

明 細 書

発明の名称：復号装置、符号化装置、復号方法及び符号化方法

技術分野

[0001] 本開示は、復号装置等に関する。

背景技術

[0002] ビデオコーディング技術は、H. 261およびMPEG-1から、H. 264/AVC (Advanced Video Coding)、MPEG-LA、H. 265/HEVC (High Efficiency Video Coding)、およびH. 266/VVC (Versatile Video Codec) へ進歩している。この進歩に伴い、様々な用途において増え続けるデジタルビデオデータ量进行处理するために、ビデオコーディング技術の改良および最適化を提供することが常に必要とされている。本開示は、ビデオコーディングにおけるさらなる進歩、改良および最適化に関する。

[0003] なお、非特許文献1は、上述されたビデオコーディング技術に関する従来の規格の一例に関する。

先行技術文献

非特許文献

[0004] 非特許文献1：H. 265 (ISO/IEC 23008-2 HEVC) / HEVC (High Efficiency Video Coding)

発明の概要

発明が解決しようとする課題

[0005] 上記のような符号化方式に関して、符号化効率の改善、画質の改善、処理量の削減、回路規模の削減、又は、フィルタ、ブロック、サイズ、動きベクトル、参照ピクチャ又は参照ブロック等の要素又は動作の適切な選択等のため、新たな方式の提案が望まれている。

[0006] 本開示は、例えば、符号化効率の改善、画質の改善、処理量の削減、回路

規模の削減、処理速度の改善、及び、要素又は動作の適切な選択等のうち1つ以上に貢献し得る構成又は方法を提供する。なお、本開示は、上記以外の利益に貢献し得る構成又は方法を含み得る。

課題を解決するための手段

[0007] 例えば、本開示の一態様に係る復号装置は、メモリと、前記メモリに接続される回路とを備え、前記回路は、前記メモリを用いて、顔動画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームから復号し、前記顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームから復号し、前記ベースデータ、前記1つ以上の拡張データ及び前記幾何学情報から、生成モデルを用いて、前記顔動画像を生成し、前記ベースデータは、前記ビットストリームにおいて、前記顔動画像のフレームである第1フレームに対応するデータに付加されており、前記1つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画像の1つ以上のフレームであり、前記第1フレームよりも後の1つ以上のフレームである1つ以上の第2フレームに対応する1つ以上のデータに付加されている、復号装置である。

[0008] 本開示における各実施の形態、またはその一部の構成もしくは方法のそれぞれは、例えば、符号化効率の改善、画質の改善、符号化／復号の処理量の削減、回路規模の削減、または、符号化／復号の処理速度の改善などのうちの、少なくともいずれか1つを可能にする。あるいは、本開示における各実施の形態、またはその一部の構成もしくは方法のそれぞれは、符号化および復号において、フィルタ、ブロック、サイズ、動きベクトル、参照ピクチャ、参照ブロックなどの構成要素／動作の適切な選択などを可能にする。なお、本開示は、上記以外の利益を提供し得る構成または方法の開示も含む。例えば、処理量の増加を抑えつつ、符号化効率を改善する構成または方法などである。

[0009] 本開示の一態様におけるさらなる利点および効果は、明細書および図面か

ら明らかにされる。かかる利点および／または効果は、いくつかの実施の形態並びに明細書および図面に記載された特徴によってそれぞれ得られるが、1つまたはそれ以上の利点および／または効果を得るために必ずしも全てが提供される必要はない。

[0010] なお、これらの全般的または具体的な態様は、システム、集積回路、コンピュータプログラム又はコンピュータ読み取り可能なＣＤ－ＲＯＭなどの記録媒体で実現されてもよく、システム、方法、集積回路、コンピュータプログラム及び記録媒体の任意な組み合わせで実現されてもよい。

発明の効果

[0011] 本開示の一態様に係る構成又は方法は、例えば、符号化効率の改善、画質の改善、処理量の削減、回路規模の削減、処理速度の改善、及び、要素又は動作の適切な選択等のうち1つ以上に貢献し得る。なお、本開示の一態様に係る構成又は方法は、上記以外の利益に貢献してもよい。

図面の簡単な説明

[0012] [図1]図1は、実施の形態における符号化復号システムの構成例を示すブロック図である。

[図2]図2は、ストリームにおけるデータの階層構造の一例を示す図である。

[図3]図3は、実施の形態における符号化装置の構成例を示すブロック図である。

[図4]図4は、実施の形態における復号装置の構成例を示すブロック図である。

[図5]図5は、顔画像の例を示す概念図である。

[図6]図6は、幾何学属性の例を示す概念図である。

[図7]図7は、顔動画画像の例を示す概念図である。

[図8]図8は、各フレームに対応するデータの構成例を示す概念図である。

[図9]図9は、ＣＰＢバッファにおけるデータのサイズを示す概念図である。

[図10]図10は、各フレームに対応するデータの別の構成例を示す概念図である。

[図11]図11は、各フレームに対応するデータのさらに別の構成例を示す概念図である。

[図12]図12は、ベースデータ及び拡張データの構成例を示す概念図である。

[図13]図13は、ベースデータ及び拡張データの別の構成例を示す概念図である。

[図14]図14は、ベースデータ及び拡張データのさらに別の構成例を示す概念図である。

[図15]図15は、ベースデータ及び拡張データのさらに別の構成例を示す概念図である。

[図16]図16は、ベースデータ及び拡張データのさらに別の構成例を示す概念図である。

[図17]図17は、各フレームに対応するデータのさらに別の構成例を示す概念図である。

[図18]図18は、各フレームに対応するデータのさらに別の構成例を示す概念図である。

[図19]図19は、顔動画像の生成の制御例を示す概念図である。

[図20]図20は、実施の形態における符号化装置の別の構成例を示すブロック図である。

[図21]図21は、実施の形態における復号装置の別の構成例を示すブロック図である。

[図22]図22は、実施の形態における符号化装置の動作例を示すフローチャートである。

[図23]図23は、実施の形態における復号装置の動作例を示すフローチャートである。

[図24]図24は、実施の形態における符号化装置のさらに別の構成例を示すブロック図である。

[図25]図25は、実施の形態における符号化装置の別の動作例を示すフロー

チャートである。

[図26]図26は、実施の形態における復号装置のさらに別の構成例を示すブロック図である。

[図27]図27は、実施の形態における復号装置の別の動作例を示すフローチャートである。

[図28]図28は、実施の形態における復号装置のさらに別の構成例を示すブロック図である。

[図29]図29は、実施の形態における復号装置のさらに別の構成例を示すブロック図である。

[図30]図30は、生成モデルとして用いられ得る様々なニューラルネットワークの例を示す図である。

[図31]図31は、実施の形態における符号化装置が動画像を符号化するための構成例を示すブロック図である。

[図32]図32は、実施の形態における復号装置が動画像を復号するための構成例を示すブロック図である。

[図33]図33は、実施の形態における符号化装置の実装例を示すブロック図である。

[図34]図34は、実施の形態における符号化装置の基本的な動作例を示すフローチャートである。

[図35]図35は、実施の形態における復号装置の実装例を示すブロック図である。

[図36]図36は、実施の形態における復号装置の基本的な動作例を示すフローチャートである。

[図37]図37は、コンテンツ配信サービスを実現するコンテンツ供給システムの全体構成図である。

[図38]図38は、webページの表示画面例を示す図である。

[図39]図39は、webページの表示画面例を示す図である。

[図40]図40は、スマートフォンの一例を示す図である。

[図41]図41は、スマートフォンの構成例を示すブロック図である。

発明を実施するための形態

[0013] [序論 (Introduction)]

例えば、ビデオ会議などにおいて、符号化装置は、顔を含む動画像をビットストリームに符号化する。そして、復号装置は、当該動画像をビットストリームから復号する。

[0014] 符号量の削減のため、符号化装置は、最初のフレームにおいて顔画像を符号化し、顔を含む領域内の幾何学属性を示す幾何学情報をフレーム毎に符号化してもよい。そして、復号装置は、最初のフレームにおいて顔画像を復号し、幾何学情報をフレーム毎に復号してもよい。そして、復号装置は、最初のフレームにおいて復号された顔画像と、フレーム毎に復号された幾何学情報とに基づいて、顔画像をフレーム毎に再現してもよい。

[0015] ここで、幾何学属性は、例えば動的属性に対応し、顔のランドマークのような点群で表現されてもよいし、複数のポリゴンの組み合わせによって物体の形状を表現するためのポリゴンモデルで表現されてもよい。また、幾何学属性は、その他の幾何学的モデルで表現されてもよい。また、幾何学属性は、顔のパーツの位置で表現されてもよい。

[0016] 幾何学情報の符号量は、顔画像の符号量よりも少ないと想定される。したがって、フレーム毎に幾何学情報の符号化及び復号を行うことで、フレーム毎に顔画像の符号化及び復号を行うよりも少ない符号量で顔動画像を再現することが可能になる。

[0017] しかしながら、最初のフレームにおいて顔画像の符号化及び復号を行うことで、最初のフレームに対応する符号量が増加し、遅延が発生する可能性がある。

[0018] そこで、例1の復号装置は、メモリと、前記メモリに接続される回路とを備え、前記回路は、前記メモリを用いて、顔動画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームから復号し、前記顔動画像の複数のフレームのそれぞれに対応する情報であり、人

物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームから復号し、前記ベースデータ、前記1つ以上の拡張データ及び前記幾何学情報から、生成モデルを用いて、前記顔動画像を生成し、前記ベースデータは、前記ビットストリームにおいて、前記顔動画像のフレームである第1フレームに対応するデータに付加されており、前記1つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画像の1つ以上のフレームであり、前記第1フレームよりも後の1つ以上のフレームである1つ以上の第2フレームに対応する1つ以上のデータに付加されている、復号装置である。

[0019] これにより、顔画像に関するベースデータ及び1つ以上の拡張データを複数のフレームに分けて復号することが可能になる場合がある。したがって、1つのフレームに対応する符号量を削減することが可能になる場合がある。よって、遅延を抑制することが可能になる場合がある。

[0020] また、例2の復号装置は、例1の復号装置であって、前記回路は、前記ベースデータと前記1つ以上の拡張データとである複数の顔画像データの少なくとも1つの制御に関する制御情報をヘッダ領域から復号する、復号装置であってもよい。

[0021] これにより、顔動画像の再現において、制御情報に従って、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0022] また、例3の復号装置は、例2の復号装置であって、前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの1つである顔画像データが含まれるか否かを示す存在情報を含む、復号装置であってもよい。

[0023] これにより、顔動画像の再現において、制御情報に従って、アクセスユニットに顔画像データが含まれるか否かの識別が可能になる場合がある。したがって、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0024] また、例4の復号装置は、例2又は例3の復号装置であって、前記制御情

報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの1つである顔画像データが含まれる場合、前記顔画像データが前記ベースデータであるか前記1つ以上の拡張データの1つである拡張データであるかに関する種別情報を含み、前記種別情報は、前記アクセスユニットが前記ベースデータを含む場合、前記アクセスユニットに含まれる前記顔画像データが、前記ベースデータであり、次のベースデータまで継続して用いられることを示し、前記種別情報は、前記アクセスユニットが前記拡張データを含む場合、前記アクセスユニットに含まれる前記顔画像データが、前記拡張データであり、前記ベースデータと共に用いられることを示す、復号装置であってもよい。

[0025] これにより、顔動画像の再現において、制御情報に従って、顔画像データがベースデータであるか拡張データであるかの識別が可能になる場合がある。そして、顔画像データがベースデータであるか拡張データであるかに従って、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0026] また、例5の復号装置は、例2～例4のいずれかの復号装置であって、前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの1つである顔画像データが含まれる場合、前記顔動画像の複数のフレームのうち前記アクセスユニットに対応するフレームの生成及び表示に前記顔画像データが適用可能であるか否かを示す適用情報を含む、復号装置であってもよい。

[0027] これにより、顔動画像の再現において、制御情報に従って、顔動画像のフレームに、当該フレームに対応するデータに付加された顔画像データを適用するか否かを適切に制御することが可能になる場合がある。

[0028] また、例6の復号装置は、例1～例5のいずれかの復号装置であって、前記ベースデータと前記1つ以上の拡張データとのそれぞれは、前記顔画像に含まれる顔の特徴を示すベクトルで表現される、復号装置であってもよい。

[0029] これにより、顔画像に関する符号量を削減することが可能になる場合があ

る。したがって、遅延を抑制することが可能になる場合がある。

[0030] また、例 7 の復号装置は、例 1 ～例 5 のいずれかの復号装置であって、前記ベースデータと前記 1 つ以上の拡張データとのそれぞれは、前記顔画像に関する画像で表現される、復号装置であってもよい。

[0031] これにより、顔動画画像の再現において、顔画像に関するベースデータ及び拡張データのそれぞれを画像データとして顔動画画像のフレームに適切に反映することが可能になる場合がある。

[0032] また、例 8 の復号装置は、例 1 ～例 7 のいずれかの復号装置であって、前記回路は、前記ベースデータと前記 1 つ以上の拡張データの少なくとも 1 つと前記幾何学情報とを前記生成モデルに入力して、前記顔動画画像のフレームを生成する、復号装置であってもよい。

[0033] これにより、顔画像に関するベースデータと拡張データとから中間画像を生成する処理を省略することが可能になる場合がある。したがって、顔動画画像を生成する処理を簡略化することが可能になる場合がある。

[0034] また、例 9 の復号装置は、例 1 ～例 7 のいずれかの復号装置であって、前記回路は、前記ベースデータと前記 1 つ以上の拡張データの少なくとも 1 つとから中間画像を生成し、前記中間画像と前記幾何学情報とを前記生成モデルに入力して、前記顔動画画像のフレームを生成する、復号装置であってもよい。

[0035] これにより、顔画像に関するベースデータと拡張データとから、顔画像に関する中間画像を適切に生成することが可能になる場合がある。そして、顔画像に関する中間画像を顔動画画像のフレームに適切に反映することが可能になる場合がある。

[0036] また、例 10 の復号装置は、例 1 ～例 7 のいずれかの復号装置であって、前記回路は、前記ベースデータを参照して、前記 1 つ以上の拡張データの 1 つである拡張データを復号し、前記拡張データと前記幾何学情報とを前記生成モデルに入力して、前記顔動画画像のフレームを生成する、復号装置であってもよい。

- [0037] これにより、ベースデータの精度よりも高い精度を有する拡張データを効率的に復号することが可能になる場合がある。そして、高精度の拡張データを用いて、顔動画像のフレームを高精度に生成することが可能になる場合がある。
- [0038] また、例 11 の復号装置は、例 1 ～例 9 のいずれかの復号装置であって、前記ベースデータは、前記顔画像に含まれる顔の一部のデータであり、前記 1 つ以上の拡張データの 1 つである拡張データは、前記顔画像に含まれる前記顔の他部のデータである、復号装置であってもよい。
- [0039] これにより、複雑な処理を伴わずに、顔画像を複数の部分に分けて復号することが可能になる場合がある。
- [0040] また、例 12 の復号装置は、例 1 ～例 9 のいずれかの復号装置であって、前記ベースデータは、前記顔画像における第 1 周波数範囲のデータであり、前記 1 つ以上の拡張データの 1 つである拡張データは、前記顔画像における前記第 1 周波数範囲よりも高い第 2 周波数範囲のデータである、復号装置であってもよい。
- [0041] これにより、顔画像における低周波数成分のデータをベースデータとして復号し、顔画像における高周波数成分のデータを拡張データとして復号することが可能になる場合がある。そして、顔動画像の再現において、低周波数成分のデータを顔動画像の第 1 フレームの生成に適用し、低周波数成分のデータと高周波数成分のデータとの両方を顔動画像の第 2 フレームの生成に適用することが可能になる場合がある。よって、遅延を抑制しつつ、顔動画像における違和感を抑制することが可能になる場合がある。
- [0042] また、例 13 の復号装置は、例 1 ～例 10 のいずれかの復号装置であって、前記ベースデータは、前記顔画像に関する画像であり、第 1 解像度を有する画像である第 1 画像に対応し、前記 1 つ以上の拡張データの 1 つである拡張データは、前記顔画像に関する画像であり、前記第 1 画像を参照して復号される画像であり、前記第 1 解像度よりも高い第 2 解像度を有する画像である第 2 画像に対応する、復号装置であってもよい。

- [0043] これにより、低解像度の顔画像をベースデータとして復号し、高解像度の顔画像を拡張データとして復号することが可能になる場合がある。そして、顔動画像の再現において、低解像度の顔画像を顔動画像の第1フレームの生成に適用し、高解像度の顔画像を顔動画像の第2フレームの生成に適用することが可能になる場合がある。よって、遅延を抑制しつつ、顔動画像における違和感を抑制することが可能になる場合がある。
- [0044] また、例14の復号装置は、例1～例10のいずれかの復号装置であって、前記ベースデータは、前記顔画像に関する画像であり、第1量子化幅で復号される画像である第1画像に対応し、前記1つ以上の拡張データの1つである拡張データは、前記顔画像に関する画像であり、前記第1画像を参照して前記第1量子化幅よりも細かい第2量子化幅で復号される画像である第2画像に対応する、復号装置であってもよい。
- [0045] これにより、粗い顔画像をベースデータとして復号し、細かい顔画像を拡張データとして復号することが可能になる場合がある。そして、顔動画像の再現において、粗い顔画像を顔動画像の第1フレームの生成に適用し、細かい顔画像を顔動画像の第2フレームの生成に適用することが可能になる場合がある。よって、遅延を抑制しつつ、顔動画像における違和感を抑制することが可能になる場合がある。
- [0046] また、例15の復号装置は、例2～例5のいずれかの復号装置であって、前記制御情報は、前記1つ以上の拡張データのそれぞれを識別するための識別情報を含む、復号装置であってもよい。
- [0047] これにより、顔動画像の再現において、制御情報に従って、各拡張データを識別することが可能になる場合がある。したがって、各拡張データを個別に指定し、各拡張データの適用を制御することが可能になる場合がある。
- [0048] また、例16の復号装置は、例2～例5及び例15のいずれかの復号装置であって、前記制御情報は、(i)前記ベースデータを含むアクセスユニットの前記ヘッダ領域に含まれる情報であり、(ii)前記1つ以上の拡張データの総数を示す情報である総数情報を含む、復号装置であってもよい。

- [0049] これにより、顔動画像の再現において、制御情報に従って、1つ以上の拡張データの総数を識別することが可能になる場合がある。したがって、1つ以上の拡張データの総数に従って、顔動画像の再現に用いられ得る1つ以上の拡張データを効率的に確定することが可能になる場合がある。
- [0050] また、例17の復号装置は、例2～例5、例15及び例16のいずれかの復号装置であって、前記制御情報は、(i)前記ベースデータを含むアクセスユニットの前記ヘッダ領域に含まれる情報であり、(ii)前記1つ以上の拡張データのうちの拡張データであって、前記1つ以上の第2フレームのうち当該拡張データを含むアクセスユニットに対応する第2フレームの生成及び表示に適用可能な拡張データを特定するための情報である特定情報を含む、復号装置であってもよい。
- [0051] これにより、顔動画像の再現において、制御情報に従って、顔動画像のフレームの生成及び表示に適用可能な拡張データを適切に特定することが可能になる場合がある。
- [0052] また、例18の復号装置は、例1～例17のいずれかの復号装置であって、前記回路は、前記メモリにおいて前記ビットストリームが記憶されるストリームバッファを制御するための少なくとも1つの制御パラメータであって、前記ストリームバッファのバッファサイズを基準サイズ以下に制御し、復号処理開始の初期遅延時間を基準遅延時間以下に制御するための少なくとも1つの制御パラメータを復号する、復号装置であってもよい。
- [0053] これにより、復号のためのリソースを削減することが可能になり、遅延時間を短くすることが可能になる。
- [0054] また、例19の符号化装置は、メモリと、前記メモリに接続される回路とを備え、前記回路は、前記メモリを用いて、顔動画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームに符号化し、前記顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームに符号化し、前記ベースデータは、前記ビットストリームにお

いて、前記顔動画像のフレームである第1フレームに対応するデータに付加され、前記1つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画像の1つ以上のフレームであり、前記第1フレームよりも後の1つ以上のフレームである1つ以上の第2フレームに対応する1つ以上のデータに付加される、符号化装置である。

[0055] これにより、顔画像に関するベースデータ及び1つ以上の拡張データを複数のフレームに分けて符号化することが可能になる場合がある。したがって、1つのフレームに対応する符号量を削減することが可能になる場合がある。よって、遅延を抑制することが可能になる場合がある。

[0056] また、例20の符号化装置は、例19の符号化装置であって、前記回路は、前記ベースデータと前記1つ以上の拡張データとである複数の顔画像データの少なくとも1つの制御に関する制御情報をヘッダ領域に符号化する、符号化装置であってもよい。

[0057] これにより、顔動画像の再現において、制御情報に従って、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0058] また、例21の符号化装置は、例20の符号化装置であって、前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの1つである顔画像データが含まれるか否かを示す存在情報を含む、符号化装置であってもよい。

[0059] これにより、顔動画像の再現において、制御情報に従って、アクセスユニットに顔画像データが含まれるか否かの識別が可能になる場合がある。したがって、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0060] また、例22の符号化装置は、例20又は例21の符号化装置であって、前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの1つである顔画像データが含まれる場合、前記顔画像データが前記ベースデータであるか前記1つ以上の拡張データの1つである拡張データであるかに関する種別情報を含み、前記種別情報は、前記アク

セスユニットが前記ベースデータを含む場合、前記アクセスユニットに含まれる前記顔画像データが、前記ベースデータであり、次のベースデータまで継続して用いられることを示し、前記種別情報は、前記アクセスユニットが前記拡張データを含む場合、前記アクセスユニットに含まれる前記顔画像データが、前記拡張データであり、前記ベースデータと共に用いられることを示す、符号化装置であってもよい。

[0061] これにより、顔動画像の再現において、制御情報に従って、顔画像データがベースデータであるか拡張データであるかの識別が可能になる場合がある。そして、顔画像データがベースデータであるか拡張データであるかに従って、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0062] また、例 2 3 の符号化装置は、例 2 0～例 2 2 のいずれかの符号化装置であって、前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの 1 つである顔画像データが含まれる場合、前記顔動画像の複数のフレームのうち前記アクセスユニットに対応するフレームの生成及び表示に前記顔画像データが適用可能であるか否かを示す適用情報を含む、符号化装置であってもよい。

[0063] これにより、顔動画像の再現において、制御情報に従って、顔動画像のフレームに、当該フレームに対応するデータに付加された顔画像データを適用するか否かを適切に制御することが可能になる場合がある。

[0064] また、例 2 4 の符号化装置は、例 1 9～例 2 3 のいずれかの符号化装置であって、前記ベースデータと前記 1 つ以上の拡張データとのそれぞれは、前記顔画像に含まれる顔の特徴を示すベクトルで表現される、符号化装置であってもよい。

[0065] これにより、顔画像に関する符号量を削減することが可能になる場合がある。したがって、遅延を抑制することが可能になる場合がある。

[0066] また、例 2 5 の符号化装置は、例 1 9～例 2 3 のいずれかの符号化装置であって、前記ベースデータと前記 1 つ以上の拡張データとのそれぞれは、前

記顔画像に関する画像で表現される、符号化装置であってもよい。

[0067] これにより、顔動画像の再現において、顔画像に関するベースデータ及び拡張データのそれぞれを画像データとして顔動画像のフレームに適切に反映することが可能になる場合がある。

[0068] また、例 2 6 の符号化装置は、例 1 9 ～例 2 5 のいずれかの符号化装置であって、前記回路は、前記顔画像に含まれる顔の一部のデータを前記ベースデータとして導出して符号化し、前記顔画像に含まれる前記顔の他部のデータを前記 1 つ以上の拡張データの 1 つである拡張データとして導出して符号化する、符号化装置であってもよい。

[0069] これにより、複雑な処理を伴わずに、顔画像を複数の部分に分けて符号化することが可能になる場合がある。

[0070] また、例 2 7 の符号化装置は、例 1 9 ～例 2 5 のいずれかの符号化装置であって、前記回路は、前記顔画像における第 1 周波数範囲のデータを前記ベースデータとして導出して符号化し、前記顔画像における前記第 1 周波数範囲よりも高い第 2 周波数範囲のデータを前記 1 つ以上の拡張データの 1 つである拡張データとして導出して符号化する、符号化装置であってもよい。

[0071] これにより、顔画像における低周波数成分のデータをベースデータとして符号化し、顔画像における高周波数成分のデータを拡張データとして符号化することが可能になる場合がある。そして、顔動画像の再現において、低周波数成分のデータを顔動画像の第 1 フレームの生成に適用し、低周波数成分のデータと高周波数成分のデータとの両方を顔動画像の第 2 フレームの生成に適用することが可能になる場合がある。よって、遅延を抑制しつつ、顔動画像における違和感を抑制することが可能になる場合がある。

[0072] また、例 2 8 の符号化装置は、例 1 9 ～例 2 5 のいずれかの符号化装置であって、前記回路は、前記顔画像に関する画像であり、第 1 解像度を有する画像である第 1 画像を前記ベースデータとして符号化し、前記顔画像に関する画像であり、前記第 1 画像を参照して符号化される画像であり、前記第 1 解像度よりも高い第 2 解像度を有する画像である第 2 画像を前記ベースデー

タとして符号化する、符号化装置であってもよい。

[0073] これにより、低解像度の顔画像をベースデータとして符号化し、高解像度の顔画像を拡張データとして符号化することが可能になる場合がある。そして、顔動画画像の再現において、低解像度の顔画像を顔動画画像の第1フレームの生成に適用し、高解像度の顔画像を顔動画画像の第2フレームの生成に適用することが可能になる場合がある。よって、遅延を抑制しつつ、顔動画画像における違和感を抑制することが可能になる場合がある。

[0074] また、例29の復号方法は、顔動画画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームから復号し、前記顔動画画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームから復号し、前記ベースデータ、前記1つ以上の拡張データ及び前記幾何学情報から、生成モデルを用いて、前記顔動画画像を生成し、前記ベースデータは、前記ビットストリームにおいて、前記顔動画画像のフレームである第1フレームに対応するデータに付加されており、前記1つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画画像の1つ以上のフレームであり、前記第1フレームよりも後の1つ以上のフレームである1つ以上の第2フレームに対応する1つ以上のデータに付加されている、復号方法である。

[0075] これにより、顔画像に関するベースデータ及び1つ以上の拡張データを複数のフレームに分けて復号することが可能になる場合がある。したがって、1つのフレームに対応する符号量を削減することが可能になる場合がある。よって、遅延を抑制することが可能になる場合がある。

[0076] また、例30の符号化方法は、顔動画画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームに符号化し、前記顔動画画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームに符号化し、前記ベースデータは、前記ビットストリームにおいて、前記顔動画画像のフレームである第1フレームに対応するデータに付加され、前記

1つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画像の1つ以上のフレームであり、前記第1フレームよりも後の1つ以上のフレームである1つ以上の第2フレームに対応する1つ以上のデータに付加される、符号化方法である。

[0077] これにより、顔画像に関するベースデータ及び1つ以上の拡張データを複数のフレームに分けて符号化することが可能になる場合がある。したがって、1つのフレームに対応する符号量を削減することが可能になる場合がある。よって、遅延を抑制することが可能になる場合がある。

[0078] さらに、これらの包括的又は具体的な態様は、システム、装置、方法、集積回路、コンピュータプログラム、又は、コンピュータ読み取り可能なCD-ROMなどの非一時的な記録媒体で実現されてもよく、システム、装置、方法、集積回路、コンピュータプログラム、及び、記録媒体の任意な組み合わせで実現されてもよい。

[0079] [用語の定義]

各用語は一例として以下のような定義であってもよい。

[0080] (1) 画像

画素の集合によって構成されたデータの単位であり、ピクチャやピクチャより小さいブロックからなり、動画の他、静止画も含む。

[0081] (2) ピクチャ

画素の集合によって構成される画像の処理単位であり、フレームやフィールドと呼ばれる場合もある。

[0082] (3) ブロック

特定数の画素を含む集合の処理単位であり、以下の例に挙げる通り、名称は問わない。また、形状も問わず、例えば、 $M \times N$ 画素からなる長方形、 $M \times M$ 画素からなる正方形はもちろん、三角形、円形、その他の形状も含む。

[0083] (ブロックの例)

- ・スライス／タイル／ブリック
- ・CTU／スーパーブロック／基本分割単位

- ・ V P D U / ハードウェアの処理分割単位
- ・ C U / 処理ブロック単位 / 予測ブロック単位 (P U) / 直交変換ブロック単位 (T U) / ユニット
- ・ サブブロック

[0084] (4) 画素 / サンプル

画像を構成する最小単位の点であって、整数位置の画素のみならず整数位置の画素に基づいて生成された小数位置の画素も含む。

[0085] (5) 画素値 / サンプル値

画素が有する固有の値であって、輝度値、色差値、R G B の階調はもちろん、d e p t h 値、又は 0、1 の 2 値も含む。

[0086] (6) フラグ

1 ビットその他、複数ビットの場合も含み、例えば、2 ビット以上のパラメータやインデックスであってもよい。また、二進数を用いた 2 値のみならず、その他の進数を用いた多値であってもよい。

[0087] (7) 信号

情報を伝達するために記号化、符号化したものであって、離散化されたデジタル信号その他、連続値を取るアナログ信号も含む。

[0088] (8) ストリーム / ビットストリーム

デジタルデータのデータ列又はデジタルデータの流れをいう。ストリーム / ビットストリームは、1 本のストリームその他、複数の階層に分けられ複数のストリームにより構成されてもよい。また、単数の伝送路でシリアル通信により伝送される場合その他、複数の伝送路でパケット通信により伝送される場合も含む。

[0089] (9) 差 / 差分

スカラー量の場合、単純差 ($x - y$) の他、差の演算が含まれていれば足り、差の絶対値 ($|x - y|$)、二乗差 ($x^2 - y^2$)、差の平方根 ($\sqrt{x - y}$)、重み付け差 ($a x - b y$: a 、 b は定数)、オフセット差 ($x - y + a$: a はオフセット) を含む。

[0090] (10) 和

スカラー量の場合、単純和 ($x + y$) の他、和の演算が含まれていれば足り、和の絶対値 ($|x + y|$)、二乗和 ($x^2 + y^2$)、和の平方根 ($\sqrt{x + y}$)、重み付け和 ($ax + by$: a 、 b は定数)、オフセット和 ($x + y + a$: a はオフセット) を含む。

[0091] (11) 基づいて (based on)

基づく対象となる要素以外を加味する場合も含む。また、直接結果を求める場合の他、中間的な結果を経由して結果を求める場合も含む。

[0092] (12) 用いて (used, using)

用いる対象となる要素以外を加味する場合も含む。また、直接結果を求める場合の他、中間的な結果を経由して結果を求める場合も含む。

[0093] (13) 禁止する (prohibit, forbid)

許されないと言い換えることができる。また、禁止していないこと又は許可されることは、必ずしも義務を意味するものではない。

[0094] (14) 制限する (limit, restriction/restrict/restricted)

許されないと言い換えることができる。また、禁止していないこと又は許可されることは、必ずしも義務を意味するものではない。さらに、量的又は質的に一部が禁止されていれば足り、全面的に禁止する場合も含まれる。

[0095] (15) 色差 (chroma)

サンプル配列または単一のサンプルが、原色に関連する2つの色差 (colour difference) 信号の1つを表すことを指定する、記号 C_b および C_r で表される形容詞である。chroma という用語の代わりに、chrominance という用語を使用することもできる。

[0096] (16) 輝度 (luma)

サンプル配列または単一のサンプルが原色に関連するモノクロ信号を表すことを指定する、記号または下付きの Y または L で表される形容詞である。luma という用語の代わりに、luminance という用語を使用する

こともできる。

[0097] [記載に関する解説]

図面において、同一の参照番号は同一または類似の構成要素を示す。また、図面における構成要素のサイズおよび相対位置は、必ずしも一定の縮尺で描かれていない。

[0098] 以下、実施の形態について図面を参照しながら具体的に説明する。なお、以下で説明する実施の形態は、いずれも包括的または具体的な例を示すものである。以下の実施の形態で示される数値、形状、材料、構成要素、構成要素の配置位置及び接続形態、ステップ、ステップの関係及び順序などは、一例であり、請求の範囲を限定する主旨ではない。

[0099] 以下では、符号化装置および復号化装置の実施の形態を説明する。実施の形態は、本開示の各態様で説明する処理および／または構成を適用可能な符号化装置および復号化装置の例である。処理および／または構成は、実施の形態とは異なる符号化装置および復号化装置においても実施可能である。例えば、実施の形態に対して適用される処理および／または構成に関して、例えば以下のいずれかを実施してもよい。

[0100] (1) 本開示の各態様で説明する実施の形態の符号化装置または復号装置の複数の構成要素のうちいずれかは、本開示の各態様のいずれかで説明する他の構成要素に置き換えまたは組み合わせられてもよい。

[0101] (2) 実施の形態の符号化装置または復号装置において、当該符号化装置または復号装置の複数の構成要素のうち一部の構成要素によって行われる機能または処理に、機能または処理の追加、置き換え、削除などの任意の変更がなされてもよい。例えば、いずれかの機能または処理は、本開示の各態様のいずれかで説明する他の機能または処理に、置き換えまたは組み合わせられてもよい。

[0102] (3) 実施の形態の符号化装置または復号装置が実施する方法において、当該方法に含まれる複数の処理のうち一部の処理について、追加、置き換えおよび削除などの任意の変更がなされてもよい。例えば、方法におけるい

ずれかの処理は、本開示の各態様のいずれかで説明する他の処理に、置き換えまたは組み合わせられてもよい。

[0103] (4) 実施の形態の符号化装置または復号装置を構成する複数の構成要素のうちの一部の構成要素は、本開示の各態様のいずれかで説明する構成要素と組み合わせられてもよいし、本開示の各態様のいずれかで説明する機能の一部を備える構成要素と組み合わせられてもよいし、本開示の各態様で説明する構成要素が実施する処理の一部を実施する構成要素と組み合わせられてもよい。

[0104] (5) 実施の形態の符号化装置または復号装置の機能の一部を備える構成要素、または、実施の形態の符号化装置または復号装置の処理の一部を実施する構成要素は、本開示の各態様いずれかで説明する構成要素と、本開示の各態様でいずれかで説明する機能の一部を備える構成要素と、または、本開示の各態様のいずれかで説明する処理の一部を実施する構成要素と組み合わせまたは置き換えられてもよい。

[0105] (6) 実施の形態の符号化装置または復号装置が実施する方法において、当該方法に含まれる複数の処理のいずれかは、本開示の各態様のいずれかで説明する処理に、または、同様のいずれかの処理に、置き換えまたは組み合わせられてもよい。

[0106] (7) 実施の形態の符号化装置または復号装置が実施する方法に含まれる複数の処理のうちの一部の処理は、本開示の各態様のいずれかで説明する処理と組み合わせられてもよい。

[0107] (8) 本開示の各態様で説明する処理および／または構成の実施の仕方は、実施の形態の符号化装置または復号装置に限定されるものではない。例えば、処理および／または構成は、実施の形態において開示する動画像符号化または動画像復号とは異なる目的で利用される装置において実施されてもよい。

[0108] [符号化復号システムの構成]

図1は、本実施の形態における符号化復号システムの構成例を示すブロッ

ク図である。符号化復号システムは、符号化装置１００及び復号装置２００を備える。本実施の形態において、符号化復号システムは、顔再現に用いられる。

[0109] 顔再現は、１人以上のソース人物の表情及びポーズを１人以上のターゲット人物の画像にマッピングすると同時に、画像においてターゲット人物のアイデンティティ及び属性が保持されることを保証する処理を指す。顔再現技術は、ビデオ会議からエンターテインメント分野まで、幅広い用途で利用されている。現在、動き表現の抽出又は潜在空間への画像特徴の投影等のような様々な手法の導入を通じてフォトリアリスティックな表現を強化する研究が数多く行われている。

[0110] 例えば、ビデオ会議アプリケーションは、符号化復号アーキテクチャで構成される。まず、ソース人物に対応するユーザの１つ以上のフレームで構成されるドライビング動画像が、符号化装置１００でキャプチャされる。その後、ドライビング動画像は、リアルタイム通信に基づいて、復号装置２００に送信される。復号装置２００は、ターゲット人物の顔動画像を再構成し表示する。

[0111] ターゲット人物の顔画像が符号化装置１００から復号装置２００に送信されてもよい。顔画像は、カメラからライブで提供されてもよいし、１つ以上の事前設定アニメ化アバター、又は、顔を含む１つ以上の事前設定ソース画像によって表現されてもよい。また、顔画像は、ユーザによって選択されてもよい。

[0112] 顔再現技術は、広告の制作、映画シーンの編集、及び、ミュージックビデオの拡張などのように、エンターテインメント業界で広く採用されている。このような用途では、ターゲットのアイデンティティ及び外見が保たれることを保証しつつ、１つ以上のドライビング動画像における感情、表情及びポーズが、ターゲットの顔に転送される。

[0113] ドライビング動画像における人物とターゲット人物とは、同じ人物であっても別の人物であってもよい。ターゲット人物は、現実の人物に限られず、

アバターなどの仮想的な人物であってもよい。

[0114] 様々なソーシャルメディアアプリケーションが普及し、その利用が増加する中、顔再現技術は、感情及び個性を象徴するような表現の生成において、独自にカスタマイズされ、柔軟性、利便性及び容易性を提供する。ドライビング動画像が、リアルタイムであったり、事前に録画されていたりするシナリオを考慮して、様々な顔再現技術が提案されている。また、出力動画像は、歪みなく自然に見えるように生成される。また、ユーザによって、出力映像の属性が調整可能であってもよい。

[0115] 図1の例において、符号化装置100は、ターゲット人物の顔画像と、複数のフレームを含むドライビング動画像とを入力として取得し、これらの情報を1つ以上のビットストリームに符号化し圧縮する。その後、通信ネットワーク又は記録媒体などの伝送チャネルを用いて、圧縮ビットストリームが復号装置200に伝送される。最後に、復号装置200は、受信ビットストリームから顔動画像を再構成する。

[0116] 例えば、顔画像は、顔を含む画像であり、基準画像又はアイデンティティ画像とも表現され得る。顔画像は、顔動画像を再構成するための静的で視覚的な特徴を表す。ドライビング動画像は、顔を含む動画像であって、カメラによって得られる撮影動画像である。ドライビング動画像は、顔画像に動きを与える役割を果たす。ビットストリームは、単にストリームとも表現される。また、1つのビットストリームが用いられることに限られず、複数のビットストリームが用いられてもよい。

[0117] 顔画像に含まれる人と、ドライビング動画像に含まれる人とは、同一人物であってもよいし、同一人物でなくてもよい。

[0118] なお、本実施の形態における符号化復号システムは、ビデオ会議、エンターテインメント業界におけるビデオの生成及び編集、ソーシャルメディア、及び、eコマース業界などに適用され得る。しかしながら、適用範囲は、これらに限定されない。

[0119] [データ構造]

図2は、ストリームにおけるデータの階層構造の一例を示す図である。ストリームは、例えばビデオシーケンスを含む。このビデオシーケンスは、例えば図2の(a)に示すように、VPS (V i d e o P a r a m e t e r S e t) と、SPS (S e q u e n c e P a r a m e t e r S e t) と、PPS (P i c t u r e P a r a m e t e r S e t) と、SEI (S u p p l e m e n t a l E n h a n c e m e n t I n f o r m a t i o n) と、複数のピクチャとを含む。

[0120] VPSは、複数のレイヤから構成されている動画像において、複数のレイヤに共通する符号化パラメータと、動画像に含まれる複数のレイヤ、または個々のレイヤに関連する符号化パラメータとを含む。

[0121] SPSは、シーケンスに対して用いられるパラメータ、すなわち、シーケンスを復号するために復号装置200が参照する符号化パラメータを含む。例えば、その符号化パラメータは、ピクチャの幅または高さを示してもよい。なお、SPSは複数存在してもよい。

[0122] PPSは、ピクチャに対して用いられるパラメータ、すなわち、シーケンス内の各ピクチャを復号するために復号装置200が参照する符号化パラメータを含む。例えば、その符号化パラメータは、ピクチャの復号に用いられる量子化幅の基準値と、重み付き予測の適用を示すフラグとを含んでもよい。なお、PPSは複数存在してもよい。また、SPSとPPSとは、単にパラメータセットと呼ばれる場合がある。

[0123] ピクチャは、図2の(b)に示すように、ピクチャヘッダと、1つ以上のスライスを含んでもよい。ピクチャヘッダは、その1つ以上のスライスを復号するために復号装置200が参照する符号化パラメータを含む。

[0124] スライスは、図2の(c)に示すように、スライスヘッダと、1つ以上のブリックとを含む。スライスヘッダは、その1つ以上のブリックを復号するために復号装置200が参照する符号化パラメータを含む。

[0125] ブリックは、図2の(d)に示すように、1つ以上のCTU (C o d i n g T r e e U n i t) を含む。

- [0126] なお、ピクチャは、スライスを含まず、そのスライスの代わりに、タイルグループを含んでいてもよい。この場合、タイルグループは、1つ以上のタイルを含む。また、ブリックにスライスが含まれていてもよい。
- [0127] CTUは、スーパーブロックまたは基本分割単位とも呼ばれる。このようなCTUは、図2の(e)に示すように、CTUヘッダと、1つ以上のCU (Coding Unit) とを含む。CTUヘッダは、1つ以上のCUを復号するために復号装置200が参照する符号化パラメータを含む。
- [0128] CUは、複数の小さいCUに分割されてもよい。また、CUは、図2の(f)に示すように、CUヘッダと、予測情報と、残差係数情報とを含む。予測情報は、そのCUを予測するための情報であって、残差係数情報は、後述の予測残差を示す情報である。なお、CUは、基本的にPU (Prediction Unit) およびTU (Transform Unit) と同一であるが、例えば後述のSBTでは、そのCUよりも小さい複数のTUを含んでいてもよい。また、CUは、そのCUを構成するVPDU (Virtual Pipeline Decoding Unit) ごとに処理されてもよい。VPDUは、例えば、ハードウェアにおいてパイプライン処理を行う際に、1ステージで処理できる固定的な単位である。
- [0129] なお、ストリームは、図2に示す各階層のうちの何れか一部の階層を有していなくてもよい。また、これらの階層の順番は、入れ替えられてもよく、何れかの階層は他の階層に置き換えられてもよい。また、符号化装置100または復号装置200などの装置によって現時点で行われる処理の対象とされているピクチャを、カレントピクチャという。その処理が符号化であれば、カレントピクチャは、符号化対象ピクチャと同義であり、その処理が復号であれば、カレントピクチャは、復号対象ピクチャと同義である。また、符号化装置100または復号装置200などの装置によって現時点で行われる処理の対象とされている例えばCUまたはCUなどのブロックを、カレントブロックという。その処理が符号化であれば、カレントブロックは、符号化対象ブロックと同義であり、その処理が復号であれば、カレントブロックは

、復号対象ブロックと同義である。

[0130] ここで、符号化及び復号に用いられるパラメータが記述される領域は、ヘッダ領域と表現され得る。例えば、ヘッダ領域は、SEIを含む領域である。ヘッダ領域は、さらに、VPS、SPS、PPS、SEI、ピクチャヘッダ、スライスヘッダ、CTUヘッダ、及び、CUヘッダを含み得る。

[0131] また、例えば、ピクチャは、Iピクチャ、Pピクチャ及びBピクチャを含む複数の種類のいずれかに分類され得る。Iピクチャは、イントラ予測ピクチャであり、イントラピクチャとも呼ばれ、他のピクチャを参照せずに符号化及び復号が行われるピクチャである。Pピクチャは、単予測ピクチャであり、他の1つのピクチャを参照して符号化及び復号が行われ得るピクチャである。Bピクチャは、双予測ピクチャであり、他の2つのピクチャを参照して符号化及び復号が行われ得るピクチャである。

[0132] また、動画画は、複数のGOP (Group of Pictures) で構成され得る。GOPは、ピクチャ集合を意味する。GOPは、1つ以上のIピクチャを含む。GOPは、1つ以上のPピクチャを含んでいてもよいし、1つ以上のBピクチャを含んでいてもよい。GOPは、映像の編集及びランダムアクセス等が可能な単位であってもよい。GOPは、一定の数のピクチャを有していてもよいし、Iピクチャ、Pピクチャ及びBピクチャに関する一定の配置順をGOP構造として有していてもよい。

[0133] [符号化及び復号の構成及び処理]

図3は、本実施の形態における符号化装置100の構成例を示すブロック図である。この例において、符号化装置100は、顔画像、ドライビング動画画像及び背景画像から、圧縮顔画像、圧縮ドライビング動画画像及び圧縮背景画像を含むビットストリームを生成する。符号化装置100は、圧縮器131、導出器132、圧縮器133及び圧縮器134を備える。各構成要素は、例えば、情報処理を行う電気回路である。圧縮器131、圧縮器133及び圧縮器134の2つ以上が統合されてもよい。

[0134] ここで、顔画像は、ターゲット人物の顔を含む。ターゲット人物の顔を含

む複数の顔画像のそれぞれが顔画像として用いられてもよい。また、ドライビング動画像の各ドライビングフレームは、ソース人物の顔を含む。また、複数の背景画像のそれぞれが背景画像として用いられてもよい。

[0135] まず、圧縮器 131 は、顔画像をビットストリームに符号化することにより圧縮する。また、導出器 132 は、カメラから直接キャプチャされたドライビング動画像の各ドライビングフレームから、ソース人物の顔を含む領域の幾何学属性を示す幾何学情報を導出する。圧縮器 133 は、幾何学情報をビットストリームに符号化することにより圧縮する。また、圧縮器 134 は、背景画像をビットストリームに符号化することにより圧縮する。

[0136] ここで、幾何学属性は、例えば動的属性に対応し、顔のランドマークのような点群で表現されてもよいし、複数のポリゴンの組み合わせによって物体の形状を表現するためのポリゴンモデルで表現されてもよい。また、幾何学属性は、その他の幾何学的モデルで表現されてもよい。また、幾何学属性は、顔のパーツの位置で表現されてもよい。

[0137] 例えば、フレーム毎（時間インスタンス毎）に、幾何学情報及び背景画像の両方が、符号化装置 100 から伝送される。また、例えば、顔画像は、最初のフレーム（最初の時間インスタンス）において伝送される。顔画像は、フレーム毎に伝送されなくてもよい。特に、顔画像は、前のフレームにおける顔画像と同じであれば、伝送されなくてもよい。

[0138] 変形例において、導出器 132 は、フレーム毎に、ドライビングフレームにおける前景領域と背景領域とを示すセグメンテーションマスクを導出し、圧縮器 133 は、セグメンテーションマスクをビットストリームに符号化することにより圧縮してもよい。そして、セグメンテーションマスクが復号装置 200 に伝送されてもよい。ドライビングフレームのセグメンテーションマスクは、再現対象の顔動画像におけるフレームのセグメンテーションマスクに対応する。

[0139] なお、背景画像は、必須の要素ではない。したがって、符号化装置 100 は、圧縮器 134 を備えなくてもよいし、ビットストリームに圧縮背景画像

が含まれなくてもよい。

- [0140] 図4は、本実施の形態における復号装置200の構成例を示すブロック図である。復号装置200は、ビットストリームから顔動画像を生成する。この例において、復号装置200は、圧縮解除器231、導出器232、圧縮解除器233、導出器237、生成器234及び圧縮解除器235を備える。各構成要素は、例えば、情報処理を行う電気回路である。圧縮解除器231、圧縮解除器233及び圧縮解除器235の2つ以上が統合されてもよい。
- [0141] 圧縮解除器231は、顔画像をビットストリームから復号することにより圧縮解除する。そして、導出器232は、顔画像から顔属性を示す顔情報を導出する。ここで、顔属性は、顔の静的で視覚的な属性である。顔属性は、アイデンティティとも表現され得る。
- [0142] 圧縮解除器233は、フレーム毎に幾何学情報をビットストリームから復号することにより圧縮解除する。そして、導出器237は、フレーム毎に幾何学情報からセグメンテーションマスクを導出する。セグメンテーションマスクは、顔動画像における前景領域及び背景領域を示す。
- [0143] 圧縮解除器235は、顔動画像に組み込まれる背景画像をビットストリームから復号する。
- [0144] 生成器234は、顔情報、幾何学情報、セグメンテーションマスク及び背景画像を入力として取り込み、複数のフレームを含む顔動画像を生成する。具体的には、生成器234は、ニューラルネットワーク等の生成モデルを用いて、顔情報、幾何学情報、セグメンテーションマスク及び背景画像から、顔動画像を生成する。これにより、符号量を抑制しつつ、顔動画像を再現することが可能になる。
- [0145] 例えば、フレーム毎に幾何学情報及び背景画像の両方が、復号装置200に伝送される。また、例えば、顔画像は、最初のフレームにおいて伝送される。顔画像は、フレーム毎に伝送されなくてもよい。特に、顔画像が伝送されない場合、前のフレームにおいて復号された顔画像が用いられてもよい。

- [0146] なお、背景画像は、必須の要素ではない。したがって、復号装置200は、圧縮解除器235を備えていなくてもよい。
- [0147] また、圧縮解除器233が、セグメンテーションマスクを復号し、復号されたセグメンテーションマスクが生成器234において用いられてもよい。また、セグメンテーションマスクは、必須の要素ではない。したがって、復号装置200は、導出器237を備えなくてもよい。また、顔情報の導出は省略されてもよい。そして、顔動画像の生成に顔情報の代わりに顔画像が用いられてもよい。したがって、復号装置200は、導出器232を備えなくてもよい。
- [0148] 例えば、生成器234は、顔画像及び幾何学情報のみに基づいて、顔動画像を生成してもよい。
- [0149] 図5は、顔画像の例を示す概念図である。この例のように、顔画像は、顔を含む画像である。
- [0150] 図6は、幾何学属性の例を示す概念図である。この例では、幾何学属性は、顔のランドマークである。例えば、ドライビング動画像のフレーム毎に幾何学属性を示す幾何学情報が導出される。
- [0151] 図7は、顔動画像の例を示す概念図である。この例のように、顔動画像は、顔を含む動画像である。顔動画像では、複数のフレームのそれぞれについて、当該フレームの幾何学属性が顔画像に反映される。これにより、顔画像に動きが与えられる。
- [0152] 上記の通り、顔画像は、最初のフレームにおいて伝送され、その後、各フレームで伝送されなくてもよい。これにより、ビット量を削減することが可能である。しかしながら、顔再現において、フレーム間のビット量の大きな違いによって、バッファサイズの大きな変動が引き起こされる場合がある。ここで、バッファサイズは、ストリームバッファのバッファサイズであって、ストリームバッファにおけるデータのサイズを意味する。
- [0153] 図8は、各フレームに対応するデータの構成例を示す概念図である。例えば、POC#0は、POC (Picture Order Count) が

0であることを意味する。この例において、符号化順と表示順とが一致しており、POC#0、POC#1、POC#2及びPOC#3の順でフレームが符号化され復号される。POC#0は、シーケンスつまりビットストリームにおいて最初のフレームであってもよいし、GOPにおいて最初のフレームであってもよい。

[0154] また、この例では、最初のフレームにおいて顔画像のデータが符号化される。また、各フレームにおいて、幾何学情報が符号化される。幾何学情報のデータのサイズは、顔画像のデータのサイズよりも小さい。したがって、顔画像が伝送される最初のフレームでは、フレームに対応するデータのサイズが大きく、顔画像が伝送されないその後の各フレームでは、フレームに対応するデータのサイズが小さい。これによって、バッファサイズの大きな変動が引き起こされ、低遅延伝送が困難になる場合がある。

[0155] 例えば、CPB (Coded Picture Buffer) のバッファサイズを小さくすることで、早期に復号が開始され、低遅延伝送が実現される。しかし、最初のフレームのデータサイズが大きい場合、バッファサイズを小さくすることは難しい。

[0156] 図9は、CPBバッファにおけるデータのサイズを示す概念図である。具体的には、CPBにおいて発生するバッファアンダーフローの例が示されている。バッファアンダーフローは、バッファレベルが0以下になった場合に発生する。この場合、CPBバッファにフルフレームデータがないため、復号画像が生成されず、復号に遅延が生じる。

[0157] 具体的には、この例において、最初のフレームのデータサイズが大きいいため、最初のフレームを復号するタイミングにおいて、最初のフレームのデータが不足している。これによって、遅延が生じる。つまり、フレーム間のデータサイズの違いによって、遅延が生じる可能性がある。すなわち、他のフレームのデータサイズに比べて、大きいデータサイズを有するフレームが取得される際に、バッファアンダーフローが生じやすくなる。したがって、フレーム間のデータサイズの違いは小さい方がよい。

- [0158] 例えば、符号化装置１００は、復号装置２００のメモリにおいてビットストリームが記憶されるストリームバッファを制御するための少なくとも１つの制御パラメータを符号化してもよい。ストリームバッファは、例えば、ＣＰＢである。そして、復号装置２００は、少なくとも１つの制御パラメータを復号してもよい。そして、復号装置２００は、少なくとも１つの制御パラメータに従ってストリームバッファを制御してもよい。
- [0159] 少なくとも１つの制御パラメータは、ストリームバッファのバッファサイズ、及び、復号処理開始の初期遅延時間を指定する少なくとも１つのパラメータを含んでいてもよい。復号処理開始の初期遅延時間は、ストリームバッファからビットストリームのデータが最初に取得されるタイミングに対応する。
- [0160] 少なくとも１つの制御パラメータは、ストリームバッファのバッファサイズが基準サイズ以下になり、復号処理開始の初期遅延時間が基準遅延時間以下になるように、設定されていてもよい。基準サイズは、標準的なサイズよりも小さい任意のサイズであってもよいし、基準遅延時間は、標準的遅延時間よりも短い任意の遅延時間であってもよい。
- [0161] また、少なくとも１つの制御パラメータは、バッファアンダーフローが生じない範囲で、ストリームバッファのバッファサイズが最も小さくなり、復号処理開始の初期遅延時間が最も短くなるように、設定されていてもよい。この場合、基準サイズは、バッファアンダーフローが生じない範囲で、最も小さいサイズとみなされてもよく、基準遅延時間は、バッファアンダーフローが生じない範囲で、最も短い遅延時間とみなされてもよい。
- [0162] ストリームバッファのバッファサイズを小さくすることで、ストリームバッファに滞留するデータのサイズが小さくなる。小さいバッファサイズでバッファアンダーフローを発生させることなくデータが伝送され処理されることにより、低遅延でデータが処理される。そのため、フレーム間で伝送データのサイズの増減は小さい方がよい。
- [0163] 図１０は、各フレームに対応するデータの別の構成例を示す概念図である

。この例では、顔画像のデータは、複数のデータに分離され、異なる複数のフレームにおいて伝送される。具体的には、顔画像のデータは、ベースデータと、1つ以上の拡張データとに分離される。ベースデータと、1つ以上の拡張データとは、顔画像のデータの異なる内容を含む。

[0164] ベースデータは、例えば、ベースピクチャとも表現されるピクチャに対応する。ベースデータから復号され出力されるピクチャは、参照テクスチャを提供し、当該参照テクスチャから顔動画像の複数のフレームに対応する複数の顔ピクチャが生成され得る。

[0165] 拡張データは、例えば、フュージョン（融合）に用いられるピクチャに対応する。拡張データは、ニューラルネットワーク等のような、顔動画像を生成するための生成モデルに入力されるピクチャであってもよい。すなわち、拡張データは、ベースデータで表現される顔画像を拡張するためのドライビングピクチャとして用いられてもよい。

[0166] フュージョン（融合）は、ベースデータと拡張データとの合算（足し合わせ）によって得られる再構成データが生成モデルに入力されることで行われてもよいし、ベースデータと拡張データとが生成モデルに入力されることで行われてもよい。

[0167] この例では、顔画像のデータが、ベースデータと、3つの拡張データとに分離される。POC # 3における最後の拡張データで顔画像のデータは完成する。POC # 3の後、顔画像のデータは伝送されない。

[0168] ここで、顔画像のデータは、顔画像データとも表現される。顔画像データは、顔画像全体のデータであってもよいし、ベースデータであってもよいし、1つ以上の拡張データのそれぞれであってもよい。また、顔画像データは、ベースデータと、1つ以上の拡張データの少なくとも1つとで再構成されるデータであってもよい。

[0169] 例えば、POC # 0に対応するベースデータ及び幾何学情報の復号後、POC # 0に対応するベースデータ及び幾何学情報から生成モデルを用いて顔動画像においてPOC # 0に対応するフレームが生成される。POC # 0に

対応するベースデータは、POC#0に対応する顔画像データともみなされる。

[0170] また、POC#1に対応する拡張データ及び幾何学情報の復号後、POC#1に対応する拡張データがベースデータに追加され、POC#1に対応する顔画像データが再構成される。POC#1に対応する顔画像データ及び幾何学情報は、顔動画像においてPOC#1に対応するフレームの生成に用いられる。

[0171] 顔動画像においてPOC#1に対応するフレームの生成では、生成モデルに、POC#1に対応する顔画像データのみでなく、POC#0に対応する顔画像データも入力されてもよい。あるいは、POC#0に対応するベースデータと、POC#1に対応する拡張データとが、生成モデルに入力されてもよい。

[0172] 同様に、POC#2に対応する拡張データ及び幾何学情報の復号後、POC#2に対応する拡張データが、POC#1に対応する顔画像データに追加され、POC#2に対応する顔画像データが再構成される。POC#2に対応する顔画像データ及び幾何学情報は、顔動画像においてPOC#2に対応するフレームの生成に用いられる。

[0173] 顔動画像においてPOC#2に対応するフレームの生成では、生成モデルに、POC#2に対応する顔画像データのみでなく、POC#0に対応する顔画像データも入力されてもよいし、POC#1に対応する顔画像データも入力されてもよい。あるいは、POC#0に対応するベースデータと、POC#1に対応する拡張データと、POC#2に対応する拡張データとが、生成モデルに入力されてもよい。

[0174] この例では、1つのフレームに対してデータサイズが大きくなることが抑制される。これにより、ストリームバッファを小さくすることが可能になり、遅延を抑制することが可能になる。

[0175] なお、図10の例のように、ベースデータ及び1つ以上の拡張データは、時間的に連続するPOCにおいて伝送されてもよい。これにより、最大CP

Bバッファサイズを削減しつつ、全ての顔画像データを迅速に揃えることが可能になる場合がある。

[0176] 拡張データが少なすぎる、つまり拡張データ数が小さすぎる場合、最大C P Bバッファサイズが十分に削減されない。一方、拡張データが多すぎる、つまり拡張データ数が大きすぎる場合、全ての顔画像データが揃うまでに時間がかかる。拡張データ数は、3であってもよい。拡張データ数は、遅延及び画質等に基づいて、ユーザによって決定されてもよい。

[0177] 図11は、各フレームに対応するデータのさらに別の構成例を示す概念図である。この例において、顔画像データは、ベースデータと、2つの拡張データとに分離される。POC#4における最後の拡張データで顔画像データは完成する。

[0178] 1つ以上の拡張データは、遅延及び画質等に基づいて、任意の順序及び任意のフレームで伝送されてもよい。

[0179] なお、図11の例のように、ベースデータ及び1つ以上の拡張データは、時間的に連続しないPOCにおいて伝送されてもよい。これにより、最大C P Bバッファサイズを削減しつつ、拡張データ数を小さくして処理量を削減することが可能になる場合がある。

[0180] また、全てのフレームのそれぞれについて、幾何学情報が伝送されなくてもよい。例えば、ベースデータが伝送されるフレームでは、幾何学情報が伝送されなくてもよい。また、拡張データが伝送されるフレームでは、幾何学情報が伝送されなくてもよい。ただし、全てのフレームのそれぞれについて、幾何学情報が伝送されることで、滑らかな動きを有する顔動画像が生成される可能性がある。

[0181] また、顔画像データの制御に関する制御情報が、SEI等であるヘッダ領域において伝送されてもよい。例えば、制御情報は、アクセスユニットに顔画像データが含まれているか否かを示す情報である存在情報を含んでいてもよい。このような存在情報は、アクセスユニットのヘッダ領域に含まれていてもよい。なお、アクセスユニットは、データ単位であり、1つのアクセス

ユニットは、1つのPOC、つまりPOCの1つの番号に対応する。

[0182] また、制御情報は、アクセスユニットにベースデータが含まれているか拡張データが含まれているかを示す情報である種別情報を含んでいてもよい。このような種別情報は、顔画像データを含むアクセスユニットのヘッダ領域に含まれていてもよい。

[0183] また、制御情報は、各拡張データを識別するための識別情報（例えばカウント情報）を含んでいてもよい。このような識別情報は、拡張データを含むアクセスユニットのヘッダ領域に含まれていてもよい。また、制御情報は、1つ以上の拡張データの総数を示す総数情報を含んでいてもよい。このような総数情報は、ベースデータを含むアクセスユニットのヘッダ領域に含まれていてもよい。

[0184] 図12は、ベースデータ及び拡張データの構成例を示す概念図である。この例では、顔画像に含まれる複数の要素が、ベースデータと拡張データとに分離される。すなわち、ベースデータは、顔の1つ以上の特徴（例えば顔の形、目及び口）を有する。拡張データは、顔の1つ以上の他の特徴（例えば眉、耳及び鼻）を有する。顔画像を再構成するため、ベースデータと拡張データとが組み合わされる。

[0185] 例えば、ビデオコーデックを用いてベースデータを符号化する前に、セグメンテーションマスクを用いて顔画像から顔の特徴（例えば眉、耳及び鼻）を除去することで、ベースデータが生成されてもよい。また、拡張データを符号化する前に、同じセグメンテーションマスクを用いて、ベースデータにおいて既に符号化された特徴を顔画像から除去することで、拡張データが生成されてもよい。

[0186] また、ベースデータの符号化には、イントラ予測が用いられてもよい。拡張データの符号化には、イントラ予測が用いられてもよいし、インター予測が用いられてもよい。例えば、拡張データは、ベースデータにおいて既に符号化されている顔の特徴に追加される他の顔の特徴を、インター予測における予測誤差データとして符号化するデータであってもよい。

- [0187] この例では、顔の複数の特徴を徐々に伝送することで、バッファアンダーフロー及び遅延の発生が抑制され得る。また、顔の部分的な特徴によって、表情が理解される可能性がある。
- [0188] なお、ベースデータに含まれる顔の特徴は、表情の印象に大きく影響し得る部分であってもよい。これにより、表情が早期に理解される可能性がある。
- [0189] また、この例において、顔動画像のフレームは、全ての顔画像データが取得された場合のみ、生成モデルによって生成されてもよい。例えば、全ての顔画像データが取得されるまで、固定のフレーム等のような、顔が含まれていないフレームが顔動画像のフレームに適用されてもよい。これにより、特徴が欠落している顔の表示が抑制され得る。
- [0190] 図 1 3 は、ベースデータ及び拡張データの別の構成例を示す概念図である。この例では、ベースデータは、低解像度の画像に対応する。拡張データは、高解像度の画像に対応し、より具体的には、低解像度の画像から高解像度の画像を再構成するための予測誤差データである。例えば、超解像技術を用いてベースデータに対応する画像の解像度を拡張することで得られる画像に拡張データを追加することによって、高精度で高解像度の顔画像が再構成され得る。
- [0191] 具体的には、例えば、ベースデータは、低解像度（例えば、フル画像の解像度の半分又は4分の1）で符号化される顔画像に対応していてもよい。拡張データは、ベースデータを参照して、フル解像度で符号化される顔画像に対応していてもよい。ここで、ベースデータの符号化には、イントラ予測が用いられてもよい。拡張データの符号化には、インター予測が用いられてもよい。
- [0192] より具体的には、拡張データの画像をインター予測で符号化するため、ベースデータの画像が参照ピクチャとして用いられてもよい。その際、R P R（Reference Picture Resampling）技術が用いられてもよい。R P R技術では、スケーリング比を用いて動き補償処理が

行われる。この場合、スケーリング比は、ベースデータの解像度と拡張データの解像度との差によって決定されてもよい。

[0193] また、複数の拡張データが、互いに異なる複数の解像度に対応していてもよい。そして、最も低い解像度から最も高い解像度へ、複数の拡張データが順に符号化されてもよい。

[0194] 図 1 4 は、ベースデータ及び拡張データのさらに別の構成例を示す概念図である。ベースデータは、顔画像の低周波数成分で構成される画像に対応する。拡張データは、顔画像の高周波数成分で構成される画像に対応する。ベースデータは、顔画像において第 1 周波数範囲の周波数成分で構成される画像に対応し、拡張データは、顔画像において第 1 周波数範囲よりも高い第 2 周波数範囲の周波数成分で構成される画像に対応する。そして、ベースデータに拡張データが追加されることで顔画像が再構成される。

[0195] 例えば、ベースデータの符号化において、顔画像からローパスフィルタ（又は類似の前処理技術）を介して低周波数成分の画像が取得される。そして、ビデオコーデックを用いて、低周波数成分の画像が、圧縮され、ベースデータとして符号化される。また、拡張データの符号化において、顔画像からハイパスフィルタを介して高周波数成分の画像が取得される。そして、ビデオコーデックを用いて、高周波数成分の画像が、圧縮され、拡張データとして符号化される。

[0196] また、ベースデータは、顔画像において閾値よりも低い周波数成分で構成される画像に対応していてもよい。拡張データは、顔画像において閾値よりも低くない周波数成分で構成される画像に対応していてもよい。

[0197] また、複数の拡張データが、互いに異なる複数の周波数範囲に対応していてもよい。そして、最も低い周波数範囲から最も高い周波数範囲へ、複数の拡張データが順に符号化されてもよい。

[0198] また、ベースデータの符号化には、イントラ予測が用いられてもよい。拡張データの符号化には、イントラ予測が用いられてもよいし、インター予測が用いられてもよい。

- [0199] この例では、顔動画像に含まれる顔が徐々にクリアになる可能性があり、顔動画像を見るユーザにとって違和感が軽減される可能性がある。
- [0200] なお、拡張データは、低周波数成分と高周波数成分との両方を含む画像に対応するデータであってもよい。具体的には、拡張データは、低周波数成分と高周波数成分との両方を含む画像に対応するデータであって、低周波数成分の画像を参照して、インター予測によって符号化される画像に対応するデータであってもよい。
- [0201] 図15は、ベースデータ及び拡張データのさらに別の構成例を示す概念図である。ベースデータは、顔画像の一部分であって、顔画像に含まれる領域に対応する。拡張データは、顔画像の他の部分であって、顔画像に含まれる他の領域に対応する。ベースデータに拡張データが結合されることで、顔画像が再構成される。
- [0202] また、顔画像が正面を向いていない場合、ベースデータは、顔の撮影された部分のみを含むデータであってもよい。拡張データは、ベースデータとは異なる向きで撮影された顔の他の部分を含むデータであってもよい。
- [0203] この例において、顔動画像のフレームは、全ての顔画像データが取得された場合のみ、ニューラルネットワークによって生成されてもよい。例えば、全ての顔画像データが取得されるまで、固定のフレーム等のような、顔が含まれていないフレームが顔動画像のフレームに適用されてもよい。これにより、部分的に領域が欠落している顔の表示が抑制され得る。
- [0204] また、ベースデータは、画像の左側のみがイントラ符号化されるGDR (Gradual Decoding Refresh) 技術を用いて符号化されてもよい。拡張データは、画像の右側のみがイントラ符号化されるGDRを用いて符号化されてもよい。そして、ベースデータ及び拡張データの復号後、完全な顔画像が再構成されてもよい。
- [0205] 図16は、ベースデータ及び拡張データのさらに別の構成例を示す概念図である。この例において、ベースデータは、大きなQP (量子化パラメータ: Quantization Parameter) 値及びイントラ予測で

符号化される顔画像のデータに対応する。拡張データは、小さなQ P値及びインター予測で符号化される顔画像のデータに対応する。

[0206] 具体的には、ベースデータの符号化に用いられるQ P値よりも、拡張データの符号化に用いられるQ P値は小さい。これにより、拡張データとして符号化される顔画像の精度が高くなる。一方、ベースデータを参照して、拡張データがインター予測で符号化される。これにより、拡張データの符号量が大きくなり過ぎることが抑制される。

[0207] また、拡張データは、ベースデータを参照してインター予測で復号される。これにより、ベースデータ及び拡張データが反映された顔画像が再構成される。

[0208] 図12～図16の複数の例において、拡張データの取得後、ベースデータと拡張データとによって再構成された顔画像データである再構成データが生成モデルに入力されてもよい。あるいは、ベースデータと拡張データとが生成モデルに入力されてもよい。あるいは、ベースデータと再構成データとが生成モデルに入力されてもよい。あるいは、複数の拡張データの取得後、ベースデータと複数の拡張データとが生成モデルに入力されてもよいし、ベースデータと複数の再構成データとが生成モデルに入力されてもよい。

[0209] 例えば、図13のように、拡張データが予測誤差データである場合、ベースデータと拡張データとによって再構成顔画像に対応する再構成データが取得されてから、再構成データが生成モデルに入力されてもよい。その際に、ベースデータと再構成データとが生成モデルに入力されてもよい。また、例えば、図12又は図15のように、ベースデータ及び拡張データのそれぞれによって画像が表現される場合、ベースデータと拡張データとが生成モデルに入力されてもよい。

[0210] 図17は、各フレームに対応するデータのさらに別の構成例を示す概念図である。この例では、POC#0のフレームに対応するデータに、SEIメッセージが付加される。このSEIメッセージは、顔画像データを完成させるPOC番号を示すパラメータを含む。

- [0211] 例えば、このパラメータを含むSEIメッセージは、ビットストリームにおいて最初のアクセスユニットで伝送されてもよいし、GOPにおいて最初のアクセスユニットで伝送されてもよいし、ベースデータを含むアクセスユニットで伝送されてもよい。
- [0212] また、SEIメッセージは、顔画像データが未完成でも顔動画像のフレームの再構成及び表示に用いられ得るかを示すパラメータを含んでいてもよい。あるいは、SEIメッセージは、顔画像データが未完成でも用いられるフレームを示すパラメータを含んでいてもよい。そして、パラメータが示すフレームにおいてのみ、当該フレームに付加された顔画像データからニューラルネットワークを用いて顔動画像のフレームが生成されてもよい。
- [0213] また、顔画像データが完成するフレームを示すPOCの代わりに、顔画像データが完成するフレームを示す画像識別子が用いられてもよい。
- [0214] また、SEIメッセージは、シンタックス構造を有しているため、SEIメッセージにおけるパラメータは、SEIメッセージのシンタックス構造におけるシンタックス要素と表現され得る。また、SEIメッセージではなく、他のヘッダ領域に、上述されたパラメータが示す内容と同じ内容を示すパラメータが含まれてもよい。
- [0215] 図18は、各フレームに対応するデータのさらに別の構成例を示す概念図である。この例では、各フレームに対応するデータに、SEIメッセージが付加される。つまり、SEIメッセージが各アクセスユニットに付加される。SEIメッセージは、SEIメッセージに対応するアクセスユニットに顔画像データが含まれるか否かを示すパラメータを含んでいてもよい。あるいは、顔画像データが含まれるアクセスユニットのみにSEIメッセージが付加されてもよい。
- [0216] アクセスユニットに顔画像データが含まれる場合、アクセスユニットに付加されたSEIメッセージには、アクセスユニットに含まれる顔画像データが、ベースデータであるか拡張データであるかを示すパラメータが含まれてもよい。あるいは、対象のアクセスユニットに、ベースデータが含まれてい

るかを示すパラメータと、拡張データが含まれているかを示すパラメータとが、それぞれ含まれていてもよい。

[0217] アクセスユニットに含まれる顔画像データが拡張データである場合、アクセスユニットに付加されたS E Iメッセージには、拡張データのカウンタ（序数）を示すパラメータが含まれてもよい。また、顔画像データがベースデータである場合、アクセスユニットに付加されたS E Iメッセージには、ベースデータに対応付けられる拡張データの総数を示すパラメータが含まれてもよい。

[0218] また、アクセスユニットに顔画像データが含まれる場合、アクセスユニットに付加されたS E Iメッセージには、アクセスユニットの顔画像データが顔動画像の対応フレームの再構成及び表示に用いられ得るかを示すパラメータが含まれていてもよい。顔画像データが顔動画像の対応フレームの再構成及び表示に用いられ得るとパラメータが示す場合のみ、アクセスユニットの顔画像データからニューラルネットワークを用いて顔動画像の対応フレームが生成されてもよい。

[0219] また、アクセスユニットに付加されたS E Iメッセージ、又は、その他のヘッダ領域に、それぞれが顔画像データを含む複数のアクセスユニットの複数の識別子を示す1つ以上のパラメータが含まれていてもよい。

[0220] 図19は、顔動画像の生成の制御例を示す概念図である。例えば、ベースデータ及び拡張データは、図12及び図15の例のように、顔画像において互いに異なる複数の部分に対応する。このような場合、ベースデータ又は拡張データに対応する不完全で部分的な画像が、表示されるよりも表示されない方がよい可能性がある。

[0221] そこで、P O C # 0では、ベースデータが取得されるが、人物の顔なしで顔動画像のフレームが生成される。また、P O C # 1では、1つ目の拡張データが取得されるが、人物の顔なしで顔動画像のフレームが生成される。

[0222] P O C # 2では、2つ目の拡張データが取得される。ベースデータ及び2つの拡張データによって、顔画像がほぼ完成され得る。したがって、P O C

2では、さらに幾何学情報が取得され、ベースデータ、2つの拡張データ、及び、幾何学情報から、ニューラルネットワークを用いて、人物の顔付きで顔動画像のフレームが生成される。

[0223] POC # 3では、3つ目の拡張データが取得される。ベースデータ及び3つの拡張データによって、顔画像が完成され得る。したがって、POC # 3では、さらに幾何学情報が取得され、ベースデータ、3つの拡張データ、及び、幾何学情報から、ニューラルネットワークを用いて、人物の顔付きで顔動画像のフレームが生成される。

[0224] また、POC # 3では、POC # 2よりも多くの拡張データによって、顔動画像のフレームが生成される。したがって、POC # 3で生成されるフレームに含まれる顔の完成度は、POC # 2で生成されるフレームに含まれる顔の完成度よりも高い。

[0225] 例えば、SEI等であるヘッダ領域において伝送される制御情報が、アクセスユニットに対応するフレームの生成及び表示に、当該アクセスユニットに含まれる顔画像データが適用可能であるか否かを示す適用情報を含んでいてもよい。また、制御情報が、ベースデータを含むアクセスユニットのヘッダ領域に含まれる情報であり、フレームに対応するデータに付加され、かつ、フレームの生成及び表示に適用可能な拡張データを特定するための情報である特定情報を含んでいてもよい。

[0226] 図10には、顔動画像の生成の別の制御例が示されている。例えば、ベースデータ及び拡張データは、図13、図14及び図16の例のように、顔画像の互いに異なる複数の品質に対応する。このような場合、ベースデータ又は拡張データに対応する不完全で低品質の画像が、表示されないよりも表示された方がよい可能性がある。

[0227] そこで、POC # 0で、ベースデータ及び幾何学情報が取得され、ベースデータ及び幾何学情報から、ニューラルネットワークを用いて、顔動画像のフレームが生成される。また、POC # 1では、1つ目の拡張データ及び幾何学情報が取得され、ベースデータ、拡張データ及び幾何学情報から、ニュ

ーラルネットワークを用いて、顔動画像のフレームが生成される。

[0228] 以降のフレームでも同様に、ニューラルネットワークを用いて顔動画像のフレームが生成される。また、顔動画像のフレームに用いられる拡張データの数が増加するにつれて、顔動画像のフレームの画質が向上する。

[0229] 図20は、本実施の形態における符号化装置100の別の構成例を示すブロック図である。符号化装置100は、顔画像及びドライビング動画像からビットストリームを生成する。この例において、符号化装置100は、前処理器141、圧縮器142、圧縮器143、導出器132及び圧縮器133を備える。各構成要素は、例えば、情報処理を行う電気回路である。圧縮器142、圧縮器143及び圧縮器133の2つ以上が統合されてもよい。

[0230] 図20の導出器132及び圧縮器133は、図3の導出器132及び圧縮器133に対応する。図20の圧縮器142及び圧縮器143は、図3の圧縮器131に対応する。

[0231] 前処理器141は、顔画像を取得し、顔画像をベースデータと1つ以上の拡張データとに分離する。圧縮器142は、ベースデータを取得し、ベースデータをビットストリーム符号化することにより圧縮する。圧縮器143は、1つ以上の拡張データを取得し、1つ以上の拡張データをビットストリームに符号化することにより圧縮する。

[0232] 導出器132は、ドライビング動画像を取得し、ドライビング動画像の各フレームから幾何学情報を導出する。圧縮器133は、幾何学情報を取得し、幾何学情報をビットストリームに符号化することにより圧縮する。

[0233] 図21は、本実施の形態における復号装置200の別の構成例を示すブロック図である。復号装置200は、ビットストリームから顔動画像を生成する。この例において、復号装置200は、圧縮解除器241、圧縮解除器242、圧縮解除器233及び生成器234を備える。各構成要素は、例えば、情報処理を行う電気回路である。圧縮解除器241、圧縮解除器242及び圧縮解除器233の2つ以上が統合されてもよい。

[0234] 図21の圧縮解除器233及び生成器234は、図4の圧縮解除器233

及び生成器 2 3 4 に対応する。図 2 1 の圧縮解除器 2 4 1 及び圧縮解除器 2 4 2 は、図 4 の圧縮解除器 2 3 1 に対応する。

[0235] 圧縮解除器 2 4 1 は、ベースデータをビットストリームから復号することにより圧縮解除する。圧縮解除器 2 4 2 は、1 つ以上の拡張データをビットストリームから復号することにより圧縮解除する。圧縮解除器 2 3 3 は、幾何学情報をビットストリームから復号することにより圧縮解除する。生成器 2 3 4 は、ベースデータ、1 つ以上の拡張データ、及び、幾何学情報から、生成モデルを用いて、顔動画像を生成し出力する。

[0236] 生成器 2 3 4 は、顔動画像の生成において、ベースデータ及び拡張データを結合して結合データを幾何学情報と共に生成モデルに入力することにより、生成モデルから出力される顔動画像を取得してもよい。

[0237] あるいは、生成器 2 3 4 は、顔動画像の生成において、ベースデータ及び拡張データを結合せずに、ベースデータ及び拡張データのそれぞれを幾何学情報と共に生成モデルに入力することにより、生成モデルから出力される顔動画像を取得してもよい。あるいは、拡張データはベースデータに対する予測残差データであり、生成器 2 3 4 は、インター予測を用いて復号された復号データを幾何学情報と共に生成モデルに入力することにより、生成モデルから出力される顔動画像を取得してもよい。

[0238] 本開示のユースケースの例は、ビデオ会議及び娯楽産業のための顔再現である。本開示のユースケースにおいて、ビットストリームは、狭帯域幅かつ低遅延の状況で伝送される。そして、本開示は、このような状況下で、ビットストリームにおける伝送量の変動を抑制し、バッファのデータ量の変動を抑制することに役立つ。すなわち、本開示では、バッファサイズを小さくすることが可能になり、低遅延の顔再現が可能になる。

[0239] 図 2 2 は、本実施の形態における符号化装置 1 0 0 の動作例を示すフローチャートである。例えば、符号化装置 1 0 0 は、ベースデータをビットストリーム符号化する (S 1 0 1)。ベースデータは、人物の顔の情報を含み、ビットストリームにおける最初のフレームに符号化される。

- [0240] また、符号化装置100は、1つ以上の拡張データをビットストリームに符号化する(S102)。1つ以上の拡張データは、同じ人物の顔の情報を含み、ビットストリームにおける1つ以上のフレームであって、最初のフレームとは異なる1つ以上のフレームに符号化される。
- [0241] さらに、符号化装置100は、ベースデータ及び1つ以上の拡張データから、生成モデルを用いて、顔動画像のフレームを生成してもよい(S103)。これにより、復号装置200によって生成されるフレームを符号化装置100によって確認することが可能になる。
- [0242] また、符号化装置100は、さらにフレーム毎に幾何学情報を符号化してもよい。そして、顔動画像のフレームの生成において、符号化装置100は、ベースデータ、1つ以上の拡張データ及び幾何学情報から、生成モデルを用いて、顔動画像のフレームを生成してもよい。
- [0243] 図23は、本実施の形態における復号装置200の動作例を示すフローチャートである。例えば、復号装置200は、ベースデータをビットストリームから復号する(S201)。ベースデータは、人物の顔の情報を含み、ビットストリームにおける最初のフレームから復号される。
- [0244] また、復号装置200、1つ以上の拡張データをビットストリームから復号する(S202)。1つ以上の拡張データは、同じ人物の顔の情報を含み、ビットストリームにおける1つ以上のフレームであって、最初のフレームとは異なる1つ以上のフレームから復号される。
- [0245] さらに、復号装置200は、ベースデータ及び1つ以上の拡張データから、生成モデルを用いて、顔画像のフレームを生成する(S203)。復号装置200は、さらに幾何学情報を復号し、ベースデータ、1つ以上の拡張データ及び幾何学情報から、生成モデルを用いて、顔動画像のフレームを生成してもよい。
- [0246] 例えば、復号装置200は、ベースデータ及び1つ以上の拡張データを用いて、同じ人物の顔の情報を含む顔画像データを再構成してもよい。そして、復号装置200は、再構成顔画像データを生成モデルに入力して、生成モ

デルから顔動画像のフレームを取得してもよい。

[0247] あるいは、復号装置 200 は、ベースデータ及び 1 つ以上の拡張データを生成モデルに入力して、生成モデルから顔動画像のフレームを取得してもよい。

[0248] 一例において、ベースデータ、1 つ以上の拡張データ、及び、再構成顔画像データは、顔の特徴を表す複数のベクトルであってもよい。別の例において、ベースデータ、1 つ以上の拡張データ、及び、再構成顔画像データは、複数の画像であってもよい。生成モデルは、ベースデータ、1 つ以上の拡張データ、再構成顔画像データ、又は、これらの組み合わせから、顔動画像のフレームを生成することに用いられ得る。

[0249] ニューラルネットワークが、生成モデルとして用いられてもよい。このようなニューラルネットワークの例は、生成ネットワークである。生成ネットワークの複数の例は、GAN (Generative Adversarial Network)、VAE (Variational Autoencoder)、自己回帰モデル、及び、拡散モデルを含む。

[0250] 生成ネットワークは、提供データセットに基づいて新しいデータを生成する機械学習フレームワークの例である。新しいデータの複数のサンプルが元のデータセットと類似していることを保証するために、生成モデルは、新しいデータの生成前に、データセットの基本的な分布を分析し学習する。

[0251] 図 24 は、本実施の形態における符号化装置 100 のさらに別の構成例を示すブロック図である。この例において、符号化装置 100 は、顔画像、ドライビング動画像及び背景画像から、圧縮顔画像、圧縮ドライビング動画像及び圧縮背景画像を含むビットストリームを生成する。符号化装置 100 は、圧縮器 131、導出器 132、圧縮器 133 及び圧縮器 134 を備える。各構成要素は、例えば、情報処理を行う電気回路である。圧縮器 131、圧縮器 133 及び圧縮器 134 の 2 つ以上が統合されてもよい。

[0252] 図 24 の圧縮器 131、導出器 132、圧縮器 133 及び圧縮器 134 は、図 3 の圧縮器 131、導出器 132、圧縮器 133 及び圧縮器 134 に対

応する。

[0253] 図25は、本実施の形態における符号化装置100の別の動作例を示すフローチャートである。例えば、図24に示された符号化装置100の複数の構成要素は、図25のフローチャートに従って動作を行う。

[0254] この例において、まず、圧縮器131は、顔画像をビットストリームに符号化することにより圧縮する(S301)。顔画像は、VVCのようなビデオコーデック方式に従って符号化されてもよい。顔画像は、ドライビング動画像のフレームでもよいし、人物の顔を含む予め用意された画像でもよいし、アバターでもよい。

[0255] また、導出器132は、ドライビング動画像から、ドライビング動画像の各フレームに対応する幾何学属性を示す幾何学情報を導出する(S302)。幾何学属性を示す幾何学情報は、単に幾何学属性とも表現される。具体的には、導出器132は、ドライビング動画像の各フレームをニューラルネットワーク等の認識モデルに入力し、認識モデルから各フレームに対応する幾何学情報を取得する。幾何学情報は、ドライビング動画像の各フレームの時間インスタンスに対応する。

[0256] ここで、幾何学属性は、例えば動的属性に対応し、顔のランドマークのような点群で表現されてもよいし、複数のポリゴンの組み合わせによって物体の形状を表現するためのポリゴンモデルで表現されてもよい。また、幾何学属性は、その他の幾何学モデルで表現されてもよい。また、幾何学属性は、顔のパーツの位置で表現されてもよい。また、幾何学属性は、幾何学属性セットとして扱われてもよい。

[0257] 例えば、幾何学属性として用いられる顔ランドマークは、顔の輪郭、目、眉、鼻、口、唇及び顎を含む顔の主要領域上の点の位置を示す。このような幾何学属性は他の人又は他の装置によって解釈可能であるため、属性の修正が可能であり、属性の処理の改善が可能である。

[0258] 圧縮器133は、エントロピー符号化等の方法を用いて幾何学情報をビットストリームに符号化することにより圧縮する(S303)。

- [0259] 圧縮器 134 は、少なくとも 1 つの背景画像をビットストリームに符号化することにより圧縮する (S304)。背景画像は、VVC のようなビデオコーデック方式に従って符号化されてもよい。背景画像は、合成顔動画像における背景領域に用いられる。つまり、背景画像は、顔を含む顔動画像にオーバーレイされる背景を示す。
- [0260] 符号化装置 100 は、復号装置 200 が行う動作と同様に、顔画像、幾何学情報及び背景画像に基づいて、合成顔動画像を生成してもよい (S305)。符号化装置 100 は、合成顔動画像を生成するため、復号装置 200 と同様の複数の構成要素を備えていてもよい。これにより、復号装置 200 で生成される合成顔動画像を符号化装置 100 で確認することが可能になる。なお、この処理は、省略されてもよい。
- [0261] 符号化装置 100 は、顔画像、幾何学情報及び背景画像をビットストリームに符号化した後、伝送チャネルを介してビットストリームを復号装置 200 に送信する。例えば、ドライビング動画像のフレーム毎、つまり、時間インスタンス毎に、符号化装置 100 から復号装置 200 へ圧縮幾何学情報がビットストリームとして伝送される。圧縮幾何学情報は、SEI (Supplemental Enhancement Information) として伝送されてもよい。
- [0262] なお、顔画像を構成するベースデータと 1 つ以上の拡張データとが顔画像として扱われてもよい。例えば、圧縮器 131 は、顔画像をベースデータと 1 つ以上の拡張データとに分離して、ベースデータと 1 つ以上の拡張データとをそれぞれ異なる複数のフレームに対応する複数のデータとしてビットストリームに符号化してもよい。
- [0263] また、背景画像は、必須の要素ではない。したがって、符号化装置 100 は、圧縮器 134 を備えなくてもよい。また、ビットストリームに圧縮背景画像が含まれなくてもよい。また、背景画像に関する入出力は、省略されてもよい。
- [0264] あるいは、顔画像と同様に、背景画像も背景ベースデータと 1 つ以上の背

景拡張データとに分離されてもよい。そして、背景画像を構成する背景ベースデータと1つ以上の背景拡張データとが背景画像として扱われてもよい。そして、圧縮器134は、背景画像を背景ベースデータと1つ以上の背景拡張データとに分離して、背景ベースデータと1つ以上の背景拡張データとをそれぞれ異なる複数のフレームに対応する複数のデータとしてビットストリームに符号化してもよい。

[0265] 図26は、本実施の形態における復号装置200のさらに別の構成例を示すブロック図である。この例において、復号装置200は、ビットストリームから合成顔動画像を生成する。復号装置200は、圧縮解除器231、導出器232、圧縮解除器233、生成器234、圧縮解除器235及び合成器236を備える。各構成要素は、例えば、情報処理を行う電気回路である。圧縮解除器231、圧縮解除器233及び圧縮解除器235の2つ以上が統合されてもよい。

[0266] 図26の圧縮解除器231、導出器232、圧縮解除器233、生成器234及び圧縮解除器235は、図4の圧縮解除器231、導出器232、圧縮解除器233、生成器234及び圧縮解除器235に対応する。

[0267] 図27は、本実施の形態における復号装置200の別の動作例を示すフローチャートである。例えば、図26に示された復号装置200の複数の構成要素は、図27のフローチャートに従って動作を行う。なお、以下では、符号化の説明と同じ説明が省略される場合がある。

[0268] 圧縮解除器231は、ビットストリームから顔画像を復号することにより圧縮解除する(S401)。顔画像は、VVCのようなビデオコーデック方式に従って復号されてもよい。その後、圧縮解除器231は、顔画像を導出器232に供給する。

[0269] 導出器232は、顔画像から顔属性を示す顔情報を導出する(S402)。ここで、顔属性を示す顔情報は、単に顔属性とも表現される。顔属性は、静的で視覚的な属性であって、アイデンティティとも表現され得る。顔属性は、髪、眼鏡、顔の毛、眉毛、目、口、鼻、肌、顔の輪郭、服装及びアクセ

サリの少なくとも1つに関する情報を含み得る。

[0270] 圧縮解除器233は、エントロピー復号等の方法を用いてビットストリームからフレーム毎に幾何学情報を復号することにより圧縮解除する(S403)。

[0271] 生成器234は、顔情報及び幾何学情報から、ニューラルネットワーク等の生成モデルを用いて、中間顔動画像を生成する(S404)。

[0272] 生成モデルは、GAN (Generative Adversarial Network)、VAE (Variational Autoencoder)、自己回帰モデル、又は、拡散モデル等でもよい。例えば、生成モデルは、提供データセットに基づいて新しいデータを生成する機械学習フレームワークであって、生成モデルにおいて、データセットの基本的な分布の分析及び学習が行われていてもよい。

[0273] 例えば、生成器234は、フレーム毎に、顔情報及び幾何学情報を生成モデルに入力して、生成モデルから中間顔動画像及びセグメンテーションマスクを取得する。より具体的には、生成器234は、フレーム毎に、顔情報及び幾何学情報を生成モデルに入力して、生成モデルから中間顔動画像のフレーム及び当該フレームのセグメンテーションマスクを取得する。このセグメンテーションマスクは、中間顔動画像(具体的には中間顔動画像のフレーム)における前景領域及び背景領域を示す。

[0274] セグメンテーションマスクは、前景領域の全ての画素値が1であり、背景領域の全ての画素値が0である2次元マップで表現されてもよいし、前景領域の全ての画素値が0であり、背景領域の全ての画素値が1である2次元マップで表現されてもよい。例えば、前景領域は、顔等を含む領域であって、動きを含む領域であり、背景領域は、顔等を含まない領域であって、動きを含まない領域である。セグメンテーションマスクは、セグメンテーション情報とも呼ばれる。

[0275] 生成器234は、顔情報に代えて、又は、顔情報に加えて、顔画像自体を用いて、中間顔動画像をレンダリングしてもよい。また、導出器232は、

生成器 2 3 4 に含まれていてもよいし、存在しなくてもよい。そして、導出器 2 3 2 において顔情報を導出するための認識モデルは、生成器 2 3 4 において中間顔動画像等を生成するための生成モデルに含まれていてもよい。導出器 2 3 2 及び顔情報に関して、他の変形例も同様である。

[0276] つまり、生成器 2 3 4 は、顔画像及び幾何学情報から生成モデルを用いてセグメンテーションマスク及び中間顔動画像を生成してもよい。その際、生成器 2 3 4 は、顔画像及び幾何学情報を生成モデルに入力して、生成モデルからセグメンテーションマスク及び中間顔動画像を取得してもよい。

[0277] 圧縮解除器 2 3 5 は、ビットストリームから少なくとも 1 つの背景画像を復号することにより圧縮解除する (S 4 0 5)。背景画像は、VVCのようなビデオコーデック方式に従って復号されてもよい。背景画像の代わりに、複数の背景画像候補から背景画像を選択するための選択パラメータが復号されてもよい。選択パラメータは、複数の背景画像候補のいずれかに対応する背景画像の識別子であってもよい。

[0278] 合成器 2 3 6 は、中間顔動画像、セグメンテーションマスク及び背景画像を用いて、中間顔動画像における背景領域に背景画像における対応領域を埋め込むことにより、合成顔動画像を生成する (S 4 0 6)。

[0279] なお、顔画像を構成するベースデータと 1 つ以上の拡張データとが顔画像として扱われてもよい。例えば、圧縮解除器 2 3 1 は、それぞれ異なる複数のフレームに対応する複数のデータとして符号化されたベースデータと 1 つ以上の拡張データとをビットストリームから復号してもよい。

[0280] また、背景画像は、必須の要素ではない。したがって、復号装置 2 0 0 は、圧縮解除器 2 3 5 及び合成器 2 3 6 を備えなくてもよい。また、ビットストリームに圧縮背景画像が含まれなくてもよい。また、背景画像に関する入出力は、省略されてもよい。

[0281] あるいは、顔画像と同様に、背景画像を構成する背景ベースデータと 1 つ以上の背景拡張データとが背景画像として扱われてもよい。そして、圧縮解除器 2 3 5 は、それぞれ異なる複数のフレームに対応する複数のデータとし

て符号化された背景ベースデータと1つ以上の背景拡張データとをビットストリームから復号してもよい。

[0282] 図28は、本実施の形態における復号装置200のさらに別の構成例を示すブロック図である。上述した例、つまり図26の例において、生成器234は、顔情報及び幾何学情報を生成モデルに入力して、生成モデルから中間顔動画像及びセグメンテーションマスクを取得する。

[0283] これに対して、この例、つまり図28の例において、生成器234は、顔情報及び幾何学情報を生成モデルに入力して、生成モデルから中間顔動画像を取得する。そして、生成器234は、中間顔動画像に対してセグメンテーション処理を行って、セグメンテーションマスクを取得する。

[0284] 具体的には、生成器234は、フレーム毎に、顔情報及び幾何学情報を生成モデルに入力して、生成モデルから中間顔動画像のフレームを取得する。そして、生成器234は、中間顔動画像の各フレームに対してセグメンテーション処理を行って、当該フレームのセグメンテーションマスクを取得する。

[0285] これにより、処理を細分化することが可能になり、処理を円滑に行うことが可能になる場合がある。生成器234の代わりに、セグメンテーション処理器（図示せず）がセグメンテーション処理を行ってもよい。

[0286] セグメンテーション処理は、ニューラルネットワーク等の機械学習モデルを用いて行われてもよい。本開示の他のセグメンテーション処理も同様である。

[0287] 復号装置200において生成される中間顔動画像及び合成顔動画像における前景領域及び背景領域は、ドライビング動画像における前景領域及び背景領域に対応する。そこで、符号化装置100が、ドライビング動画像に対してセグメンテーション処理を行い、ドライビング動画像のセグメンテーションマスクを符号化してもよい。そして、復号装置200が、セグメンテーションマスクを復号し、セグメンテーションマスクを用いて、合成顔動画像を生成してもよい。

- [0288] 具体的には、符号化装置100において、導出器132が、フレーム毎に、ドライビング動画像に対してセグメンテーション処理を行って、ドライビング動画像における前景領域及び背景領域を示すセグメンテーションマスクを生成してもよい。また、圧縮器133が、セグメンテーションマスクをビットストリームに符号化することにより圧縮してもよい。
- [0289] そして、復号装置200において、圧縮解除器233が、セグメンテーションマスクをビットストリームから復号することにより圧縮解除してもよい。さらに、合成器236が、セグメンテーションマスク等を用いて、合成顔動画像を生成してもよい。これにより、復号装置200における処理量が削減される可能性がある。
- [0290] また、符号化装置100において、導出器132とは異なるセグメンテーション処理器（図示せず）が、セグメンテーション処理を行ってもよい。また、圧縮器133とは異なる圧縮器（図示せず）が、セグメンテーションマスクをビットストリームに符号化してもよい。また、復号装置200において、圧縮解除器233とは異なる圧縮解除器（図示せず）がセグメンテーションマスクをビットストリームから復号してもよい。
- [0291] また、セグメンテーションマスクは、符号化装置100から復号装置200へフレーム毎にSEIにおいて伝送されてもよい。
- [0292] 顔画像における背景領域の各画素サンプルには、背景のための規定の背景カラーコードが割り当てられていてもよい。これにより、背景領域の各画素サンプルに規定の背景カラーコードが割り当てられた中間顔動画像が生成される。したがって、セグメンテーション処理が行われなくても、中間顔動画像における背景領域が効率的に特性することが可能になる。
- [0293] 図29は、本実施の形態における復号装置200のさらに別の構成例を示すブロック図である。この例では、生成器234は、顔情報、幾何学情報及び背景画像から生成モデルを用いて合成顔動画像を生成する。具体的には、生成器234は、顔情報、幾何学情報及び背景画像を生成モデルに入力して、生成モデルから合成顔動画像を取得することにより、合成顔動画像を生成

する。これにより、処理が簡素化され得る。この場合、復号装置 200 は、別途の合成器 236 を備えていなくてもよい。

[0294] 〔生成モデルの例〕

図 30 は、生成モデルとして用いられ得る様々なモデルの例を示す図である。例えば、ニューラルネットワークが生成モデルとして用いられる。具体的には、図 30 において、敵対的生成ネットワーク、変分オートエンコーダ、フローベース生成モデル、及び、拡散モデルが示されている。

[0295] 敵対的生成ネットワークでは、入力データの特徴の学習を介して入力データに類似する新たなデータインスタンスが生成される。具体的には、生成モデルにおける教師なしタスクが、2 種類のサブモデルによって教師ありタスクへ変換される。

[0296] 例えば、生成器サブモデルが、偽物のサンプルを生成し、識別器サブモデルが、本物の入力と、生成器サブモデルで生成された偽物のサンプルとを識別する。そして、本物の入力と偽物のサンプルとに対する正確なラベルの割り当てにおける識別器サブモデルの識別確率を最大にすると同時に、本物の入力と偽物のサンプルとの分布差を最小にするミニマックスゲームを介し、出力画像が生成される。

[0297] 変分オートエンコーダでは、まず、できるだけ正確に潜在空間からデータを再構成するための多変量潜在分布に入力データが圧縮される。これにより、データ圧縮及び次元削減が効率的に行われる。フローベース生成モデルでは、1 つ以上の可逆変換のシーケンスを介してソース分布が訓練データの分布に変換される。これにより、データ分布の学習及び最終目標の尤度の正確な算出が可能になる。

[0298] 拡散モデルでも、訓練データに類似する新たなデータインスタンスが生成される。まず、拡散モデルでは、オリジナルデータを復元する試みにおいてノイズ除去処理を始める前に、摂動及びノイズの反復注入を介して訓練データの構造を劣化させる。結果として、各ステップの潜在状態が前のステップの潜在状態にのみ依存するマルコフ連鎖を介して、データが潜在分布に繰り

返しマッピングされる。そして、階層的なノイズ除去によりデータが復元される。

[0299] 例えば、ニューラルネットワークは、顔パラメータのための固定フォーマットで表現される幾何学情報と、ピクチャとを用いて、出力ピクチャを生成するために用いられ得る顔ピクチャ生成ニューラルネットワークであってもよい。すなわち、ニューラルネットワークは、出力動画像に含まれる1つのピクチャである出力ピクチャを構成する複数のサンプル値を生成する処理に対応する。

[0300] 上述されたニューラルネットワークの代替例は、上述されたニューラルネットワークの任意の組み合わせであってもよい。また、他の種類の生成モデル等が用いられてもよい。

[0301] また、ニューラルネットワーク等の機械学習モデルは、セグメンテーション処理に用いられてもよい。また、ニューラルネットワーク等の機械学習モデルは、幾何学情報の導出に用いられてもよいし、顔情報の導出に用いられてもよい。

[0302] [動画像の符号化及び復号の構成例]

図31は、本実施の形態における符号化装置100が動画像を符号化するための構成例を示すブロック図である。例えば、符号化装置100は、動画像に含まれる画像をブロック単位でVVCに従って符号化するための複数の構成要素として、図31に示された複数の構成要素を含んでいてもよい。符号化装置100は、上述された複数の構成要素に加えて、図31に示された複数の構成要素を含んでいてもよいし、上述された複数の構成要素の少なくとも一部が、図31に示された複数の構成要素に統合されてもよい。

[0303] 図31に示すように、符号化装置100は、分割部102と、減算部104と、変換部106と、量子化部108と、エントロピー符号化部110と、逆量子化部112と、逆変換部114と、加算部116と、ブロックメモリ118と、ループフィルタ部120と、フレームメモリ122と、イントラ予測部124と、インター予測部126と、予測制御部128と、予測パ

ラメータ生成部 130 とを備える。なお、イントラ予測部 124 及びインター予測部 126 のそれぞれは、予測処理部の一部として構成されている。

[0304] 分割部 102 は、画像を複数のブロックに分割し、分割に関するパラメータをエントロピー符号化部 110 に提供する。減算部 104 は、カレントブロックから予測画像ブロックを減算して予測残差ブロックを取得する。変換部 106 は、予測残差ブロックに対して変換を行って変換係数ブロックを取得する。量子化部 108 は、変換係数ブロックに対して量子化を行って量子化係数ブロックを取得する。エントロピー符号化部 110 は、量子化係数ブロック及びパラメータに対してエントロピー符号化を行ってビットストリームを生成する。

[0305] 逆量子化部 112 は、量子化係数ブロックに対して逆量子化を行って変換係数ブロックを取得する。逆変換部 114 は、変換係数ブロックに対して逆変換を行って予測残差ブロックを取得する。加算部 116 は、予測残差ブロックに予測画像ブロックを加算して再構成画像ブロックを取得する。ブロックメモリ 118 には、再構成画像ブロックが記憶される。ループフィルタ部 120 は、再構成画像ブロックにループフィルタを適用する。フレームメモリ 122 には、ループフィルタが適用された再構成画像ブロックが記憶される。

[0306] イントラ予測部 124 は、ブロックメモリ 118 を参照してイントラ予測を行って予測画像ブロックを生成する。インター予測部 126 は、フレームメモリ 122 を参照してインター予測を行って予測画像ブロックを生成する。予測制御部 128 は、イントラ予測部 124 が生成する予測画像ブロック、又は、インター予測部 126 が生成する予測画像ブロックを減算部 104 及び加算部 116 に提供する。予測パラメータ生成部 130 は、イントラ予測又はインター予測に関するパラメータをエントロピー符号化部 110 に提供する。

[0307] 図 32 は、本実施の形態における復号装置 200 が動画像を復号するための構成例を示すブロック図である。例えば、復号装置 200 は、動画像に含

まれる画像をブロック単位でVVCに従って復号するための複数の構成要素として、図32に示された複数の構成要素を含んでもよい。復号装置200は、上述された複数の構成要素に加えて、図32に示された複数の構成要素を含んでもよいし、上述された複数の構成要素の少なくとも一部が、図32に示された複数の構成要素に統合されてもよい。

- [0308] 図32に示すように、復号装置200は、エントロピー復号部202と、逆量子化部204と、逆変換部206と、加算部208と、ブロックメモリ210と、ループフィルタ部212と、フレームメモリ214と、イントラ予測部216と、インター予測部218と、予測制御部220と、予測パラメータ生成部222と、分割決定部224とを備える。なお、イントラ予測部216及びインター予測部218のそれぞれは、予測処理部の一部として構成されている。
- [0309] エントロピー復号部202は、ビットストリームに対してエントロピー復号を行って量子化係数ブロック及びパラメータを取得する。逆量子化部204は、量子化係数ブロックに対して逆量子化を行って変換係数ブロックを取得する。逆変換部206は、変換係数ブロックに対して逆変換を行って予測残差ブロックを取得する。加算部208は、予測残差ブロックに予測画像ブロックを加算して再構成画像ブロックを取得する。ループフィルタ部212は、再構成画像ブロックにループフィルタを適用する。
- [0310] ブロックメモリ210には、再構成画像ブロックが記憶される。フレームメモリ214には、ループフィルタが適用された再構成画像ブロックが記憶される。
- [0311] イントラ予測部216は、ブロックメモリ210を参照してイントラ予測を行って予測画像ブロックを生成する。インター予測部218は、フレームメモリ214を参照してインター予測を行って予測画像ブロックを生成する。予測制御部220は、イントラ予測部216が生成する予測画像ブロック、又は、インター予測部218が生成する予測画像ブロックを加算部208に提供する。予測パラメータ生成部222は、イントラ予測又はインター予

測に関するパラメータを予測制御部 220 等に提供する。

[0312] 分割決定部 224 は、分割に関するパラメータに従って、画像をブロック単位で復号するためのブロックを決定する。

[0313] [組み合わせ]

本開示における複数の構成例が任意に組み合わせられてもよい。また、本開示における複数の動作例が任意に組み合わせられてもよい。また、本開示の複数の例において重複する説明が省略されている場合がある。また、符号化の構成及び処理に対応する構成及び処理が復号に適用されてもよいし、復号の構成及び処理に対応する構成及び処理が符号化に適用されてもよい。また、本開示の複数の例に含まれる例の一部のみが実施されていてもよい。

[0314] [実施例]

図 33 は、符号化装置 100 の実装例を示すブロック図である。符号化装置 100 は、回路 151 及びメモリ 152 を備える。例えば、上述された符号化装置 100 の複数の構成要素は、回路 151 及びメモリ 152 によって実装される。

[0315] 回路 151 は、情報処理を行う電気回路であり、メモリ 152 にアクセス可能である。例えば、回路 151 は、本開示の符号化方法を実行する専用の回路であってもよいし、本開示の符号化方法に対応するプログラムを実行する汎用の回路であってもよい。また、回路 151 は、CPU 等のプロセッサであってもよい。さらに、回路 151 は、複数の回路の集合体であってもよい。

[0316] メモリ 152 は、回路 151 が画像を符号化するための情報が記憶される専用又は汎用のメモリである。メモリ 152 は、電気回路であってもよく、回路 151 に接続されていてもよい。また、メモリ 152 は、回路 151 に含まれてもよい。また、メモリ 152 は、複数の回路の集合体であってもよい。また、メモリ 152 は、磁気ディスク又は光ディスク等であってもよいし、ストレージ又は記録媒体等と表現されてもよい。また、メモリ 152 は、不揮発性メモリであってもよいし、揮発性メモリであってもよい。

- [0317] 例えば、メモリ 152 には、画像等の符号化対象データが記憶されてもよいし、ビットストリーム等の符号化済みデータが記憶されてもよい。また、メモリ 152 には、回路 151 に画像処理を行わせるためのプログラムが記憶されていてもよい。また、メモリ 152 には、回路 151 に生成モデルが記憶されていてもよい。
- [0318] 図 34 は、符号化装置 100 の基本的な第 1 動作例を示すフローチャートである。符号化装置 100 の回路 151 は、この例の動作において、メモリ 152 を用いて、以下を行う。
- [0319] 具体的には、回路 151 は、顔動画像に関する顔画像のベースデータと、顔画像の 1 つ以上の拡張データをビットストリームに符号化する（S501）。
- [0320] また、ベースデータは、ビットストリームにおいて第 1 フレームに対応するデータに付加される。1 つ以上の拡張データは、ビットストリームにおいて 1 つ以上の第 2 フレームに対応する 1 つ以上のデータに付加される。第 1 フレームは、顔動画像のフレームである。1 つ以上の第 2 フレームは、顔動画像の 1 つ以上のフレームであり、第 1 フレームよりも後の 1 つ以上のフレームである。
- [0321] また、回路 151 は、幾何学情報をビットストリームに符号化する（S502）。ここで、幾何学情報は、顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である。
- [0322] これにより、顔画像に関するベースデータ及び 1 つ以上の拡張データを複数のフレームに分けて符号化することが可能になる場合がある。したがって、1 つのフレームに対応する符号量を削減することが可能になる場合がある。よって、遅延を抑制することが可能になる場合がある。
- [0323] なお、幾何学情報は、顔動画像の複数の第 3 フレームのそれぞれに対応していてもよい。第 1 フレームは、複数の第 3 フレームに含まれていてもよいし、複数の第 3 フレームに含まれていなくてもよい。同様に、1 つ以上の第 2 フレームは、複数の第 3 フレームに含まれていてもよいし、複数の第 3 フ

フレームに含まれていなくてもよい。

[0324] また、ビットストリームにおいて、顔動画像のフレームに対応するデータには、当該フレームに対応する幾何学情報が含まれていてもよいし、当該フレームに対応する復号時刻及び表示時刻等の符号化パラメータが含まれていてもよい。ビットストリームにおいて、顔動画像のフレームに対応するデータは、顔動画像のフレームに対応するアクセスユニットであってもよい。

[0325] 例えば、回路151は、ベースデータと1つ以上の拡張データとである複数の顔画像データの少なくとも1つの制御に関する制御情報をヘッダ領域に符号化してもよい。これにより、顔動画像の再現において、制御情報に従って、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0326] また、例えば、制御情報は、ヘッダ領域によって制御されるアクセスユニットに複数の顔画像データの1つである顔画像データが含まれるか否かを示す存在情報を含んでもよい。これにより、顔動画像の再現において、制御情報に従って、アクセスユニットに顔画像データが含まれるか否かの識別が可能になる場合がある。したがって、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0327] また、例えば、制御情報は、ヘッダ領域によって制御されるアクセスユニットに複数の顔画像データの1つである顔画像データが含まれる場合、顔画像データがベースデータであるか拡張データであるかに関する種別情報を含んでもよい。拡張データは、1つ以上の拡張データの1つである。

[0328] 具体的には、種別情報は、アクセスユニットがベースデータを含む場合、アクセスユニットに含まれる顔画像データが、ベースデータであり、次のベースデータまで継続して用いられることを示してもよい。また、種別情報は、アクセスユニットが拡張データを含む場合、アクセスユニットに含まれる顔画像データが、拡張データであり、ベースデータと共に用いられることを示してもよい。

[0329] これにより、顔動画像の再現において、制御情報に従って、顔画像データ

がベースデータであるか拡張データであるかの識別が可能になる場合がある。そして、顔画像データがベースデータであるか拡張データであるかに従って、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0330] また、例えば、制御情報は、ヘッダ領域によって制御されるアクセスユニットに複数の顔画像データの1つである顔画像データが含まれる場合、適用情報を含んでもよい。ここで、適用情報は、顔動画像の複数のフレームのうちアクセスユニットに対応するフレームの生成及び表示に顔画像データが適用可能であるか否かを示す。これにより、顔動画像の再現において、制御情報に従って、顔動画像のフレームに、当該フレームに対応するデータに付加された顔画像データを適用するか否かを適切に制御することが可能になる場合がある。

[0331] また、例えば、ベースデータと1つ以上の拡張データとのそれぞれは、顔画像に含まれる顔の特徴を示すベクトルで表現されてもよい。これにより、顔画像に関する符号量を削減することが可能になる場合がある。したがって、遅延を抑制することが可能になる場合がある。

[0332] また、例えば、ベースデータと1つ以上の拡張データとのそれぞれは、顔画像に関する画像で表現されてもよい。これにより、顔動画像の再現において、顔画像に関するベースデータ及び拡張データのそれぞれを画像データとして顔動画像のフレームに適切に反映することが可能になる場合がある。

[0333] また、例えば、回路151は、顔画像に含まれる顔の一部のデータをベースデータとして導出して符号化してもよい。また、回路151は、顔画像に含まれる顔の他部のデータを1つ以上の拡張データの1つである拡張データとして導出して符号化してもよい。これにより、複雑な処理を伴わずに、顔画像を複数の部分に分けて符号化することが可能になる場合がある。

[0334] なお、顔の一部は、顔の複数の特徴のうちの一部であってもよいし、顔の複数の領域のうちの一部であってもよい。顔の他部は、顔の複数の特徴のうちの他部であってもよいし、顔の複数の領域のうちの他部であってもよい。

また、ベースデータは、第1部分のデータであってもよい。そして、1つ以上の拡張データは、第1部分とは異なる1つ以上の第2部分の1つ以上のデータであってもよい。複数の拡張データが用いられる場合、互いに異なる複数の第2部分が用いられてもよい。

[0335] また、例えば、回路151は、顔画像における第1周波数範囲のデータをベースデータとして導出して符号化してもよい。また、回路151は、顔画像における第1周波数範囲よりも高い第2周波数範囲のデータを1つ以上の拡張データの1つである拡張データとして導出して符号化してもよい。

[0336] これにより、顔画像における低周波数成分のデータをベースデータとして符号化し、顔画像における高周波数成分のデータを拡張データとして符号化することが可能になる場合がある。そして、顔動画画像の再現において、低周波数成分のデータを顔動画画像の第1フレームの生成に適用し、低周波数成分のデータと高周波数成分のデータとの両方を顔動画画像の第2フレームの生成に適用することが可能になる場合がある。よって、遅延を抑制しつつ、顔動画画像における違和感を抑制することが可能になる場合がある。

[0337] なお、第1周波数範囲のデータは、最低周波数成分のデータ、つまり直流(DC)成分のデータを含んでいてもよい。また、1つ以上の拡張データは、第1周波数範囲よりも高い1つ以上の第2周波数範囲の1つ以上のデータであってもよい。複数の拡張データが用いられる場合、互いに異なる複数の第2周波数範囲が用いられてもよい。

[0338] また、例えば、回路151は、第1画像をベースデータとして符号化してもよい。また、例えば、回路151は、第2画像をベースデータとして符号化してもよい。ここで、第1画像は、顔画像に関する画像であり、第1解像度を有する画像である。第2画像は、顔画像に関する画像であり、第1画像を参照して符号化される画像であり、第1解像度よりも高い第2解像度を有する画像である。

[0339] これにより、低解像度の顔画像をベースデータとして符号化し、高解像度の顔画像を拡張データとして符号化することが可能になる場合がある。そし

て、顔動画像の再現において、低解像度の顔画像を顔動画像の第1フレームの生成に適用し、高解像度の顔画像を顔動画像の第2フレームの生成に適用することが可能になる場合がある。よって、遅延を抑制しつつ、顔動画像における違和感を抑制することが可能になる場合がある。

[0340] なお、1つ以上の拡張データは、第1解像度よりも高い1つ以上の第2解像度を有する1つ以上の第2画像に対応していてもよい。複数の拡張データが用いられる場合、互いに異なる複数の第2解像度が用いられてもよい。また、複数の拡張データは、複数の第2画像に対応していてもよい。また、第2画像の符号化において、第1画像に代えて他の第2画像が参照されてもよい。

[0341] また、1つ以上の第2フレームは、表示順において、第1フレームよりも後であってもよいし、符号化順又は復号順において、第1フレームよりも後であってもよい。

[0342] また、ベースデータは、顔画像の情報である第1情報を含んでいてもよい。1つ以上の拡張データのそれぞれは、顔画像の情報であり第1情報とは異なる情報である第2情報を含んでいてもよい。

[0343] また、符号化装置100は、入力端子と、エントロピー符号化器と、出力端子とを備えてもよい。そして、回路151によって行われる動作が、エントロピー符号化器によって行われてもよい。また、入力端子には、エントロピー符号化器の動作に用いられるデータが入力されてもよい。そして、出力端子から、エントロピー符号化器の動作によって得られたデータが出力されてもよい。

[0344] 図35は、復号装置200の実装例を示すブロック図である。復号装置200は、回路251及びメモリ252を備える。例えば、上述された復号装置200の複数の構成要素は、回路251及びメモリ252によって実装される。

[0345] 回路251は、情報処理を行う電気回路であり、メモリ252にアクセス可能である。例えば、回路251は、本開示の復号方法を実行する専用の回

路であってもよいし、本開示の復号方法に対応するプログラムを実行する汎用の回路であってもよい。また、回路251は、CPU等のプロセッサであってもよい。さらに、回路251は、複数の回路の集合体であってもよい。

[0346] メモリ252は、回路251が画像を復号するための情報が記憶される専用又は汎用のメモリである。メモリ252は、電気回路であってもよく、回路251に接続されていてもよい。また、メモリ252は、回路251に含まれてもよい。また、メモリ252は、複数の回路の集合体であってもよい。また、メモリ252は、磁気ディスク又は光ディスク等であってもよいし、ストレージ又は記録媒体等と表現されてもよい。また、メモリ252は、不揮発性メモリであってもよいし、揮発性メモリであってもよい。

[0347] 例えば、メモリ252には、ビットストリーム等の復号対象データが記憶されてもよいし、画像等の復号済みデータが記憶されてもよい。また、メモリ252には、回路251に画像処理を行わせるためのプログラムが記憶されていてもよい。また、メモリ252には、回路251に生成モデルが記憶されていてもよい。

[0348] 図36は、復号装置200の基本的な第1動作例を示すフローチャートである。復号装置200の回路251は、この例の動作において、メモリ252を用いて、以下を行う。

[0349] 具体的には、回路251は、顔動画像に関する顔画像のベースデータと、顔画像の1つ以上の拡張データとをビットストリームから復号する(S601)。

[0350] また、ベースデータは、ビットストリームにおいて第1フレームに対応するデータに付加されている。1つ以上の拡張データは、ビットストリームにおいて1つ以上の第2フレームに対応する1つ以上のデータに付加されている。第1フレームは、顔動画像のフレームである。1つ以上の第2フレームは、顔動画像の1つ以上のフレームであり、第1フレームよりも後の1つ以上のフレームである。

[0351] また、回路251は、幾何学情報をビットストリームから復号する(S6

02)。ここで、幾何学情報は、顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である。

また、回路251は、ベースデータ、1つ以上の拡張データ及び幾何学情報から、生成モデルを用いて、顔動画像を生成する(S603)。

[0352] これにより、顔画像に関するベースデータ及び1つ以上の拡張データを複数のフレームに分けて復号することが可能になる場合がある。したがって、1つのフレームに対応する符号量を削減することが可能になる場合がある。よって、遅延を抑制することが可能になる場合がある。

[0353] なお、幾何学情報は、顔動画像の複数の第3フレームのそれぞれに対応していてもよい。第1フレームは、複数の第3フレームに含まれていてもよいし、複数の第3フレームに含まれていなくてもよい。同様に、1つ以上の第2フレームは、複数の第3フレームに含まれていてもよいし、複数の第3フレームに含まれていなくてもよい。

[0354] また、ビットストリームにおいて、顔動画像のフレームに対応するデータには、当該フレームに対応する幾何学情報が含まれていてもよいし、当該フレームに対応する復号時刻及び表示時刻等の符号化パラメータが含まれていてもよい。ビットストリームにおいて、顔動画像のフレームに対応するデータは、顔動画像のフレームに対応するアクセスユニットであってもよい。

[0355] 例えば、回路251は、ベースデータと1つ以上の拡張データとである複数の顔画像データの少なくとも1つの制御に関する制御情報をヘッダ領域から復号してもよい。これにより、顔動画像の再現において、制御情報に従って、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0356] また、例えば、制御情報は、ヘッダ領域によって制御されるアクセスユニットに複数の顔画像データの1つである顔画像データが含まれるか否かを示す存在情報を含んでもよい。これにより、顔動画像の再現において、制御情報に従って、アクセスユニットに顔画像データが含まれるか否かの識別が可能になる場合がある。したがって、顔画像データに対して適切な処理を適用

することが可能になる場合がある。

[0357] また、例えば、制御情報は、ヘッダ領域によって制御されるアクセスユニットに複数の顔画像データの1つである顔画像データが含まれる場合、顔画像データがベースデータであるか拡張データであるかに関する種別情報を含んでいてもよい。拡張データは、1つ以上の拡張データの1つである。

[0358] 具体的には、種別情報は、アクセスユニットがベースデータを含む場合、アクセスユニットに含まれる顔画像データが、ベースデータであり、次のベースデータまで継続して用いられることを示してもよい。また、種別情報は、アクセスユニットが拡張データを含む場合、アクセスユニットに含まれる顔画像データが、拡張データであり、ベースデータと共に用いられることを示してもよい。

[0359] これにより、顔動画像の再現において、制御情報に従って、顔画像データがベースデータであるか拡張データであるかの識別が可能になる場合がある。そして、顔画像データがベースデータであるか拡張データであるかに従って、顔画像データに対して適切な処理を適用することが可能になる場合がある。

[0360] また、例えば、制御情報は、ヘッダ領域によって制御されるアクセスユニットに複数の顔画像データの1つである顔画像データが含まれる場合、適用情報を含んでいてもよい。ここで、適用情報は、顔動画像の複数のフレームのうちアクセスユニットに対応するフレームの生成及び表示に顔画像データが適用可能であるか否かを示す。これにより、顔動画像の再現において、制御情報に従って、顔動画像のフレームに、当該フレームに対応するデータに付加された顔画像データを適用するか否かを適切に制御することが可能になる場合がある。

[0361] また、例えば、ベースデータと1つ以上の拡張データとのそれぞれは、顔画像に含まれる顔の特徴を示すベクトルで表現されてもよい。これにより、顔画像に関する符号量を削減することが可能になる場合がある。したがって、遅延を抑制することが可能になる場合がある。

- [0362] また、例えば、ベースデータと1つ以上の拡張データとのそれぞれは、顔画像に関する画像で表現されてもよい。これにより、顔動画像の再現において、顔画像に関するベースデータ及び拡張データのそれぞれを画像データとして顔動画像のフレームに適切に反映することが可能になる場合がある。
- [0363] また、例えば、回路251は、ベースデータと1つ以上の拡張データの少なくとも1つと幾何学情報とを生成モデルに入力して、顔動画像のフレームを生成してもよい。これにより、顔画像に関するベースデータと拡張データとから中間画像を生成する処理を省略することが可能になる場合がある。したがって、顔動画像を生成する処理を簡略化することが可能になる場合がある。なお、幾何学情報が対応するフレームが、顔動画像のフレームとして生成され得る。
- [0364] また、例えば、回路251は、ベースデータと1つ以上の拡張データの少なくとも1つとから中間画像を生成してもよい。そして、回路251は、中間画像と幾何学情報とを生成モデルに入力して、顔動画像のフレームを生成してもよい。
- [0365] これにより、顔画像に関するベースデータと拡張データとから、顔画像に関する中間画像を適切に生成することが可能になる場合がある。そして、顔画像に関する中間画像を顔動画像のフレームに適切に反映することが可能になる場合がある。なお、幾何学情報が対応するフレームが、顔動画像のフレームとして生成され得る。
- [0366] また、例えば、回路251は、ベースデータを参照して、1つ以上の拡張データの1つである拡張データを復号してもよい。そして、回路251は、拡張データと幾何学情報とを生成モデルに入力して、顔動画像のフレームを生成してもよい。
- [0367] これにより、ベースデータの精度よりも高い精度を有する拡張データを効率的に復号することが可能になる場合がある。そして、高精度の拡張データを用いて、顔動画像のフレームを高精度に生成することが可能になる場合がある。なお、幾何学情報が対応するフレームが、顔動画像のフレームとして

生成され得る。

[0368] また、例えば、ベースデータは、顔画像に含まれる顔の一部のデータであってもよい。また、1つ以上の拡張データの1つである拡張データは、顔画像に含まれる顔の他部のデータであってもよい。これにより、複雑な処理を伴わずに、顔画像を複数の部分に分けて復号することが可能になる場合がある。

[0369] なお、顔の一部は、顔の複数の特徴のうちの一部であってもよいし、顔の複数の領域のうちの一部であってもよい。顔の他部は、顔の複数の特徴のうちの他部であってもよいし、顔の複数の領域のうちの他部であってもよい。また、ベースデータは、第1部分のデータであってもよい。そして、1つ以上の拡張データは、第1部分とは異なる1つ以上の第2部分の1つ以上のデータであってもよい。複数の拡張データが用いられる場合、互いに異なる複数の第2部分が用いられてもよい。

[0370] また、例えば、ベースデータは、顔画像における第1周波数範囲のデータであってもよい。また、1つ以上の拡張データの1つである拡張データは、顔画像における第1周波数範囲よりも高い第2周波数範囲のデータであってもよい。

[0371] これにより、顔画像における低周波数成分のデータをベースデータとして復号し、顔画像における高周波数成分のデータを拡張データとして復号することが可能になる場合がある。そして、顔動画像の再現において、低周波数成分のデータを顔動画像の第1フレームの生成に適用し、低周波数成分のデータと高周波数成分のデータとの両方を顔動画像の第2フレームの生成に適用することが可能になる場合がある。よって、遅延を抑制しつつ、顔動画像における違和感を抑制することが可能になる場合がある。

[0372] なお、第1周波数範囲のデータは、最低周波数成分のデータ、つまり直流(DC)成分のデータを含んでいてもよい。また、1つ以上の拡張データは、第1周波数範囲よりも高い1つ以上の第2周波数範囲の1つ以上のデータであってもよい。複数の拡張データが用いられる場合、互いに異なる複数の

第2周波数範囲が用いられてもよい。

[0373] また、例えば、ベースデータは、第1画像に対応していてもよい。1つ以上の拡張データの1つである拡張データは、第2画像に対応していてもよい。ここで、第1画像は、顔画像に関する画像であり、第1解像度を有する画像である。第2画像は、顔画像に関する画像であり、第1画像を参照して復号される画像であり、第1解像度よりも高い第2解像度を有する画像である。

[0374] これにより、低解像度の顔画像をベースデータとして復号し、高解像度の顔画像を拡張データとして復号することが可能になる場合がある。そして、顔動画像の再現において、低解像度の顔画像を顔動画像の第1フレームの生成に適用し、高解像度の顔画像を顔動画像の第2フレームの生成に適用することが可能になる場合がある。よって、遅延を抑制しつつ、顔動画像における違和感を抑制することが可能になる場合がある。

[0375] なお、1つ以上の拡張データは、第1解像度よりも高い1つ以上の第2解像度を有する1つ以上の第2画像に対応していてもよい。複数の拡張データが用いられる場合、互いに異なる複数の第2解像度が用いられてもよい。また、複数の拡張データは、複数の第2画像に対応していてもよい。また、第2画像の復号において、第1画像に代えて他の第2画像が参照されてもよい。

[0376] また、例えば、ベースデータは、第1画像に対応していてもよい。1つ以上の拡張データの1つである拡張データは、第2画像に対応していてもよい。ここで、第1画像は、顔画像に関する画像であり、第1量子化幅で復号される画像である。第2画像は、顔画像に関する画像であり、第1画像を参照して第1量子化幅よりも細かい第2量子化幅で復号される画像である。

[0377] これにより、粗い顔画像をベースデータとして復号し、細かい顔画像を拡張データとして復号することが可能になる場合がある。そして、顔動画像の再現において、粗い顔画像を顔動画像の第1フレームの生成に適用し、細かい顔画像を顔動画像の第2フレームの生成に適用することが可能になる場合

がある。よって、遅延を抑制しつつ、顔動画像における違和感を抑制することが可能になる場合がある。

[0378] なお、細かい量子化幅は、小さい量子化幅であってもよい。また、1つ以上の拡張データは、第1量子化幅よりも細かい1つ以上の第2量子化幅で復号される1つ以上の第2画像に対応していてもよい。複数の拡張データが用いられる場合、互いに異なる複数の第2量子化幅が用いられてもよい。また、複数の拡張データは、複数の第2画像に対応していてもよい。また、第2画像の復号において、第1画像に代えて他の第2画像が参照されてもよい。

[0379] また、例えば、制御情報は、1つ以上の拡張データのそれぞれを識別するための識別情報を含んでいてもよい。これにより、顔動画像の再現において、制御情報に従って、各拡張データを識別することが可能になる場合がある。したがって、各拡張データを個別に指定し、各拡張データの適用を制御することが可能になる場合がある。

[0380] また、例えば、制御情報は、(i) ベースデータを含むアクセスユニットのヘッダ領域に含まれる情報であり、(ii) 1つ以上の拡張データの総数を示す情報である総数情報を含んでいてもよい。これにより、顔動画像の再現において、制御情報に従って、1つ以上の拡張データの総数を識別することが可能になる場合がある。したがって、1つ以上の拡張データの総数に従って、顔動画像の再現に用いられ得る1つ以上の拡張データを効率的に確定することが可能になる場合がある。

[0381] また、例えば、制御情報は、(i) ベースデータを含むアクセスユニットのヘッダ領域に含まれる情報である特定情報を含んでいてもよい。また、特定情報は、(ii) 1つ以上の拡張データのうちの拡張データであって、1つ以上の第2フレームのうち当該拡張データを含むアクセスユニットに対応する第2フレームの生成及び表示に適用可能な拡張データを特定するための情報であってもよい。

[0382] これにより、顔動画像の再現において、制御情報に従って、顔動画像のフレームの生成及び表示に適用可能な拡張データを適切に特定することが可能

になる場合がある。

- [0383] また、例えば、回路251は、メモリ252においてビットストリームが記憶されるストリームバッファを制御するための少なくとも1つの制御パラメータを復号してもよい。また、少なくとも1つの制御パラメータは、ストリームバッファのバッファサイズを基準サイズ以下に制御し、復号処理開始の初期遅延時間を基準遅延時間以下に制御するための少なくとも1つのパラメータであってもよい。これにより、復号のためのリソースを削減することが可能になり、遅延時間を短くすることが可能になる。
- [0384] また、例えば、1つ以上の第2フレームは、表示順において、第1フレームよりも後であってもよいし、符号化順又は復号順において、第1フレームよりも後であってもよい。
- [0385] また、ベースデータは、顔画像の情報である第1情報を含んでもよい。1つ以上の拡張データのそれぞれは、顔画像の情報であり第1情報とは異なる情報である第2情報を含んでもよい。
- [0386] また、例えば、復号装置200は、入力端子と、エントロピー復号器と、出力端子とを備えてもよい。そして、回路251によって行われる動作が、エントロピー復号器によって行われてもよい。また、入力端子には、エントロピー復号器の動作に用いられるデータが入力されてもよい。そして、出力端子から、エントロピー復号器の動作によって得られたデータが出力されてもよい。
- [0387] また、例えば、ビットストリームを記憶するコンピュータ読み取り可能な非一時的な記録媒体が用いられてもよい。ビットストリームは、顔動画像に関する顔画像のベースデータと、顔画像の1つ以上の拡張データと、幾何学情報とを含んでもよい。
- [0388] また、ベースデータは、ビットストリームにおいて第1フレームに対応するデータに付加されている。1つ以上の拡張データは、ビットストリームにおいて1つ以上の第2フレームに対応する1つ以上のデータに付加されている。第1フレームは、顔動画像のフレームである。1つ以上の第2フレーム

は、顔動画像の1つ以上のフレームであり、第1フレームよりも後の1つ以上のフレームである。また、幾何学情報は、顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である。

[0389] そして、ビットストリームは、(i) ベースデータ、1つ以上の拡張データ及び幾何学情報を復号し、(ii) ベースデータ、1つ以上の拡張データ及び幾何学情報から、生成モデルを用いて、顔動画像を生成する処理を復号装置200に実行させてもよい。

[0390] これにより、上述の復号装置及び復号方法に対応する1つ以上のビットストリームを記憶する記録媒体を実現することが可能になる場合がある。したがって、上述の復号装置200等と同様の効果を記録媒体によって得ることが可能になる場合がある。

[0391] [その他の例]

上述された各例における符号化装置100及び復号装置200は、それぞれ、画像符号化装置及び画像復号装置として利用されてもよいし、動画像符号化装置及び動画像復号装置として利用されてもよい。また、符号化装置100に含まれる複数の構成要素と、復号装置200に含まれる複数の構成要素とは、互いに対応する動作を行ってもよい。

[0392] また、符号化するという表現は、格納する、含める、書き込む、記述する、信号化する、送り出す、通知する、又は、保存する等の表現に置き換えられてもよく、これらの表現が相互に置き換えられてもよい。例えば、情報を符号化することは、ビットストリームに情報を含めることであってもよい。また、情報をビットストリームに符号化することは、情報を符号化して、符号化された情報を含むビットストリームを生成することを意味してもよい。

[0393] また、復号するという表現は、読み出す、読み解く、読み取る、読み込む、導出する、取得する、受け取る、抽出する、又は、復元する等の表現に置き換えられてもよく、これらの表現が相互に置き換えられてもよい。例えば、情報を復号することは、ビットストリームから情報を取得することであっ

てもよい。また、ビットストリームから情報を復号することは、ビットストリームを復号して、ビットストリームに含まれる情報を取得することを意味してもよい。

[0394] また、例えば、ビットストリームに含まれる符号化情報及び圧縮情報等が、単に、情報と表現される場合がある。

[0395] また、上述された各例の少なくとも一部が、符号化方法として利用されてもよいし、復号方法として利用されてもよいし、エントロピー符号化方法として利用されてもよいし、エントロピー復号方法として利用されてもよいし、その他の方法として利用されてもよい。

[0396] また、各構成要素は、専用のハードウェアで構成されるか、各構成要素に適したソフトウェアプログラムを実行することによって実現されてもよい。各構成要素は、CPU又はプロセッサなどのプログラム実行部が、ハードディスク又は半導体メモリなどの記録媒体に記録されたソフトウェアプログラムを読み出して実行することによって実現されてもよい。

[0397] 具体的には、符号化装置100及び復号装置200のそれぞれは、処理回路(Processing Circuitry)と、当該処理回路に電気的に接続された、当該処理回路からアクセス可能な記憶装置(Storage)とを備えていてもよい。例えば、処理回路は回路151又は251に対応し、記憶装置はメモリ152又は252に対応する。

[0398] 処理回路は、専用のハードウェア及びプログラム実行部の少なくとも一方を含み、記憶装置を用いて処理を実行する。また、記憶装置は、処理回路がプログラム実行部を含む場合には、当該プログラム実行部により実行されるソフトウェアプログラムを記憶する。

[0399] 上述したソフトウェアプログラムの一例は、ビットストリームである。ビットストリームは、符号化された画像と、画像を復号する復号処理を行うためのシンタックスを含む。ビットストリームは、シンタックスに基づく処理を復号装置200に実行させることで、画像を復号装置200に復号させる。また、例えば、上述された符号化装置100又は復号装置200などを実

現するためのソフトウェアは、次のようなプログラムである。

- [0400] 例えば、このプログラムは、コンピュータに、顔動画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームに符号化し、前記顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームに符号化し、前記ベースデータは、前記ビットストリームにおいて、前記顔動画像のフレームである第1フレームに対応するデータに付加され、前記1つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画像の1つ以上のフレームであり、前記第1フレームよりも後の1つ以上のフレームである1つ以上の第2フレームに対応する1つ以上のデータに付加される、符号化方法を実行させてもよい。
- [0401] また、例えば、このプログラムは、コンピュータに、顔動画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームから復号し、前記顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームから復号し、前記ベースデータ、前記1つ以上の拡張データ及び前記幾何学情報から、生成モデルを用いて、前記顔動画像を生成し、前記ベースデータは、前記ビットストリームにおいて、前記顔動画像のフレームである第1フレームに対応するデータに付加されており、前記1つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画像の1つ以上のフレームであり、前記第1フレームよりも後の1つ以上のフレームである1つ以上の第2フレームに対応する1つ以上のデータに付加されている、復号方法を実行させてもよい。
- [0402] また、上述された各構成要素は、回路であってもよい。これらの回路は、全体として1つの回路を構成してもよいし、それぞれ別々の回路であってもよい。また、各構成要素は、汎用的なプロセッサで実現されてもよいし、専用のプロセッサで実現されてもよい。
- [0403] また、特定の構成要素が実行する処理を別の構成要素が実行してもよい。

また、処理を実行する順番が変更されてもよいし、複数の処理が並行して実行されてもよい。また、本開示の複数の例のうちの任意の2つ以上が、適宜、組み合わされて、実施されてもよい。また、符号化復号装置が、符号化装置100及び復号装置200を備えていてもよい。

[0404] また、本開示における複数の構成要素の全てが実装されなくてもよく、本開示における複数の構成要素の一部のみが実装されてもよい。同様に、本開示における複数の処理の全てが実行されなくてもよく、本開示における複数の処理の一部のみが実行されてもよい。

[0405] また、説明に用いられた第1及び第2等の序数は、適宜、付け替えられてもよい。また、構成要素などに対して、序数が新たに与えられてもよいし、取り除かれてもよい。また、これらの序数は、要素を識別するため、要素に付けられる場合があり、意味のある順序に対応しない場合がある。

[0406] また、例えば、第1要素、第2要素及び第3要素の少なくとも1つ（又は1つ以上）という表現は、第1要素、第2要素、第3要素、又は、これらの任意の組み合わせに対応する。

[0407] 以上、符号化装置100及び復号装置200の態様について、複数の例に基づいて説明したが、符号化装置100及び復号装置200の態様は、これらの例に限定されるものではない。本開示の趣旨を逸脱しない限り、当業者が思いつく各種変形を各例に施したものや、異なる例における構成要素を組み合わせて構築される形態も、符号化装置100及び復号装置200の態様の範囲内に含まれてもよい。

[0408] ここで開示された1以上の態様を本開示における他の態様の少なくとも一部と組み合わせて実施してもよい。また、ここで開示された1以上の態様のフローチャートに記載の一部の処理、装置の一部の構成、シンタックスの一部などを他の態様と組み合わせて実施してもよい。

[0409] [実施及び応用]

以上の各実施の形態において、機能的又は作用的なブロックの各々は、通常、MPU (micro processing unit) 及びメモリ等

によって実現可能である。また、機能ブロックの各々による処理は、ROM等の記録媒体に記録されたソフトウェア（プログラム）を読み出して実行するプロセッサなどのプログラム実行部として実現されてもよい。当該ソフトウェアは、配布されてもよい。当該ソフトウェアは、半導体メモリなどの様々な記録媒体に記録されてもよい。なお、各機能ブロックをハードウェア（専用回路）によって実現することも可能である。

[0410] 各実施の形態において説明した処理は、単一の装置（システム）を用いて集中処理することによって実現してもよく、又は、複数の装置を用いて分散処理することによって実現してもよい。また、上記プログラムを実行するプロセッサは、単数であってもよく、複数であってもよい。すなわち、集中処理を行ってもよく、又は分散処理を行ってもよい。

[0411] 本開示の態様は、以上の実施例に限定されることなく、種々の変更が可能であり、それらも本開示の態様の範囲内に包含される。

[0412] さらにここで、上記各実施の形態で示した動画像符号化方法（画像符号化方法）又は動画像復号化方法（画像復号方法）の応用例、及び、その応用例を実施する種々のシステムを説明する。このようなシステムは、画像符号化方法を用いた画像符号化装置、画像復号方法を用いた画像復号装置、又は、両方を備える画像符号化復号装置を有することを特徴としてもよい。このようなシステムの他の構成について、場合に応じて適切に変更することができる。

[0413] [使用例]

図37は、コンテンツ配信サービスを実現する適切なコンテンツ供給システムex100の全体構成を示す図である。通信サービスの提供エリアを所望の大きさに分割し、各セル内にそれぞれ、図示された例における固定無線局である基地局ex106、ex107、ex108、ex109、ex110が設置されている。

[0414] このコンテンツ供給システムex100では、インターネットex101に、インターネットサービスプロバイダex102又は通信網ex104、

及び基地局 e x 1 0 6 ~ e x 1 1 0 を介して、コンピュータ e x 1 1 1、ゲーム機 e x 1 1 2、カメラ e x 1 1 3、家電 e x 1 1 4、及びスマートフォン e x 1 1 5 などの各機器が接続される。当該コンテンツ供給システム e x 1 0 0 は、上記のいずれかの装置を組合せて接続するようにしてもよい。種々の実施において、基地局 e x 1 0 6 ~ e x 1 1 0 を介さずに、各機器が電話網又は近距離無線等を介して直接的又は間接的に相互に接続されていてもよい。さらに、ストリーミングサーバ e x 1 0 3 は、インターネット e x 1 0 1 等を介して、コンピュータ e x 1 1 1、ゲーム機 e x 1 1 2、カメラ e x 1 1 3、家電 e x 1 1 4、及びスマートフォン e x 1 1 5 などの各機器と接続されてもよい。また、ストリーミングサーバ e x 1 0 3 は、衛星 e x 1 1 6 を介して、飛行機 e x 1 1 7 内のホットスポット内の端末等と接続されてもよい。

[0415] なお、基地局 e x 1 0 6 ~ e x 1 1 0 の代わりに、無線アクセスポイント又はホットスポット等が用いられてもよい。また、ストリーミングサーバ e x 1 0 3 は、インターネット e x 1 0 1 又はインターネットサービスプロバイダ e x 1 0 2 を介さずに直接通信網 e x 1 0 4 と接続されてもよいし、衛星 e x 1 1 6 を介さず直接飛行機 e x 1 1 7 と接続されてもよい。

[0416] カメラ e x 1 1 3 はデジタルカメラ等の静止画撮影、及び動画撮影が可能な機器である。また、スマートフォン e x 1 1 5 は、2 G、3 G、3. 9 G、4 G、そして今後は 5 G と呼ばれる移動通信システムの方式に対応したスマートフォン機、携帯電話機、又は PHS (P e r s o n a l H a n d y p h o n e S y s t e m) 等である。

[0417] 家電 e x 1 1 4 は、冷蔵庫、又は家庭用燃料電池コージェネレーションシステムに含まれる機器等である。

[0418] コンテンツ供給システム e x 1 0 0 では、撮影機能を有する端末が基地局 e x 1 0 6 等を通じてストリーミングサーバ e x 1 0 3 に接続されることで、ライブ配信等が可能になる。ライブ配信では、端末（コンピュータ e x 1 1 1、ゲーム機 e x 1 1 2、カメラ e x 1 1 3、家電 e x 1 1 4、スマート

フォン e x 1 1 5、及び飛行機 e x 1 1 7 内の端末等)は、ユーザが当該端末を用いて撮影した静止画又は動画コンテンツに対して上記各実施の形態で説明した符号化処理を行ってもよく、符号化により得られた映像データと、映像に対応する音を符号化した音データと多重化してもよく、得られたデータをストリーミングサーバ e x 1 0 3 に送信してもよい。即ち、各端末は、本開示の一態様に係る画像符号化装置として機能する。

[0419] 一方、ストリーミングサーバ e x 1 0 3 は要求のあったクライアントに対して送信されたコンテンツデータをストリーム配信する。クライアントは、上記符号化処理されたデータを復号化することが可能な、コンピュータ e x 1 1 1、ゲーム機 e x 1 1 2、カメラ e x 1 1 3、家電 e x 1 1 4、スマートフォン e x 1 1 5、又は飛行機 e x 1 1 7 内の端末等である。配信されたデータを受信した各機器は、受信したデータを復号化処理して再生する。即ち、各機器は、本開示の一態様に係る画像復号装置として機能してもよい。

[0420] [分散処理]

また、ストリーミングサーバ e x 1 0 3 は複数のサーバ又は複数のコンピュータであって、データを分散して処理したり記録したり配信するものであってもよい。例えば、ストリーミングサーバ e x 1 0 3 は、CDN (C o n t e n t s D e l i v e r y N e t w o r k) により実現され、世界中に分散された多数のエッジサーバとエッジサーバ間をつなぐネットワークによりコンテンツ配信が実現されていてよい。CDNでは、クライアントに応じて物理的に近いエッジサーバが動的に割り当てられる。そして、当該エッジサーバにコンテンツがキャッシュ及び配信されることで遅延を減らすことができる。また、いくつかのタイプのエラーが発生した場合又はトラフィックの増加などにより通信状態が変わる場合に複数のエッジサーバで処理を分散したり、他のエッジサーバに配信主体を切り替えたり、障害が生じたネットワークの部分を迂回して配信を続けることができるので、高速かつ安定した配信が実現できる。

[0421] また、配信自体の分散処理にとどまらず、撮影したデータの符号化処理を

各端末で行ってもよいし、サーバ側で行ってもよいし、互いに分担して行ってもよい。一例として、一般に符号化処理では、処理ループが2度行われる。1度目のループでフレーム又はシーン単位での画像の複雑さ、又は、符号量が検出される。また、2度目のループでは画質を維持して符号化効率を向上させる処理が行われる。例えば、端末が1度目の符号化処理を行い、コンテンツを受け取ったサーバ側が2度目の符号化処理を行うことで、各端末での処理負荷を減らしつつもコンテンツの質と効率を向上させることができる。この場合、ほぼリアルタイムで受信して復号する要求があれば、端末が行った一度目の符号化済みデータを他の端末で受信して再生することもできるので、より柔軟なリアルタイム配信も可能になる。

[0422] 他の例として、カメラ $e \times 113$ 等は、画像から特徴量抽出を行い、特徴量に関するデータをメタデータとして圧縮してサーバに送信する。サーバは、例えば特徴量からオブジェクトの重要性を判断して量子化精度を切り替えるなど、画像の意味（又は内容の重要性）に応じた圧縮を行う。特徴量データはサーバでの再度の圧縮時の動きベクトル予測の精度及び効率向上に特に有効である。また、端末でVLC（可変長符号化）などの簡易的な符号化を行い、サーバでCABAC（コンテキスト適応型二値算術符号化方式）など処理負荷の大きな符号化を行ってもよい。

[0423] さらに他の例として、スタジアム、ショッピングモール、又は工場などにおいては、複数の端末によりほぼ同一のシーンが撮影された複数の映像データが存在する場合がある。この場合には、撮影を行った複数の端末と、必要に応じて撮影をしていない他の端末及びサーバを用いて、例えばGOP（Group of Picture）単位、ピクチャ単位、又はピクチャを分割したタイル単位などで符号化処理をそれぞれ割り当てて分散処理を行う。これにより、遅延を減らし、よりリアルタイム性を実現できる。

[0424] 複数の映像データはほぼ同一シーンであるため、各端末で撮影された映像データを互いに参照し合えるように、サーバで管理及び／又は指示をしてもよい。また、各端末からの符号化済みデータを、サーバが受信し複数のデー

タ間で参照関係を変更、又はピクチャ自体を補正或いは差し替えて符号化しなおしてもよい。これにより、一つ一つのデータの質と効率を高めたストリームを生成できる。

[0425] さらに、サーバは、映像データの符号化方式を変更するトランスコードを行ったうえで映像データを配信してもよい。例えば、サーバは、MPEG系の符号化方式をVP系（例えばVP9）に変換してもよいし、H. 264をH. 265に変換してもよい。

[0426] このように、符号化処理は、端末、又は1以上のサーバにより行うことが可能である。よって、以下では、処理を行う主体として「サーバ」又は「端末」等の記載を用いるが、サーバで行われる処理の一部又は全てが端末で行われてもよいし、端末で行われる処理の一部又は全てがサーバで行われてもよい。また、これらに関しては、復号処理についても同様である。

[0427] [3D、マルチアングル]

互いにほぼ同期した複数のカメラ $\times 113$ 及び／又はスマートフォン $\times 115$ などの端末により撮影された異なるシーン、又は、同一シーンを異なるアングルから撮影した画像或いは映像を統合して利用することが増えてきている。各端末で撮影した映像は、別途取得した端末間の相対的な位置関係、又は、映像に含まれる特徴点が一致する領域などに基づいて統合される。

[0428] サーバは、2次元の動画像を符号化するだけでなく、動画像のシーン解析などに基づいて自動的に、又は、ユーザが指定した時刻において、静止画を符号化し、受信端末に送信してもよい。サーバは、さらに、撮影端末間の相対的な位置関係を取得できる場合には、2次元の動画像だけでなく、同一シーンが異なるアングルから撮影された映像に基づき、当該シーンの3次元形状を生成できる。サーバは、ポイントクラウドなどにより生成した3次元のデータを別途符号化してもよいし、3次元データを用いて人物又はオブジェクトを認識或いは追跡した結果に基づいて、受信端末に送信する映像を、複数の端末で撮影した映像から、選択、又は、再構成して生成してもよい。

[0429] このようにして、ユーザは、各撮影端末に対応する各映像を任意に選択してシーンを楽しむこともできるし、複数画像又は映像を用いて再構成された3次元データから選択視点の映像を切り出したコンテンツを楽しむこともできる。さらに、映像と共に、音も複数の相異なるアングルから收音され、サーバは、特定のアングル又は空間からの音に対応する映像と多重化して、多重化された映像と音とを送信してもよい。

[0430] また、近年ではVirtual Reality (VR) 及びAugmented Reality (AR) など、現実世界と仮想世界とを対応付けたコンテンツも普及してきている。VRの画像の場合、サーバは、右目用及び左目用の視点画像をそれぞれ作成し、Multi-View Coding (MVC) などにより各視点映像間で参照を許容する符号化を行ってもよいし、互いに参照せずに別ストリームとして符号化してもよい。別ストリームの復号時には、ユーザの視点に応じて仮想的な3次元空間が再現されるように互いに同期させて再生するとよい。

[0431] ARの画像の場合には、サーバは、現実空間のカメラ情報に、仮想空間上の仮想物体情報を、3次元的位置又はユーザの視点の動きに基づいて重畳する。復号装置は、仮想物体情報及び3次元データを取得又は保持し、ユーザの視点の動きに応じて2次元画像を生成し、スムーズにつなげることで重畳データを作成してもよい。または、復号装置は仮想物体情報の依頼に加えてユーザの視点の動きをサーバに送信してもよい。サーバは、サーバに保持される3次元データから受信した視点の動きに合わせて重畳データを作成し、重畳データを符号化して復号装置に配信してもよい。なお、重畳データは、RGB以外に透過度を示す α 値を有し、サーバは、3次元データから作成されたオブジェクト以外の部分の α 値が0などに設定し、当該部分が透過する状態で、符号化してもよい。もしくは、サーバは、クロマキーのように所定の値のRGB値を背景に設定し、オブジェクト以外の部分は背景色にしたデータを生成してもよい。

[0432] 同様に配信されたデータの復号処理はクライアントである各端末で行って

も、サーバ側で行ってもよいし、互いに分担して行ってもよい。一例として、ある端末が、一旦サーバに受信リクエストを送り、そのリクエストに応じたコンテンツを他の端末で受信し復号処理を行い、ディスプレイを有する装置に復号済みの信号が送信されてもよい。通信可能な端末自体の性能によらず処理を分散して適切なコンテンツを選択することで画質のよいデータを再生することができる。また、他の例として大きなサイズの画像データをTV等で受信しつつ、鑑賞者の個人端末にピクチャが分割されたタイルなど一部の領域が復号されて表示されてもよい。これにより、全体像を共有化しつつ、自身の担当分野又はより詳細に確認したい領域を手元で確認することができる。

[0433] 屋内外の近距離、中距離、又は長距離の無線通信が複数使用可能な状況下で、MPEG-DASH (Dynamic Adaptive Streaming over HTTP) などの配信システム規格を利用して、シームレスにコンテンツを受信することが可能かもしれない。ユーザは、ユーザの端末、屋内外に配置されたディスプレイなどの復号装置又は表示装置を自由に選択しながらリアルタイムで切り替えてもよい。また、自身の位置情報などを用いて、復号する端末及び表示する端末を切り替えながら復号を行うことができる。これにより、ユーザが目的地へ移動している間に、表示可能なデバイスが埋め込まれた隣の建物の壁面又は地面の一部に情報をマップ及び表示することが可能になる。また、符号化データが受信端末から短時間でアクセスできるサーバにキャッシュされている、又は、コンテンツ・デリバリー・サービスにおけるエッジサーバにコピーされている、などの、ネットワーク上での符号化データへのアクセス容易性に基づいて、受信データのビットレートを切り替えることも可能である。

[0434] [Webページの最適化]

図38は、コンピュータex111等におけるwebページの表示画面例を示す図である。図39は、スマートフォンex115等におけるwebページの表示画面例を示す図である。図38及び図39に示すようにwebペ

ージが、画像コンテンツへのリンクであるリンク画像を複数含む場合があり、閲覧するデバイスによってその見え方は異なる。画面上に複数のリンク画像が見える場合には、ユーザが明示的にリンク画像を選択するまで、又は画面の中央付近にリンク画像が近付く或いはリンク画像の全体が画面内に入るまで、表示装置（復号装置）は、リンク画像として各コンテンツが有する静止画又はIピクチャを表示してもよいし、複数の静止画又はIピクチャ等でgifアニメのような映像を表示してもよいし、ベースレイヤのみを受信し、映像を復号及び表示してもよい。

[0435] ユーザによりリンク画像が選択された場合、表示装置は、ベースレイヤを最優先にしつつ復号を行う。なお、webページを構成するHTML（HyperText Markup Language）にスケーラブルなコンテンツであることを示す情報があれば、表示装置は、エンハンスメントレイヤまで復号してもよい。さらに、リアルタイム性を担保するために、選択される前又は通信帯域が非常に厳しい場合には、表示装置は、前方参照のピクチャ（Iピクチャ、Pピクチャ、前方参照のみのBピクチャ）のみを復号及び表示することで、先頭ピクチャの復号時刻と表示時刻との間の遅延（コンテンツの復号開始から表示開始までの遅延）を低減できる。またさらに、表示装置は、ピクチャの参照関係を敢えて無視して、全てのBピクチャ及びPピクチャを前方参照にして粗く復号し、時間が経ち受信したピクチャが増えるにつれて正常の復号を行ってもよい。

[0436] [自動走行]

また、車の自動走行又は走行支援のため2次元又は3次元の地図情報などのような静止画又は映像データを送受信する場合、受信端末は、1以上のレイヤに属する画像データに加えて、メタ情報として天候又は工事の情報なども受信し、これらに対応付けて復号してもよい。なお、メタ情報は、レイヤに属してもよいし、単に画像データと多重化されてもよい。

[0437] この場合、受信端末を含む車、ドローン又は飛行機などが移動するため、受信端末は、当該受信端末の位置情報を送信することで、基地局ex106

～e x 1 1 0を切り替えながらシームレスな受信及び復号の実行を実現できる。また、受信端末は、ユーザの選択、ユーザの状況及び／又は通信帯域の状態に応じて、メタ情報をどの程度受信するか、又は地図情報をどの程度更新していくかを動的に切り替えることが可能になる。

[0438] コンテンツ供給システムe x 1 0 0では、ユーザが送信した符号化された情報をリアルタイムでクライアントが受信して復号し、再生することができる。

[0439] [個人コンテンツの配信]

また、コンテンツ供給システムe x 1 0 0では、映像配信業者による高画質で長時間のコンテンツのみならず、個人による低画質で短時間のコンテンツのユニキャスト、又はマルチキャスト配信が可能である。このような個人のコンテンツは今後も増加していくと考えられる。個人コンテンツをより優れたコンテンツにするために、サーバは、編集処理を行ってから符号化処理を行ってもよい。これは、例えば、以下のような構成を用いて実現できる。

[0440] 撮影時にリアルタイム又は蓄積して撮影後に、サーバは、原画データ又は符号化済みデータから撮影エラー、シーン探索、意味の解析、及びオブジェクト検出などの認識処理を行う。そして、サーバは、認識結果に基づいて手動又は自動で、ピントずれ又は手ブレなどを補正したり、明度が他のピクチャに比べて低い又は焦点が合っていないシーンなどの重要性の低いシーンを削除したり、オブジェクトのエッジを強調したり、色合いを変化させるなどの編集を行う。サーバは、編集結果に基づいて編集後のデータを符号化する。また撮影時刻が長すぎると視聴率が下がることも知られており、サーバは、撮影時間に応じて特定の時間範囲内のコンテンツになるように上記のように重要性が低いシーンのみならず動きが少ないシーンなどを、画像処理結果に基づき自動でクリップしてもよい。または、サーバは、シーンの意味解析の結果に基づいてダイジェストを生成して符号化してもよい。

[0441] 個人コンテンツには、そのままでは著作権、著作者人格権、又は肖像権等の侵害となるものが写り込んでいるケースもあり、共有する範囲が意図した

範囲を超えてしまうなど個人にとって不都合な場合もある。よって、例えば、サーバは、画面の周辺部の人の顔、又は家の中などを取えて焦点が合わない画像に変更して符号化してもよい。さらに、サーバは、符号化対象画像内に、予め登録した人物とは異なる人物の顔が映っているかどうかを認識し、映っている場合には、顔の部分にモザイクをかけるなどの処理を行ってもよい。または、符号化の前処理又は後処理として、著作権などの観点からユーザが画像を加工したい人物又は背景領域を指定してもよい。サーバは、指定された領域を別の映像に置き換える、又は焦点をぼかすなどの処理を行ってもよい。人物であれば、動画像において人物をトラッキングして、人物の顔の部分の映像を置き換えることができる。

[0442] データ量の小さい個人コンテンツの視聴はリアルタイム性の要求が強いため、帯域幅にもよるが、復号装置は、まずベースレイヤを最優先で受信して復号及び再生を行う。復号装置は、この間にエンハンスメントレイヤを受信し、再生がループされる場合など2回以上再生される場合に、エンハンスメントレイヤも含めて高画質の映像を再生してもよい。このようにスケーラブルな符号化が行われているストリームであれば、未選択時又は見始めた段階では粗い動画だが、徐々にストリームがスマートになり画像がよくなるような体験を提供することができる。スケーラブル符号化以外にも、1回目に再生される粗いストリームと、1回目の動画を参照して符号化される2回目のストリームとが1つのストリームとして構成されていても同様の体験を提供できる。

[0443] [その他の実施応用例]

また、これらの符号化又は復号処理は、一般的に各端末が有するLSIex500において処理される。LSI (large scale integration circuitry) ex500 (図37参照) は、ワンチップであっても複数チップからなる構成であってもよい。なお、動画像符号化又は復号用のソフトウェアをコンピュータex111等で読み取り可能な何らかの記録メディア (CD-ROM、フレキシブルディスク、又はハー

ドディスクなど）に組み込み、そのソフトウェアを用いて符号化又は復号処理を行ってもよい。さらに、スマートフォン $e \times 115$ がカメラ付きである場合には、そのカメラで取得した動画データを送信してもよい。このときの動画データはスマートフォン $e \times 115$ が有する $LSI \ e \times 500$ で符号化処理されたデータである。

[0444] なお、 $LSI \ e \times 500$ は、アプリケーションソフトをダウンロードしてアクティベートする構成であってもよい。この場合、端末は、まず、当該端末がコンテンツの符号化方式に対応しているか、又は、特定サービスの実行能力を有するかを判定する。端末がコンテンツの符号化方式に対応していない場合、又は、特定サービスの実行能力を有さない場合、端末は、コーデック又はアプリケーションソフトをダウンロードし、その後、コンテンツ取得及び再生する。

[0445] また、インターネット $e \times 101$ を介したコンテンツ供給システム $e \times 100$ に限らず、デジタル放送用システムにも上記各実施の形態の少なくとも動画像符号化装置（画像符号化装置）又は動画像復号化装置（画像復号装置）のいずれかを組み込むことができる。衛星などを利用して放送用の電波に映像と音が多重化された多重化データを載せて送受信するため、コンテンツ供給システム $e \times 100$ のユニキャストがし易い構成に対してマルチキャスト向きであるという違いがあるが符号化処理及び復号処理に関しては同様の応用が可能である。

[0446] [ハードウェア構成]

図40は、図37に示されたスマートフォン $e \times 115$ のさらに詳細を示す図である。また、図41は、スマートフォン $e \times 115$ の構成例を示す図である。スマートフォン $e \times 115$ は、基地局 $e \times 110$ との間で電波を送受信するためのアンテナ $e \times 450$ と、映像及び静止画を撮ることが可能なカメラ部 $e \times 465$ と、カメラ部 $e \times 465$ で撮像した映像、及びアンテナ $e \times 450$ で受信した映像等が復号されたデータを表示する表示部 $e \times 458$ とを備える。スマートフォン $e \times 115$ は、さらに、タッチパネル等であ

る操作部 e x 4 6 6 と、音声又は音響を出力するためのスピーカ等である音声出力部 e x 4 5 7 と、音声を入力するためのマイク等である音声入力部 e x 4 5 6 と、撮影した映像或いは静止画、録音した音声、受信した映像或いは静止画、メール等の符号化されたデータ、又は、復号化されたデータを保存可能なメモリ部 e x 4 6 7 と、ユーザを特定し、ネットワークをはじめ各種データへのアクセスの認証をするための S I M (S u b s c r i b e r I d e n t i t y M o d u l e) e x 4 6 8 とのインタフェース部であるスロット部 e x 4 6 4 とを備える。なお、メモリ部 e x 4 6 7 の代わりに外付けメモリが用いられてもよい。

[0447] 表示部 e x 4 5 8 及び操作部 e x 4 6 6 等を統括的に制御する主制御部 e x 4 6 0 と、電源回路部 e x 4 6 1、操作入力制御部 e x 4 6 2、映像信号処理部 e x 4 5 5、カメラインタフェース部 e x 4 6 3、ディスプレイ制御部 e x 4 5 9、変調／復調部 e x 4 5 2、多重／分離部 e x 4 5 3、音声信号処理部 e x 4 5 4、スロット部 e x 4 6 4、及びメモリ部 e x 4 6 7 とが同期バス e x 4 7 0 を介して接続されている。

[0448] 電源回路部 e x 4 6 1 は、ユーザの操作により電源キーがオン状態にされると、スマートフォン e x 1 1 5 を動作可能な状態に起動し、バッテリーパックから各部に対して電力を供給する。

[0449] スマートフォン e x 1 1 5 は、C P U、R O M 及び R A M 等を有する主制御部 e x 4 6 0 の制御に基づいて、通話及データ通信等の処理を行う。通話時は、音声入力部 e x 4 5 6 で收音した音声信号を音声信号処理部 e x 4 5 4 でデジタル音声信号に変換し、変調／復調部 e x 4 5 2 でスペクトラム拡散処理を施し、送信／受信部 e x 4 5 1 でデジタルアナログ変換処理及び周波数変換処理を施し、その結果の信号を、アンテナ e x 4 5 0 を介して送信する。また受信データを増幅して周波数変換処理及びアナログデジタル変換処理を施し、変調／復調部 e x 4 5 2 でスペクトラム逆拡散処理し、音声信号処理部 e x 4 5 4 でアナログ音声信号に変換した後、これを音声出力部 e x 4 5 7 から出力する。データ通信モード時は、本体部の操作部 e x 4 6 6

等の操作に基づいてテキスト、静止画、又は映像データが操作入力制御部 $e \times 462$ を介して主制御部 $e \times 460$ に送出される。同様の送受信処理が行われる。データ通信モード時に映像、静止画、又は映像と音声を送信する場合、映像信号処理部 $e \times 455$ は、メモリ部 $e \times 467$ に保存されている映像信号又はカメラ部 $e \times 465$ から入力された映像信号を上記各実施の形態で示した動画像符号化方法によって圧縮符号化し、符号化された映像データを多重／分離部 $e \times 453$ に送出する。音声信号処理部 $e \times 454$ は、映像又は静止画をカメラ部 $e \times 465$ で撮像中に音声入力部 $e \times 456$ で収音した音声信号を符号化し、符号化された音声データを多重／分離部 $e \times 453$ に送出する。多重／分離部 $e \times 453$ は、符号化済み映像データと符号化済み音声データを所定の方式で多重化し、変調／復調部（変調／復調回路部） $e \times 452$ 、及び送信／受信部 $e \times 451$ で変調処理及び変換処理を施してアンテナ $e \times 450$ を介して送信する。

[0450] 電子メール又はチャットに添付された映像、又はウェブページにリンクされた映像を受信した場合等において、アンテナ $e \times 450$ を介して受信された多重化データを復号するために、多重／分離部 $e \times 453$ は、多重化データを分離することにより、多重化データを映像データのビットストリームと音声データのビットストリームとに分け、同期バス $e \times 470$ を介して符号化された映像データを映像信号処理部 $e \times 455$ に供給するとともに、符号化された音声データを音声信号処理部 $e \times 454$ に供給する。映像信号処理部 $e \times 455$ は、上記各実施の形態で示した動画像符号化方法に対応した動画像復号化方法によって映像信号を復号し、ディスプレイ制御部 $e \times 459$ を介して表示部 $e \times 458$ から、リンクされた動画像ファイルに含まれる映像又は静止画が表示される。音声信号処理部 $e \times 454$ は、音声信号を復号し、音声出力部 $e \times 457$ から音声が出力される。リアルタイムストリーミングがますます普及しだしているため、ユーザの状況によっては音声の再生が社会的にふさわしくないこともあり得る。そのため、初期値としては、音声信号は再生せず映像データのみを再生する構成の方が望ましく、ユーザが

映像データをクリックするなど操作を行った場合にのみ音声を同期して再生してもよい。

[0451] またここではスマートフォン `ex 115` を例に説明したが、端末としては符号化器及び復号化器を両方持つ送受信型端末の他に、符号化器のみを有する送信端末、及び、復号化器のみを有する受信端末という3通りの他の実装形式が考えられる。デジタル放送用システムにおいて、映像データに音声データが多重化された多重化データを受信又は送信するとして説明した。ただし、多重化データには、音声データ以外に映像に関連する文字データなどが多重化されてもよい。また、多重化データではなく映像データ自体が受信又は送信されてもよい。

[0452] なお、CPUを含む主制御部 `ex 460` が符号化又は復号処理を制御するとして説明したが、種々の端末はGPU (Graphics Processing Unit) を備えることも多い。よって、CPUとGPUで共通化されたメモリ、又は共通に使用できるようにアドレスが管理されているメモリにより、GPUの性能を活かして広い領域を一括して処理する構成でもよい。これにより符号化時間を短縮でき、リアルタイム性を確保し、低遅延を実現できる。特に動き探索、デブロックフィルタ、SAO (Sample Adaptive Offset)、及び変換・量子化の処理を、CPUではなく、GPUでピクチャなどの単位で一括して行くと効率的である。

産業上の利用可能性

[0453] 本開示は、動画像を符号化する符号化装置等に利用可能であり、ビデオ会議システム等に適用可能である。

符号の説明

[0454] 100 符号化装置
 102 分割部
 104 減算部
 106 変換部
 108 量子化部

110 エントロピー符号化部
112、204 逆量子化部
114、206 逆変換部
116、208 加算部
118、210 ブロックメモリ
120、212 ループフィルタ部
122、214 フレームメモリ
124、216 イントラ予測部
126、218 インター予測部
128、220 予測制御部
130、222 予測パラメータ生成部
131、133、134、142、143 圧縮器
132、232、237 導出器
141 前処理器
151、251 回路
152、252 メモリ
200 復号装置
202 エントロピー復号部
224 分割決定部
231、233、235、241、242 圧縮解除器
234 生成器
236 合成器

請求の範囲

- [請求項1] メモリと、
 前記メモリに接続される回路とを備え、
 前記回路は、前記メモリを用いて、
 顔動画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームから復号し、
 前記顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームから復号し、
 前記ベースデータ、前記1つ以上の拡張データ及び前記幾何学情報から、生成モデルを用いて、前記顔動画像を生成し、
 前記ベースデータは、前記ビットストリームにおいて、前記顔動画像のフレームである第1フレームに対応するデータに付加されており、
 前記1つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画像の1つ以上のフレームであり、前記第1フレームよりも後の1つ以上のフレームである1つ以上の第2フレームに対応する1つ以上のデータに付加されている、
 復号装置。
- [請求項2] 前記回路は、前記ベースデータと前記1つ以上の拡張データとである複数の顔画像データの少なくとも1つの制御に関する制御情報をヘッダ領域から復号する、
 請求項1に記載の復号装置。
- [請求項3] 前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの1つである顔画像データが含まれるか否かを示す存在情報を含む、
 請求項2に記載の復号装置。
- [請求項4] 前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニ

ットに前記複数の顔画像データの1つである顔画像データが含まれる場合、前記顔画像データが前記ベースデータであるか前記1つ以上の拡張データの1つである拡張データであるかに関する種別情報を含み、

前記種別情報は、前記アクセスユニットが前記ベースデータを含む場合、前記アクセスユニットに含まれる前記顔画像データが、前記ベースデータであり、次のベースデータまで継続して用いられることを示し、

前記種別情報は、前記アクセスユニットが前記拡張データを含む場合、前記アクセスユニットに含まれる前記顔画像データが、前記拡張データであり、前記ベースデータと共に用いられることを示す、

請求項2又は3に記載の復号装置。

[請求項5] 前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの1つである顔画像データが含まれる場合、前記顔動画像の複数のフレームのうち前記アクセスユニットに対応するフレームの生成及び表示に前記顔画像データが適用可能であるか否かを示す適用情報を含む、

請求項2又は3に記載の復号装置。

[請求項6] 前記ベースデータと前記1つ以上の拡張データとのそれぞれは、前記顔画像に含まれる顔の特徴を示すベクトルで表現される、

請求項1～3のいずれか1項に記載の復号装置。

[請求項7] 前記ベースデータと前記1つ以上の拡張データとのそれぞれは、前記顔画像に関する画像で表現される、

請求項1～3のいずれか1項に記載の復号装置。

[請求項8] 前記回路は、前記ベースデータと前記1つ以上の拡張データの少なくとも1つと前記幾何学情報とを前記生成モデルに入力して、前記顔動画像のフレームを生成する、

請求項1～3のいずれか1項に記載の復号装置。

- [請求項9] 前記回路は、前記ベースデータと前記1つ以上の拡張データの少なくとも1つとから中間画像を生成し、前記中間画像と前記幾何学情報とを前記生成モデルに入力して、前記顔動画像のフレームを生成する、
- 請求項1～3のいずれか1項に記載の復号装置。
- [請求項10] 前記回路は、前記ベースデータを参照して、前記1つ以上の拡張データの1つである拡張データを復号し、前記拡張データと前記幾何学情報とを前記生成モデルに入力して、前記顔動画像のフレームを生成する、
- 請求項1～3のいずれか1項に記載の復号装置。
- [請求項11] 前記ベースデータは、前記顔画像に含まれる顔の一部のデータであり、
- 前記1つ以上の拡張データの1つである拡張データは、前記顔画像に含まれる前記顔の他部のデータである、
- 請求項1～3のいずれか1項に記載の復号装置。
- [請求項12] 前記ベースデータは、前記顔画像における第1周波数範囲のデータであり、
- 前記1つ以上の拡張データの1つである拡張データは、前記顔画像における前記第1周波数範囲よりも高い第2周波数範囲のデータである、
- 請求項1～3のいずれか1項に記載の復号装置。
- [請求項13] 前記ベースデータは、前記顔画像に関する画像であり、第1解像度を有する画像である第1画像に対応し、
- 前記1つ以上の拡張データの1つである拡張データは、前記顔画像に関する画像であり、前記第1画像を参照して復号される画像であり、前記第1解像度よりも高い第2解像度を有する画像である第2画像に対応する、
- 請求項1～3のいずれか1項に記載の復号装置。

- [請求項14] 前記ベースデータは、前記顔画像に関する画像であり、第1量子化幅で復号される画像である第1画像に対応し、
前記1つ以上の拡張データの1つである拡張データは、前記顔画像に関する画像であり、前記第1画像を参照して前記第1量子化幅よりも細かい第2量子化幅で復号される画像である第2画像に対応する、
請求項1～3のいずれか1項に記載の復号装置。
- [請求項15] 前記制御情報は、前記1つ以上の拡張データのそれぞれを識別するための識別情報を含む、
請求項2又は3に記載の復号装置。
- [請求項16] 前記制御情報は、(i)前記ベースデータを含むアクセスユニットの前記ヘッダ領域に含まれる情報であり、(ii)前記1つ以上の拡張データの総数を示す情報である総数情報を含む、
請求項2又は3に記載の復号装置。
- [請求項17] 前記制御情報は、(i)前記ベースデータを含むアクセスユニットの前記ヘッダ領域に含まれる情報であり、(ii)前記1つ以上の拡張データのうちの拡張データであって、前記1つ以上の第2フレームのうち当該拡張データを含むアクセスユニットに対応する第2フレームの生成及び表示に適用可能な拡張データを特定するための情報である特定情報を含む、
請求項2又は3に記載の復号装置。
- [請求項18] 前記回路は、前記メモリにおいて前記ビットストリームが記憶されるストリームバッファを制御するための少なくとも1つの制御パラメータであって、前記ストリームバッファのバッファサイズを基準サイズ以下に制御し、復号処理開始の初期遅延時間を基準遅延時間以下に制御するための少なくとも1つの制御パラメータを復号する、
請求項1～3のいずれか1項に記載の復号装置。
- [請求項19] メモリと、
前記メモリに接続される回路とを備え、

前記回路は、前記メモリを用いて、

顔動画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームに符号化し、

前記顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームに符号化し、

前記ベースデータは、前記ビットストリームにおいて、前記顔動画像のフレームである第1フレームに対応するデータに付加され、

前記1つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画像の1つ以上のフレームであり、前記第1フレームよりも後の1つ以上のフレームである1つ以上の第2フレームに対応する1つ以上のデータに付加される、

符号化装置。

[請求項20] 前記回路は、前記ベースデータと前記1つ以上の拡張データとである複数の顔画像データの少なくとも1つの制御に関する制御情報をヘッダ領域に符号化する、

請求項19に記載の符号化装置。

[請求項21] 前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの1つである顔画像データが含まれるか否かを示す存在情報を含む、

請求項20に記載の符号化装置。

[請求項22] 前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの1つである顔画像データが含まれる場合、前記顔画像データが前記ベースデータであるか前記1つ以上の拡張データの1つである拡張データであるかに関する種別情報を含み、

前記種別情報は、前記アクセスユニットが前記ベースデータを含む場合、前記アクセスユニットに含まれる前記顔画像データが、前記ベ

ースデータであり、次のベースデータまで継続して用いられることを示し、

前記種別情報は、前記アクセスユニットが前記拡張データを含む場合、前記アクセスユニットに含まれる前記顔画像データが、前記拡張データであり、前記ベースデータと共に用いられることを示す、

請求項 20 又は 21 に記載の符号化装置。

[請求項23] 前記制御情報は、前記ヘッダ領域によって制御されるアクセスユニットに前記複数の顔画像データの 1 つである顔画像データが含まれる場合、前記顔動画画像の複数のフレームのうち前記アクセスユニットに対応するフレームの生成及び表示に前記顔画像データが適用可能であるか否かを示す適用情報を含む、

請求項 20 又は 21 に記載の符号化装置。

[請求項24] 前記ベースデータと前記 1 つ以上の拡張データとのそれぞれは、前記顔画像に含まれる顔の特徴を示すベクトルで表現される、

請求項 19 ～ 21 のいずれか 1 項に記載の符号化装置。

[請求項25] 前記ベースデータと前記 1 つ以上の拡張データとのそれぞれは、前記顔画像に関する画像で表現される、

請求項 19 ～ 21 のいずれか 1 項に記載の符号化装置。

[請求項26] 前記回路は、前記顔画像に含まれる顔の一部のデータを前記ベースデータとして導出して符号化し、前記顔画像に含まれる前記顔の他部のデータを前記 1 つ以上の拡張データの 1 つである拡張データとして導出して符号化する、

請求項 19 ～ 21 のいずれか 1 項に記載の符号化装置。

[請求項27] 前記回路は、前記顔画像における第 1 周波数範囲のデータを前記ベースデータとして導出して符号化し、前記顔画像における前記第 1 周波数範囲よりも高い第 2 周波数範囲のデータを前記 1 つ以上の拡張データの 1 つである拡張データとして導出して符号化する、

請求項 19 ～ 21 のいずれか 1 項に記載の符号化装置。

[請求項28] 前記回路は、前記顔画像に関する画像であり、第1解像度を有する画像である第1画像を前記ベースデータとして符号化し、前記顔画像に関する画像であり、前記第1画像を参照して符号化される画像であり、前記第1解像度よりも高い第2解像度を有する画像である第2画像を前記ベースデータとして符号化する、

請求項19～21のいずれか1項に記載の符号化装置。

[請求項29] 顔動画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームから復号し、

前記顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームから復号し、

前記ベースデータ、前記1つ以上の拡張データ及び前記幾何学情報から、生成モデルを用いて、前記顔動画像を生成し、

前記ベースデータは、前記ビットストリームにおいて、前記顔動画像のフレームである第1フレームに対応するデータに付加されており、

前記1つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画像の1つ以上のフレームであり、前記第1フレームよりも後の1つ以上のフレームである1つ以上の第2フレームに対応する1つ以上のデータに付加されている、

復号方法。

[請求項30] 顔動画像に関する顔画像のベースデータと、前記顔画像の1つ以上の拡張データとをビットストリームに符号化し、

前記顔動画像の複数のフレームのそれぞれに対応する情報であり、人物の顔を含む領域内の幾何学属性を示す情報である幾何学情報を前記ビットストリームに符号化し、

前記ベースデータは、前記ビットストリームにおいて、前記顔動画像のフレームである第1フレームに対応するデータに付加され、

前記 1 つ以上の拡張データは、前記ビットストリームにおいて、前記顔動画像の 1 つ以上のフレームであり、前記第 1 フレームよりも後の 1 つ以上のフレームである 1 つ以上の第 2 フレームに対応する 1 つ以上のデータに付加される、
符号化方法。