

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2009-182776

(P2009-182776A)

(43) 公開日 平成21年8月13日(2009.8.13)

(51) Int.Cl.

H04N 7/32 (2006.01)

F I

H04N 7/137

Z

テーマコード (参考)

5C059

5C159

審査請求 未請求 請求項の数 36 O L (全 21 頁)

(21) 出願番号 特願2008-20545 (P2008-20545)
 (22) 出願日 平成20年1月31日 (2008.1.31)

(71) 出願人 000005108
 株式会社日立製作所
 東京都千代田区丸の内一丁目6番6号
 (74) 代理人 110000350
 ポレール特許業務法人
 (72) 発明者 黒田 雄樹
 神奈川県横浜市戸塚区吉田町292番地
 株式会社日立製作所組込みシステム基盤研
 究所内
 (72) 発明者 高橋 昌史
 神奈川県横浜市戸塚区吉田町292番地
 株式会社日立製作所組込みシステム基盤研
 究所内

最終頁に続く

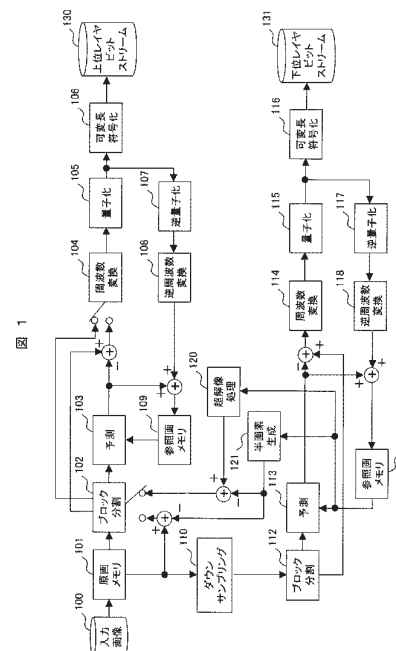
(54) 【発明の名称】 符号化装置、復号化装置、および、動画像符号化方法、動画像復号化方法

(57) 【要約】

【課題】 スケーラビリティを利用した符号化装置によるデータ伝送において、データ圧縮率を向上させ、スケーラブルなデータ構造により、伝送帯域の効率的な利用をおこなう。

【解決手段】 各々のレイヤ毎に、画像データを圧縮符号化する周波数変換部、量子化部、可変長符号化部と、予測のための画像を生成する逆周波数変換部、逆量子化部とを備え、画像の解像度を縮小するダウンサンプリング部と、画像の解像度を向上させる超解像処理部や半画素生成処理部を設ける。そして、入力画像をダウンサンプリング部により、ダウンサンプリングして、解像度の低い下位レイヤの周波数変換部に入力し、下位レイヤの予測のために用いられる画像を、超解像処理部や半画素生成処理部によって、超解像処理や半画素生成処理をおこなって、それらの生成画像または差分画像を、解像度の高い上位レイヤの周波数変換部に入力する。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

画像を入力して、複数の解像度の異なるレイヤからなるストリームとして、各々符号化して出力する符号化装置において、

各々のレイヤ毎に、画像データを圧縮する圧縮部と、

圧縮されたデータを伸張する伸張部とを備え、

前記画像の解像度を縮小するダウンサンプリング部と、

伸張された画像の解像度を向上させるアップサンプリング部とを有し、

入力された画像を前記ダウンサンプリング部により、ダウンサンプリングして、解像度の低い下位レイヤの圧縮部に入力し、

前記下位レイヤの圧縮部により、圧縮された画像を前記下位レイヤの伸張部に入力して、伸張された画像を前記アップサンプリング部により、アップサンプリングして、解像度の高い上位レイヤの圧縮部に入力することを特徴とする符号化装置。

10

【請求項 2】

前記アップサンプリング部は、アップサンプリング処理として超解像処理をおこなうことを特徴とする請求項 1 記載の符号化装置。

【請求項 3】

前記アップサンプリング部は、アップサンプリング処理として半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）をおこなうことを特徴とする請求項 1 記載の符号化装置。

20

【請求項 4】

前記アップサンプリング部は、アップサンプリング処理として超解像処理と、半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）とをおこない、

前記超解像処理をおこなった画像と、前記半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）とをおこなった画像との差分画像を、前記上位レイヤの圧縮部に入力することを特徴とする請求項 1 記載の符号化装置。

【請求項 5】

前記アップサンプリング部は、アップサンプリング処理として半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）をおこない、

前記上位レイヤの圧縮部に入力される元の画像と、前記半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）とをおこなった画像との差分画像を、前記上位レイヤの圧縮部に入力することを特徴とする請求項 1 記載の符号化装置。

30

【請求項 6】

前記アップサンプリング部は、アップサンプリング処理として超解像処理をおこない、

前記上位レイヤの圧縮部に入力される元の画像と、前記超解像処理をおこなった画像との差分画像を、前記上位レイヤの圧縮部に入力することを特徴とする請求項 1 記載の符号化装置

【請求項 7】

前記上位レイヤの圧縮部で、インター予測またはイントラ予測をおこなわないことを特徴とする請求項 4 ないし請求項 6 記載のいずれかの符号化装置。

40

【請求項 8】

前記上位レイヤの圧縮部で、インター予測またはイントラ予測をおこなって、上位レイヤのストリームを送信する場合、

前記上位レイヤの圧縮部で、インター予測またはイントラ予測をおこなわず、上位レイヤのストリームを送信する場合、

上位レイヤのストリームを送信しない場合の各々の画質に対する符号量のコストを算出し、

前記コストの低い場合の動作を選択することを特徴とする請求項 1 記載の符号化装置。

【請求項 9】

下位レイヤのストリームに対して、エラー耐性を強化して送信することを特徴とする請

50

求項 1 記載の符号化装置。

【請求項 10】

複数のレイヤのストレームのパケットを一つのヘッダ情報によりまとめて、同じ伝送路で送信することを特徴とする請求項 1 記載の符号化装置。

【請求項 11】

複数のレイヤのストレームを異なる伝送路で送信することを特徴とする請求項 1 記載の符号化装置。

【請求項 12】

前記各々のレイヤの圧縮部において、複数種類の画像ブロック分割サイズで画像信号を圧縮することを特徴とする請求項 1 記載の符号化装置。

10

【請求項 13】

前記異なるレイヤの圧縮部において、各々の圧縮部で異なった圧縮符号化方法により符号化されることを特徴とする請求項 1 記載の符号化装置。

【請求項 14】

前記異なるレイヤの圧縮部において、各々の圧縮部で異なった量子化パラメータにより量子化されることを特徴とする請求項 1 記載の符号化装置。

【請求項 15】

複数の解像度の異なるレイヤからなるストリームを入力して、各々解像度の異なる画像を出力する復号化装置において、

各々のレイヤ毎に、圧縮されたデータを伸張する伸張部を備え、

20

伸張された画像の解像度を向上させるアップサンプリング部を有し、

解像度の高い上位レイヤの伸張部により伸張された画像と、解像度の低い下位レイヤの伸張部により、伸張された画像を前記アップサンプリング部により、アップサンプリングして出力される画像とを加算して、画像を生成することを特徴とする復号化装置。

【請求項 16】

前記アップサンプリング部は、アップサンプリング処理として超解像処理をおこなうことを特徴とする請求項 15 記載の復号化装置。

【請求項 17】

前記アップサンプリング部は、アップサンプリング処理として半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）をおこなうことを特徴とする請求項 15 記載の復号化装置。

30

【請求項 18】

複数のレイヤの伸張部から出力される画像に対して、少なくとも一つの画像に関しては予測をおこなわずに復号することを特徴とする請求項 15 記載の復号化装置。

【請求項 19】

画像を入力して、複数の解像度の異なるレイヤからなるストリームとして、各々符号化して出力する動画像符号化方法において、

動画像符号化方法を実行する符号化装置は、

各々のレイヤ毎に、画像データを圧縮する圧縮部と、

圧縮されたデータを伸張する伸張部とを備え、

40

各々のレイヤ毎に、前記圧縮部により画像データを圧縮する手順と、

各々のレイヤ毎に、圧縮されたデータを前記伸張部に伸張して、前記圧縮部でおこなわれる予測のために用いる手順と、

入力された画像をダウンサンプリングして、解像度の低い下位レイヤの圧縮部に入力する手順と、

前記下位レイヤの圧縮部により、圧縮された画像を前記下位レイヤの伸張部に入力して、伸張された画像をアップサンプリングして、解像度の高い上位レイヤの圧縮部に入力する手順とを有することを特徴とする動画像符号化方法。

【請求項 20】

前記伸張された画像をアップサンプリングする処理は、超解像処理であることを特徴と

50

する請求項 19 記載の動画像符号化方法。

【請求項 21】

前記伸張された画像をアップサンプリングする処理は、半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）であることを特徴とする請求項 19 記載の動画像符号化方法。

【請求項 22】

前記伸張された画像をアップサンプリングする処理は、超解像処理と、半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）であり、

前記超解像処理をおこなった画像と、前記半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）とをおこなった画像との差分画像を、前記上位レイヤの圧縮部に入力することを特徴とする請求項 19 記載の動画像符号化方法。

10

【請求項 23】

前記伸張された画像をアップサンプリングする処理は、半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）であり、

前記上位レイヤの圧縮部に入力される元の画像と、前記半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）とをおこなった画像との差分画像を、前記上位レイヤの圧縮部に入力することを特徴とする請求項 19 記載の動画像符号化方法。

【請求項 24】

前記伸張された画像をアップサンプリングする処理は、超解像処理であり、

前記上位レイヤの圧縮部に入力される元の画像と、前記超解像処理をおこなった画像との差分画像を、前記上位レイヤの圧縮部に入力することを特徴とする請求項 19 記載の動画像符号化方法

20

【請求項 25】

前記上位レイヤの圧縮部で、インター予測またはイントラ予測をおこなわないことを特徴とする請求項 19 ないし請求項 24 記載のいずれかの動画像符号化方法。

【請求項 26】

前記上位レイヤの圧縮部で、インター予測またはイントラ予測をおこなって、上位レイヤのストリームを送信する場合、

前記上位レイヤの圧縮部で、インター予測またはイントラ予測をおこなわず、上位レイヤのストリームを送信する場合、

30

上位レイヤのストリームを送信しない場合の各々の画質に対する符号量のコストを算出し、

前記コストの低い場合の動作を選択することを特徴とする請求項 19 記載の動画像符号化方法。

【請求項 27】

下位レイヤのストリームに対して、エラー耐性を強化して送信することを特徴とする請求項 19 記載の動画像符号化方法。

【請求項 28】

複数のレイヤのストリームのパケットを一つのヘッダ情報によりまとめて、同じ伝送路で送信することを特徴とする請求項 19 記載の動画像符号化方法。

40

【請求項 29】

複数のレイヤのストリームを異なる伝送路で送信することを特徴とする請求項 19 記載の動画像符号化方法。

【請求項 30】

前記各々のレイヤの圧縮部において、複数種類の画像ブロック分割サイズで画像信号を圧縮することを特徴とする請求項 19 記載の動画像符号化方法。

【請求項 31】

前記異なるレイヤの圧縮部において、各々の圧縮部で異なった圧縮符号化方法により符号化されることを特徴とする請求項 19 記載の動画像符号化方法。

【請求項 32】

50

前記異なるレイヤの圧縮部において、各々の圧縮部で異なった量子化パラメータにより量子化されることを特徴とする請求項 19 記載の動画像符号化方法。

【請求項 33】

複数の解像度の異なるレイヤからなるストリームを入力して、各々解像度の異なる画像を出力する動画像復号化方法において、

動画像復号化方法を実行する復号化装置は、各々のレイヤ毎に、圧縮されたデータを伸張する伸張部を備え、

伸張された画像の解像度を向上させるアップサンプリング部を有し、

解像度の高い上位レイヤの伸張部により伸張された画像と、解像度の低い下位レイヤの伸張部により、伸張された画像をアップサンプリングして出力される画像とを加算して、
画像を生成する手順を有することを特徴とする動画像復号化方法。

10

【請求項 34】

前記伸張された画像をアップサンプリングする処理は、超解像処理であることを特徴とする請求項 33 記載の動画像復号化方法。

【請求項 35】

前記アップサンプリング部は、アップサンプリング処理として半画素生成処理、または、 $1/n$ 画素生成処理（ n は 2 より大きい整数）をおこなうことを特徴とする請求項 33 記載の動画像復号化方法。

【請求項 36】

複数のレイヤの伸張部から出力される画像に対して、少なくとも一つの画像に関しては
予測をおこなわずに復号することを特徴とする請求項 33 記載の動画像復号化方法。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、符号化装置、復号化装置、および、動画像符号化方法、動画像復号化方法に係り、特に、階層の異なるデータを同時に送信して、データの特性に応じて利用するスケラビリティに応用して、データ伝送効率を向上させるのに好適な符号化装置、復号化装置、および、動画像符号化方法、動画像復号化方法に関する。

【背景技術】

【0002】

今日、伝送帯域の限られた状況、また、データを受信する端末の種類が複数存在する状況において、映像データや音声データを伝送するために、スケラビリティと呼ばれる機能が導入されている。このスケラビリティは、ビットストリームを部分的に復号することによって S N R (Signal to Noise Ratio)、空間解像度、時間方向の解像度などを段階的に可変とする機能であり、M P E G - 2 や M P E G - 4、H. 264 の拡張規格 S V C といった動画像圧縮規格に採用されている。スケラビリティの実現には、階層符号化と呼ばれる手法が用いられている。例えば、M P E G - 2 では、S N R スケラブルプロファイル、空間スケラブルプロファイル、H i g h プロファイルが定義されている。

30

【0003】

伝送帯域の限られた状況において、単純にそれぞれの階層を別々に符号化して伝送する手法（サイマルキャスト）と比較して、少しでも符号化効率を向上させ、符号化に必要なヘッダ情報やサイド情報の重複を避けることは重要である。スケラビリティは、各階層間でこれらの情報の相互利用をおこなうことによって符号化効率を向上する。なお、スケラビリティについては、以下の非特許文献 1 に解説がある。

40

【0004】

S N R スケラブルプロファイルは、後述する S N R スケラビリティと呼ばれる技術を採用し、グレースフルデグラデーション機能を実現する。グレースフルデグラデーションとは、例えば、電波の受信強度に応じた緩やかな画質の劣化を実現する機能である。通常のデジタル放送において、電波の受信強度がある規定値以下になった場合、それを境界としてまったく放送が受信できなくなるという問題があるが、放送に S N R スケラビ

50

リティを適用した場合は、画像の品質が緩やかに劣化することで、受信状態に応じた画質の再生が可能となる。

【0005】

S N Rスケラビリティは、周波数領域（M P E G - 2ではD C T係数）の階層符号化を実現する。基本的な考え方は、ベースとなる画質（基準となる低S N R）を提供する信号を高い優先度の伝送路で送り出し、さらに空間解像度は同一で、ベースの画質を向上させるための補助信号（エンハンスメント信号）を低い優先度の伝送路で送り出す。これにより、両方の信号を受信できる地域では高画質な復号が可能となる。また、受信状態が悪化しエンハンスメント信号を受け取れない場合でも、ベースとなる信号は高優先度で伝送、受信されるため、基準となる画質を再生する信頼度を向上させることができる。

10

【0006】

ここで、図7および図8を用いてS N Rスケラビリティの符号化と復号化の例について説明する。

図7は、S N Rスケラビリティ符号化装置の構成例を示す図である。

図8は、S N Rスケラビリティ復号化装置の構成例を示す図である。

【0007】

図7に示されるように、符号化装置に、入力画像500を入力し、低S N Rビットストリームを生成するために、M C部507で生成される予測画像との差分画像データをD C T部501へ入力する。ここで、予測画像は、逆量子化部504、I D C T部505、F M（フレームメモリ）部506、M C（Motion Compensation）部507とで形成されるループ処理によって生成される予測のための画像である。

20

【0008】

D C T部501に入力されたデータに対し、D C T処理部501でD C Tをおこなった後、量子化部502において粗く量子化し、V L C部503によって低S N Rのビットストリーム541を生成する。また、量子化部502の出力を逆量子化部504で逆量子化し、粗く量子化したD C T係数を生成する。このD C T係数をD C T部511が出力するD C T係数との差分をとり、量子化部512へ入力する。また、逆量子化部504の出力を逆量子化部514の出力に足しこみ、I D C T部515へ入力する。逆量子化部504とI D C T部505とF M（フレームメモリ）部506とM C（Motion Compensation）部507とで形成されるループ処理は、M P E G - 2におけるローカルデコードであり、デコーダ側で復元される画像をエンコーダ側でも生成し、復元された画像は、次フレームの予測をおこなうために用いられる。

30

【0009】

一方で、エンハンスメントストリーム540については、D e l a y部510へ入力画像500を入力したのち、M C部517で生成される予測画像との差分をとり、差分画像がD C T部511に入力される。ここで、予測画像は、逆量子化部514、I D C T部515、F M（フレームメモリ）部516、M C（Motion Compensation）部517とで形成されるループ処理によって生成される予測のための画像である。

【0010】

D C T部511に入力された画像は、D C T部511により周波数変換される。そして、D C T部511によって出力されたD C T係数から、I Q部504の出力するD C T係数の差分をとり、それを量子化部512へ入力する。ここで生成された差分信号は、低S N R信号を足しこむことによって、高S N R信号を得るための信号である。そして、量子化部512で量子化されたデータをV L C部513で可変長符号化することにより、エンハンスメントビットストリーム540を生成する。また、予測のために、I D C T部515とF M（フレームメモリ）部516とM C（Motion Compensation）部517によって、エンハンスメント画像データのローカルデコードをおこない、M C部517から出力されるデータを、D C T部511に入力する。

40

【0011】

S N Rスケラビリティの復号化の場合には、図8に示されるS N Rスケラビリティ

50

復号装置により、低S N Rビットストリーム540をV L D部600によって復号処理をおこない、I Q部601によって逆量子化をおこなう。一方、エンハンスメントビットストリーム541をV L D部602によって復号処理をおこない、I Q部603によって逆量子化をおこなう。そして、I Q部601から出力されるD C T係数とI Q部603から出力されるD C T係数を足し合わせ、I D C T部604によってI D C T処理をおこなう。以降は、シングルレイヤの場合と同様に復号をおこなう。

【0012】

空間スケーラブルプロファイルは、例えば、S D T V (Standard Definition Television) とH D T V (High Definition Television) など、解像度の異なる画像間における表示の両立を実現する手法である。

10

【0013】

空間スケーラビリティは、解像度の異なる階層間での画像の相似性を利用して、符号化効率を高めるための技術である。ここで、図9および図10を用いて空間スケーラビリティの符号化と復号化の例について説明する。

図9は、空間スケーラビリティ符号化装置の構成例を示す図である。

図10は、空間スケーラビリティ復号化装置の構成例を示す図である。

【0014】

図9に示されるように、空間スケーラビリティ符号化装置では、まず、下位レイヤのビットストリームを生成するために、入力画像500をダウンサンプリング部700でダウンサンプリングする。この画像とM C部507によって生成された予測画像の差分をとり、D C T部501に入力する。ここで、図8の場合と同様、予測画像は、逆量子化部504、I D C T部505、F M (フレームメモリ)部506、M C (Motion Compensation)部507とで形成されるループ処理によって生成される予測のための画像である。

20

【0015】

D C T部501に入力されたデータに対し、D C T処理部501でD C Tをおこなった後、量子化部502において粗く量子化をおこない、その後、V L C部503で可変長符号化をおこない、下位レイヤビットストリーム711を伝送する。また、量子化部502の出力を、I Q部504で逆量子化、I D C T部505でI D C T処理したものを、アップサンプリング部701で入力画像と同じ解像度に変換し、この画像を上位レイヤの動き補償における2枚の参照画像と共に、第3の参照画像として用いる。この下位レイヤからの予測を、空間予測と呼ぶ。空間予測参照画像は、上位レイヤで符号化する画像と同一の時間における画像のため、動き補償の必要はない。ここで、M C部517の出力である動き補償画像とアップサンプリングフィルタ部701の出力である空間予測画像は、0から1までの値をとる変数Wを用いて、それぞれ(W)および(1-W)の重み付けで加算され、上位レイヤの予測に利用される。

30

【0016】

空間スケーラビリティの復号をおこなう場合、図10に示される空間スケーラビリティ復号化装置では、下位レイヤビットストリーム711をV L D部800とI Q部801とI D C T部802で復号をおこなう。また、下位レイヤの復号画像をアップサンプリング部805でアップサンプリングし、上位レイヤビットストリームの復号に利用する。上位レイヤビットストリーム710は、V L D部810とI Q部811とI D C T部812で復号をおこない、下位レイヤの復元画像をアップサンプリングした画像と上位レイヤで動き補償された画像をそれぞれ参照画像として、符号化伝送された変数Wから、(W)、(1-W)の重み付け加算をおこなうことにより、高品質で高解像度の画像を生成する。

40

【0017】

さらに、このS N Rスケーラビリティと空間スケーラビリティを組み合わせ利用できるH i g hプロファイルも存在する。

【0018】

また、最近のテレビ受像機は大画面化が進んでおり、放送や通信、蓄積媒体などから入力された画像信号をそのまま表示するのではなく、デジタル信号処理によって水平・垂直

50

方向の画素数を増やして表示することが一般的におこなわれている。この際、一般的に知られている sinc 関数を用いた補間ローパスフィルタやスプライン関数等によって画素数を増やすだけでは解像度を上げることはできない。

【0019】

そこで、特許文献1、特許文献2、非特許文献2に記載されているように、入力された複数の画像フレームを合成して1枚のフレームとすることにより、高解像度化しながら画素数を増やす技術（以下、「超解像処理」という）が提案されている。

【0020】

【特許文献1】特開平8-336046号公報

【特許文献2】特開平9-69755号公報

【非特許文献1】テレビジョン学会編 「総合マルチメディア選書MPEG」、平成8年4月20日、5-6 スケーラビリティ、p.117-132

【非特許文献2】青木伸 “複数のデジタル画像データによる超解像処理”，Ricoh Technical Report pp.19-25, No.24, NOVEMBER, 1998

【非特許文献3】Gary J. Sullivan and Thomas Wiegand, "Rate-Distortion Optimization for Video Compression", IEEE Signal Processing Magazine, vol. 15, no. 6, pp.74-90, November, 1998.

【発明の開示】

【発明が解決しようとする課題】

【0021】

前述のとおり、複数種類の受信端末向けに複数解像度のデータのサイマルキャストイングをおこなう場合、限られた伝送帯域を大幅に占有し、伝送効率が悪いという問題点がある。

【0022】

また、従来のスケーラビリティを利用した伝送方法でも、よりデータ圧縮率を向上させ、伝送効率をよくしたいという要請がある。

【0023】

本発明は、上記問題点を解決するためになされたもので、その目的は、スケーラビリティを利用した符号化装置、および、復号化装置によるデータ伝送において、データ圧縮率を向上させ、スケーラブルなデータ構造により、伝送帯域の効率的な利用をおこなうことにある。

【課題を解決するための手段】

【0024】

本発明の符号化装置は、画像を入力して、複数の解像度の異なるレイヤからなるストリームとして、各々符号化して出力する符号化装置において、各々のレイヤ毎に、画像データを圧縮符号化する周波数変換部、量子化部、可変長符号化部と、予測のための画像を生成する逆周波数変換部、逆量子化部とを備え、画像の解像度を縮小するダウンサンプリング部と、画像の解像度を向上させる超解像処理部や半画素生成処理部を設ける。そして、入力画像をダウンサンプリング部により、ダウンサンプリングして、解像度の低い下位レイヤの周波数変換部に入力し、下位レイヤの予測ために用いられる画像を、超解像処理部や半画素生成処理部によって、超解像処理や半画素生成処理をおこなって、それらの生成画像または差分画像を、解像度の高い上位レイヤの周波数変換部に入力する。

【0025】

また、本発明の復号化装置では、複数の解像度の異なるレイヤからなるストリームを入力して、各々解像度の異なる画像を出力する復号化装置において、各々のレイヤ毎に、圧縮されたデータを復号する逆周波数変換部、逆量子化部を備え、下位レイヤのストリームから生成され、復号の際に予測に用いられる画像を、超解像処理部や半画素生成処理部によって、超解像処理や半画素生成処理をおこなって、それらの生成画像を、解像度の高い上位レイヤのストリームから生成された画像と加えた画像を生成する。

【発明の効果】

【0026】

本発明によれば、スケーラビリティを利用した符号化装置、および、復号化装置によるデータ伝送において、データ圧縮率を向上させ、スケーラブルなデータ構造により、伝送帯域の効率的な利用をおこなうことにある。

【発明を実施するための最良の形態】

【0027】

以下、本発明に係る各実施形態を、図1ないし図6を用いて説明する。

【0028】

〔実施形態1〕

以下、本発明の第一の実施形態に係る映像の符号化装置と復号化装置を、図1ないし図4を用いて説明する。

10

【0029】

先ず、図2を用いて本発明の第一の実施の形態に係る映像の符号化装置のハードウェア構成について説明する。

図2は、本発明の第一の実施の形態に係る映像の符号化装置のハードウェア構成について説明するブロック図である。

【0030】

本実施形態の映像の符号化装置201は、図2に示されるように、相互に通信可能に接続されたプロセッサ202、メモリ203、入力インターフェース（入力I/F）204および出力インターフェース（出力I/F）206からなる。入力I/F204は、入力装置205に接続されており、また、出力I/F206は、出力装置207に接続されている。

20

【0031】

プロセッサ202は、本実施形態の映像の符号化の処理をおこなうプロセッサであり、メモリ203に格納されたプログラムを実行して、入力I/F204から受けたデータを符号化し、出力I/F206に送る。メモリ203には、プロセッサ202が実行するプログラムが格納される。また、プロセッサ202が処理する対象となるデータが一時的に格納される。映像の符号化装置201には、プロセッサ202およびメモリ203が複数設けられてもよい。例えば、本発明の映像の符号化の処理をおこなうプログラムの一部のみを実行する専用のプロセッサが設けられてもよい。また、同一の処理をおこなう専用のプロセッサが複数設けられてもよい。

30

【0032】

プロセッサ202およびメモリ203は、単一のチップ上に実装されてもよい。

【0033】

入力I/F204は、プロセッサ202が処理する映像データを入力装置205から受信するインターフェースである。

【0034】

入力装置205は、映像の符号化装置201が処理する映像信号を入力I/F204に入力する装置であり、例えば、レンズや映像撮像素子である。この場合、入力I/F204は、例えば、A/Dコンバータである。また、入力装置205は、符号化されていない映像データが格納された記憶装置であってもよい。この場合、入力I/F204は、例えば、SATA（Serial ATA Attachment）インターフェースである。

40

【0035】

出力I/F206は、プロセッサ202によって符号化されたデータを出力装置207に送信するインターフェースである。

【0036】

出力装置207は、映像符号化装置201が符号化したデータの出力先であり、例えば、符号化したデータを格納する記憶装置である。この場合、出力I/F206は、例えば、SATAインターフェースである。また、出力装置207は、LANやIPネットワーク等（図示省略）を介して出力I/F206と接続されたコンピュータ装置であってもよい。

50

い。この場合、出力 I / F 2 0 6 は、ネットワークインターフェースである。また、出力装置 2 0 7 は、電話通信網（図示省略）を介して出力 I / F 2 0 6 と接続された受像装置であってもよい。この場合、出力 I / F 2 0 6 は、電話信号の送信機である。また、出力装置 2 0 7 は、デジタル T V 受像機であってもよい。この場合、出力 I / F 2 0 6 は、デジタル T V 信号の送信機が処理するインターフェースになる。

【0037】

映像の符号化装置 2 0 1 は、入力 I / F 2 0 4 および出力 I / F 2 0 6 をそれぞれ複数備え、それぞれの入力 I / F 2 0 4 および出力 I / F 2 0 6 に、異なる種類の入力装置および出力装置が接続されてもよい。例えば、映像符号化装置 2 0 1 は、二つの出力 I / F 2 0 6 を備え、一方には、ハードディスク装置が接続され、もう一方には光磁気ディスク装置が接続されてもよい。また、一方には、ハードディスク装置が接続され、もう一方には L A N 等を介してコンピュータ装置が接続されてもよい。

10

【0038】

次に、図 1 を用いて本発明の第一の実施形態に係る映像の符号化装置の構成と動作について説明する。

図 1 は、本発明の第一の実施形態に係る映像の符号化装置の構成を示すブロック図である。

【0039】

映像の符号化装置は、図 1 に示されるように、原画メモリ 1 0 1 と、ブロック分割部 1 0 2 および 1 1 2 と、予測部 1 0 3 および 1 1 3 と、周波数変換部 1 0 4 および 1 1 4 と、量子化部 1 0 5 および 1 1 5 と、可変長符号化部 1 0 6 および 1 1 6 と、量子化部 1 0 7 および 1 1 7 と、逆周波数変換部 1 0 8 および 1 1 8 と、参照画メモリ 1 0 9 および 1 1 9 と、ダウンサンプリング部 1 1 0 と、超解像処理部 1 2 0 と、半画素生成部 1 2 1 を備える。

20

【0040】

原画メモリ 1 0 1 は、入力された原画像を一時的な保持するメモリである。ブロック分割部 1 0 2 および 1 1 2 は、画像の一部をブロック単位で処理をおこなうために画像を複数の画像ブロックに分割する。予測部 1 0 3 および 1 1 3 は、同一フレーム内（イントラ予測）または前後のフレーム間（インタラ予測）での予測をおこなう。周波数変換部 1 0 4 および 1 1 4 は、画像信号に対して、例えば、D C T（Discrete Cosine Transform：離散コサイン変換）などの周波数変換をおこなう。量子化部 1 0 5 および 1 1 5 は、入力される周波数係数に対して量子化処理をおこなう。可変長符号化部 1 0 6 および 1 1 6 は、量子化されたデータに対してビット数を圧縮するための可変長符号化をおこなう。

30

【0041】

逆量子化部 1 0 7 および 1 1 7 は、量子化とは逆の逆量子化処理をおこなう。逆周波数変換部 1 0 8 および 1 1 8 は、周波数係数から画像を得る逆周波数変換をおこなう。参照画メモリ 1 0 9 および 1 1 9 は、参照画像を保持するための一時的なメモリである。

【0042】

ダウンサンプリング部 1 1 0 は、入力画像を下位レイヤのビットストリームのためもの符号化をおこなうためにダウンサンプリング処理する。超解像処理部 1 2 0 は、複数のフレーム画像から高解像度画像を生成する超解像処理をおこなう。半画素生成部 1 2 1 は、半画素精度の画素を生成する半画素生成処理をおこなう。超解像処理部 1 2 0 と半画素生成部 1 2 1 は、下位レイヤのストリーム生成のための画像を、上位レイヤのストリームを生成するためにアップサンプリングする処理をおこなうアップサンプリング部を形成する。

40

【0043】

なお、これら各部はハードウェアによって構成されてもよいし、ソフトウェアによって構成されていてもよい。また、ハードウェアとソフトウェアとを組み合わせたモジュールであってもよい。また、周波数変換部の処理は特に限定されず、D C T、D S T（Discrete Sine Transform：離散サイン変換）、D F T（Discrete Fourier Transform：離散フ

50

ーリエ変換)、ウェーブレット変換、整数変換処理等がおこなわれてもよい。

【0044】

ブロック分割部102、予測部103、周波数変換部104、量子化部105は、可変長符号化部106は、上位レイヤビットストリームを生成するための圧縮機能を実現する一つの圧縮部であり、一方、ブロック分割部112、予測部113、周波数変換部114、量子化部115、可変長符号化部116は、上位レイヤビットストリームを生成するための圧縮機能を実現する一つの圧縮部である。

【0045】

また、逆量子化部107、逆周波数変換部108、参照画メモリ109は、上位レイヤビットストリームを生成するための予測を実現するための伸張をおこなう一つの伸張部であり、逆量子化部117、逆周波数変換部118、参照画メモリ119は、下位レイヤビットストリームを生成するための予測を実現するための伸張をおこなう一つの伸張部である。

10

【0046】

ここで、ブロック分割部102と112、予測部103と113、周波数変換部104と114、量子化部105と115、可変長符号化部106と116、逆量子化部107と117、逆周波数変換部108と118、参照画メモリ109と119はそれぞれ一つのハードウェアで構成され、時分割で使用されてもよい。

【0047】

また、半画素生成部121は、予測部103や予測部113内に存在する半画素生成部を用いて処理されてもよい。

20

【0048】

次に、図1に示した映像の符号化装置の動作について説明する。

【0049】

原画メモリ101に格納された入力画像100は、ダウンサンプリング部110に入力される。ダウンサンプリング部110は、入力画像に対してローパスフィルタ処理をおこない、その後、画素の間引き処理をおこなう。間引き処理がおこなわれた画像は、下位レイヤのブロック分割部112に入力され、複数の画像ブロック単位に分割される。分割された画像ブロックは、予測部113によってフレーム内予測(イントラ予測)またはフレーム間予測(インター予測)がおこなわれ、予測画像が作成される。画面間予測をおこなう場合は、参照画メモリ119に格納されている画像データが参照される。

30

【0050】

そして、ブロック分割部112の画像と予測部113で作成された予測画像の差分をとった画像が周波数変換部114に入力される。周波数変換部114は、入力画像を画像ブロック毎に周波数変換する。ここで、周波数変換処理はDCT処理やDST処理やDFT処理やウェーブレット変換処理であってもよく、また、整数変換処理であってもよい。周波数変換された係数データは、量子化部115に入力される。量子化部115は、量子化パラメータを用いて、入力された周波数変換データを量子化する。量子化されたデータは、可変長符号化部116に入力される。可変長符号化部116は、入力データを動き補償情報や符号化モード情報等のサイド情報とともに可変長符号化し、下位レイヤビットストリーム131を作成する。下位レイヤビットストリーム131は、図2に示されるバッファとしての役割を果たすメモリ203に一旦格納される。そして、メモリ203から、入力された符号化データが、伝送回線速度にあわせて出力される。なお、このとき、後述する上位レイヤビットストリーム130と同じ伝送線路で伝送されてもよいし、別々の伝送線路によって伝送されてもよい。ここで、伝送線路とは無線による伝送や有線による伝送を問わず、また、ハードディスクドライブやDVDといった記録媒体を介して伝送されてもよい。さらに、下位レイヤビットストリーム131と上位レイヤビットストリーム130のパケットは、一つのヘッダ情報に対してまとめて送信されてもよく、また、下位レイヤビットストリーム131と上位レイヤビットストリームの一部が交互に送信されてもよい。

40

50

【0051】

一方、量子化部115によって生成された量子化データは、逆量子化部117に入力される。逆量子化部117は、量子化データを逆量子化する。逆量子化された係数データは、逆周波数変換部118に入力される。逆周波数変換部118は、このデータを逆周波数変換して差分ブロック画像として復元する。復元された差分ブロック画像は、予測部113から出力される予測画像と加算され、局部復号画像ブロックとして参照画メモリ119に格納される。そして、この局部復号画像ブロックが予測部113に入力される。局部復元画像ブロックは、また超解像処理部120と半画素生成部121に入力される。ここで、超解像処理部120と半画素生成部121と予測部113に入力される局部復号画像ブロックは、必ずしも参照画メモリ119から出力される必要はなく、参照画メモリに格納されるデータが直接入力されてもよい。

10

【0052】

超解像処理部120は、入力された画像を複数フレーム使用し、高解像度画像を生成する。ここで、本実施形態において超解像処理部の処理内容は、必ずしも従来技術の所で示したような超解像処理である必要はなく、入力される画像の解像度を向上させる処理であれば、処理内容を特に限定しない。

【0053】

一方、半画素生成部121は入力される局部復号画像ブロックに対し、半画素生成処理によって半画素精度の画像を生成する。ただし、半画素生成処理の代わりに1/4画素生成処理、1/8画素生成処理、…のように、それより画素の精度が細くなる1/n画素生成処理（nは2より大きい整数）がおこなわれてもよい。また、超解像処理の解像度に合わせた、解像度を向上させる処理であれば、特に処理内容は限定されない。ここで、半画素生成処理とは画素データに対してフィルタ処理をおこなうことで、1画素の2倍の精度を持つ画像を生成する処理を指し、具体的にはMPEG-2やMPEG-4やH.264/AVCにおける補間フィルタなどを意味する。より詳細な値を示すならば、MPEG-4のASPにおいて整数画素から半画素生成をおこなう場合の8タップのフィルタ係数は $\begin{bmatrix} -2 & 6 & -12 & 40 & 40 & -12 & 6 & -2 \end{bmatrix} / 64$ であり、半画素から1/4画素生成をおこなう場合の2タップのフィルタ係数は $\begin{bmatrix} 1 & 1 \end{bmatrix} / 2$ である。また、H.264/AVCにおいて整数画素から半画素生成をおこなう場合の6タップのフィルタ係数は $\begin{bmatrix} 1 & -5 & 20 & 20 & -5 & 1 \end{bmatrix} / 32$ であり、半画素から1/4画素生成をおこなう場合の2タップのフィルタ係数は $\begin{bmatrix} 1 & 1 \end{bmatrix} / 2$ である。

20

30

【0054】

ここで、生成される半画素精度画像（場合によっては、それより精度が細くなる1/n画素生成精度画像（nは2より大きい整数））と、超解像処理部120の出力画像の差分をとり、その差分データをブロック分割部102へ入力する。または、原画メモリ101に格納された入力画像と半画素生成部121で生成した半画素精度画像の差分をとり、その差分データをブロック分割部102へ入力する。

【0055】

ブロック分割部102は、超解像処理部120の出力と半画素生成部121の出力の差分データ、また、原画メモリ101に格納されているデータと半画素生成部121の出力の差分データを受信する。ブロック分割部102は、所定単位時間あたりの差分データの和をとり、この差分画像データ量があらかじめ設定した任意の値よりも小さい場合に、半画素生成（または、それより画素の精度が細くなる1/n画素生成（nは2より大きい整数））を実行するための情報のみを上位レイヤビットストリーム130として送信することもできる。なお、前述の所定単位時間は、1フレームを単位としてもよい。また、入力画像がインタレース信号の場合には1フィールドを単位としてもよい。あるいは、1フレームを複数のブロックに分割した際のブロックを所定単位時間としてもよい。さらに、フレーム単位で符号化モード（Iピクチャ、Pピクチャ、Bピクチャ）を変更する場合には、1GOP（Group Of Pictures）を所定単位時間としてもよい。また、差分画像を圧縮せずに、ヘッダ情報と差分である画素データをそのまま送信してもよい。

40

50

【0056】

このようにして、下位レイヤからアップサンプリングされた差分画像データは、ブロック分割部102で分割される。このとき、ブロック分割部112でダウンサンプリング画像を分割したブロック単位と異なるブロック単位で処理をおこなってもよい。そして、分割した画像を予測部103および周波数変換部104に入力する。

【0057】

予測部103では、フレーム内予測（イントラ予測）もしくはフレーム間予測（インター予測）がおこなわれ、予測画像が作成される。フレーム間予測をおこなう場合には、参照画メモリ109に格納されている画像データが参照される。

【0058】

周波数変換部104には、ブロック分割部102の画像と予測部103で作成された予測画像の差分をとった画像か、あるいは、ブロック分割部102の出力画像が入力される。周波数変換部104は、入力された画像を画像ブロック毎に周波数変換して、量子化部105に出力する。量子化部105は、量子化パラメータを用いて、入力された周波数変換データを量子化する。量子化されたデータは、可変長符号化部106に入力される。可変長符号化部106は、入力データを動き補償情報や符号化モード情報等のサイド情報とともに可変長符号化し、上位レイヤビットストリーム130を作成する。

【0059】

このとき、ブロック分割部に入力される画像の選択、ブロック分割の大きさの選択や、周波数変換部104に入力される画像の選択、上位レイヤビットストリームを送信しない場合など、複数の符号化方法における符号量と画質のコストを算出し、最もコストの低い圧縮符号化方法となったデータを上位レイヤビットストリーム130として送信してもよい。例えば、画質の向上に対して、どれくらい符号量を増加させなければならないかを、コストとして算出してもよい。このとき、非特許文献3に示すR-D Optimization技術を用いて、ビットレートと画質の値から算出するコストを予め算出しておき、符号化方法を選択してもよい。

【0060】

上位レイヤビットストリーム130は、図2に示されるバッファとしての役割を果たすメモリ203に一旦格納される。そして、メモリ203から、入力された符号化データが、伝送回線速度にあわせて出力される。なお、このとき、下位レイヤビットストリーム130と共に同じ伝送線路によって送信されてもよいし、別々の伝送線路によって送信されてもよい。ここで伝送線路とは無線による伝送や有線による伝送を問わず、また、ハードディスクドライブやDVDといった記録媒体を介して伝送されてもよい。

【0061】

ここで、下位レイヤビットストリームを生成する圧縮符号化方法と上位レイヤビットストリームを生成する圧縮符号化方法は、必ずしも同じである必要はなく、また、上位レイヤビットストリームは、必ずしも伝送する必要はない。デコーダの構成が複数の動画像圧縮規格に対応した構成であれば、複数の動画像圧縮規格のツールをハイブリッドに使用した符号化ならびに復号化をおこなってもよい。例えば、半画素生成部121の出力画像を生成するためにMPEG-4ASPの半画素生成フィルタを用い、その後の周波数変換部104では、H.264/AVCの整数変換処理をおこなってもよい。

【0062】

本実施形態では、下位レイヤビットストリームと上位レイヤビットストリームを組み合わせた構成としているが、ストリームの種類は二つである必要はなく、3ストリーム以上の構成としてもよい。

【0063】

また、本実施形態における下位レイヤビットストリーム131は上位レイヤビットストリーム130に比べて、例えば、CRCをおこなうなど手段により、ストリームの冗長性を持たせるなどして、エラー耐性を強化して伝送されてもよく、これによって、低ビットレートで必要最低限伝送を実現したいデータの伝送を強化できる。

【0064】

また、本実施形態における、量子化部115と量子化部105の量子化パラメータは同じである必要はなく、例えば、量子化部115では低周波数領域の係数に対する量子化値を細かくとり、量子化部105では高周波数領域の係数に対する量子化値を細かくした量子化パラメータとしてもよい。

【0065】

さらに、下位レイヤビットストリームを生成するための量子化部115では、粗い量子化パラメータで符号化をおこない、上位レイヤビットストリームを生成するための量子化部105で量子化パラメータを細かく設定してもよい。

【0066】

10

次に、図4を用いて本発明の第一の実施形態に係る映像の復号化装置を説明する。

図4は、本発明の第一の実施の形態に係る映像の符号化装置のハードウェア構成について説明するブロック図である。

【0067】

本実施形態の映像の復号化装置401は、相互に通信可能に接続されたプロセッサ402、メモリ403、入力インターフェース（入力I/F）404および出力インターフェース（出力I/F）406からなる。入力I/F404は、入力装置405に接続されており、また、出力I/F406は、出力装置407に接続されている。

【0068】

プロセッサ402は、本実施形態の映像の復号化の処理をおこなうプロセッサであり、メモリ403に格納されたプログラムを実行して、入力I/F404から受けたデータを復号化し、出力I/F406に送る。

20

【0069】

メモリ403には、プロセッサ402が実行するプログラムが格納される。また、プロセッサ402が処理する対象となるデータが一時的に格納される。

【0070】

映像の復号化装置401には、プロセッサ402およびメモリ403が複数設けられてもよい。例えば、本発明の映像復号化の処理をおこなうプログラムの一部のみを実行する専用のプロセッサが設けられてもよい。

【0071】

30

プロセッサ402およびメモリ403は、単一のチップ上に実装されてもよい。

【0072】

入力I/F404は、プロセッサ402が処理する符号化されたデータを入力装置405から受信するインターフェースである。

【0073】

入力装置405は、映像復号化装置401が処理する符号化されたデータを入力I/F404に入力する装置であり、例えば、符号化されたデータが格納された記憶装置である。この場合、入力I/F404は、例えば、SATAインターフェースである。また、入力装置405は、LANやIPネットワーク等（図示省略）を介して入力I/F404と接続されたコンピュータ装置であってもよい。この場合、入力I/F404は、ネットワークインターフェースである。また、入力装置405は、電話通信網（図示省略）を介して入力I/F404と接続された映像データ送信装置であってもよい。この場合、入力I/F404は、電話信号の受信機である。また、入力装置405は、デジタルTV放送局であってもよい。この場合、入力I/F404は、デジタルTVチューナーの提供するインターフェースである。

40

【0074】

出力I/F406は、プロセッサ402によって復号化されたデータを出力装置407に送信するインターフェースである。

【0075】

出力装置407は、映像の復号化装置401が復号化したデータの出力先であり、例え

50

ば、映像を出力するディスプレイである。この場合、出力 I / F 4 0 6 は、例えば、ビデオカードである。また、出力装置 4 0 7 は、復号化したデータをストリーム記録する記憶装置である。この場合、出力 I / F 4 0 6 は、例えば、S A T A インターフェースである。

【0076】

映像の復号化装置 4 0 1 は、入力 I / F 4 0 4 および出力 I / F 4 0 6 をそれぞれ複数備え、それぞれの入力 I / F 4 0 4 および出力 I / F 4 0 6 に、異なる種類の入力装置および出力装置が接続されてもよい。例えば、映像の復号化装置 4 0 1 は、二つの入力 I / F 4 0 4 を備え、一方にはハードディスク装置が接続され、もう一方には光ディスク装置が接続されてもよい。また、一方にはハードディスク装置が接続され、もう一方には L A N 等を介してコンピュータ装置が接続されてもよい。

10

【0077】

次に、図 3 を用いて本発明の一実施形態に係る映像の復号化装置の構成と動作について説明する。

図 3 は、本発明の一実施形態に係る映像の復号化装置の構成を示すブロック図である。

【0078】

本実施形態の映像復号化装置 4 0 1 は、図 3 に示されるように、可変長復号部 3 0 0 および 3 1 0 と、逆量子化部 3 0 1 および 3 1 1 と、逆周波数変換部 3 0 2 および 3 1 2 と、参照画メモリ 3 0 3 および 3 1 3 と、予測部 3 0 4 および 3 1 4 と、半画素生成部 3 2 1 からなる。ここで、可変長復号部 3 0 0 および 3 1 0、逆量子化部 3 0 1 および 3 1 1、逆周波数変換部 3 0 2 および 3 1 2、予測部 3 0 4 および 3 1 4 はメモリ 4 0 3 に格納されプロセッサ 4 0 2 によって実行されるプログラムであり、参照画メモリ 3 0 3 および 3 1 3 は、メモリ 4 0 3 の一部の領域である。また、これらのプログラムの一部または全部はその処理をおこなう専用のハードウェアによって実現されてもよい。

20

【0079】

また、可変長復号部 3 0 0 および 3 1 0、逆量子化部 3 0 1 および 3 1 1、逆周波数変換部 3 0 2 および 3 1 2、予測部 3 0 4 および 3 1 4 はそれぞれ、同一のハードウェアによって実現されてもよく、それらのハードウェアを時分割に使用することで処理が実行されてもよい。

【0080】

本実施形態の映像の復号化装置 4 0 1 は、図 1 に示した構成の映像の符号化装置 2 0 1 によって符号化されたストリームを復号化することができる。

30

【0081】

以下、映像の復号化装置 4 0 1 の各部の機能と動作を、符号化されたストリームを復号化する手順に沿って説明する。

【0082】

可変長復号部 3 1 0 は、入力された下位レイヤビットストリーム 1 3 1 のデータを解析し、フラグやデータの情報からフレーム間予測（インター予測）やフレーム内予測（イントラ予測）に関するモードの制御を予測部 3 1 4 に対しておこなう。予測部 3 1 4 は、モードがインター予測の場合、可変長復号部 3 1 0 からの動きベクトル等の情報と参照画メモリ 3 1 3 の画像データを用いてフレーム間予測をおこなう。また、モードがイントラ予測モードである場合、予測部 3 1 4 は、可変長復号部 3 1 0 から伝送されたイントラ符号化モード等の情報と参照画メモリ 3 1 3 の画像とを用いてフレーム内予測をおこなう。

40

【0083】

可変長復号部 3 1 0、逆量子化部 3 1 1、逆周波数変換部 3 1 2 の基本的な機能は、従来の復号化装置と同じものであり、それぞれ、ストリームの復号と情報解析、変換係数の逆量子化および逆周波数変換等をおこなう。

【0084】

参照画メモリ 3 1 3 と予測部 3 1 4 を用いて生成された予測画像と、逆周波数変換部 3

50

12の出力である差分画像を足し合わせるにより、低解像度画像331を生成し、出力装置407に伝送して、復号化された画像の画面への表示出力などをおこなう。

【0085】

同様に、上位レイヤビットストリームについても、可変長復号化部300、逆量子化部301、逆周波数変換部302、参照画メモリ303、予測部304、を用いて同様に復号化がおこなわれる。このとき、上位レイヤビットストリーム130が、予測をおこなわずに圧縮され送信されてきた場合は、逆周波数変換部302の出力画像に予測部304の出力は足し合わされない、あるいは、値がすべて0の画像データが足し合わされる。これにより、上位レイヤビットストリームを復号化した画像データが復元される。これに、半画素生成部321から出力された半画素精度画像が足し合わされることにより、高解像度画像330が生成される。半画素生成部321では、図1に示した画素生成部121で処理される半画素生成処理と同じ処理がおこなわれる。ここで、半画素生成部121は、下位レイヤのストリームにより生成された画像を、上位レイヤのストリームから生成された画像と足し合わせるためのアップサンプリングする処理をおこなうアップサンプリング部を形成している。

10

【0086】

以上説明した本実施形態の映像の符号化装置201および映像の復号化装置401は、同一のハードウェアとして実装されてもよい。この場合、メモリ203（または、メモリ403）には、図1および図3において説明した各プログラムが格納されると共に、図1および図3において説明した各領域が確保される。

20

【0087】

以上説明したように、本実施形態によれば、入力画像をダウンサンプリングして符号化した下位レイヤビットストリームと、そのストリームを復号し、入力画像と半画素生成処理の生成画像の差分をとった画像、または、超解像処理と半画素生成処理の生成画像の差分をとった画像を圧縮符号化した上位レイヤビットストリームとを伝送することができる。その結果、画像データの高い圧縮によって、従来よりも効率的な伝送帯域の利用と、サイマルキャストに比べてより少ない伝送レートでの通信を実現することができる。これにより、データを受信する各受信端末に応じた好適な伝送を実現することが可能となる。

【0088】

〔実施形態2〕

以下、本発明の第二の実施形態に係る映像の符号化装置と復号化装置を、図5および図6を用いて説明する。

30

【0089】

本実施形態で説明する符号化装置と復号化装置は、第一の実施形態と比べて、下位レイヤビットストリームから生成される画像をアップサンプリングする処理の形態を変えたものである。

【0090】

先ず、図5を用いて本発明の第二の実施形態に係る映像の符号化装置の構成と動作について説明する。

図5は、本発明の第二の実施形態に係る映像の符号化装置の構成を示すブロック図である。

40

【0091】

図1に示した第一の実施形態の符号化装置の構成に対して、原画メモリ101に格納されている入力画像と超解像処理部120からの出力画像の差分をとり、ブロック分割部102へ入力している点が異なっている。その他の処理は、第一の実施形態の符号化装置と同様である。

【0092】

この構成においても、従来技術に比べて復号化した際のデコード画像において、より高画質な動画像符号化が実現でき、復元時の画質が最大で原画像と同等の画質の伝送が可能となる。

50

【 0 0 9 3 】

次に、図 6 を用いて本発明の第二の実施形態に係る映像の復号化装置の構成と動作について説明する。

図 6 は、本発明の第二の実施形態に係る映像の復号化装置の構成を示すブロック図である。

【 0 0 9 4 】

図 3 に示した第一の実施形態の復号化装置の構成に対して、半画素生成部 3 2 0 が超解像処理部 1 0 2 0 となっている点が異なっている。下位レイヤビットストリーム 1 3 1 をデコードした低解像度画像 3 3 1 を超解像処理部 1 0 2 0 で超解像処理したものと、上位レイヤビットストリーム 9 0 0 をデコードした画像とを足し合わせることで、高画質な高解像度画像を復元することができる。

10

【 0 0 9 5 】

以上本実施形態の映像の符号化装置と復号化装置を用いれば、第一の実施形態と同様に、従来技術と比べて、より高画質な画像データを用いた伝送の実現が可能となる。

【図面の簡単な説明】

【 0 0 9 6 】

【図 1】本発明の第一の実施形態に係る映像の符号化装置の構成を示すブロック図である。

。

【図 2】本発明の第一の実施の形態に係る映像の符号化装置のハードウェア構成について説明するブロック図である。

20

【図 3】本発明の一実施形態に係る映像の復号化装置の構成を示すブロック図である。

【図 4】本発明の第一の実施の形態に係る映像の符号化装置のハードウェア構成について説明するブロック図である。

【図 5】本発明の第二の実施形態に係る映像の符号化装置の構成を示すブロック図である。

。

【図 6】本発明の第二の実施形態に係る映像の復号化装置の構成を示すブロック図である。

。

【図 7】S N R スケーラビリティ符号化装置の構成例を示す図である。

【図 8】S N R スケーラビリティ復号化装置の構成例を示す図である。

【図 9】空間スケーラビリティ符号化装置の構成例を示す図である。

30

【図 1 0】空間スケーラビリティ復号化装置の構成例を示す図である。

【符号の説明】

【 0 0 9 7 】

1 0 0, 5 0 0 … 入力画像

1 0 1 … 原画メモリ

1 0 2, 1 1 2 … ブロック分割部

1 0 3, 1 1 3, 3 0 4, 3 1 4 … 予測部

1 0 4, 1 1 4 … 周波数変換部

1 0 5, 1 1 5 … 量子化部

1 0 6, 1 1 6 … 可変長符号化部

40

1 0 7, 1 1 7, 3 0 1, 3 1 1 … 逆量子化部

1 0 8, 1 1 8, 3 0 2, 3 1 2 … 逆周波数変換部

1 0 9, 1 1 9, 3 0 3, 3 1 3 … 参照画メモリ

1 1 0, 7 0 0 … ダウンサンプリング部

1 2 0 … 超解像処理部

1 2 1, 3 2 1 … 半画素生成部

1 3 0, 7 1 0, 9 0 0 … 上位レイヤビットストリーム

1 3 1, 7 1 1 … 下位レイヤビットストリーム

2 0 1 … 映像符号化装置

2 0 2, 4 0 2 … プロセッサ

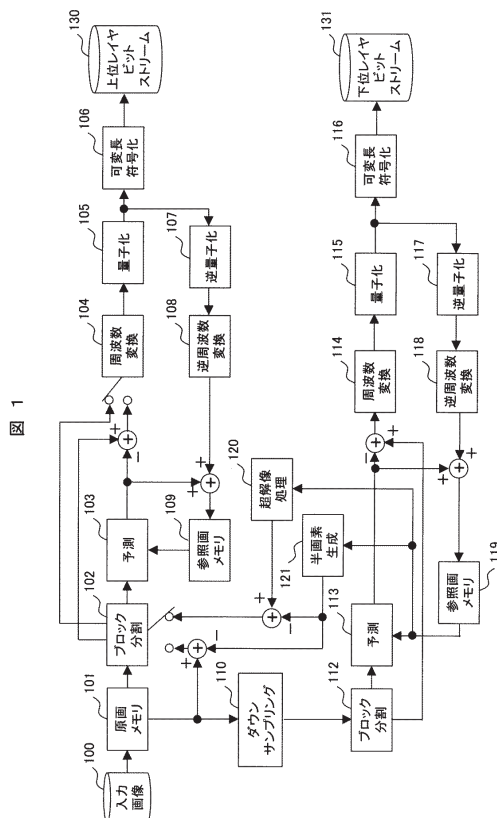
50

203, 403…メモリ
 204, 404…入力インターフェース (I/F)
 205, 405…入力装置
 206, 406…出力インターフェース (I/F)
 207, 407…出力装置
 300, 310…可変長復号化部
 330, 1010…高解像度画像
 331…低解像度画像
 401…映像復号化装置
 501, 511…DCT部
 502, 512…Q部
 503, 513…VLC部
 504, 514, 601, 603, 801, 811…IQ部
 505, 515, 604, 802, 812…IDCT部
 506, 516, 605, 803, 813…FM部
 507, 517, 606, 804, 814…MC部
 510…Delay部
 540…エンハンスメントビットストリーム
 541…低SNRビットストリーム
 600, 602, 800, 810…VLD部
 701…アップサンプリング部。

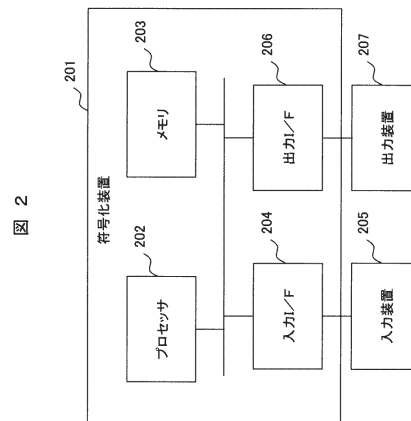
10

20

【図1】



【図2】



【図 3】

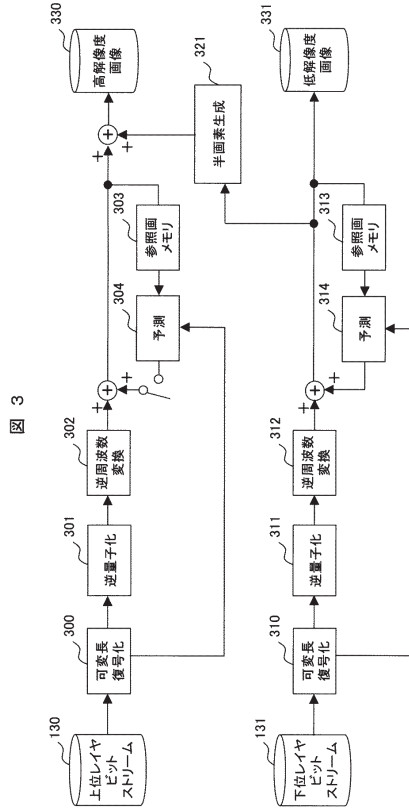


図 3

【図 4】

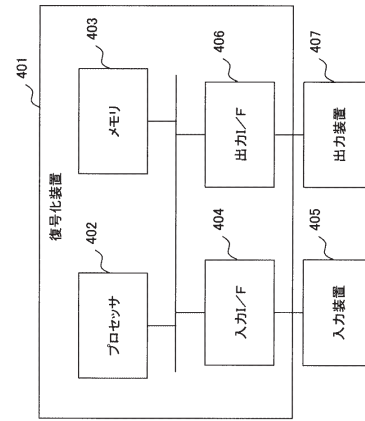


図 4

【図 5】

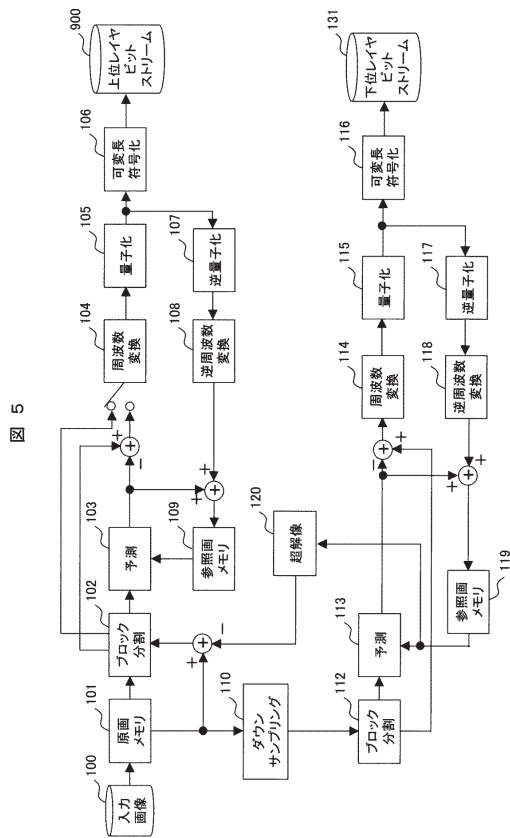


図 5

【図 6】

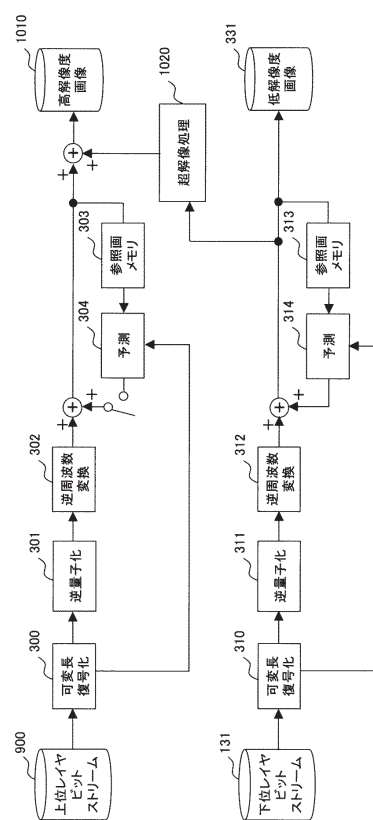
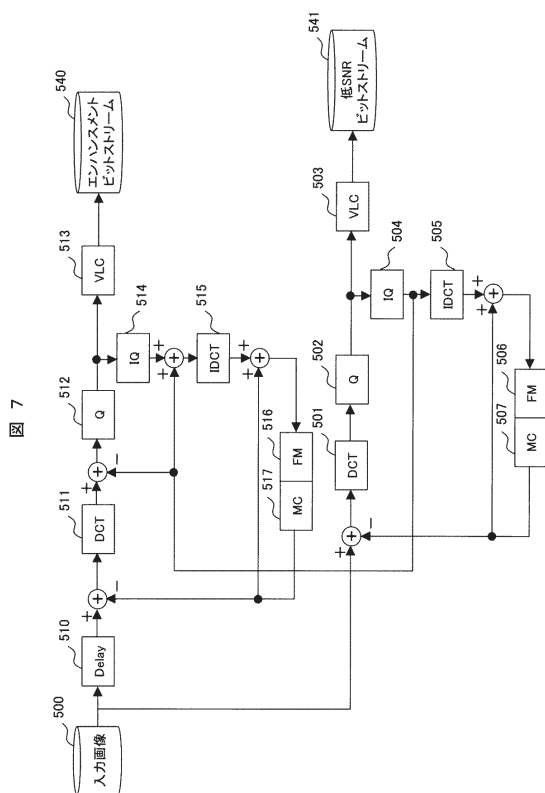
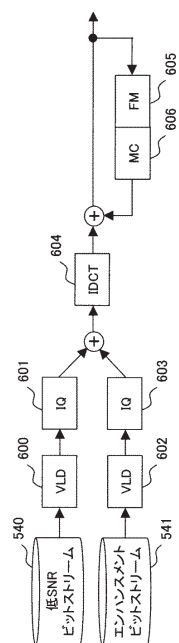


図 6

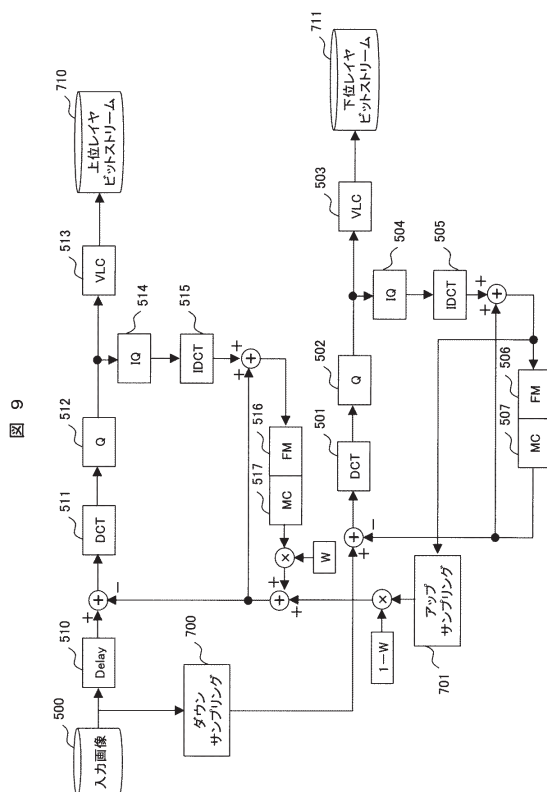
【图 7】



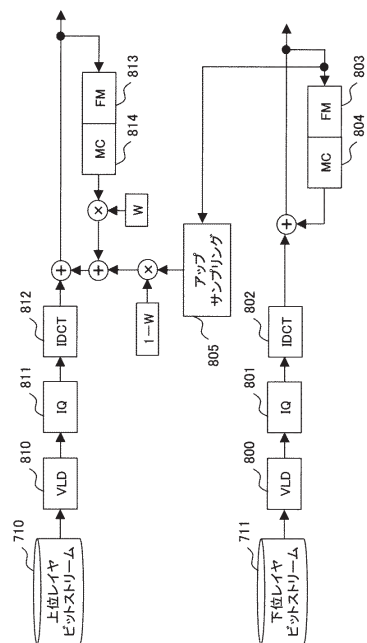
【图 8】



【图 9】



【 図 1 0 】



フロントページの続き

(72)発明者 中田 啓明

神奈川県横浜市戸塚区吉田町2 9 2 番地 株式会社日立製作所組込みシステム基盤研究所内

(72)発明者 江浜 真和

神奈川県横浜市戸塚区吉田町2 9 2 番地 株式会社日立製作所組込みシステム基盤研究所内

(72)発明者 影山 昌広

神奈川県横浜市戸塚区吉田町2 9 2 番地 株式会社日立製作所組込みシステム基盤研究所内

F ターム(参考) 5C059 LB05 LB15 MA00 MA04 MA05 MA12 MA17 MA21 MA23 MA24

MA32 MC11 ME01 PP05 PP06 PP07 RF00 SS02 SS08 SS13

TA17 TA46 TA76 TC18 TD02 UA02 UA05 UA12

5C159 LB05 LB15 MA00 MA04 MA05 MA12 MA17 MA21 MA23 MA32

MA41 MC11 ME01 PP05 PP06 PP07 RF00 SS02 SS08 SS13

TA17 TA46 TA76 TC18 TD02 UA02 UA05 UA12

Notice

This translation is machine-generated. It cannot be guaranteed that it is intelligible, accurate, complete, reliable or fit for specific purposes. Critical decisions, such as commercially relevant or financial decisions, should not be based on machine-translation output.

DESCRIPTION JP2009182776A

10 Encoding device, decoding device, video encoding method, and video decoding method

[0001]

14 The present invention relates to an encoding device, a decoding device, a video encoding method, and a video decoding method, and in particular to an encoding device, a decoding device, a video encoding method, and a video decoding method that are suitable for improving data transmission efficiency by applying scalability in which data of different layers is simultaneously transmitted and utilized according to the characteristics of the data.

[0002]

22 Nowadays, a function called scalability has been introduced to transmit video and audio data in a situation where the transmission band is limited and where there are multiple types of terminals receiving the data. This scalability is a function that gradually varies the SNR (Signal to Noise Ratio), spatial resolution, and time resolution by partially decoding the bit stream, and is adopted in video compression standards such as MPEG-2, MPEG-4, and SVC, an extension standard of H.264. To achieve scalability, a technique called hierarchical coding is used. For example, MPEG-2 defines an SNR scalable profile, a spatial scalable profile, and a High profile.

[0003]

32 In situations where transmission bandwidth is limited, it is important to improve coding efficiency as much as possible and to avoid duplication of header information and side information required for coding, compared to a method (simulcast) in which each layer is simply coded and transmitted separately.

36 Scalability improves coding efficiency by making mutual use of this information between layers. The scalability is explained in the following non-patent document 1.

[0004]

- ⁴¹ The SNR scalable profile employs a technique called SNR scalability, which will be described later, to realize a graceful degradation function.
- ⁴³ Graceful degradation is a function that realizes gradual deterioration of image quality according to the reception strength of radio waves, for example. In normal digital broadcasting, there is a problem that when the radio wave reception strength falls below a certain specified value, the broadcast cannot be received at all from that point on. However, when SNR scalability is applied to broadcasting, the image quality deteriorates gradually, making it possible to reproduce image quality according to the reception conditions.

[0005]

- ⁵² SNR scalability realizes hierarchical coding in the frequency domain (DCT coefficients in MPEG-2).
- ⁵⁴ The basic idea is to send a signal providing base image quality (reference low SNR) over a high priority transmission path, and then send an auxiliary signal (enhancement signal) to improve the base image quality while maintaining the same spatial resolution over a low priority transmission path. This allows high-quality decoding in areas where both signals can be received. Furthermore, even if the reception conditions deteriorate and the enhancement signal cannot be received, the base signal is transmitted and received with high priority, so that the reliability of reproducing the reference image quality can be improved.

[0006]

- ⁶⁴ Here, an example of encoding and decoding of SNR scalability will be described with reference to FIG. 7 and FIG.
- ⁶⁶ FIG. 7 is a diagram showing an example of the configuration of an SNR scalability encoding device. FIG. 8 is a diagram showing an example of the configuration of an SNR scalability decoding device.

[0007]

- ⁷² As shown in FIG. 7, an input image 500 is input to the encoding device, and differential image data from a predicted image generated by an MC unit 507 is input to a DCT unit 501 in order to generate a low SNR bitstream.
- ⁷⁵ Here, the predicted image is an image for prediction generated by a loop process formed by an inverse quantization unit 504, an IDCT unit 505, an FM (frame memory) unit 506, and an MC (Motion Compensation) unit 507.

[0008]

80 Data input to the DCT unit 501 undergoes DCT in the DCT processing unit 501 , and then the data is roughly quantized in the quantization unit 502 , and a low SNR bit stream 541 is generated by the VLC unit 503 .

84 The output of the quantization unit 502 is inversely quantized by an inverse quantization unit 504 to generate coarsely quantized DCT coefficients. The difference between this DCT coefficient and the DCT coefficient output by the DCT section 511 is calculated and input to the quantization section 512 . In addition, the output of the inverse quantization unit 504 is added to the output of the inverse quantization unit 514 and input to the IDCT unit 515 . The loop processing formed by the inverse quantization unit 504, IDCT unit 505, FM (frame memory) unit 506, and MC (Motion Compensation) unit 507 is local decoding in MPEG-2, in which the image restored on the decoder side is also generated on the encoder side, and the restored image is used to predict the next frame.

[0009]

96 On the other hand, for the enhancement stream 540 , after the input image 500 is input to the delay unit 510 , a difference is taken between the input image 500 and a predicted image generated by the MC unit 517 , and the difference image is input to the DCT unit 511 .

99 Here, the predicted image is an image for prediction generated by a loop process formed by an inverse quantization unit 514 , an IDCT unit 515 , an FM (frame memory) unit 516 , and an MC (Motion Compensation) unit 517 .

[0010]

105 The image input to the DCT unit 511 is frequency-converted by the DCT unit 511 .

106 Then, the difference between the DCT coefficients output by the DCT section 511 and the DCT coefficients output by the IQ section 504 is calculated, and this difference is input to the quantization section 512 . The generated differential signal is a signal for obtaining a high SNR signal by adding a low SNR signal. The data quantized by the quantization unit 512 is then variable-length coded by the VLC unit 513 to generate an enhancement bitstream 540 . For prediction, the enhancement image data is locally decoded by an IDCT unit 515 , an FM (frame memory) unit 516 , and an MC (Motion Compensation) unit 517 , and the data output from the MC unit 517 is input to the DCT unit 511 .

[0011]

117 In the case of SNR scalability decoding, the SNR scalability decoding device shown in FIG. 8 performs decoding processing on the low SNR bit stream 540 by the VLD unit 600 and performs inverse quantization by the IQ unit 601 .

120 On the other hand, the enhancement bit stream 541 is decoded by the VLD unit 602 and inverse quantized by the IQ unit 603 . Then, the DCT coefficients output from the IQ unit

601 and the DCT coefficients output from the IQ unit 603 are added together, and an IDCT process is performed by the IDCT unit 604 . Thereafter, decoding is performed in the same manner as in the single layer case.

[0012]

130 The spatial scalable profile is a method for realizing compatibility between displays of images with different resolutions, such as standard definition television (SDTV) and high definition television (HDTV).

[0013]

136 Spatial scalability is a technique for improving coding efficiency by utilizing image similarity between layers with different resolutions.

138 Here, an example of encoding and decoding of spatial scalability will be described with reference to FIG. 9 and FIG.

140 FIG. 9 is a diagram showing an example of the configuration of a spatial scalability coding device. FIG. 10 is a diagram showing an example of the configuration of a spatial scalability decoding device.

[0014]

146 As shown in FIG. 9, in the spatial scalability coding apparatus, first, an input image 500 is downsampled by a downsampling unit 700 to generate a bitstream of a lower layer.

148 The difference between this image and a predicted image generated by the MC unit 507 is taken and input to the DCT unit 501 . Here, similarly to the case of FIG. 8, the predicted image is an image for prediction generated by a loop process formed by an inverse quantization unit 504, an IDCT unit 505, an FM (frame memory) unit 506, and an MC (Motion Compensation) unit 507.

[0015]

156 The data input to the DCT unit 501 undergoes DCT in the DCT processing unit 501 , then undergoes coarse quantization in the quantization unit 502 , and then undergoes variable length coding in the VLC unit 503 , and transmits the lower layer bit stream 711 .

159 In addition, the output of the quantization unit 502 is inverse quantized by the IQ unit 504 and subjected to IDCT processing by the IDCT unit 505, and then converted to the same resolution as the input image by the upsampling unit 701. This image is used as a third reference image together with the two reference images in motion compensation of the upper layer. This prediction from the lower layer is called spatial prediction. The spatial prediction reference image is an image at the same time as the image to be coded in the upper layer, so no motion compensation is required. Here, the motion compensation image

output from the MC unit 517 and the spatial prediction image output from the upsampling filter unit 701 are added together with weights (W) and (1-W), respectively, using a variable W that ranges from 0 to 1, and are used for predicting the upper layer.

[0016]

179 When decoding spatial scalability, the spatial scalability decoding device shown in FIG. 10 decodes the lower layer bit stream 711 in a VLD unit 800, an IQ unit 801, and an IDCT unit 802.

182 Furthermore, the decoded image of the lower layer is upsampled by the upsampling unit 805 and used for decoding the upper layer bit stream. The upper layer bit stream 710 is decoded by the VLD unit 810, the IQ unit 811, and the IDCT unit 812, and a high-quality, high-resolution image is generated by performing weighted addition of (W) and (1-W) from the encoded and transmitted variable W using the upsampled image of the restored image of the lower layer and the motion compensated image of the upper layer as reference images.

[0017]

191 Furthermore, there is also a High profile that can utilize a combination of this SNR scalability and spatial scalability.

[0018]

196 In addition, recent television sets have become larger in screen size, and rather than simply displaying the image signals input from broadcasts, communications, storage media, etc., the number of pixels in the horizontal and vertical directions is generally increased through digital signal processing.

200 In this case, the resolution cannot be increased simply by increasing the number of pixels using a commonly known interpolation low-pass filter using a sinc function, a spline function, or the like.

[0019]

206 Therefore, as described in Patent Document 1, Patent Document 2, and Non-Patent Document 2, a technology has been proposed that increases the number of pixels while increasing the resolution by synthesizing multiple input image frames into a single frame (hereinafter referred to as "super-resolution processing").

[0020]

213 特開平 8－3 3 6 0 4 6 号公報

214 特開平 9－6 9 7 5 5 号公報

- 215 テレビジョン学会編「総合マルチメディア選書MPEG」、平成8年4月20日、5－6 ス
ケーラビリティ、p.117-132
- 215 青木伸 “複数のデジタル画像データによる超解像処理”, Ricoh Technical Report pp.19-25,
No.24, NOVEMBER, 1998
- 219 Gary J. Sullivan and Thomas Wiegand, "Rate-Distortion Optimization for Video Compression",
IEEE Signal Processing Magazine, vol. 15, no. 6, pp.74-90, November, 1998.

[0021]

- 224 As described above, when simulcasting data of multiple resolutions to multiple types of
receiving terminals, a problem arises in that the limited transmission band is largely
occupied, resulting in poor transmission efficiency.

[0022]

- 230 Furthermore, even in conventional transmission methods utilizing scalability, there is a
demand for improving the data compression rate and transmission efficiency.

[0023]

- 235 The present invention has been made to solve the above problems, and its object is to
improve the data compression rate in data transmission using an encoding device and a
decoding device that utilize scalability, and to efficiently utilize the transmission bandwidth
by using a scalable data structure.

[0024]

- 242 The coding device of the present invention inputs an image, codes each of the images as a
stream consisting of a number of layers of different resolutions, and outputs the coded data.
For each layer, the coding device is provided with a frequency conversion unit, a quantization
unit, and a variable-length coding unit which compress and code the image data, as well as
an inverse frequency conversion unit and an inverse quantization unit which generate
images for prediction. It is also provided with a downsampling unit which reduces the
resolution of the image, and a super-resolution processing unit and a half-pixel generation
processing unit which improve the resolution of the image.
- 250 The input image is then downsampled by a downsampling unit and input to a frequency
conversion unit for a lower layer with lower resolution, and the image used for predicting the
lower layer is subjected to super-resolution processing and half-pixel generation processing
by a super-resolution processing unit and a half-pixel generation processing unit, and these
generated images or difference images are input to a frequency conversion unit for a higher
layer with higher resolution.

[0025]

260 In addition, in the decoding device of the present invention, a stream consisting of a number of layers with different resolutions is input, and images each having a different resolution are output. For each layer, the decoding device is provided with an inverse frequency conversion unit and an inverse quantization unit that decodes compressed data, and images generated from the lower layer stream and used for prediction during decoding are subjected to super-resolution processing and half-pixel generation processing by a super-resolution processing unit and a half-pixel generation processing unit, and these generated images are added to images generated from the higher resolution upper layer stream to generate an image.

[0026]

271 According to the present invention, in data transmission by an encoding device and a decoding device utilizing scalability, the data compression rate is improved and a scalable data structure is used to efficiently utilize the transmission band.

[0027]

277 Hereinafter, embodiments of the present invention will be described with reference to FIGS. 1 to 6. FIG.

[0028]

282 [First embodiment] Hereinafter, a video encoding device and a video decoding device according to a first embodiment of the present invention will be described with reference to Figs.

[0029]

288 First, the hardware configuration of the video encoding device according to the first embodiment of the present invention will be described with reference to FIG.

290 FIG. 2 is a block diagram illustrating a hardware configuration of a video encoding device according to a first embodiment of the present invention.

[0030]

295 As shown in FIG. 2, the video encoding device 201 of this embodiment comprises a processor 202, a memory 203, an input interface (input I/F) 204, and an output interface (output I/F) 206, all of which are communicatively connected to one another.

298 The input I/F 204 is connected to an input device 205, and the output I/F 206 is connected to an output device 207.

[0031]

³⁰³ The processor 202 is a processor that performs the video encoding process of this embodiment, executes a program stored in the memory 203 , encodes data received from an input I/F 204 , and sends the encoded data to an output I/F 206 .

³⁰⁶ The memory 203 stores the programs executed by the processor 202 . In addition, data to be processed by the processor 202 is temporarily stored. The video encoding device 201 may be provided with a plurality of processors 202 and memories 203 . For example, a dedicated processor may be provided for executing only a part of the program that performs the video encoding process of the present invention. Furthermore, multiple dedicated processors for performing the same processing may be provided.

[0032]

³¹⁵ The processor 202 and the memory 203 may be implemented on a single chip.

[0033]

³¹⁹ The input I/F 204 is an interface that receives video data to be processed by the processor 202 from an input device 205 .

[0034]

³²⁴ The input device 205 is a device that inputs a video signal to be processed by the video encoding device 201 to the input I/F 204, and is, for example, a lens or a video imaging element.

³²⁷ In this case, the input I/F 204 is, for example, an A/D converter.

³²⁸ The input device 205 may also be a storage device in which unencoded video data is stored.

³²⁹ In this case, the input I/F 204 is, for example, a Serial AT Attachment (SATA) interface.

[0035]

³³³ The output I/F 206 is an interface that transmits the data encoded by the processor 202 to an output device 207 .

[0036]

³³⁸ The output device 207 is an output destination of the data encoded by the video encoding device 201, and is, for example, a storage device that stores the encoded data.

³⁴⁰ In this case, the output I/F 206 is, for example, a SATA interface.

³⁴¹ Furthermore, the output device 207 may be a computer device connected to the output I/F

206 via a LAN, an IP network, or the like (not shown). In this case, the output I/F 206 is a network interface. The output device 207 may also be an image receiving device connected to the output I/F 206 via a telephone communication network (not shown). In this case, the output I/F 206 is a transmitter of a telephone signal. The output device 207 may also be a digital TV set. In this case, the output I/F 206 becomes an interface that is processed by a transmitter of a digital TV signal.

[0037]

352 The video encoding device 201 includes a plurality of input I/Fs 204 and a plurality of output I/Fs 206, and different types of input devices and output devices may be connected to the respective input I/Fs 204 and output I/Fs 206.

355 For example, the video encoding device 201 may include two output I/Fs 206, one of which may be connected to a hard disk device and the other to a magneto-optical disk device. Also, a hard disk device may be connected to one side, and a computer device may be connected to the other side via a LAN or the like.

[0038]

362 Next, the configuration and operation of a video encoding device according to a first embodiment of the present invention will be described with reference to FIG.

364 FIG. 1 is a block diagram showing the configuration of a video encoding device according to a first embodiment of the present invention.

[0039]

369 As shown in FIG. 1, the video encoding device comprises an original image memory 101, block division units 102 and 112, prediction units 103 and 113, frequency conversion units 104 and 114, quantization units 105 and 115, variable length coding units 106 and 116, quantization units 107 and 117, inverse frequency conversion units 108 and 118, reference image memories 109 and 119, downsampling unit 110, super-resolution processing unit 120, and half-pixel generation unit 121.

[0040]

378 The original image memory 101 is a memory for temporarily storing an input original image.

379 The block division units 102 and 112 divide an image into a plurality of image blocks in order to process a portion of the image in block units.

381 The prediction units 103 and 113 perform prediction within the same frame (intra prediction) or between adjacent frames (inter prediction). The frequency transform units 104 and 114 perform frequency transform such as DCT (Discrete Cosine Transform) on the image signals. The quantization units 105 and 115 perform quantization processing on the input frequency

coefficients. Variable-length coding units 106 and 116 perform variable-length coding on the quantized data to compress the number of bits.

[0041]

³⁹⁴ The inverse quantization units 107 and 117 perform inverse quantization processing that is the inverse of quantization.

³⁹⁶ The inverse frequency transform units 108 and 118 perform an inverse frequency transform to obtain an image from the frequency coefficients. The reference image memories 109 and 119 are temporary memories for holding reference images.

[0042]

⁴⁰² The downsampling unit 110 performs downsampling processing on the input image in order to encode the bitstream of the lower layer as well.

⁴⁰⁴ The super-resolution processor 120 performs super-resolution processing to generate a high-resolution image from a plurality of frame images. The half-pixel generating unit 121 performs half-pixel generation processing to generate pixels with half-pixel accuracy. The super-resolution processing unit 120 and the half-pixel generating unit 121 form an upsampling unit that performs processing for upsampling an image for generating a lower layer stream in order to generate a higher layer stream.

[0043]

⁴¹³ Each of these units may be configured by hardware or software.

⁴¹⁴ It may also be a module that combines hardware and software. Furthermore, the processing in the frequency conversion section is not particularly limited, and DCT, DST (Discrete Sine Transform), DFT (Discrete Fourier Transform), wavelet transform, integer transform processing, etc. may be performed.

[0044]

⁴²¹ The block division unit 102, prediction unit 103, frequency conversion unit 104, quantization unit 105, and variable length coding unit 106 are a single compression unit that realizes the compression function for generating an upper layer bit stream, while the block division unit 112, prediction unit 113, frequency conversion unit 114, quantization unit 115, and variable length coding unit 116 are a single compression unit that realizes the compression function for generating an upper layer bit stream.

[0045]

⁴³⁰ In addition, the inverse quantization unit 107, the inverse frequency conversion unit 108, and

the reference image memory 109 form a single decompression unit that performs decompression to realize prediction for generating an upper layer bitstream, and the inverse quantization unit 117, the inverse frequency conversion unit 118, and the reference image memory 119 form a single decompression unit that performs decompression to realize prediction for generating a lower layer bitstream.

[0046]

440 Here, the block division units 102 and 112, the prediction units 103 and 113, the frequency conversion units 104 and 114, the quantization units 105 and 115, the variable length coding units 106 and 116, the inverse quantization units 107 and 117, the inverse frequency conversion units 108 and 118, and the reference image memories 109 and 119 are each formed as a single piece of hardware, and may be used in a time-division manner.

[0047]

448 Furthermore, the half pixel generator 121 may be processed using a half pixel generator that exists in the prediction unit 103 or the prediction unit 113 .

[0048]

453 Next, the operation of the video encoding device shown in FIG. 1 will be described.

[0049]

457 An input image 100 stored in an original image memory 101 is input to a downsampling unit 110 .

459 The downsampling unit 110 performs low-pass filtering on the input image, and then performs pixel thinning processing.

461 The image that has been subjected to the thinning process is input to the block division section 112 in the lower layer, and divided into a plurality of image blocks.

463 The divided image blocks are subjected to intra-frame prediction (intra-prediction) or inter-frame prediction (inter-prediction) by the prediction unit 113 to generate a predicted image.

465 When inter-picture prediction is performed, image data stored in the reference image memory 119 is referenced.

[0050]

470 Then, an image obtained by calculating the difference between the image from the block division unit 112 and the predicted image created by the prediction unit 113 is input to the frequency transformation unit 114 .

473 The frequency transform unit 114 performs frequency transform on the input image for each

image block.

475 Here, the frequency transformation process may be DCT, DST, DFT, or wavelet transformation, or may be integer transformation.

478 The frequency-converted coefficient data is input to the quantization unit 115. The quantization unit 115 quantizes the input frequency transformation data using the quantization parameter. The quantized data is input to a variable-length coding unit 116. The variable length coding unit 116 performs variable length coding on the input data together with side information such as motion compensation information and coding mode information, to generate a lower layer bit stream 131. The lower layer bitstream 131 is temporarily stored in a memory 203 that serves as a buffer as shown in FIG. The input coded data is then output from memory 203 in accordance with the transmission line speed. At this time, it may be transmitted over the same transmission line as the upper layer bit stream 130 described later, or may be transmitted over a separate transmission line. Here, the transmission line may refer to wireless or wired transmission, and may also refer to transmission via a recording medium such as a hard disk drive or a DVD. Furthermore, packets of the lower layer bitstream 131 and the enhancement layer bitstream 130 may be transmitted together with respect to one header information, or portions of the lower layer bitstream 131 and the enhancement layer bitstream may be transmitted alternately.

[0051]

496 On the other hand, the quantized data generated by the quantization unit 115 is input to the inverse quantization unit 117.

498 The inverse quantization unit 117 inverse quantizes the quantized data. The dequantized coefficient data is input to the inverse frequency transform unit 118. The inverse frequency transform unit 118 performs inverse frequency transform on this data to restore it as a difference block image. The reconstructed difference block image is added to the predicted image output from the prediction unit 113, and the result is stored in the reference image memory 119 as a locally decoded image block. This locally decoded image block is then input to the prediction unit 113. The locally restored image block is also input to a super-resolution processor 120 and a half-pixel generator 121. Here, the local decoded image block input to the super-resolution processing unit 120, the half pixel generation unit 121, and the prediction unit 113 does not necessarily have to be output from the reference image memory 119, and data stored in the reference image memory may be directly input.

[0052]

512 The super-resolution processor 120 uses multiple frames of the input image to generate a high-resolution image.

514 Here, in this embodiment, the processing content of the super-resolution processing unit does not necessarily have to be super-resolution processing as described in the prior art section, and the processing content is not particularly limited as long as it is processing that

improves the resolution of the input image.

[0053]

524 On the other hand, the half-pixel generating unit 121 generates an image with half-pixel accuracy by performing half-pixel generation processing on the input locally decoded image block.

527 However, instead of the half pixel generation process, 1/4 pixel generation process, 1/8 pixel generation process, . . . , or 1/n pixel generation process (n is an integer greater than 2) with finer pixel accuracy may be performed. Furthermore, the content of the processing is not particularly limited as long as it is processing that improves the resolution in accordance with the resolution of the super-resolution processing. Here, the half-pixel generation process refers to a process of generating an image having twice the accuracy of one pixel by performing a filter process on pixel data, and specifically refers to an interpolation filter in MPEG-2, MPEG-4, or H.264/AVC. To give more detailed values, the 8-tap filter coefficient when generating half pixels from integer pixels in the MPEG-4 ASP is $[-2 - 12 - 12 - 2]/64$, and the 2-tap filter coefficient when generating $1/4$ pixels from half pixels is $[1 + 1]/2$. In addition, in H.264/AVC, the 6-tap filter coefficient when generating half pixels from integer pixels is $[1 - 5 \ 20 \ 20 - 5 \ 1]/32$, and the 2-tap filter coefficient when generating $1/4$ pixels from half pixels is $[1 \ 1]/2$.

[0054]

543 Here, the difference is taken between the generated half-pixel precision image (or, in some cases, a 1/n-pixel generated precision image (n is an integer greater than 2) which has even finer precision) and the output image of the super-resolution processing unit 120, and this difference data is input to the block division unit 102.

547 Alternatively, the difference between the input image stored in the original image memory 101 and the half-pixel precision image generated by the half-pixel generation unit 121 is calculated, and the difference data is input to the block division unit 102 .

[0055]

553 The block division unit 102 receives difference data between the output of the super-resolution processing unit 120 and the output of the half-pixel generation unit 121 , and also receives difference data between the data stored in the original image memory 101 and the output of the half-pixel generation unit 121 .

557 The block division unit 102 sums up the differential data per specified unit time, and if the amount of differential image data is smaller than a predetermined arbitrary value, it can also transmit only the information for performing half-pixel generation (or 1/n-pixel generation (n is an integer greater than 2) which provides finer pixel accuracy) as the upper layer bit stream 130. The above-mentioned predetermined unit time may be one frame. Furthermore,

when the input image is an interlaced signal, one field may be used as a unit. Alternatively, one frame may be divided into a plurality of blocks, and each block may be used as the predetermined unit time. Furthermore, when the coding mode (I picture, P picture, B picture) is changed on a frame-by-frame basis, one GOP (Group Of Pictures) may be set as a predetermined unit time. Also, the difference image may not be compressed, and the header information and the pixel data which is the difference may be transmitted as is.

[0056]

576 In this manner, the differential image data upsampled from the lower layer is divided by the block division unit 102 .

578 At this time, the processing may be performed in block units different from the block units into which the downsampled image is divided by the block dividing section 112 . Then, the divided images are input to the prediction unit 103 and the frequency transformation unit 104 .

[0057]

585 The prediction unit 103 performs intra-frame prediction (intra-prediction) or inter-frame prediction (inter-prediction) to generate a predicted image.

587 When performing inter-frame prediction, image data stored in the reference image memory 109 is referenced.

[0058]

592 The frequency transform unit 104 receives an image obtained by calculating the difference between the image from the block division unit 102 and the predicted image created by the prediction unit 103 , or an output image from the block division unit 102 .

595 The frequency transform unit 104 performs frequency transform on the input image for each image block, and outputs the transformed image to the quantization unit 105 . The quantization unit 105 quantizes the input frequency transformation data using the quantization parameter. The quantized data is input to a variable-length coding unit 106 . The variable length coding unit 106 performs variable length coding on the input data together with side information such as motion compensation information and coding mode information, to generate an enhancement layer bit stream 130 .

[0059]

605 At this time, the cost of the code amount and image quality for multiple encoding methods may be calculated, such as when selecting an image to be input to the block division unit, selecting the size of the block division, selecting an image to be input to the frequency conversion unit 104, or when not transmitting an upper layer bit stream, and the data

resulting from the compression encoding method with the lowest cost may be transmitted as the upper layer bit stream 130.

613 For example, the amount of code that must be increased to improve image quality may be calculated as the cost. At this time, the R-D Optimization technique described in Non-Patent Document 3 may be used to calculate a cost based on the bit rate and image quality in advance, and then the encoding method may be selected.

[0060]

622 The enhancement layer bitstream 130 is temporarily stored in a memory 203 that serves as a buffer as shown in FIG.

624 The input coded data is then output from memory 203 in accordance with the transmission line speed. At this time, the lower layer bit stream 130 may be transmitted via the same transmission line, or may be transmitted via a separate transmission line. Here, the transmission line may refer to wireless or wired transmission, and may also refer to transmission via a recording medium such as a hard disk drive or DVD.

[0061]

632 Here, the compression encoding method for generating the lower layer bitstream and the compression encoding method for generating the upper layer bitstream do not necessarily have to be the same, and the upper layer bitstream does not necessarily have to be transmitted.

636 If the decoder is configured to support a plurality of video compression standards, encoding and decoding may be performed using a hybrid of tools for a plurality of video compression standards. For example, an MPEG-4 ASP half-pixel generation filter may be used to generate an output image from the half-pixel generation unit 121, and the subsequent frequency conversion unit 104 may perform integer conversion processing of H.264/AVC.

[0062]

644 In this embodiment, a configuration is used in which a lower layer bit stream and an upper layer bit stream are combined, but the number of types of streams does not need to be two, and a configuration of three or more streams may be used.

[0063]

650 Furthermore, in this embodiment, the lower layer bit stream 131 may be transmitted with enhanced error resistance compared to the upper layer bit stream 130, for example by providing stream redundancy by means of performing CRC, thereby enhancing the transmission of data that needs to be transmitted at a low bit rate at the minimum necessary level.

[0064]

658 In addition, in this embodiment, the quantization parameters of quantization unit 115 and quantization unit 105 do not need to be the same. For example, quantization unit 115 may take fine quantization values for coefficients in the low frequency region, while quantization unit 105 may take fine quantization parameters for coefficients in the high frequency region.

[0065]

665 Furthermore, the quantization unit 115 for generating the lower layer bit stream may perform coding using a coarse quantization parameter, and the quantization unit 105 for generating the upper layer bit stream may set a fine quantization parameter.

[0066]

671 Next, a video decoding device according to a first embodiment of the present invention will be described with reference to FIG.

673 FIG. 4 is a block diagram illustrating the hardware configuration of the video encoding device according to the first embodiment of the present invention.

[0067]

678 A video decoding device 401 of this embodiment comprises a processor 402, a memory 403, an input interface (input I/F) 404, and an output interface (output I/F) 406, all of which are communicatively connected to one another.

681 The input I/F 404 is connected to an input device 405, and the output I/F 406 is connected to an output device 407.

[0068]

686 The processor 402 is a processor that performs the video decoding process of this embodiment, executes a program stored in the memory 403, decodes data received from the input I/F 404, and sends the data to the output I/F 406.

[0069]

692 The memory 403 stores the programs executed by the processor 402.

693 Also, data to be processed by the processor 402 is temporarily stored.

[0070]

697 The video decoding device 401 may be provided with a plurality of processors 402 and memories 403 .

697 For example, a dedicated processor may be provided for executing only a part of the program for performing the video decoding process of the present invention.

[0071]

704 The processor 402 and the memory 403 may be implemented on a single chip.

[0072]

708 The input I/F 404 is an interface that receives encoded data to be processed by the processor 402 from an input device 405 .

[0073]

713 The input device 405 is a device that inputs the encoded data to be processed by the video decoding device 401 to the input I/F 404, and is, for example, a storage device in which the encoded data is stored.

716 In this case, the input I/F 404 is, for example, a SATA interface.

717 Furthermore, the input device 405 may be a computer device connected to the input I/F 404 via a LAN, an IP network, or the like (not shown).

719 In this case, the input I/F 404 is a network interface.

720 The input device 405 may also be a video data transmission device connected to the input I/F 404 via a telephone communication network (not shown).

722 In this case, the input I/F 404 is a receiver of a telephone signal.

723 The input device 405 may also be a digital TV broadcast station.

724 In this case, the input I/F 404 is an interface provided by a digital TV tuner.

[0074]

728 The output I/F 406 is an interface that transmits the data decoded by the processor 402 to an output device 407 .

[0075]

733 The output device 407 is an output destination of the data decoded by the video decoding device 401, and is, for example, a display that outputs video.

735 In this case, the output I/F 406 is, for example, a video card.

736 The output device 407 is a storage device that records the decoded data in a stream.

737 In this case, the output I/F 406 is, for example, a SATA interface.

[0076]

742 The video decoding device 401 includes a plurality of input I/Fs 404 and a plurality of output I/Fs 406, and different types of input devices and output devices may be connected to the respective input I/Fs 404 and output I/Fs 406.

745 For example, the video decoding device 401 may be equipped with two input I/Fs 404, one of which may be connected to a hard disk device and the other to an optical disk device. Also, a hard disk device may be connected to one side, and a computer device may be connected to the other side via a LAN or the like.

[0077]

752 Next, the configuration and operation of a video decoding device according to an embodiment of the present invention will be described with reference to FIG.

754 FIG. 3 is a block diagram showing the configuration of a video decoding device according to an embodiment of the present invention.

[0078]

759 As shown in FIG. 3, the video decoding device 401 of this embodiment comprises variable length decoding units 300 and 310, inverse quantization units 301 and 311, inverse frequency transformation units 302 and 312, reference image memories 303 and 313, prediction units 304 and 314, and a half-pixel generation unit 321.

763 Here, the variable length decoding units 300 and 310, the inverse quantization units 301 and 311, the inverse frequency conversion units 302 and 312, and the prediction units 304 and 314 are programs stored in the memory 403 and executed by the processor 402, and the reference image memories 303 and 313 are partial areas of the memory 403. Furthermore, some or all of these programs may be realized by dedicated hardware that performs the processing.

[0079]

772 In addition, the variable length decoding units 300 and 310, the inverse quantization units 301 and 311, the inverse frequency conversion units 302 and 312, and the prediction units 304 and 314 may each be realized by the same hardware, and processing may be performed by using that hardware in a time-division manner.

[0080]

779 The video decoding device 401 of this embodiment can decode a stream that has been coded by the video coding device 201 having the configuration shown in FIG.

[0081]

785 The function and operation of each part of the video decoding device 401 will be explained below along with the procedure for decoding an encoded stream.

[0082]

790 The variable length decoding unit 310 analyzes the data of the input lower layer bit stream 131, and controls the mode related to inter-frame prediction (inter-prediction) and intra-frame prediction (intra-prediction) for the prediction unit 314 based on flag and data information.

794 When the mode is inter prediction, the prediction unit 314 performs inter-frame prediction using information such as motion vectors from the variable length decoding unit 310 and image data in the reference image memory 313 .

797 Furthermore, when the mode is the intra prediction mode, the prediction unit 314 performs intra-frame prediction using information such as the intra coding mode transmitted from the variable length decoding unit 310 and the image in the reference image memory 313 .

[0083]

803 The basic functions of the variable length decoding unit 310, the inverse quantization unit 311, and the inverse frequency transformation unit 312 are the same as those of conventional decoding devices, and each unit performs stream decoding and information analysis, inverse quantization of transformation coefficients, and inverse frequency transformation, etc.

[0084]

811 A low-resolution image 331 is generated by adding together the predicted image generated using the reference image memory 313 and the prediction unit 314 and the difference image output by the inverse frequency conversion unit 312. The low-resolution image 331 is then transmitted to the output device 407, where the decoded image is displayed on the screen.

[0085]

818 Similarly, the enhancement layer bit stream is also decoded in the same manner using a variable length decoding unit 300 , an inverse quantization unit 301 , an inverse frequency conversion unit 302 , a reference image memory 303 , and a prediction unit 304 .

821 In this case, if the upper layer bit stream 130 is compressed and transmitted without performing prediction, the output of the prediction unit 304 is not added to the output image of the inverse frequency transform unit 302, or image data whose values are all 0 is added.

824 In this way, image data is restored by decoding the enhancement layer bit stream.

825 A half-pixel precision image output from a half-pixel generation unit 321 is added to this to generate a high-resolution image 330 .

825 The half-pixel generating section 321 performs the same half-pixel generating process as that performed by the pixel generating section 121 shown in FIG.

829 Here, the half pixel generation unit 121 forms an upsampling unit that performs upsampling processing to add an image generated from a lower layer stream to an image generated from a higher layer stream.

[0086]

835 The video encoding device 201 and the video decoding device 401 of the present embodiment described above may be implemented as the same hardware.

837 In this case, the memory 203 (or the memory 403) stores the programs described with reference to FIGS. 1 and 3, and also secures the areas described with reference to FIGS.

[0087]

842 As described above, according to this embodiment, it is possible to transmit a lower layer bit stream obtained by downsampling and encoding an input image, and an upper layer bit stream obtained by decoding the bit stream and compressing and encoding an image obtained by taking the difference between the input image and the image generated by the half-pixel generation process, or an image obtained by taking the difference between the images generated by the super-resolution process and the half-pixel generation process.

848 As a result, the high compression of image data makes it possible to utilize the transmission band more efficiently than before and to realize communication at a lower transmission rate than simulcast. This makes it possible to realize suitable transmission for each receiving terminal that receives the data.

[0088]

855 Second Embodiment Hereinafter, a video encoding device and a video decoding device according to a second embodiment of the present invention will be described with reference to FIGS.

[0089]

861 The encoding device and decoding device described in this embodiment are different from those in the first embodiment in that the form of the process of upsampling an image generated from a lower layer bitstream is changed.

[0090]

867 First, the configuration and operation of a video encoding device according to the second embodiment of the present invention will be described with reference to FIG.

867 FIG. 5 is a block diagram showing the configuration of a video encoding device according to the second embodiment of the present invention.

[0091]

874 The configuration of the encoding device of the first embodiment shown in Figure 1 differs in that the difference between the input image stored in the original image memory 101 and the output image from the super-resolution processing unit 120 is taken and input to the block division unit 102.

878 The other processes are similar to those of the encoding device of the first embodiment.

[0092]

882 Even with this configuration, it is possible to realize video coding with higher image quality in the decoded image when compared with the prior art, and it is possible to transmit a restored image with image quality equivalent to that of the original image at the maximum.

[0093]

888 Next, the configuration and operation of a video decoding device according to a second embodiment of the present invention will be described with reference to FIG.

890 FIG. 6 is a block diagram showing the configuration of a video decoding device according to the second embodiment of the present invention.

[0094]

895 The difference with respect to the configuration of the decoding device of the first embodiment shown in FIG. 3 is that the half pixel generating unit 320 is replaced by a super-resolution processing unit 1020 .

898 A high-resolution image with high image quality can be restored by adding together a low-resolution image 331 decoded from the lower layer bit stream 131 that has been super-resolution processed by the super-resolution processing unit 1020 and an image decoded from the upper layer bit stream 900.

[0095]

905 As described above, by using the video encoding device and decoding device of this embodiment, it is possible to realize transmission using image data with higher image quality than with conventional techniques, similar to the first embodiment.

[0096]

912 1 is a block diagram showing a configuration of a video encoding device according to a first embodiment of the present invention.

914 1 is a block diagram illustrating a hardware configuration of a video encoding device according to a first embodiment of the present invention.

916 1 is a block diagram showing a configuration of a video decoding device according to an embodiment of the present invention;

918 1 is a block diagram illustrating a hardware configuration of a video encoding device according to a first embodiment of the present invention.

920 FIG. 11 is a block diagram showing a configuration of a video encoding device according to a second embodiment of the present invention.

922 FIG. 11 is a block diagram showing the configuration of a video decoding device according to a second embodiment of the present invention.

924 FIG. 1 is a diagram illustrating an example of the configuration of an SNR scalability encoding device.

926 FIG. 1 is a diagram illustrating an example of the configuration of an SNR scalability decoding device.

928 FIG. 1 is a diagram illustrating an example of the configuration of a spatial scalability coding device.

930 FIG. 1 is a diagram illustrating an example of the configuration of a spatial scalability decoding device.

932 Explanation of symbols

[0097]

936 100,500... Input image 101... Original image memory 102,112... Block division unit 103,113,304,314... Prediction unit 104,114... Frequency conversion unit 105,115... Quantization unit 106,116... Variable length coding unit 107,117,301,311... Inverse quantization unit 108,118,302,312... Inverse frequency conversion unit 109,119,303,313... Reference image memory 110,700... Downsampling unit 120... Super-resolution processing unit 121,321... Half pixel generation unit 130,710,900... Upper layer bit stream 131,711... Lower layer bit stream 201... Video coding device 202,402... Processor 203,403... Memory 204, 404...input interface (I/F) 205, 405...input device 206, 406...output interface (I/F) 207, 407...output device 300, 310...variable length decoding section 330, 1010...high resolution image 331...low resolution image 401...video decoding device 501, 511...DCT section 502, 512...Q section 503, 513...VLC section 504, 514, 601, 603, 801, 811...IQ section 505, 515, 604, 802, 812...IDCT section 506, 516, 605, 803, 813...FM section 507, 517, 606, 804, 814...MC section 510...Delay section 540: Enhancement bit stream 541: Low SNR bit stream 600, 602, 800, 810: VLD section 701: Upsampling section.