Bildes, mit folgenden Merkmalen:
 einem Rückwärts-Prädiktor (**60**) zum Erzeugen eines ersten Prädiktionsbildes für das erste Bild basierend auf dem zweiten Bild; und
 einem Vorwärts-Prädiktor (**61**) zum Erzeugen eines zweiten Prädiktionsbildes für das zweite Bild basierend auf dem ersten Bild;
 einem Subtrahierer (**44**) zum Subtrahieren des ersten Prädiktionsbildes von dem ersten Bild, um ein Hochpassbild zu erhalten, das Unterschiede des ersten und des zweiten Bildes umfasst; und
 einem Addierer (**46**) zum Addieren des Prädiktionsbildes zu dem zweiten Bild, um ein Tiefpassbild zu erhalten, das Gemeinsamkeiten des ersten und des zweiten Bildes darstellt.

23. Invers-Filtervorrichtung zum Invers-Filtern eines Hochpassbildes und eines Tiefpassbildes, um ein erstes und ein zweites Bild zu erhalten, mit folgenden Merkmalen:
 einem Vorwärts-Prädiktor zum Erzeugen eines ersten Prädiktionsbildes für das Tiefpassbild aus dem Hochpassbild;
 einem Subtrahierer (**101**) zum Subtrahieren des Prädiktionsbildes von dem Tiefpassbild;
 einem Rückwärts-Prädiktor (**60**) zum Erzeugen eines Prädiktionsbildes für das Hochpassbild unter Verwendung eines Ausgangssignals aus dem Subtrahierer; und
 einem Addierer (**102**) zum Addieren des Hochpassbildes und des zweiten Prädiktionsbildes, das aus dem Rückwärts-Prädiktor (**60**) erhalten wird,
 wobei ein Ausgangssignal aus dem Addierer (**102**) das erste Bild darstellt, während ein Ausgangssignal aus dem Subtrahierer (**101**) das zweite Bild darstellt.

24. Verfahren zum Filtern eines ersten und eines zweiten Bildes, mit folgenden Schritten:
 Erzeugen (**60**) eines ersten Prädiktionsbildes für das erste Bild basierend auf dem zweiten Bild; und
 Erzeugen (**61**) eines zweiten Prädiktionsbildes für das zweite Bild basierend auf dem ersten Bild;
 Subtrahieren (**44**) des ersten Prädiktionsbildes von dem ersten Bild, um ein Hochpassbild zu erhalten, das Unterschiede des ersten und des zweiten Bildes umfasst; und
 Addieren (**46**) des Prädiktionsbildes zu dem zweiten Bild, um ein Tiefpassbild zu erhalten, das Gemeinsamkeiten des ersten und des zweiten Bildes darstellt.

25. Verfahren zum Invers-Filtern eines Hochpassbildes und eines Tiefpassbildes, um ein erstes und ein zweites Bild zu erhalten, mit folgenden Schritten:
 Erzeugen eines ersten Prädiktionsbildes für das Tiefpassbild aus dem Hochpassbild;
 Subtrahieren (**101**) des Prädiktionsbildes von dem Tiefpassbild;
 Erzeugen (**60**) eines zweiten Prädiktionsbildes für das Hochpassbild unter Verwendung eines Ausgangssignals aus dem Subtrahierer; und
 Addieren (**102**) des Hochpassbildes und des zweiten Prädiktionsbildes,
 wobei ein Ausgangssignal aus dem Addieren (**102**) das erste Bild darstellt, während ein Ausgangssignal aus dem Subtrahieren (**101**) das zweite Bild darstellt.

26. Computer-Programm mit einem Programmcode zum Durchführen des Verfahrens nach Patentanspruch 20, 21, 24 oder 25, wenn das Computer-Programm auf einem Computer abläuft.

Es folgen 9 Blatt Zeichnungen

FIG 1

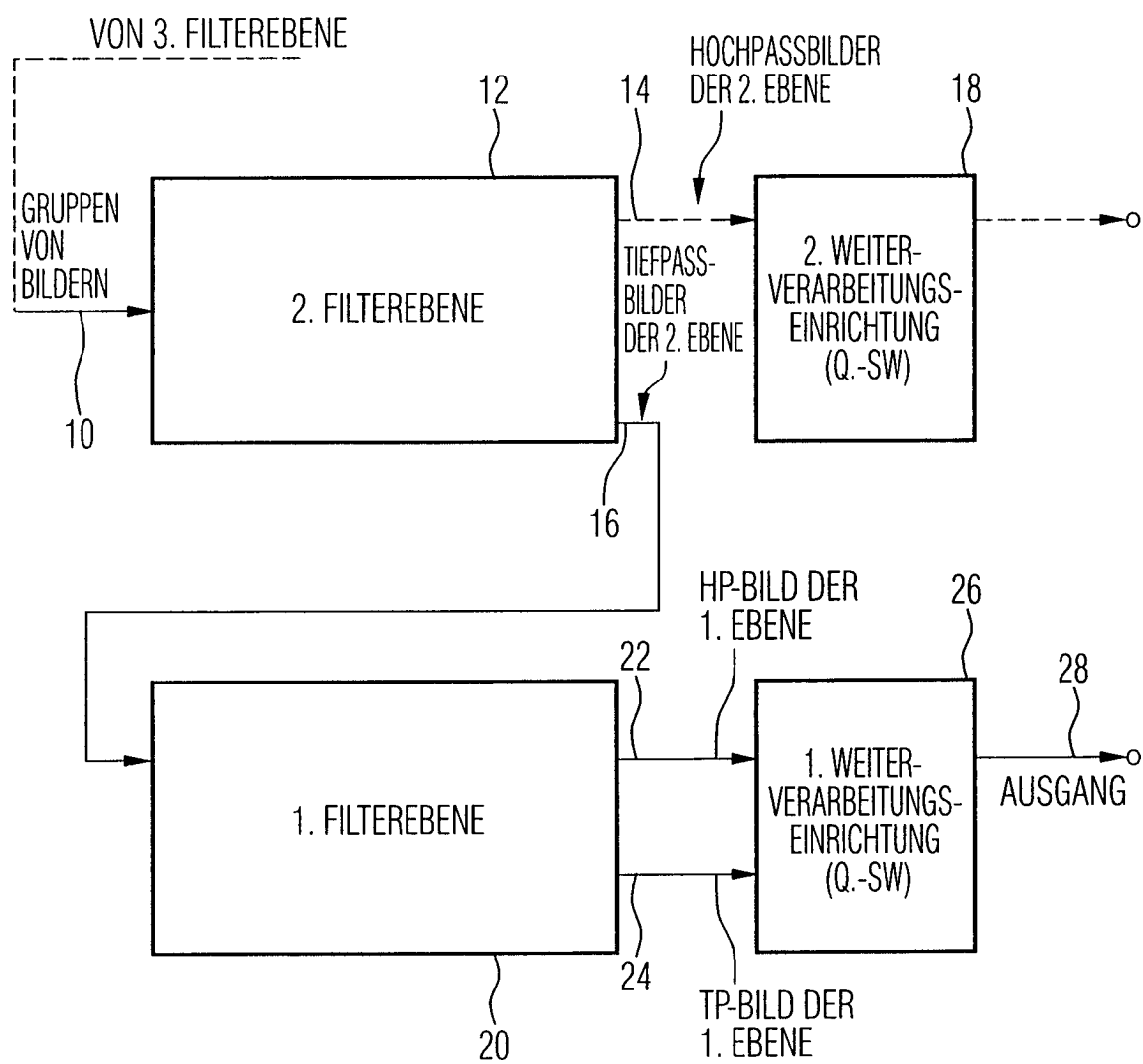


FIG 2

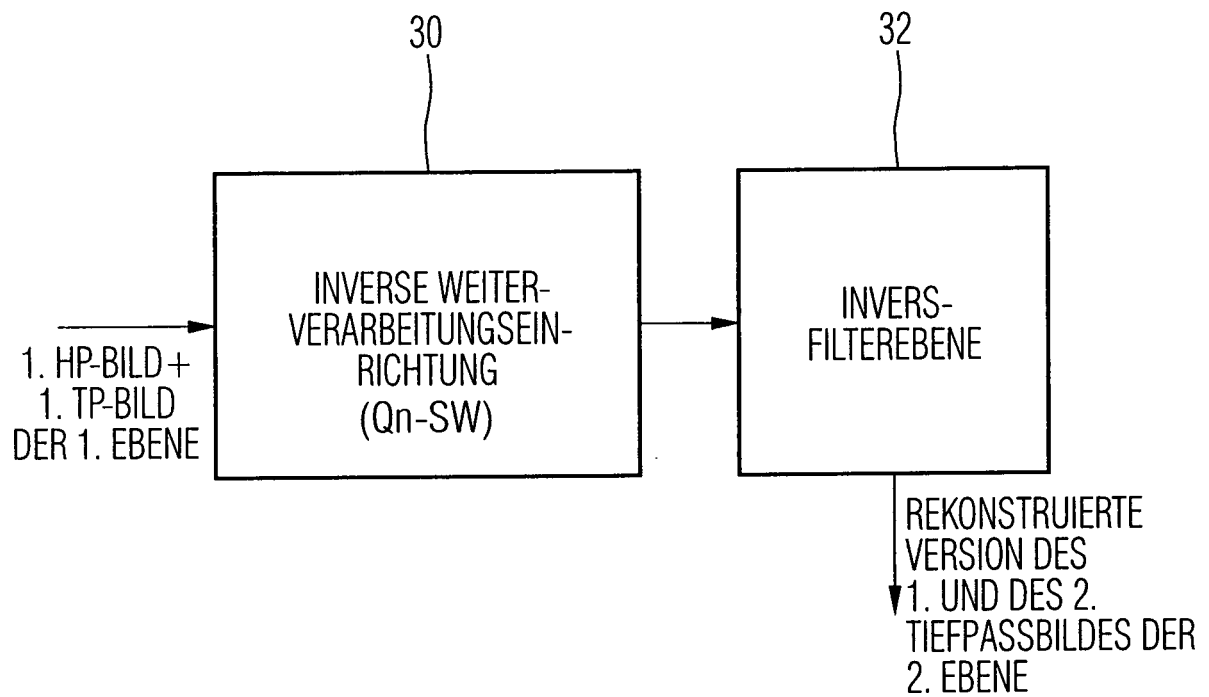


FIG 3

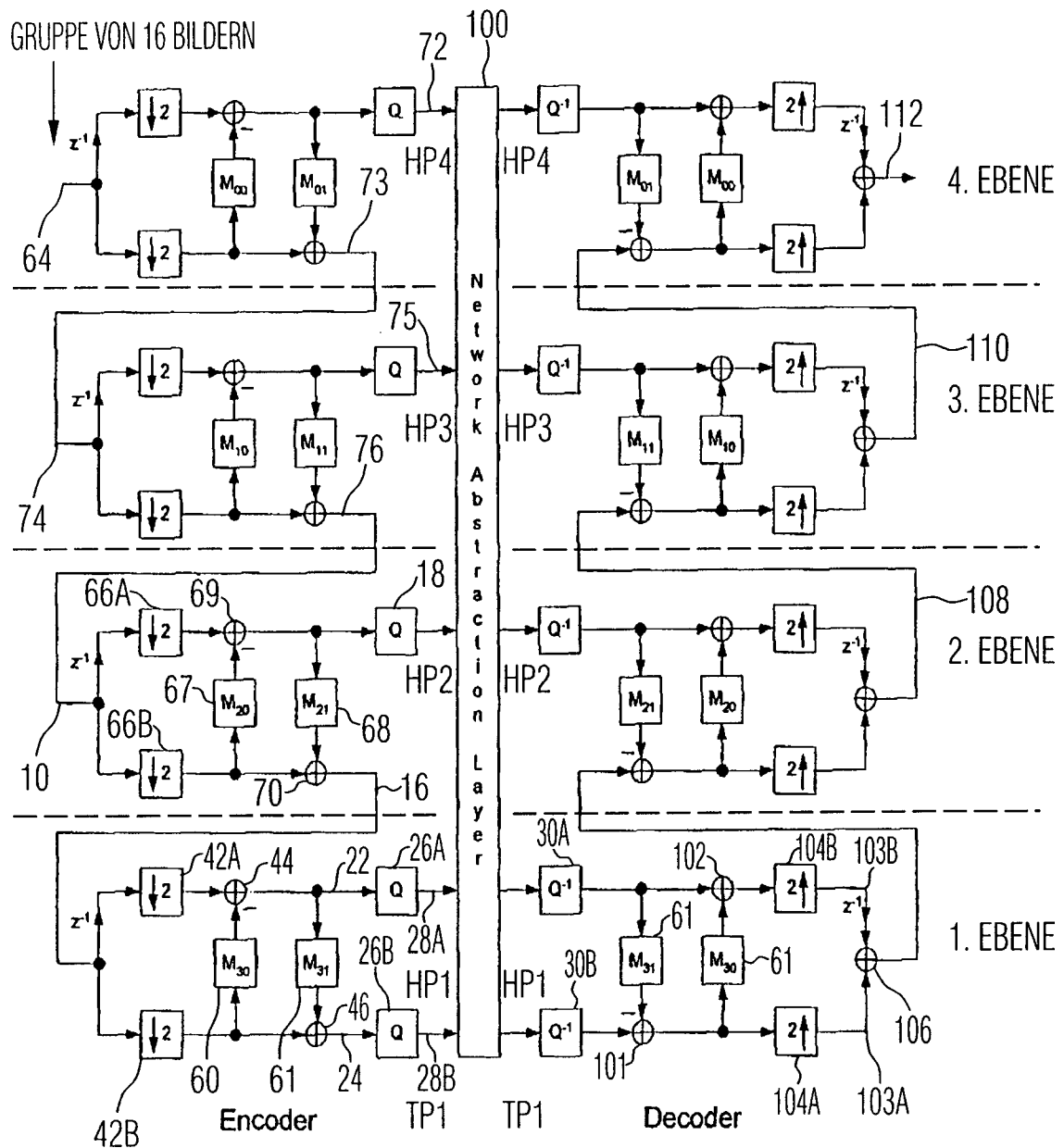


FIG 4

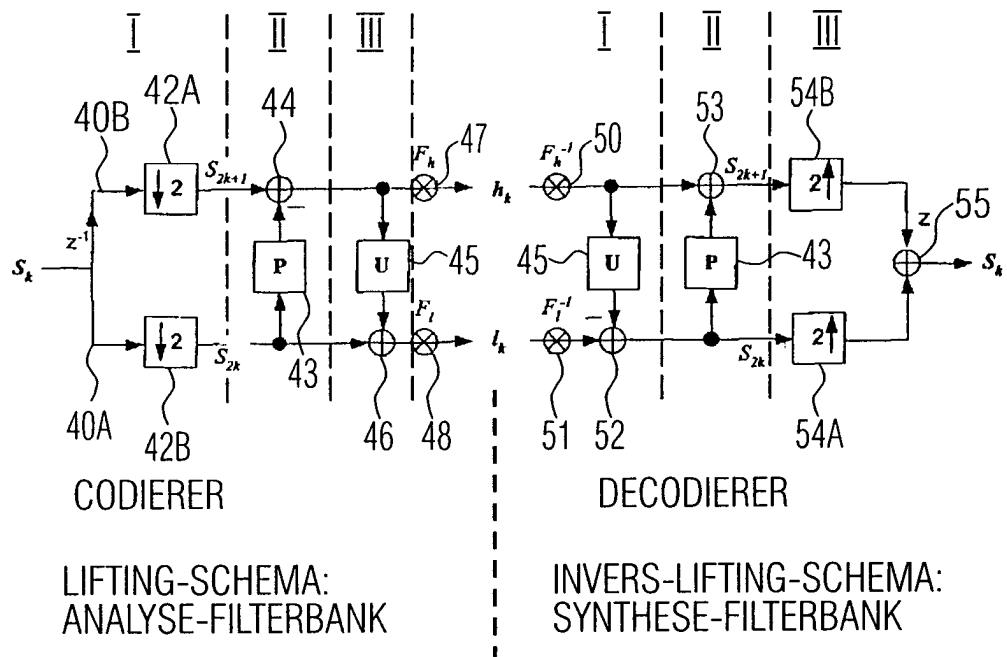
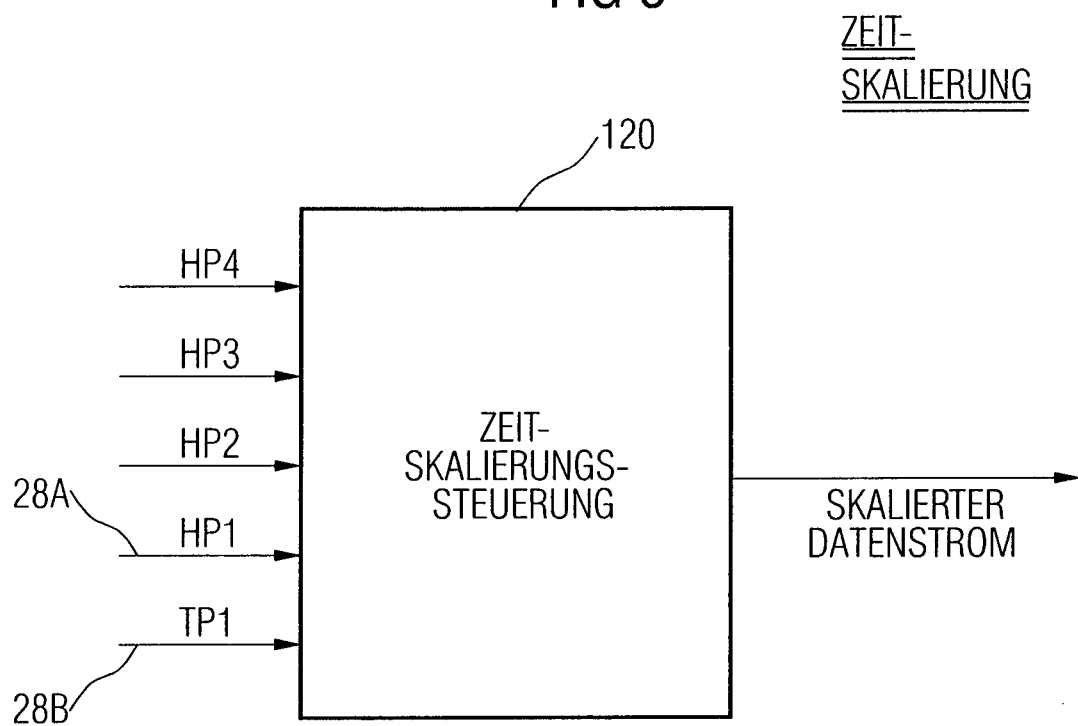


FIG 5



BASIS-SKALIERUNGSSCHICHT: TP1 + HP1 + SEITENINFOS

1. ERW. -SKAL.-SCHICHT:	HP2 + SEITENINFOS
2. ERW. -SKAL.-SCHICHT:	HP3 + SEITENINFOS
3. ERW. -SKAL.-SCHICHT:	HP4 + SEITENINFOS

FIG 6

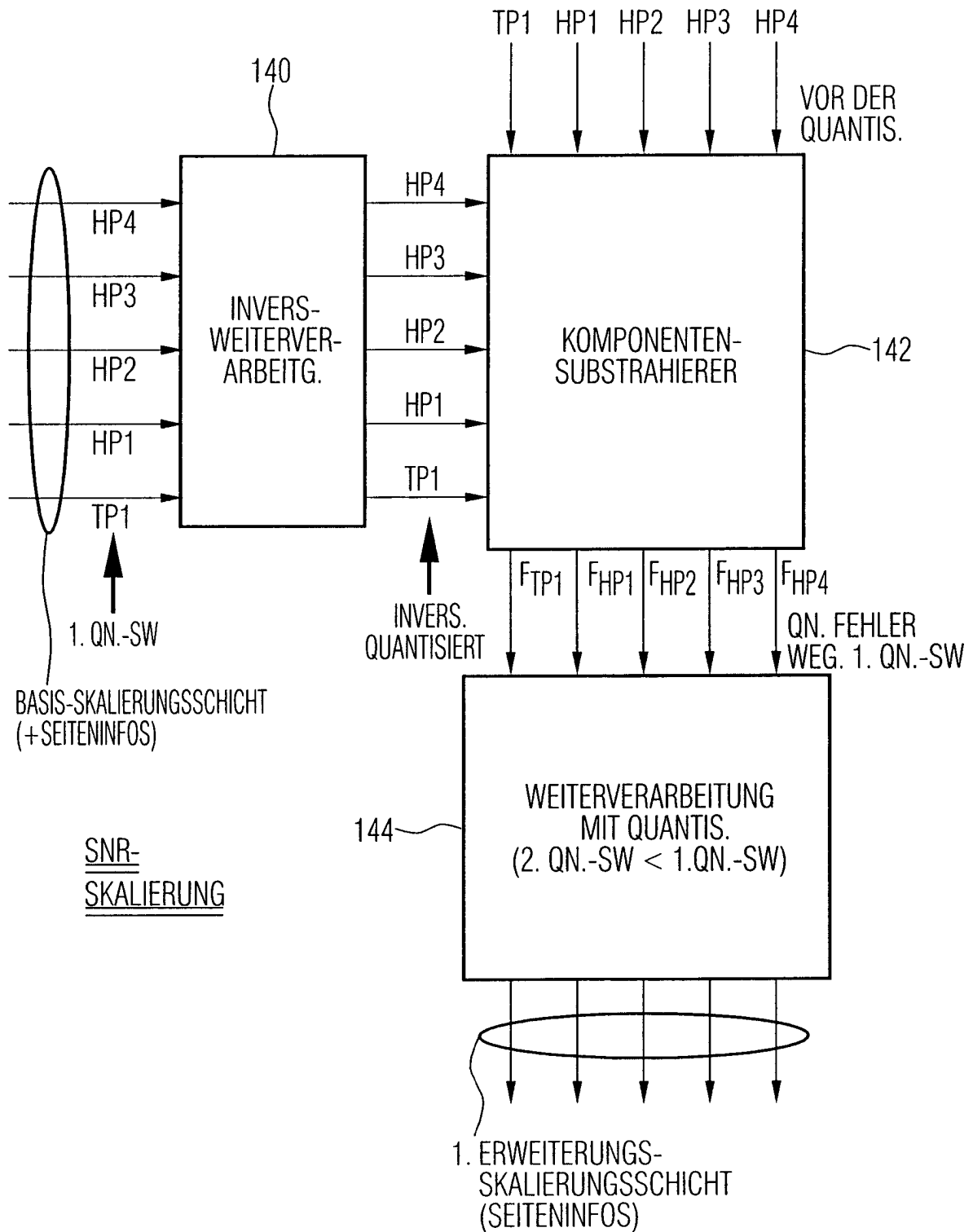


FIG 7

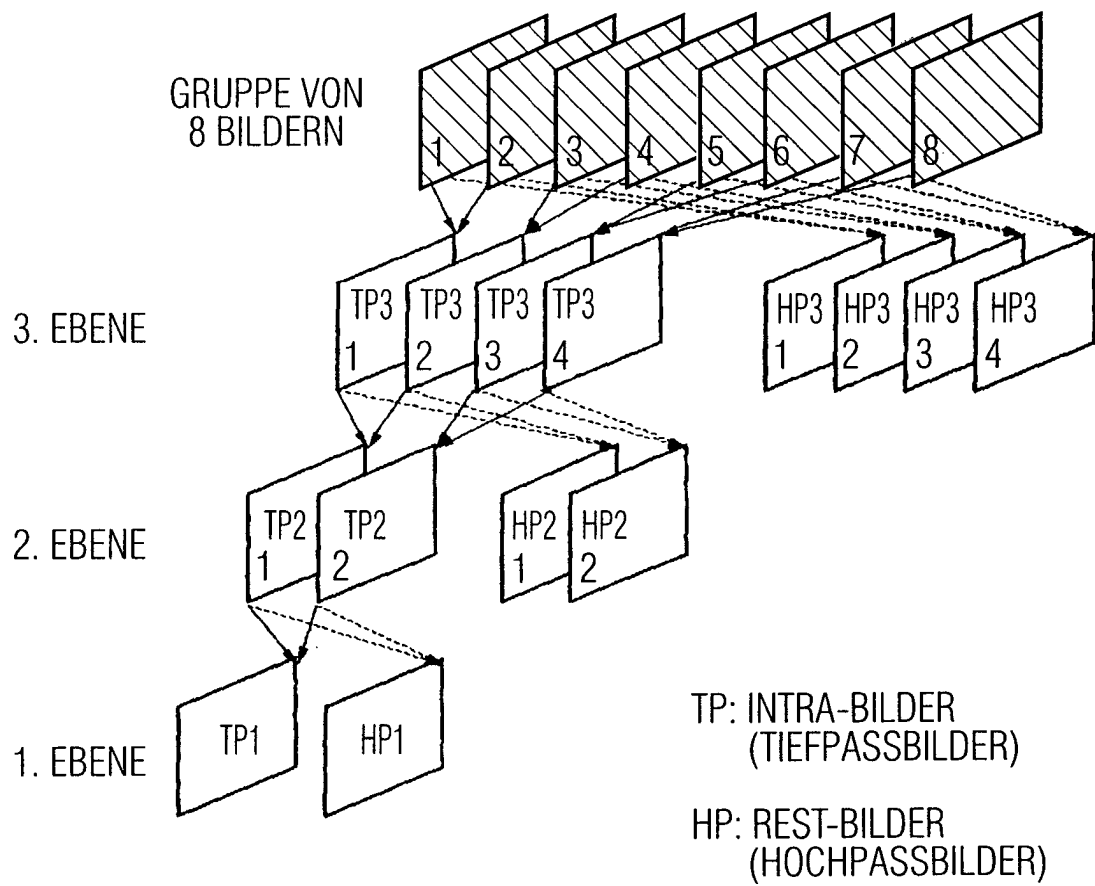


FIG 8

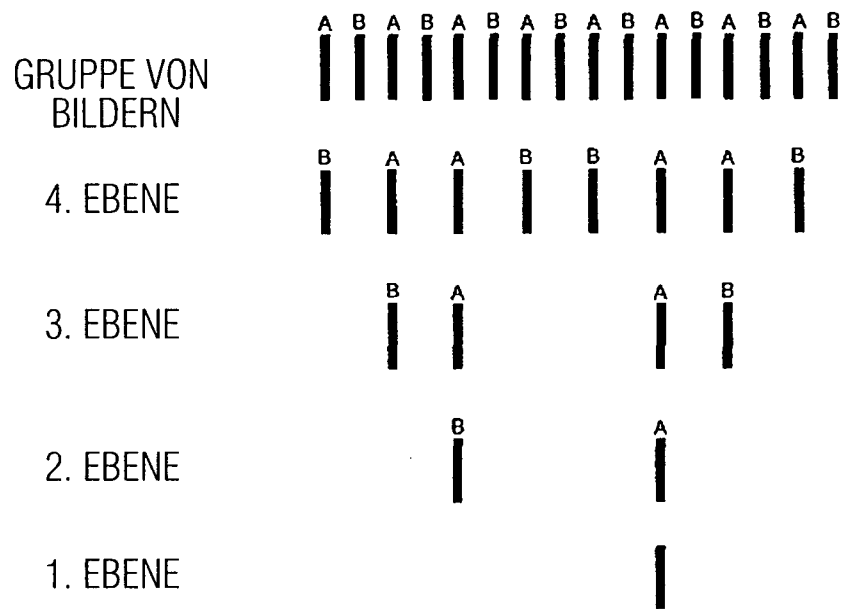


FIG 9 (STAND DER TECHNIK)

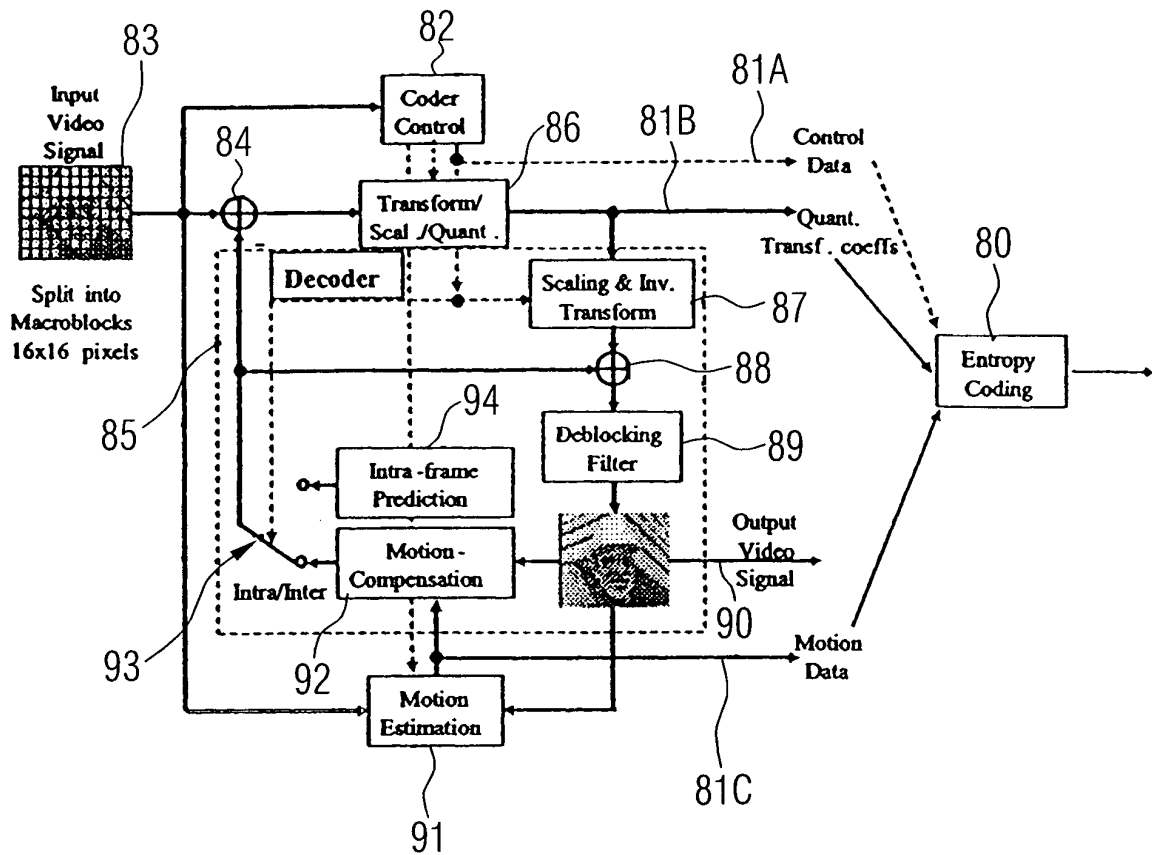


FIG 10 (STAND DER TECHNIK)

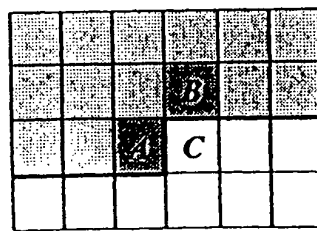


FIG 11 (STAND DER TECHNIK)

